

**A WATER QUALITY ASSESSMENT MODEL FOR THE CROCODILE RIVER AND
INYAKA DAM**

by

LAZARUS KATLEGO MOGANE

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

in the subject

Life Sciences

at the

UNIVERSITY OF SOUTH AFRICA

FLORIDA CAMPUS

Supervisors: Prof. Tracy M. Masebe (University of South Africa)

Co-Supervisors: Prof. Titus A.M. Msagati (University of South Africa)

Dr. Esper J. Ncube (University of Pretoria/Rand Water)

FEBRUARY 2026

“Those who complete the course will do so only because they do not, as fatigue sets in, convince themselves that the road ahead is still too long, the inclines too steep, the loneliness impossible to bear, and the prize itself, of doubtful value.”

Thabo Mvuyelwa Mbeki

(24 September 2008)

DEDICATIONS

This research work is dedicated to my late grandmother **Lesi Ruth Mogane**, my first friend and motherlike nurturer and the progenitor of those who raised me to responsible age, **Griedah Taelo** and **Attie Christinah Mogane**. First among equals, I would like to give recognition to the role played by **Griedah Taelo Mogane**, for being a mother to me when she had to and raising me with all her love.

The inspiration of this work has been and continues to be centred around my son **Bokang Mogane Mogane**, for when my vision blurs, I would love for him to sit on my shoulders and see further.

I also dedicate this thesis to my collateral siblings, **Dr. Constance Mafuwane, Walter, Thabang, Stanley, and Fortunate aka Chido**, for being the best encouragement towards our dreams. I especially dedicate the thesis to **Walter Segage** for his selflessness, guidance, and patience throughout my life.

The thesis is also dedicated to all my available friends, the best cheerleaders of my journey. First among equals, **Motlalepula Simango**; for being there in my lowest of moments and providing refuge and advice; **Dr Rifilwe Modiba** for going all out to ensure my mental health with all the brotherly talks and prayers; **Sipho Ntimane** for your persistence in maintaining contact with consistent encouragement, I appreciate you; my dearest friend **Nobethu Masanabo** for always lightening my day with positive talks and silly jokes; my work colleagues who always appreciated my journey and contributed positively, **Mpho Adivhaho Makhavhu, Rirhandzu Ntusi, Lebogang Mdluli, and Thembi Khoza**.

And finally, this study would not have been without The Creator of the ends of the Earth, The Infinite Energy of Isaiah 40:28-31.

DECLARATION

Name: Lazarus Katlego Mogane

Student number: 61804649

Degree: Doctor of Philosophy in Life Sciences

Dissertation title:

A WATER QUALITY ASSESSMENT MODEL FOR THE CROCODILE RIVER AND INYAKA DAM

I declare that the above thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

In addition, I declare that I submitted the thesis to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any other higher education institution.



Signature

February 2026

Date

ACKNOWLEDGEMENTS

I want to express my sincere appreciation to my supervisor, mentor and support system, Prof. Tracy Masebe, for always being encouraging on a personal and professional level, her technical advice, and for managing the whole project.

I would also like to express my heartfelt gratitude to Dr Esper Ncube's best-laid foundations of the project, her technical inputs in the respective chapters and for her strong words of encouragement throughout the project.

Prof Titus Msagati contributed immensely to the chapters through his specialist technical knowledge and sincere talks.

Would like to express many thanks to Khomotso Chokoe for her warm welcome and hospitality in her laboratories.

Special thank you to the Bushbuckridge Local Municipality for allowing me to use their laboratory for some sections of the project.

The IUCMA deserve a mention, especially Thabo Rasiuba for all his work towards the study permit applications.

KEY TERMS DESCRIBING THE TOPIC OF THE DISSERTATION

TITLE OF THESIS:

A WATER QUALITY ASSESSMENT MODEL FOR THE CROCODILE RIVER AND INYAKA DAM

KEY TERMS:

Crocodile River: A river originating from Dullstroom in the Mpumalanga province, passing through Mbombela and Kruger National Park, and confluences the Komati River

Inyaka Dam: Is an earth-filled dam located in the Bushbuckridge area, Mpumalanga province built on the Marite River and primarily used for irrigation and potable water supply.

Water quality: The measure of water's suitability for specific use determined by its physicochemical and biological characteristics.

Water quality assessment: The evaluation of water quality against set standards for different uses.

Aquatic ecosystems: A water-based environment that provides habitat for animals, plants, microbes, and is divided into fresh and marine waters.

Lotic: A type of freshwater ecosystem including flowing aquatic systems such as rivers and streams.

Lentic: A type of freshwater ecosystem including still aquatic systems such as dams, lakes, ponds, etc.

Water quality indicator: A measurable physical, chemical, or biological characteristic that can demonstrate the health of an aquatic system.

Physicochemical: Characteristics of water that include the physical and chemical properties or content.

Microbiological: Characteristics of water that include the microbiological properties or content.

Summer: In the context of South Africa, a season characterised by warm weather and high rainfall, and longer daylight.

Winter: A season characterised by cool weather, minimal rainfall, and shorter daylight.

Sequencing: A systematic way of determining the order of the nucleotide bases in DNA to understand the genetics, compare species, understand community dynamics, etc.

16S metagenomics: A DNA sequencing method based on the 16Svedberg units (16S) gene to profile microbial community composition.

Bacterial community dynamics: Types, numbers, and interactions of bacteria within a mixed matrix, such as water, over time, and driven by environmental factors and nutrients.

Artificial neural networks: A non-linear signal processing model inspired by the biological neural process.

Modelling: The usage of mathematical models to perform calculations or make predictions.

Forecasting: Making predictions based on historical and present data.

PUBLISHED ARTICLES

Mogane, L.K., Masebe, T., Msagati, T.A.M. Ncube E. A comprehensive review of water quality indices for lotic and lentic ecosystems. *Environ Monit Assess* **195**, 926 (2023). <https://doi.org/10.1007/s10661-023-11512-2>

Lazarus Katlego Mogane, Esper J. Ncube, Titus A.M. Msagati, Tracy M. Masebe, Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa, *Heliyon*, Volume 10, Issue 19, 2024, e38246. <https://doi.org/10.1016/j.heliyon.2024.e38246>

TABLE OF CONTENTS

DEDICATIONS.....	ii
DECLARATION	iii
ACKNOWLEDGEMENTS	iv
KEY TERMS DESCRIBING THE TOPIC OF THE DISSERTATION	v
PUBLISHED ARTICLES	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
EXECUTIVE SUMMARY	xvii
1. Introduction.....	1
1.1. Background.....	1
1.2. Water Quality requirements versus South African water quality guidelines	2
1.3. Water resource quality objectives and targets.....	3
1.3.1. The Crocodile River.....	3
1.3.2. The Inyaka Dam.....	4
1.4. Microbial ecology of freshwater resources	5
1.5. Major bacterial lineages in freshwater systems	6
1.5.1. Actinobacteria	6
1.5.2. Bacteroidota	7
1.5.3. Proteobacteria	7
1.6. Freshwater ecosystems: temporal and spatial variations of microbial communities	8
1.7. Water Quality Index (WQI)	8
1.8. Microbial indicators of water quality	11
1.9. Selected physicochemical parameters.....	12
1.9.1. Temperature	12
1.9.2. Turbidity	13
1.9.3. Electrical Conductivity	13
1.9.4. pH.....	14
1.9.5. Iron.....	14
1.9.6. Manganese	15
1.9.7. Nitrate-nitrogen.....	15

1.9.8. Phosphate	16
1.10. Research problem statement.	16
1.11. Study rationale.	17
1.12. Objectives of the study, Hypotheses and Research Questions.....	18
1.12.1. Overall objective of the study	18
1.13. Specific Objectives, Hypotheses and Research Questions	18
1.14. Study assumptions.	19
1.15. Delimitations of the study	19
1.16. Limitations of the study	19
1.17. Conceptual framework.....	20
1.18. Thesis structure	21
1.19. References.....	25
2. Literature Review.....	39
2.1. Preamble	39
2.2. A comprehensive review of water quality indices for lotic and lentic ecosystems.....	40
2.3. Abstract.....	40
2.4. Background.....	41
2.6. Method and approach of review.....	43
2.7. Research questions, inclusion and exclusion criteria for articles, and criteria for articles and models.....	44
2.8. Articles search strategy.	44
2.9. The typical steps for developing a WQI.	53
i. Parameter Selection	55
ii. Transformation of parameters to a standard scale.....	57
iii. Establishing parameter weights	59
iv. Aggregation of transformed parameters to produce the final index.....	60
Categories of aggregation methods.....	61
i. Additive	61
ii. Multiplicative.....	62
iii. Logical	62
i. WQI Classification Schemes	63
2.10. Selected common WQI used for evaluation of lotic and lentic ecosystems.	65
2.10.1. Canadian Council of Ministers of the Environment Index (CCME) index.....	65
2.10.2. Oregon WQI (OWQI)	67
2.10.3. National Sanitation Foundation (NSF) WQI.....	69
2.10.3. The Scottish Research Development Department (SRDD) Index	70
2.10.4. House Index	71

2.11. Critique of WQI	73
2.11.1. Issues involved in WQI development.	73
i. Parameter selection	73
ii. Lack of universal WQI.....	73
iii. WQI ambiguity and eclipsing.	74
iv. Uncertainty issues.	74
v. The Metaphoring classification schemes problem.	75
2.12. Conclusions.....	76
2.13. Future directions	77
2.14. References.....	79
3. Methodology	91
3.1. Introduction.....	91
3.2. Study area Description.....	91
3.3. Data collection and processing	93
3.4. Experimental methods and data analysis	94
3.4.1. Selected microbial parameters.	94
3.4.2. Selected physicochemical parameters.....	95
3.5. Statistical analysis of physical, chemical, and microbiological data	96
3.6. DNA Extraction	96
3.7. DNA Sequencing	97
3.8. Statistical and Bioinformatic Analysis of Microbial Ecology Data.....	97
3.9. Model development workflow.....	98
3.10 Ethical Considerations	99
3.11 References.....	100
4. Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa	101
4.1. Preamble	101
4.2. Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa.	102
4.3. Abstract.....	102
4.4. Background.....	103
4.5. Materials and methods	104
4.5.1. Study area description and land use information.	104
4.5.2. Sample collection and handling.	106
4.5.3. Experimental methods.....	106
4.5.4. Statistical analysis	107
4.6. Results and Discussion	109

4.6.1. Water quality properties of the Inyaka Dam	109
4.6.2. Multivariate Statistical Analysis	115
4.6.3. Water quality properties of the Crocodile River	119
4.6.4. Multivariate Statistical Analysis	124
4.7. Conclusions.....	130
4.8. Future directions	130
4.9. CRediT author statement.	131
4.10. Statements and declarations	131
4.11. References.....	132
5. Exploring the physicochemical drivers of bacterial community dynamics in lotic and lentic freshwater ecosystems	136
5.1. Preamble	136
5.2. Abstract.....	137
5.3. Background.....	138
5.4. Materials and Methods.....	140
5.4.1. Study area.....	140
5.4.2. Water sample collection and handling	142
5.4.3. Experimental methods.....	142
5.5. Results and Discussion	143
5.5.1. Physicochemical properties of surface water in the Inyaka Dam and Crocodile River	143
5.5.2. Dynamics of Bacterial Community Structure in Relation to Water Quality Variations ...	145
5.6. Conclusions.....	160
5.7. Statements and Declarations	160
5.8. Appendices.....	161
5.9. References.....	162
6. Development of a water quality prediction tool using an Artificial Neural Networks-based Machine Learning Algorithm	169
6.1. Preamble	169
6.2. Abstract.....	170
6.3. Background.....	171
6.4. Materials and methods	173
6.4.1. Study site.....	173
6.4.2. Data generation from water quality parameters	174
6.4.3. Data Standardisation	174
6.4.4. Data Preparation, Training, Validation, and Testing	175
6.4.5. ANN Model network architecture.....	175
6.5. Validity and Reliability.....	181

6.6. Results and Discussion	183
6.6.1. Model training and performance evaluation for physical parameters.....	183
6.6.2. Model training and performance evaluation for chemical parameters.....	188
6.6.3. Model training and performance evaluation for microbiological parameters.	192
6.6.4. Model testing and performance appraisal.	193
6.6.5. Inherent aspects that influence model performance	196
6.6.6. Limitations of the developed models.	198
6.7. Conclusions and prospects	198
6.8. Statements and Declarations	199
6.9. References.....	200
7: Synthesis of chapter findings, Conclusions and Recommendations.....	206
7.1. Preamble	206
7.2. Synthesis of chapter findings	207
7.2.1. Original Contribution to Knowledge	209
7.3. Conclusions and Recommendations	210
Appendices.....	214
Appendix 1: University of South Africa Research Ethics Clearance Letter (2020/CAES_HREC/047).....	214
Appendix 2: Gatekeeper letter from Inkomati-Usuthu Catchment Management Agency (IUCMA) to conduct research on catchments under IUCMA	217
Appendix 3: Gatekeeper letter from Bushbuckridge Local Municipality to conduct research experiments in the municipality's premises.....	218
Appendix 4: Chapter 4 supplementary tables and figures	219
Appendix 5: Chapter 6 Tables	227

LIST OF TABLES

Table 1.1: Water quality objectives for the Crocodile River (DWS, 2016)	4
Table 1.2: Target water quality guidelines for the Sabie catchment to which the Inyaka dam belongs (DWS, 2016).....	5
Table 1.3: Commonly used functions to compute WQI.	10
Table 1.4: Common microbial indicators of water quality	12
Table 2.1: Overview of included studies, appraisal of the stages of development, system of application, and where and when studies were first published.....	46
Table 4.1: Water quality parameters mean measurements and standard deviations for all observed seasons.....	109
Table 4.2: Mean and standard deviation values of the physicochemical and microbiological properties of the Crocodile River for four seasons.	119
Table 5.1: Characteristics of the observed major phyla in relation to physicochemical properties of water.....	148
Table 6.1: Model training specifications and training evaluation results for pH prediction..	184
Table 6.2: Model training specifications and training evaluation results for phosphate prediction	188
Table 6.3: Model training specifications and training evaluation results for <i>E. coli</i> prediction	192
Table S4.1: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in summer.....	223
Table S4.2: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in autumn.....	223
Table S4.3: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in winter.....	224
Table S4.4: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in spring.....	224
Table S4.5: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in summer.....	225
Table S4.6: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in autumn.....	225
Table S4.7: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in winter.....	226
Table S4.8: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in spring.....	226

LIST OF FIGURES

Fig. 1.1 Conceptual framework of the study	20
Fig. 2.1 PRISMA flow diagram of searching, screening, and article selection.....	45
Fig. 2.2 The general WQI development steps.....	52
Fig. 2.3 Summary of the number of parameters used per WQI included in the study	56
Fig. 2.4 Percentage of the frequency of use for individual water quality parameters in all selected WQIs	57
Fig. 2.5 Usage of parameter transformation method among the analysed WQIs.....	59
Fig. 2.6 Aggregation category usage among the analysed WQIs.....	61
Fig. 2.7 Different classification schemes used among the WQI models to explain the “metaphoring problem” contributing to WQI uncertainty.....	64
Fig. 3.1 Map of the Crocodile River sampling areas and land use activities.....	90
Fig. 3.2 Map of the Inyaka Dam and surrounding land use activities.....	91
Fig. 3.3 Model development workflow involving the number of parameters used as input variables, design of the model architecture, data sorting, training, testing and model performance evaluation.....	96
Fig. 4.1 Maps of the study area, and sampling sites (Esri: https://www.arcgis.com). Fig. 4.1 frame A shows the Inyaka Dam, and frame B shows the Crocodile River.	105
Fig. 4.2 Variations in the collective physical parameter throughout the different seasons of the Inyaka Dam.....	111
Fig. 4.3 Collective chemical parameter variations during the observed seasons in the Inyaka Dam.....	113
Fig. 4.4 Comparison of the variations in the collective microbiological parameters during the different seasons in the Inyaka Dam.....	114
Fig. 4.5 Correlation matrix for all measured water quality parameters during summer at Inyaka Dam.....	116
Fig. 4.6 A comparison of 3D PCA plots for analysed parameters in the autumn and spring seasons for the Inyaka Dam	118
Fig. 4.7 Comparison of the variations in the collective physical parameters during different seasons in the Crocodile River.....	121
Fig. 4.8 Comparison of the variations in the collective microbiological parameters during the observed seasons in the Crocodile River	124
Fig. 4.9 Correlation matrix for all observed water quality parameters in summer in the Crocodile River.....	125
Fig. 4.10 PCA plots for analysed parameters for the summer and winter seasons in the Crocodile River.....	127
Fig. 4.11 PCA plots for analysed parameters for the summer and winter seasons in the Crocodile River.....	128
Fig. 5.1 Study area maps of the Inyaka Dam (A) and the Crocodile River (B). Both frames illustrate the position of the sampling points SS-25 and C-40 along with associated tributaries and treatment works around the sites.....	138
Fig. 5.2 Comparison of seasonal physicochemical parameters in the Inyaka Dam collected during winter and summer. The chart shows key water quality indicators: Electrical conductivity (EC), Iron (Fe), pH, Phosphate (PO_4^{3-}), Nitrate-Nitrogen ($\text{NO}_3\text{-N}$), Manganese	

(Mn), and Temperature plotted in seasonally segregated bars to enable distinction of seasonal variation.....142

Fig. 5.3 Comparative seasonal analysis of physicochemical parameters in the Crocodile River The chart shows key water quality indicators: Electrical conductivity (EC), Iron (Fe), pH, Phosphate (PO_4^{3-}), Nitrate-Nitrogen ($\text{NO}_3\text{-N}$), Manganese (Mn), and Temperature plotted in seasonally segregated bars to enable distinction of temporal variation.....143

Fig. 5.4 Comparative relative abundance according to microbial phyla in the Crocodile River and Inyaka Dam, illustrating variation in the microbial community composition at the phylum level, with *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, and *Chloroflexi* being prominent in both aquatic habitats.....144

Fig. 5.5 Seasonal shifts in relative abundance of dominant taxa at a phylum level in the Crocodile River and Inyaka Dam. Stacked bar charts illustrate the relative abundance of major freshwater bacterial phyla from samples collected in summer and winter at both sites, while the seasonal comparisons illustrate temporal compositional shifts related to environmental variability.....146

Fig. 5.6 Sunburst charts illustrating the relative abundance of bacterial communities in the Crocodile River during winter (left) and summer (right). Both charts display a hierarchical classification from domain to phylum and class taxa. *Proteobacteria* is the dominant phylum across all seasons, with Gammaproteobacteria being the most abundant in both winter (32%) and summer (59%).....150

Fig. 5.7 Sunburst diagrams illustrating seasonal variations in the relative abundance of bacterial class compositions in the Inyaka Dam during winter (left) and summer (right). Both charts represent a classification depth of up to class level, with the percentages denoting the relative abundance of each class and highlighting observable bacterial community shifts between seasons.....152

Fig. 5.8 Comparison of alpha diversity indices ACE, Chao1 and Jackknife across the Crocodile River and Inyaka Dam. The box plots illustrate the distribution of the observed values for the richness indices in the aquatic ecosystems. Each richness index has abundance categories: High (red), Mid (blue), and Low (green). 154

Fig. 5.9 Comparison of alpha diversity indices between the Crocodile River and Inyaka Dam The box plots represent the Shannon diversity index (left) and the Simpson diversity index (right) categorised by bacterial community abundance groupings: High (brown), Mid (blue), and Low (green).....156

Fig 5.10 Beta diversity plot of Principal coordinate analysis (PCoA) based on Jensen-Shannon divergence, inclusive of unclassified OTUs. The PCoA plot reveals distinct clustering of microbial communities from the Crocodile River (green) and the Inyaka Dam (blue). The first and second principal coordinates explain a total variation of 86.20% with each contributing 71.027% and 15.739% respectively.....157

Fig. 6.1 Water quality prediction model development framework..... 173

Fig. 6.2 Map of the study area in the Crocodile River situated within the northeastern parts of South Africa..... 174

Fig. 6.3 One output fully connected MLP-ANN which can be employed by the developed model for selecting of one desired output parameter 176

Fig. 6.4 Basic multiple outputs fully connected feedforward MLP-ANN with error backpropagation. The number of nodes in the hidden layer increased as the number of output nodes	177
Fig. 6.5 Model validation procedure to ensure the reliability of the selected model	182
Fig. 6.6 Illustration of the successful cost functions for pH prediction training. The graphs denote the closeness of the validation and training experiments, the closer the graphs are to each other, the more reliable the trained model.	185
Fig. 6.7 MSLE cost functions for the training and validation of the turbidity prediction model.	186
Fig. 6.8 MSLE cost functions for temperature prediction model training and validation.	187
Fig. 6.9 Loss functions for phosphate prediction model training and validation show similar learning patterns exhibited by different models.	189
Fig. 6.10 Illustration of loss function results from training of the nitrate-nitrogen prediction model.	190
Fig. 6.11 Illustration of the loss function results from the training of the Iron prediction model.	190
Fig. 6.12 Loss function results from training the Manganese prediction model.	191
Fig. 6.13 Model prediction testing time series graphs for all physical parameters under study.	194
Fig. 6.14 Model prediction testing graphs for all chemical parameters under study.	195
Fig. 6.15 Model prediction testing graphs for all microbiological parameters under study. Parts A, B, C, and D, specifically demonstrate the prediction of known data points prior to future predictions.	196
Fig. S4.1 Illegal waste disposal site near the Crocodile River in the City of Mbombela (Photo by LK Mogane, ...	219
Fig. S4.2 The Maviljan wastewater treatment plant in Bushbuckridge, Mpumalanga, which discharges its effluent in the Inyaka Dam (Photo by LK Mogane, 2021).	219
Fig. S4.3 Correlation matrix for all measured parameters in the autumn season in the Inyaka Dam.	220
Fig. S4.4 Correlation matrix for all measured parameters in the winter season in the Inyaka Dam.	220
Fig. S4.5 Correlation matrix for all measured parameters in the spring season in the Inyaka Dam.	221
Fig. S4.6 Correlation matrix for all measured parameters in the autumn season in the Crocodile River.	221
Fig. S4.7 Correlation matrix for all measured parameters in the spring season in the Crocodile River.	222
Fig. S4.8 Correlation matrix for all measured parameters in the winter season in the Crocodile River.	222

EXECUTIVE SUMMARY

“Water, water everywhere, nor any drop to drink”. These are the famous words from Samuel Taylor Coleridge’s poem, “The Rime of the Ancient Mariner,” penned in the late 18th century, which highlight the lived realities of many South Africans. There is widespread consensus that South Africa is a water-stressed country with an average annual rainfall of only 464 mm, compared to the global average of 860 mm/year. Access to clean and safe water has become a luxury, especially in the rural areas of South Africa, where municipal services are either inadequate or non-existent. As such, these rural households rely on unsafe sources such as rivers and lakes. This study was motivated by the need to monitor the quality of surface water used by rural communities with limited municipal services. However, since the water quality monitoring process has financial implications, a more proactive and financially viable solution needed to be sought. Here, we present a desktop-based system that employs historical and ongoing data to predict water quality and provide decision-making support to officials on how and when to act during potential pollution events.

This study was conducted within the context of the Mpumalanga province in the Ehlanzeni District, where there are notable potable water shortages, especially in the rural local municipalities. The study aimed to develop a water quality assessment model for the Crocodile River and Inyaka Dam. This was achieved by addressing the following objectives: (1) To investigate the microbial and physicochemical water quality characteristics of the Crocodile River and Inyaka Dam across all four seasons; (2) To investigate the bacterial community structure and the dominant bacterial taxa in the Crocodile River and Inyaka Dam across two seasons, and (3) To develop a water quality assessment prototype for the Crocodile River and the Inyaka Dam.

To achieve these objectives, a quantitative method approach was employed. Data obtained from water samples were analysed as per the objectives using parametric and multivariate statistical analyses, such as Pearson correlation, bar charts, and principal component analysis, using OriginPro and SPSS. To achieve the second objective, 16S metagenomic sequencing was conducted using the PacBio platform. All bioinformatic analyses were conducted on the web-based QIIME 2 platform, while R was used to generate alpha and beta diversity graphs. Lastly, the water quality assessment model was coded and programmed on the Python platform and launched on Microsoft Windows-based computers.

The key findings of this study were as follows:

Literature exploration revealed that existing models of water quality assessment were vulnerable to distortions, such as eclipsing and ambiguity, which increased their uncertainty at various stages of model development and during water quality assessments. In this regard, this study recommended the incorporation of statistical methods and machine learning techniques into the development of these models to reduce uncertainty.

The investigation of microbial and physicochemical characteristics of water quality in the Crocodile River and Inyaka Dam revealed that water pollution in these aquatic systems stems from both natural processes, such as stratification, and industrial activities like wastewater discharge and agriculture. The water quality in the Inyaka Dam is highly influenced by temperature stratification, which causes degradation in winter when surface temperature decreases. In contrast, the Crocodile River is largely influenced by changes in the flow regime between the observed seasons.

A further exploration of the influence of physicochemical properties on the composition of bacterial communities revealed that certain microbial communities remain relatively stable despite seasonal fluctuations. Key physicochemical drivers in the Crocodile River, such as flow rate, nutrient concentrations of PO_4^{3-} , $\text{NO}_3\text{-N}$, Mn, Fe, and EC, influenced microbial distribution and diversity, irrespective of the season, favouring a more diverse and spatially variable microbial assemblage. On the other hand, the Inyaka Dam supported a more stable community structure across all seasons, largely due to water column stability and nutrient stratification. Although there were dissimilarities between the Crocodile River and Inyaka Dam, the study revealed that the phylum *Proteobacteria* dominated both aquatic systems across all observed seasons. Moreover, *Gammaproteobacteria* were more abundant across all seasons in both ecosystems, suggesting *Proteobacteria's* ubiquity and adaptation capabilities. These findings increase our understanding of the microbial biogeography, especially in these ecological habitats, and they also provide grounds for the development of an ecosystem-specific microbial metrics system for water quality management.

This study successfully developed a predictive tool using an Artificial Neural Networks (ANNs) to forecast water quality parameters, demonstrating the potential of machine learning in improving water resource management. Despite the promising results, its effectiveness is constrained by the availability and diversity of data. The tool represents a significant step

forward in the prediction of water quality, offering a more efficient and potentially more accurate alternative to traditional methods in decision support. However, to fully realise its potential, more research is needed to address its limitations, particularly to expand the data set and refine the model to improve its accuracy and applicability in various scenarios.

Overall, the principal contribution of this study lies in the integration of four dimensions of water quality assessment often considered separately. These include assessment methodology, environmental variability, microbial ecology, and predictive modelling. The contribution is therefore not the production of an ANN program but the development of an integrated evidence base for context-sensitive and predictive water quality assessment.

Keywords: Water quality, Crocodile River, Inyaka Dam, Model, Physicochemical parameters, Microbial communities

1. Introduction

1.1. Background

Water forms part of the basic needs to sustain life for all living organisms, thus the necessity to maintain this resource so that it is utilised sustainably and appropriately (Juwana et al., 2014; Musie & Gonfa, 2023). For this reason, water quality management and environmental protection have become the pillars of water quality management policy frameworks worldwide (Sapkal and Valunjkar, 2013; Ejiohuo et al., 2025). In South Africa, water quality management schemes are very important due to water scarcity and unpredictable rainfall patterns (Pitman, 2011; Soyapi, 2017; Nhemachena et al., 2020). On average, South Africa receives an annual rainfall of 465mm compared with the annual world average of 860mm (Pitman, 2011). According to Van Rooyen et al. (2010), the unreliability and shortage of rain have been attributed to the dominance of high-pressure systems, which make it unfavourable for rain to form. Furthermore, although high evaporation rates have been observed in the eastern and southern regions of Southern Africa, 65% of South Africa receives less than 500mm of rain per annum. In addition to these hydrological conditions, there is still the issue of an ever-increasing demand for adequate, sustainable, safe, and clean freshwater supply sources due to their decreasing trend (Walter et al., 2011; du Plessis, 2019). Moreover, the deterioration of surface water quality has been aggravated due to the contamination from discharges of sub-standard effluents, emanating from municipal channels, industrial and agricultural activities, as well as anthropogenic activities. Due to the contribution of anthropogenic water contamination problems, the trend and extent of water quality differ significantly in various parts of lotic and lentic systems (Gavrilaş et al., 2025). This implies that there is a serious need to ensure that surface water undergoes appropriate treatment processes before it is distributed for human consumption (Akbar, 2013). Moreover, maintaining good water quality is fundamental to the protection of aquatic life and biodiversity (Srebotnjak et al., 2011; Ejiohuo et al., 2025). Consequently, the significance of ensuring water quality for ecological and human health is of paramount importance, and this can be monitored by water quality indices (WQIs), which normally employ modelling approaches to define and simplify water quality data (Srebotnjak et al., 2012; Chidiac et al., 2023)

The concepts regarding WQI have been around since the mid-twentieth century, when initially introduced by Horton (1965), and since then, there has been widespread use of WQIs to manage

water quality in marine and freshwater resources. The evidence of this widespread usage of WQIs lies in the number of indices produced over the past years (Cox, 2003; Wang et al., 2013; Liangliang and Daoliang, 2014; Sutadian et al., 2018; Uddin et al., 2023). However, the recent decades have witnessed an evolution in the water quality field with the rise of computational statistics. Several computational models, such as MINITAB, MATLAB, R Statistical package, etc, have been developed and widely applied in water quality management programmes and have thus far given authenticity to water quality indices (Sravya and Sreejani, 2016; Lokman et al., 2025).

1.2. Water Quality requirements versus South African water quality guidelines

The prime aim of having water quality standards is to ensure and protect water quality from pollution problems for the betterment of public health and welfare (Ejiohuo et al., 2025). This is in line with public interests in drinking water supplies to maintain and improve the biological integrity of water. The World Health Organisation (WHO) has set an international reference for water quality requirements, even though individual countries have the prerogative to develop their own country-based or regional water quality standards and regulations (US-EPA, 2012). In most cases, water quality requirements are based on determinants that characterise water quality (WHO, 2011; Nawaz et al., 2023). In 2017, the WHO has introduced the development of Water Safety Plans (WSP), which consider all risk aspects that can endanger potable water from source to tap water (Figueras and Borrego, 2010; WHO, 2017). In South Africa, the Department of Water and Sanitation (DWS) is the custodian of water policing, and they have developed the South African Water Quality guidelines, which are grouped into four categories according to water use. The different categories include domestic (drinking, cooking, bathing, laundry, and gardening), large-scale agricultural activities (irrigation and livestock), industrial, aquatic ecosystems and recreational guidelines categories (Mathebula, 2015). These categories are structured like the resource quality objectives (RQOs) with ideal (fit for use), tolerable (fit for consumption for a limited period) and unacceptable as classes of water quality.

1.3. Water resource quality objectives and targets

The National Water Act (NWA) No. 36 of 1998 makes provision for every water resource to have a list of prioritised parameters for continuous water quality monitoring. The Inkomati-Usuthu Catchment Management Agency (IUCMA) manages both the Crocodile River and Inyaka Dam and has since developed RQOs for every catchment under its management. Section 13(1) of the NWA of 1998 requires that a CMA develop and implement RQOs. As such, in 2016, the water RQOs were gazetted (Department of Water and Sanitation, 2016).

1.3.1. The Crocodile River

The Crocodile River catchment is located entirely within the Mpumalanga province in South Africa. The river originates near Dullstroom where it flows into the Kweni Dam passing through Mbombela before joining the Komati, then enters Mozambique at Komatipoort. The Elands River and Kaap River form part of the two major tributaries whereas the Lungsklip River, Nels River, Houtbosloopspruit, Gladdespruit, White River, and Besterspruit form part of the minor tributaries. The upper part of the catchment is largely dominated by man-made forests as well as commercial and emerging agriculture.

The Crocodile River is monitored by the IUCMA monthly for physicochemical and microbiological parameters, which have been set as objectives or targets adequate for water use. Table 1.1 depicts a set of parameters and their set targets for the Crocodile River, which were chosen on a risk basis about the area. There are several wastewater treatment plants which discharge directly into the river, hence the inclusion of total suspended solids (TSS), ammonia (NH_3), and phosphate (PO_4^{3-}). The metals were selected based on the related industrial activities in the surroundings.

Table 1.1: Water quality objectives for the Crocodile River (DWS, 2016). MPN denotes most Probable Number.

Variable	Ideal	Acceptable	Tolerable
pH (pH units)	6.5-8.0		
Electrical conductivity (mS/m)	30	50	60
TSS (mg/L)	5	15	25
NH ₃ -N (mg/L)	0.015	0.058	0.1
PO ₄ ³⁻ (mg/L)	0.03	0.05	0.1
SO ₄ ²⁻ (mg/L)	20	40	60
NO ₃ -NO ₂ (mg/L)	0.5	2	4
Chloride (mg/L)	25	40	50
Fluoride (mg/L)	0.2	0.5	0.75
Manganese (Mn) (mg/L)	0.02	0.10	0.30
Iron (Fe) (mg/L)	0.01	0.55	0.75
Arsenic (mg/L)	0.01	0.05	0.08
Aluminium (Al) (mg/L)	0.03	0.07	0.09
<i>E. coli</i> (MPN/100mL)	10	80	120

1.3.2. The Inyaka Dam

The Inyaka Dam is an earth fill type of dam that is built along the Ngwaritsane River in the Bushbuckridge area of Mpumalanga. The dam was built with the primary purpose of storing water for irrigation. The dam is surrounded by domestic settlements, agricultural activities, and a wastewater treatment plant. As of 2014, the dam served as an abstraction point for the Inyaka water treatment plant which supplied treated water to over 750,821 people in Bushbuckridge (Statistics South Africa, 2022). Given the population supplied by the Inyaka Dam, the Dam is one of the most important water resources that is unfortunately, under-represented in literature. The Inyaka dam is associated with fewer target parameters compared to the Crocodile River, reflecting differences in catchment characteristics. In addition, the Inyaka Dam has limited industrial activity around the rivers that flow into it. However, there are small-scale agricultural activities in the rivers, presenting a potential risk of nutrient-rich catchment, hence the inclusion of ortho-phosphate, ammonia, nitrites, and nitrates as priority parameters. Table 1.2 presents both the nutrient-related parameters relevant to agricultural inputs and the common physicochemical and microbiological parameters applicable to the Sabie catchment. The exclusion of metals is also noted, as there are no industrial activities to contribute to a risk.

Table 1.2: Target water quality guidelines for the Sabie catchment to which the Inyaka dam belongs (DWS, 2016). MPN denotes Most Probable Number.

Variable	Target Water Quality Guidelines
pH (pH units)	6.5-8.5
Electrical conductivity (mS/m)	0-40
Ammonia (mg/L)	0-1.0
Chemical Oxygen Demand (COD)	0-10
Nitrate and Nitrite (mg/L)	0-6
Soluble Orthophosphate (mg/L)	0.005-0.025
Total Suspended Solids (mg/L)	0-5
<i>E. coli</i> (MPN/100mL)	0

1.4. Microbial ecology of freshwater resources

Microbial ecology is a collective term for microbial diversity, community structure and microbial community responses to environmental changes (Logue and Lindstrom, 2008; Jiang et al., 2025). Aquatic ecosystems are the most diverse habitats, and they include fresh surface waters (lentic and lotic) (Jordaan, 2015) that support diverse microbial communities with different physiological characteristics as well as different growth rates, which are influenced by changing environmental conditions (Kirchman, 2012; Sehnal et al., 2021). Freshwater resources are the most important building blocks of life, even though they make up less than 1% of the Earth's surface water (Harrison et al., 2016; Dudgeon, 2019). Several studies (Hahn, 2006; Musie & Gonfa, 2023) define freshwater as water with a low salinity (<1g/L), which formulates the basis of daily life, and provides resources for domestic use, agricultural, and industrial uses. Since the end of the Second World War, the world's population has grown significantly (United Nations, 2008; Klein et al., 2010), with low and middle-income countries seeing more population growth than high-income countries (Joseph and McGinley, 2008; Paprotny, 2021). These changes have culminated in increased food demand, which in turn pressurises the already burdened water resources (Ban et al., 2025). Freshwater resources are also significant recreational and cultural resources for humans (W. Zhang et al., 2023). As such, sustainable management of freshwater resources is fundamental for clean water supply as it drives both the economic and ecological systems (Hahn, 2006; Asaeda et al., 2009; Olley et al., 2024).

1.5. Major bacterial lineages in freshwater systems

Freshwater ecosystem harbours a diverse group of prokaryotic organisms that differ in physiology, metabolism, morphology, and geographical preference (Cabello-Yeves & Rodriguez-Valera, 2019; Musie & Gonfa, 2023). Until the recent past (~30 years), the most abundant microbial taxa comprised the river-based bacterioplankton that resembled lake-based bacteria, which were regarded as typical freshwater bacteria (Zwart et al., 2002; Lozupone and Knight, 2007; Newton et al., 2011; Jiao et al., 2018). According to Kirchman (2012), the most dominant type of bacteria is *Proteobacteria*, which includes the *Betaproteobacteria* as the principal taxa followed by *Gammaproteobacteria* and *Alphaproteobacteria*. Moreover, other bacterial groups within the microbial ecology of freshwater include *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Verrucomicrobia*, with *Proteobacteria*, and its subgroups such as the *Alpha*, *Beta*, and *Gammaproteobacteria*, being dominant bacterial communities in rivers (Liu et al., 2012; Betiku et al., 2021; Zemskaya et al., 2022). Metagenomic studies have also reported the general dominance of the freshwater lineage *Actinobacteria*-1 (ac1) of the phylum *Actinobacteria* as discussed in the sections that follow (Ghai et al., 2011; Kang et al., 2017).

1.5.1. Actinobacteria

Actinobacteria are one of the other numerically dominant phyla in freshwater ecosystems (Betiku et al., 2021; Zemskaya et al., 2022). These include Gram-positive bacteria representing one of the biggest taxa of the major lineages. They are recognised in the Bacteria domain and include five subclasses, six orders, and 14 suborders (Ludwig et al., 2012; Boukhatem et al., 2022). Most *Actinobacteria* are free-living organisms found in both terrestrial and aquatic ecosystems (Macagnan et al., 2006; Shahrajabian & Sun, 2025). In addition, *Actinobacteria* have a high guanine-plus-cytosine (G+C) content in their genomes (Seshadri et al., 2022). To date, most actinobacterial genomes that have been sequenced belong to organisms relevant to human and veterinary medicine, ecology, and industrial microbiology (Ventura et al., 2007; Nouiouei et al., 2018). A study conducted by Jordaan et al. (2019) reported *Actinobacteria* among seven other dominant phyla in freshwater. This finding aligns with the works of Drury et al. (2013) and Ibekwe et al. (2016).

1.5.2. Bacteroidota

The *Bacteroidota* is a diverse bacterial phylum with its name having been changed several times over the past years (Thomas et al., 2011). Previously, this phylum was known as the *Cytophaga-Flexibacter-Bacteroides* (CFB) group to reflect the diversity of bacterial organisms classified in this phylogenetic group (Woese, 1987; Woese et al., 1990) and recently renamed Bacteroidota (formerly Bacteroidetes) (Shin et al., 2024). The Bergey's manual of Systematic Bacteriology (Bergey, 2011) reported that the phylum consists of four classes, namely *Bacteroidia*, *Flavobacteria*, *Sphingobacteria*, and *Cytophagia*. According to NCBI records, the four classes shared among them 7000 species, with *Flavobacteria* being the largest of the four (NCBI, 2023). These gram-negative bacteria have been reported to colonise different ecological systems, including the gastrointestinal tract, soil, ocean, and freshwater (Thomas et al., 2011; Shin et al., 2024). Further, using Fluorescence in Situ Hybridization (FISH), *Bacteroidetes* were found in all lake water samples taken from four different countries (Austria, Switzerland, Germany, and Siberia) with a reported median abundance between 2% and 12% (Glockner et al., 1999) and they accounted for more than 40% of bacteria detected in river samples from Spain (Simek, 2001). Another important aspect of the *Bacteroidota* is that they are highly dependent on organic matter load or cyanobacterial blooms; as such, the lineages are different from other common freshwater groups because they have been reported not to show any temporal or lake-specific occurrence patterns (Eiler and Bertilsson, 2007; Newton et al., 2011; Zheng et al., 2022).

1.5.3. Proteobacteria

The phylum *Proteobacteria* was formerly called the “purple bacteria” and comprises four recognised groups, viz, alpha, beta, gamma, and delta, which were classified according to the structure of their 16S rRNA genetic sequences (Woese, 1987; Sharma et al., 2022). In 2005, the phylum was also re-established using phylogenetic analysis of the 16S rRNA genetic sequences with five classes containing all the known gram-negative bacterial groups (alpha, beta, gamma, and delta) (Garrity et al., 2005). In addition, recent molecular analyses using complete multi-genome and multiprotein alignment have enhanced the possibilities of more detailed probing of these class delineations (Williams et al., 2007; 2010; Yutin et al., 2012). According to Imhoff et al., (2018), *Proteobacteria* consist of phototrophs, chemolithotrophs, and chemoorganotrophs found in both oxic and anoxic habitats. Members of the *Betaproteobacteria* are active in the nitrogen cycle by providing plants with fixed nitrogen via

the oxidation of ammonium to nitrate (Newton, 2008; Chiang et al., 2018). However, the *Alpha* and *Gammaproteobacteria* are less commonly found in freshwaters (Newton, 2008; Chiang et al., 2018). Despite this, *Alphaproteobacteria* play an important role in freshwater by degrading organic compounds (Newton et al., 2011; Chiang et al., 2018), whereas *Gammaproteobacteria* are adapted to high nutrient conditions and members of the *Enterobacteriaceae* family. These are commonly found in human and animal excreta and are used as indicators of water quality (Stoeckel and Harwood, 2007; Newton et al., 2011; Chiang et al., 2018).

1.6. Freshwater ecosystems: temporal and spatial variations of microbial communities

Temporal variation refers to the changes accumulated with elapsing time, whereas spatial variation refers to geographical changes. In the context of freshwater bacterial communities, temporal and spatial variation refers to the dissimilarities in microbial communities that occur in space and time between and within habitats as influenced by a myriad of factors (Logue and Lindstrom, 2008; Wu et al., 2024). In other words, freshwater bacterial communities, in their complexity and diversity, have low evenness as the communities tend to be dominated by a few different groups (Kirchman, 2012; Schmidt et al., 2020; Feng et al., 2025). In most cases, the dominant species will do so for short periods at varying depths, which results in a succession process of dominant communities (Yannarell and Kent 2009; Yang et al., 2023). Several reports have demonstrated that several factors control bacterial abundance and community structure, including water chemistry, temperature (Yang et al., 2023), and quantity of dissolved organic matter (DOM) (Louca et al., 2016; Hou et al., 2017; Yang et al., 2023; Feng et al., 2025). This means that the dynamics of freshwater bacterial ecosystems are controlled by many fast-changing niches, which are used by different species (Yannarell and Kent, 2009). Furthermore, these rapid developments and the collapse of niches cause dramatic shifts in the variance of bacterial community structures over a short period. It has, however, been revealed that bacterial communities mainly experience varying changes between long periods of stability and rapid turnover (Barbour et al., 2022; Stanić et al., 2025).

1.7. Water Quality Index (WQI)

Water quality indices (WQIs) are mechanisms used to monitor water quality based on an expression of a mathematical or statistical formula for defining the level of water quality (Bordalo et al., 2006; Sutadian et al., 2018; Uddin et al., 2021). These indices also provide a

measure of how water quality parameters perform in comparison to set water quality guidelines or objectives for a defined area (Kannel et al., 2007; Mophin-Kani and Murugesan, 2011; Mogane et al., 2023). In these indices, complex data is compacted into simplified mathematical numbers which are given water quality classes such as excellent, very good, good, moderate, and poor (Akbar, 2013; Sutadian et al., 2018; Uddin et al., 2021). The quality of water indices estimation endeavour single value, which decreases the big quantity of parameters and represents data in a simple way. Usually, a higher mathematical number or score will denote better water quality (good to very good) and a lower score points towards degraded water quality (moderate, poor) (Lumb et al., 2011, Uddin et al., 2021). Indices are considered the most powerful tool in relaying useful information to policymakers and the public (Mophin-Kani and Murugesan, 2011; Chidiac et al., 2023). Conversely, while this is a normal technical language within the water resource quality community, it does not always provide meaningful information to communities and lawmakers. As such, the prime objective of water quality indices should be to turn complex data sets into information that can be easily comprehended and used by the public (Pinto et al., 2013; Fortes et al., 2023). While most WQI utilise a variety of parameters to generalise the state of resource water quality, WQI based on some very important parameters can provide a strong single indicator of water quality (Thakor et al., 2011; Salih et al., 2012; Uddin et al., 2021). There are various numbers of WQI aggregation functions that have been used in the past. Table 1.3 provides a list of commonly used functions:

Table 1.3: Commonly used functions to compute WQI.

Q_i = subindex/transformed parameter i , and W_i = weight of the transformed parameter.

Name of function	Formula	Remarks
Weighted sum index	$WQI = \sum_{i=1}^n Q_i \cdot W_i$	This method deals away with ambiguity but suffers from eclipsing
Root mean square function	$WQI = \sqrt{0.5 \{Min(Q_i)^2 + [(1/n) \sum Q_i^2]\}}$	In this function, an index is zero if any one sub index is zero. This characteristic helps to eliminate the eclipsing and ambiguity problem
Weighted geometric function	$WQI = \prod_{i=1}^n Q_i^{W_i}$	In this function, an index is zero if any one sub index is zero. This characteristic helps to eliminate the eclipsing and ambiguity problem
Weighted Solway function	$WQI = \frac{1}{100} \{W_i \cdot Q_i\}^2$	In this function, an index is zero if any one sub index is zero. This characteristic helps to eliminate the eclipsing and ambiguity problem
Unweighted geometric	$WQI = \{\sum_{i=1}^n Q_i\}^{1/n}$	In this function, an index is zero if any one sub index is zero. This characteristic helps to eliminate the eclipsing and ambiguity problem
Un-weighted harmonic square mean function	$WQI = \sqrt{\frac{n}{\sum_{i=1}^n (Q_i)^2}}$	In this function, an index is zero if any one sub index is zero. This characteristic helps to eliminate the eclipsing and ambiguity problem
Maximum operator index	$WQI = \text{Max} \{I_1, I_2, I_3 \dots I_n\}$	This function is suited for applications in which an index must report if at least one recommended unit is violated.

1.8. Microbial indicators of water quality

There are several microbial groups which are known to be indicator microorganisms or index microorganisms, which are also termed process indicators, faecal indicators, and model organisms (Odonkor and Ampofo, 2013; Richiardi et al., 2023). Process indicators are a group of microorganisms which are used to determine the efficacy of a treatment process, such as heterotrophic bacteria and total coliforms, for chlorine disinfection. Faecal indicators, on the other hand, demonstrate the presence of faecal pollution and organisms of these groups include thermotolerant bacterial groups such as coliforms. Index and model organisms include organisms which are indicative of the presence of a specific pathogen, such as F-RNA coliphages being models of human enteric viruses, and *E. coli* as an index of Salmonella (Odonkor and Ampofo, 2013; Richiardi et al., 2023). The most common group of microbial indicators of water quality are the coliforms, which have been used for more than 100 years as an indicator for microbes originating from faeces (Savichtcheva et al., 2006; Tambi et al., 2023). Table 1.4 presents common microbial indicators of water quality, and their descriptions as determined by the WHO.

Table 1.4: Common microbial indicators of water quality

Microbial indicator	Description
Coliforms	These are Gram-negative, rod-shaped, facultative anaerobic microorganisms that belong to the <i>Enterobacteriaceae</i> family which are characterised by the production of the enzyme galactosidase. In a growth medium such as the Colilert-18 [®] , these organisms will grow within 18-22 hours at 36-38°C.
<i>Escherichia coli</i>	Thermophilic coliforms known to produce indole from tryptophan. Recent descriptions, however, have defined <i>E. coli</i> as coliforms that can produce β -glucuronidase. This group is said to be the most appropriate indicator of faecal contamination from humans and warm-blooded animals.
Faecal coliforms (Thermotolerant coliforms)	These are coliforms which ferment lactose to produce acid and gas at 44-45°C within 24-26 hours.
Faecal Streptococci	These are Gram-positive cocci which belong to two genera, <i>Enterococcus</i> and <i>Streptococcus</i> which possess the Lancefield group D antigen. They can only be cultured from selective media such as m Enterococcus agar and azide dextrose broth.
Bifidobacteria	These are Gram-positive bacilli which are highly pleomorphic, catalase-negative, obligate anaerobes non-motile and non-spore-forming and most numerous in the faeces of warm-blooded animals

(WHO, 2001)

1.9. Selected physicochemical parameters

1.9.1. Temperature

Temperature plays an important role in water chemical reactions and rates of reactions, sustaining aquatic life, and the condition of water for functions beneficial to humans and the environment (Metcalf and Eddy, 2004; Larance et al., 2025). The optimum temperature for bacteria ranges from 25°C - 37°C and a sharp increase in temperature halts aerobic digestion and nitrification. Similarly, a temperature decrease of 15°C causes inactivity in some methane-

producing bacteria. According to the South African Water Quality Guidelines (vol. 7): Aquatic Ecosystems (DWS, 1996), water temperature should not vary from the average by more than 10% or $>2^{\circ}\text{C}$. Changes in water temperature can indirectly affect water quality by influencing microbial growth, chemical reaction rates, and the effectiveness of water treatment processes (Abalasei et al., 2025; Khaliq et al., 2024). On the other hand, temperature is a critical parameter of aquatic ecosystem functioning because it influences metabolic activity, microbial processes, reproduction, and the distribution of aquatic organisms (Wang et al., 2026)

1.9.2. Turbidity

Turbidity can be loosely explained as the measure of the light-transmitting properties of water (Mathebula, 2015). High turbidity in water is an indication of colloidal and residual suspended matter. Turbidity is important for many municipal and industrial applications since suspended solids can shield microorganisms in the water or cause a high disinfection demand, which will interfere with the efficacy of the disinfection (Bratby, 2006; Bratby, 2015). High rainfall events tend to influence turbidity, especially with rivers, because more suspended solids flow into the water body or suspended solids are brought back from the bottom sediments into suspension (Dallas and Day, 2004; Li et al., 2024). Elevated turbidity is also known to reduce light penetration and can consequently interfere with aquatic primary production (Shen et al., 2019). Additionally, high concentration of suspended materials may affect aquatic organisms through physical abrasion, sedimentation, and modification of habitat conditions (Krupa et al., 2025).

1.9.3. Electrical Conductivity

Electrical conductivity refers to the ability of water to pass an electric current (Gapparov & Isakova, 2023). Conductivity in water is greatly influenced by the number of dissolved ions, such as chloride, sulphides, carbonate compounds, etc, and cations such as calcium, iron, etc. Conductivity is also highly influenced by temperature. A higher temperature will decrease the viscosity of water, thus causing higher conductivity (Clor et al., 2012; Gapparov & Isakova, 2023; Yang et al., 2022). The survival, growth and reproduction of any aquatic organism is highly dependent on the dissolved ion concentration in any water source (Farias et al., 2024). For this reason, the limit of TDS for source water to be regarded as freshwater is 1000mg/L. For electrical conductivity, the Sabie-Sand and Crocodile River catchments the resource

quality objective is set at 30mS/m (DWS, 2016). For humans, elevated conductivity may affect the suitability of water for domestic, agricultural, and industrial uses (Loock et al., 2015).

1.9.4. pH

Potential of Hydrogen (pH) is the intensity of factor of acidity in a water body (DiBartola, 2006;Bezuidenhout, 2013) and is a crucial parameter for evaluating the quality of natural and waste waters. The concentration of hydrogen is influenced by the extent to which water molecules dissociate to form hydrogen and hydroxyl ions (Jay-Gerin, 2025). The set resource objectives for pH in the study areas range between 6.5 and 8.5. Extreme pH conditions may influence the solubility of and mobility of metals and potentially altering their availability in water (Król et al., 2020). In aquatic ecosystems, pH plays a major role because organisms have different tolerances to acidic and alkaline conditions. As such, changes in pH influences physiological processes and alter chemical availability, and toxicity of other elements in water (Saalidong et al., 2022).

1.9.5. Iron

Iron (Fe) occurs in nature in three oxidation states, zero (0), Fe II and Fe III, with Fe III being the most common form (Chen et al., 2018). Fe is an essential micronutrient in all living organisms and forms part of the haem. In water, it occurs as suspended iron hydroxides, dissolved ferric iron (Fe^{2+}) and ferrous iron (Fe^{3+}) (Dallas and Day, 2004; Zhang et al., 2024). Fe is an abundant element which makes up to 5.6% of the earth's crust and water contamination by iron can originate from geologic sources, industrial effluents and domestic waste (Colter and Mahler, 2006; Hu et al., 2020). Fe is, however, an essential element for haemoglobin and myoglobin, and its deficiency may lead to anaemia (Zheng et al., 2023). However, Fe overload causes serious health problems in humans like liver cancer, diabetes, heart disease and infertility. In freshwater, high concentrations of Iron causes respiratory dysfunction, compromised anti-oxidative defense system, and severe histological damage in gills and liver of fish (Singh et al., 2019). Furthermore, the presence of high concentrations in water changes the aesthetic value of water, stains clothes and corrodes water pipelines (Wojtkowska et al., 2022).

1.9.6. Manganese

Manganese (Mn) occurs naturally in many surface waters and in soils that may erode into the surface waters (Crapnell & Banks, 2022). However, human activities also account for much of the contamination of water in some areas (WHO, 2011). The current acceptable level for Mn in South Africa is set at 0.10 milligrams per litre (mg/L) and ideal levels are set at 0.02 mg/l (IUCMA, 2014). Manganese is usually present in water sources in the reduced form (Mn II) or oxidized form (Mn IV). Manganese (II) is however the most soluble form and its oxidation results in the formation of a dark brown precipitate which causes the scaling of pipes (Gerke et al., 2016). The removal of soluble Mn from surface water to be treated for drinking can be affected by applying chlorine to the influent directly upstream (Hoyland et al., 2014). While manganese is an essential element at appropriate concentrations, elevated presence in water may pose health concerns (Podgorski & Berg, 2026). For aquatic ecosystems, high manganese levels may alter water chemistry and contribute to the formation of precipitates that affect aquatic habitats and infrastructure (Matveeva et al., 2022; Wu et al., 2022).

1.9.7. Nitrate-nitrogen.

Nitrogen is a naturally occurring element essential for plant growth and crop production (Farhan et al., 2024). It is also a key component of proteins and nucleic acids. Several reports have however stressed the potential dangers of reactive nitrogen (N_r) (Nohel et al., 2008; Donatela et al., 2010; Nieder & Benbi, 2022). This is exacerbated by the fact that nitrogen is extremely abundant (78%) in the Earth's atmosphere and exists in the unreactive form as N_2 . The remaining 22% of nitrogen formulates the reactive nitrogen which is useful to plants and animals. N_r includes oxidised and reduced nitrogen compounds like nitrates, ammonia, organic nitrogen compounds and nitric acid (Sutton et al., 2011; Nieder & Benbi, 2022). Agriculture has been reported to play a major role in surface water contamination because of direct irrigation run-offs and high usage of nitrogen-based fertilisers (Gomes et al., 2025). As much as nitrogen is essential to the functioning of all living ecosystems. However, too much of it may lead to the acceleration of processes such as eutrophication (Sutton, 2011; de Vries, 2021). The overall result of this is a disruption of ecological function, loss of biodiversity, toxic algal blooms, and serious degradation of water quality for either use, be it agriculture, drinking, or just recreation (Tian et al., 2007; Lassaletta et al., 2009; Mian et al., 2010; Akinnawo, 2023).

1.9.8. Phosphate.

The natural occurrence of phosphorus is limited in surface water and just a small amount may trigger algal blooms, especially in water bodies with high nitrogen and carbon (Mylavarapu, 2017). Algal blooms are normally activated by an estimated phosphorus concentration of 80ug/L (Guo et al., 2023). The phosphorus concentration in surface water bodies must be <25ug/L to significantly avoid eutrophication and improve water clarity (Shock and Pratt, 2003; (Savic et al., 2022).

1.10. Research problem statement.

Approximately 12 million rural South Africans lack access to safe drinking water, relying instead on rivers for various domestic purposes (Thwala, 2010; Gakuba et al., 2015; DWS, 2018; Bazaanah & Mothapo, 2024; Tshona et al., 2025). The rural areas of the City of Mbombela and Bushbuckridge Local Municipality face similar challenges. Compounding the issue, the Crocodile River and Inyaka Dam are part of catchments surrounded by extensive domestic, agricultural, and industrial activities (MTPA, 2012; IUCMA, 2014, 2016), making them highly susceptible to pollution from multiple sources. Consequently, water sampling and analysis occur more frequently than standard practice to monitor these water bodies effectively.

The Inyaka Dam and Crocodile River, located in different municipalities—Bushbuckridge Local Municipality (BLM) and City of Mbombela (COM)—illustrate a common South African challenge described by Burke and Bokako (2004) as the "data-rich-but-information-poor" syndrome. Despite the abundance of daily water quality data, it is often underutilised, preventing water managers and decision-makers from implementing effective measures to mitigate pollution and protect aquatic ecosystems (Neswiswi, 2014; Bashir et al., 2020).

Using *E. coli* as an indicator organism tends to overlook other environmental contamination sources (Luyt et al., 2012; (Deblais et al., 2024; Richiardi et al., 2023). Therefore, including a broader range of microorganisms is critical, especially when assessing large water systems. Although bacteria are natural components of ecosystems, microbiological contamination throughout watersheds is a significant concern (Bayoumi and Patko, 2012; Haldar et al., 2022). Aquatic microbial diversity plays an essential role in ecosystem functioning (Cotner & Biddanda, 2002; Gessner et al., 2010). While progress has been made in understanding the link between microbial diversity and ecosystem function (Smith, 2007; Singer et al., 2010; Comte

et al., 2013), the response of microbial diversity to environmental changes requires further study (Allison and Martiny, 2008; Langenheder et al., 2012; Shade et al., 2012; Logue et al., 2015).

Changes in the physicochemical properties of water can cause systematic variations in freshwater ecosystems (Saturday et al., 2021). Variations in parameters such as temperature, conductivity, turbidity, pH, and elevated levels of inorganic and organic nutrients, as well as salts like magnesium, iron, manganese, sulphates, and phosphates, pose significant threats to ecosystem health (Dey et al., 2021).

1.11. Study rationale.

Addressing water quality challenges fundamentally requires a multi-faceted approach involving pollution prevention, effective water treatment, and the restoration or preservation of aquatic ecosystems (Namugize and Jewitt, 2018; Ejiohuo et al., 2025). However, these strategies must be complemented by a robust and inclusive monitoring system that goes beyond simply generating statistical data. Such systems should provide actionable insights that enable effective decision-making.

Neswiswi (2014) highlights that water quality managers and decision-makers often lack the scientific expertise needed to interpret complex statistical data. This disconnect hampers their ability to identify pollution sources, leaving communities vulnerable to waterborne diseases and accelerating the degradation of ecosystems. Water quality models can distil intricate data into concise, meaningful information, serving as a practical decision-making support tool. Building on this concept, the water quality assessment model proposed in this study aims to overcome the limitations of relying solely on *E. coli* as an indicator organism for biological quality of water. By incorporating a diverse range of bacterial and physicochemical indicators of water quality, this model will offer a more comprehensive understanding of water quality dynamics. Using historical data to forecast water quality will provide water quality managers with sufficient data to make informed decisions during emergencies, thereby saving time and costs.

1.12. Objectives of the study, Hypotheses and Research Questions.

1.12.1. Overall objective of the study

To develop a water quality assessment model for the Crocodile River and the Inyaka dam in the Mpumalanga province of South Africa.

1.13. Specific Objectives, Hypotheses and Research Questions

- **Specific objective 1:** To investigate the microbial and physicochemical water quality characteristics of the Crocodile River and Inyaka Dam across all four seasons.

Hypothesis: Seasonal weather variations have an impact on the water quality of surface water.

Research Question: How does water quality vary across seasons and what are the main influencers?

- **Specific objective 2:** To investigate the bacterial community structure and the dominant bacterial taxa in the Crocodile River and Inyaka Dam across four seasons.

Hypothesis: The bacterial community structure and dominant bacterial taxa would not be uniform in different seasons.

Research Question: What are the effects of significantly variable water quality on freshwater ecosystems and ecological function?

- **Specific objective 3:** To develop an inclusive water quality index model for the Crocodile River and the Inyaka Dam.

Research Question:

Would an artificial intelligence (AI)-based water quality prediction model help to identify trends in water quality to obtain necessary information on pollution sources?

1.14. Study assumptions.

The selected physical and chemical water quality parameters are the foundation of this study. The main assumption of the study is that there will be significant temperature variations across South African weather seasons which will affect microbial growth dynamics. In addition, there would be considerable rainfall in the autumn and summer seasons which will disturb settled sediment in the water resources and increase turbidity and river flows. Furthermore, high microbial activity is expected to occur in the warmer and rainy seasons of autumn and summer.

1.15. Delimitations of the study

“We need to know precisely what the researcher intends to do. We also need to know precisely what the researcher does not intend to do.... The limits of the problem should be as carefully bounded for a research effort as a parcel of land is for a real estate transfer. ... Only a researcher who thinks carefully about the problem and its focal centre can distinguish between what is relevant and what is not relevant to the problem. All irrelevancies to the problem must be firmly ruled out in the statement of delimitations.” (Leedy and Ormrod, 2005).

In accordance with the preceding citation, the investigation activities of this study are focused on lentic and lotic ecosystems. However, a lentic system can be anything ranging from a natural lake, dam, pond, seep, basin marshes and seeps. Likewise, a lotic system includes all flowing inland water bodies such as rivers, streams, and creeks. As such, this study will only be considering a dam (Inyaka Dam) and a river (Crocodile River) as lentic and lotic systems respectively.

1.16. Limitations of the study

The model envisaged to be produced by this study is only limited to water quality parameters which are of concern to the study areas. This means that the model, in its current state, can only be applied in the study areas or regions where similar water quality issues are experienced. While the varying dynamics on the two ecosystems are taken into cognizance, the study used the same water quality parameters of interest in both these ecosystems to see through the study objectives.

1.17. Conceptual framework

The following diagram presents the study's conceptual framework and outlines the interconnection between the specific objectives of the study.

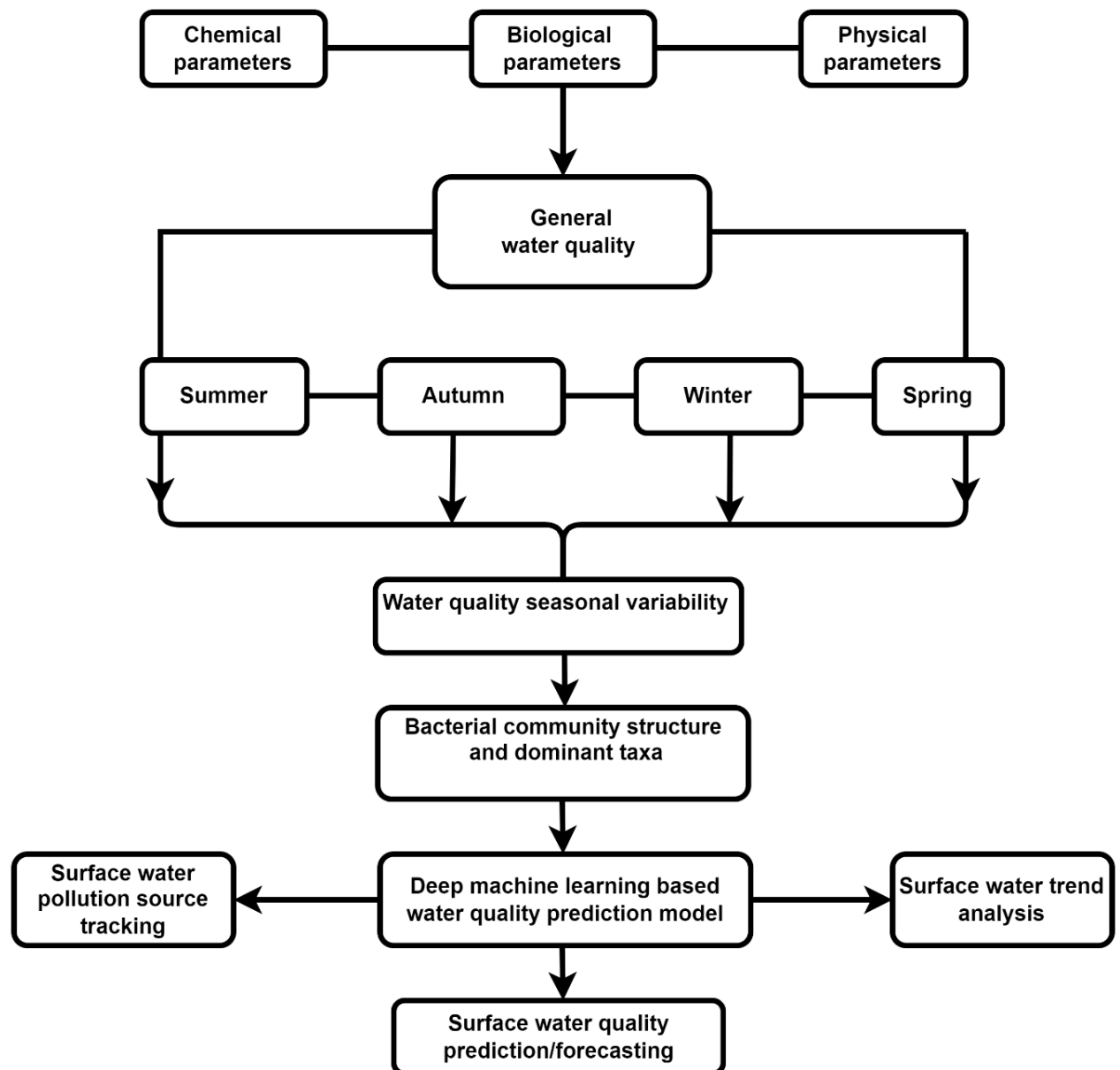


Fig. 1.1 Conceptual framework of the study

1.18. Thesis structure

The thesis is configured into seven chapters.

Chapter 1: Introduction

Attempts to identify, define, and clarify the fundamental and pertinent issues central to the study. It covers the following: a detailed background of the study, research statement, study rationale, conceptual framework, objectives, hypotheses, and research question

Chapter 2: Literature Review

The second chapter provides a comprehensive review of water quality index models used to assess lotic and lentic ecosystems. The literature review presents an appraisal of 17 general water quality assessment WQIs based on five developmental steps which are: parameter selection, transformation to a common scale, generation of parameter weights, aggregation to compute the final index value, and water quality classification schemes. The scope of application for each WQI, which forms part of the pertinent aspects of the review, is also discussed. The chapter further discusses the limitations of WQI models and suggests statistical and machine learning techniques aimed to ameliorate these limitations. This chapter is based on the following article:

Mogane, L.K., Masebe, T., Msagati, T.A.M. Ncube E. A comprehensive review of water quality indices for lotic and lentic ecosystems. *Environ Monit Assess* **195**, 926 (2023). <https://doi.org/10.1007/s10661-023-11512-2>

Chapter 3: Methodology

This chapter details the research methodology that was used in the research exercise. It further details the techniques used to gather data and answer research objectives and questions.

Chapter 4: Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa

This chapter highlights the significant role of seasonal variability in shaping water quality in the Crocodile River and Inyaka Dam. It identifies rainfall and temperature as key drivers behind these variations, with distinct water quality parameters influencing each aquatic system differently. The results obtained in this chapter underscore how these environmental factors

not only fluctuate seasonally but also directly affect the physical, chemical, and biological characteristics of the water in both ecosystems. The findings emphasise the intricate relationship between climate patterns and water quality, illustrating how changes in rainfall and temperature create ripple effects on aquatic ecosystem dynamics. Seasonal shifts impact nutrient loads, turbidity, microbial activity, and other critical water quality indicators, ultimately influencing the ecological balance and health of these water resources. The chapter also draws attention to the broader implications of these seasonal variations, particularly in the context of climate change. It underscores the importance of understanding seasonal dynamics to anticipate potential water quality challenges and stresses the need for adaptive, ecosystem-specific water management strategies.

This chapter is based on the following journal article:

Lazarus Katlego Mogane, Esper J. Ncube, Titus A.M. Msagati, Tracy M. Masebe, Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa, *Heliyon*, Volume 10, Issue 19, 2024, e38246. <https://doi.org/10.1016/j.heliyon.2024.e38246>

Chapter 5: Exploring the physicochemical drivers of bacterial community dynamics in lotic and lentic freshwater ecosystems.

This chapter utilises advanced sequencing tools to analyse the bacterial community structure within the Crocodile River and Inyaka Dam. The study focuses on identifying the dominant bacterial taxa present throughout the different weather seasons, providing insights into the seasonal dynamics of microbial communities in these aquatic ecosystems. In addition to profiling bacterial communities, the chapter examines the chemical and physical factors that influence or promote the dominance of specific bacterial taxa. By correlating the water quality data presented in Chapter 4 with the seasonal bacterial community profiles, this analysis aims to uncover the relationships between environmental parameters such as temperature, pH, nutrient concentrations, turbidity, and microbial composition. The investigation provides a comprehensive understanding of how external factors drive shifts in microbial communities and shape ecosystem health. By integrating microbial data with environmental variables, the findings contribute to a deeper understanding of microbial ecology in freshwater systems and highlight the role of hydrological context in shaping bacterial diversity. This knowledge is

critical for informing adaptive ecosystem-specific water quality management strategies and addressing the ecological impacts of environmental changes on freshwater ecosystems.

The chapter is based on the following proposed research paper:

Lazarus Katlego Mogane, Titus A.M. Msagati, Esper Ncube, and Tracy Masebe. Exploring the physicochemical drivers of bacterial community dynamics in lotic and lentic freshwater ecosystems.

Target journal: *Microbial Ecology* (Springer Nature)

Chapter 6: Development of a Water Quality Prediction Model using an Artificial Neural Network (ANN)-based Machine Learning Algorithm

Chapter 6 concludes the series of experimental studies by presenting the development of a water quality prediction tool. Built using the Python platform, the tool employs a feed-forward, fully connected, multi-layer perceptron architecture based on artificial neural networks (ANN). The ANN approach enables the model to capture complex, non-linear relationships between input variables and predicted outcomes, making it a powerful tool for water quality forecasting.

This chapter not only describes the technical details of the tool but also emphasizes its potential applications. By leveraging historical and real-time data, the tool can assist water resource managers in anticipating water quality changes, identifying pollution risks, and implementing timely interventions. Additionally, the predictive capability of the model aligns with the broader goal of improving water resource management, supporting sustainable development, and mitigating the impacts of environmental variability and climate change.

The development of this model marks a significant step toward integrating advanced machine learning techniques into water quality monitoring and decision-making processes. It provides a foundation for further refinement and adaptation to various aquatic systems, paving the way for more resilient and data-driven water management strategies.

This chapter is based on the following proposed research article:

Lazarus Katlego Mogane, Titus A.M. Msagati, Esper Ncube, and Tracy Masebe. Development of a Water Quality Prediction Model using an Artificial Neural Networks (ANN)-based Machine Learning Algorithm using the Crocodile River and Inyaka Dam

Submitted to *IWA Journal of Hydroinformatics* (Under Review)

Chapter 7: Synthesis of chapter findings, Conclusions and Recommendations.

This chapter provides a comprehensive summary of the findings and insights presented throughout the study, synthesising the key conclusions drawn from each chapter. It reflects on the core objectives, hypotheses, and research questions, evaluates how they were achieved, and connects the individual experimental outcomes to the broader goal of advancing water quality monitoring and management. In summarising the chapters, this section highlights critical discoveries, such as the seasonal variability of water quality, the role of environmental parameters in shaping microbial communities, and the application of advanced predictive models for water quality assessment. It consolidates these findings into a cohesive narrative, cementing their significance for water resource management, ecosystem preservation, and public health protection. Additionally, this chapter addresses the limitations encountered during the study, offering an evaluation of the challenges faced, including financial, data availability, sampling constraints, and the inherent complexity of modelling dynamic aquatic systems. In doing so, it sets the stage for future research opportunities, proposing areas where further investigation is needed to build upon the foundation laid by this study.

1.19. References

- Abalasei, M. E., Toma, D., Dorus, M., & Teodosiu, C. (2025). The Impact of Climate Change on Water Quality: A Critical Analysis. In *Water (Switzerland)* (Vol. 17, Number 21). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/w17213108>
- Akbar, T.A. (2013). Development and application of water quality classification models, University of Calgary, PhD thesis.
- Akinawo, S. O. (2023). Eutrophication: Causes, consequences, physical, chemical and biological techniques for mitigation strategies. *Environmental Challenges*, 12, 100733. <https://doi.org/10.1016/J.ENVC.2023.100733>
- Allison, S.D., and Martiny, J.B.H. (2008). Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Science USA*, 105, 11512–11519.
- Asaeda, T., Manatunge, J., Priyadarshana, T., Park, B.K. (2009). Problems, restoration, and conservation of lakes and rivers. In: Wolanski E, 2009. *Oceans and Aquatic Ecosystems*, Vol. 1. Encyclopaedia of Life Support Systems.
- Ban, Y., Tang, Q., Liu, X., Li, X., Ren, Y., Deng, H., Zhao, G., Zhang, Y., Zhang, Y., & Leng, G. (2025). Intensification of water scarcity threatens future food security in northwestern China. *Journal of Hydrology: Regional Studies*, 58. <https://doi.org/10.1016/j.ejrh.2025.102177>
- Barbour, K. M., Weihe, C., Allison, S. D., & Martiny, J. B. H. (2022). Bacterial community response to environmental change varies with depth in the surface soil. *Soil Biology and Biochemistry*, 172, 108761. <https://doi.org/10.1016/J.SOILBIO.2022.108761>
- Bashir, I., Lone, F. A., Bhat, R. A., Mir, S. A., Dar, Z. A., & Dar, S. A. (2020). Concerns and threats of contamination on aquatic ecosystems. In *Bioremediation and Biotechnology: Sustainable Approaches to Pollution Degradation* (pp. 1–26). Springer International Publishing. https://doi.org/10.1007/978-3-030-35691-0_1
- Bayoumi-Hamuda, H.E.A.F., and Patko, I. (2012). Ecological monitoring of Danube water quality in Budapest region. *American Journal of Environmental Science*, 8: 202–211.
- Bazaanah, P., & Mothapo, R. A. (2024). Sustainability of drinking water and sanitation delivery systems in rural communities of the Lepelle Nkumpi Local Municipality, South Africa. *Environment, Development and Sustainability*, 26(6), 14223–14255. <https://doi.org/10.1007/s10668-023-03190-4>
- Betiku, O. C., Sarjeant, K. C., Ngatia, L. W., Aghimien, M. O., Odewumi, C. O., & Latinwo, L. M. (2021). Evaluation of microbial diversity of three recreational water bodies using 16S rRNA metagenomic approach. *Science of The Total Environment*, 771, 144773. <https://doi.org/10.1016/J.SCITOTENV.2020.144773>
- Bezuidenhout, N. (2013). An investigation into the Cyanobacteria and related Cyanotoxins in the Vaalkop Treatment Plant, Rustenburg, Unpublished Thesis, University of Johannesburg, pp. 42-44.

- Bordalo, A.A., Teixeira, R., Wiebe, W.J.A. (2006). Water Quality Index Applied to an International Shared River Basin: The Case of the Douro River, *Environmental Management*, 38, 910-920.
- Boukhatem, Z. F., Merabet, C., & Tsaki, H. (2022). Plant Growth Promoting Actinobacteria, the Most Promising Candidates as Bioinoculants? In *Frontiers in Agronomy* (Vol. 4). Frontiers Media S.A. <https://doi.org/10.3389/fagro.2022.849911>
- Bratby, J. (2015). Turbidity: Measurement of Filtrate and Supernatant Quality? *Progress in Filtration and Separation*, 637–657. <https://doi.org/10.1016/B978-0-12-384746-1.00016-1>
- Burke, J., Bokako, T. (2004). Optimisation of surface water quality monitoring within the City of Johannesburg. *Proceedings of the 2004 Water Institute of Southern Africa (WISA) Biennial Conference*, 26 May 2004, Cape Town, South Africa
- Cabello-Yeves, P. J., & Rodriguez-Valera, F. (2019). Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome*, 7(1). <https://doi.org/10.1186/s40168-019-0731-5>
- Chen, R., Liu, H., Tong, M., Zhao, L., Zhang, P., Liu, D., & Yuan, S. (2018). Impact of Fe(II) oxidation in the presence of iron-reducing bacteria on subsequent Fe(III) bio-reduction. *Science of The Total Environment*, 639, 1007–1014. <https://doi.org/10.1016/J.SCITOTENV.2018.05.241>
- Chiang, E., Schmidt, M. L., Berry, M. A., Biddanda, B. A., Burtner, A., Johengen, T. H., Palladino, D., & Denef, V. J. (2018). Verrucomicrobia are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats. *PLoS ONE*, 13(3). <https://doi.org/10.1371/journal.pone.0195112>
- Chidiac, S., El Najjar, P., Ouaini, N., El Rayess, Y., & El Azzi, D. (2023). A comprehensive review of water quality indices (WQIs): history, models, attempts and perspectives. In *Reviews in Environmental Science and Biotechnology* (Vol. 22, Issue 2, pp. 349–395). Springer Science and Business Media B.V. <https://doi.org/10.1007/s11157-023-09650-7>
- Clor, L.E., McCleskey, R.B., Huebner, M.A., Lowenstern, J.B., Heasler, H.P., Maloney, T., Evans, W.C. (2012). Water chemistry and electrical conductivity database for river in Yellowstone National Park, Wyoming: *U.S. Geological Survey Data Series* 632, p.18.
- Colter, A., Mahler, R.L. (2006). Iron in Drinking Water, A Pacific Northwest Extension Publication. University of Idaho, Oregon State University and Washington State University in cooperation with University of Alaska, Fairbanks, Accessed 12/11/2018, <http://www.cals.uidaho.edu/edComm/pdf/pnw/pnw589.pdf>.
- Comte, J.R.M., Fauteux, L., and Del Giorgio, P.A. (2013). Links between metabolic plasticity and functional redundancy in freshwater bacterioplankton communities. *Frontiers of Microbiology*, 4:112.
- Cotner, J.B., and Biddanda, B.A. (2002). Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. *Ecosystems* 5, 105–121

- Cox, B.A. (2003). A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Science of the Total Environment*, 314–316, 335–377.
- Crapnell, R. D., & Banks, C. E. (2022). Electroanalytical overview: The determination of manganese. *Sensors and Actuators Reports*, 4, 100110. <https://doi.org/10.1016/J.SNR.2022.100110>
- Dallas, H.F., Day, J.A. (2004). The effect of water quality variables on aquatic ecosystems: A review. Report No. TT 224/04. *Water Research Commission*. Pretoria, South Africa.
- de Vries, W. (2021). Impacts of nitrogen emissions on ecosystems and human health: A mini review. *Current Opinion in Environmental Science & Health*, 21, 100249. <https://doi.org/10.1016/J.COESH.2021.100249>
- Deblais, L., Ahmedo, B. U., Ojeda, A., Mummed, B., Wang, Y., Mekonnen, Y. T., Demisie Weldesenbet, Y., Hassen, K. A., Brhane, M., McKune, S., Havelaar, A. H., Liang, S., & Rajashekara, G. (2024). Assessing fecal contamination from human and environmental sources using *Escherichia coli* as an indicator in rural eastern Ethiopian households—a cross-sectional study from the EXCAM project. *Frontiers in Public Health*, 12. <https://doi.org/10.3389/fpubh.2024.1484808>
- Department of Water Affairs and Forestry. (1996). Water Quality Guidelines for Aquatic Ecosystems, vol. 7. https://www.dws.gov.za/iwqs/wq_guide/edited/Pol_saWQguideFRESH_vol7_Aquaticecosystems.pdf
- Department of Water and Sanitation. (2016). Classes of water resources and resource quality objectives for the catchments of the Inkomati, *Government Gazette*, 40531, vol. 618: 4-79.
- Department of Water and Sanitation, 2018. National water and sanitation master plan (NW&SMP) vol. 1. Call to Action version 10.1:pp. 2.
- Dey, S., Botta, S., Kallam, R., Angadala, R., & Andugala, J. (2021). Seasonal variation in water quality parameters of Gudlavalluru Engineering College pond. *Current Research in Green and Sustainable Chemistry*, 4, 100058. <https://doi.org/10.1016/J.CRGSC.2021.100058>
- DiBartola, S. P. (2006). Introduction to acid-base disorders. *Fluid, Electrolyte and Acid-Base Disorders in Small Animal Practice*, 229–251. <https://doi.org/10.1016/B0-72-163949-6/50012-6>
- Drury, B., Rosi-Marshall, E., Kelly, J.J. (2013). Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Applied Environmental Microbiology* 79(6):1897–905. <https://doi.org/10.1128/AEM.03527-12> PMID: 23315724
- du Plessis, A. (2019). Evaluation of Southern and South Africa's Freshwater Resources. In: Water as an Inescapable Risk. *Springer Water*. Springer, Cham. https://doi.org/10.1007/978-3-030-03186-2_7

- Dudgeon, D. (2019). Multiple threats imperil freshwater biodiversity in the Anthropocene. *Current Biology*, 29(19), R960–R967. <https://doi.org/10.1016/J.CUB.2019.08.002>
- Ejiohuo, O., Onyeaka, H., Akinsemolu, A., Nwabor, O. F., Siyanbola, K. F., Tamasiga, P., & Al-Sharify, Z. T. (2025). Ensuring water purity: Mitigating environmental risks and safeguarding human health. In *Water Biology and Security* (Vol. 4, Issue 2). KeAi Communications Co. <https://doi.org/10.1016/j.watbs.2024.100341>
- Eiler, A., Bertilsson, S. (2007). Flavobacteria blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. *Applied and Environmental Microbiology*, 73(11): 3511–3518.
- Farhan, M., Sathish, M., Kiran, R., Mushtaq, A., Baazeem, A., Hasnain, A., Hakim, F., Naqvi, S. A. H., Mubeen, M., Iftikhar, Y., Abbas, A., Hassan, M. Z., & Moustafa, M. (2024). Plant Nitrogen Metabolism: Balancing Resilience to Nutritional Stress and Abiotic Challenges. *Phyton-International Journal of Experimental Botany*, 93(3), 581–609. <https://doi.org/10.32604/PHYTON.2024.046857>
- Farias, L., Beszteri, B., Burfeid Castellanos, A. M., Doliwa, A., Enss, J., Feld, C. K., Grabner, D., Lampert, K. P., Mayombo, N. A. S., Prati, S., Schürings, C., Smollich, E., Schäfer, R. B., Sures, B., & Le, T. T. Y. (2024). Influence of salinity on the thermal tolerance of aquatic organisms. *Science of The Total Environment*, 953, 176120. <https://doi.org/10.1016/J.SCITOTENV.2024.176120>
- Feng, C., Lu, J., Liu, T., Shi, X., Zhao, S., Lv, C., Shi, Y., Zhang, Z., Jin, Y., Pang, J., & Sun, A. (2025). Microbial community dynamics in shallow-water grass-type lakes: Habitat succession of microbial ecological assembly and coexistence mechanisms. *Ecotoxicology and Environmental Safety*, 291, 117819. <https://doi.org/10.1016/J.ECOENV.2025.117819>
- Figueras, M.J., Borrego, J.J. (2010). New perspectives on monitoring drinking water microbial quality. *International Journal of Environmental Research and Public Health*, 7(12):4179-4202.
- Fortes, A. C. C., Barrocas, P. R. G., & Kligerman, D. C. (2023). Water quality indices: Construction, potential, and limitations. *Ecological Indicators*, 157, 111187. <https://doi.org/10.1016/J.ECOLIND.2023.111187>
- Gakuba, E., Moodley, B., Ndungu, P., and Birungi, G. (2015). Occurrence and significance of polychlorinated biphenyls in water, sediment pore water and surface sediments of Umgeni River, KwaZulu-Natal, South Africa. *Environmental, Monitoring and Assessment* 187 (9) 1-14.
- Ghai, R., Rodriguez-Valera, F., McMahon, K.D., Toyama, D., Rinke, R., Cristina Souza de Oliveira, T., et al. (2011). Metagenomics of the water column in the pristine upper course of the Amazon River. *PLoS ONE*, 6: e23785.

- Gapparov, A., & Isakova, M. (2023). Study on the characteristics of water resources through electrical conductivity: A case study of Uzbekistan. *IOP Conference Series: Earth and Environmental Science*, 1142(1). <https://doi.org/10.1088/1755-1315/1142/1/012057>
- Gavrilaş, S., Burescu, F. L., Chereji, B. D., & Munteanu, F. D. (2025). The Impact of Anthropogenic Activities on the Catchment's Water Quality Parameters. In *Water (Switzerland)* (Vol. 17, Issue 12). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/w17121791>
- Gessner, M.O., Swan, C.M., Dang, C.K., Mckie, B.G., Bardgett, R.D., Wall, D.H., et al. (2010). Diversity meets decomposition. *Trends of Ecological Evolution*, 25: 372–380.
- Gerke, T. L., Little, B. J., & Barry Maynard, J. (2016). Manganese deposition in drinking water distribution systems. *Science of The Total Environment*, 541, 184–193. <https://doi.org/10.1016/J.SCITOTENV.2015.09.054>
- Glöckner, F.O., Fuchs, B.M., and Amann, R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Applied Environmental Microbiology*, 65, 3721–3726.
- Gomes, M., Ralph, T. J., Humphries, M. S., Graves, B. P., Kobayashi, T., & Gore, D. B. (2025). Waterborne contaminants in high intensity agriculture and plant production: A review of on-site and downstream impacts. *Science of The Total Environment*, 958, 178084. <https://doi.org/10.1016/J.SCITOTENV.2024.178084>
- Guo, Y., Fu, W., Xiong, N., He, J., & Zheng, Z. (2023). Phosphorus Threshold for the Growth of *Microcystis wesenbergii*, *Microcystis aeruginosa*, and *Chlorella vulgaris* Based on the Monod Formula. *Water (Switzerland)*, 15(24). <https://doi.org/10.3390/w15244249>
- Hahn, M.W. (2006). The microbial diversity of inland waters. *Current Opinion in Biotechnology*, 17(3): 256–261
- Halder, K., Kujawa-Roeleveld, K., Hofstra, N., Datta, D. K., & Rijnaarts, H. (2022). Microbial contamination in surface water and potential health risks for peri-urban farmers of the Bengal delta. *International Journal of Hygiene and Environmental Health*, 244, 114002. <https://doi.org/10.1016/J.IJHEH.2022.114002>
- Horton, R.K. (1965). An Index number system for rating water quality. *Journal of Water Pollution Control Federation*, 37(3): 300-306.
- Hoyland, V. W., Knocke, W. R., Falkinham, J. O., Pruden, A., & Singh, G. (2014). Effect of drinking water treatment process parameters on biological removal of manganese from surface water. *Water Research*, 66, 31–39. <https://doi.org/10.1016/J.WATRES.2014.08.006>
- Hu, Y., Wu, G., Li, R., Xiao, L., & Zhan, X. (2020). Iron sulphides mediated autotrophic denitrification: An emerging bioprocess for nitrate pollution mitigation and sustainable wastewater treatment. *Water Research*, 179, 115914. <https://doi.org/10.1016/J.WATRES.2020.115914>

- Ibekwe, A.M., Jincai, M., Murinda, S.E. (2016). Bacterial community composition and structure in an urban river impacted by different pollutant sources. *Science of the Total Environment*, 566–567:1176–85.
- Imhoff, J. F., Rahn, T., Künzel, S., & Neulinger, S. C. (2018). Photosynthesis is widely distributed among Proteobacteria as demonstrated by the phylogeny of PufLM reaction center proteins. *Frontiers in Microbiology*, 8(JAN). <https://doi.org/10.3389/fmicb.2017.02679>
- Inkomati-Usuthu Catchment Management Agency (IUCMA). (2014). Annual water quality status report for the Inkomati-Usuthu water management area, IUCMA, Nelspruit, South Africa.
- Inkomati-Usuthu Catchment Management Agency (IUCMA). (2016). Ecstatus of the Sabie-Sand River Catchment: Inkomati River System Phase II. IUCMA, Nelspruit, South Africa.
- Jay-Gerin, J.-P. (2025). Fundamentals of Water Radiolysis. *Encyclopedia*, 5(1), 38. <https://doi.org/10.3390/encyclopedia5010038>
- Jiang, W., Wang, S., Gu, F., Yang, X., Qi, Q., & Liang, Q. (2025). Advances in synthetic microbial ecosystems approach for studying ecological interactions and their influencing factors. In *Engineering Microbiology* (Vol. 5, Issue 2). Elsevier B.V. <https://doi.org/10.1016/j.engmic.2025.100205>
- Jiao, C., Zhao, D., Huang, R., Cao, X., Zeng, J., Lin, Y., & Zhao, W. (2018). Abundant and rare bacterioplankton in freshwater lakes subjected to different levels of tourism disturbances. *Water (Switzerland)*, 10(8). <https://doi.org/10.3390/w10081075>
- Jordaan K, 2015. Molecular profiling of microbial population dynamics in environmental water. PhD thesis, North-West University
- Jordaan, K., Comeau, A.M., Khasa, D.P., Bezuidenhuot, C.C. (2019). An integrated insight into the response of bacterial communities to anthropogenic contaminants in a river: A case study of the Wonderfonteinspruit catchment area, South Africa. *PloS ONE*, 14(5):1-24.
- Juwana, I., Muttill, N., Perera, B.J.C. (2014). Application of West Java water sustainability index to Citarum catchment in West Java, Indonesia. *Water Science and Technology: Water Supply*, 14: 1150-1159
- Kang, I., Kim, S., Islam, M.R., Cho, J.C., (2017). The first complete genome sequences of the acI lineage, the most abundant freshwater *Actinobacteria*, obtained by whole genome amplification of dilution-to-extinction cultures. *Science Reports*, 7:42252
- Kannel, P.R., Lee, S., Lee, Y.S., Kanel, S.R., and Khan, S.P. (2007). Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. *Environmental Monitoring and Assessment*, 132, 93–110
- Khalik, I., Chollet Ramampandra, E., Vorburger, C., Narwani, A., & Schuwirth, N. (2024). The effect of water temperature changes on biological water quality assessment. *Ecological Indicators*, 159. <https://doi.org/10.1016/j.ecolind.2024.111652>
- Kirchman DL, 2012. Processes in microbial ecology. New York: Oxford University Press Inc.

- Król, A., Mizerna, K., & Bożym, M. (2020). An assessment of pH-dependent release and mobility of heavy metals from metallurgical slag. *Journal of Hazardous Materials*, 384, 121502. <https://doi.org/10.1016/J.JHAZMAT.2019.121502>
- Krupa, P. M., Kennedy, A. J., Mcqueen, A. D., & Wilkens, J. L. (2025). *Effects of Suspended Sediment on Aquatic Organisms A Literature Review and Database Effort Environmental Laboratory*. <https://www.fisheries.noaa.gov>
- Langenheder, S., Berga, M., Ostman, O., and Szekely, A.J. (2012). Temporal variation of beta-diversity and assembly mechanisms in a bacterial metacommunity. *International Society for Microbial Ecology Journal*, 6:1107–1114
- Larance, S., Wang, J., Delavar, M. A., & Fahs, M. (2025). Assessing Water Temperature and Dissolved Oxygen and Their Potential Effects on Aquatic Ecosystem Using a SARIMA Model. *Environments - MDPI*, 12(1). <https://doi.org/10.3390/environments12010025>
- Leedy, P., Ormrod, J.E. (2005). Practical Research, Planning and Design. Eighth edition. Upper Saddle River, New York: Pearson Education Inc
- Li, H., Lv, J., He, X., Bao, Y., & Nsabimana, G. (2024). Precipitation-dependent sensitivity of suspended sediment concentration to turbidity in a mountainous river in southwestern China. *Ecological Indicators*, 159, 111644. <https://doi.org/10.1016/J.ECOLIND.2024.111644>
- Liangliang, G., Daoliang, L. (2014). A review of hydrological/water-quality models. *Frontiers of Agricultural Science and Engineering*, 1: 267–276
- Liu, Z., Huang, S., Sun, G., Xu, Z., and Xu, M. (2012). Phylogenetic diversity, composition and distribution of bacterioplankton community in the Dongjiang River, China. *FEMS Microbiology Ecology*, 80: 30–44
- Logue, J.B., and Lindström, E.S. (2008). Biogeography of bacterioplankton in inland waters. *Freshwater Reviews* 1: 99-114
- Logue, J. B., Findlay, S. E. G., & Comte, J. (2015). Editorial: Microbial responses to environmental changes. In *Frontiers in Microbiology* (Vol. 6, Issue DEC). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2015.01364>
- Lokman, A., Ismail, W. Z. W., & Aziz, N. A. A. (2025). A Review of Water Quality Forecasting and Classification Using Machine Learning Models and Statistical Analysis. *Water*, 17(15), 2243. <https://doi.org/10.3390/w17152243>
- Loock, M. M., Beukes, J. P., & van Zyl, P. G. (2015). Conductivity as an indicator of surface water quality in the proximity of ferrochrome smelters in South Africa. *Water SA*, 41(5), 705–711. <https://doi.org/10.4314/wsa.v41i5.14>
- Louca, S., Parfrey, L.W., Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, 353: 1272–1277
- Ludwig, W., Euzéby, J., Schumann, P., Buss H.J., Trujillo, M.E., Kämpfer, P., Whiteman, W.B. (2012). Road map of the phylum *Actinobacteria*, p 1–28. In Goodfellow M, Kämpfer P, Busse

- HJ, Trujillo ME, Suzuki KI, Ludwig W, Whitman WB (ed), Bergey's manual of systematic bacteriology, *Springer-Verlag* vol. 5, New York, NY
- Luyt, C.D., Tandlich, R., Muller, W.J., Wilhelmi, B.S. (2012). Microbial monitoring of surface water in South Africa. *International Journal of Environmental Research and Public Health*, 9: 2670-2693
- Macagnan, D., Romeiro, R.D.S., de Souza, J.T., Pomella, A.W.V. (2006). Isolation of *Actinomycetes* and endospore-forming bacteria from the cacao pod surface and their antagonistic activity against the witches' broom and black pod pathogens. *Phytoparasitica* 3:122–132
- Mathebula, B. (2015). Assessment of the surface water quality of the main rivers feeding into the Katse dam, Lesotho, MSc Thesis, University of Pretoria.
- Matveeva, V. A., Alekseenko, A. V., Karthe, D., & Puzanov, A. V. (2022). Manganese Pollution in Mining-Influenced Rivers and Lakes: Current State and Forecast under Climate Change in the Russian Arctic. *Water (Switzerland)*, 14(7). <https://doi.org/10.3390/w14071091>
- Metcalf, L., and Eddy, H.P. (2004). Wastewater engineering. Fourth Edition, McGraw-Hill Publishers, New York.
- Mogane, L. K., Masebe, T., Msagati, T. A. M., & Ncube, E. (2023). A comprehensive review of water quality indices for lotic and lentic ecosystems. In *Environmental Monitoring and Assessment* (Vol. 195, Issue 8). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s10661-023-11512-2>
- Mophin-Kani K, and Murugesan AG, 2011.Evaluation and classification of water quality of Perennial River Tamirabarani through aggregation of water quality index. *International Journal of Environmental Protection*, 1(5): 24-33.
- Mpumalanga Tourism and Parks Agency (MTPA). (2012). Ecstatus of the Sabie/Sand River Catchments. MTPA Scientific Services, Nelspruit, South Africa.
- Musie, W., & Gonfa, G. (2023). Fresh water resource, scarcity, water salinity challenges and possible remedies: A review. In *Heliyon* (Vol. 9, Issue 8). Elsevier Ltd. <https://doi.org/10.1016/j.heliyon.2023.e18685>
- Nawaz, R., Nasim, I., Irfan, A., Islam, A., Naeem, A., Ghani, N., Irshad, M. A., Latif, M., Nisa, B. U., & Ullah, R. (2023). Water Quality Index and Human Health Risk Assessment of Drinking Water in Selected Urban Areas of a Mega City. *Toxics*, 11(7). <https://doi.org/10.3390/toxics11070577>
- Nemugize, J.N., Jewitt, G.P.W. (2018). Sensitivity analysis for water quality monitoring frequency in the application of a water quality index for the uMngeni River and its tributaries, KwaZulu-Natal, South Africa. *Water SA*, vol. 44(4):516-527.
- Neswiswi, A.B. (2014). Development of water quality index for the Juskei river catchment, Johannesburg, MSc, University of Witwatersrand.

- Newton, R.J. (2008). Cosmopolitan freshwater bacterial dynamics in lakes across time and space. PhD Dissertation. University of Wisconsin-Madison.
- Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*, 75: 14–49.
- Nhemachena, C., Nhamo, L., Matchaya, G., Nhemachena, C. R., Muchara, B., Karuaihe, S. T., & Mpandeli, S. (2020). Climate change impacts on water and agriculture sectors in southern Africa: Threats and opportunities for sustainable development. In *Water (Switzerland)* (Vol. 12, Issue 10, pp. 1–17). MDPI AG. <https://doi.org/10.3390/w12102673>
- Nieder, R., & Benbi, D. K. (2022). Reactive nitrogen compounds and their influence on human health: an overview. In *Reviews on Environmental Health* (Vol. 37, Issue 2, pp. 229–246). De Gruyter Open Ltd. <https://doi.org/10.1515/reveh-2021-0021>
- Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., Pukall, R., Klenk, H. P., Goodfellow, M., & Göker, M. (2018). Genome-based taxonomic classification of the phylum actinobacteria. *Frontiers in Microbiology*, 9(AUG). <https://doi.org/10.3389/fmicb.2018.02007>
- Olley, J., Cvitanovic, M., Ginige, T., & Bunt-MacRury, L. (2024). A systematic literature review of sustainable water management in South Africa. *Sustainable Water Resources Management*, 10(5). <https://doi.org/10.1007/s40899-024-01135-x>
- Paprotny, D. (2021). Convergence Between Developed and Developing Countries: A Centennial Perspective. *Social Indicators Research*, 153(1), 193–225. <https://doi.org/10.1007/s11205-020-02488-4>
- Pinto, U., Maheshwari, B.L., and Ollerton, R.L. (2013). Analysis of long-term water quality for effective river health monitoring in peri-urban landscapes—a case study of the Hawkesbury–Nepeanriver system in NSW Australia. *Environmental, Monitoring and Assessment*, 185: 4551–4569
- Pitman, W.V. (2011). Overview of the water resource assessment in South Africa: Current state and future challenges. *Water SA*, 37(5):659–664
- Podgorski, J., & Berg, M. (2026). Manganese, a hidden threat to global water quality and health. *Nature Water*. <https://doi.org/10.1038/s44221-026-00676-z>
- Richiardi, L., Pignata, C., Fea, E., Bonetta, S., & Carraro, E. (2023). Are Indicator Microorganisms Predictive of Pathogens in Water? In *Water (Switzerland)* (Vol. 15, Issue 16). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/w15162964>
- Salih, N.M., Khwakaram, A.I., and Hama, N.Y. (2012). Determination of Water Quality Index (WQI) for Qalyasan stream in Sulaimani city/Kurdistan region of Iraq. *Plant, Animal and Environmental Sciences*, 2:148–157

- Saaliidong, B. M., Aram, S. A., Otu, S., & Lartey, P. O. (2022). Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems. *PLoS ONE*, 17(1 1). <https://doi.org/10.1371/journal.pone.0262117>
- Sapkal, R.S., Valunekar, S.S. (2013). Development and sensitivity analysis of water quality index for evaluation of surface water for drinking purpose, *International Journal of Civil Engineering and Technology*, vol. 4(4):119-134
- Saturday, A., Lyimo, T. J., Machiwa, J., & Pamba, S. (2021). Spatio-temporal variations in physicochemical water quality parameters of Lake Bunyonyi, Southwestern Uganda. *SN Applied Sciences*, 3(7). <https://doi.org/10.1007/s42452-021-04672-8>
- Savic, R., Stajic, M., Blagojević, B., Bezdan, A., Vranesovic, M., Jokanović, V. N., Baumgartel, A., Kovačić, M. B., Horvatinec, J., & Ondrasek, G. (2022). Nitrogen and Phosphorus Concentrations and Their Ratios as Indicators of Water Quality and Eutrophication of the Hydro-System Danube–Tisza–Danube. *Agriculture (Switzerland)*, 12(7). <https://doi.org/10.3390/agriculture12070935>
- Schmidt, M. L., Biddanda, B. A., Weinke, A. D., Chiang, E., Januska, F., Props, R., & Denef, V. J. (2020). Microhabitats are associated with diversity-productivity relationships in freshwater bacterial communities. *FEMS Microbiology Ecology*, 96(4). <https://doi.org/10.1093/FEMSEC/FIAA029>
- Sehnal, L., Brammer-Robbins, E., Wormington, A. M., Blaha, L., Bisesi, J., Larkin, I., Martyniuk, C. J., Simonin, M., & Adamovsky, O. (2021). Microbiome Composition and Function in Aquatic Vertebrates: Small Organisms Making Big Impacts on Aquatic Animal Health. In *Frontiers in Microbiology* (Vol. 12). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2021.567408>
- Seshadri, R., Roux, S., Huber, K. J., Wu, D., Yu, S., Udway, D., Call, L., Nayfach, S., Hahnke, R. L., Pukall, R., White, J. R., Varghese, N. J., Webb, C., Palaniappan, K., Reimer, L. C., Sardà, J., Bertsch, J., Mukherjee, S., Reddy, T. B. K., ... Ivanova, N. N. (2022). Expanding the genomic encyclopedia of Actinobacteria with 824 isolate reference genomes. *Cell Genomics*, 2(12). <https://doi.org/10.1016/j.xgen.2022.100213>
- Shade, A., Peter, H., Allison, S.D., Baho, D.L., Berga, M., Burgmann, H., et al. (2012). Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*, 3:417
- Shahrajabian, M. H., & Sun, W. (2025). Microbial interaction of Actinobacteria strains with various plants, promote growth and development in natural farming, and alleviating biotic and abiotic stresses. In *Discover Sustainability* (Vol. 6, Issue 1). Springer Nature. <https://doi.org/10.1007/s43621-025-01413-4>
- Sharma, V., Vashishtha, A., Jos, A. L. M., Khosla, A., Basu, N., Yadav, R., Bhatt, A., Gulani, A., Singh, P., Lakhera, S., & Verma, M. (2022). Phylogenomics of the Phylum Proteobacteria: Resolving the Complex Relationships. *Current Microbiology*, 79(8). <https://doi.org/10.1007/s00284-022-02910-9>

- Shen, X., Sun, T., Su, M., Dang, Z., & Yang, Z. (2019). Short-term response of aquatic ecosystem metabolism to turbidity disturbance in experimental estuarine wetlands. *Ecological Engineering*, 136, 55–61. <https://doi.org/10.1016/J.ECOLENG.2019.06.005>
- Shin, J. H., Tillotson, G., MacKenzie, T. N., Warren, C. A., Wexler, H. M., & Goldstein, E. J. C. (2024). Bacteroides and related species: The keystone taxa of the human gut microbiota. *Anaerobe*, 85, 102819. <https://doi.org/10.1016/J.ANAEROBE.2024.102819>
- Sigee, D.C. (2005). Freshwater microbiology: biodiversity and dynamic interactions of microorganisms in the aquatic environment. West Sussex: John Wiley & Sons Ltd
- Simek, K., Armengol, J., Comerma, M., Garcia, J.C., Kojecka, P., Nedoma, J., and Hejzlar, J. (2001). Changes in the epilimnetic bacterial community composition, production, and protist-induced mortality along the longitudinal axis of a highly eutrophic reservoir. *Microbial Ecology*, 42: 359–371
- Singer, G., Besemer, K., Schmitt-Kopplin, P., Hodl, I., and Battin, T.J. (2010). Physical heterogeneity increases biofilm resource use and its molecular diversity in stream mesocosms. *PLoS ONE*, 5: e9988
- Singh, M., Barman, A. S., Devi, A. L., Devi, A. G., & Pandey, P. K. (2019). Iron mediated hematological, oxidative and histological alterations in freshwater fish Labeo rohita. *Ecotoxicology and Environmental Safety*, 170, 87–97. <https://doi.org/10.1016/J.ECOENV.2018.11.129>
- Smith, V.H. (2007). Microbial diversity-productivity relationships in aquatic ecosystems. *FEMS Microbiology Ecology*, 62: 181–186.
- Soyapi, C.B. (2017). Water security and the right to water in southern Africa: An overview. *Potchefstroom Electronic Law Journal*, vol. 20: 1-26
- Sravva, P.V.R., and Sreejani, T.P. (2016). Water Quality Modelling- Statistical Approach. *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3 (2): 6-10.
- Srebotnjak, T., Carr, G., de Sherbinin, A., Rickwood, C. (2012). A global water quality index and hot-deck imputation of missing data, *Ecological Indicators*, 17:108-119.
- Stanić, I., Kajan, K., Selak, L., & Orlić, S. (2025). Environmental drivers of microbial assembly and stability in lakes across biogeographical regions. *Ecological Indicators*, 172, 113324. <https://doi.org/10.1016/J.ECOLIND.2025.113324>
- Stoeckel, D.M., Harwood, V.J. (2007). Performance, design, and analysis in microbial source tracking studies. *Applied and Environmental Microbiology*, 73(8): 2405–2415
- Sutadian, A.D., Muttill, N., Yilmaz, A.G., Perera, B.J.C. (2016). Development of river water quality indices-a review. *Environmental, Monitoring and Assessment*, 188:58

- Tambi, A., Brighu, U., & Gupta, A. B. (2023). Methods for detection and enumeration of coliforms in drinking water: a review. *Water Supply*, 23(10), 4047–4058. <https://doi.org/10.2166/ws.2023.247>
- Thakor, F.J., Bhoi, D.K., Dabhi, H.R., Pandya, S.N., and Chauhan, N.B. (2011). Water Quality Index (I) of Pariyej Lake Dist. Kheda – Gujarat. *Current World Environment*, 6: 225- 231
- Thwala, W.D. (2010). Community participation is a necessity for project success: A case study of rural water supply project in Jeppes Reef, South Africa. *African Journal of Agricultural Research*, 5 (10): 970–979
- Tshona, S. S., Lungisa, S., & Mgweba, L. (2025). Thirsting for solutions: Unpacking inadequate water provision in rural communities. *Africa's Public Service Delivery & Performance Review*, 13(1). <https://doi.org/10.4102/apsdpr.v13i1.873>
- Uddin, M.G., Nash, S., and Olbert, A.I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. <https://doi.org/10.1016/j.ecolind.2020.107218>
- Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). A sophisticated model for rating water quality. *Science of the Total Environment*, 868. <https://doi.org/10.1016/j.scitotenv.2023.161614>
- van Rooyen, J., de Lange, M., and Hassan, R. (2010). Water resource situation strategies and allocation regimes in South Africa. *Transforming Water Management in South Africa. Springer*, 2:19-32.
- Ventura, M., Canchaya, C., Fitzgerald, G.F., Gupta, R.S., van Sinderen, D. (2007). Genomics as a means to understand bacterial phylogeny and ecological adaptation: The case of *Bifidobacteria*. *Antonie Van Leeuwenhoek*, 91:351–372
- Walter, T., Kloos, J., and Tsegai, D. (2011). Options for improving water use efficiency under worsening scarcity: Evidence from the Middle Olifants Sub-Basin in South Africa. *Water SA*, 37:357-370
- Wang, P., Bakari, J. S., Han, Y., Hu, H., Liu, Z., Zhang, Y., Chen, Z., Huang, C., Wang, M., Chen, H., Jing, X., & Su, S. (2026). Effects of Thermal Stress on Growth and Reproduction of *Procambarus clarkii* and Aquaculture Best Practices. In *Animals* (Vol. 16, Number 3). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/ani16030495>
- Williams, K.P., Sobral, B.W., and Dickerman, A.W. (2007). A robust species tree for the *Alphaproteobacteria*. *Journal of Bacteriology*, 189: 4578–4586
- World Health Organization (WHO). (2001). Water Quality: Guidelines, Standards and Health. Chapter 13. Edited by Lorna Fewtrell and Jamie Bartram. Published by IWA Publishing, London, UK. ISBN: 1 900222 28 0
- World Health Organisation (WHO). (2011). Guidelines for drinking water quality 4th edn. World Health Organisation. Geneva, Switzerland

- World Health Organisation (WHO). (2017). Guidelines for drinking water quality, 4th Edition: Incorporating the first addendum. Creative Commons Attribution-Non-Commercial, Switzerland, pp.: 45-49
- Woese, C.R. (1987). Bacterial evolution. *Microbiology Reviews*, 51: 221–271
- Woese, C.R., Yang, D., Mandelco, L., and Stetter, K.O. (1990). The flexibacter- flavobacter connection. *Systematic and Applied Microbiology*, 13: 161–165
- Wojtkowska, M., Malesińska, A., Machowska, A., Puntorieri, P., Barbaro, G., Fiamma, V., & Biedugnis, S. (2022). The Influence of Water Quality Change on the Corrosion Process in Galvanized Pipes of Fire Protection Installations. *Sustainability (Switzerland)*, 14(13). <https://doi.org/10.3390/su14137708>
- Wu, R., Yao, F., Li, X., Shi, C., Zang, X., Shu, X., Liu, H., & Zhang, W. (2022). Manganese Pollution and Its Remediation: A Review of Biological Removal and Promising Combination Strategies. In *Microorganisms* (Vol. 10, Number 12). MDPI. <https://doi.org/10.3390/microorganisms10122411>
- Wu, D., Zou, Y., Xiao, J., Mo, L., Lek, S., Chen, B., Fu, Q., & Guo, Z. (2024). The spatiotemporal variations of microbial community in relation to water quality in a tropical drinking water reservoir, Southmost China. *Frontiers in Microbiology*, 15. <https://doi.org/10.3389/fmicb.2024.1354784>
- Yang, G., Xu, Y., Huo, L., Wang, H., & Guo, D. (2022). Analysis of Temperature Effect on Saturated Hydraulic Conductivity of the Chinese Loess. *Water (Switzerland)*, 14(9). <https://doi.org/10.3390/w14091327>
- Yang, R., Fan, X., Zhao, L., & Yang, K. (2023). Identification of major environmental factors driving phytoplankton community succession before and after the regime shift of Erhai Lake, China. *Ecological Indicators*, 146, 109875. <https://doi.org/10.1016/J.ECOLIND.2023.109875>
- Yutin, N., Puigbo, P., Koonin, E.V., and Wolf, Y.I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE*, 7: e36972
- Zemskaya, T. I., Bukin, S. V., Bukin, Y. S., Chernitsina, S. M., Pogodaeva, T. V., Rusanov, I. I., Shubenkova, O. V., Zakharenko, A. S., & Pimenov, N. V. (2022). Taxonomic diversity and metabolic activity of microbial communities in rivers and estuarine waters of Southern Baikal in summer. *Journal of Great Lakes Research*, 48(1), 125–142. <https://doi.org/10.1016/J.JGLR.2021.10.019>
- Zhang, H., Rush, Z., Penn, Z., Dunn, K., Asmus, S., Cooke, C., Cord, Z., Coulter, S., & Morris, C. (2024). Films Floating on Water Surface: Coupled Redox Cycling of Iron Species (Fe (III)/Fe (II)) at Soil/Water and Water/Air Interfaces. *Water (Switzerland)*, 16(9). <https://doi.org/10.3390/w16091298>
- Zhang, W., ElDidi, H., Masuda, Y. J., Meinzen-Dick, R. S., Swallow, K. A., Ringler, C., DeMello, N., & Aldous, A. (2023). Community-Based Conservation of Freshwater Resources: Learning

from a Critical Review of the Literature and Case Studies. *Society and Natural Resources*, 36(6), 733–754. <https://doi.org/10.1080/08941920.2023.2191228>

Zheng, F., Zhang, T., Yin, S., Qin, G., Chen, J., Zhang, J., Zhao, D., Leng, X., An, S., & Xia, L. (2022). Comparison and interpretation of freshwater bacterial structure and interactions with organic to nutrient imbalances in restored wetlands. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.946537>

Zheng, H., Yang, F., Deng, K., Wei, J., Liu, Z., Zheng, Y. C., & Xu, H. (2023). Relationship between iron overload caused by abnormal hepcidin expression and liver disease: A review. In *Medicine (United States)* (Vol. 102, Issue 11). Lippincott gart and Wilkins. <https://doi.org/10.1097/MD.00000000000033225>

Zwart, G., Crump, B.C., Kamst-van Agterveld, M.P., Hagen, F., and Han, S.K. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology*, 28: 141–155.

2. Literature Review¹

2.1. Preamble

This chapter provides a comprehensive review of the literature on 17 water quality index (WQI) models designed to evaluate the overall water quality of lotic and lentic ecosystems. These models are typically developed using a standardised five-step process. This process includes the subjective selection of parameters, transformation of these parameters to a common scale, assignment of weights to the parameters, aggregation of parameter values, and classification of water quality to compute the final index value. While not all researchers strictly adhere to all five steps, this process has gained widespread acceptance as the universal methodology for developing WQIs. However, no universally accepted water quality index currently exists. Most indices are tailored to address specific regional water quality issues, which limits their applicability in other geographic areas. This presents a significant challenge when attempting to use these indices in diverse environmental contexts.

This review delves deeper into five widely used water quality indices for general water quality assessments. It concludes that the vast number of existing indices, each with a limited scope of application, underscores the need for future research to focus on adapting existing WQIs for broader applicability across both lotic and lentic ecosystems. Of note, the Canadian Council of Ministers of the Environment (CCME) model is the only index among the 17 reviewed that can be effectively applied to both lentic and lotic ecosystems. Furthermore, the parameters within the CCME model can be customized to align with the specific regional requirements, making it a versatile tool for assessing water quality.

¹ The contents of this chapter are published in *Environ Monit Assess* 195, 926 (2023). <https://doi.org/10.1007/s10661-023-11512-2>

2.2. A comprehensive review of water quality indices for lotic and lentic ecosystems

2.3. Abstract

Freshwater resources play a pivotal role in sustaining life and meeting various domestic, agricultural, economic, and industrial demands. As such, there is a significant need to monitor the water quality of these resources. Water quality index (WQI) models have gradually gained popularity since their maiden introduction in the 1960s for evaluating and classifying the water quality of aquatic ecosystems. WQIs transform complex water quality data into a single dimensionless number to enable accessible communication of the water quality status of water resource ecosystems.

To screen relevant articles for this study in the open literature, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method was employed to include or exclude articles. A total of 17 peer-reviewed articles were used in the final paper synthesis.

Among the reviewed reported WQIs, only the Canadian Council for Ministers of the Environment (CCME) index, Irish water quality index (IEWQI) and Hahn index were used to assess both lotic and lentic ecosystems. Furthermore, the CCME index is the only exception from rigidity because it does not specify parameters to select. Except for the West-Java WQI and the IEWQI, none of the reviewed WQI performed sensitivity and uncertainty analysis to improve the acceptability and reliability of the WQI. It has been proven that all stages of WQI development have a level of uncertainty which can be determined using statistical and machine learning tools. Extreme gradient boosting (XGB) has been reported as an effective machine learning tool to deal with uncertainties during parameter selection, the establishment of parameter weights, and determining accurate classification schemes. Considering the IEWQI model architecture and its effectiveness in coastal and transitional waters, this review recommends that future research in lotic or lentic ecosystems focus on addressing the underlying uncertainty issues associated with the WQI model in addition to the use of machine learning techniques to improve the predictive accuracy, robustness and increase the domain of application.

2.4. Background

Freshwater resources play a pivotal role in sustaining life as well as in meeting various domestic, agricultural, economic, and industrial needs. However, there is an ever-increasing concern about water security in terms of its availability, affordability, and quality, especially in arid and semi-arid regions of the world (Busico et al., 2020; Aragaw and Gnanachandrasamy, 2021). The concerns emanate from two fronts, which are – (i) the dwindling of freshwater resources due to climate change and other reasons and (ii) the continuous pollution of water resources due to anthropogenic, industrial, and agricultural activities which have become a serious environmental issue of concern, requiring appropriate strategies for constant monitoring accompanied by enforcement of regulatory policies to sustain such ecosystems (Gupta et al., 2009; Mahlathi et al., 2016; El-Batrawy et al., 2018; Nagy-Kovacs et al., 2019, Sandhu et al., 2019). It is also important to note that the deterioration of water quality decreases the portion of available safe and clean water, while the demand and dependence on water by humans and animals remain the same (Young and Beck, 1974; Peters and Meybeck, 2000; Kanakoudis and Tsitsifli, 2020). However, in the case of humans, they have adapted to the situation by developing survival mechanisms for dealing with polluted water resources through treatment processes for various uses, while animals do not have such options (Rangeti et al., 2015). As it stands, the current situation has become such that, rivers, streams, and dams are not only reliable sources of freshwater supply but also the disposal points of either untreated or partially treated wastewater effluents (Bartram and Balance, 1996; Das and Acharya, 2003; Tukura et al., 2009; Edokpayi et al., 2017). This has led to the mass contamination of water resources and has affected the normal functioning of aquatic ecosystems in many places (Kumarasamy and Macholo, 2018). Therefore, monitoring water quality to ensure the safety of consumers and the ecosystem has been an issue of paramount importance. To ensure that the monitoring and safeguarding of water resources are effective, it is imperative to put strategies in place to understand, improve, and mitigate such degrading effects. As such, various water quality indices (WQIs) have been devised to cater for such purposes.

A water quality index (WQI) model is a tool that converts large water quality data into a single value called the index score. The WQI model involves five stages of development, which include parameter selection, generation of sub-index functions, the establishment of parameter weights, aggregation of subindex values and determination of classification schemes. The importance of WQIs for the evaluation of water quality is highlighted by the number of studies

that seek to put to light the limitation (Abbasi and Abbasi, 2012; Sutadian et al., 2016; Uddin et al., 2021; Gupta and Gupta et al., 2021) and possible solutions (Malek et al., 2022; Uddin et al., 2022a, b, c, 2023a, b) in order to improve the accuracy, robustness, reliability, and wide acceptability. The earlier models of WQIs involved several subjective methods such as the Delphi Technique and expert opinions in the development stages, especially during parameter selection. This has been reported by most studies as a source of uncertainty and contributed to low model acceptability (Uddin et al., 2021; Sutadian et al., 2016).

The attractive aspect of using water quality indices in water resources management is that they present a qualitative method of aggregating or summarising water quality datasets from different parameters in a simpler, easier, and more understandable way (Couillard and Lefebvre, 1985; Cude, 2001; Tanner et al., 2011; Hoseinzadeh et al., 2015; Barakat et al., 2018). Since 1965, various water quality indices have been developed; with each customised based on the relevant water resource in a specific region (Horton, 1965). Most of these indices tend to differ based on the water quality parameters used to develop them, the calculation algorithm, and the scale used to rate the water quality (Tyagi et al., 2013; Feng et al., 2016, Malek et al., 2022, Uddin et al., Uddin et al., 2023c). These indices simplify complex water quality data for political decision-makers, water resource managers who are not technically inclined, and the public (Mladenovic-Ranisavljevic and Žerajic, 2017). Beyond simplifying complex water quality data, WQIs have been used as vital tools to gain knowledge about pollutants and their transport processes and predict the quality of water resources (Kumarasamy and Macholo, 2018).

Although there is not any globally standardised or harmonised methodology for developing a WQI (Sutadian et al., 2016), water quality indices usage in the assessment of surface water has been well documented (Banerjee and Srivastava, 2009; Alobaidy et al., 2010; Massoud, 2012; Sutadian et al., 2016; Sener et al., 2017). However, a limited number of systematic reviews on the development and use of WQIs have so far been published and are available in the open literature and they include those that deal with exploring the different types of WQIs (Gupta and Gupta, 2021; Uddin et al., 2021), assessing steps in their development, advantages (Sutadian et al., 2016), and disadvantages (Uddin et al., 2021; Chidiac et al., 2023). For example, Lumb et al. (2011) reviewed other WQI models developed between 1960 and 2010, where they demonstrated the importance of the steps used in the formulation of the WQI. In addition, Sutadian et al. (2016) reviewed 30 models of WQIs, the country where they were developed and applied, and more recently, a review of the accuracy of commonly used WQIs

by Uddin et al., (2021). The available WQIs used to evaluate surface water quality are based on physical and chemical parameters and very few microbial pathogens such as *Escherichia coli*, faecal coliforms, and total coliforms.

Notwithstanding, the introduction and wide application of machine learning tools have significantly evolved the architecture of the modern WQI models (Gazzaz et al., 2012, Najafsadeh et al., 2021; Malek et al., 2022; Uddin et al., 2022b). Many WQI models are designed for a specific domain of application such as the Oregon index (Dunnette, 1979), NSF index, Malaysian index (DoEM, 2002), West Java index (Sutadian et al., 2018), etc., and precisely to address a specific regional water quality problem such as the Irish Water Quality Index (IEWQI) (Uddin et al., 2023a). Although selected WQI models have been customised for application in other regions such as the NSF and the SRDD, the domain of application has always been the same. However, the success of the IEWQI especially with reducing uncertainty while being applied in multiple domains, should serve as a useful benchmark for future WQI model developers. The present study for this thesis, seeks to investigate the possibility of applying one WQI model to assess both lotic and lentic systems. In addition, the authors would like to present workable solutions to allow the seamless application of WQI models in both lotic and lentic systems with significant efficacy.

2.6. Method and approach of review

The current study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009), which is an evidence-based system composed of a set of items for reporting systematic reviews and meta-analyses. By primary design, PRISMA is used to report review studies that evaluate the effects of interventions such as aetiology, prevalence, diagnosis, or prognosis. However, the methodology can also be employed in review studies that have an objective other than that of evaluating interventions. The method is described in full by Page et al. (2021).

2.7. Research questions, inclusion and exclusion criteria for articles, and criteria for articles and models.

WQI models are primarily developed for specific regions and to solve regional water challenges. It is only after it has demonstrated reliable performance and less uncertainty that other regions will attempt to customise and use that model for their own water quality challenges. However, if a model has been optimised to only evaluate the water quality of a river, lake or marine system, the pertinent question is, “can that WQI model be used to assess the quality of both lotic and lentic ecosystems with equal efficacy?”

As part of the article filtering process, the present study devised the inclusion or exclusion criteria for articles found in all accessed databases. The inclusion criteria for relevant articles included articles where WQIs were developed through four stages or steps: parameter selection, the transformation of parameters to a standard scale, the weighting of parameters, and aggregation, and used for the general assessment of water quality. The exclusion criteria involved articles that applied an existing WQI and used a WQI model with a specific water assessment use. Furthermore, only peer-reviewed original articles were included in the final review synthesis.

2.8. Articles search strategy.

The article search parameters were defined to address the objectives of this study. The key phrases for the searches included the following: “water quality index for lentic systems,” “water quality index for lotic systems,” and “water quality index development.” The “AND” Boolean operator was applied to all the search phrases to narrow the search results. Five article databases ((Springer: <https://link.springer.com/>) (Date accessed: 08/07/2022); (MDPI: <https://www.mdpi.com/>) (Date accessed: 09/07/2022) ; (Scopus: <https://www.scopus.com/>) (Date accessed: 12/07/2022); (Taylor and Francis Online: <https://www.tandfonline.com/>) (Date accessed: 15/07/2022); and (Google Scholar: <https://scholar.google.com/>) (Date accessed: 16/07/2022)) were used for article search. The following Fig. 2.1 is a representation of the article filtering process that ends with 17 articles used for the final synthesis. These 17 articles are also presented in detail in Table 2.1.

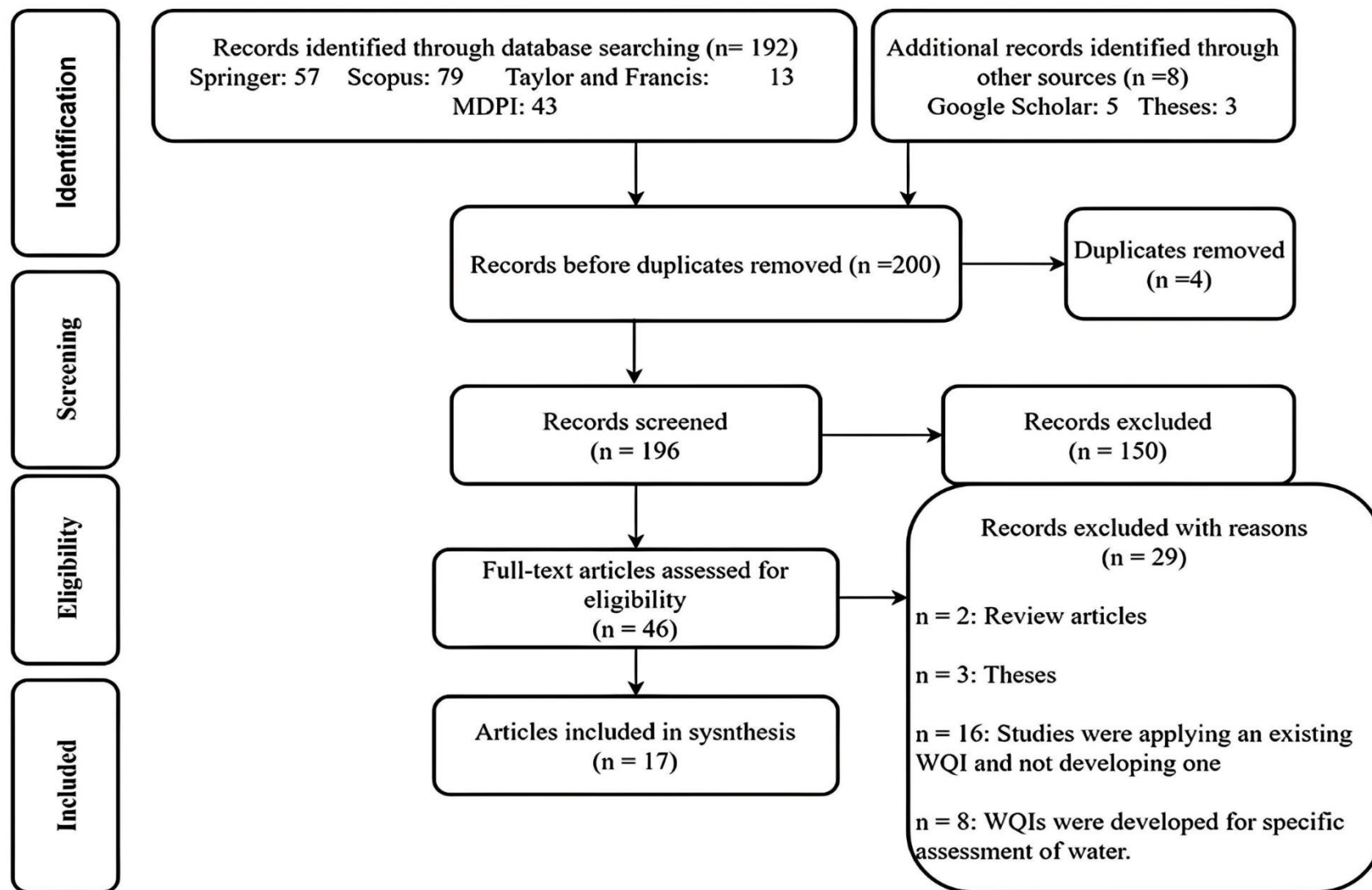


Fig. 2.1 PRISMA flow diagram of searching, screening, and article selection

Table 2.1: Overview of included studies, appraisal of the stages of development, system of application, and where and when studies were first published

Name of WQI	Selected parameters and method of selection	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
Liou's Index (Liou et al., 2004)	Literature & Principal Component Analysis (PCA) 9 parameters: DO, BOD ₅ , ammonia nitrogen, SS, turbidity, FC, Temperature, toxicity, & pH.	Parameters are taken directly as sub-indices using permissible limits	PCA	Geometric mean	Lotic systems	Taiwan
Pesce Index (Pesce and Wunderlin, 2000)	GEMS/Water UNEP Program recommendations 20 parameters: Ammonia, BOD ₅ , Calcium, Chloride, COD, DO, Hardness, Magnesium, Nitrates, Nitrites, oil and grease, pH, Orthophosphate, Solids: dissolved, Solids: total, sulphates, surfactants, temperature, total coliforms, & turbidity.	Ranking using permissible limits	Researchers' experience	Arithmetic Mean	Lotic systems	Argentina

Table 2.1 (continued)

Name of WQI	Parameter Selection method	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
Hanh Index (Hanh et al., 2011)	PCA 8 parameters: DO, turbidity, SS, COD, BOD ₅ , ammonium nitrogen, Orthophosphate, & total coliforms.	Piecewise linear membership rating functions	No weighting	Linear product power & linear sum power.	Lotic and lentic systems	Vietnam
Indian Pollution Index (Sargaonkar & Deshpande, 2003)	Indian water resources quality objectives. 13 parameters: Turbidity, pH, colour, DO, BOD ₅ , TDS, hardness, Chloride, Sulfate, nitrate, total coliforms, Arsenic, & Fluoride.	Rating curves with permissible limits	No weighting	Arithmetic mean	Lotic systems	India
Prati Index (Prati et al., 1971)	Through Literature 13 parameters: pH, DO, BOD, COD, SS, ammonia, nitrates, Chlorine, Iron, Mn, alkyl benzene sulphonates & carbon chloroform extract.	Linear and parabolic functions	No weighting	Arithmetic mean	Lotic systems	Italy

Table 2.1 (continued)

Name of WQI	Parameter Selection method	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
National Sanitation Foundation (NSF) (Brown et al., 1970)	Delphi Technique 11 parameters: DO, pH, BOD ₅ , temperature, total phosphorus, nitrates, turbidity, total solids, pesticides, toxicity, & faecal coliforms.	Delphi Technique	Delphi Technique	Originally Arithmetic but later changed to Multiplicative (Brown et al., 1973)	Lotic systems	USA
Canadian Council of Ministers of the Environment (CCME) Index (CCME, 2001)	Not set, can be customised by the user	Standard guidelines or water resource quality objectives.	No weighting	Sum root of squares	Lotic and lentic systems	Canada
The Scottish Research Development Department (SRDD) Index (SRDD, 1976)	Delphi Technique 10 parameters: DO, BOD ₅ , pH, phosphate, SS, temperature, conductivity, total oxidised, free, and saline ammonia, & <i>E. coli</i>	Delphi Technique	Delphi Technique	Arithmetic mean	Lotic systems	Scotland
Malaysian index (DoEM, 2002)	Expert panel opinion 6 parameters: pH, DO, COD, Ammoniacal Nitrogen, SS, & BOD	Parameters are taken directly as sub-indices.	Unequal weighting through expert opinions	Arithmetic mean	Lentic systems	Malaysia

Table 2.1 (continued)

Name of WQI	Parameter Selection method	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
Hallock Index (Hallock, 2002)	Delphi Technique 8 parameters: DO, pH, total nitrogen, total phosphorus, total suspended solids, turbidity, & faecal coliforms.	Rating curves	Delphi Technique	Arithmetic mean	Lotic systems	Washington State
Said's index (Said et al., 2004)	Environmental importance 5 parameters: DO, turbidity, Total phosphorus, specific conductivity, & faecal coliforms	Parameters are taken directly as sub-indices	Not required	Mathematical function	Lotic systems	Florida State
Universal Water Quality index (UWQI) (Banda & Kumarasamy, 2020)	Expert opinion 13 parameters: Ammonia, Calcium, Chloride, Chlorophyll a, Conductivity, Fluoride, Hardness, Magnesium, Manganese, Nitrate, pondus Hydrogenium, Sulphate, & Turbidity	Parameter relative environmental importance	Target water quality ranges (TWQR)	Weighted arithmetic	Lotic systems	South Africa

Table 2.1 (continued)

Name of WQI	Parameter Selection method	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
West-Java WQI (Sutadian et al., 2018)	Availability of monitoring data against standards; Statistical assessment. 13 parameters: Temperature, SS, COD, DO, nitrite, total phosphate, detergents, phenols, chloride, Zinc, Lead, Mercury, & faecal coliforms.	Linear scaling for temperature & mathematical function for other parameters	Evaluation of expert opinion with Analytical Hierarchy Process (AHP)	Multiplicative	Lotic systems	Indonesia
Dinius Index (Dinius, 1987)	Delphi Technique 12 parameters: DO, BOD ₅ , total coliforms, <i>E. coli</i> , Alkalinity, Hardness, Chloride, specific conductance, pH, nitrate, temperature, and colour.	Rating functions	Expert opinion	Multiplicative	Lotic systems	Alabama State

Table 2.1 (continued)

Name of WQI	Parameter Selection method	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
House Index (House, 1989)	Interviews with water stakeholders 9 parameters: DO, ammoniacal nitrogen, SS, pH, temperature, BOD ₅ , nitrates, Chlorides, & total coliforms	Rating curves	Questionnaire survey	Arithmetic mean	Lotic systems	United Kingdom
River Ganga Index of Ved Prakash et al. (Abassi & Abassi, 2012)	Delphi technique 4 parameters: DO, BOD, pH, & faecal coliforms	Rating curves from expert opinion data	Relative weight assignment	Weighted Arithmetic mean	Lotic systems	India

Table 2.1 (continued)

Name of WQI	Parameter Selection method	Transformation	Weighting	Aggregation Method	System of application	Country of first Application
Irish Water Quality Index (IEWQI) (Uddin et al., 2023a)	Random forest machine learning algorithm 9 parameters: Salinity, DO, BOD ₅ , pH, water temperature, transparency, total oxidised nitrogen, dissolved inorganic nitrogen, molybdate reactive phosphorus.	Interpolation rescaling functions	Random forest machine learning and rank sum mathematical function	Quadratic mean	Lotic and lentic systems	Ireland

2.9. The typical steps for developing a WQI.

The development of a WQI involves five steps which include:

1. Parameter selection: is the process where water quality parameters or variables are selected for inclusion based on their importance in a specific region. This step involves the use of the Delphi Technique (Mladenović-Ranisavljević and Žerajić, 2017), a panel of experts to give professional judgement (House, 1980), statistical methods (Sutadian et al., 2018; Guo et al., 2022; Parween et al., 2022), and machine learning techniques (Uddin et al., 2023a).
2. Transformation to a standard scale includes converting the various parameter units to unitless sub-indices. Various methods have been applied among existing WQI models which involve the use of rating functions (Fathi et al., 2022), interpolated functions (Parween et al., 2022; Uddin et al., 2022a, 2023a), expert opinions (Dinius, 1987), and taking parameter concentrations directly as subindex values (Liou et al., 2004; Said et al., 2004)
3. Establishing parameter weights involves assigning weightage to the parameters based on their overall relative importance to the assessment (Uddin et al., 2023a). Various methods exist in literature including subjective techniques such as Delphi technique (Dadollahi-Sohrab et al., 2012), expert opinions (Sutadian et al., 2018), mathematical functions such as the rank sum technique (Uddin et al., 2023a) and statistical methods such as AHP (Sutadian et al., 2018);
4. Aggregating the transformed parameters to produce the final index value includes computing a single comprehensive index value by combining the sub-index and assigned weightings. Different aggregation methods exist in literature and include three main categories: additive or arithmetic (Horton, 1965; Prati et al., 1971; House, 1989), multiplicative (Asadollah et al., 2021; Parween et al., 2022), and logical or a combination of arithmetic and multiplicative (Almeida et al., 2012, Dadollahi et al., 2012).
5. Water quality index score and classification schemes. These schemes consist of five classification categories (Horton, 1965; Brown, 1970; CCME, 2001, Sutadian et al., 2018) and sometimes less (DoEM, 2002; Uddin et al., 2022a) or more categories (SRDD, 1976).

These steps are depicted in Fig. 2.2 for further clarity (Abbasi and Abbasi, 2012; Sutadian et al., 2018; Uddin et al., 2021, Uddin et al., 2023a).

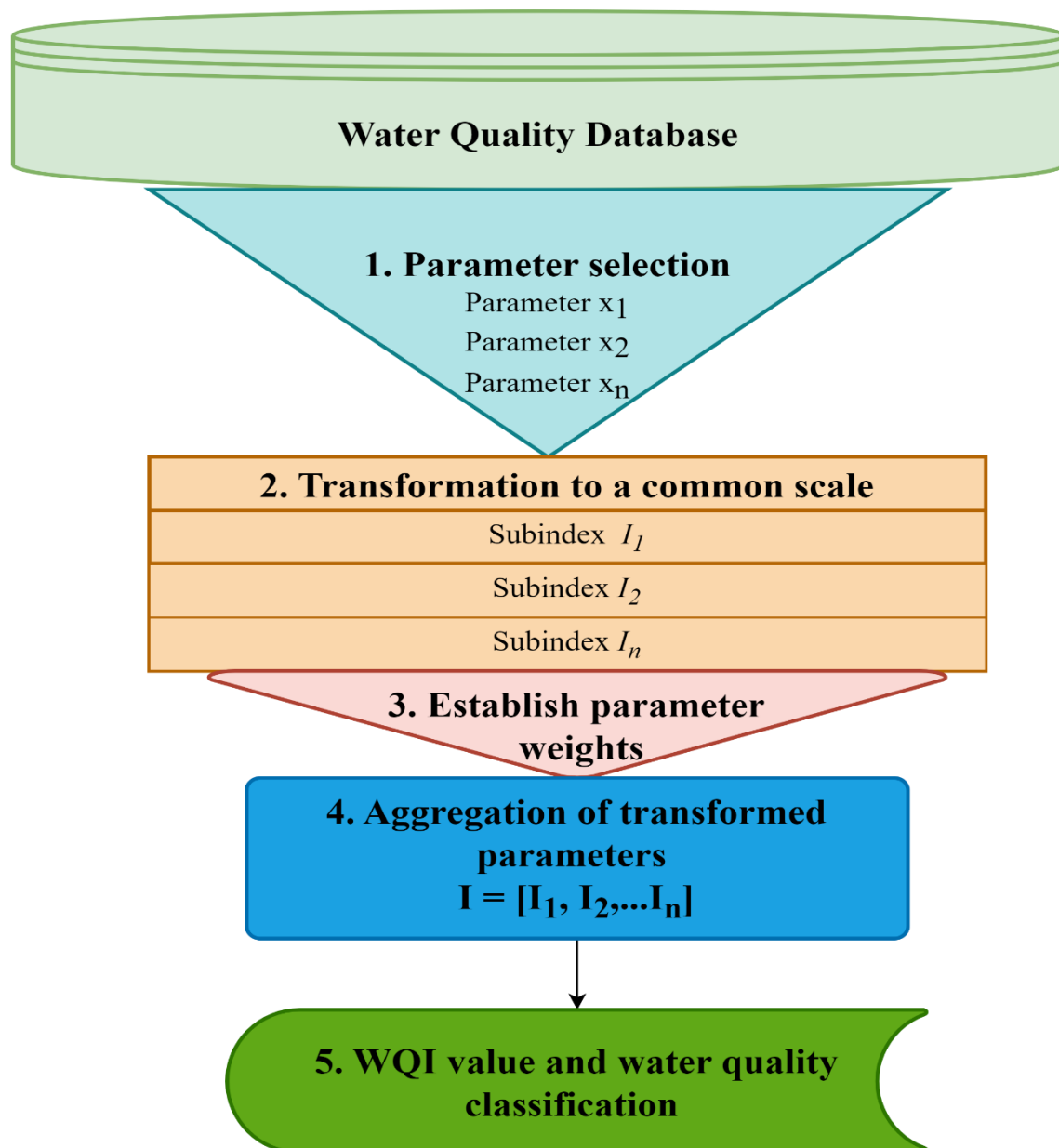


Fig. 2.2 The general WQI development steps

The development steps are further discussed in sequence. This discussion is followed by an in-depth look at some of the most used WQIs.

i. Parameter Selection

The consensus is that it is impossible to continuously monitor all water quality parameters because of financial implications and time. As such, the most critical parameters are considered instead. This stage is the most challenging because omitting parameters may lead to a loss of information and misinformation about water quality (Rangeti et al., 2015). This stage is also prone to subjectivity and uncertainty (Sutadian et al., 2016; Uddin et al., 2021, 2023a). It is therefore recommended that the opinions of local water quality stakeholders, experts, and government officials be considered.

The original WQI by Horton (1965) used a committee of experts to deliberate whether a particular variable could be added. This was later criticised by Joung et al. (1979) because experts disagreed on the relevance and relative significance of parameters. These disagreements opened loopholes for criticism by another committee of experts. The Delphi Technique is another method that considers the expert contributions, only there is no debate, but a series of questionnaires sent to the participants without them converging in one area (Delbecq et al., 1975). As Lohani and Todino (1984) argued, these approaches of parameter selection are highly subjective because their professional backgrounds prejudiced the judgment of these experts. This led to the suggestion of complex statistical methods such as factor analysis (FA) and PCA, which are more robust and eliminate the compelling issue of biases (Lohani and Todino, 1984; Jolliffe, 2005; Rangeti et al., 2015; Sutadian et al., 2018). Despite the accuracy of these statistical methods (Kumar et al., 2019; Ma et al., 2020; Chakravarty and Gupta, 2021; Parween et al., 2022), their adoption for use has been less because of the complex statistics they come with. Many studies (Medeiros et al., 2017; Sutadian et al., 2018) have continued to use expert judgment for parameter selection.

Both FA and PCA have been recommended for use in modelling studies (May et al., 2011), especially with intelligent learning systems such as artificial neural networks (ANN) (Singh et al., 2009; Rangeti et al., 2015). However, recent research has reported that these existing methods contribute significantly to the unreliability of a model and the inappropriateness of the selected parameters (Uddin et al., 2021; 2023b). In a recent study, the authors compared different methods for optimising parameter selection which included filter, wrapper, and embedded methods. The authors reported that embedded-based methods such as random forest and extra tree and filter-based mutual information methods outperform the commonly used filter-based methods such as PCA and Pearson correlation. In addition, the authors reported

that the usage of these methods helped to improve model performance by reducing model uncertainty due to less robust parameter selection methods (Uddin et al., 2023b).

Among the 17 WQIs assessed, the number of selected parameters was different. The Pesce index had the highest number of parameters (20), while the river Ganga index of Ved Prakash et al. had only four parameters (Fig. 2.3) The CCME requires a minimum of four parameters, but the number is user dependent. Fig. 2.3 presents a summary of the number of parameters used for each analysed WQI model and Fig. 2.4 illustrates the frequency of use for each parameter among the 17 studied WQI models.

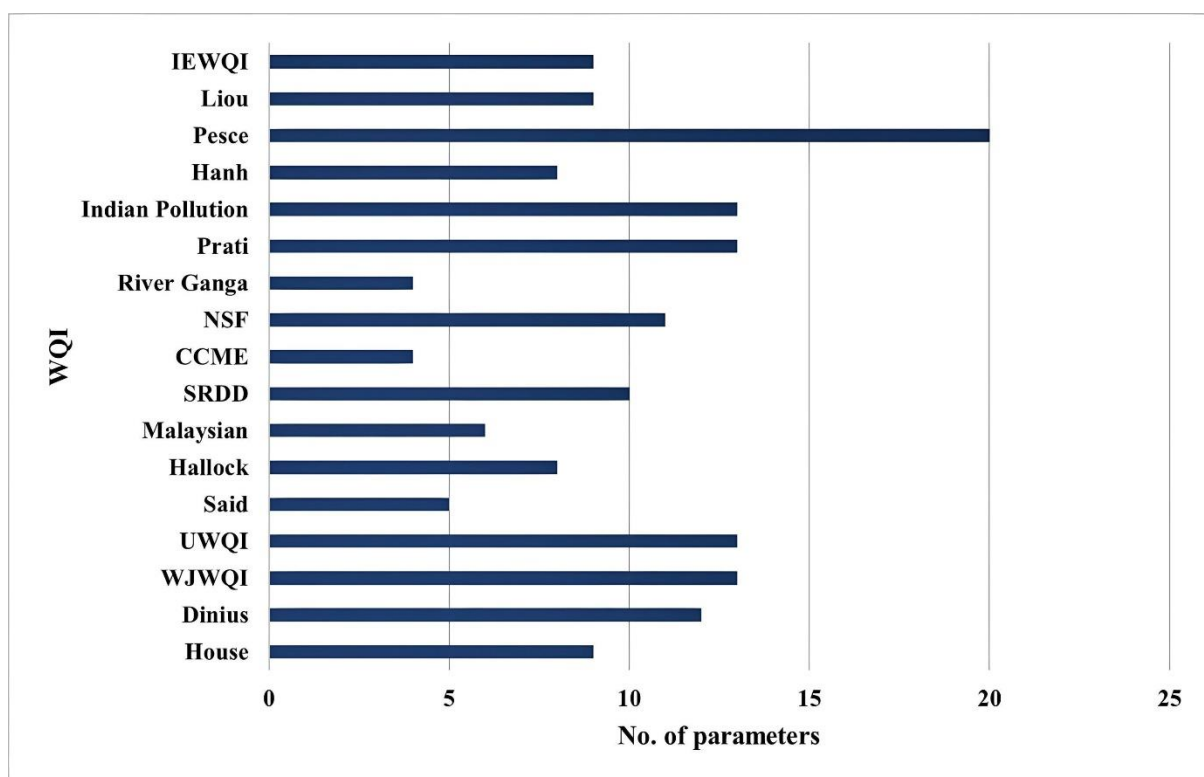


Fig. 2.3 Summary of the number of parameters used per WQI included in the study

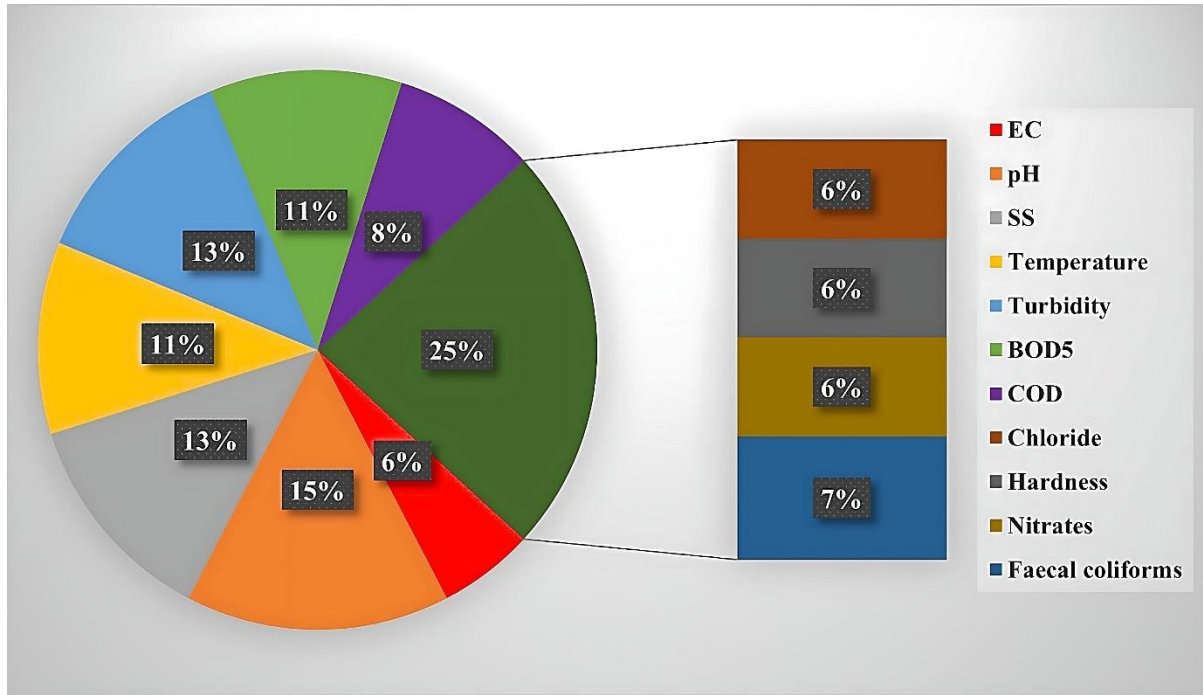


Fig. 2.4 Percentage of the frequency of use for individual water quality parameters in all selected WQIs

Water quality is a collective term for water's physical, chemical, and biological quality. As such, any assessment of water quality must include all three categories of water quality. Among the analysed WQIs, physical parameters were mainly used (56%) as compared to chemical parameters (37%) and biological parameters (7%) (Fig 2.4). This denotes that some WQIs do not include all categories of water quality.

ii. Transformation of parameters to a standard scale

The part of parameter transformation aims to convert the different selected parameters to a single dimensionless scale. This is important because their different scales will only be possible to aggregate the parameters together. For example, turbidity is presented in nephelometric turbidity units (NTU), while Manganese (Mn) and Iron (Fe) in milligrams per litre (mg/L) (Rangeti et al., 2015; Sutadian et al., 2016). Moreover, different parameters have different ranges of acceptable fit-for-use standards. For instance, dissolved oxygen (DO) seldom gets readings beyond 0-12 mg/L, whereas Sodium can be between 0-1000 mg/L (Abbasi and Abbasi, 2012). In essence, different parameters have different impacts concerning concentration. Parameter transformation eliminates the units of the various parameters and produces a new scale that is without dimension but two end-points (Ott, 1978; Dunnette, 1979),

one with the lowest endpoint representing unacceptable values and the other with the highest end and representing acceptable quality (Richardson, 1997). While this step is crucial for aggregation, a few WQI models, such as the CCME, do not consider transforming parameters but use multivariate statistical procedures to aggregate the actual values of the parameters. Said et al. (2004) also developed a new WQI and proposed a mathematical equation for directly standardising the index without generating subindices. To create subindex values, rating functions are developed. These curves are mathematical relationships between a parameter's concentration value and the water quality. Common ways to determine subindex functions include using water quality standards, expert judgment, and statistical methods (Harkins, 1974; Lohani and Todino, 1984, Sutadian et al., 2016; Uddin et al., 2021). Despite the wide use of mathematical functions, many researchers have reported that sub-index functions are a source of model uncertainty (Sutadian et al., 2016; Gupta and Gupta, 2021; Uddin et al., 2021, 2022b, 2023a). As such, a recent study developed a hybrid method that involved the use of linear interpolation rescaling functions with threshold water quality guidelines to ameliorate model uncertainty (Uddin et al., 2022a).

Among the analysed WQIs, rating functions were used (52.9%) for transformation. The Liou, Indian Pollution, and River Ganga indices used various methods. 29.4% of the analysed WQIs used standard permissible limits, while very few indices (17.6%) used expert opinions (Fig. 2.5). This is presented in Fig. 2.5 in sequence which illustrates the frequency of use of transformation methods among the analysed WQIs.

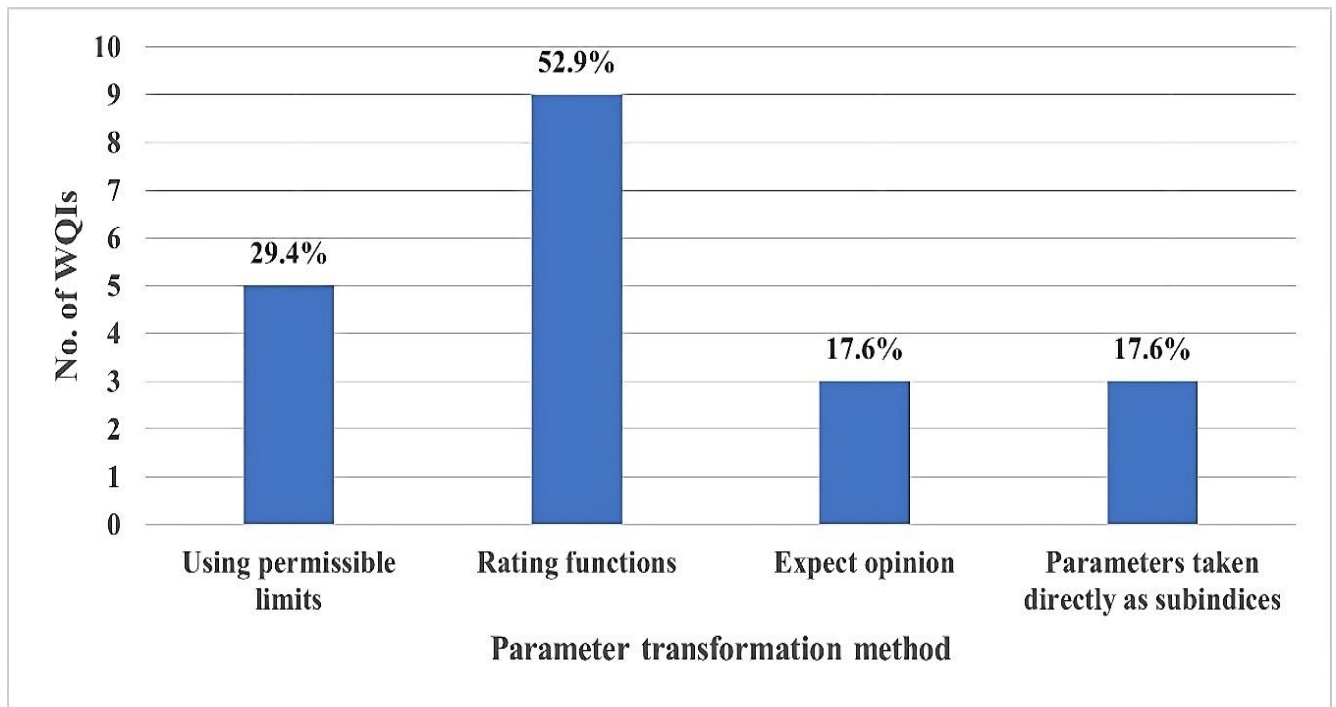


Fig. 2.5 Usage of parameter transformation method among the analysed WQIs

iii. Establishing parameter weights

The weight of a parameter is assigned based on its relative importance and influence on the overall value of the water quality index (Dzwairo et al., 2012). For purposes of credibility, well-formulated techniques are used. To correctly assign a weight, one needs to have knowledge of the parameter, its threshold standard limits, and its normal concentration range in a particular water resource (Kumar and Dua, 2009). If the parameters are equally important, equal weights are assigned, and unequal weights are allocated if the parameters have lesser or greater importance. Sapkal and Valunekar (2013) considered, along with the latter, treatment methods of parameters. For example, if a parameter requires advanced treatment methods for removal, a lower weighting is assigned, and a higher weighting is allocated if the parameter requires conventional treatment. Other methods to assign parameter weights include participatory-based approaches, which may involve water quality experts or managers, policymakers, environmental protection agencies, and other key stakeholders. However, the Delphi method and analytical hierarchy process (AHP) have been widely used (Kodikara et al., 2010, Sutadian et al., 2018). The AHP uses pairwise comparison matrices where the respondents specify their preferences based on other choices. In recent years, studies have reported these methods to be significant sources of uncertainty due to inappropriate weight

estimation (Uddin et al., 2021; Uddin et al., 2022a). Uddin et al. (2021) went further to document the sources of eclipsing and uncertainty for different WQI models. As such, different robust techniques such as the rank sum method (Uddin et al., 2022c), random forest machine learning (Uddin et al., 2023a), extreme gradient boosting (XGB) machine learning (Uddin et al., 2022a) have been developed to ameliorate uncertainty in WQI models. In addition, recent studies have used machine learning approaches such as XGB, SVM, LSBoost, and DNN to estimate model performance with regards to uncertainty (Uddin et al., 2023b)

iv. Aggregation of transformed parameters to produce the final index.

Most aggregation methods are possible and have been applied to obtain a single value representative of the overall quality of water (Abassi and Abassi, 2012; Fu and Wang, 2012). This is the most crucial step in the whole procedure because of the potential loss of information and data distortions such as ambiguity, eclipsing, and rigidity (Ball and Church, 1980; Couillard and Lefebvre, 1985, Abassi and Abassi, 2012). Ambiguity occurs when the overall index value is above the limit value, when none of the considered individual parameter scores does not exceed the limit. At the same time, eclipsing occurs when the overall index does not exceed standard limits, but one or more of the considered individual parameters exceed the set limits. Rigidity, however, occurs when a need arises to add more parameters to an existing index to address new water quality concerns and the model does not allow such addition.

There are three most common categories of aggregation methods: (i). additive or arithmetic, (ii). Multiplicative or geometric, and (iii). logical. These are documented by (Abassi and Abassi, 2012; Uddin et al., 2021). Additive or arithmetic aggregation methods involve combining the transformed parameters through summation. This method characterised the early days of WQI models (Horton, 1965; Brown et al., 1970; Prati et al., 1971; Ott, 1978). The weighted arithmetic mean is the most used additive aggregative method. Multiplicative or geometric aggregation methods, such as the indices of Walski & Parker (1974) and Dinius (1987), combine the transformed parameters through product operation, while logical techniques, such as the Smith index, combine the transformed parameters through a logical operation such as minimum and maximum (Smith, 1990). The additive and multiplicative aggregation methods have been identified as sources of eclipsing (Uddin et al., 2021). However, the logical aggregation approach was developed specifically to solve the eclipsing

problem (Smith, 1990). The eclipsing data distortion during aggregation results in the overall index value overestimation or underestimation as observed by Uddin et al., (2022a; 2023a). Moreover, studies have achieved success using machine learning approaches such as ANN (Gazazz et al., 2012; Uddin et al., 2022b).

Most (58.8%) of the analysed WQIs used additive methods, while 41.2% used multiplicative methods (Fig. 2.6) to resolve the final index.

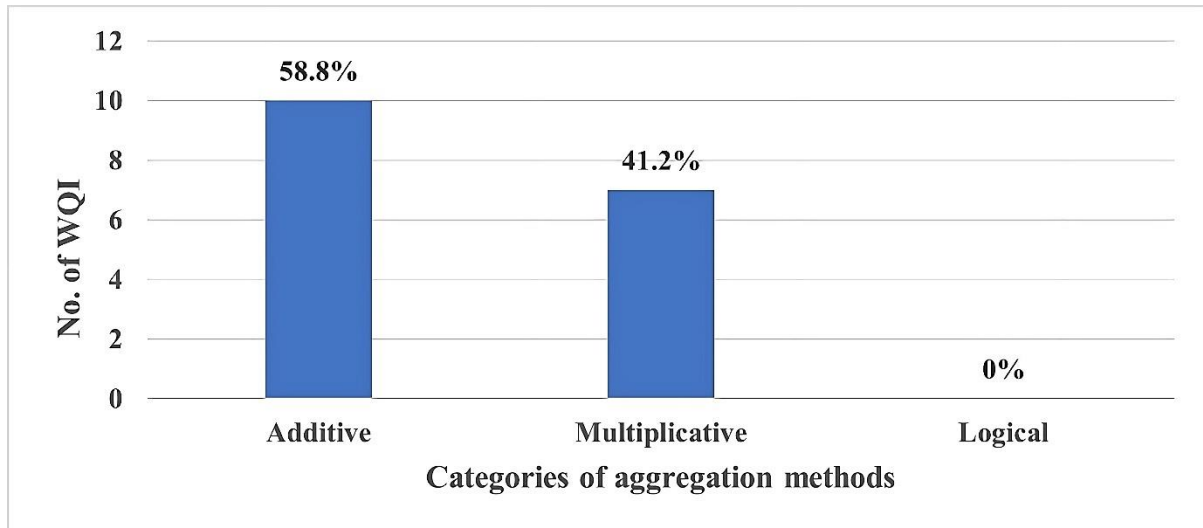


Fig. 2.6 Aggregation category usage among the analysed WQIs

Categories of aggregation methods

i. Additive

The additive method is the most common one recorded in the available literature. The method involves using summation to combine the transformed values of the parameters to produce the index value (Abbasi and Abbasi, 2012). Indices of general water quality assessment using the additive method include Prati Index (Prati, 1971) and Pesce Index (Pesce, 2000). One of the most used additive methods is the weighted arithmetic mean. The weighted arithmetic mean measures the central tendency of a set of observational data when not all observations have the same importance. This method has been used to aggregate transformed parameters of existing water quality indices, such as that of Brown et al. (1970), Dunnette (1979), and Sargaokar and Deshpande (2003). It has been recognized for offering simplicity, where the final value of the index is calculated by summing the weighted transformed parameters. Although this method deals away with eclipsing (Rangeti et al., 2015), it has been criticised for lacking sensitivity

(Liou et al., 2004; Juwana et al., 2012; Sutadian et al., 2016). An example of this aggregation method is shown in Equation 1:

$$WQI = \sum_{i=1}^n Q_i \cdot W_i \quad (2.1)$$

Where Q_i = subindex/transformed parameter i , n = number of transformed parameters and W_i = weight of the transformed parameter.

ii. Multiplicative

The multiplicative aggregation method involves combining the sub-index values through a product operation. In this category, the weighted geometric mean is the most used method. Indices of Dinius (Dinius 1987), Walski and Parker (Walski and Parker, 1974), Liou (Liou et al., 2004), and the SRDD (SRDD, 1976) which are for the general assessment of water quality have employed this model. The weighted geometric mean, in comparison to the weighted arithmetic mean, is more viable and unbiased (Landwehr and Deininger, 1974; Joung et al., 1979) and has been used as an alternative for many studies (McClelland, 1974; Waski and Parker, 1974; Almeida et al., 2012). This method has since been adopted by the National Sanitation Foundation of the United States, commonly known as the NSF-WQI. In the geometric mean function illustrated in equation 2, the final index is zero if any one sub-index is zero. This characteristic helps to eliminate the eclipsing and ambiguity problem (Liou et al., 2004).

$$WQI = \prod_{i=1}^n Q_i^{W_i} \quad (2.2)$$

Where Q_i = subindex/transformed parameter i , n = number of transformed parameters and W_i = weight of the transformed parameter.

iii. Logical

The logical aggregation method involves combining the subindices using logical operators. The most common logical operators are the minimum and maximum operators, notably used in the Smith index (Smith, 1990).

Minimum operator function

The minimum operator is an aggregation function which avoids eclipses and ambiguity in the final index by using the lowest subindex values to produce the index value. This function was initially created and applied in New Zealand by Smith (1990) to assess the water quality of lotic systems. However, it was used to evaluate surface water quality in India (Shah and Joshi, 2015). The mathematical expression of the function is given below:

$$WQI = \text{Min} (S_i, S_{i+1}, S_{i+2}, \dots, S_{\text{sub}n}) \quad (2.3)$$

Where S_i is the subindex value for the i th parameter and n is the number of subindices.

Maximum operator function

Unlike the minimum operator, the maximum operator aggregation function performs the summation of subindices in an increasing scale manner. None of the published WQI models have used this function for aggregation. However, it has been reported to be suited to applications where an index must report if any of the recommended limits are violated (Abassi and Abassi, 2012). The general function of the maximum operator is presented in equation 4:

$$I = \max (I_1, I_2, \dots, I_n) \quad (2.4)$$

In the maximum operator function, I assume the largest of the subindex values, and $I = 0$ if $I_i = 0$ for all i

i. WQI Classification Schemes

The final stage of WQI development involves classifying or categorising water quality based on the index value. These quality classes group the water quality status of water into categories such as “excellent,” “good,” “fair,” and “poor”. However, different WQIs use different classification schemes with varying interpretations for the same water quality parameters. Like the other stages in the WQI development, this stage also presents its own distortions that primarily stems out from what has been recently adopted as the “metaphoring problem” (Uddin et al., 2023b). The “metaphoring problem” denotes that different WQI models employ different classification schemes to explain the WQI value. According to Uddin et al. (2021), the traditional WQI model does not express the actual state of water quality because of the use of

various classification schemes, thus prompting model uncertainty and unreliability. Uddin et al. (2022a) advances further ideas that the current WQI model provides ambiguous information to end users of the model and impedes them from responding as quickly as required. Fig 2.7 gives an illustration of the different classification schemes used among the 17 WQI model, with five classes being the most (47.1%) used, albeit with varying interpretations in selected instances.

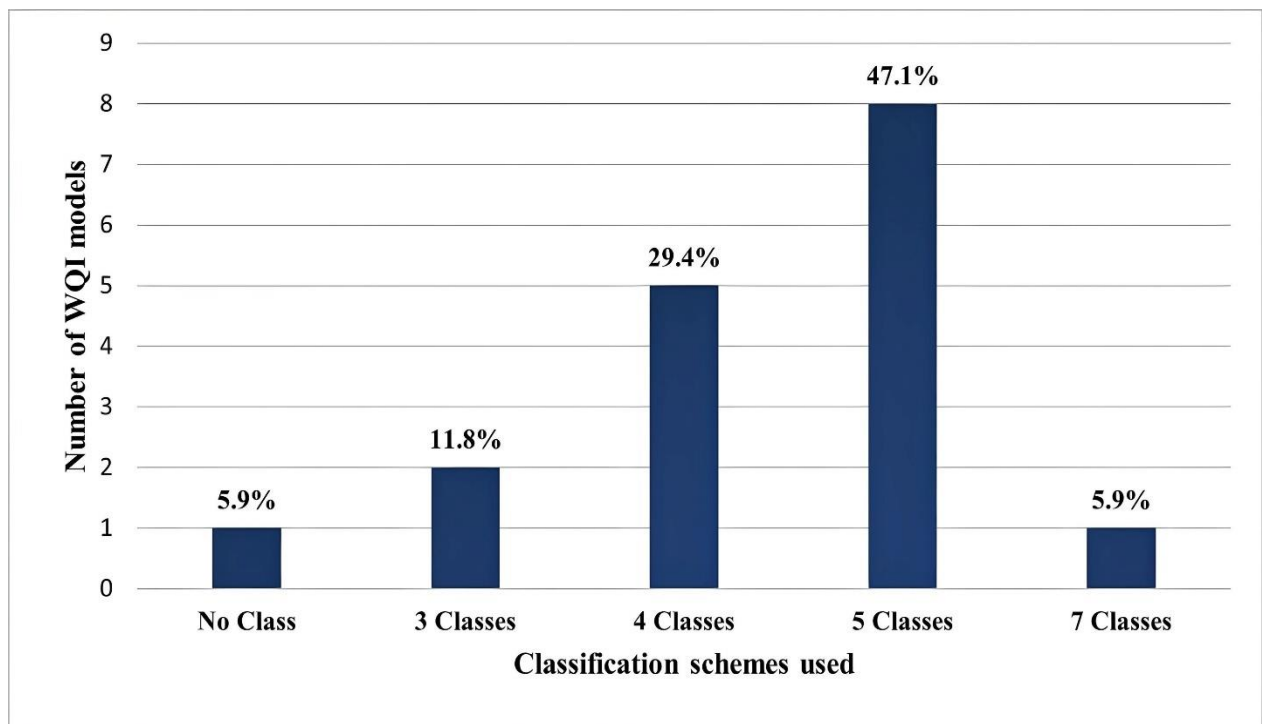


Fig. 2.7 Different classification schemes used among the WQI models to explain the “metaphoring problem” contributing to WQI uncertainty

It is against this backdrop that recent studies have used more robust, reliable, and precise machine learning (ML) techniques such as K-Nearest Neighbour (KNN), artificial neural networks (ANN), Decision Tree (DTs), Gaussian Naïve Bayes (GNB), subversion (SVN), Random Forest (RF), and eXtreme Gradient Boosting (XGB) (Najafzadeh and Niazmardi, 2021; Malek et al., 2022, Uddin et al., 2023c). ML classifiers have been effectively used in the recent past to predict the correct classification when assessing water quality (Shakhari and Banerjee, 2019; Najafzadeh and Niazmardi, 2021). In addition, it has recently been proven that the use of XGB yields the most accurate, precise, and specific water quality classification schemes (Malek et al., 2022). Furthermore, a recent study piloted in coastal waters evaluated

the performance of ML classifiers such as GNB, SVN, KNN and XGB. The authors reported that XGB yielded the most accurate classification for most water quality classes except for the “poor” class (Uddin et al., 2023c).

2.10. Selected common WQI used for evaluation of lotic and lentic ecosystems.

Although there are many WQI models developed over the years, some of them are modifications or derivations of the original ones. For example, the NSF was derived from the original Horton index, while the Hahn index was derived from the Said index, and the CCME index was modified from the British Columbia WQI. However, in very few instances you will find the original index being more commonly used than the modified index. This may be attributed to the fact that modified or derived index models always come with an upgrade or addresses a fallacy of the original model. The following section discusses some of the commonly used WQIs to evaluate water quality of lotic and lentic systems.

2.10.1. Canadian Council of Ministers of the Environment Index (CCME) index

The CCME index was initially introduced in Canada in the 1990s as the British Columbia Water Quality Index and used both as a water quality public communication tool and to identify watersheds for priority action. The index was modified and endorsed by the Canadian Council of Ministers of the Environment (CCME) in 2001 and referred to as the CCME WQI (CCME, 2001). Since then, the CCME index has been used in most studies conducted outside Canada in Turkey, India, Iran, and Albania (Boyacioglu, 2010; Sharma and Kansal, 2011; Damo and Icka, 2013; Moustafei, 2014) to evaluate the quality of both lentic and lotic ecosystems (Davies, 2006; Giriappanavar and Patil, 2013; Mostafaei et al., 2014).

i. Parameter selection

The CCME was designed to be flexible regarding the number of parameters included in the calculation of the final WQI. It, however, requires that minimum of four parameters. This was to accommodate for easy modification to suit local conditions. The CCME does not have a specific parameter selection method, meaning the user can decide on the process according to their needs.

ii. *CCME Transformation to a standard scale*

The CCME WQI does not standardise the parameters to a common scale. Instead, it uses standard guidelines or water resource quality objectives.

iii. *Establishing parameter weights*

The weighting process is not conducted because there are no sub-indices.

iv. *Aggregation*

Aggregation is conducted through a sum root of squares mathematical equation (Equation 10)

The aggregation function for the CCME is based on scope, frequency, and amplitude, denoted by F_1 , F_2 , and F_3 , respectively. It requires that all the parameters be standardised, and the three factors calculated. Scope (F_1) refers to the percentage number of parameters that exceed freshwater ecosystem guidelines and is calculated using the equation below.

$$F_1 = \left(\frac{\text{Number of failed parameters}}{\text{Total number of parameters}} \right) \times 100 \quad (2.5)$$

Frequency (F_2) is the percentage number of tests for each parameter that does not comply with the guidelines. This is calculated using Equation 5.

$$F_2 = \left(\frac{\text{Number of failed tests}}{\text{Total number of tests}} \right) \times 100 \quad (2.6)$$

Amplitude (F_3) is the extent to which the failed tests exceed the limit guidelines. Unlike F_1 and F_2 , the calculation of F_3 involves two steps that determine the excursion and the normalised sum of the excursion (nse). The excursion refers to the number of times a parameter's test value is greater than that parameter's objective. This is calculated using Equation 6. The nse is calculated by adding all the excursions from individual tests from their objectives and dividing them by the total number of tests using Equation 7. The amplitude (F_3) is then calculated using Equation 8.

$$\text{Excursion}_i = \left\lceil \frac{\text{Failed test value}_i}{\text{Objective}_i} \right\rceil - 1 \quad (2.7)$$

$$\text{nse} = \left(\frac{\sum_{i=1}^n \text{excursions}}{\text{number of tests}} \right) \quad (2.8)$$

$$F_3 = \left\lceil \frac{\text{nse}}{0.01\text{nse} + 0.001} \right\rceil \quad (2.9)$$

The final CCME index is then calculated using the sum root of the squares of all the factors with Equation 9.

$$\text{CCME WQI} = 100 - \left(\frac{\sqrt{(F_1)^2 + (F_2)^2 + (F_3)^2}}{1.732} \right) \quad (2.10)$$

In this index, the constant value of 1.732 is used to normalize the resultant values to a 0-100 range where zero (0) depicts poor quality and 100, the best quality of water (Lumb et al., 2011; Rangeti et al., 2015; Sutadian et al., 2016).

The final index is interpreted with value ranges between 0 to 100, with five water quality classes: poor quality: 0-44; marginal quality: 45-64; fair quality: 65-79; good quality: 80-94; and excellent quality: 95-100.

The scale of application of the CCME WQI has been used in lotic (CCME, 2001) and lentic ecosystems (Giriyappanavar and Patil, 2013).

2.10.2. Oregon WQI (OWQI)

The OWQI was created to express ambient water quality for recreational uses in 1979 (Dunnette, 1979). The index was discontinued in 1983 because it needed to be more cost-effective (Sutadian et al., 2016). However, Cude (2001) updated the index to interpret the overall quality of water and to communicate the water quality status of Oregon rivers.

i. Parameter Selection

The selection of parameters involved a four-stage elimination process. Stage one involved a thorough literature review of previous WQIs, which accumulated 90 parameters. Stage two involved using rejection criteria: data availability, the significance of parameters, and not being present in harmful concentrations. Stage two reduced the parameters from 90 to 30. Stage three involved conducting the Delphi method but only through the Oregon Department of Environmental Quality (ODEQ) members as respondents. This process trimmed down the parameters to 14. The last stage involved conducting another rejection test called redundancy test and impairment categories which reduced the parameters to six, after which two (total phosphorus and temperature) were added in 2001 to a total of eight (Sutadian et al., 2016)

based on a better understanding of the importance of these parameters to the streams of Oregon (Cude, 2001).

ii. Transformation to a standard scale

The current version of the OWQI uses non-linear regression rating curves. In this step, the parameter measurements for each sub-index are converted to a relative quality rating between 10 (worst case) and 100 (ideal).

iii. Establishing weights

Parameter weights were initially assigned using the Delphi technique on the six parameters. Unequal weights were set as follows: DO (0.4), FC (0.2), pH (0.1), Ammonia+Nitrate Nitrogen (0.1), TS (0.1), and BOD (0.1). However, in the Cude (2001) update, it was argued that unequal weights are only suitable for WQIs, which are developed for a specific use, where some parameters play more important roles than others. Equal weights assignment was then adopted for this index.

iv. Aggregation

The original aggregation of OWQI was through the weighted arithmetic mean (Equation 1). However, this aggregation method experienced some eclipsing problems, so the updated version adopted the unweighted harmonic square mathematical formula presented as Equation 11 below:

$$WQI = \sqrt{\frac{n}{\sum_{i=1}^n (Qi)^2}} \quad (2.11)$$

The final interpretation of the OWQI is categorised into five classes, namely: excellent (90-100), good (85-89), fair (80-84), poor (60-79), and extremely poor (10-59).

The scale of application for the OWQI is limited to the Oregon streams, and any attempt to apply this WQI to different waterbody types should be done with caution (Cude, 2001). Therefore, the OWQI was only developed to work in local lotic systems.

2.10.3. National Sanitation Foundation (NSF) WQI

The NSF WQI was developed in the 1970s in the USA in a process that involved over 100 water quality experts throughout the USA (Brown et al., 1970). Due to its credibility, its direct application or modification has been recorded in most places outside the USA (Mojahedi and Attari, 2009; Benvenuti et al., 2015; Fathi et al., 2018).

i. Parameter Selection

The selection of parameters was based on the consensus of the water quality experts using the Delphi technique. A total of nine parameters were selected from a set of 30 frequently measured parameters in the USA. It was later updated by adding pesticides and toxic elements.

ii. Transformation to a standard scale

Sub-index generation for the NSF WQI was done through the Delphi technique. The information acquired from the Delphi technique was later used to provide rating curves that represented the guidelines for the parameter in question (Lumb et al., 2012).

iii. Establishing weights

The Delphi technique applied to decide on the weighting of the selected parameters. The final weights were as follows: DO (0.17), FC (0.16), BOD5 (0.11), Temperature (0.10), NO₃ (0.10), Turbidity (0.08), TS (0.07), pH (0.11), and FC (0.16).

iv. Aggregation

The aggregation method proposed by Brown et al. (1970) took the structure of an additive model, which was later found to be insensitive because one wrong parameter automatically renders the WQI zero. To address this, a multiplicative variation of the NSF WQI was proposed (Brown et al., 1973).

This aggregation method was interpreted as follows: excellent (90-100), good (70-89), medium (50-69), bad (25-49), and very bad (0-24) (Brown and McClelland, 1974).

The scale of application for the NSF WQI was not specified, but the index has been used in lotic systems.

2.10.3. The Scottish Research Development Department (SRDD) Index

The SRDD was initially developed in 1976 for Scotland as a modification or adaptation of the NSF WQI. However, it has been reported in the literature to being in use in other countries to evaluate river basins (Bordalo et al., 2006; Carvalho et al., 2011; Dadolahi-Sohrab et al., 2012).

i. Parameter selection

Following similar methods as the NSF WQI, the SRDD also used the Delphi technique to select ten parameters with the local water experts as the respondents to the questionnaires (SRDD, 1976).

ii. Transformation of parameters

The Delphi technique of respondents' judgment is used to develop the sub-indices of the SSRD index. The possible lowest and highest values of the sub-indices were decided to range between 0 and 100.

iii. Establishing weights

Parameter weights were assigned using the Delphi technique. The weights for each parameter were as follows: DO (0.18), BOD5 (0.15), free and saline ammonia (0.12), pH (0.09), total oxidised nitrogen (0.08), phosphate (0.08), SS (0.07), temperature (0.05), conductivity (0.06) and *E. coli* (0.01).

iv. Aggregation

The final aggregation used in the SRDD was a modified additive method like the one used by Bordalo et al. (2006) and Carvalho et al. (2011). This SRDD index had seven classes of water quality, namely: clean (90-100), good (80-89), good water with some treatment (70-79), tolerable (40-69), polluted (30-39), severely polluted (20-29) and water akin to piggery waste (0-19).

The SRDD was adapted from the NSF WQI, so the scale of application was the same as the NSF WQI. Application of the SRDD in lotic systems from reported literature expands to Thailand (Bordalo et al., 2001), Portugal (Carvalho et al., 2011), and Iran (Dadolahi-Sohrab et al., 2012).

2.10.4. House Index

The House index was developed in the late 1980s by House (1989) as a group of four indices that could be used separately or as a hybrid index when more information on river quality was needed. The first index, called general WQI, was for general water quality assessment to indicate river health. The other three, which are the potable water supply index (PWSI), aquatic toxicity index (ATI), and potable sapidity index (PSI), were developed to assess the suitability of potable water supply and toxicity in aquatic and wildlife populations (House 1989).

i. Parameter selection

Parameter selection was through interviews with stakeholders in the water sector, including water authorities and bulk water supply water boards. The parameters were chosen by collating the information from the interviews, especially those that the stakeholders routinely monitored. To assess river health, the general WQI had nine parameters: DO, ammoniacal nitrogen, BOD₅, suspended solids, NO₃, pH, temperature, Chloride, and total coliforms. The PWSI, on the other hand, had 13 parameters: DO, ammoniacal nitrogen, BOD₅, suspended solids, NO₃, pH, temperature, Chloride, total coliforms, Fluoride, colour, and dissolved iron. The ATI comprised 12 parameters: dissolved copper, polyaromatic hydrocarbons (PAH), and total pesticides. A similar number of parameters were considered for the PSI and included total copper, zinc, cadmium, mercury, lead, chromium, arsenic, cyanide, phenols, total hydrocarbons, PAHs, and total pesticides (House 1989).

ii. Transformation of parameters

Rating curves were the preferred method to transform the parameters to a standard scale. The rating curves were developed using the individual parameter's water quality objectives or compliance criteria. Moreover, when the parameter had two or more quality standards, the median was computed and converted into specific sub-index values (House, 1989; Sutadian et al., 2016).

iii. Establishing weights

Assigning parameter weights to reflect the relative importance of individual parameters was obtained through a questionnaire survey sent to the operational management participants of the involved stakeholders. The participants were, however, only asked to rank the nine parameters which formed the general WQI and the 13 parameters that formed the PWSI. There was no

weight assignment for the ATI and PSI because all the selected parameters had equal importance and were considered harmful to human and aquatic life (House, 1989; Sutadian et al., 2016).

iv. Aggregation

To compute the final index, a modification of the aggregation method was first used by the SRDD in developing the SRDD index (House, 1989). The final index interpretation is divided into four class categories, namely: Class I (71-100), which represents the water of a high quality that is suitable for potable water supply, game fisheries, direct contact recreation, and industrial uses; Class II (51-70) which represents the water of reasonable quality and suitable potable water supply with conventional treatment, fisheries, indirect contact recreation, and most industrial uses; Class III (31-50) which represents water that is generally polluted but useful for potable water supply with advanced treatment, indirect contact sport, and breeding fish population; Class IV which generally indicates badly polluted water and requiring a sizeable investment in treatment infrastructure but can be used for sewage transport and navigation and non-contact recreational activities (House, 1989).

This review could not trace formal reports of the hybrid House index application or applied with modifications in other regions or water types. However, the general WQI has been used elsewhere outside the United Kingdom in river systems (Cavalho et al., 2011). According to recorded literature, the scale of application of the House index has been limited to lotic systems in different regions.

2.11. Critique of WQI

The original architecture of the WQI model by Horton (1965) set the tone for all subsequent WQI models. The methods to select parameters, generate sub-index values, assign parameter weights, aggregate, and determine classification schemes have always carried the data distortions of eclipsing, ambiguity, and ultimately uncertainty. This has led to the development of multiple WQI models across the world, each presenting its own sources of uncertainty, thus perpetuating the lack of a universal WQI model. The following section seeks to appraise the current issues associated with the WQI development and the current research efforts towards creating a more accurate, robust, and acceptable WQI model.

2.11.1. Issues involved in WQI development.

i. Parameter selection

Indices naturally contain fewer data than actual raw data, and they only incorporate variables deemed necessary for a catchment because of limitations of time, resources, and complexity, among others. This may become a problem if certain important variables are left out of the index by mistake because the index will present a contrary picture to the actual water quality (Rangeti et al., 2015; Uddin et al., 2021). For example, according to Zainudin (2010), in Malaysia, they used a WQI which excluded coliform bacteria which is an essential indicator of the microbiological safety of the water.

ii. Lack of universal WQI

Another critical limitation of WQI is the reported disagreements of the same WQI. For instance, for their evaluation of the surface water quality of the Ganges River, Sharma et al. (2014) used two different scales for the same index. These disagreements have been noted in some cases: when the same index is used but other limits for classes are used (Ramakrishnaiah et al., 2009; Yadav et al., 2010); the same index is used, but the number and type of variables differ. These disagreements add to the need for a universal water quality index.

iii. WQI ambiguity and eclipsing.

Ambiguity in WQI arises due to the selection of parameters and their weightings, which can vary between WQI models. Different models can have different parameter sets and weightings based on their specific objectives and geographical locations. As a result, WQI values can be difficult to interpret and compare across different models and locations. (Gupta and Gupta 2021; Uddin et al., 2021). The authors also discuss several approaches that have been proposed to address ambiguity and uncertainty in WQI. These include incorporating stakeholder perspectives into WQI models, conducting sensitivity analyses to test the robustness of different models, and using machine learning techniques to improve the accuracy and predictive power of WQI models (Uddin et al., 2022b; Uddin et al., 2023a)

Eclipsing occurs when a single parameter or group of parameters dominates the calculation of WQI values, causing other parameters to be ignored or given less weight in the final index value. It can also occur due to the subjective selection of parameters and weightings used in WQI models. For example, if a WQI model places greater emphasis on a single parameter such as dissolved oxygen, it may overshadow the impacts of other parameters such as pH or total dissolved solids. This can lead to a misleading assessment of water quality and mask underlying issues (Sutadian et al., 2016; Gupta and Gupta, 2021; Uddin et al., 2021)

To address the issue of eclipsing, Gupta and Gupta (2021) suggest the use of multi-criteria decision-making approaches in WQI models. Multi-criteria decision-making approaches allow for the simultaneous consideration of multiple parameters and criteria, avoiding the problem of eclipsing. Additionally, other recent studies have suggested the use of sensitivity and uncertainty analyses (Sutadian et al., 2018; Uddin et al., 2022a) to identify the parameters and weightings that have the greatest impact on WQI values. Positive results of the use of these suggestions were revealed by Uddin et al. (2023a) with the IEWQI, which presents a novel approach for rating water quality using a combination of fuzzy logic and analytical hierarchy process (AHP) techniques coupled with machine learning to appraise model performance.

iv. Uncertainty issues.

Uncertainty in a model is a fundamental feature related to the model's specific parameters. As a result, during an analysis of index uncertainty, much focus is given to how the parameter

variation could affect the subindex values and the final index value (Uddin et al., 2021). Several studies have reported that uncertainty in an index is associated with the various stages of development of the WQI model (Juwana et al., 2016; Seifi et al., 2020). Thus, the purpose of analysing uncertainty is to determine the source of uncertainty in the whole model and the impact thereof on the final index value (Akhtar et al., 2021). Considering the preceding, the design and development of any WQI model must include a comprehensive uncertainty analysis to improve confidence when applying the model (Sutadian et al., 2018; Uddin et al., 2021). Uddin et al. (2023a) highlighted the importance of sensitivity analysis in quantifying uncertainty and assessing the robustness of any proposed model. Sensitivity analysis can help to identify the parameters and weightings that have the greatest impact on the final water quality rating, allowing for the identification of potential sources of uncertainty and the development of more robust and reliable models. In addition, Uddin et al. (2023c) demonstrated the effectiveness of machine-learning techniques in improving the predictive accuracy of the WQI model, proving that machine-learning models outperform traditional statistical approaches in terms of predictive accuracy and robustness.

v. The Metaphoring classification schemes problem.

This “metaphoring problem” of water quality classification schemes, as referred to by (Uddin et al., 2023c), was initially pointed out by Uddin et al. (2021) as the authors discussed the importance of water quality classification schemes, which provide a framework for interpreting water quality data and communicating the results to stakeholders and the public. The authors note that different countries and regions have developed their own water quality classification schemes, which can vary in terms of the number and type of parameters considered, the thresholds used to classify water quality, and the associated management actions. Considering this, recent studies have highlighted the importance of selecting an appropriate water quality classification scheme that is relevant to the specific context and management goals (Malek et al., 2022; Uddin et al., 2023c).

The recent study has demonstrated the persistence of this problem with Fig. 7 showing five different classification schemes for the 17 WQI models analysed. Recent studies have demonstrated that the current WQI model fuels uncertainty (Uddin et al., 2023c) due to high bias and overfitting (Malek et al., 2022). To resolve this, Malek et al. (2022) demonstrated the potential of machine learning algorithms such as Multilayer Perceptron Neural Networks,

KNN, XGB, SVN, GNB, Decision Tree, and RF to accurately predict water quality classification. The authors revealed that XGB had the highest accuracy and precision. A recent study also supported this finding when they reported that XGB outperformed all other classifier algorithms with 99.9% accuracy (Uddin et al., 2023c).

2.12. Conclusions

Water quality indices are data management tools for communicating complex water quality data to water authorities and the public. They are also used to assess the water quality status or health condition of watersheds at certain times and locations. Although there is yet to be a universally accepted WQI, there is consensus on the development methods. The objective of this review was to analyse water quality indices developed for surface water general evaluation and establish whether the WQIs can be used to evaluate both lentic and lotic ecosystems simultaneously. The following are the main conclusions of the review:

1. Most analysed WQIs are developed in a generic manner involving the five key steps, which include: (1) selection of parameters, (2) transformation of parameters to a single scale, (3) assigning parameter weightings, and (4) aggregation of parameters to determine the water quality index value, and (5) determining water quality classification schemes. Over 70% of the analysed WQI models used all five steps, and the other 30% used at least two steps. Although most models were developed using the generic method, most are site or region-specific and primarily address regional water quality challenges.
2. The selection of parameters is made through subjective methods from experts and local water quality stakeholders. For most WQIs, parameter selection is based on water quality risks, and the high variability confirms this in the number of parameters included in the analysed WQIs. In addition, physical parameters played a significant role as they were used the most, while biological parameters such as faecal coliforms and *E. coli* were used the least. This is an improvement opportunity for WQI models to align the model development process with international standards to improve the acceptability and wide use of a model.

3. Majority of the analysed WQIs used mathematical functions to transform the parameters, while expert opinions were considered the least. Subjectivity has been a constant impediment to the proper adoption of WQI models. Using less subjective methods demonstrates a significant leap towards more robust and impartial methods in the WQI development process.
4. The issues of eclipsing and uncertainty are significant in that they affect the accuracy of a model. These limitations are observed in most WQIs. As such, further studies that seek to reuse an existing WQI model or develop a new one can create a way to score these limitations. This would help users to know how dependable and acceptable the final index is.
5. It has been proven that despite the use of statistical methods such as cluster analysis (CA), factor analysis (FA), and analytic hierarchy process (AHP), WQI models continue to suffer from either eclipsing, ambiguity, or uncertainty limitations because natural ecosystems tend to be too complex for these statistical methods.
6. Apart from the CCME, IEWQI and the Hahn indices, the 14 other WQIs are designed only to evaluate one type of freshwater ecosystem. This is a limitation for any institution wishing to solve similar water quality challenges in a different ecosystem type. As such, these indices can be revisited and applied in another system to evaluate if the same effect can be achieved. This will reduce the burden of using more than one index for the same region instead of adapting one index for both ecosystem types.

2.13. Future directions

Models available in the literature have relied on expert opinion for parameter selection and weightage of parameters. This has been a significant contributor to issues of model uncertainty, reliability and adoption by water quality management practitioners and institutions. In addition, the preceding has always been a major limiting factor in the effort towards a universally acceptable WQI. In the more recent studies, such as the development of the West-Java WQI (Sutadian et al., 2018), we observed the effective application of statistical methods in parameter selection (cluster analysis) and assigning of weights (analytic hierarchy process) as a substitute

for the more subjective methods of expert opinions, while the IEWQI (Uddin et al., 2023a) used fuzzy logic and AHP coupled with machine learning. This improved the acceptability of the West-Java WQI and the IEWQI as reliable tools for managing water resources. Although this is positive, other studies (Elsayed et al., 2021) have argued that natural ecosystems are too complex for these statistical models and suggested models based on machine learning, such as artificial neural networks (ANNs). This is because ANNs can generalise non-linear patterns within a database and solve complex problems (Isiyaka et al., 2019; Adnan et al., 2019). Furthermore, ANNs have been effectively applied to evaluate the accuracy and reduce uncertainty in the final aggregation process (Šiljić et al., 2018; Uddin et al., 2021; Uddin et al., 2023a) and classification schemes (Gupta and Gupta, 2021; Uddin et al., 2023c). Using statistical methods coupled with machine learning techniques requires further exploration in the development process of WQI to eliminate WQI inaccuracies and uncertainties and improve the scope of application.

CREdiT Statement

Lazarus Katlego Mogane: was responsible for conceptualisation, data collection and curation, writing of the first draft and final revision. Esper J. Ncube: was responsible for the conceptualisation, editing, and reviews. Titus A.M. Msagati: was responsible for the conceptualisation, editing, and reviews. Tracy M. Masebe: was responsible for the conceptualisation, Methodology, editing, and reviews. All the authors read and approved the final version of the chapter output.

2.14. References

- Abbasi, T., Abbasi, S.A. (2012). Water quality indices. Burlington: Elsevier Science.
<https://doi.org/10.1016/b978-0-444-54304-2.00016-6>
- Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M., Kisi, O. (2019). Daily streamflow prediction using optimally pruned extreme learning machine. *Journal of Hydrology*, 577, 123981. <https://doi.org/10.1016/j.jhydrol.2019.123981>
- Akhtar, N., Ishak, M.I.S., Ahmad, M.I., Umar, K., Md Yusuff, M.S., Anees, M.T., Qadir, A., Ali Almanasir, Y.K. (2021). Modification of the water quality index (wqi) process for simple calculation using the multi-criteria decision-making (MCDM) method: A review. *Water*, 13, 905. <https://doi.org/10.3390/w13070905>
- Almeida, C., González, S., Mallea, M., and González, P. (2012). A recreational water quality index using chemical, physical, and microbiological parameters. *Environmental Science and Pollution Research*, 19(8), 3400–3411. <https://doi.org/10.1007/s11356-012-0865-5>
- Alobaidy, A.H.M.J., Abid, H.S., and Maulood, B.K. (2010). Application of water quality index for assessment of Dokan Lake ecosystem, Kurdistan region, Iraq. *Journal of Water Resources Protection*, 2(9), 792-798. doi:10.4236/jwarp.2010.29093
- Aragaw, T.T., Gnanachandrasamy, G. (2021). Evaluation of groundwater quality for drinking and irrigation purposes using GIS-based water quality index in urban area of Abaya-Chemo sub-basin of Great Rift Valley, Ethiopia. *Applied Water Science*, 11, 148. <https://doi.org/10.1007/s13201-021-01482-6>
- Asadollah, S.B.H.S., Sharafati, A., Motta, D., Yaseen, Z.M., 2021. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*, 9: 104599. <https://doi.org/10.1016/j.jece.2020.104599>.
- Ball, R.O., Church, R.L. (1980). Water quality indexing and scoring. *Journal of the Environmental Engineering Division.*, 106(4), 757-771. <https://doi.org/10.1061/JEEGAV.0001067>
- Banda, T.D., Kumarasamy, M. (2020). Development of a Universal Water Quality Index (UWQI) for South African River Catchments. *Water*, 12, 1534. <https://doi.org/10.3390/w12061534>

- Banerjee, T., Srivastava, R.K. (2009). Application of water quality index for assessment of surface water quality surrounding integrated industrial estate-Pantnagar. *Water Science and Technology*, 60, 2041–2053. <https://doi.org/10.2166/wst.2009.537>
- Barakat, A., Meddah, R., Afdali, M., and Touhami, F. (2018). Physicochemical and microbial assessment of spring water quality for drinking supply in Piedmont of Béni-Mellal Atlas (Morocco). *Physics and Chemistry of the Earth Parts A/B/C*, 104, 39–46. <https://doi.org/10.1016/j.pce.2018.01.006>
- Bartram, J., and Balance, R. (1996). Water quality monitoring: A practical guide to the design and implementation of freshwater quality studies and monitoring programmes. United Nations Environmental Programme and the World Health Organization. UNEP/WHO; ISBN: 0419223207
- Benvenuti, T., Kieling-Rubio, M., Klauck, C., and Rodrigues, M. (2015). Evaluation of water quality at the source of streams of the Sinos River Basin, southern Brazil. *Brazilian Journal of Biology*, 7(2), 98-104. <https://doi.org/10.1590/1519-6984.1513>
- Bordalo, A., Teixeira, R., and Wiebe, W. (2006). A water quality index applied to an international shared river basin: the case of the Douro River. *Environmental Management*, 38(6), 910–920. <https://doi.org/10.1007/s00267-004-0037-6>
- Bordalo, A.A., Nilsumranchit, W., and Chalermwat, K. (2001). Water quality and uses of the Bangpakong River (Eastern Thailand). *Water Research*, 35(15), 3635–3642. [https://doi.org/10.1016/s0043-1354\(01\)00079-3](https://doi.org/10.1016/s0043-1354(01)00079-3)
- Boyacioglu, H. (2010). Utilization of the water quality index method as a classification tool. *Environmental Monitoring and Assessment*, 167(1-4), 115–124. <https://doi.org/10.1007/s10661-009-1035-1>
- Brown, R.M., and McClelland, N.I. (1974). Up from chaos: the water quality index as an effective instrument in water quality management, pp.: 27. Ann Arbor, Michigan: National Sanitation Foundation
- Brown, R.M., McClelland, N.I., Deininger, R.A., and Landwehr, J.M. (1973). Validating the WQI. The paper presented at national meeting of American Society of Civil Engineers on water resources engineering, Washington, DC
- Brown, R.M., McClelland, N.I., Deininger, R.A., and Tozer, R.G. (1970). A water quality index—do we dare? *Water and Sewage Works*, 117(10), 339–343
- Busico, G.; Kazakis, N., Cuoco, E., Colombani, N., Tedesco, D.; Voudouris, K., and Mastrocicco, M. (2020). A novel hybrid method of specific vulnerability to

- anthropogenic pollution using multivariate statistical and regression analyses. *Water Research*, 171, 115386. <https://doi.org/10.1016/j.watres.2019.115386>
- Canadian Council of Ministers of the Environment (CCME) 2001. Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, user's manual. In: Canadian environmental quality guidelines. Winnipeg: CCME
- Carvalho, L., Cortes, R., and Bordalo, A. (2011). Evaluation of the ecological status of an impaired watershed by using a multi-index approach. *Environmental Monitoring and Assessment*, 174(1-4), 493–508. <https://doi.org/10.1007/s10661-010-1473-9>
- Chakravarty, T., Gupta S. (2021). Assessment of water quality of a hilly river of south Assam, northeast India using water quality index and multivariate statistical analysis. *Environmental Challenges*, 5: 100392. <https://doi.org/10.1016/j.envc.2021.100392>
- Chidiac, S., El Najjar, P., Ouaini, N., El Rayess, Y., and El Azzi, D. (2023). A comprehensive review of water quality indices (WQIs): history, models, attempts and perspectives. *Reviews in Environmental Science and Biotechnology*. <https://doi.org/10.1007/s11157-023-09650-7>
- Couillard, D., and Lefebvre, Y. (1985). Analysis of water indices. *Journal of Environmental Management.*, 21: 161-179
- Cude, C.G. (2001). Oregon water quality index: a tool for evaluating water quality management effectiveness. *Journal of American Water Resources Association*, 37, 125–137. <https://doi.org/10.1111/j.1752-1688.2001.tb05480.x>
- Dadollahi-Sohrab, A., Arjomand, F., and Fadaei-Nasab, M. (2012). Water quality index as a simple indicator of watersheds pollution in southwestern part of Iran. *Water and Environment Journal*, 26(4), 445–454. <https://doi.org/10.1111/j.1747-6593.2011.00303.x>
- Damo, R., and Icka, P. (2013). Evaluation of water quality index for drinking water. *Polish Journal of Environmental Studies*, 22(4), 1045–1051
- Das, J., Acharya, B.C. (2003). Hydrology and assessment of lotic water quality in Cuttack City, India. *International Journal of Engineering Research and Applications*, 150:163–175. <https://doi.org/10.1023/A:1026193514875>
- Davies, J.M. (2006). Application and tests of the Canadian Water Quality Index for assessing changes in water quality in lakes and rivers of central North America. *Lake and Reservoir Management*, 22(4), 308–320. <https://doi.org/10.1080/07438140609354365>

- Delbecq, A.L., Ven, A.H.V.D., and Gustafon, D.H. (1975). Group techniques for program planning: A guide to nominal group and Delphi processes, Glenview, Illinois
- Department of Environment Malaysia (DoEM). (2002). Malaysia environmental quality report (2001) Putrajaya, Malaysia: Department of Environment, Ministry of Science, Technology and Environment
- Dinius, S.H. (1987). Design of an index of water quality. *Water Resources Bulletin, American Water Resources Association*, 23(5), 823–843
- Dunnette, D. (1979). A geographically variable water quality index used in Oregon. *Journal of Water Pollution Control Federation*, 51, 53-61
- Dzwairo, B., Otieno, F.A.O., Ochieng, G.M., and Bezuidenhout, J.J. (2012). Raw water quality weight factors: Vaal Basin, South Africa. *Water Science and Technology*, 66, 1061-1068. <https://doi.org/10.2166/wst.2012.281>
- Edokpayi, J.N., Odiyo, J.O., and Durowoju, O.S. (2017). Chapter 18: Impact of wastewater on surface water quality in developing countries: A case study of South Africa. *IntechOpen*, pp. 401-416. <http://dx.doi.org/10.5772/66561>
- El-Batrawy, O.A., Ibrahim, M.S., Fakhry, H., El-Aassar, M.R., El-Zeiny, A.M., El-Hamid, H.T.A., El-Alfy, M.A. (2018). Anthropogenic impacts on water quality of river Nile and marine environment, Rosetta branch using geospatial analyses. *Journal of Environmental Science*, 47, 89–101. <https://doi.org/10.21608/joese.2018.158202>
- Elsayed, S., Ibrahim, H., Hussein, H., Elsherbiny, O., Elmetwalli, A.H., Moghanm, F.S., Ghoneim, A.M., Danish, S., Datta, R., Gad, M. (2021). Assessment of water quality in Lake Qaroun using ground-based remote sensing data and artificial neural networks. *Water*, 13, 3094. <https://doi.org/10.3390/w13213094>
- Fathi, E., Zamani-Ahmadmarmoodi, R., and Zare-Bidaki, R. (2018). Water quality evaluation using water quality index and multivariate methods, Beheshtabad River, Iran. *Water Science and Technology*, 8, 210. <https://doi.org/10.1007/s13201-018-0859-7>
- Fathi, P., Ebrahimi Dorche, E., Zare Shahraki, M., Stribling, J., Beyraghdar Kashkooli, O., Esmaeili Ofogh, A., Bruder, A. (2022). Revised Iranian water quality index (RIWQI): a tool for the assessment and management of water quality in Iran. *Environmental Monitoring and Assessment*. 194 (7): 504. [10.1007/s10661-022-10121-9](https://doi.org/10.1007/s10661-022-10121-9)
- Feng, Y., Danying, Q., Bao, Q., Lin, M., Xigang, X., You, Z., and Xiaogang, W. (2016). Improvement of CCME WQI using grey relational method. *Journal of Hydrology*, 543, 316–323. <https://doi.org/10.1016/j.jhydrol.2016.10.007>

- Fu, L., Wang, Y.G. (2012). Statistical tools for analysing water quality data. In: Voudouris K, Voutsas D (eds.), *Water Quality Monitoring and Assessment*. Rijeka, Croatia: *InTechOpen*, pp.: 143-168
- Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H., Ramli, M.F. (2012). Artificial neural network modelling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*. 64 : 2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Giriyappanavar, B.S., Patil, R.R. (2013). Water quality assessment of Gullava Reservoir of Belgium District Using NSF-WQI. *International Journal of Science and Research*, 2(4), 71-72. <https://www.doi.org/10.36106/ijsr>
- Guo, D., Lintern, A., Webb, J.A., Ryu, D., Liu, S., Bende-Michl, U., Leahy, P., Wilson, P., Western, A.W. (2019). Key factors affecting temporal variability in stream water quality. *Water Resources Research*. 55, 112–129. <https://doi.org/10.1029/2018WR023370>
- Gupta, S., Gupta, S.K. (2021). A critical review on water quality index tool: genesis, evolution, and future directions. *Ecological Informatics*, 63: 101299 <https://doi.org/10.1016/j.coinf.2021.101299>
- Gupta, V.K., Carrott, P.J.M., Suhas, M.L.R. (2009). Low-cost adsorbents: Growing approach to wastewater treatment. Critical Reviews in *Journal of Environmental Science Technology*, 39, 783–842. <https://doi.org/10.1080/10643380801977610>
- Hallock, D. (2002). A water quality index for ecology's stream monitoring program. Washington: Department of ecology publications.
- Hanh, P.T.M., Sthiannopkao, S., Ba, D.T., and Kim, K.W. (2011). Development of water quality indexes to identify pollutants in Vietnam's surface water. *Journal of Environmental Engineering*, 137, 273-83
- Harkins, R.D. (1974). An objective water quality index. *Journal of Water Pollution Control Federation*, 46, 588-591.
- Horton, R.K. (1965). An index number system for rating water quality. *Journal of Water Pollution Control Federation*, 37(3), 300–306.
- Hoseinzadeh, E., Khorsandi, H., Wei, C., and Alipour, M. (2015). Evaluation of Aydughmush river water quality using the national sanitation foundation water quality index (NSFWQI), river pollution index (RPI), and forestry water quality index (FWQI).

- Desalination and Water Treatment*, 54(11), 2994–3002.
<https://doi.org/10.1080/19443994.2014.913206>
- House, M.A. (1980). A water quality index for rivers. *Water International* 5, 16–21.
<https://doi.org/10.1080/02508068008685872>.
- House, M.A. (1989). A water quality index for river management. *Journal of the Institution of Water and Environmental Management*, 3, 336–344. <https://doi.org/10.1111/j.1747-6593.1989.tb01538.x>
- Isiyaka, H.A., Mustapha, A., Juahir, H., Phil-Eze, P. (2019). Water quality modelling using artificial neural network and multivariate statistical techniques. *Modelling, Earth Systems, and Environment*, 5, 583–593. <https://doi.org/10.1007/s40808-018-0551-9>
- Jolliffe, I. (2005). Principal component analysis. In: Everitte B, Howell D (eds), *Encyclopedia of statistics in behavioural science*. New York: John Wiley & Sons.
doi.org/10.1002/0470013192.bsa501
- Joung, H.M., Miller, W.W., Mahannah, C.N., and Guittjens, J.C. (1979). A generalised water quality index based on multivariate factor analysis. *Journal of Environmental Quality*, 8, 95-100. <https://doi.org/10.2134/jeq1979.00472425000800010021x>
- Juwana, I., Muttill, N., and Perera, B.J.C. (2012). Indicator-based water sustainability assessment—a review. *Science of the Total Environment*, 438(0), 357–371.
<https://doi.org/10.1016/j.scitotenv.2012.08.093>
- Juwana, I., Muttill, N., Perera, B.J.C. (2016). Uncertainty and sensitivity analysis of West Java Water Sustainability Index - A case study on Citarum catchment in Indonesia. *Ecological Indicators*, 61, 170–178. <https://doi.org/10.1016/j.ecolind.2015.08.034>.
- Kanakoudis, V., and Tsitsifli, S. (2020). Insights on the Water–Energy–Food Nexus. *Water*. 12(10), 2882. <https://doi.org/10.3390/w12102882>
- Kodikara, P.N., Perera, B.J.C., and Kularathna, M. (2010). Stakeholder preference elicitation and modelling in multicriteria decision analysis—a case study on urban water supply. *European Journal of Operational Research*, 206(1), 209–220.
<https://doi.org/10.1016/j.ejor.2010.02.016>
- Kumar, A., Dua, A. (2009). Water quality index for assessment of water quality of river Ravi at Madhopur (India). *Global Journal of Environmental Sciences*, 8, 49-57.
<https://doi.org/10.4314/GJES.V8I1.50824>
- Kumar, B., Singh, U.K., Ojha, S.N. (2019). Evaluation of geochemical data of Yamuna River using WQI and multivariate statistical analyses: a case study. *International*

- Kumarasamy, M.V., and Macholo, T.C. (2018). Water quality modelling study for Umhlangane river, South Africa. *Indian Journal of Science and Technology*, 11 (23), 1-7. <https://doi.org/10.17485/ijst/2018/v11i23/125653>
- Landwehr, J.M., and Deininger, R.A. (1974). Discussion: an objective water quality index. *Journal of Water Pollution Control Federation*, 46, 1804-1809.
- Liou, S.M., Lo, S.L., and Wang, S.H. (2004). A generalized water quality index for Taiwan. *Environmental Monitoring and Assessment*, 96(1-3), 35–52. <https://doi.org/10.1023/b:emas.0000031715.83752.a1>
- Lohani, B.N., and Todino, G. (1984). Water quality index for Chao Phraya River. *Journal of Environmental Engineering*, 110, 1162-1177. [https://doi.org/10.1061/\(asce\)0733-9372\(1984\)110:6\(1163\)](https://doi.org/10.1061/(asce)0733-9372(1984)110:6(1163))
- Lumb, A., Sharma, T.C., Bibeault, J.F., and Klawunn, P. (2012). A comparative study of USA and Canadian water quality index models. *Water Quality, Exposure, and Health*, 3, 203-2016. <https://doi.org/10.1007/s12403-011-0056-5>
- Lumb, A., Sharma, T.C., and Bibeault, J.F. (2011). A review of genesis and evolution of water quality index (WQI) and some future directions. *Water Quality Exposure and Health*, 3, 11-24. <https://doi.org/10.1007/s12403-011-0040-0>
- Ma, X., Wang, L., Yang, H., Li, N., Gong, C. (2020). Spatiotemporal analysis of water quality using multivariate statistical techniques and the water quality identification index for the qinhuai river basin, east China. *Water* (Switzerland) 12. <https://doi.org/10.3390/w12102764>
- Mahlathi, C., Siyakatshana, N., and Chirwa, E. (2016). Water quality modelling and optimisation of wastewater treatment network using mixed integer programming. *Water SA*, 42(4), 650-657. <https://doi.org/10.4314/wsa.v42i4.16>
- Malek, N.H.A., Wan Yaacob, W.F., Md Nasir, S.A., Shaadan, N. (2022). Prediction of water quality classification of the Kelantan River Basin, Malaysia, using machine learning techniques. *Water* (Basel) <https://doi.org/10.3390/w14071067>
- Massoud, M.A. (2012). Assessment of water quality along a recreational section of the Damour River in Lebanon using the water quality index. *Environmental Monitoring and Assessment*, 184, 4151–4160. <https://doi.org/10.1007/s10661-011-2251-z>

- May, R., Dandy, G., and Maier, H. (2011). Review of input variable selection methods for artificial neural networks. In Suzuki K (ed.), *Artificial neural networks – methodological advances and biomedical applications. InTech Open*, pp.:19-44. <https://doi.org/10.5772/16004>
- McClelland, N.I. (1974). Water quality index application in the Kansas River basin. EPA-907/9-74-001. US Environmental Protection Agency, Kansas City, Missouri
- Medeiros, A.C., Faial, K.R.F., do Carmo Freitas Faial, K., da Silva Lopes, I.D., de Oliveira Lima, M., Guimarães, R.M., Mendonça, N.M. (2017). Quality Index of the surface water of Amazonian rivers in industrial areas in Pará Brazil. *Marine Pollution Bulletin*, 123: 56–164. <https://doi.org/10.1016/j.marpolbul.2017.09.002>
- Mladenovic-Ranisavljevic I.I., and Zerjic, S.A. (2017). Comparison of different models of water quality index in the assessment of surface water quality. *International Journal of Science and Technology*, 15, 665-674 <https://doi.org/10.1007/s13762-017-1426-8>
- Moher, D., Liberati, A., Tetzlaff, J., and Altman, D.G., PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 2009 Jul 21, 6(7): e1000097. doi: 10.1371/journal.pmed.1000097.
- Mojahedi, S.A., and Attari, J. (2009). A comparative study of water quality indices for Karun River. *World Environmental and Water Resources Congress: Great Rivers*, 1–9. [https://doi.org/10.1061/41036\(342\)246](https://doi.org/10.1061/41036(342)246)
- Mostafaei, A. (2014). Application of multivariate statistical methods and water quality index to evaluation of water quality in the Kashkan River. *Environmental Management*, 53(4), 865–881. <https://doi.org/10.1007/s00267-014-0238-6>
- Nagy-Kovács, Z., Davidesz, J., Czihat-Mártonné, K., Till, G., Fleit, E., Grischek, T. (2019). water quality changes during riverbank filtration in Budapest, Hungary. *Water*, 11, 302. doi:10.3390/w11020302
- Najafzadeh, M., Homaei, F., Farhadi, H. (2021). Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: integration of remote sensing and data-driven models. *Artificial Intelligence Review*, 54: 4619–4651. <https://doi.org/10.1007/s10462-021-10007-1>.
- Najafzadeh, M., Niazmardi, S. (2021). A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. *Natural Resources Research*, 30: 3761–3775. <https://doi.org/10.1007/s11053-021-09895-5>
- Ott, W.R. (1978). Environmental indices theory and practice. Michigan: Ann Arbor Science.

- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal*, 372: n71. doi: 10.1136/bmj. n71. PMID: 33782057; PMCID: PMC8005924.
- Parmar, V., and Singh, B. (2011). Suitability of Gangetic water for pisciculture at Patna using WQI approach. *International Journal of Environmental Sciences and Research*, 1, 39-43.
- Parween, S., Siddique, N.A., Mahammad Diganta, M.T., Olbert, A.I., Uddin, Md.G. (2022). Assessment of urban river water quality using modified NSF water quality index model at Siliguri city, West Bengal, India. *Environmental and Sustainability Indicators*, 16, 100202. <https://doi.org/10.1016/j.indic.2022.100202>
- Pesce, S.F., and Wunderlin, D.A. (2000). Use of water quality indices to verify the impact of Coardoba City (Argentina) on Suquoa River. *Water Research*, 34(11): 2915-26. [https://dx.doi.org/10.1016/S0043-1354\(00\)00036-1](https://dx.doi.org/10.1016/S0043-1354(00)00036-1)
- Peters, N.E., and Meybeck, M. (2000). Water Quality Degradation Effects on Freshwater Availability: Impacts of Human Activities. *Water International*, 25, 185-193. <https://doi.org/10.1080/02508060008686817>
- Prati, L., Pavanello, R., and Pesarin, F. (1971). Assessment of surface water quality by a single index of pollution. *Water Research*, 5 (9), 741-751. [https://doi.org/10.1016/0043-1354\(71\)90097-2](https://doi.org/10.1016/0043-1354(71)90097-2)
- Ramakrishnaiah, C., Sadashivaiah, C., and Ranganna, G. (2009). Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, *Indian Journal of Chemistry*, 6, 523–530. <https://doi.org/10.1155/2009/757424>
- Rangeti, I., Dzwauro, B., Barratt, G.J., and Otieno, F.A.O. (2015). Ecosystem-specific water quality indices. *African Journal of Aquatic Science*, pp.: 1-8. <http://dx.doi.org/10.2989/16085914.2015.1054341>
- Richardson, A.M., (1997). Development of an estuarine water quality index (eWQI) for New South Wales. BSc Honours dissertation, University of Sydney, Australia

- Said, A., Stevens, D.K., and Sehlke, G. (2004). An innovative index for evaluating water quality in streams. *Environ. Management*, 34(3), 406–414. <https://doi.org/10.1007/s00267-004-0210-y>
- Sandhu, C., Grischek, T., Börnick, H., Feller, J., Sharma, S.K. (2019). A water quality appraisal of some existing and potential riverbank filtration sites in India. *Water*, 11, 215. <https://doi.org/10.3390/w11020215>
- Sapkal, R.S., and Valunjkar, S.S. (2013). Development and sensitivity analysis of water quality index for evaluation of surface water for drinking purpose, *International Journal of Civil Engineering and Technology*, 4(4), 119-134
- Sargaonkar, A., and Deshpande, V. (2003). Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. *Environmental Monitoring and Assessment*, 89(1), 43–67.
- Scottish Research Development Department (SRDD). (1976). Development of a water quality index. Scottish research development department, applied research & development report number ARD3 pp.: 61. Edinburg, UK: Engineering Division
- Seifi, A., Dehghani, M., Singh, V.P. (2020). Uncertainty analysis of water quality index (WQI) for groundwater quality evaluation: Application of Monte-Carlo method for weight allocation. *Ecological Indicators*, 117 <https://doi.org/10.1016/j.ecolind.2020.106653>
- Sener, S., Sener, E., and Davraz, A. (2017). Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SWTurkey). *Science of the Total Environment*, 584-585, 131-144. doi.org/10.1016/j.scitotenv.2017.01.102
- Shah, K.A., Joshi, G.S. (2015). Evaluation of water quality index for river Sabarmati, Gujarat, India. *Applied Water Science*, 7, 1349-1358. <https://doi.org/10.1007/s13201-015-0318-7>
- Shakhari, S., Baberjee, I. (2019). A multi-class classification system for continuous water quality monitoring. *Heliyon*, 5: e01822. <https://doi.org/10.1016/j.heliyon.2019.e01822>.
- Sharma, D., and Kansal, A. (2011). Water quality analysis of River Yamuna using water quality index in the national capital territory, India (2000–2009). *Applied Water Science*, 1(3-4), 147–157. doi:10.1007/s13201-011-0011-4
- Sharma, P., Meher, P.K., Kumar, A., Gautam, Y.P., and Mishra, K.P. (2014). Changes in water quality index of Ganges River at various locations in Allahabad. *Sustain. Water Quality and Ecology*, 3(4), 67–76. <https://doi.org/10.1016/j.swaqe.2014.10.002>

- Singh, K.P., Basant, A., Malik, A., and Jain, G. (2009). Artificial neural network modelling of the river water quality- a case study. *Ecological Modelling*, 220, 888-895. <https://doi.org/10.1016/j.ecolmodel.2009.01.004>
- Šiljić Tomić, A., Antanasijević, D., Ristić, M., Perić-Grujić, A., Pocajt, V. (2018). Application of experimental design for the optimization of artificial neural network-based water quality model: A case study of dissolved oxygen prediction. *Environmental Science and Pollution Research*, 25, 9360–9370. <https://doi.org/10.1007/s11356-018-1246-5>
- Smith, D.G. (1990). A better water quality indexing system for rivers and streams. *Water Research*, 24, 1237-1244. [https://doi.org/10.1016/0043-1354\(90\)90047-A](https://doi.org/10.1016/0043-1354(90)90047-A)
- Soumalia, K.I., Niandou, A.S., Naimi, M., Mohamed, C., Schimmel, K., Luster-Teasley, S., and Sheick, N.N. (2019). A systematic review and meta-analysis of water quality indices. *Journal of Agricultural Science and Technology B*, 9, 1-14. doi:10.17265/2161-6264/2019.01.001
- Sutadian, A.D., Muttill, N., Yilmaz, A.G., and Perera, B.J.C. (2016). Development of river water quality indices-a review. *Environmental, Monitoring, and Assessment*, 188 (58), 29-57. <https://doi.org/10.1007/s10661-015-5050-0>
- Sutadian, A.D., Mutil, N., Yilmaz, A.G., and Perera, B.J.C. (2018). Development of a water quality index for rivers in West Java province, Indonesia. *Ecological Indicators*, 85, 966-982. <https://doi.org/10.1016/j.ecolind.2017.11.049>
- Tanner, M.U., Ustun, B., and Erdinçler, A. (2011). A simple tool for the assessment of water quality in polluted lagoon systems: A case study for Kucukcekmece Lagoon, Turkey. *Ecological Indices*, 11, 749–756. <https://doi.org/10.1016/j.ecolind.2010.08.003>
- Tukura, B.W., Kagbu, J.A., and Gimba, C.E. (2009). Effects of pH and seasonal variations on dissolved and suspended heavy metals in dam surface water. *Chemistry Class Journal*, 6, 27–30.
- Tyagi, S.H., Sharma, B., Singh, P., and Dobhal, R. (2013). Water quality assessment in terms of water quality index. *American Journal of Water Resources*, 1, 34–38. <https://doi.org/10.12691/ajwr-1-3-3>
- Uddin, M.G., Nash, S., and Olbert, A.I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. <https://doi.org/10.1016/j.ecolind.2020.107218>
- Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I. (2022a). A comprehensive method for improvement of water quality index (WQI) models for coastal water quality

- assessment. *Water Research*. 219, 118532.
<https://doi.org/10.1016/j.watres.2022.118532>
- Uddin, Md.G., Nash, S., Talas, M., Diganta, M., Rahman, A., Olbert, A.I. (2022b). Robust machine learning algorithms for predicting coastal water quality index. *Journal of Environmental Management*. 321: 115923
<https://doi.org/10.1016/j.jenvman.2022.115923>.
- Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I. (2022c). Development of a water quality index model - a comparative analysis of various weighting methods. *Mediterranean Geosciences Union Annual Meeting (MedGU-21)*. Springer, Istanbul, pp. 1–6.
- Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I., (2023a). A sophisticated model for rating water quality. *Science of the Total Environment*, 868:161614.
<http://dx.doi.org/10.1016/j.scitotenv.2023.161614>
- Uddin, G., Nash, S., Rahman, A., Olbert, A.I., (2023b). Assessing optimization techniques for improving water quality model. *Journal of Cleaner Production*, 385: 135671.
<https://doi.org/10.1016/j.jclepro.2022.135671>
- Uddin, M.D.G., Nash, S., Rahman, A., Olbert, A.I. (2023c). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection*. 169, 808–828.
<https://doi.org/10.1016/j.psep.2022.11.073>
- Walski, T. M., & Parker, F. L. (1974). Consumers water quality index. *Journal of the Environmental Engineering Division, ASCE*, 100(3), 593–611.
- Yadav, A.K., Khan, P., and Sharma, S.K. (2010). Water quality index assessment of groundwater in Todaraisingh Tehsil of Rajasthan state, India-A greener approach. *Journal of Chemistry*, 7, 428–432. <https://doi.org/10.1155/2010/419432>
- Young, P., and Beck, B. (1974). The modelling and control of water quality in river system, *Automatica*, 1, 455–68. [https://doi.org/10.1016/0005-1098\(74\)90047-8](https://doi.org/10.1016/0005-1098(74)90047-8)
- Zainudin, Z. (2010). Benchmarking river water quality in Malaysia. *Jurutera*, February 2010: 12-15.

3. Methodology

3.1. Introduction

This chapter provides a detailed description of the study's methodological approaches used to address the research problem and achieve the stated objectives. It further describes research design, data collection methods, secondary data sources, and the analytical techniques employed. The chapter begins with restating the research questions to remind the reader how the methodological approach addresses the pertinent issues. This study adopted a quantitative method to achieve the objectives and address the research questions.

Research Questions

1. How does water quality vary across seasons and what are the main influencers?
2. What are the effects of significantly variable water quality on freshwater ecosystems and ecological function?
3. Would an artificial intelligence (AI)-based water quality prediction model help to identify trends in water quality to obtain necessary information on pollution sources?

3.2. Study area Description

- i. The Crocodile River

The Crocodile River catchment is located entirely within the Mpumalanga province in South Africa (Fig. 3.1). The river originates near Dullstroom where it flows into the Kweni Dam passing through Mbombela before joining the Komati, then enters Mozambique at Komatipoort (IUCMA, 2014). The Elands River and Kaap River form part of the two major tributaries whereas the Lungsklip River, Nels River, Houtbosloopspruit, Gladdespruit, White River, and Besterspruit form part of the minor tributaries (IUCMA, 2014). The upper part of the catchment is largely dominated by man-made forests as well as commercial and emerging agriculture (DWA, 2011; ICMA, 2011).

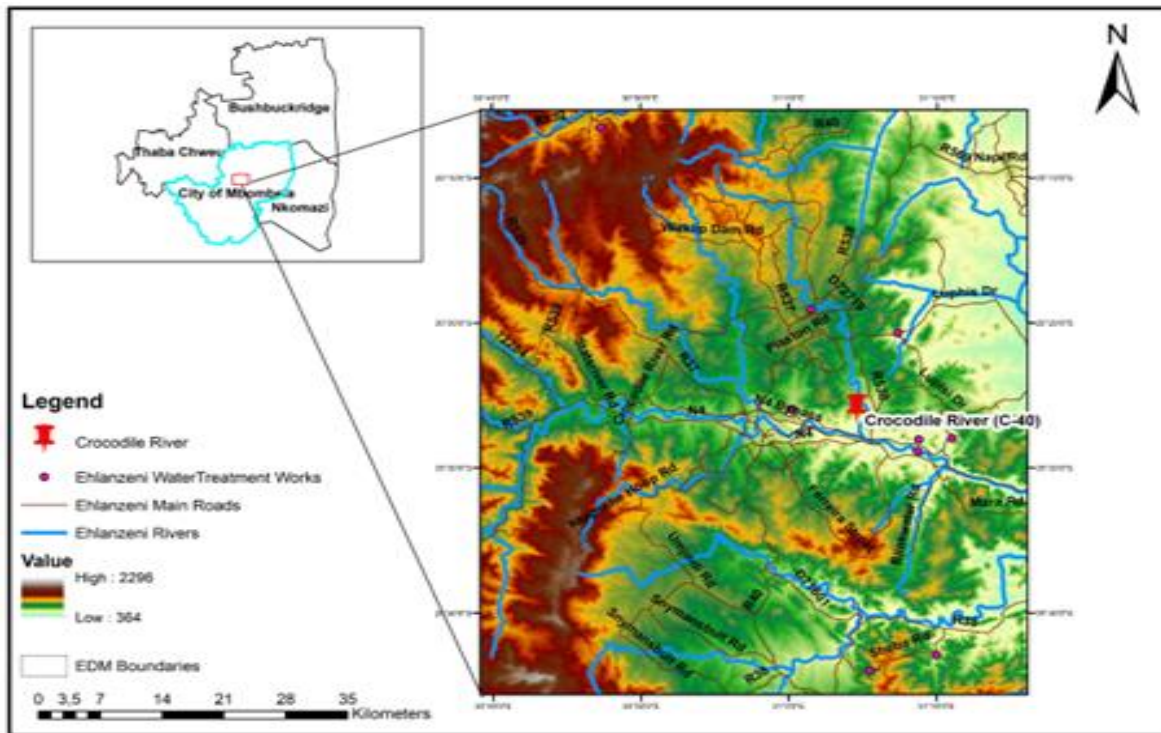


Fig. 3.1 Map of the Crocodile River sampling areas and land use activities (Esri: <https://www.arcgis.com>)

The catchment largely receives rainfall in the summer season and has a subtropical climate. In addition, the Crocodile River carry with it some considerable ecological significance as it passes along the southern boundary of the Kruger National Park (KNP) (IUCMA, 2014). There are notable activities in the surrounding areas of the catchment that may affect the quality of water. These include a pulp and paper mill and a ferrochrome smelter along the Elands tributary, mining in the Kaap tributary and a sugar mill along the lower Crocodile River (DWAf, 2004; ICMA, 2011; DWA, 2011; IUCMA, 2014). In 2004, the Department of Water Affairs and Forestry (DWAf) reported that the water quality in the catchment was good, but this was changed in 2011 when the same institution reported that the water quality may be degrading (DWAf, 2004; DWA, 2011). It is also worth noting that South Africa has signed an agreement with Mozambique regarding quantity and quality of the water passing through the borders (TPTC, 2002).

ii. The Inyaka Dam

The Inyaka Dam is an earth fill type of dam that is built along the Ngwaritsane River in the Bushbuckridge area of Mpumalanga. The dam was built with the primary purpose of storing water for irrigation. The dam is surrounded by domestic settlements, agricultural activities, and

a wastewater treatment plant. As of 2014, the dam served as an abstraction point for the Inyaka water treatment plant which supplied treated water to over 750,821 people in Bushbuckridge (Statistics South Africa, 2022). Given the population supplied by the Inyaka Dam, the Inyaka Dam is one of the most important water resources that is unfortunately, under-represented in literature. Fig. 3.2 shows the map of the Inyaka Dam.

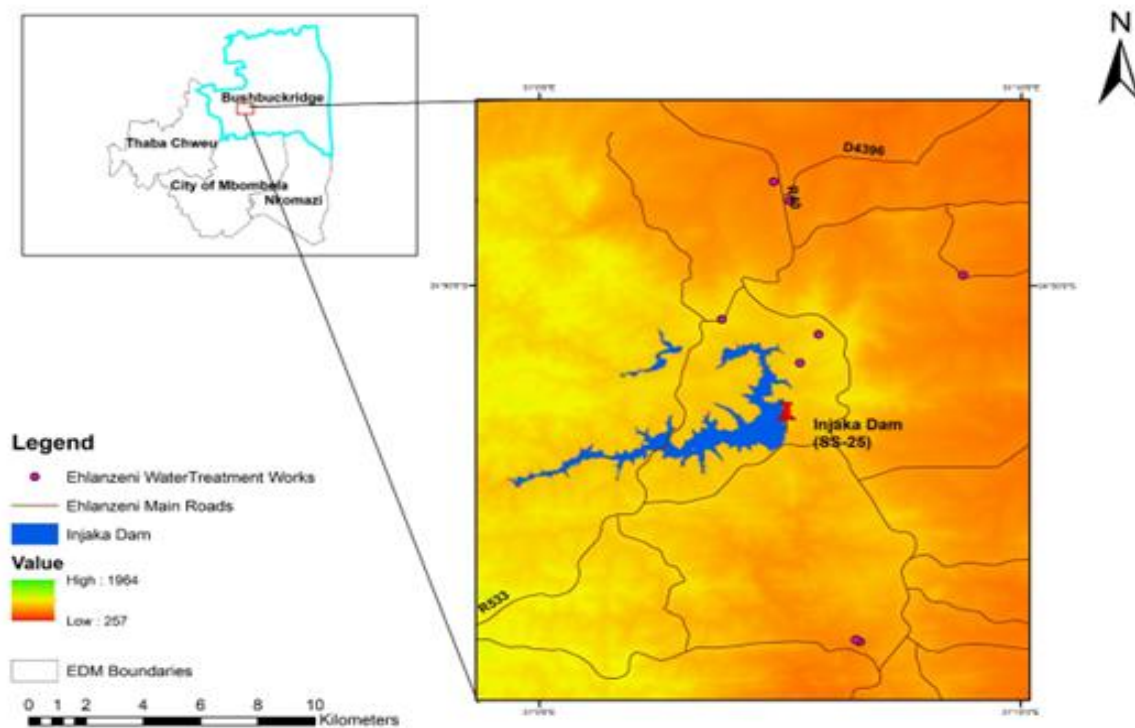


Fig. 3.2 Map of the Inyaka Dam and surrounding land use activities (Esri: <https://www.arcgis.com>)

3.3. Data collection and processing

Composite water samples were collected in the summer (December to February), autumn (March to May), winter (June to August), and spring (September to November) seasons of 2021 from the Crocodile River and Inyaka Dam. Samples were collected with sterile 500mL bottles for bacteriological analysis and 1L polypropylene bottles for chemical analysis. The samples were analysed within 6 hours after transport to the Inyaka laboratory or stored at $<8^{\circ}\text{C}$ overnight when analysis was not possible. In total, 104 samples were collected at the Inyaka Dam throughout the four seasons (summer = 27; autumn = 22; winter = 31; spring = 24). In addition, the total number of samples collected through the four seasons in the Crocodile River was 102 (summer = 22; autumn = 18; winter = 37; spring = 25).

3.4. Experimental methods and data analysis

Objective 1: To investigate the microbial and physicochemical water quality characteristics of the Crocodile River and Inyaka Dam across all four seasons.

3.4.1. Selected microbial parameters.

Microbiological parameters were selected from those prescribed in the WHO Water Quality Guidelines, Standards and Health (WHO, 2001). The proposed microbial parameters are based on the ease of the detection method and the cost of analysis.

i. Total Coliforms and *E. coli*

For the detection of coliforms, raw water samples were aliquoted into sterile 120mL anti-foam vessels (IDEXX, Maine) and graduated at 100mL. For purposes of dilution, three anti-foam vessels were prepared. The first vessel contained 100mL of the sample, the second, 10mL of the sample and the third, 1mL, thus making a serial dilution of 1:1, 1:10 and 1:100, respectively. After mixing the water sample with Colilert-18 reagent, the Colilert Quanti-Trays (IDEXX, Maine) were sealed using a Model 2X Quanti-Tray sealer (IDEXX, Maine) and incubated at 36 °C for 18-22 hours. All yellow wells were recorded as positive for total coliforms. For quality control purposes, sterile distilled water and reference species (*Klebsiella variicola*-positive control, *E. coli*-positive control and *Pseudomonas aeruginosa*-negative control) were run concurrently with the samples. The IDDEX Quanti Tray/2000 MPN table was used to determine the result as count MPN/100mL. Results were recorded on designated forms and transferred to a Microsoft Excel worksheet.

ii. Faecal streptococci

A series of dilutions (1:1 through to 1:1000) were prepared for each sample using sterile 14mL volume McCartney bottles and 9mL sterile deionised water to give counts of between 30-300 colonies. Kenner Faecal (KF) Streptococcus agar base (Oxoid, UK) medium was dispensed within 2 hours of plating 1mL of the sample to avoid contamination. Petri dishes were incubated at 35°C for 48 hours. Plates were counted for all red to pink colonies with a colony counter (Gallenkamp, England) and recorded as colony-forming units (CFU) per 1mL of river water sampled. For plates with dilutions, the results were computed using the respective

dilution factor. Results were recorded on designated forms in CFU. This information was also recorded on a Microsoft Excel worksheet.

iii. Bifidobacteria

55.5g of Bifidus Selective Medium (BSM) agar (Sigma-Aldrich, Missouri, USA) was weight out, dissolved into 1L of deionised water, and sterilised at 121°C. After cooling the agar to 55 °C BSM supplement was added to the agar and homogenised. Approximately 1mL of the sample was plated, and the agar was poured onto sterile petri dishes. Plates were observed for brown or violet colonies. Results were recorded on designated forms with the following notations: No growth for negative samples; brown or violet for positive samples. This information was also recorded on a Microsoft Excel 365 sheet.

3.4.2. Selected physicochemical parameters.

The selected physicochemical parameters used in this study included Temperature, Turbidity, Conductivity, pH, Iron (Fe), Manganese (Mn), Nitrate-Nitrogen (NO₃-N), and Phosphate (PO₄³).

i. Determination of pH

To determine pH, a HACH TitraLab AT1000 Series Autotitrator (Hach, Colorado, USA), was used. pH calibration standards 4.0, 7.0 and 10.0 were used to calibrate the instrument before analysis and pH standard 7.0 was used after every 10th sample to verify the instrument. Data was recorded in pH units on designated analysis sheets and transferred to a Microsoft Excel worksheet.

ii. Determination of turbidity

To determine the turbidity of surface water, a HACH 2100N portable turbidimeter (HACH, Colorado, USA) was used. Standard 1.00 NTU was used to calibrate the instrument before use, and the same standard was used to verify the instrument after every 10th sample. Data were recorded in nephelometric turbidity units (NTU) on designated analysis sheets and transferred to a Microsoft Excel worksheet.

iii. Determination of electrical conductivity

To determine conductivity using a HACH TitrLab AT1000 Series Autotitrator (Hach, Colorado, USA), the instrument was first calibrated and passed calibration specifications. The instrument was calibrated at 25°C with a 0.01M (147mS/m) potassium chloride standard. The calibration was accepted if the cell constant was between 0.3 and 0.5. Data was recorded in mS/m units on designated analysis sheets and transferred to a Microsoft Excel worksheet.

iv. Determination of Fe, Mn, NO₃-N, PO₄³⁻,

The HACH DR900 multi-parameter spectrophotometer was used to analyse water samples for iron, Manganese, Nitrogen and Phosphate. HACH Methods adapted from the Standard Methods for the Examination of Water and Wastewater were used. Data was recorded in mg/L units on designated analysis sheets and transferred to a Microsoft Excel worksheet.

3.5. Statistical analysis of physical, chemical, and microbiological data

Water quality data was populated using Microsoft Excel 365. Seasonal parameter means and standard deviations were computed in SPSS 26. The Pearson correlation matrix at 95% and 99% confidence levels determined the relationships between water quality parameters. OriginPro graphing and analysis software version 2024b was used to compute factor analysis (FA) and construct principal component analysis (PCA) biplots to determine which parameters contributed the most to water quality deterioration.

Objective 2: To investigate and profile the bacterial community structure and the dominant bacterial taxa in the Crocodile River and Inyaka dam across the summer and winter seasons

3.6. DNA Extraction

DNA was extracted from 100mL of surface water samples which was filtered through a 0.45 nitrate cellulose membrane filter (Whatman GE Healthcare Life Sciences, Buckinghamshire, UK) and lyzed in a 1mg/mL lysozyme solution containing 0.30-0.50mm glass beads (Sigma-Aldrich Corporation, St. Louis, MO, USA) for disruption of the bacterial cell wall. The lysis

solution was agitated in a vortex for 5min. 1mg/mL of Proteinase K was added to the lysis solution. DNA was eluted into 1.5mL nuclease free tubes. The concentration of the DNA was determined using Nanodrop ND100 (NanoDrop Technologies, Wilmington, DE, USA). Isolated DNA was stored at -20 °C for further analysis.

3.7. DNA Sequencing

Genomic DNA samples were commercially sequenced by Inqaba Biotechnical Industries Pty Ltd, South Africa, for a full-length 16S rRNA on a PacBio single-molecule real-time (SMRT) sequencing platform (Pacific Biosciences Inc., Menlo Park, CA, USA). High-performance liquid chromatography (HPLC) purified and barcoded 16S rRNA-specific forward primer 12F (5'-CTACACGACGCTCTTCCGATCTGTAGACGCTGAGRGTTYGATYMTGGCTCAG-3') and reverse primer 32R

(5'-AAGCAGTGGTATCAACGCAGAGCTCGTCAGATRGYTACCTTGTTACGACTT-3') were used to generate SMRTbell libraries. The libraries were sequenced on a PacBio RS II system with SPRQ chemistry.

3.8. Statistical and Bioinformatic Analysis of Microbial Ecology Data

Microbial sequence data were processed using the Quantitative Insights into Microbial Ecology (QIIME) software suite (version 1.7) through the cloud-based Nephele Microbiome Analysis platform (Weber et al., 2018). Sequence clustering was performed at 100% identity to obtain representative sequences using PyNAST and UCLUST software (Caporaso et al., 2010; Chigwada et al., 2023). QIIME 2 cloud-based Krona visualisation was used to generate abundance charts at the phylum and class resolutions (<https://view.qiime2.org/>). Diversity analyses included both alpha and beta indices, such as the Chao1, ACE, Jackknife, Shannon-Weiner, Simpson, and Principal coordinate analysis (PCoA). All graphs were generated using R package (version 3.1.2), Microsoft Excel 365, and OriginPro 2024b (student version). The PCoA was conducted based on Jensen-Shannon divergence metrics using species-level taxonomic resolution and included unclassified operational taxonomic units (OTUs). Analysis of similarities (ANOSIM) was used to determine significant differences in microbial community structure between and within the sites.

Objective 3: To develop an inclusive microbial water quality index model for the Crocodile River and the Inyaka Dam.

3.9. Model development workflow.

The development of the machine learning algorithm (MLA) based predictive model followed a set of predefined processes. The step-by-step processes undertaken to compute the final model are outlined in Fig. 3.3. The coding of the software was completed entirely on Python[®] programming platform.

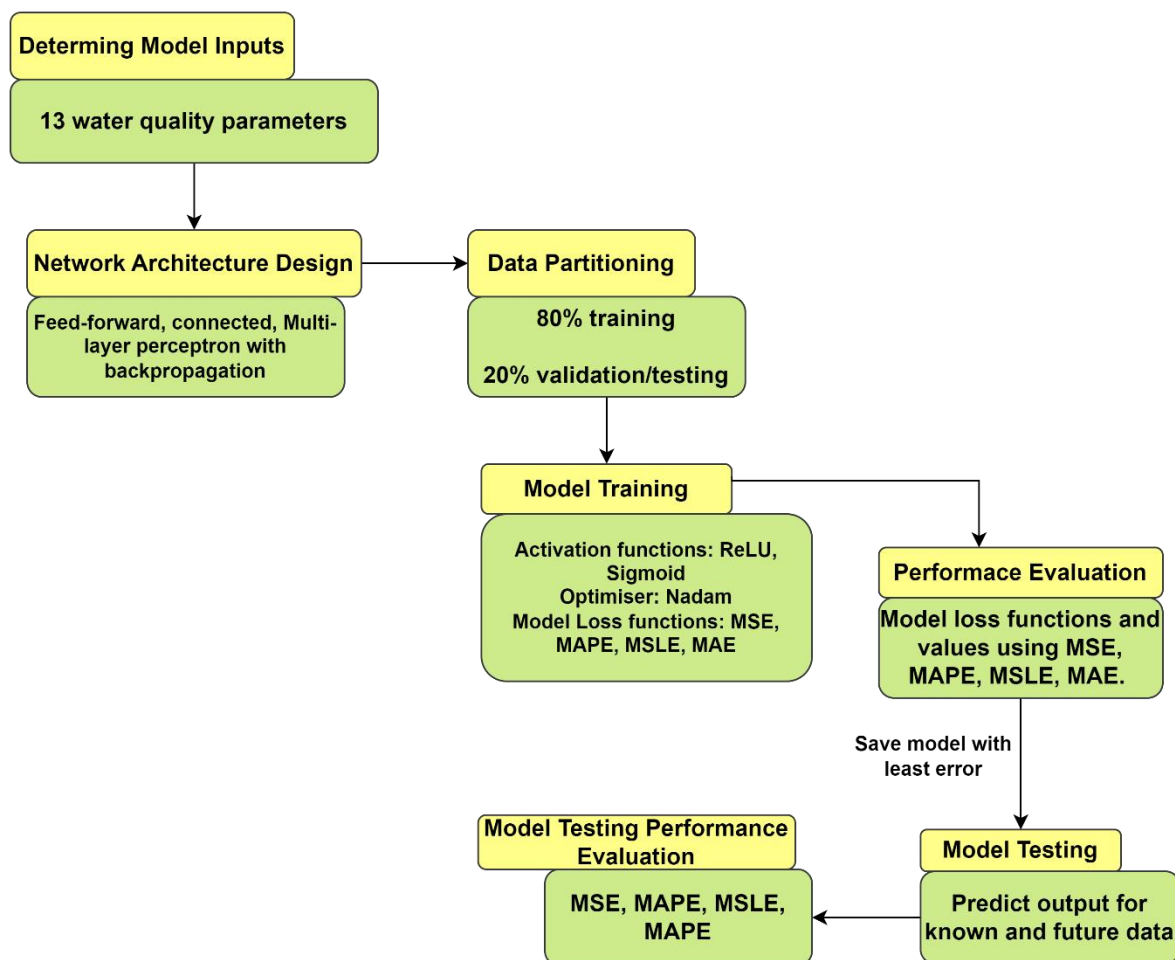


Fig. 3.3 Model development workflow involving the number of parameters used as input variables, design of the model architecture, data sorting, training, testing and model performance evaluation

The input variables for the model consisted of thirteen preselected water quality parameters whose data was collected between January and December of 2021. However, deep machine learning algorithms naturally requires large volumes of data for the model to have acceptable

accuracy. As such, historical data from 2016 to 2020 was added to what was already collected in 2021. The following water quality parameters were considered: pH, turbidity, electrical conductivity (EC), temperature, rainfall, Manganese (Mn), Iron (Fe), Phosphate (PO_4^{3-}), Nitrate-nitrogen ($\text{NO}_3\text{-N}$), total coliforms, *E. coli*, Bifidobacteria, and faecal streptococci. The model is designed to produce at least one outputs. However, when more than one outputs are expected, there would be lesser data to train the model.

A Microsoft Excel database of the water quality data was created from the historical and recent data. Data was partitioned into 80% training and 20% validation. For the model training parameters, the rectified linear unit (ReLU) was used as the default activation function while the Nesterov-accelerated Adaptive Moment estimation (Nadam) which is an extension of the Adaptive Movement Estimation (Adam) algorithm, was used as a default model optimizer. After training, the performance and reliability of the model was evaluated using the mean squared error (MSE), mean squared logarithmic error (MSLE), mean absolute percentage error (MAPE), mean absolute error (MAE). The model loss against epoch graphs was also used to determine how far apart the training data is from the validation data, or how likely the model will be reliable during prediction.

3.10 Ethical Considerations

Ethical clearance for this study was obtained from UNISA-College of Agriculture and Environmental Sciences (CAES) Health Research Ethics Committee with reference number: 2020/CAES_HREC/047 ([Appendix 1](#)). Permission to access all sampling sites was acquired through a written consent letter from the Catchment Management Agency (CMA), the Inkomati-Usuthu CMA, which manages the Crocodile River and the Inyaka dam ([Appendix 2](#)). Furthermore, permission to use the Bushbuckridge Local Municipality laboratory was acquired by means of written notice ([Appendix 3](#)).

3.11 References

- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335–336.
- Chigwada, A. D., Ogola, H. J. O., & Tekere, M. (2023). Multivariate analysis of enriched landfill soil consortia provide insight on the community structural perturbation and functioning during low-density polyethylene degradation. *Microbiological Research*, 274. <https://doi.org/10.1016/j.micres.2023.127425>
- Department of Water Affairs and Forestry (DWAF). (2004). Internal Strategic Perspectives: Inkomati Water Management Area – Version 1. Prepared by Tlou & Matji (Pty) Ltd. DWAF Report No. P WMA 05/000/00/0303. Department of Water Affairs and Forestry, Pretoria.
- Department of Water Affairs (DWA). (2011). Planning Level Review of Water Quality in South Africa. Sub-series No. WQP 2.0. Pretoria, South Africa. Authors: Priya Moodley, Jan
- Inkomati Catchment Management Agency (ICMA). (2011). The Inkomati Catchment Management Strategy: a first generations Catchment Management Strategy for the Inkomati Water Management Area. Inkomati Catchment Management Agency, Nelspruit, South Africa.
- Inkomati-Usuthu Catchment Management Agency (IUCMA), 2014. Annual water quality status report for the Inkomati-Usuthu water management area, IUCMA, Nelspruit, South Africa.
- Statistics South Africa (Stats SA). (2022). Census 2022 Statistical Release: 0301.4. Pretoria, South Africa.
- Tripartite Permanent Technical Committee (TPTC), (2002). Tripartite interim agreement between the Republic of Mozambique and the Republic of South Africa and the Kingdom of Swaziland for co-operation on the protection and sustainable utilisation of the water resources of the Incomati and Maputo watercourses. Johannesburg.

4. Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa

4.1. Preamble

The development of water quality indices relies on the availability of water quality data, whether historical or current. This chapter presents a technical paper that explores the impact of seasonal variations in key water quality parameters, specifically rainfall and temperature, on the overall health of watersheds. The study compared how these seasonal variations affect two distinct aquatic systems—a river and a dam—while considering their differing ecological dynamics.

The selection of parameters for monitoring was guided by two criteria: (i) catchment Resource Quality Objectives (RQOs) and (ii) pollution risks associated with the parameters. Despite the unique characteristics of the two water resources, the same microbial parameters were monitored in both systems to ensure consistency in comparative analysis. Data collection occurred over four seasons, from January to December 2021, to capture seasonal parametric variations and assess their significance in overall water quality. With permission, monthly historical rainfall data from 2016 to 2021 was obtained from the South African Weather Services (SAWS). Historical water quality data was also sourced publicly from the Inkomati-Usuthu Catchment Management Agency (IUCMA).

This chapter underscores the seasonal variability of water quality in the Crocodile River and Inyaka Dam. It identifies rainfall and temperature as primary drivers of these variations, with different parameters contributing to variability in each water resource. The findings highlight the interconnectedness of climate patterns and water quality, emphasising how seasonal shifts are critical in understanding aquatic ecosystem dynamics. Furthermore, this study draws attention to the growing influence of climate change on water security. Seasonal variability underscores the urgency of developing and implementing ecosystem-specific water quality management strategies that are resilient to the impacts of climate change. Such strategies are essential to ensuring sustainable water security and maintaining the ecological integrity of aquatic systems. This research contributes valuable insights to inform water resource

management and policy development by addressing the unique challenges posed by seasonal and climate-driven variability.

4.2. Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa.²

4.3. Abstract

This study assessed seasonal variations in water quality and their impact on the Inyaka Dam and Crocodile River in Mpumalanga Province, South Africa. A total of 206 water samples were collected across four seasons to analyse parameters including pH, turbidity, electrical conductivity (EC), temperature, Manganese (Mn), Iron (Fe), phosphate (PO_4^{3-}), nitrate-nitrogen ($\text{NO}_3\text{-N}$), *E. coli*, total coliforms, faecal streptococci, and Bifidobacteria. Principal component analysis (PCA) and Pearson correlation identified major pollutants and their seasonal variations.

Findings indicated that water quality in both the Inyaka Dam and the Crocodile River exceeded permissible limits. In the Inyaka Dam, summer rainfall (202.68 mm) spurred microbial growth, including *E. coli* and Bifidobacteria, while winter saw elevated levels of PO_4^{3-} , EC, Fe, and $\text{NO}_3\text{-N}$. The Crocodile River exhibited its poorest water quality in summer, with high levels of conductivity, turbidity, Fe, Mn, and $\text{NO}_3\text{-N}$, driven largely by rainfall. Winter pollution in the river was marked by *E. coli*, PO_4^{3-} , total coliforms, $\text{NO}_3\text{-N}$, and Bifidobacteria.

The study highlights significant pollution in the Crocodile River, particularly in summer, linked to rainfall and effluent discharges. Microbial pollution persisted across seasons, influenced by both weather and point-source contamination. The winter season exacerbated water quality deterioration in the Crocodile River due to reduced flow, while Inyaka Dam's winter pollution was attributed to lake stratification.

² The contents of this chapter are published in Heliyon, Volume 10, Issue 19, 2024, e38246. <https://doi.org/10.1016/j.heliyon.2024.e38246>

4.4. Background

Water security has recently emerged as a major global challenge, exacerbated by climate change, pollution, and population growth, which contribute to over-abstraction of water resources (Ahmed et al., 2020; Mishra et al., 2021). Although the United Nations General Assembly declared access to clean and safe water a human right in July 2016, a significant number of people still lack access to this essential resource (United Nations Water, 2021). Surface water quality has long been subjected to continuous degradation due to both natural processes and human activities, posing a serious concern for sustainable human and economic development (Uddin et al., 2021a). Many third-world countries face the dilemma of balancing water quality with the need to prioritise water supply, while even some developed nations struggle to maintain water quality due to nutrient enrichment in their water resources (Ortega et al., 2016; Abbasi & Abbasi, 2012).

Natural factors affecting water quality include variations in the hydrological cycle, climate, stratification, and topography (Magesh et al., 2013; Uddin et al., 2018) which can degrade water quality due to seasonal fluctuations in physicochemical and biological characteristics. Additionally, anthropogenic factors such as undertreated or untreated industrial effluents, agricultural runoff, and improper waste disposal are significant contributors to water quality degradation (Hasan et al., 2019; Uddin & Jeong, 2021). Contaminants from these activities including metals, pesticides, nutrients, salts, and other hazardous chemicals are often released into water bodies (Varol & Şen, 2012). Furthermore, nitrates from sewage effluent are common in municipal wastewater that is discharged into freshwater systems (Pisciotta et al., 2015).

In South Africa, like many other parts of the world, industries are located near rivers and dams, discharging semi-treated or untreated effluents into these bodies of water (Rahman et al., 2021). For instance, the Crocodile River in Mbombela, Mpumalanga, is surrounded by informal dumping sites, agricultural activities, and industrial operations, making it a direct discharge point for wastewater. The Inyaka Dam, in Bushbuckridge, also in Mpumalanga, was primarily constructed for irrigation and is surrounded by agricultural activities and a wastewater treatment plant that discharges effluent directly into the dam (DWA, 2007).

Both the Inyaka Dam and Crocodile River are subject to target water quality guidelines that monitor the quality of effluent discharges from surrounding activities (Department of Water and Sanitation (DWS) South Africa, 2016). National monitoring programmes such as the National Microbiological Monitoring Programme (NMMP), National Chemical Monitoring

Programme (NCMP), National Toxicity Monitoring Programme (NTMP), National Eutrophication Monitoring Programme (NEMP), and the River Eco-status Monitoring Programme (REMP), along with municipalities and catchment agencies, conduct routine water quality monitoring at standardised sampling sites. Despite these monitoring programmes and legislation like Chapter 3 of the South African National Water Act 36 of 1998 and Chapter 7 of the National Environmental Management Act 107 of 1998, which provides guidelines for pollution control and subsequent enforcement steps, South Africa still experiences significant pollution levels in its water resources (Mnguni, 2022; Phungela et al., 2022). These examples underscore the critical role of water quality monitoring plays in restoring and maintaining healthy ecosystems (Saiful Islam et al., 2019; Sarkar et al., 2016).

To the best of our knowledge, very few studies, if any, have reported on the temporal variation in the physicochemical and microbiological properties of water conducted in the Crocodile River in Mbombela and the Inyaka Dam in Bushbuckridge. This study, therefore, assessed and compared the temporal variations and potential pollutants in the two water resources over different seasons. Understanding temporal variations in water quality is essential for improving watershed management.

4.5. Materials and methods

4.5.1. Study area description and land use information.

i. The Crocodile River

The Crocodile River catchment is located entirely within the Mpumalanga province in South Africa (Fig. 4.1). The river originates from areas that lie in the vicinity of Dullstroom, where it flows into the Kweni Dam, passing through Mbombela. The catchment receives rainfall in the summer and has a subtropical climate. In addition, the Crocodile River bears with it some considerable ecological significance as it passes along the southern boundary of the Kruger National Park (KNP). There are notable activities in the surrounding areas of the catchment that may affect water quality. Closer to the sampling areas of the study are a dumping site (Fig. S1) and a wastewater treatment plant that discharges its effluent into the river. Fig. 4.1 shows the sampling site with coordinates (25° 29' 18''S 31° 10' 23''E)

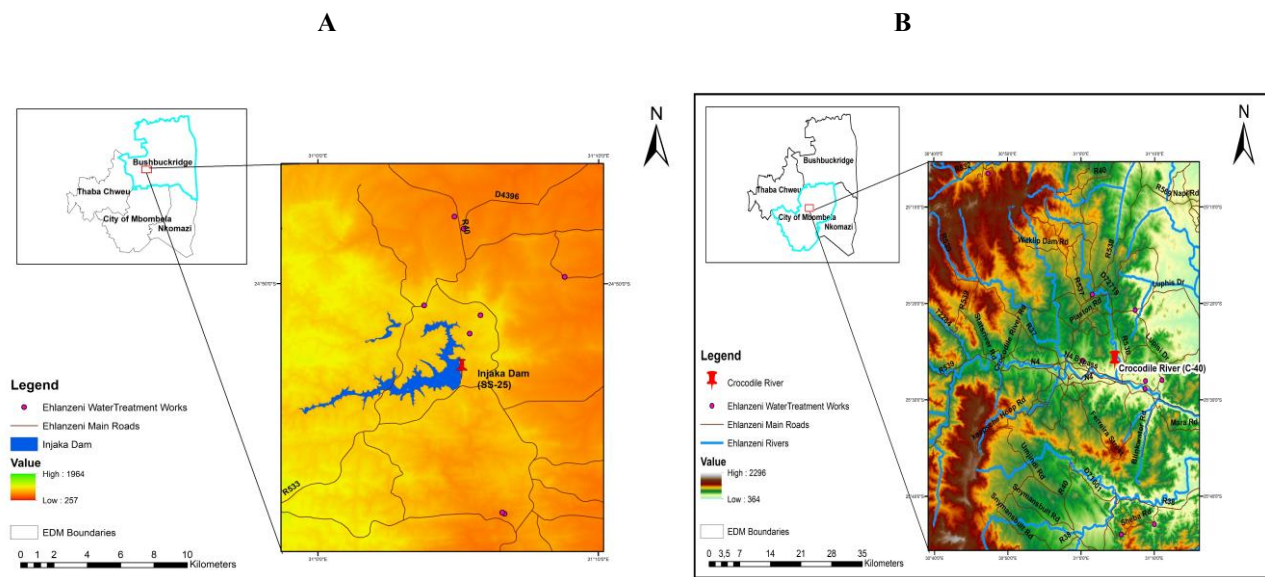


Fig. 4.1 Maps of the study area, and sampling sites (Esri: <https://www.arcgis.com>). Fig. 4.1 frame A shows the Inyaka Dam, and frame B shows the Crocodile River.

Dense industrial, agricultural, and domestic land use activities surround the Crocodile River, which is used as a discharge point. According to (Phungela et al., 2022), wastewater treatment systems are problematic for the water quality of the Crocodile River.

ii. Inyaka Dam

The Inyaka Dam is an earth-fill Dam constructed along the Marite River in the Bushbuckridge area of Mpumalanga province and was built to store water for irrigation purposes. As of 2014, the dam served as an abstraction point for the Inyaka water treatment plant, which supplied treated water to more than 850,000 people in Bushbuckridge (DWA, 2007). Given the population supplied by the Inyaka Dam, the dam is one of the most critical water resources and is unfortunately under-represented in the literature. Frame A shows the sampling point with coordinates (24°53'26.60''S 31°04'45.37''E).

Along the Inyaka Dam, there are noticeable scattered small-scale agricultural activities. This is because the dam was initially intended for irrigation, but most of the water has been repurposed for domestic and drinking use because of the growing population of Bushbuckridge. Moreover,

the Maviljan wastewater treatment plant uses the dam as a discharge point (Appendix 4, Fig. S4.2), which is currently under-treating the effluent discharged to the dam.

4.5.2. Sample collection and handling.

Composite water samples were collected in the summer (December to February), autumn (March to May), winter (June to August), and spring (September to November) of 2021 from the Crocodile River and Inyaka Dam. Water samples were collected at nationally established sites designated SS-25 and C-40 in the Department of Water and Sanitation monitoring programme. Water samples were collected from the surface layer in 500 mL sterile bottles directly from the river and dam. A total number of 104 and 102 samples were collected from the Inyaka Dam and Crocodile River, respectively every month of each season. The samples were analysed within six hours after being transported to the laboratory or stored at $<8^{\circ}\text{C}$ overnight when analysis was impossible.

4.5.3. Experimental methods

Thirteen parameters were selected based on resource quality objectives and risk assessment from surrounding activities. These parameters included *Escherichia coli* (*E. coli*), Total coliforms (TC), Faecal streptococci (FS), Bifidobacteria, pH, Electrical conductivity (EC), Turbidity, Temperature, Average rainfall (AR), Iron (Fe), Manganese (Mn), Nitrate-nitrogen ($\text{NO}_3\text{-N}$), and Phosphate (PO_4^{3-}). Only temperature was recorded on site with an infrared thermometer. Average rainfall data were obtained with permission from the South African Weather Services (SAWS).

The concentrations of Fe, Mn, $\text{NO}_3\text{-N}$ and PO_4 were measured using a Hach DR3900 spectrophotometer (Hach, Loveland, CO) using the powder pillow standard methods. The EC and pH of the samples were analysed with a HACH TitraLab AT1000 Series auto titrator (Hach, Loveland, CO), while turbidity was rapidly determined using a Hach TU5200 spectrophotometer (Hach, Loveland, CO).

E. coli and TC were biochemically determined using IDEXX Colilert-18[®] (IDEXX, Westbrook, ME) based on the ISO 9308-2 standard method. Briefly, surface water samples were analysed in serial dilutions of 100, 10 and 1mL. Colilert-18 medium snap packs were

dissolved in the samples and incubated for 18 – 22 hours before reading results for TC under visible light and *E. coli* under ultra-violet light.

Bifidobacteria counts were obtained by incubating the pour-plated samples for 72 hours at 35°C in a Bifidus Selective Medium (BSM) enriched with BSM supplement (Condalab, Torrenjon de Ardoz, Madrid). All purple-brown colonies were counted and recorded as bifidobacteria.

To determine FS counts, samples were pour-plated and incubated for 48 hours at 35°C in a Kenner Faecal (KF) Streptococcus agar base enriched with 1% TTC supplement (Condalab, Torrenjon de Ardoz, Madrid). All colonies with red to pink colour were counted and recorded as faecal streptococci.

All study area maps throughout this document were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used here under license.

4.5.4. Statistical analysis

Water quality data were populated using Microsoft Excel 2019. The seasonal parameter means, and standard deviations were computed using IBM SPSS 26. The Pearson correlation matrix at 95% and 99% confidence levels determined the relationships between water quality parameters. OriginPro 2024b was used to compute factor analysis (FA) tables and construct principal component analysis (PCA) plots to determine which parameters contributed the most to water quality deterioration.

PCA computation procedure

- i. Briefly, the source data from the variables is standardised so that each of the parameters contributes equally to the analysis. Standardisation to a common scale is done using the following equation:

$$Z = \frac{x - \bar{x}}{\sigma} \quad (4.1)$$

Where x is the parameter value; $\bar{x}$ is the mean value of all parameters; and σ is the standard deviation

- ii. From this, the covariance matrix is computed to identify correlations of the parameters with each other

- iii. From the covariance matrix, the eigenvectors and eigenvalues are generated and ranked from highest (high variance) to lowest (least variance).
- iv. Lastly, the eigenvalues with the highest variance are plotted along the principal component axes.

4.6. Results and Discussion

4.6.1. Water quality properties of the Inyaka Dam

Water quality compliance was evaluated by measuring the critical physical, chemical, and biological properties of the water and comparing them to recommended targets. In this study, selected physicochemical and microbiological parameters of water, such as pH, EC, turbidity, temperature, PO_4^{3-} , $\text{NO}_3\text{-N}$, Fe, Mn, *E. coli*, TC, Bifidobacteria, and FS, were measured. The measurements for all four seasons, and the permissible limits, are presented in Table 4.1.

Table 4.1: Water quality parameters mean measurements and standard deviations for all observed seasons

Parameters	Summer			Autumn			Winter			Spring			Limits	
	N (Total)	Mean	SD	N (Total)	Mean	SD	N (Total)	Mean	SD	N (total)	Mean	SD	RQO	TWQG
<i>E. coli</i>	27	1194,07	5042,18	22	274,27	123,24	31	6,52	7,95	24	9,54	13,40	130	130
PO_4^{3-} ,	27	0,07	0,07	22	0,04	0,05	31	0,03	0,04	24	0,02	0,03	0,125	0,025
$\text{NO}_3\text{-N}$	27	0,57	0,62	22	0,36	0,56	31	1,51	2,43	24	0,95	1,04	6	6
pH	27	7,48	0,44	22	7,21	0,53	31	7,31	0,30	24	7,35	0,41	6.5-8.0	6,5 – 8,5
EC	27	20,19	19,27	22	11,27	12,78	31	25,16	18,19	24	22,30	19,88	30	40
Aver. Rainfall	27	204,59	135,03	22	66,66	61,48	31	7,09	8,48	24	51,28	61,13	N/A	N/A
Bifidobacteria	27	2,70	2,13	22	5,64	4,86	31	0,58	0,67	24	0,29	0,46	-	-
FS	27	7,22	8,35	22	3,41	4,10	31	2,81	3,11	24	1,75	1,75	-	30
Fe	27	1,09	0,60	22	0,74	0,51	31	0,88	0,29	24	0,76	0,10	-	0,01
Mn	27	0,04	0,01	22	0,08	0,04	31	0,14	0,18	24	0,09	0,03	0,18	
TC	27	16698,59	4834,54	22	11796,18	514,10	31	1933,16	672,21	24	8282,00	960,19	-	-
Temperature	27	25,01	0,95	22	24,45	2,07	31	18,54	2,05	24	21,72	1,23		
Turbidity	27	3,16	1,14	22	3,38	1,26	31	2,99	1,04	24	3,31	1,92	-	-

FS=Faecal streptococci, TC= Total coliforms

From Table 4.1, it is possible to see non-compliance for certain parameters, especially during summer, autumn and winter. The noncomplying parameters included *E. coli* in summer (1.15225×10^3), autumn (2.7427×10^2 MPN/100mL), and Mn in winter (0.19mg/L). Although variation was observed for other parameters such as pH, EC, Nitrates, and FS, the concentrations remained within the permissible limits.

i. Physical parameters

The mean pH of the Inyaka Dam ranged from 7.21 to 7.49 across the observed seasons. The actual seasonal mean pH values were (7.21), (7.31), (7.35), and (7.49), respectively (Table 4.1). In line with the target water quality guidelines for the Inyaka catchment, the pH was within acceptable ideal limits of 6.5 - 8.0 (Department of Water and Sanitation (DWS) South Africa, 2016)

The mean EC ranged from 11.27 to 25.16 mS/cm across all four seasons. The lowest mean EC was observed in autumn at 11.27, while the other seasons were consistently in the same range at 21.12, 22.30, and 25.16 mS/cm, respectively (Table 4.1). This was within the prescribed acceptable limits of ≤ 55 mS/m. Although these values are typical of a lentic system, there is a massive difference between autumn and the other seasons. This trend can only be attributed to seasonal weather variations and yearly rainfall patterns.

The turbidity values ranged from 2.55 to 4.10 NTU for the Inyaka Dam. Interestingly, the highest turbidity measures were recorded in winter and spring, when rainfall was minimal. This result was inconsistent with the findings of previous studies (Scott Winton et al., 2019a). The seasonal mean temperatures regarding their seasonal variations for the surveyed seasons were 24.17°C, 24.4°C, 19.46°C, 21.62°C, respectively (Table 4.1), these values are conducive to aquatic life and household activities.

In South Africa, temperature is only regulated as an aquatic ecosystem driver (Department of Water and Sanitation (DWS) South Africa, 2016) It states that temperature variations should occur infrequently and should not vary by more than 2°C. According to (Rahman et al., 2021), slight variations in water temperature may not be as significant in freshwater because of aquatic life's wide range of temperature tolerance. However, studies by (Machender et al., 2014) and (Ahmed et al., 2016) reported that temperature variations play a significant role in highly polluted water because they affect dissolved oxygen concentrations.

Despite the small range in values, high physical parameter variability was observed in the summer season compared to autumn, spring, and winter, with rainfall influencing patterns of variability across all seasons (Fig. 4.2). However, parameter concentrations were below the permissible ranges. Fig. 4.2 also provides a comparison of the seasonal variability of the physical parameters in the Inyaka Dam.

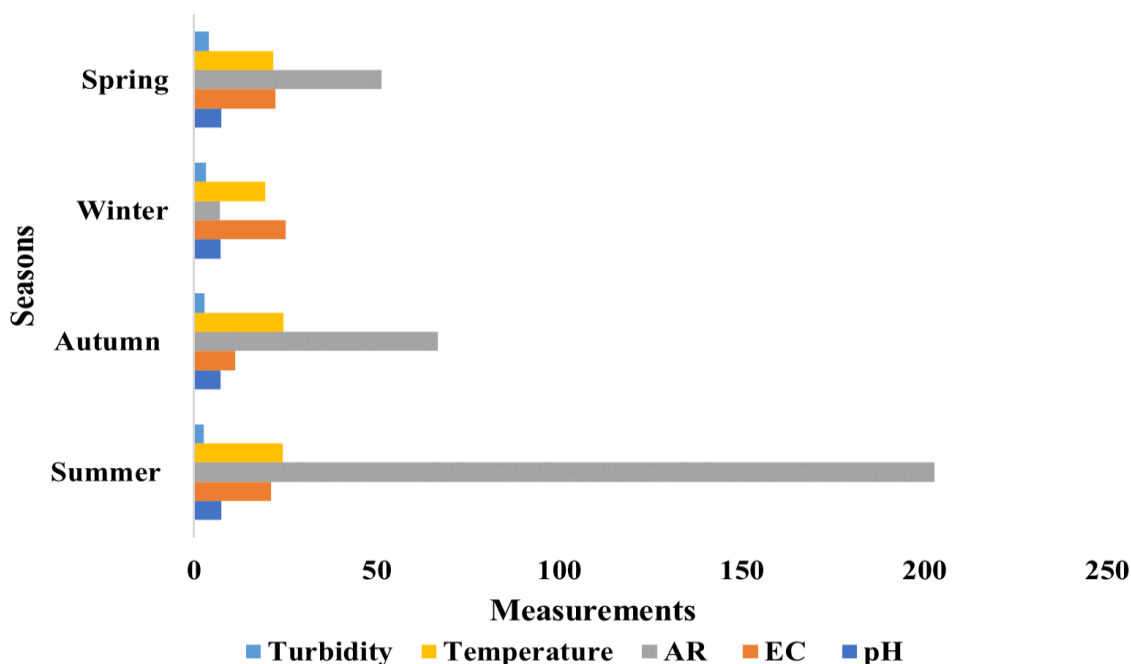


Fig. 4.2 Variations in the collective physical parameter throughout the different seasons of the Inyaka Dam

ii. Chemical parameters

Phosphates and Nitrates

Phosphate and Nitrate compounds create significant problems when released into water resources insufficiently treated or without prior treatment. These nutrients are also significant nutrients required by living organisms for their physiological processes. However, these are considered pollutants under heavy loads and above the standard limits. High nutrient loads favour the growth of aquatic plants and accelerate the growth of algal blooms, bad odour and water decolouration (Moloantoa et al., 2022).

The Inyaka dam receives under-treated effluent discharged from the Maviljan wastewater plant and is surrounded by agricultural activities. It is expected that the concentration of nutrients upstream of the dam will be high. However, there are no visible or reported cases of

eutrophication in the Inyaka dam. The mean concentration of PO_4^{3-} throughout the surveyed seasons ranged from 0.02 to 0.07mg/L. According to the set limits of the Inyaka dam, 50% of PO_4^{3-} data may not exceed 0.125 mg/L (Table 4.1). The highest PO_4^{3-} was recorded in the rainy summer months, consistent with other studies (Umer et al., 2020). Additionally, Benariba et al. (2022) reported fair to low readings of phosphates across all studied seasons.

The nitrate concentration of the Inyaka Dam was measured against domestic target water quality set at 6mg/L. The observed mean values across the seasons ranged from 0.36 to 1.51mg/L. The highest reading was observed in winter, with less mean rainfall (7.09mm). The concentrations are within limits set out for the water resource and resonate with other studies (Isiuku & Enyoh, 2020; Wirnkor et al., 2018).

Iron and Manganese

Iron is found in freshwater at levels ranging between 0.5 to 50mg/L (WHO, 2003). The presence of iron in water promotes the growth of iron bacteria, which combine Iron and Manganese with oxygen to form rust on (Telegdi et al., 2017). The mean Fe measurements ranged from 0.76 to 1.59mg/L, with the highest level observed in summer. This contrasts with what was reported in Lake Beseka in Ethiopia (Umer et al., 2020), where Fe concentrations ranged between 0.40–3.80 mg/L with the highest concentrations recorded in the dry season.

Deposits of high manganese concentration can be found near the bottoms of deep-fill dams such as the Inyaka Dam. This may be attributed to temperature stratification during the summer months when the warmer water remains in the upper layer of the dam. In winter, the upper layer is cooler and able to mix with the lower layer, thus freeing trapped manganese at the bottom (Winton et al., 2019b). This process explains the observed mean Mn measurements for all the observed seasons, which were (0.03), (0.07), (0.19), and (0.11), respectively. The highest measurements were observed in winter when the rainfall and temperature were minimal (Table 4.1). The standard limits for Mn are provided in the South African water quality guidelines for irrigation and domestic use. They were set at 0.18 mg/L and 0.02 mg/L, respectively (DWAf, 1996b). Comparatively, the combined chemical parameters showed high concentrations in winter, whereas spring and autumn showed the lowest concentrations (Fig. 4.3).

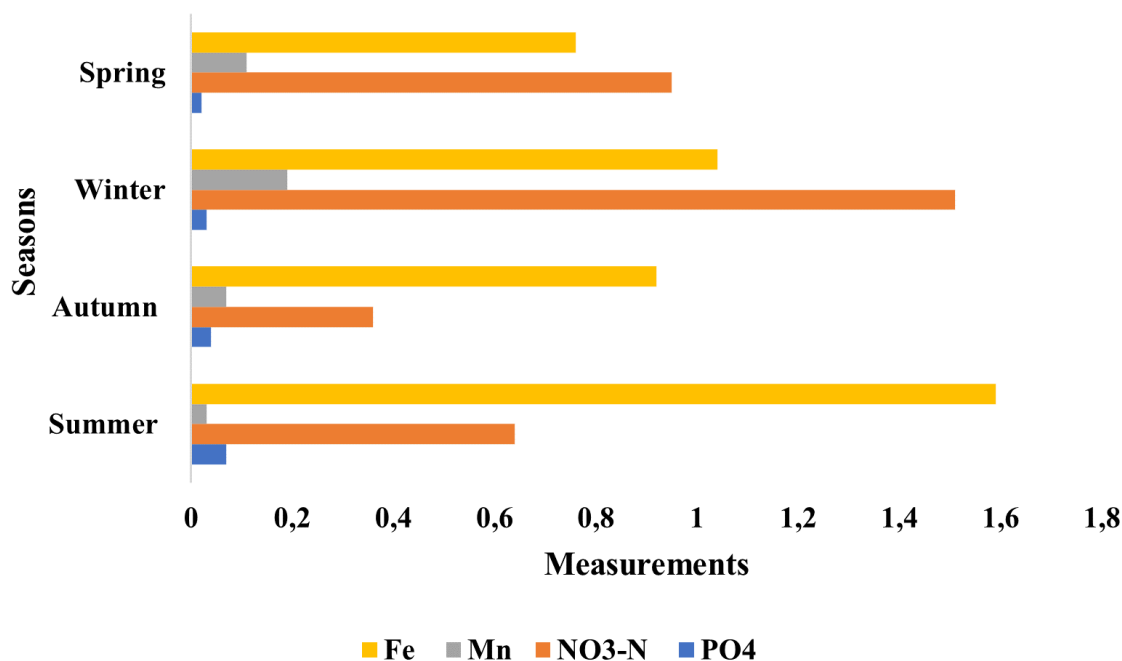


Fig. 4.3 Collective chemical parameter variations during the observed seasons in the Inyaka Dam.

iii. Microbiological parameters

Total coliforms and E. coli

Total coliforms are a group of coliform bacteria found naturally in soil, water, warm-blooded animals, and humans (Bai et al., 2022). *E. coli* forms a significant part of this group because it is also used to indicate faecal pollution by mammals in water resources (Nowicki et al., 2021). The standard limit for *E. coli* is set at 1.3×10^2 CFU/100mL (Table 1) according to the RQOs for the Inyaka Dam. During the surveyed periods, *E. coli* compliance was observed only in winter and spring (6.52; 9.54 MPN/100mL) when rainfall was minimal (Table 4.1). This result indicates less sediment run-off and disturbance, leading to cleaner water in the upper layer of the dam. In addition, these observations are congruent with those reported by (Saiful Islam et al., 2019) and in both dry and wet seasons. The observed mean TC measurements for all surveyed seasons were $(1.25910 \times 10^3 \pm 388.32)$, (7.5443×10^2) , (3.59306×10^3) , and (1.262130×10^4) , respectively (Table 4.1). The highest mean TC measurements were observed in spring when precipitation and temperature were moderate.

Faecal streptococci and Bifidobacteria

Very few studies have investigated the effect of these groups of microorganisms on freshwater quality (Pagals & Grabow, 1996; Qalmoun et al., 2022). Although these groups are not severe contaminants, their monitoring is significant in environmental and recreational waters (DWAf, 1996b). The mean measurement for FS ranged between 1.30 and 4.10 CFU/mL. Although these microorganisms are not regulated in the available RQOs for surface water, the values are well below the 1.3×10^2 CFU/100mL limit set for *E. coli*. The seasonal mean measurements for FS were (4.10), (2.43), (3.50), and (1.30CFU/mL), respectively. Higher measurements were observed in summer and winter when climatic conditions differed significantly. On the other hand, Bifidobacteria mean measurements ranged from 0.00 to 3.14 CFU/mL, with the lowest levels observed in spring and highest in autumn (Table 4.1). The seasonal mean measurements for Bifidobacteria regarding seasonal variations were (2.1), (3.14), (0.31), and (0.00), respectively. A higher combined microbial variability was also observed in the summer months (Fig. 4.4). This trend can be attributed to increased temperatures and high rainfall during the summer months.

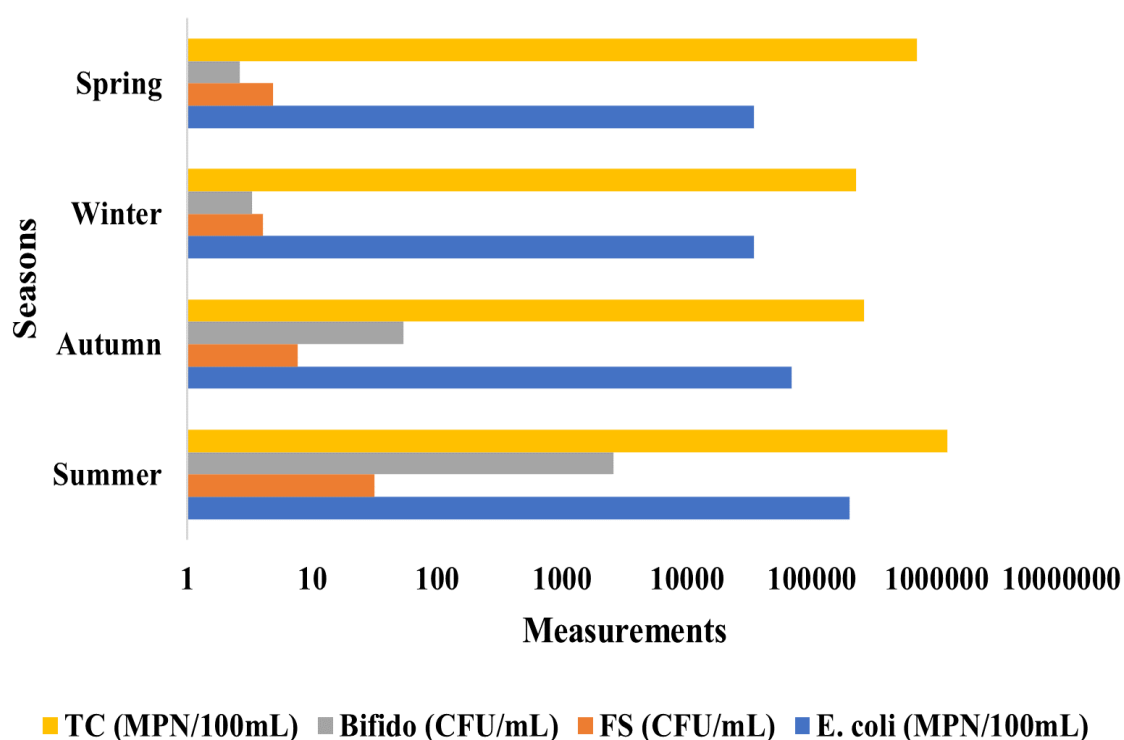


Fig. 4.4 Comparison of the variations in the collective microbiological parameters during the different seasons in the Inyaka Dam

4.6.2. Multivariate Statistical Analysis

i. Correlation Matrix

All- physical parameters, excluding pH, showed a strong and significant linear relationship with either chemical or microbiological parameters during summer. Summer pH was only positively correlated with Mn ($r = 0.469$ and $p < 0.05$). The EC had a strong positive correlation with $\text{NO}_3\text{-N}$ ($r = 0.960$ and $p < 0.01$), whereas rainfall correlated with *E. coli* ($r = 0.677$ and $p < 0.01$). Temperature and turbidity were the only physical parameters to present significant correlations with more than one parameter. The highest correlation coefficient for temperature was recorded for Fe ($r = -0.660$ and $p < 0.01$). Rainfall greatly influences the turbidity of water, especially in the summer and rainy months, because of sediment disturbance (Göransson et al., 2013). Rainfall and turbidity had a positive correlation ($r = 0.756$ and $p < 0.01$) in the summer months at the Inyaka Dam. The correlation structure of the microbiological parameters was rather interesting, as the chemical parameters were correlated with the microbiological parameters. Only rainfall and turbidity correlated with *E. coli* significantly ($r = 0.677$ and $p < 0.01$), whereas FS, Bifidobacteria, and TC all positively correlated with *E. coli* ($r = 0.468$ and $p < 0.05$), ($r = 0.507$ and $p < 0.01$), and ($r = 0.413$ and $p < 0.05$), respectively (Fig. 4.5). This can be attributed to the fact that freshwater sediment contains a variety of microorganisms, and when disturbed by runoff due to heavy rains, microorganisms are released to the surface of the water (Zeki, 2022).

In winter, only EC had multiple correlations with Bifidobacteria ($r = -0.450$ and $p < 0.05$), temperature ($r = 0.476$ and $p < 0.01$), Fe ($r = 0.583$ and $p < 0.01$), PO_4 ($r = -0.484$ and $p < 0.01$), and $\text{NO}_3\text{-N}$ ($r = 0.500$ and $p < 0.01$). In addition, EC showed positive and negative linear relationships, which denotes that any significant change in the EC of the Inyaka Dam may affect other parameters either positively or negatively (Appendix 4 Fig. S4.4). The autumn and spring seasons received almost the same volume of rainfall (66.66 and 51.21). However, in autumn, *E. coli* showed more correlated relationships than any other parameter, whereas in spring, $\text{NO}_3\text{-N}$ showed more meaningful correlations. *E. coli* showed correlations with PO_4 ($r = 0.612$ and $p < 0.01$), FS ($r = 0.579$ and $p < 0.01$), TC ($r = 1.000$ and $p < 0.01$), and turbidity ($r = 0.467$ and $p < 0.05$). To support these findings, a study conducted in the Turkish lake Imrahor reported that surface runoff due to rainfall causes high microbial counts (Zeki, 2022). In spring, $\text{NO}_3\text{-N}$ showed correlations with EC ($r = 0.955$ and $p < 0.01$), Bifidobacteria ($r = -0.505$ and $p < 0.05$), TC ($r = 0.538$ and $p < 0.01$), Mn ($r = 0.586$ and $p < 0.01$), and turbidity ($r = 0.516$ and

$p < 0.01$) (Appendix 4, Figs. S4.3 and S4.5). The positive correlation of $\text{NO}_3\text{-N}$ with the microbiological indicators, Mn, and turbidity indicates a surface drainage, sewage system or animal waste contamination (Kothari et al., 2021). Fig. 4.5 presents the Pearson correlation coefficients (r) of all the measured parameters during the summer.

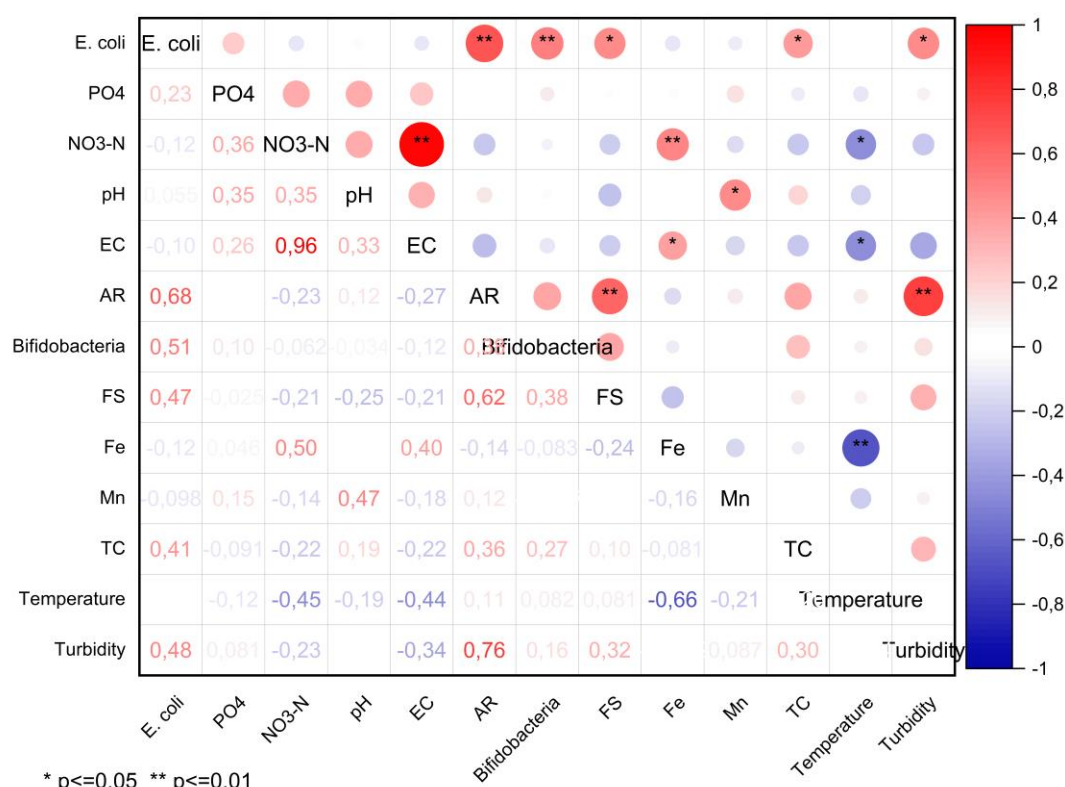


Fig. 4.5 Correlation matrix for all measured water quality parameters during summer at Inyaka Dam

ii. Principal Component Analysis (PCA)

PCA is a statistical method that demonstrates the relationship between measured parameters by reducing them to fewer factors (Rahman et al., 2021). Fig. 4.6 and Fig. 4.7 present the PCA analyses and the temporal variations in water quality among the observed seasons. Three components (PC1, PC2, PC3) were used to create the 3D plot. In summer, the three components represented 60.4%, 52.5% in winter 67.2% in autumn, and 63.1% in spring, of the total variation (Appendix 4, Tables S4.1, S4.2, S4.3, and S4.4).

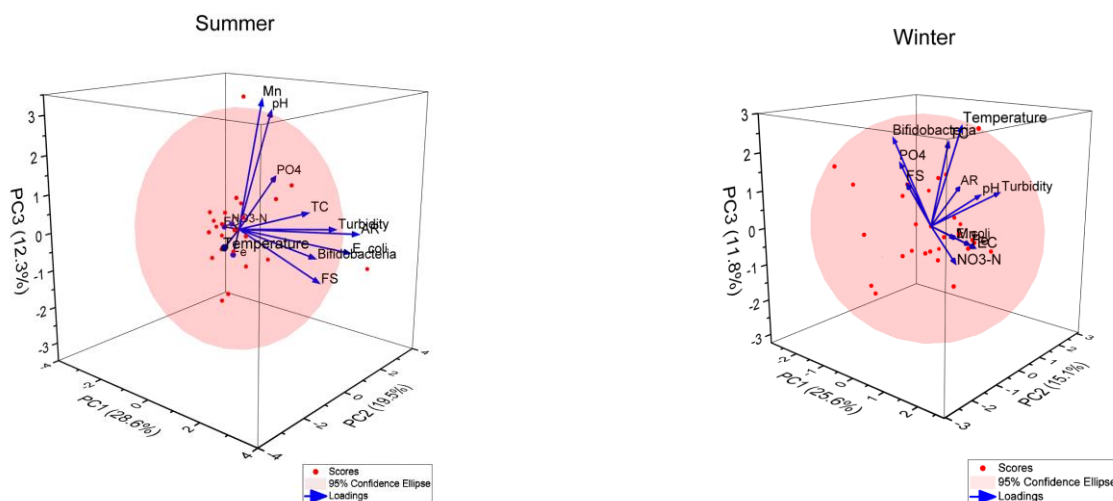


Fig. 4.6 A comparison of 3D PCA plots for analysed parameters in the summer and winter seasons for the Inyaka Dam

The effect of seasonal sampling is illustrated in Fig. 4.6. In the summer season, *E. coli*, AR, turbidity, Bifidobacteria, and FS correlate more to the component (PC1) and are the most influential in determining water quality. In winter, turbidity, TC, FS, and Mn were the most influential parameters regarding the first component (PC1) largely due to stratification. In addition, the results show that rainfall affects the direction of the linear relationship between Bifidobacteria and the first component in the summer (0.57) and winter (-0.76) (Appendix 4, Tables [S4.1](#) & [S4.2](#)). The autumn and spring PCA biplot comparison presented in Fig. 4.7 shows different relationships between the parameters.

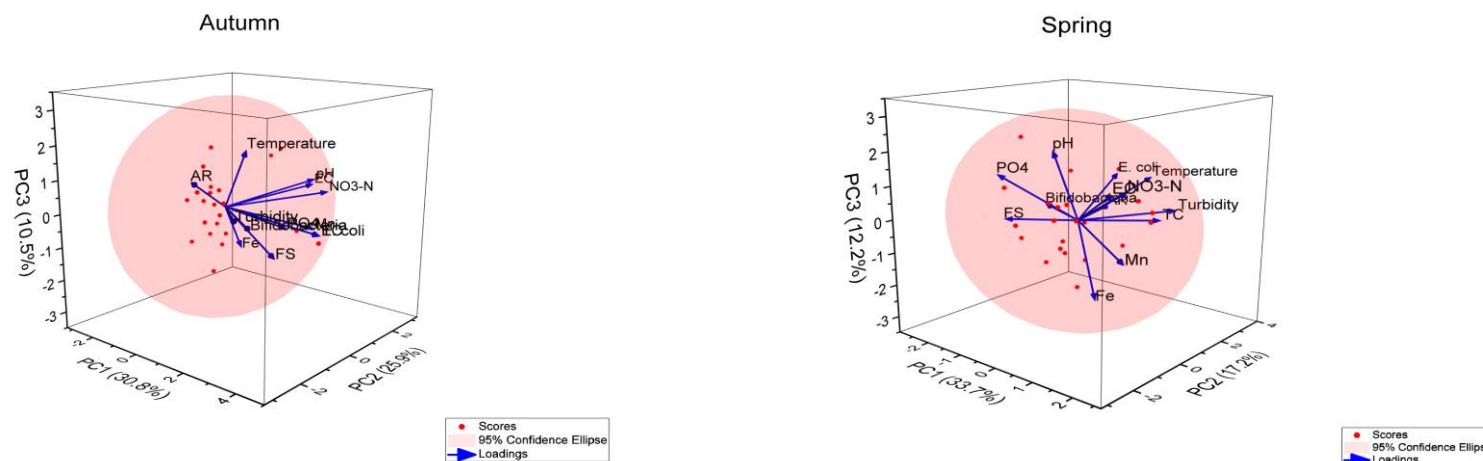


Fig. 4.6 A comparison of 3D PCA plots for analysed parameters in the autumn and spring seasons for the Inyaka Dam

The autumn and spring seasons received almost the same average rainfall (66.66 and 51.21mm). This denotes that rainfall was not responsible for the parametric variations between the two seasons. However, the Inyaka Dam is a lentic system that is influenced by stratification. As such, during spring low temperatures, Mn, NO₃-N, EC, turbidity, and TC concentrations were high, a phenomenon also observed in winter. The autumn season displayed the opposite of what was observed in the winter and spring seasons, with lower concentrations of NO₃-N, Mn, EC, turbidity, and TC. Notwithstanding, PCA showed that the water quality in spring was influenced by turbidity, TC, FS, and Mn, while autumn was influenced by TC, *E. coli*, PO₄, FS, and turbidity.

4.6.3. Water quality properties of the Crocodile River

Similar water quality properties were compared at the Crocodile River to those at the Inyaka Dam. The results for all four seasons are presented in Table 4.2. The parameters were measured against the Crocodile River RQOs and TWQG, where RQOs were unavailable. Table 4.2 demonstrates several non-compliances during summer for EC, PO_4^{3-} , *E. coli*, Mn, Iron, and FS. This is due to the greater amount of rainfall (120.10mm) received in summer compared to autumn, winter, and spring (Table 4.2). Moreover, parameter values decrease with the decreasing rainfall received in subsequent seasons.

Table 4.2: Mean and standard deviation values of the physicochemical and microbiological properties of the Crocodile River for four seasons.

	Summer			Autumn			Winter			Spring			Limits	
	N (Total)	Mean	SD	N (Total)	Mean	SD	N (Total)	Mean	SD	N (total)	Mean	SD	RQOs	TWQG
<i>E. coli</i>	22	201858.00	5293.13	18	69388.72	1165.83	37	34440.73	112.65	25	45598.84	771.92	130	130
PO_4^{3-}	22	0.39	0.32	18	0.22	0.10	37	0.24	0.23	25	0.38	0.61	0.125	0.025
$\text{NO}_3\text{-N}$	22	4.20	5.92	18	1.75	1.32	37	1.79	0.74	25	3.10	3.31	6	6
pH	22	7.68	0.28	18	7.79	0.39	37	7.59	0.31	25	7.82	0.22	6.5-8.0	6.5 – 8.5
EC	22	204.75	58.71	18	91.66	64.86	37	90.47	123.03	25	140.34	102.51	55	40
Aver. Rainfall	22	120.10	62.11	18	51.23	32.79	37	7.89	9.12	25	43.89	35.38	N/A	N/A
Bifidobacteria	22	2595.05	6154.44	18	53.67	111.18	37	3.30	2.47	25	2.64	3.67	-	-
FS	22	31.64	13.93	18	7.67	6.49	37	4.03	5.97	25	4.88	4.13	-	30
Fe	22	2.14	2.37	18	0.65	0.17	37	0.39	0.18	25	0.73	0.18	-	0.01
Mn	22	0.65	0.72	18	0.19	0.06	37	0.18	0.07	25	0.23	0.08	0.18	
Coliforms	22	1220190.14	2257.93	18	261620.39	251.63	37	224903.16	550.35	25	696279.12	1787.05	-	-
Temperature	22	23.11	1.24	18	22.40	1.07	37	16.49	4.05	25	22.31	1.36		
Turbidity	22	52.32	52.45	18	13.66	10.87	37	4.52	1.70	25	18.02	8.46	-	-

i. Physical parameters

The mean pH measured across the observed seasons ranged from 7.59 to 7.82, with the highest value recorded in spring. The seasonal mean pH regarding seasonal variation for all observed seasons was (7.68), (7.79), (7.59) and (7.82), respectively (Table 4.2). These pH measurements were well within the set limits for the Crocodile River (Table 4.2). In addition, the pH range reported in this study is consistent with recent studies conducted in the Crocodile River (Che et al., 2021; Phungela et al., 2022).

The electrical conductivity of a river is greatly influenced by river flow, run-off of sediments due to heavy rainfall, and anthropogenic activities that disturb settled sediments. The highest EC measured across the observed seasons was in summer (204.75 mS/m) when the rainfall was also highest (120.10 mm). The lowest (90.47 mS/m) was measured in winter when rainfall was minimal (7.89 mm) compared to the other seasons (Table 4.2). This is consistent with studies conducted in River Rwizi, Uganda (Ojok et al., 2017). According to the Crocodile River standard limits, the observed EC measurements across all seasons denote that the water is unsuitable for contact recreation and inconducive for aquatic ecosystems. The EC of water is strongly related to turbidity in a manner of direct proportion. When turbidity is high, the EC also increases because of the disturbance of settled salts.

Rainfall is another parameter that significantly influences turbidity and EC because it is responsible for the run-off of salts and disturbance of settled sediments, especially in river systems (Zeki, 2022). The highest measured rainfall was observed in the summer months, which is typical in the region. The mean rainfall for all seasons was (120.10), (51.23), (7.89), and (43.89mm), respectively (Table 4.2). The lowest mean rainfall was observed in the winter, which is typically a dry season.

The seasonal mean temperatures ranged between 16.49°C and 23.11°C. The lowest mean temperature was measured in winter, whereas the highest was measured in summer. The temperature measurements in the spring and autumn months were almost similar (22.31°C and 22.40°C), which means that temperature did not play a significant role in influencing meaningful variation in spring and autumn. To support the foregoing observations, the collective physical parameters showed greater variability in the summer months, with EC and rainfall being influential across all seasons (Fig. 4.8).

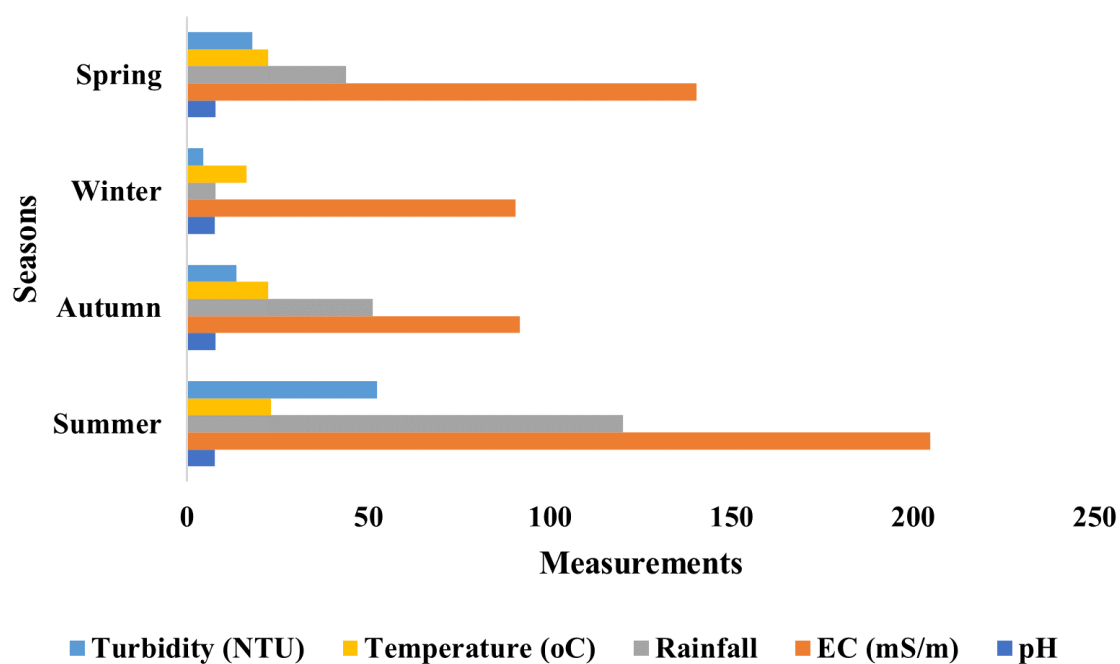


Fig. 4.7 Comparison of the variations in the collective physical parameters during different seasons in the Crocodile River

ii. Chemical parameters

Phosphate and Nitrate

Flowing systems such as rivers are supposed to have the least risk of eutrophication compared to lakes, ponds, and dams. Although a wastewater treatment plant discharges effluent into the Crocodile River, there are no reported incidences of eutrophication in the area. Nonetheless, the phosphate measurements for the observed seasons ranged between 0.22 and 0.39mg/L (Table 4.2), which is well above the specified limits for the Crocodile River. A similar study (Phungela et al., 2022) also reported high measurements of phosphates at the exact location and attributed this to the nearby wastewater treatment plant. The highest tolerable limit for phosphate is 0.125mg/L (Table 4.2). The seasonal mean concentrations of nitrate were (4.20), (1.75), (1.79), (3.10). The highest nitrate measurements were obtained in summer (4.20 mg/L) and spring (3.10 mg/L) (Table 4.2), where rainfall measurements were higher compared to winter and autumn. These measurements were well below the limit of 6mg/L for nitrate in the Crocodile River.

Iron and Manganese

Fe and Mn concentrations in the eastern part of the Crocodile River are considered risky parameters that require frequent monitoring. This is due to the manganese processing company upstream of this study location. Sources of manganese in surface water can originate from both natural and anthropogenic activities (Matveeva et al., 2022). The seasonal mean Mn for all observed seasons was (0.65), (0.19), (0.18), and (0.23) respectively (Table 4.2). In contrast to the Inyaka Dam, only the winter season in the Crocodile River complied with the aquatic ecosystem standard limit of 0.18 mg/L (Table 4.2). The Crocodile River experienced high manganese concentrations during the summer rainy seasons (0.65 mg/L). This is in congruence with previous studies reporting a range between 0.22-0.34 mg/L (Nde et al., 2021).

The Fe concentrations ranged from 0.39 to 2.14 mg/L. The limit for domestic use is set at 0.01 mg/L (DWAF, 1996a) thus, the water is not suitable for domestic use. The highest measurements were obtained in summer when rainfall was highest in the Crocodile River, while the lowest concentration was measured in the winter. This is similar to the observations made by (Ojok et al., 2017) in River Rwizi, Uganda, where Fe concentrations exceeded permissible limits. Consistent with the foregoing observations, the summer months with high rainfall were responsible for the highest variability (Fig. 4.9).

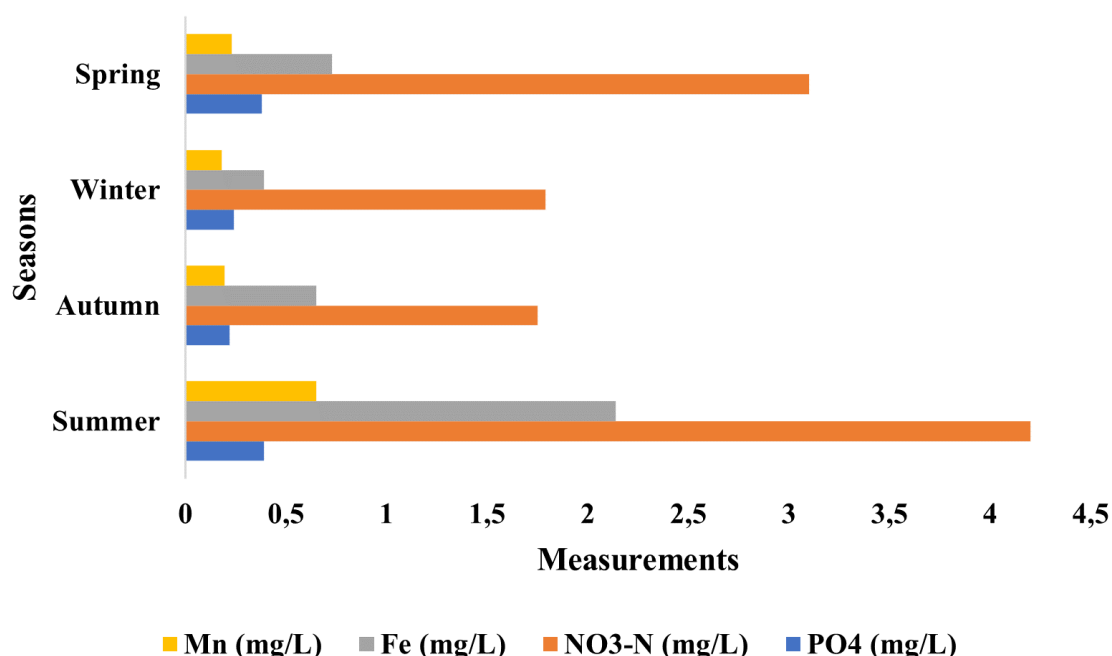


Fig. 4.9 Comparison of variations in the collective chemical parameters during different seasons in the Crocodile River

iii. Microbiological parameters

Total coliforms and E. coli

Total coliforms are used to evaluate the treatment efficiency in water treatment plants and only indicate gaps for improvement in the treatment process. *E. coli*, on the other hand, is used as the preferred indicator of pollution of water resources (Nowicki et al., 2021). The general limit for *E. coli* in the Crocodile River is 1.3×10^2 CFU/100mL (Table 4.2), and for the current investigation, the mean *E. coli* measurements ranged between 3.444×10^4 and 2.01858×10^5 MPN/100mL (Table 4.2). Microbial growth variability is high in lotic systems such as rivers as illustrated in Fig 4.10. This variation pattern is typical of bacterial communities, especially in the dry and wet seasons (Nguyen et al., 2022). The lowest measurement was observed in winter (3.444073×10^4), while the highest was observed in the summer (2.01858×10^5). All measurements exceeded the prescribed limits, as detailed in Table 4.2.

Faecal streptococci and Bifidobacteria

Faecal streptococci (FS) are used to indicate the presence of pollution of animal faecal origin (Scott Winton et al., 2019a) and are preferred indicators of pollution in recreational waters such as marine environments, because they can survive longer than coliform bacteria in water. The maximum limit for FS in recreational waters is at most 30 CFU/mL (Table 4.2). The results obtained from the different seasons in the current observation (31.64, 7.67, 4.03, and 4.88 CFU/mL) show only non-compliance in the summer. Interestingly, summer is usually the season when recreational activities increase. Bifidobacteria measurements ranged between 2.64 and 2595 CFU/mL, with the highest measurement obtained in the summer and the lowest in winter. The limit for Bifidobacteria is not set for any water resource in South Africa because it is considered an unsuitable indicator of pollution (Lamendella et al., 2008).

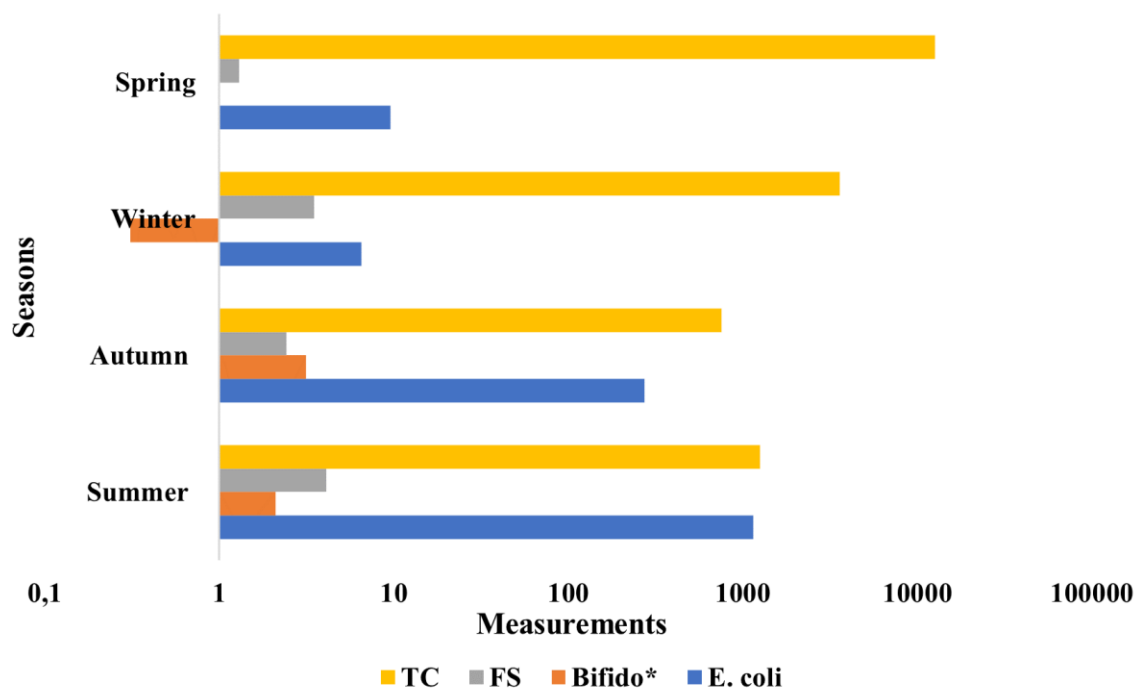


Fig. 4.8 Comparison of the variations in the collective microbiological parameters during the observed seasons in the Crocodile River

4.6.4. Multivariate Statistical Analysis

i. Correlation Matrix

All the studied parameters except temperature showed a correlation with one or more parameters in the summer. $\text{NO}_3\text{-N}$ and EC showed more meaningful and significant linear relationships with other parameters. However, $\text{NO}_3\text{-N}$ showed no correlation with any of the microbiological parameters, whereas EC only showed a correlation with FS ($r = 0.568$ and $p < 0.01$). Rainfall also showed a positive correlation with Fe ($r = 0.485$ and $p < 0.05$), Mn ($r = 0.520$ and $p < 0.05$), and turbidity ($r = 0.548$ and $p < 0.01$). The pH measurement presented a negative correlation with EC ($r = -0.675$ and $p < 0.01$), Fe ($r = -0.687$ and $p < 0.01$), Mn ($r = -0.634$ and $p < 0.01$), and turbidity ($r = -0.694$ and $p < 0.01$) (Fig. 4.11). Congruent with these findings, a study by (Kothari et al., 2021)) reported negative Fe ($r = -0.3$), and low correlation values EC ($r = 0.2$) and turbidity ($r = 0.01$).

In winter, *E. coli* had the most meaningful linear relationships with other parameters, while rainfall did not correlate with all the parameters. *E. coli* correlated with PO_4 ($r = 0.907$ and $p < 0.01$), $\text{NO}_3\text{-N}$ ($r = -0.609$ and $p < 0.01$), TC ($r = 0.807$ and $p < 0.01$), Bifidobacteria ($r = 0.470$

and $p < 0.01$), and EC ($r = -0.337$ and $p < 0.05$), NO₃-N presented a correlation with EC ($r = 0.669$ and $p < 0.01$), Bifidobacteria ($r = -0.404$ and $p < 0.05$), FS ($r = 0.361$ and $p < 0.05$), and TC ($r = -0.622$ and $p < 0.01$) while the correlation of NO₃-N with microbiological parameters in winter was negative (Appendix 4, Fig. S4.8). The recorded winter rainfall was low for the Crocodile River (7.89 mm). As such, this correlation pattern can be attributed to possible contamination from sewage systems, agricultural activities, or animal wastes (Kothari et al., 2021).

Like the Inyaka Dam, the spring and autumn seasons in the Crocodile River showed less difference in the amount of rainfall received (43.89 and 51.23 mm). Rainfall only showed a correlation in the spring season with Mn ($r = -0.691$ and $p < 0.01$), while there was no correlation with any of the parameters in autumn. Only *E. coli* showed more than one meaningful linear relationship with PO₄³⁻ ($r = 0.554$ and $p < 0.05$) and turbidity ($r = 0.658$ and $p < 0.01$) in the autumn season. In spring, *E. coli* showed correlations with EC ($r = -0.431$ and $p < 0.05$), Bifidobacteria ($r = 0.587$ and $p < 0.01$), and TC ($r = 0.647$ and $p < 0.01$) (Appendix 4, Figs. S4.6 and S4.7). These observations are consistent with those reported by (Solaiman et al., 2020) where *E. coli* correlated with EC ($r = 0.25$ and $p < 0.001$).

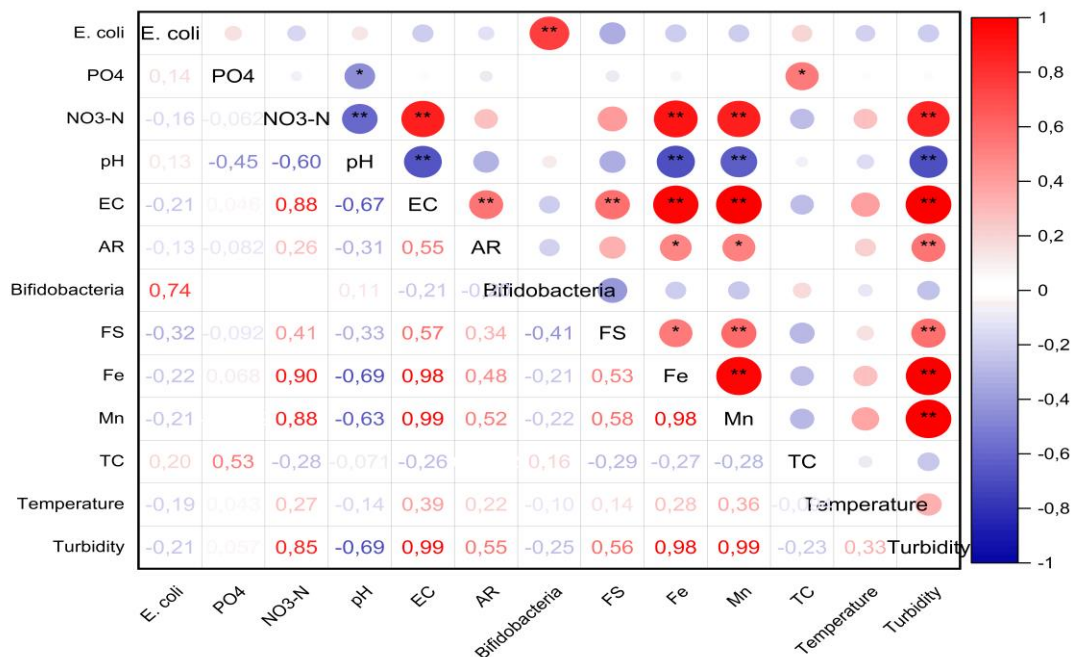


Fig. 4.9 Correlation matrix for all observed water quality parameters in summer in the Crocodile River

ii. Principal Component Analysis (PCA)

The first, second, and third components (PC1, PC2, PC3) were used to compute the PCA plots for all observed seasons, with summer explaining 74.7%, winter 64.2%, autumn 58.2%, and spring 58.8% of the total variance (Appendix 4, Tables [S4.5](#), [S4.6](#), [S4.7](#), and [S4.8](#)). In addition, Figs. 4.12 & 4.13 illustrate the parametric variance and influence across the four seasons. The longer the arrow, the higher the correlation and subsequent influence on water quality.

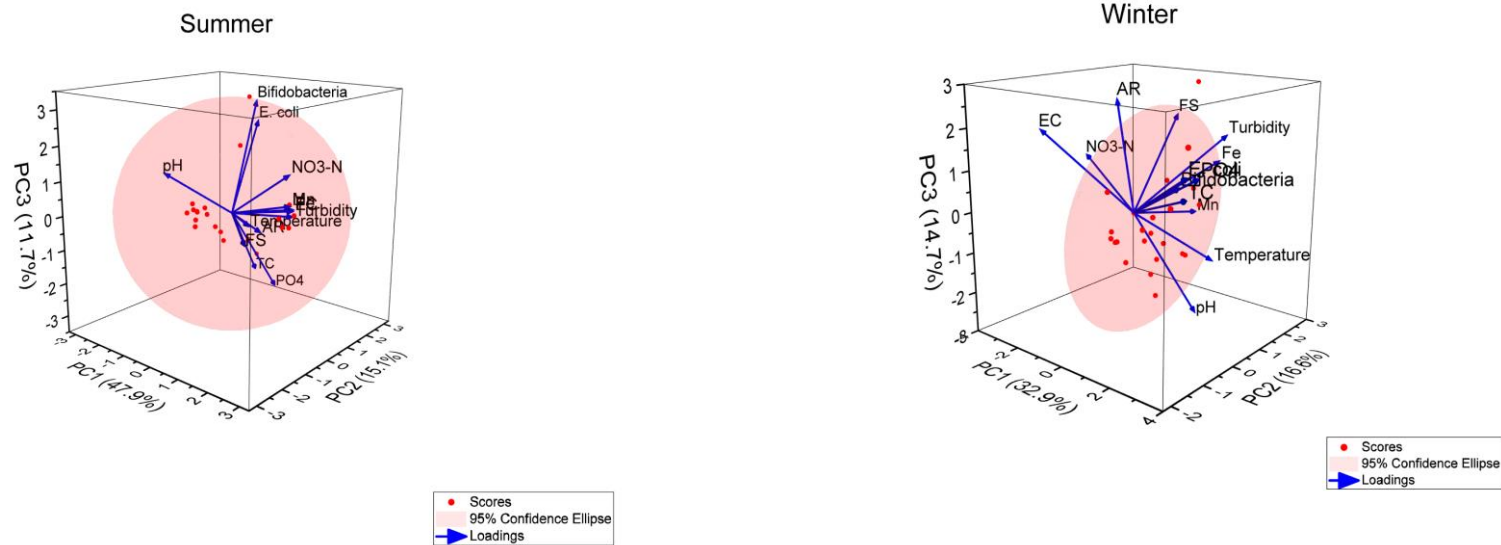


Fig. 4.10 PCA plots for analysed parameters for the summer and winter seasons in the Crocodile River.

The Crocodile River presented a different parametric variance and influence compared to the pattern observed in the Inyaka Dam. In the summer season, eight (turbidity, EC, Fe, Mn, NO₃-N, pH, FS, and AR) essential parameters influenced water quality compared to five in the Inyaka Dam.

Differences in the number of influential parameters can be attributed to other factors, such as ecosystem dynamics. The winter season revealed five parameters (*E. coli*, PO₄³⁻, TC, NO₃-N, and Bifidobacteria) with the greatest influence on water quality.

The autumn and spring seasons in the Crocodile River differed significantly from those observed in the Inyaka Dam. Fig. 4.13 illustrates the comparison of the two seasons in the Crocodile River.

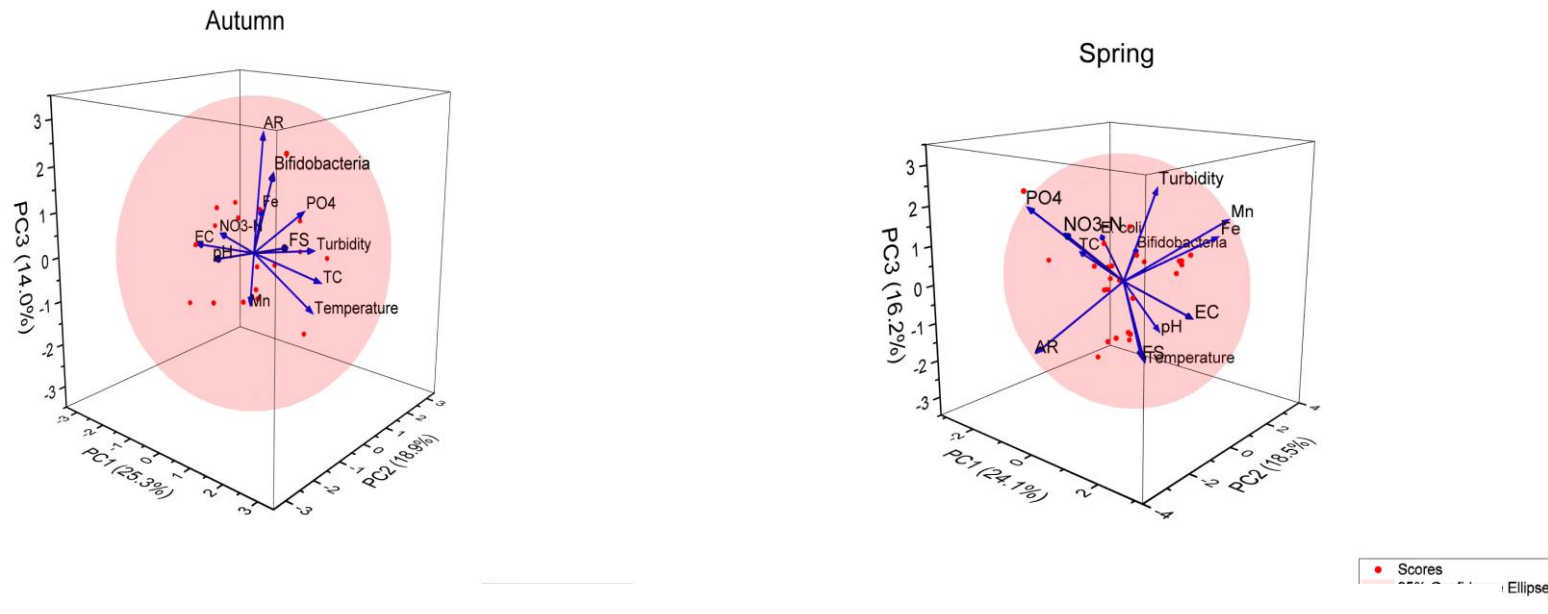


Fig. 4.11 PCA plots for analysed parameters for the summer and winter seasons in the Crocodile River

In autumn, turbidity, *E. coli*, and PO₄ levels were responsible for water quality deterioration. Despite receiving almost similar volumes of rainfall (51.23 and 43.89 mm), the water quality in spring was highly influenced by different parameters such as Mn, Fe, and rainfall. According to (Che et al., 2021), Mn levels in the Crocodile River vary during the wet seasons. This explains why Mn contributes to water quality degradation. These observations are also consistent with other studies (Somerset et al., 2015) that reported significant changes in physicochemical and heavy metal concentrations in South African rivers.

4.7. Conclusions

This study examined the impact of weather variations on water quality in the Inyaka Dam and Crocodile River, South Africa, highlighting the threat to water security intensified by anthropogenic activities and climate change. Water quality is affected by temperature and rainfall, which are key contributors to climate variability. The study found that water pollution in these aquatic systems stems from both natural processes, such as stratification, and human activities like wastewater treatment and agriculture.

In the Inyaka Dam, summer rainfall significantly increased microbial counts, including *E. coli*, and turbidity. Pearson correlation analysis showed strong positive correlations between rainfall and *E. coli* (0.68) and turbidity (0.76) (Fig. 5). Conversely, an increase in temperature led to a decrease in pollutants like Fe, EC, and NO₃-N due to negative correlations. In contrast, the Crocodile River's summer season was marked by high levels of EC, turbidity, Fe, Mn, and NO₃-N, which negatively correlated with pH.

Winter conditions in both water bodies featured low rainfall and temperatures, with continued water quality deterioration in the Crocodile River due to reduced water flow. In winter, the Crocodile River exhibited high levels of *E. coli*, total coliforms, Bifidobacteria, and PO₄³⁻. In contrast, the Inyaka Dam showed decreased levels of PO₄³⁻ and Bifidobacteria, while Fe, EC, and NO₃-N increased.

The study concludes that water quality degradation, particularly in summer, is driven by rainfall and effluent discharges, with pollution levels often exceeding permissible limits. The findings underscore the need for tailored water quality management strategies that consider the unique characteristics of lentic (Inyaka Dam) and lotic (Crocodile River) systems. It advocates for ecosystem-specific, climate-resilient approaches to effectively mitigate water pollution and enhance water security. This research emphasises the importance of distinguishing between different aquatic ecosystems when designing water pollution control strategies and policies.

4.8. Future directions

Despite the availability of well-defined local water quality monitoring guidelines and standard limits, certain water quality parameters are crucial to the overall quality of surface waters. This is particularly important because most surface water resources serve as discharge points or

receiving sources for effluent discharges and agricultural runoff. Therefore, parameters such as dissolved oxygen (DO), 5-day biological oxygen demand (BOD₅), and chemical oxygen demand (COD) should be included in any study related to surface water pollution. This study did not assess these parameters, which is considered a major limitation of the findings. Consequently, future studies should emphasise these parameters in addition to those outlined in local water quality monitoring programs.

4.9. CRediT author statement.

Lazarus Katlego Mogane: was responsible for conceptualisation, data collection and curation, writing of the first draft and final revision. **Esper J. Ncube:** was responsible for the conceptualisation, editing, and reviews. **Titus A.M. Msagati:** was responsible for the conceptualisation, editing and reviews. **Tracy M. Masebe:** was responsible for the conceptualisation, Methodology, editing, and reviews. All the authors read and approved the final draft of the manuscript.

4.10. Statements and declarations

Competing interests: The authors declared no conflicting interests regarding publishing this paper.

Data availability: The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Funding: No funding was obtained for this study.

4.11. References

- Abbasi, T., & Abbasi, S. A. (2012). Why Water-Quality Indices. In *Water Quality Indices*. <https://doi.org/10.1016/b978-0-444-54304-2.00001-4>
- Ahmed, K. S., Rahman, A. K. M. L., Sarkar, M., Islam, J. B., Jahan, I. A., Moniruzzaman, M., Saha, B., & Bhounik, N. C. (2016). Assessment on the level of contamination of Turag river at Tongi area in Dhaka. In *Bangladesh J. Sci. Ind. Res* (Vol. 51, Issue 3). www.banglajol.info
- Ahmed, T., Zounemat-Kermani, M., & Scholz, M. (2020). Climate change, water quality and water-related challenges: A review with focus on Pakistan. In *International Journal of Environmental Research and Public Health* (Vol. 17, Issue 22, pp. 1–22). MDPI AG. <https://doi.org/10.3390/ijerph17228518>
- Bai, V. R., Kit, A. C., Kangadharan, G., Gopinath, R., Varadarajan, P., & Hao, A. J. (2022). Experimental study on total coliform violations in the complied NH₂CL, O₃, and UV treated municipal water supply system. *European Physical Journal Plus*, 137(6). <https://doi.org/10.1140/epjp/s13360-022-02891-5>
- Benariba, H., Habi, M., Morsli, B., & Moulla, A. S. (2022). Assessment of surface water quality in a semi-arid Mediterranean region. Case study of Sikkak dam (north-western Algeria). *Water Practice and Technology*, 17(5), 1177–1196. <https://doi.org/10.2166/wpt.2022.040>
- Che, N. S., Bett, S., Okpara, E. C., Olagbaju, P. O., Fayemi, O. E., & Mathuthu, M. (2021). An Assessment of Land Use and Land Cover Changes and Its Impact on the Surface Water Quality of the Crocodile River Catchment, South Africa. *River Deltas Research - Recent Advances*, 1–34. www.intechopen.com
- Department of Water and Sanitation (DWS) South Africa. (2016). *Classes of water resources and resource quality objectives for the catchments of the Inkomati*. www.gpwonline.co.za
- DWA. (2007). Compiled for: Department of Water Affairs and Forestry Directorate Water Abstraction and Instream Water Use Sub-directorate Environment and Recreation Private Bag X 313 PRETORIA 0001.
- DWAF. (1996a). Department of Water Affairs and Forestry: Water Quality Guidelines for Domestic Use.
- DWAF. (1996b). WATER QUALITY GUIDELINES VOLUME 2 RECREATIONAL USE Department of Water Affairs and Forestry.
- Hasan, Md. K., Shahriar, A., & Jim, K. U. (2019). Water pollution in Bangladesh and its impact on public health. *Heliyon*, 5(8), e02145. <https://doi.org/10.1016/j.heliyon.2019.e02145>
- Isiuku, B. O., & Enyoh, C. E. (2020). Pollution and health risks assessment of nitrate and phosphate concentrations in water bodies in South Eastern, Nigeria. *Environmental Advances*, 2. <https://doi.org/10.1016/j.envadv.2020.100018>
- Kothari, V., Vij, S., Sharma, S. K., & Gupta, N. (2021). Correlation of various water quality parameters and water quality index of districts of Uttarakhand. *Environmental and Sustainability Indicators*, 9. <https://doi.org/10.1016/j.indic.2020.100093>

- Lamendella, R., Santo Domingo, J. W., Kelty, C., & Oerther, D. B. (2008). Bifidobacteria in feces and environmental waters. *Applied and Environmental Microbiology*, 74(3), 575–584. <https://doi.org/10.1128/AEM.01221-07>
- Machender, G., Dhakate, R., & Narsimha Reddy, M. (2014). Hydrochemistry of groundwater (GW) and surface water (SW) for assessment of fluoride in Chinnaeru river basin, Nalgonda district, (AP) India. *Environmental Earth Sciences*, 72(10), 4017–4034. <https://doi.org/10.1007/s12665-014-3291-9>
- Magesh, N. S., Krishnakumar, S., Chandrasekar, N., & Soundranayagam, J. P. (2013). Groundwater quality assessment using WQI and GIS techniques, Dindigul district, Tamil Nadu, India. *Arabian Journal of Geosciences*, 6(11), 4179–4189. <https://doi.org/10.1007/s12517-012-0673-8>
- Matveeva, V. A., Alekseenko, A. V., Karthe, D., & Puzanov, A. V. (2022). Manganese Pollution in Mining-Influenced Rivers and Lakes: Current State and Forecast under Climate Change in the Russian Arctic. *Water (Switzerland)*, 14(7). <https://doi.org/10.3390/w14071091>
- Mishra, B. K., Kumar, P., Saraswat, C., Chakraborty, S., & Gautam, A. (2021). Water security in a changing environment: Concept, challenges and solutions. *Water (Switzerland)*, 13(4). <https://doi.org/10.3390/w13040490>
- Mnguni, E. S. (2022). Pollution of the Vaal River system in South Africa: A case study. *WIT Transactions on Ecology and the Environment*, 257(2022), 17–26. <https://doi.org/10.2495/WMEI220021>
- Moloantoa, K. M., Khetsa, Z. P., van Heerden, E., Castillo, J. C., & Cason, E. D. (2022). Nitrate Water Contamination from Industrial Activities and Complete Denitrification as a Remediation Option. In *Water (Switzerland)* (Vol. 14, Issue 5). MDPI. <https://doi.org/10.3390/w14050799>
- Nguyen, T. M. H., Le, T. P. Q., Hoang, V. Van, Nhu, D. Le, & Ha, H. T. T. (2022). Fecal indicator bacteria diversity and decay in an estuarine mangrove ecosystem of the Xuan Thuy National Park, Vietnam. *Journal of Water and Health*, 20(6), 915–926. <https://doi.org/10.2166/wh.2022.310>
- Nowicki, S., DeLaurent, Z. R., De Villiers, E. P., Githinji, G., & Charles, K. J. (2021). The utility of *Escherichia coli* as a contamination indicator for rural drinking water: Evidence from whole genome sequencing. *PLoS ONE*, 16(1 January). <https://doi.org/10.1371/journal.pone.0245910>
- Ojok, W., Wasswa, J., & Ntambi, E. (2017). Assessment of Seasonal Variation in Water Quality in River Rwizi Using Multivariate Statistical Techniques, Mbarara Municipality, Uganda. *Journal of Water Resource and Protection*, 09(01), 83–97. <https://doi.org/10.4236/jwarp.2017.91007>
- Ortega, D. J. P., Pérez, D. A., Américo, J. H. P., De Carvalho, S. L., & Segovia, J. A. (2016). Development of index of resilience for surface water in watersheds. *Journal of Urban and Environmental Engineering*, 10(1), 72–82. <https://doi.org/10.4090/juee.2016.v10n1.007282>
- Pagals, P., & Grabow, W. O. K. (1996). An evaluation of sorbitol-fermenting bifidobacteria as specific indicators of human faecal pollution of environmental water. *Water SA*, 22(3), 235–238.

- Phungela, T. T., Maphanga, T., Chidi, B. S., Madonsela, B. S., & Shale, K. (2022). The impact of wastewater treatment effluent on Crocodile River quality in Ehlanzeni District, Mpumalanga Province, South Africa. *South African Journal of Science*, 118(7–8). <https://doi.org/10.17159/sajs.2022/12575>
- Pisciotta, A., Cusimano, G., & Favara, R. (2015). Groundwater nitrate risk assessment using intrinsic vulnerability methods: A comparative study of environmental impact by intensive farming in the Mediterranean region of Sicily, Italy. *Journal of Geochemical Exploration*, 156, 89–100. <https://doi.org/10.1016/j.gexplo.2015.05.002>
- Qalmoun, A., Bouzrarf, K., & Belqat, B. (2022). Assessment of the ecological status of the Oum Er-rabie River basin (Central Morocco) through physicochemical, bacteriological parameters and biotic indices. *Biologia*, 77(9), 2533–2547. <https://doi.org/10.1007/s11756-022-01128-1>
- Rahman, A., Jahanara, I., & Jolly, Y. N. (2021). Assessment of physicochemical properties of water and their seasonal variation in an urban river in Bangladesh. *Water Science and Engineering*, 14(2), 139–148. <https://doi.org/10.1016/j.wse.2021.06.006>
- Saiful Islam, M., Afroz, R., & Mia, B. (2019). Investigation of surface water quality of the Buriganga River in Bangladesh: Laboratory and Spatial analysis approach. In *J. Biol. Sci* (Vol. 28, Issue 2).
- Sarkar, M., Islam, J. B., & Akter, S. (2016). Pollution and ecological risk assessment for the environmentally impacted Turag River, Bangladesh. *J. Mater. Environ. Sci*, 7(7), 2295–2304.
- Scott Winton, R., Calamita, E., & Wehrli, B. (2019a). Reviews and syntheses: Dams, water quality and tropical reservoir stratification. In *Biogeosciences* (Vol. 16, Issue 8, pp. 1657–1671). Copernicus GmbH. <https://doi.org/10.5194/bg-16-1657-2019>
- Scott Winton, R., Calamita, E., & Wehrli, B. (2019b). Reviews and syntheses: Dams, water quality and tropical reservoir stratification. In *Biogeosciences* (Vol. 16, Issue 8, pp. 1657–1671). Copernicus GmbH. <https://doi.org/10.5194/bg-16-1657-2019>
- Solaiman, S., Allard, S. M., Theresa Callahan, M., Jiang, C., Handy, E., East, C., Haymaker, J., Bui, A., Craddock, H., Murray, R., Kulkarni, P., Anderson-Coughlin, B., Craighead, S., Gartley, S., Vanore, A., Duncan, R., Foust, D., Taabodi, M., Sapkota, A., ... Johanna Björkroth, E. (2020). *Longitudinal Assessment of the Dynamics of Escherichia coli, Total Coliforms, Enterococcus spp., and Aeromonas spp. in Alternative Irrigation Water Sources: a CONSERVE Study*. <https://doi.org/10>
- Somerset, V., Van der Horst, C., Silwana, B., Walters, C., & Iwuoha, E. (2015). Biomonitoring and Evaluation of Metal Concentrations in Sediment and Crab Samples from the North-West Province of South Africa. *Water Air and Soil Pollution*, 226. <https://doi.org/10.1007/s11270-015-2329-2>
- Telegdi, J., Shaban, A., & Trif, L. (2017). Microbiologically influenced corrosion (MIC). *Trends in Oil and Gas Corrosion Research and Technologies: Production and Transmission*, 191–214. <https://doi.org/10.1016/B978-0-08-101105-8.00008-5>
- Uddin, M. G., Moniruzzaman, M., Quader, M. A., & Hasan, M. A. (2018). Spatial variability in the distribution of trace metals in groundwater around the Rooppur nuclear power plant in

- Ishwardi, Bangladesh. *Groundwater for Sustainable Development*, 7, 220–231. <https://doi.org/10.1016/j.gsd.2018.06.002>
- Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. <https://doi.org/10.1016/j.ecolind.2020.107218>
- Uddin, M. J., & Jeong, Y. K. (2021). Urban river pollution in Bangladesh during last 40 years: potential public health and ecological risk, present policy, and future prospects toward smart water management. In *Heliyon* (Vol. 7, Issue 2). Elsevier Ltd. <https://doi.org/10.1016/j.heliyon.2021.e06107>
- Umer, A., Assefa, B., & Fito, J. (2020). Spatial and seasonal variation of lake water quality: Beseka in the Rift Valley of Oromia region, Ethiopia. *International Journal of Energy and Water Resources*, 4(1), 47–54. <https://doi.org/10.1007/s42108-019-00050-8>
- United Nations Water. (2021). *Summary Progress Update 2021: SDG 6-water and sanitation for all*.
- Varol, M., & Şen, B. (2012). Assessment of nutrient and heavy metal contamination in surface water and sediments of the upper Tigris River, Turkey. *Catena*, 92, 1–10. <https://doi.org/10.1016/j.catena.2011.11.011>
- WHO. (2003). *Iron in Drinking-water Background document for development of WHO Guidelines for Drinking-water Quality*.
- Wirnkor, V. A., Ngozi, V. E., Amaobi, C., Ebere, E. C., Collins Emeka, A., & Ebere, C. (2018). Water pollution scenario at river Uramurukwa flowing through owerri metropolis, Imo State, Nigeria. *International Journal of Advanced Scientific Research Wwww.Allscientificjournal.Com*, 40, 40–46. www.allscientificjournal.com
- Zeki, S. (2022). A preliminary evaluation of microbial water quality in the irrigation pond. *Water Environment Research*, 94(7). <https://doi.org/10.1002/wer.10757>

5. Exploring the physicochemical drivers of bacterial community dynamics in lotic and lentic freshwater ecosystems

5.1. Preamble

This chapter adopts a multidisciplinary approach to examine the structure and seasonal dynamics of bacterial communities in aquatic ecosystems using high-throughput sequencing and metagenomic analysis. Emphasis is placed on how temporal variation shapes microbial community composition, diversity, and functional potential, providing an opportunity to unravel previously inaccessible layers of biodiversity.

5.2. Abstract

The freshwater ecosystem is a habitat for diverse microbial communities essential in several natural processes, such as nutrient cycling and ecosystem stability. The drivers of bacterial community dynamics across hydrologically distinct systems, such as lotic and lentic, are regional and system dependent. This study investigates the influence of variations in physicochemical parameters on community structure and bacterial diversity in the Inyaka Dam and Crocodile River using 16S rRNA sequencing. Relative abundance results revealed that the phylum *Proteobacteria* and class *Gammaproteobacteria* are dominant in the Crocodile River and Inyaka Dam across all seasons. Moreover, Alpha diversity indices (Chao1, ACE, Jackknife, Shannon-Weiner, and Simpson) revealed significantly higher bacterial richness and evenness in the Crocodile River compared to the Inyaka Dam, irrespective of the season. This observation is associated with high nutrient levels in Mn, Fe, PO_4^{3-} , and electrical conductivity. The Inyaka Dam maintained a homogeneous water column with nutrients below permissible limits, which accounted for the observed low species abundance. Beta diversity was assessed using Principal coordinate analysis (PCoA), supported by ANOSIM ($R = 0.749$, $p = 0.003$), which demonstrated distinct microbial community assemblages between the sites, indicating strong environmental filtering. The study findings highlight the significance of hydrological context in structuring freshwater microbiomes and underscore the critical role of bacteria as sensitive indicators of environmental variation. Thus, there is a need to develop a microbial metrics system for ecosystem monitoring and management.

5.3. Background

Diversity of microorganisms in freshwater is determined primarily by environmental variations (Jiang et al., 2021; Li et al., 2023). Understanding these relationships is essential for improving water and environmental quality, specifically considering climate change, natural human-induced disasters, and water scarcity (Leal Filho et al., 2022). Climate change-related stress affects the stability of microbial structures and their maintenance of beneficial interactions with the ecosystems (Guigard et al., 2025). These stresses lead to alterations of the physicochemical makeup of water, extreme temperature fluctuations, and the introduction of contaminants (Dong et al., 2025). Moreover, it has been demonstrated that water scarcity and anthropogenic activities affect the diversity, distribution, and abundance of microbial communities (Feng et al., 2025; Regina et al., 2021).

There is adequate literature regarding microbial communities and their temporal and spatial patterns (Attiani et al., 2025; Ladau & Elie-Fadrosh, 2019; Wang et al., 2024). However, observations of microorganisms have relied on a few sampled sites, while the spatial and temporal scales of their distributions extend over analogous orders of magnitude to those of the hydrosphere and the atmosphere (Wild et al., 2022). It has also been observed that the composition of aquatic microbial communities changes over time, particularly the dominance of both prokaryotic phyla and genera, as well as their functional groups (Ghate et al., 2021; Karlicki et al., 2024; Yu et al., 2024).

The seasonal variation in bacterial community structure within diverse environments such as soil, fresh, and marine waters has emerged as a critical area of investigation, providing insights into the ecological responses of microorganisms to changing environmental conditions (Gong et al., 2022; Shang et al., 2022). Recent advancements in sequencing technologies have enhanced the ability to delve into the complexities of microbial communities with unprecedented resolution and depth. These technologies have enabled researchers to explore the taxonomic composition and functional potential of bacterial communities in various ecosystems, shedding light on the intricate interplay between microbial diversity and environmental factors.

The significance of freshwater ecosystems is often undervalued in comparison to their marine counterparts (Birnie-Gauvin et al., 2023). However, it is the freshwater biome that supports biomes with the greatest biodiversity (Hitt et al., 2015). As a result of anthropogenic actions, rivers are being altered, leading to deeply exacerbated ecological changes and ultimately

resulting in species loss (Priya et al., 2023). These systems and their alterations have consequences for the health of the entire planet as more pollutants alter the functioning of the atmosphere and ecosystems (Adla et al., 2022). However, rivers are systems that might be recoverable if conservation is prioritised in time (Zhang & Fryirs, 2024). As such, intervention tools such as recovery modelling have been employed to understand the biodiversity residing in such environments are essential (Cepic et al., 2022).

Seasonal bacterial composition has been acknowledged in unfished and fished water bodies, thus highlighting the effect of anthropogenic activities (Frenk et al., 2015; Marmen et al., 2021). Moreover, bacterial dominance in both abundance and composition has revealed different behaviour between fished and non-fished sites and seasons (Ngisiange et al., 2024). Salinity has also been reported to cause clear variations in bacterial community composition, with respect to beta-diversity, even in water bodies in close surroundings (Eswaran & Khandeparker, 2020; Zhang, Qi, et al., 2021). Generally, clear seasonal variations in bacterial composition are expected because of stratification, sediment-water-column exchange processes, and physical impacts switching on and off, such as temperature affecting oxygen usage and subsequently bacterial activity (Kritzberg & Bååth, 2022; Shi et al., 2022).

Several studies have highlighted the significance of seasonal fluctuations in shaping bacterial communities (Dong et al., 2023; Zhang et al., 2019). For instance, the work of Smith et al. (2021) demonstrated distinct shifts in the composition of soil bacteria across seasons, emphasising the impact of temperature, precipitation, and nutrient availability on community dynamics. Similarly, research conducted by Chun et al. (2021) and Gao et al. (2023) in aquatic environments revealed a seasonal succession of bacterial taxa, indicating a strong correlation between water temperature, nutrient levels, and community structure. Understanding the seasonal dynamics of bacterial communities is crucial for multiple reasons. Firstly, it provides valuable insights into the adaptive strategies employed by bacteria to thrive in specific environmental conditions (Ramasamy et al., 2023). Secondly, it contributes to the comprehension of ecosystem functioning, as microbial communities play pivotal roles in nutrient cycling, organic matter decomposition, and pollution source tracking (Li et al., 2020). Lastly, this knowledge has implications for various fields, including agriculture, environmental management, and public health.

In this context, our study aims to contribute to the growing body of knowledge on seasonal bacterial community structure profiling using 16S metagenomics. By examining microbial

communities across different seasons, the study seeks to elucidate the temporal patterns and driving factors influencing bacterial composition in the Crocodile River and Inyaka Dam in the Mpumalanga Province of South Africa. The findings from this research have the potential to enhance an understanding of the ecological dynamics of bacterial communities and as well as informing strategies for sustainable water resource stewardship.

5.4. Materials and Methods

5.4.1. Study area

i. The Inyaka Dam and Crocodile River

The Inyaka Dam is an earth-fill Dam constructed along the Marite River in the Bushbuckridge area of Mpumalanga province, northeast of South Africa. The Dam was initially built to store water for irrigation purposes to support local small-scale farmers. The Dam currently supplies just under a million people with water for drinking and domestic uses. Frame A of Fig. 5.1 illustrates the location (24°53'26.60" S 31°04'45.37"E) of the dam and the surrounding water treatment plants. The Crocodile River is also located within the Mpumalanga province in the eastern part of South Africa. The river receives rainfall in summer and has a subtropical climate. It also bears some considerable ecological significance as it passes along the southern boundary of the Kruger National Park (KNP). Moreover, it is also a transboundary river shared between the Kingdom of Eswatini and the Republic of Mozambique. Frame B of Fig. 5.1 illustrates the location (25° 29'18"S 31° 10'23"E) along with the minor tributaries of the river.

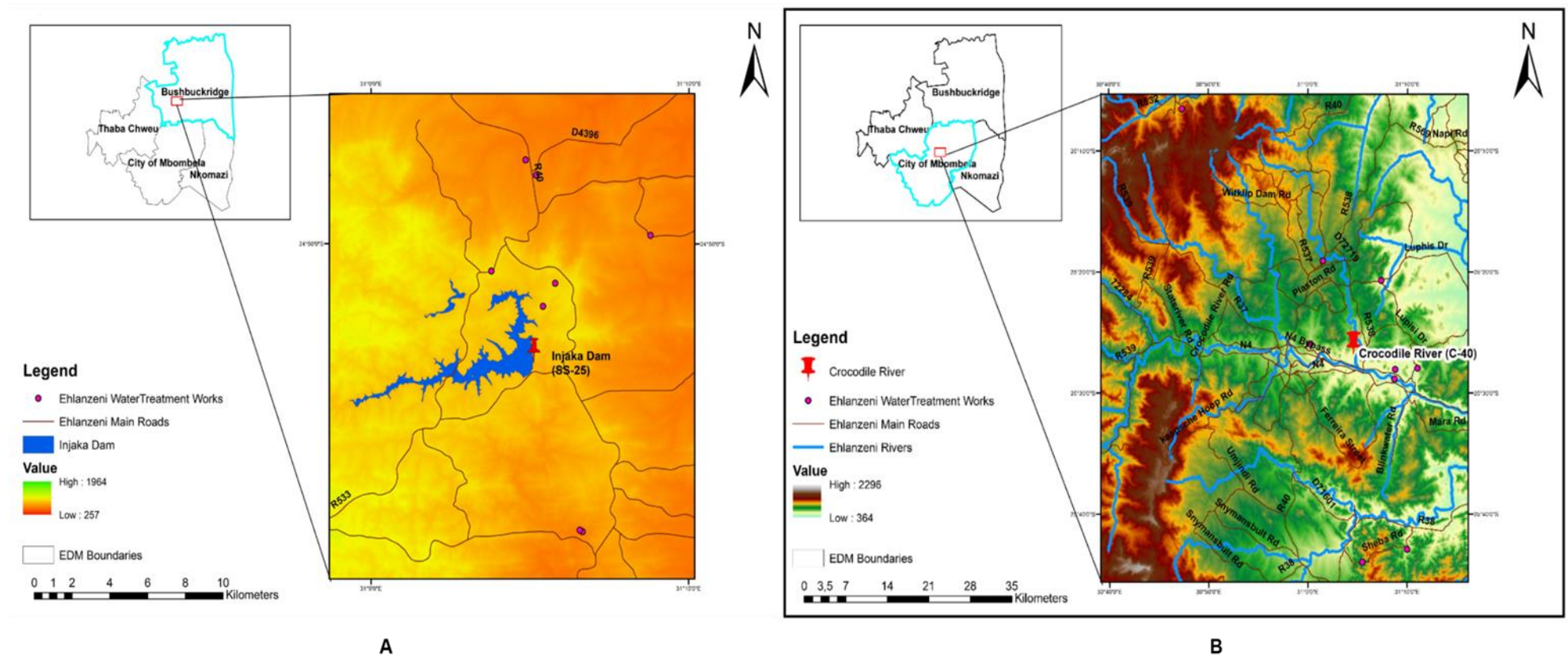


Fig. 5.1 Study area maps of the Inyaka Dam (A) and the Crocodile River (B). Both frames illustrate the position of the sampling points SS-25 and C-40 along with associated tributaries and treatment works around the sites.

5.4.2. Water sample collection and handling

Briefly, water samples were collected in the summer (December to February) and winter (June to August) of 2021 from the Crocodile River and Inyaka Dam. These samples were collected from nationally established sites designated SS-25 and C-40 in the Department of Water and Sanitation monitoring programme. Sealed and sterile 500 mL plastic bottles were used to collect water samples directly from the surface of the water. A total of 8 samples were used for 16S metagenomics due to budget considerations.

5.4.3. Experimental methods

5.4.3.1. Physicochemical and bacteriological analysis

Twelve water quality indicators were assessed from water samples used for the study. These included pH, turbidity, electrical conductivity (EC), temperature, Manganese, Phosphate, Nitrate-Nitrogen, Iron, *E. coli*, total coliforms, Bifidobacteria, and Faecal streptococci.

5.4.3.2. DNA extraction and 16S rRNA sequencing

DNA was extracted from 100mL of surface water samples and filtered through a 0.45 nitrate cellulose membrane filter (Whatman GE Healthcare Life Sciences, Buckinghamshire, UK). Using a Macherey-Nagel microbial DNA extraction kit (Dueren, Germany), cells were lysed in a 1mg/mL lysozyme solution containing 0.30-0.50mm glass beads (Sigma-Aldrich Corporation, St. Louis, MO, USA) for disruption of the bacterial cell wall. The lysis solution was agitated in a vortex for 5 min. 1mg/mL of Proteinase K was added to the lysis solution. DNA was eluted into 1.5 mL nuclease-free tubes. The concentration of the DNA was determined using Nanodrop ND100 (NanoDrop Technologies, Wilmington, DE, USA). The extracted DNA was stored at -20 °C for further analysis.

Genomic DNA samples were commercially sequenced at Inqaba Biotechnical Industries Pty Ltd, South Africa, for a near full-length 16S rRNA on a PacBio single-molecule real-time (SMRT) sequencing platform (Pacific Biosciences Inc., Menlo Park, CA, USA). High-performance liquid chromatography (HPLC) purified and barcoded 16S rRNA-specific forward primer 12F

(5'-CTACACGACGCTCTTCCGATCTGTAGACGCTGAGRGTTYGATYMTGGCTCAG-3') and reverse primer 32R (5'-

AAGCAGTGGTATCAACGCAGAGCTCGTCAGATRGYTACCTTGTTACGACTT-3')

were used to generate SMRTbell libraries. The libraries were sequenced on a PacBio RS II system with SPRQ chemistry.

5.4.3.3. Statistical and bioinformatic analysis

Physicochemical characteristics of the two ecosystems were illustrated using comparative bar graphs to highlight seasonal variations. Microbial sequence data were processed using the Quantitative Insights into Microbial Ecology (QIIME) software suite (version 1.7) through the cloud-based Nephel Microbiome Analysis platform (Weber et al., 2018). Sequence clustering was performed at 100% identity to obtain representative sequences using PyNAST and UCLUST software (Caporaso et al., 2010; Chigwada et al., 2023). QIIME 2 cloud-based Krona visualisation was used to generate abundance charts at the phylum and class resolutions (<https://view.qiime2.org/>). Diversity analyses included both alpha and beta indices, such as the Chao1, ACE, Jackknife, Shannon-Weiner, Simpson, and Principal coordinate analysis (PCoA). All graphs were generated using R package (version 3.1.2), Microsoft Excel 365, and OriginPro 2024b (student version). The PCoA was conducted based on Jensen-Shannon divergence metrics using species-level taxonomic resolution and included unclassified operational taxonomic units (OTUs). Finally, an analysis of similarities (ANOSIM) was conducted to determine significant differences in microbial community structure between and within the sites.

5.5. Results and Discussion

5.5.1. Physicochemical properties of surface water in the Inyaka Dam and Crocodile River

This study evaluated specific water quality indicators in the Inyaka Dam and Crocodile River during summer and winter. Between these seasons, a total of 117 samples were collected to assess the variation in water quality indicators. Figs. 5.2 and 5.3 provide a summary of the collected water quality data, while [Appendix 5.1](#) provides details on the data means and standard deviations from the observed seasons.

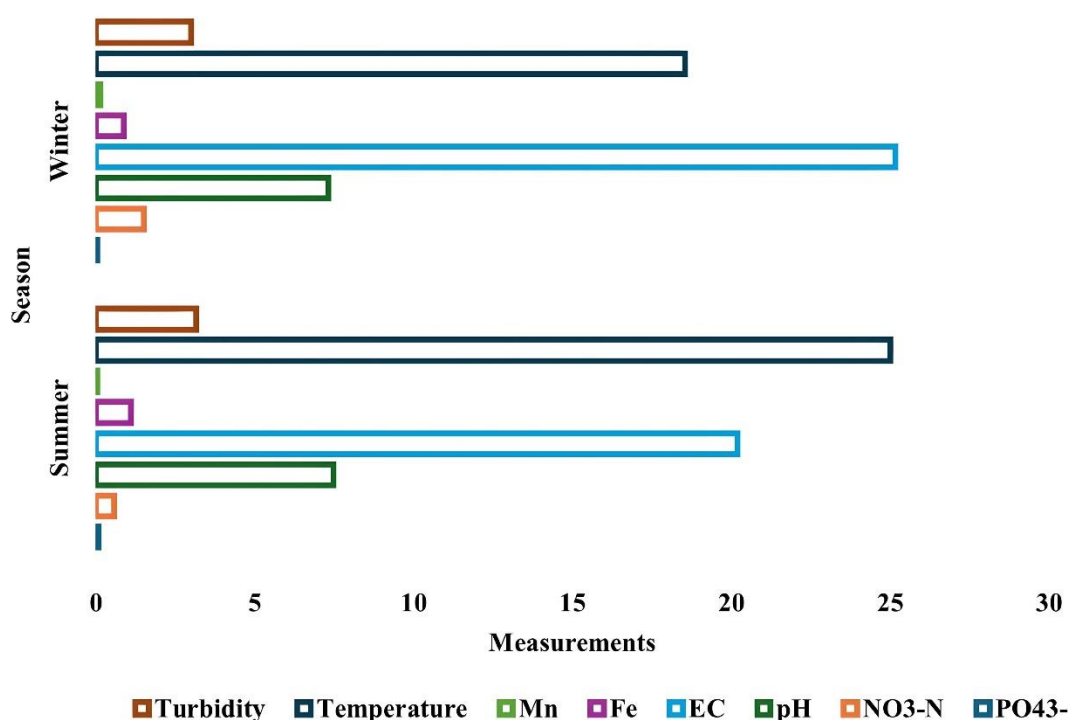


Fig. 5.2 Comparison of seasonal physicochemical parameters in the Inyaka Dam collected during winter and summer. The chart shows key water quality indicators: Electrical conductivity (EC), Iron (Fe), pH, Phosphate (PO_4^{3-}), Nitrate-Nitrogen ($\text{NO}_3\text{-N}$), Manganese (Mn), and Temperature plotted in seasonally segregated bars to enable distinction of seasonal variation.

The concentrations of $\text{NO}_3\text{-N}$, Mn, and EC showed considerable increases from summer (0.57 mg/L, 0.04 mg/L, 20.19 uS/cm to winter (1.51 mg/L, 0.18 mg/L, 25.16 uS/cm), suggesting a regression in water quality in the Inyaka Dam. Although this trend is not universal with all lentic systems in South Africa, recent studies have demonstrated that lake water quality tends to be worse in winter due to stratification and reduced precipitation (Mahdzir et al., 2024; Mogane et al., 2024). The increase in electrical conductivity from 20.19 uS/cm suggests the easing off stratification and decreased water flow, resulting in pollutant build-up and the water layers mixing, thus disturbing settled sediment (Fig. 5.2).

In South Africa, the water quality of lotic systems degrades more in the summer due to increased precipitation and run-off from farming activities degradation happens during the dries seasons for the lentic systems (Avenant et al., 2024; Shabalala et al., 2013). Experimental observations of the Crocodile River (Fig. 5.3) revealed significant parameter variations between the summer and winter seasons, with observed increases from winter to summer. The highest change was observed for $\text{NO}_3\text{-N}$ from 1.79 to 4.20 mg/L while EC and turbidity

changed from 90.47 to 204.75 mS/m and 4.52 to 52.32 NTU, respectively. On the other hand, Mn and Fe increased from near-compliant values of 0.18 and 0.39 mg/L to 0.65 and 2.14 mg/L. These observations are congruent with similar studies on lotic systems in countries such as Iraq (Moyel et al., 2023) and China (Qin et al., 2020).

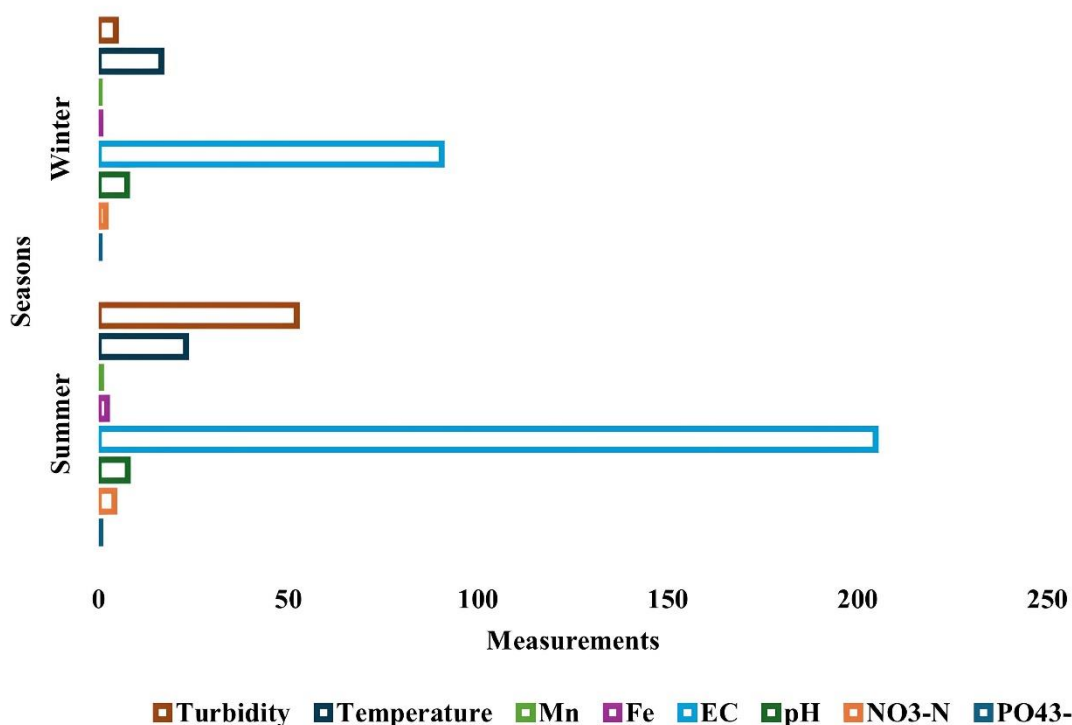


Fig. 5.3 Comparative seasonal analysis of physicochemical parameters in the Crocodile River The chart shows key water quality indicators: Electrical conductivity (EC), Iron (Fe), pH, Phosphate (PO_4^{3-}), Nitrate-Nitrogen ($\text{NO}_3\text{-N}$), Manganese (Mn), and Temperature plotted in seasonally segregated bars to enable distinction of temporal variation.

5.5.2. Dynamics of Bacterial Community Structure in Relation to Water Quality Variations

5.5.2.1. Phyla Abundance

Major bacteriological lineages in freshwater ecosystems have been reported to include *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Verrucomicrobia* (Jordaan & Bezuidenhout, 2013; Kim et al., 2021). The compositions and proportions of these bacterial lineages are influenced by the prevailing climate and the type of freshwater ecosystem (lake or river) (Farkas et al., 2022). In both the Crocodile River and the Inyaka Dam, *Proteobacteria*

were the most abundant taxon, with *Actinobacteria* the second most abundant. In the Crocodile River, *Proteobacteria* accounted for approximately 35-40% of the relative abundance, while the abundance estimations for *Actinobacteria* were approximately 25-30%. However, in the Inyaka Dam, *Actinobacteria* were more abundant (30-35%) compared to *Proteobacteria*, which accounted for 25-30%. Major site-specific differences were observed, with *Verrucomicrobia*, *Planctomycetes*, and *Chloroflexi* being more prominent in the Inyaka Dam, while *Cyanobacteria* were only found in the Inyaka Dam. Other minor phyla, such as *Acidobacteria* and *Firmicutes*, showed small variations between the sites. Figs. 5.4 and 5.5 compare the phyla composition of the Inyaka Dam and the Crocodile River.

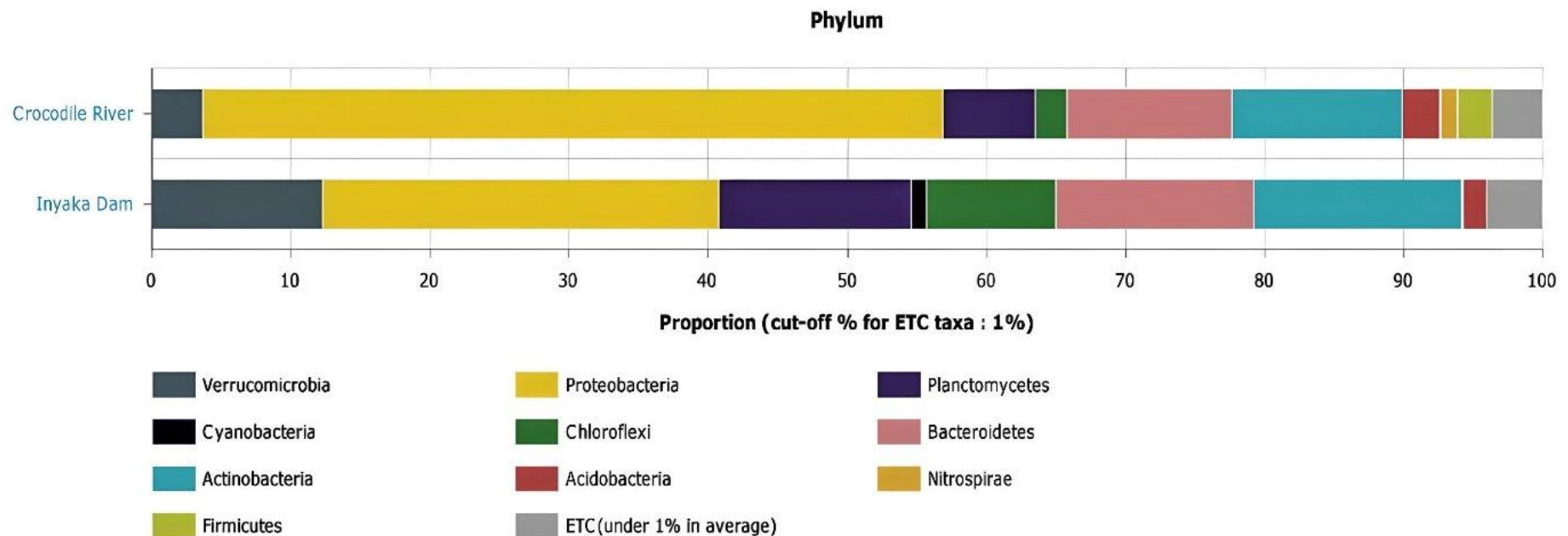


Fig. 5.4 Comparative relative abundance according to microbial phyla in the Crocodile River and Inyaka Dam, illustrating variation in the microbial community composition at the phylum level, with *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, and *Chloroflexi* being prominent in both aquatic habitats.

As seen in Fig. 5.5, the Crocodile River is consistently dominated by bacterial species from *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes* in the summer, with *Proteobacteria* showing extreme dominance (approx. 45,000 counts). Although *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes* continue to dominate in the winter months with variable proportions, there is significant variability in the relative abundance of *Proteobacteria* and *Bacteroidetes* with percentage decreases of 32.44% and 17.30%, respectively, suggesting seasonal sensitivity. This trend is supported by the fact that these two phyla thrive well in eutrophic or nutrient-rich waters (Chiang et al., 2018; Zheng et al., 2022). The mean concentrations of PO_4^{3-} , $\text{NO}_3\text{-N}$, Fe, and Mn (0.39, 4.20, 2.14, and 0.65 mg/L) were highest in summer as compared to winter (0.03, 1.51, 0.88, and 0.14 mg/L) in the Crocodile River ([Appendix 5.1](#)). In addition, these observations are congruent with the winter and summer physicochemical and microbial observations (Table 4.2).

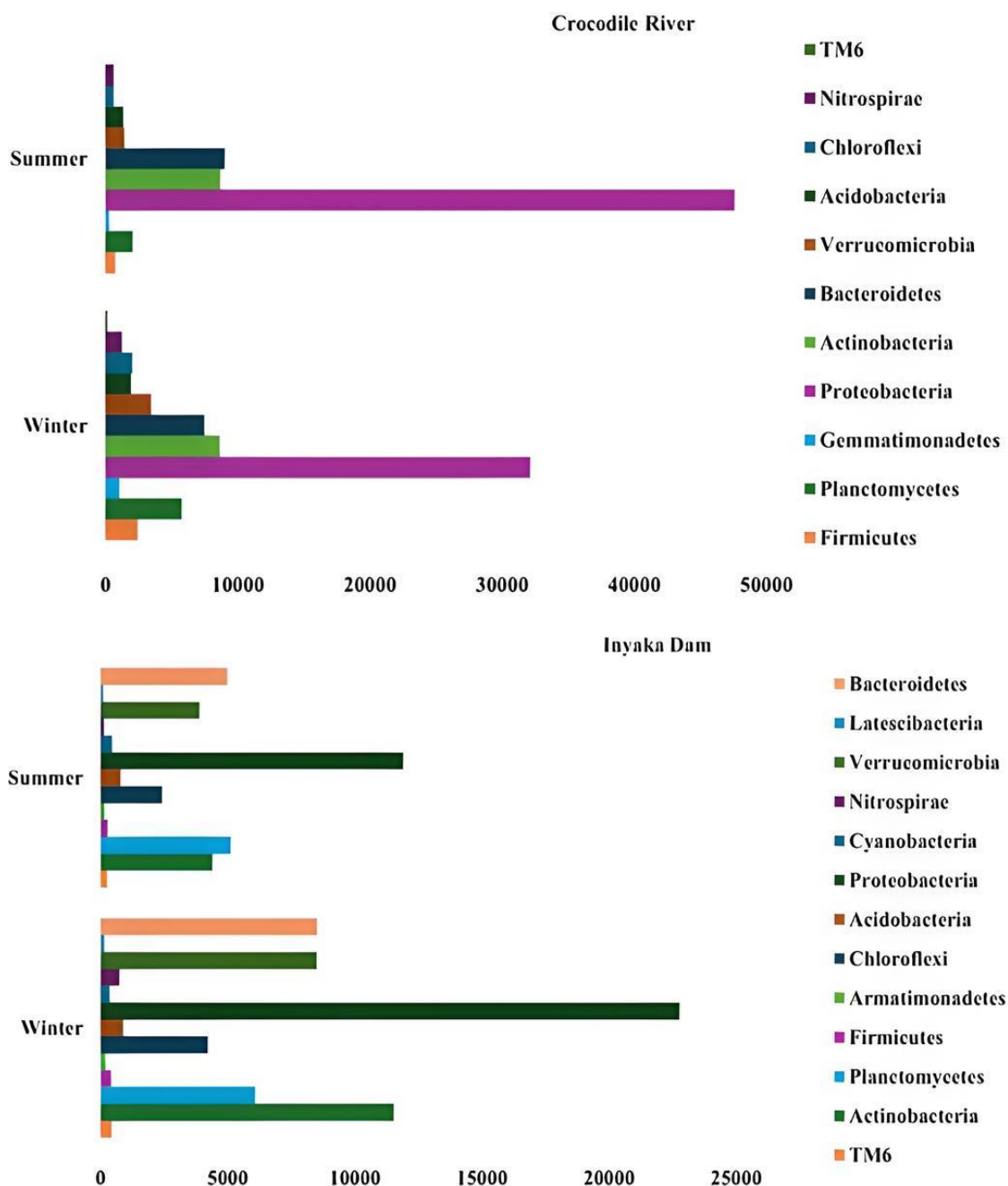


Fig. 5.5 Seasonal shifts in relative abundance of dominant taxa at a phylum level in the Crocodile River and Inyaka Dam. Stacked bar charts illustrate the relative abundance of major freshwater bacterial phyla from samples collected in summer and winter at both sites, while the seasonal comparisons illustrate temporal compositional shifts related to environmental variability.

On the other hand, *Actinobacteria* recorded the lowest seasonal abundance with just a 0.35% percentage change. In addition, the taxonomic classes with membership to the phylum *Proteobacteria* are known to be ubiquitous in aquatic environments and are reactive to increases in organic matter (Y. Yang et al., 2019). Other noticeable groups in the Crocodile River included *Firmicutes*, *Planctomycetes*, *Verrucomicrobia*, *Chloroflexi*, *Acidobacteria*, and

Nitrospirae, which demonstrated an opposite trend, being relatively abundant in winter compared to summer. The mean pH in the Crocodile River in winter was 7.59, well within the tolerance range of these phyla (neutral to alkaline). In addition, the mean temperature (16.43 °C) in the winter supported the growth of these phyla, as they fall within the mesophilic temperature tolerance range (Table 5.1). It has been demonstrated that the Inyaka Dam becomes more polluted in winter due to the easing off in dam stratification, which allows the water to mix and disturb settled sediments (Kim et al., 2025). This phenomenon results in elevated levels of physical and chemical parameters such as electrical conductivity, Manganese and Iron. Similarly, the Inyaka Dam showed a greater abundance of microbial communities in the winter compared to summer, with *Proteobacteria* and *Actinobacteria* being consistently dominant throughout the observed seasons. Other noticeable taxa exhibiting a similar trend included *Planctomycetes*, *Verrucomicrobiota*, and *Bacteroidetes* with moderate abundances (Fig. 5.5). Most of these phyla were present in significant abundance in both sites and seasons, albeit in varying abundance. Despite the overwhelming presence of *Proteobacteria* in both seasons, owing to their broad tolerance range, the presence of *Planctomycetes*, *Verrucomicrobiota*, and *Actinobacteria* highlights the pristine conditions of the Inyaka Dam. Members of these phyla thrive well in oligotrophic waters (Table 5.1) (Andrei et al., 2019; Chiang et al., 2018; Lipko & Belykh, 2021). In both seasons, the Inyaka Dam demonstrated water quality in pristine conditions except for Iron (0.04 in summer and 0.88 mg/L in winter).

Table 5.1: Characteristics of the observed major phyla in relation to physicochemical properties of water.

Phylum	Optimal pH	Temperature	Common Habitat	Nutrient Preference	Reference
<i>Proteobacteria</i>	Neutral to alkaline (6.5 – 8.5)	Mesophilic (20 – 30 °C)	Ubiquitous in freshwater	Broad (Oligo to Eutrophic waters with high organic content	(Chiang et al., 2018)
<i>Actinobacteria</i>	Slightly acidic (6.0 – 7.5)	Mesophilic (20 – 30 °C)	Oligotrophic/Nutrient- poor lakes and rivers	Oligotrophic waters with dissolved organic matter	(Lipko & Belykh, 2021)
<i>Bacteroidetes</i> (<i>Bacteroidota</i>)	Neutral to alkaline (7.0 – 8.5)	Mesophilic (20 – 30 °C)	Eutrophic environments and algal blooms	Eutrophic waters with high organic matter content	(Zheng et al., 2022)
<i>Chloroflexi</i>	Neutral to slightly alkaline (7.0 – 8.0)	Thermophilic (20 – 35 °C)	Low light zones of stratified freshwater bodies, sediments	High organic matter and low oxygen zones	(Mehrshad et al., 2018)
<i>Verrucomicrobiota</i>	Neutral to alkaline (7.0 – 8.5)	Mesophilic (20 – 30 °C)	Free living in freshwater columns	Oligotrophic to mesotrophic waters	(Chiang et al., 2018)
<i>Nitrospirae</i> (<i>Nitrospirota</i>)	Neutral to slightly acidic (6.5 – 7.5)	Mesophilic (20 – 30 °C)	Sediments and hypolimnetic zones of lakes	Ammonia and nitrite-saturated waters	(Ghimire-Kafle et al., 2023)
<i>Planctomycetes</i>	Neutral to slightly acidic (6.0 – 7.5)	Mesophilic to psychrotolerant (5 – 25 °C)	Particle-associated niches in lakes and rivers	Oligotrophic waters	(Andrei et al., 2019)

5.5.2.2. Microbial seasonal class distribution and abundance

The phylum *Proteobacteria* dominates the Crocodile River and Inyaka Dam irrespective of the prevailing season (Fig. 5.5), accounting for almost half of the bacterial relative abundance. However, within the phylum *Proteobacteria*, the classes *Gammaproteobacteria* and *Alphaproteobacteria* are consistently dominant with varying abundance throughout the observed seasons. In winter, *Gammaproteobacteria* accounted for 32% of total bacteria classes while *Alphaproteobacteria* accounted for 19% in the Crocodile River (Fig. 5.6). This was lower in comparison to summer, where *Gammaproteobacteria* accounted for a whole 59% compared to only 19% for *Alphaproteobacteria* in the Crocodile River. The increase in microbial abundance in the Crocodile River coincides with the increase in pollution due to increased nutrients, temperature, and rainfall (Fig. 5.4). This increase in the abundance of *Gammaproteobacteria* is significant in that this class includes some common bacteria of public health importance, such as those of the genera *Escherichia*, *Vibrio*, *Shigella*, *Salmonella*, *Pseudomonas*, etc (Rizzatti et al., 2017). In addition, this demonstrates that these bacteria enjoy high-nutrient environments with increased water temperatures (Ji et al., 2024). On the other hand, the class, *Alphaproteobacteria*, seems to be following the opposite trend as the season changes. It was observed that the abundance of *Alphaproteobacteria* decreased from 17% to 6% from winter to summer, suggesting that the physicochemical properties of water are not ideal for its proliferation. This is in congruence with the nature of members of the *Alphaproteobacteria* as they are known to be phototrophic and prefer low-nutrient environments (Imhoff, 2006). Several other studies have reported that the relative abundance of *Alphaproteobacteria* is usually not greater than 5.5% within a community (Lew et al., 2018) and that the low availability of organic nutrients may favour some members of the *Alphaproteobacteria* (Newton et al., 2011).

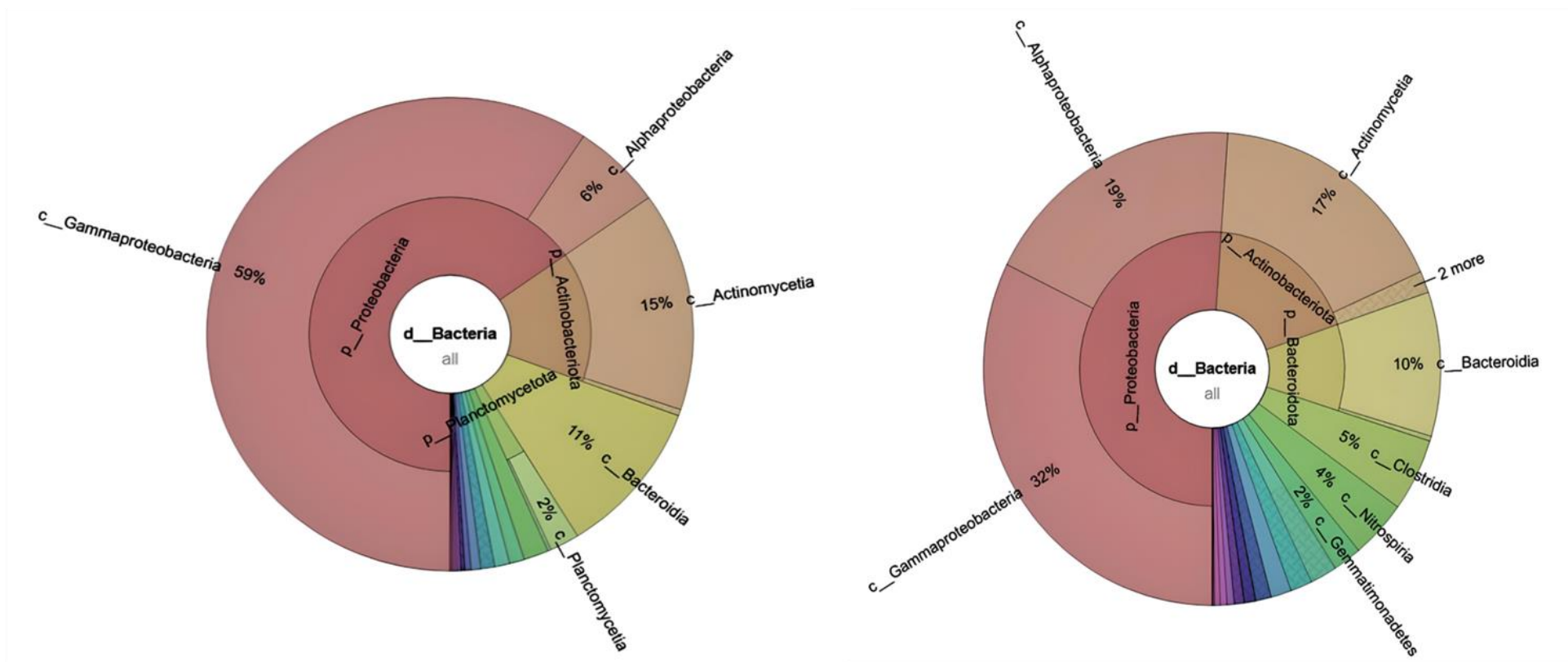


Fig. 5.6 Sunburst charts illustrating the relative abundance of bacterial communities in the Crocodile River during winter (left) and summer (right). Both charts display a hierarchical classification from domain to phylum and class taxa. *Proteobacteria* is the dominant phylum across all seasons, with *Gammaproteobacteria* being the most abundant in both winter (32%) and summer (59%).

In Fig. 5.5, the present study has demonstrated the ubiquity of the phylum *Proteobacteria* in both the Crocodile River and Inyaka Dam with less effect from seasonal variations. This study has also demonstrated that the classes *Gammaproteobacteria* and *Alphaproteobacteria* are dominant within the *Proteobacteria* phylum in all observed seasons. Although there was some observed variation of water quality in the Inyaka Dam between the winter and summer seasons, water quality was still within permissible limits except for Fe which recorded means of 0.88 and 1.09 mg/L for winter and summer respectively (Fig. 5.2). Previous studies have established that an increase in iron causes brownification in water systems and affects structure and function of aquatic ecosystems (Kritzberg et al., 2020; Xiao & Riise, 2021) and may negatively affect species abundance (Heikkinen et al., 2022). As such, the negligible variation in the physicochemical characteristics of water in the Inyaka Dam may be accountable to the low abundance of bacteria (Fig. 5.7) that has been observed to be ubiquitous in ecosystems with variable seasonal water quality. Based on the water quality data presented in Fig. 5.2, the Inyaka Dam is oligotrophic, thus limiting mass proliferation of *Gammaproteobacteria*, contrary to the observations in the Crocodile River.

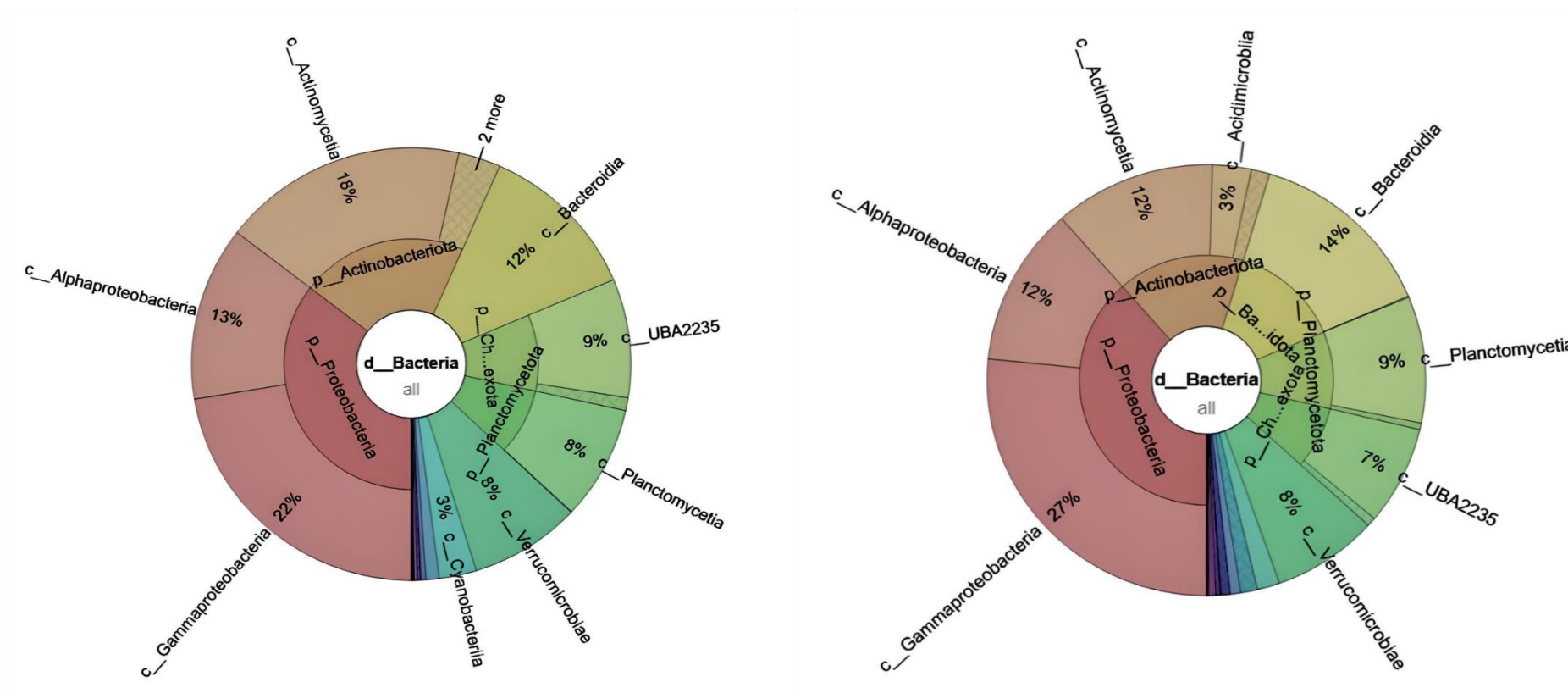


Fig. 5.7 Sunburst diagrams illustrating seasonal variations in the relative abundance of bacterial class compositions in the Inyaka Dam during winter (left) and summer (right). Both charts represent a classification depth of up to class level, with the percentages denoting the relative abundance of each class and highlighting observable bacterial community shifts between seasons.

5.5.2.3. Microbial diversity

i. Alpha diversity with ACE, Chao1, and Jackknife indices

To evaluate microbial richness within the two sites, alpha diversity was assessed using three non-parametric richness indices: Abundance-based Coverage Estimator (ACE), Jackknife, and Chao1. The boxplots in Fig. 5.8 display the distribution of richness values for microbial communities from the Crocodile River and Inyaka Dam. The data demonstrate substantially higher species richness in the Crocodile River compared to the Inyaka Dam across all three indices. The ACE reported a median of 2950 operational taxonomic units (OTUs) in the Crocodile River compared to 1800 in the Inyaka Dam. Similarly, the Chao1 index indicated a median richness of approximately 2850 OTUs versus 1705 OTUs in the Inyaka Dam, while the Jackknife estimator yielded 3000 and 1950 OTUs for the Crocodile River and Inyaka Dam, respectively. These differences demonstrate a consistent increase in OTUs in the Crocodile River, further suggesting a more stable and homogenous richness across the sampled seasons. The data collectively suggest that the Crocodile River host a richer microbial community, due to distinct environmental factors influencing microbial proliferation. We have observed slightly elevated levels of physicochemical indicators of water quality, such as PO_4^{3-} , Fe, EC, and Mn in the Crocodile River, suggesting that the Crocodile River is in a mesotrophic state, thus able to support above-normal microbial proliferation. Similar studies have also reported that mesotrophic aquatic ecosystems tend to be like moderately eutrophic ecosystems in terms of microbial community richness (Geng et al., 2022; Wang et al., 2022). The Inyaka Dam, on the contrary, presents similar characteristics of an oligotrophic state lake (Fig. 5.2) with all tested water quality indicators within permissible limits. Previous studies have shown that aquatic ecosystems with low nutrients, such as nitrates and phosphates, tend to have low microbial abundance and richness (Logue et al., 2012; Xie et al., 2024). This is confirmed in Fig. 5.8 with all indices consistently showing low levels of OTUs in the Inyaka Dam compared to the Crocodile River.

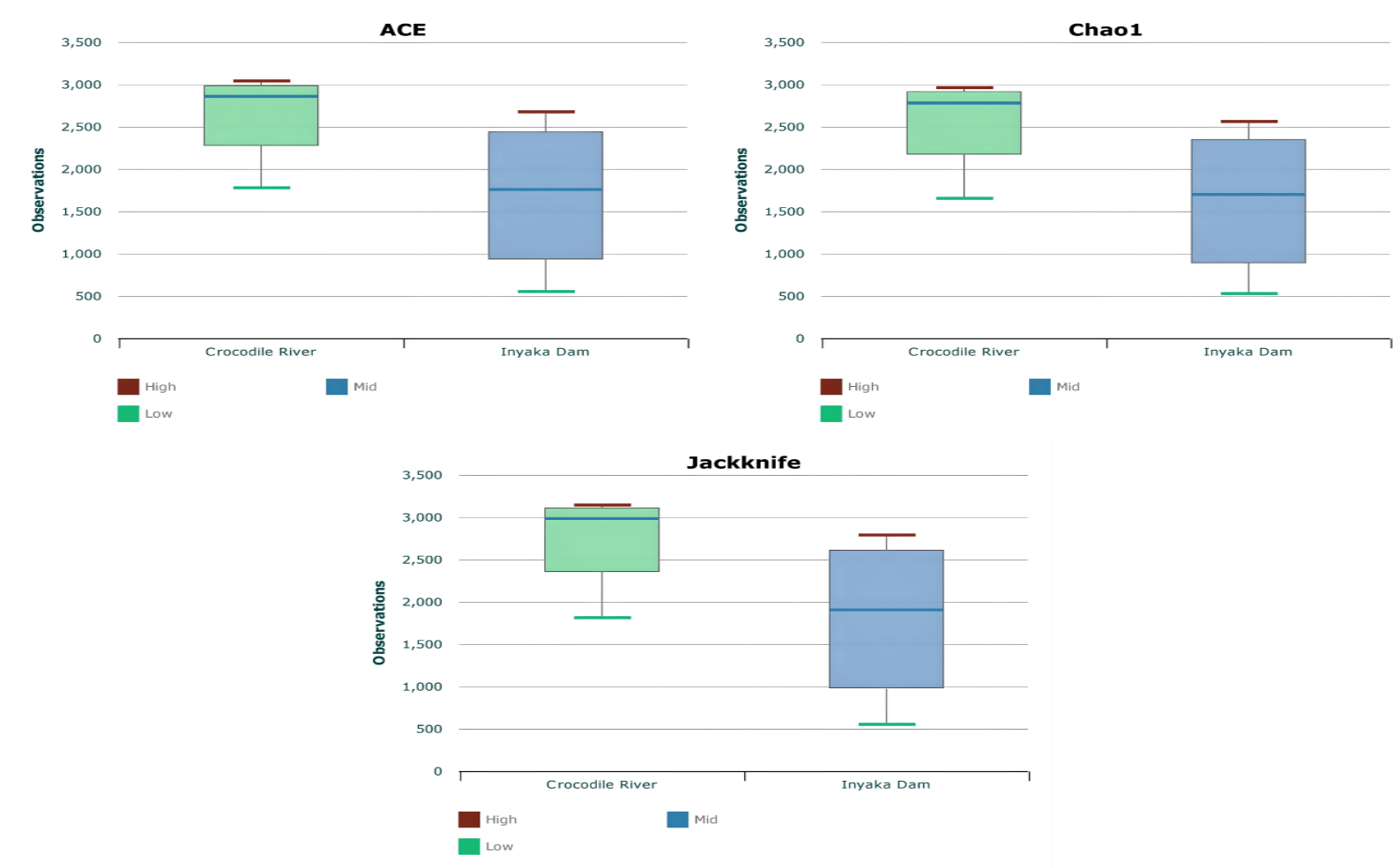


Fig. 5.8 Comparison of alpha diversity indices ACE, Chao1 and Jackknife across the Crocodile River and Inyaka Dam. The box plots illustrate the distribution of the observed values for the richness indices in the aquatic ecosystems. Each richness index has abundance categories: High (red), Mid (blue), and Low (green).

ii. *Alpha diversity with Shannon and Simpson Indices (Evenness)*

Alpha diversity was further evaluated using the Shannon and Simpson diversity indices to determine the species evenness in the obtained microbial communities within the Crocodile River and Inyaka Dam (Fig. 5.9). In addition to elucidating the number of present species, these indices also provide an idea of the distribution patterns of individual microorganisms (Cassol et al., 2025). The Shannon index values were consistently higher in samples from the Crocodile River compared to those from the Inyaka Dam, indicating a more diverse and evenly distributed microbial community. The Shannon values in the Crocodile River ranged between 5.0 to 6.0, with a mean value of 6.0, while the Inyaka Dam exhibited a lower median of about 5.3 and a range of between 4.1 to 6.0. By comparison, this suggests that the Crocodile River has greater diversity and evenness, thus reflecting a stable habitat compared to the Inyaka Dam. The Simpson index values in the Crocodile River ranged between 0.0 and 0.05, while the Inyaka Dam presented higher values, ranging between 0.0 and 0.05. This data further supports the Shannon index observations, although it emphasises dominance and the probability of encountering the same species twice in the Crocodile River. This is supported by the observed and consistent dominance of the phylum *Proteobacteria* across all seasons (Fig. 5.5). The broader range in the Inyaka Dam shows reduced alpha diversity, which could be indicative of the oligotrophic state, thus affecting microbial balance (Hou et al., 2025).

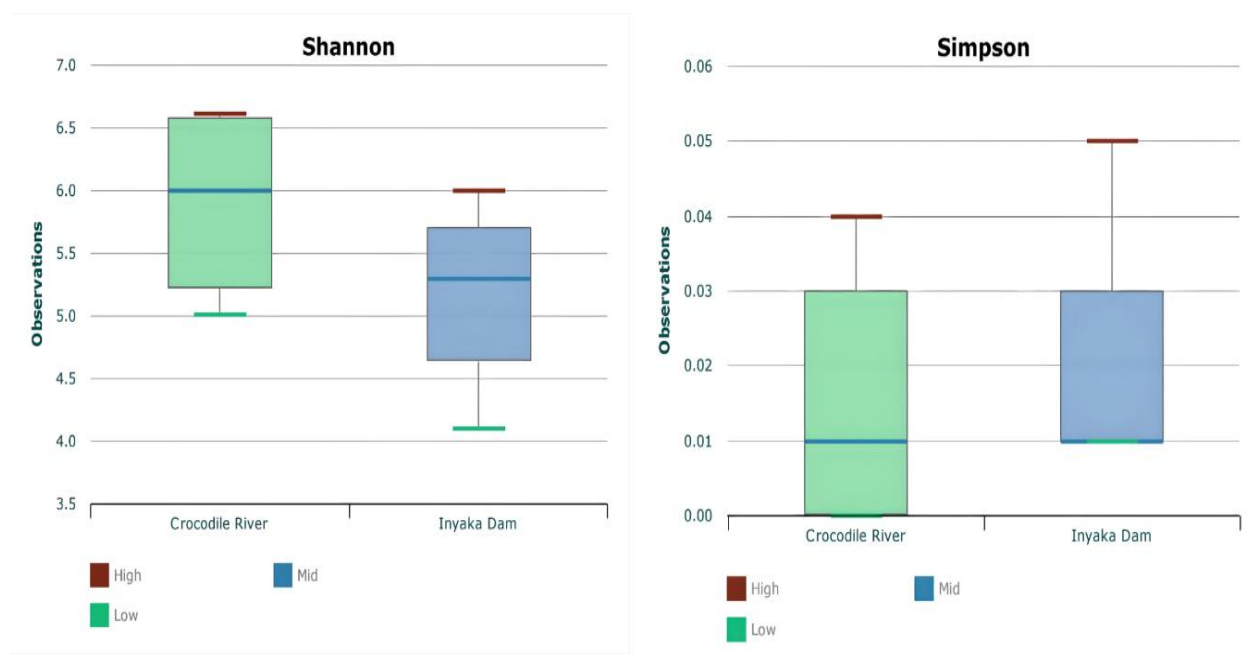


Fig. 5.9 Comparison of alpha diversity indices between the Crocodile River and Inyaka Dam. The box plots represent the Shannon diversity index (left) and the Simpson diversity index (right) categorised by bacterial community abundance groupings: High (brown), Mid (blue), and Low (green).

iii. *Beta diversity and bacterial community composition dissimilarity*

To investigate the dissimilarities in microbial community composition between the Crocodile River and Inyaka Dam, Principal Coordinate Analysis (PCoA) was performed based on the Jensen-Shannon divergence at the species level. This analysis included unclassified OTUs to ensure a comprehensive representation of microbial diversity across samples. As illustrated in Fig. 5.10, the PCoA reveals a clear spatial separation between microbial communities from the Inyaka Dam and the Crocodile River along the first principal coordinate (PC1), which accounts for 71.027% of the total variance. The second component (PC2) contributes an additional 15.229%, which captures within-group variation in the Crocodile River. This indicates distinct taxonomic compositions between the two aquatic systems. In addition, the clustering of samples within each group and the variable segregation patterns between the groups highlight significant differences in microbial community structure between the lotic and lentic systems. Analysis of Similarity (ANOSIM) was conducted to validate these observations and yielded a Global R of 0.749, $p = 0.003$, indicating a significant dissimilarity in microbial composition. This R-value also suggested that inter-group variation is greater than intra-group variation (Somerfield et al., 2021).

Lentic systems such as the Inyaka Dam are typically characterised by stable and stratified conditions with lower turbulence and gradual nutrient cycling capable of supporting a potentially distinct microbial community. As observed in Fig. 5.10, the tight clustering in the Inyaka Dam samples (blue) suggests consistent environmental conditions. Contrarywise, lotic systems such as the Crocodile River feature continuous flow and dynamic nutrient input to support a more heterogeneous microbial community (Chen et al., 2025). This supports the observed inter-site dissimilarity in the Crocodile River (green) and points to environmental variability driven by physicochemical parameters such as temperature, dissolved oxygen, nutrient concentrations, and flow dynamics (Bilbao et al., 2023).

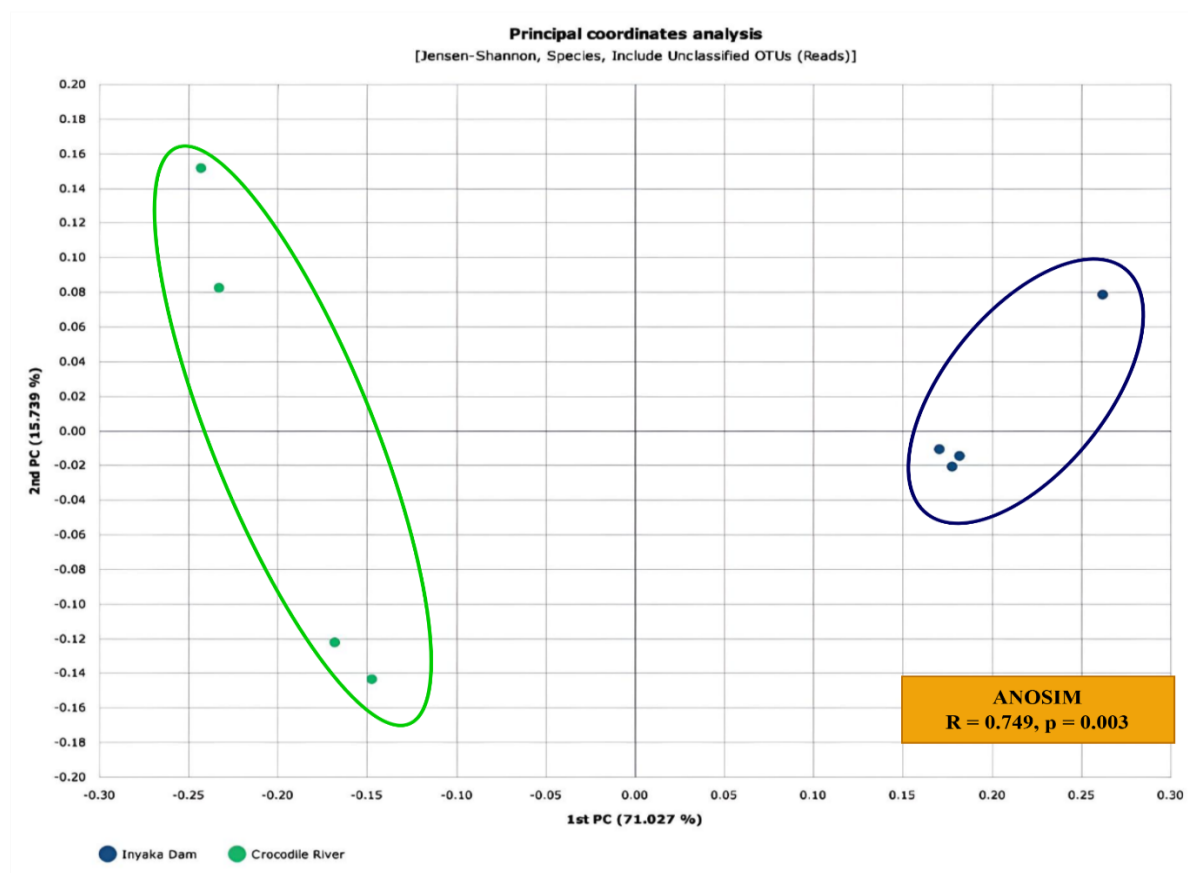


Fig 5.10 Beta diversity plot of Principal coordinate analysis (PCoA) based on Jensen-Shannon divergence, inclusive of unclassified OTUs. The PCoA plot reveals distinct clustering of microbial communities from the Crocodile River (green) and the Inyaka Dam (blue). The first and second principal coordinates explain a total variation of 86.20% with each contributing 71.027% and 15.739% respectively.

5.6. Conclusions

This study provides an assessment of the bacterial community structure and diversity across lotic and lentic freshwater ecosystems, the Crocodile River and Inyaka Dam, with a primary focus on the influence of physicochemical parameters. The integrative data analysis used in this study revealed clear and significant differences in bacterial compositions between the two sites. Key physicochemical drivers in the lotic system, such as flow rate, nutrient concentrations of PO_4^{3-} , $\text{NO}_3\text{-N}$, Mn, Fe, and EC, were influencers of microbial distribution and diversity, irrespective of season, favouring a more diverse and spatially variable microbial assemblage. The lentic system supported a more homogenous community structure across all seasons, largely due to water column stability and nutrient stratification. Despite the observed dissimilarities between the systems, the study revealed that the phylum *Proteobacteria* dominated both systems across all observed seasons. Moreover, *Gammaproteobacteria* were more abundant across all seasons in both ecosystems, suggesting their ubiquity and adaptation capabilities. These findings underscore the importance of hydrological context in shaping bacterial communities and highlight the need for ecosystem-specific monitoring frameworks. In addition, it enhances our comprehension of microbial biogeography in freshwater habitats and offers valuable insights for biodiversity conservation, water quality assessment, and ecological restoration in freshwater ecosystems. Furthermore, the findings of this study support the development of an ecosystem-specific microbial metrics system for water quality management. However, to advance progress on this subject, future work should be prioritised towards incorporating temporal sampling and functional metagenomics to unravel pathways through which environmental gradients mediate microbial communities.

5.7. Statements and Declarations

Competing interests: The authors declared no conflicting interests regarding publishing this paper.

Data availability: The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Funding: No funding was obtained for this study

Author contributions

Lazarus Katlego Mogane: was responsible for conceptualisation, data collection and curation, data analysis and visualization, writing of the first draft and final revision. **Esper J. Ncube:** was responsible for the conceptualisation, editing, and reviews. **Titus A.M. Msagati:** was responsible for the conceptualisation, editing and reviews. **Tracy M. Masebe:** was responsible for the conceptualisation, Methodology, editing, and reviews. All the authors read and approved the final draft of the manuscript.

5.8. Appendices

Appendix 5.1: Physicochemical properties of the Crocodile River and Inyaka Dam

	Crocodile River				Inyaka Dam			
	Summer		Winter		Summer		Winter	
Parameter	Mean	SD	Mean	SD	Mean	SD	Mean	SD
PO ₄ ³⁻	0.39	0.32	0.24	0.23	0.07	0.07	0.03	0.04
NO ₃ -N	4.20	5.92	1.79	0.74	0.57	0.62	1.51	2.43
pH	7.68	0.28	7.59	0.31	7.48	0.44	7.31	0.3
EC	204.75	58.71	90.47	123.03	20.19	19.27	25.16	18.19
Fe	2.14	2.37	0.39	0.18	1.09	0.6	0.88	0.29
Mn	0.65	0.72	0.18	0.07	0.04	0.01	0.14	0.18
Temperature	23.11	1.24	16.49	4.05	25.01	0.95	18.54	2.05
Turbidity	52.32	52.45	4.52	1.70	3.16	1.14	2.99	1.04

5.9. References

- Adla, K., Dejan, K., Neira, D., & Dragana, Š. (2022). Degradation of ecosystems and loss of ecosystem services. *One Health: Integrated Approach to 21st Century Challenges to Health*, 281–327. <https://doi.org/10.1016/B978-0-12-822794-7.00008-3>
- Andrei, A. Ș., Salcher, M. M., Mehrshad, M., Rychtecký, P., Znachor, P., & Ghai, R. (2019). Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. *ISME Journal*, 13(4), 1056–1071. <https://doi.org/10.1038/s41396-018-0332-5>
- Attiani, V., Smidt, H., & van der Wielen, P. W. J. J. (2025). Investigating spatial and temporal dynamics in microbial community composition of multiple full-scale slow sand filters in drinking water treatment. *Water Research*, 282, 123751. <https://doi.org/10.1016/J.WATRES.2025.123751>
- Avenant, M., Börnick, H., Graumnitz, S., Nyoka, N., Opeolu, B., Voua Otomo, P., Schubert, A., Schubert, S., Vos, T., & Jungmann, D. (2024). Investigating a surface water quality monitoring approach for QwaQwa, South Africa, by combining biological in vitro tests and chemical analyses. *Frontiers in Water*, 6. <https://doi.org/10.3389/frwa.2024.1408856>
- Bilbao, J., Pavloudi, C., Blanco-Rayón, E., Franco, J., Madariaga, I., & Seoane, S. (2023). Phytoplankton community composition in relation to environmental variability in the Urdaibai estuary (SE Bay of Biscay): Microscopy and eDNA metabarcoding. *Marine Environmental Research*, 191. <https://doi.org/10.1016/j.marenvres.2023.106175>
- Birnie-Gauvin, K., Lynch, A. J., Franklin, P. A., Reid, A. J., Landsman, S. J., Tickner, D., Dalton, J., Aarestrup, K., & Cooke, S. J. (2023). The RACE for freshwater biodiversity: Essential actions to create the social context for meaningful conservation. *Conservation Science and Practice*, 5(4). <https://doi.org/10.1111/csp2.12911>
- Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., & Knight, R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2), 266–267. <https://doi.org/10.1093/bioinformatics/btp636>
- Cassol, I., Ibañez, M., & Bustamante, J. P. (2025). Key features and guidelines for the application of microbial alpha diversity metrics. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-024-77864-y>
- Cepic, M., Bechtold, U., & Wilfing, H. (2022). Modelling human influences on biodiversity at a global scale—A human ecology perspective. *Ecological Modelling*, 465. <https://doi.org/10.1016/j.ecolmodel.2021.109854>
- Chen, Y. T., Reid, T., & Weisener, C. (2025). Microbial community and functional shifts across agricultural and urban landscapes within a Lake Erie watershed. *Journal of Environmental Management*, 373. <https://doi.org/10.1016/j.jenvman.2024.123683>
- Chiang, E., Schmidt, M. L., Berry, M. A., Biddanda, B. A., Burtner, A., Johengen, T. H., Palladino, D., & Denef, V. J. (2018). Verrucomicrobia are prevalent in north-temperate freshwater lakes

- and display class-level preferences between lake habitats. *PLoS ONE*, 13(3). <https://doi.org/10.1371/journal.pone.0195112>
- Chigwada, A. D., Ogola, H. J. O., & Tekere, M. (2023). Multivariate analysis of enriched landfill soil consortia provide insight on the community structural perturbation and functioning during low-density polyethylene degradation. *Microbiological Research*, 274. <https://doi.org/10.1016/j.micres.2023.127425>
- Chun, S. J., Cui, Y., Baek, S. H., Ahn, C. Y., & Oh, H. M. (2021). Seasonal succession of microbes in different size-fractions and their modular structures determined by both macro- and micro-environmental filtering in dynamic coastal waters. *Science of the Total Environment*, 784. <https://doi.org/10.1016/j.scitotenv.2021.147046>
- Dong, L., Lin, L., Xu, Q., Zhao, L., Qi, X., Yin, G., Pan, X., Xu, A., & Liu, Y. (2025). Changes in microbial community structure and responses to environmental factors in heavily polluted rivers remediated by bio-wax. *Journal of Hazardous Materials Advances*, 20. <https://doi.org/10.1016/j.hazadv.2025.100930>
- Dong, W., Cui, Z., Zhao, M., & Li, J. (2023). Seasonal and Spatial Variations of Bacterial Community Structure in the Bailang River Estuary. *Journal of Marine Science and Engineering*, 11(4). <https://doi.org/10.3390/jmse11040825>
- Eswaran, R., & Khandeparker, L. (2020). Influence of salinity stress on bacterial community composition and β -glucosidase activity in a tropical estuary: Elucidation through microcosm experiments. *Marine Environmental Research*, 159, 104997. <https://doi.org/10.1016/J.MARENVR.2020.104997>
- Farkas, M., Szoboszlay, S., Vörös, L., Lovász, Z., Méhes, N., Mátyás, K., Sebők, R., Kaszab, E., Háhn, J., Tóth, G., Harkai, P., Ferincz, Á., Táncsics, A., & Kriszt, B. (2022). Bacterial Community Dynamics along a River-Wetland-Lake System. *Water (Switzerland)*, 14(21). <https://doi.org/10.3390/w14213519>
- Feng, C., Lu, J., Liu, T., Shi, X., Zhao, S., Lv, C., Shi, Y., Zhang, Z., Jin, Y., Pang, J., & Sun, A. (2025). Microbial community dynamics in shallow-water grass-type lakes: Habitat succession of microbial ecological assembly and coexistence mechanisms. *Ecotoxicology and Environmental Safety*, 291, 117819. <https://doi.org/10.1016/J.ECOENV.2025.117819>
- Frenk, S., Dag, A., Yermiyahu, U., Zipori, I., Hadar, Y., & Minz, D. (2015). Seasonal effect and anthropogenic impact on the composition of the active bacterial community in Mediterranean orchard soil. *FEMS Microbiology Ecology*, 91(9). <https://doi.org/10.1093/femsec/fiv096>
- Gao, N., Wang, Y., Fang, T., Liang, Y., Cui, K., Chen, J., & Lu, W. (2023). Seasonal Succession of Bacterial Communities in the Land-Based Container with Recycling Water Aquaculture System. *Aquaculture Research*, 2023. <https://doi.org/10.1155/2023/4537061>
- Geng, M., Zhang, W., Hu, T., Wang, R., Wang, J., & Cheng, X. (2022). Eutrophication causes microbial community homogenization via modulating generalist species. *Water Research*, 210. <https://doi.org/10.1016/j.watres.2021.118003>

- Ghate, S. D., Shastry, R. P., Arun, A. B., & Rekha, P. D. (2021). Unraveling the bacterial community composition across aquatic sediments in the Southwestern coast of India by employing high-throughput 16S rRNA gene sequencing. *Regional Studies in Marine Science*, 46. <https://doi.org/10.1016/j.rsma.2021.101890>
- Ghimire-Kafle, S., Weaver, M. E., & Bollmann, A. (2023). Ecophysiological and Genomic Characterization of the Freshwater Complete Ammonia Oxidizer *Nitrospira* sp. Strain BO4. *Applied and Environmental Microbiology*, 89(2). <https://doi.org/10.1128/aem.01965-22>
- Gong, D., Guo, Z., Wei, W., Bi, J., Wang, Z., & Ji, X. (2022). Phytoplankton Community Structure and Its Relationship with Environmental Factors in Nanhai Lake. *Diversity*, 14(11). <https://doi.org/10.3390/d14110927>
- Guigard, L., Nazaret, F., Almario, J., Bertolla, F., Boubakri, H., Cantarel, A. A. M., Cournoyer, B., Favre-Bonté, S., Florio, A., Galia, W., Hazard, C., Henry, G., Belaroussi, A. H., Chong, S. K. T. F., Lavire, C., Lobreau, C., Luis, P., Maréchal, M., Meyer, T., ... Shade, A. (2025). The connections of climate change with microbial ecology and their consequences for ecosystem, human, and plant health. In *Journal of Applied Microbiology* (Vol. 136, Number 7). Oxford University Press. <https://doi.org/10.1093/jambio/lxaf168>.
- Heikkinen, K., Saari, M., Heino, J., Ronkanen, A. K., Kortelainen, P., Joensuu, S., Vilmi, A., Karjalainen, S. M., Hellsten, S., Visuri, M., & Marttila, H. (2022). Iron in boreal river catchments: Biogeochemical, ecological and management implications. In *Science of the Total Environment* (Vol. 805). Elsevier B.V. <https://doi.org/10.1016/j.scitotenv.2021.150256>
- Hitt, N. P., Bonneau, L. K., Jayachandran, K. V., & Marchetti, M. P. (2015). Network of Conservation Educators & Practitioners Freshwater Ecosystems and Biodiversity. In *in Conservation* (Vol. 5).
- Hou, X., Zhang, L., Zhao, Y., Li, J., Jiang, Z., Wang, S., Li, X., Wang, X., & Liu, X. (2025). Changes in microbial communities across the whole A2/O wastewater treatment process and their drivers—Reduced community diversity but increased proportion of certain pathogens. *Water Research*, 268. <https://doi.org/10.1016/j.watres.2024.122790>
- Imhoff, J. F. (2006). The phototrophic alpha-proteobacteria. *The Prokaryotes*, 41–64.
- Ji, L., Zhang, H., Wang, Z., Tian, Y., Tian, W., & Liu, Z. (2024). Temperature alters bacterial community structure in sediment of mountain stream. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-82497-2>
- Jiang, T., Sun, S., Chen, Y., Qian, Y., Guo, J., Dai, R., & An, D. (2021). Microbial diversity characteristics and the influence of environmental factors in a large drinking-water source. *Science of the Total Environment*, 769. <https://doi.org/10.1016/j.scitotenv.2020.144698>
- Jordaan, K., & Bezuidenhout, C. C. (2013). The impact of physico-chemical water quality parameters on bacterial diversity in the Vaal River, South Africa. *Water SA*, 39(3), 385–396. <https://doi.org/10.4314/wsa.v39i3.7>

- Karlicki, M., Bednarska, A., Hałakuc, P., Maciszewski, K., & Karnkowska, A. (2024). Spatio-temporal changes of small protist and free-living bacterial communities in a temperate dimictic lake: insights from metabarcoding and machine learning. *FEMS Microbiology Ecology*, 100(8). <https://doi.org/10.1093/femsec/fiae104>
- Kim, B. K., Kim, D. H., & Hwang, J. H. (2025). Impact of water residence time and stratification on water quality improvement of an artificial brackish waterway. *Journal of Hydrology: Regional Studies*, 57. <https://doi.org/10.1016/j.ejrh.2024.102091>
- Kim, S., Islam, M. R., Kang, I., & Cho, J. C. (2021). Cultivation of Dominant Freshwater Bacterioplankton Lineages Using a High-Throughput Dilution-to-Extinction Culturing Approach Over a 1-Year Period. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.700637>
- Kritzberg, E., & Bååth, E. (2022). Seasonal variation in temperature sensitivity of bacterial growth in a temperate soil and lake. *FEMS Microbiology Ecology*, 98(10). <https://doi.org/10.1093/femsec/fiac111>
- Kritzberg, E. S., Hasselquist, E. M., Škerlep, M., Löfgren, S., Olsson, O., Stadmark, J., Valinia, S., Hansson, L. A., & Laudon, H. (2020). Browning of freshwaters: Consequences to ecosystem services, underlying drivers, and potential mitigation measures. In *Ambio* (Vol. 49, Number 2, pp. 375–390). Springer. <https://doi.org/10.1007/s13280-019-01227-5>
- Ladau, J., & Elie-Fadrosh, E. A. (2019). Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. In *Trends in Microbiology* (Vol. 27, Number 8, pp. 662–669). Elsevier Ltd. <https://doi.org/10.1016/j.tim.2019.03.003>
- Leal Filho, W., Totin, E., Franke, J. A., Andrew, S. M., Abubakar, I. R., Azadi, H., Nunn, P. D., Ouweneel, B., Williams, P. A., & Simpson, N. P. (2022). Understanding responses to climate-related water scarcity in Africa. *Science of The Total Environment*, 806, 150420. <https://doi.org/10.1016/J.SCITOTENV.2021.150420>
- Lew, S., Glińska-Lewczuk, K., & Ziemińska-Buczyńska, A. (2018). Prokaryotic community composition affected by seasonal changes in physicochemical properties of water in peat bog lakes. *Water (Switzerland)*, 10(4). <https://doi.org/10.3390/w10040485>
- Li, C., Quan, Q., Gan, Y., Dong, J., Fang, J., Wang, L., & Liu, J. (2020). Effects of heavy metals on microbial communities in sediments and establishment of bioindicators based on microbial taxa and function for environmental monitoring and management. *Science of the Total Environment*, 749. <https://doi.org/10.1016/j.scitotenv.2020.141555>
- Li, Y., Wang, F., Yang, H., Li, H., & Hu, C. (2023). Balanced biogeographic and local environmental effects determine the patterns of microbial diversity in biocrusts at multi-scales. *Frontiers in Microbiology*, 14. <https://doi.org/10.3389/fmicb.2023.1284864>
- Lipko, I. A., & Belykh, O. I. (2021). Environmental Features of Freshwater Planktonic Actinobacteria. *Contemporary Problems of Ecology*, 14(2), 158–170. <https://doi.org/10.1134/S1995425521020074>

- Logue, J. B., Langenheder, S., Andersson, A. F., Bertilsson, S., Drakare, S., Lanzén, A., & Lindström, E. S. (2012). Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species-area relationships. *ISME Journal*, 6(6), 1127–1136. <https://doi.org/10.1038/ismej.2011.184>
- Mahdzir, M. M., Yahaya, S. H., Fauadi, M. H. F. M., & Yeow, T. T. (2024). A comprehensive review of water quality control on lentic ecosystem through the water quality, IoT and aeration models. In *Multidisciplinary Reviews* (Vol. 7, Number 5). Malque Publishing. <https://doi.org/10.31893/multirev.2024098>
- Marmen, S., Fadeev, E., Al Ashhab, A., Benet-Perelberg, A., Naor, A., Patil, H. J., Cytryn, E., Viner-Mozzini, Y., Sukenik, A., Lalzar, M., & Sher, D. (2021). Seasonal Dynamics Are the Major Driver of Microbial Diversity and Composition in Intensive Freshwater Aquaculture. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.679743>
- Mehrshad, M., Salcher, M. M., Okazaki, Y., Nakano, S. I., Šimek, K., Andrei, A. S., & Ghai, R. (2018). Hidden in plain sight - Highly abundant and diverse planktonic freshwater Chloroflexi. *06 Biological Sciences 0602 Ecology 05 Environmental Sciences 0502 Environmental Science and Management. Microbiome*, 6(1). <https://doi.org/10.1186/s40168-018-0563-8>
- Mogane, L. K., Ncube, E. J., Msagati, T. A. M., & Masebe, T. M. (2024). Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa. *Heliyon*, 10(19). <https://doi.org/10.1016/j.heliyon.2024.e38246>
- Moyel, M. S., Akbar, M. M., & Hussain, N. A. (2023). Evaluation of Spatio-Temporal Changes in Water Quality in the Middle Section of the Shatt Al-Arab River, Southern Iraq. *Journal of Ecological Engineering*, 24(4), 297–311. <https://doi.org/10.12911/22998993/161014>
- Newton, J. R., Jones, E. S., Alexander, E., McMahon, D. K., & Stefan, B. (2011). A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews*, 75(1), 14–49. <https://doi.org/10.1128/mmbr.00028-10>
- Ngisiange, N., Tarimo, B., Daudi, L., Mwangi, S., Malesa, F., George, R., Kyewalyanga, M. S., Gullström, M., Osore, M., Mwaluma, J., & Winder, M. (2024). Seasonal fish larvae abundance and composition in seagrass habitats of coastal East Africa. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-62012-3>
- Priya, A. K., Muruganandam, M., Sivarethinamohan, R., Sujatha, S., Gaddam, M. K. R., Priya, V., Gomathi, R., Gokulan, R., Gurugubelli, T. R., & Muniasamy, S. K. (2023). Impact of climate change and anthropogenic activities on aquatic ecosystem – A review. *Environmental Research*, 238, 117233. <https://doi.org/10.1016/j.envres.2023.117233>
- Qin, G., Liu, J., Xu, S., & Wang, T. (2020). Water quality assessment and pollution source apportionment in a highly regulated river of Northeast China. *Environmental Monitoring and Assessment*, 192(7). <https://doi.org/10.1007/s10661-020-08404-0>
- Ramasamy, K. P., Mahawar, L., Rajasabapathy, R., Rajeshwari, K., Miceli, C., & Pucciarelli, S. (2023). Comprehensive insights on environmental adaptation strategies in Antarctic bacteria

- and biotechnological applications of cold adapted molecules. In *Frontiers in Microbiology* (Vol. 14). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2023.1197797>
- Regina, A. L. A., Medeiros, J. D., Teixeira, F. M., Côrrea, R. P., Santos, F. A. M., Brantes, C. P. R., Pereira, I. A., Stapelfeldt, D. M. A., Diniz, C. G., & da Silva, V. L. (2021). A watershed impacted by anthropogenic activities: Microbial community alterations and reservoir of antimicrobial resistance genes. *Science of the Total Environment*, 793. <https://doi.org/10.1016/j.scitotenv.2021.148552>
- Rizzatti, G., Lopetuso, L. R., Gibiino, G., Binda, C., & Gasbarrini, A. (2017). Proteobacteria: A common factor in human diseases. In *BioMed Research International* (Vol. 2017). Hindawi Limited. <https://doi.org/10.1155/2017/9351507>
- Shabalala, A. N., Combrinck, L., & McCrindle, R. (2013). Effect of farming activities on seasonal variation of water quality of Bonsma Dam, KwaZulu-Natal. *South African Journal of Science*, 109(7–8). <https://doi.org/10.1590/sajs.2013/20120052>
- Shang, Y., Wu, X., Wang, X., Wei, Q., Ma, S., Sun, G., Zhang, Huanxin, Wang, L., Dou, H., & Zhang, Honghai. (2022). Factors affecting seasonal variation of microbial community structure in Hulun Lake, China. *Science of the Total Environment*, 805. <https://doi.org/10.1016/j.scitotenv.2021.150294>
- Shi, P., Wang, H., Feng, M., Cheng, H., Yang, Q., Yan, Y., Xu, J., & Zhang, M. (2022). The Coupling Response between Different Bacterial Metabolic Functions in Water and Sediment Improve the Ability to Mitigate Climate Change. *Water (Switzerland)*, 14(8). <https://doi.org/10.3390/w14081203>
- Smith, L. C., Orgiazzi, A., Eisenhauer, N., Cesarz, S., Lochner, A., Jones, A., Bastida, F., Patoine, G., Reitz, T., Buscot, F., Rillig, M. C., Heintz-Buschart, A., Lehmann, A., & Guerra, C. A. (2021). Large-scale drivers of relationships between soil microbial properties and organic carbon across Europe. *Global Ecology and Biogeography*, 30(10), 2070–2083. <https://doi.org/10.1111/geb.13371>
- Somerfield, P. J., Clarke, K. R., & Gorley, R. N. (2021). A generalised analysis of similarities (ANOSIM) statistic for designs with ordered factors. *Austral Ecology*, 46(6), 901–910. <https://doi.org/10.1111/aec.13043>
- Wang, Y., Guo, M., Li, X., Liu, G., Hua, Y., Zhao, J., Huguet, A., & Li, S. (2022). Shifts in microbial communities in shallow lakes depending on trophic states: Feasibility as an evaluation index for eutrophication. *Ecological Indicators*, 136. <https://doi.org/10.1016/j.ecolind.2022.108691>
- Wang, Z., Fuad, M. T. I., Liu, J., Lin, K., Liu, L., Gao, C., Wang, W., & Liu, X. (2024). Spatial Patterns of Microbial Communities in Intertidal Sediments of the Yellow River Estuary, China. *Microbial Ecology*, 87(1). <https://doi.org/10.1007/s00248-025-02494-4>
- Weber, N., Liou, D., Dommer, J., MacMenamin, P., Quiñones, M., Misner, I., Oler, A. J., Wan, J., Kim, L., Coakley McCarthy, M., Ezeji, S., Noble, K., & Hurt, D. E. (2018). Nephele: a cloud

- platform for simplified, standardized and reproducible microbiome data analysis. *Bioinformatics*, 34(8), 1411–1413. <https://doi.org/10.1093/bioinformatics/btx617>
- Wild, B., Gerrits, R., & Bonneville, S. (2022). The contribution of living organisms to rock weathering in the critical zone. In *npj Materials Degradation* (Vol. 6, Number 1). Nature Publishing Group. <https://doi.org/10.1038/s41529-022-00312-7>.
- Xiao, Y., & Riise, G. (2021). Coupling between increased lake color and iron in boreal lakes. *Science of the Total Environment*, 767. <https://doi.org/10.1016/j.scitotenv.2021.145104>.
- Xie, G., Zhang, Y., Gong, Y., Luo, W., & Tang, X. (2024). Extreme trophic tales: deciphering bacterial diversity and potential functions in oligotrophic and hypereutrophic lakes. *BMC Microbiology*, 24(1). <https://doi.org/10.1186/s12866-024-03488-x>
- Yang, Y., Gao, Y., Huang, X., Ni, P., Wu, Y., Deng, Y., & Zhan, A. (2019). Adaptive shifts of bacterioplankton communities in response to nitrogen enrichment in a highly polluted river. *Environmental Pollution*, 245, 290–299. <https://doi.org/10.1016/j.envpol.2018.11.002>
- Yu, X., Li, Y., Wu, Y., Gao, H., Liu, W., Liu, H., Gong, S., & Wu, H. (2024). Seasonal changes of prokaryotic microbial community structure in Zhangjiayan Reservoir and its response to environmental factors. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-55702-5>
- Zhang, M., Wu, Z., Sun, Q., Ding, Y., Ding, Z., & Sun, L. (2019). The spatial and seasonal variations of bacterial community structure and influencing factors in river sediments. *Journal of Environmental Management*, 248. <https://doi.org/10.1016/j.jenvman.2019.109293>
- Zhang, N., & Fryirs, K. (2024). Quantifying trajectories of geomorphic river recovery through analysis of assemblages of geomorphic units: Aiding detection to inform river management. *Geomorphology*, 455. <https://doi.org/10.1016/j.geomorph.2024.109202>
- Zhang, X., Qi, L., Li, W., Hu, B. X., & Dai, Z. (2021). Bacterial community variations with salinity in the saltwater-intruded estuarine aquifer. *Science of The Total Environment*, 755, 142423. <https://doi.org/10.1016/J.SCITOTENV.2020.142423>
- Zheng, F., Zhang, T., Yin, S., Qin, G., Chen, J., Zhang, J., Zhao, D., Leng, X., An, S., & Xia, L. (2022). Comparison and interpretation of freshwater bacterial structure and interactions with organic to nutrient imbalances in restored wetlands. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.946537>

6. Development of a water quality prediction tool using an Artificial Neural Networks-based Machine Learning Algorithm

6.1. Preamble

In the contemporary era, the sustainable management of water resources has become a critical concern, necessitating innovative approaches and technologies. Recognising the significance of predicting water quality for effective resource management, this chapter develops a cutting-edge water quality prediction tool. Leveraging the power of Artificial Neural Networks (ANNs) and machine learning, this chapter aims to construct a robust and accurate water quality prediction tool capable of forecasting various water quality parameters. Through the fusion of advanced computational techniques, data analytics, and environmental science, the proposed tool seeks to contribute valuable insights for proactive decision-making in the realm of water quality management. This preamble sets the stage for a comprehensive exploration into the development process of an ANN-based machine learning tool dedicated to predicting water quality, underscoring its potential implications for sustainable and informed water resource stewardship.

6.2. Abstract

Traditional water quality monitoring methods are often too expensive and slow for rural areas with limited funding. To address this, the study developed a water quality prediction tool using an Artificial Neural Network (ANN) algorithm. This tool analyses twelve water quality parameters, such as pH, electrical conductivity, turbidity, and various chemical and microbiological indicators, chosen based on regional risks and water quality goals.

The models were trained on 80% of the data and validated with the remaining 20%, going through 16 optimization iterations for each parameter. Evaluation was done using metrics like mean squared error (MSE) and mean absolute percentage error (MAPE), mean squared logarithmic error (MSLE), and mean absolute error (MAE). It was cross validated with weather data from the South African Weather Services.

The tool proved reliable for pH (MSLE=0.0), turbidity (MAPE=36.54%), temperature (MSLE=0.02), phosphate (MSLE=0.04), nitrate (MSLE=0.23), iron (MSLE=0.11), manganese (MSLE=0.06), but was less effective for predicting electrical conductivity and microbiological parameters due to limited data. Overall reliability with regards to performance of the prediction tool was 62%. Consequently, this modelling tool is currently best used for reference, and future work should aim to gather more comprehensive data to improve its predictive accuracy.

6.3. Background

In recent decades, the pollution of water resources has placed many water authorities and stakeholders in the spotlight regarding water quality monitoring processes. For the longest time, the quality status of water resources has been assessed against predetermined water quality standards, which is normally performed by collecting spatial and temporal water quality data to identify and evaluate water quality characteristics (Chounlamany et al., 2017). Several recent studies have challenged these traditional methods, highlighting their many limitations (Aldrees et al., 2022; Alnuwaiser et al., 2022; Sattari et al., 2016). For instance, traditional watershed quality monitoring methods are laborious and require expensive site inspections and laboratory investigations, which are time-consuming (Shah, et al., 2021a; Shah et al., 2021b), and have ambiguities due to human errors, which account for the inability to generalise the results (Sattari et al., 2016). Due to the limitations of traditional methods and the need for accurate and reliable water quality information, the past five decades have seen a rise in the development and usage of water quality index (WQI) models (Mogane et al., 2023; Sutadian et al., 2018; Uddin et al., 2021). A WQI is a dimensionless value assigned to an aggregated set of parameters for classifying water quality (Gazzaz et al., 2012; Sutadian et al., 2016; Uddin et al., 2021b). However, WQI models have also presented shortcomings that have reduced their reliability and acceptability (Gupta & Gupta, 2021; Malek et al., 2022; Sutadian et al., 2016; M. G. Uddin et al., 2021b). However, recent studies have used statistical approaches (Sutadian et al., 2017) and machine learning (Izhar Shah et al., 2021; Najafzadeh et al., 2019; Rustam et al., 2022; Uddin et al., 2022, 2023a, 2023c, 2023b) to address the shortcomings of WQI models. These hybrid WQI models are more accurate, robust, and demonstrate measurable performance (Uddin et al., 2022, 2023a).

Despite the advances made in improving the reliability and acceptability of WQI models, their usability requires a specific level of expertise in statistical and artificial intelligence (AI) tools. In addition, some steps involved in the development of the WQI model, such as the generation of sub-index values, employ various formulas for every sub-index value produced, thus requiring time and effort (Bui et al., 2020; Gazzaz et al., 2012; Madni et al., 2023). In a study conducted in the Malaysian Kinta River (Gazzaz et al., 2012), the authors expressed the need for a more direct and rapid method for computing and forecasting WQI values. They recommended the use of artificial neural networks (ANN) to reduce computation time and effort experienced when dealing with WQI models.

AI tools, such as machine learning algorithms (MLAs), have emerged as better options for assessing water quality monitoring data (Gazzaz et al., 2012; Rustam et al., 2022; Uddin et al., 2022, 2023). Deep machine learning tools have become popular, especially for modelling highly complicated processes and relationships. The most common type of deep MLAs is the ANN. The ANN is a digital representation of a biological neural network and consists of multiple layers of neurons, each connected to the neurons in the previous and subsequent layers. As such, ANNs are a subset of machine learning algorithms (MLAs) that are designed to learn and make predictions on complex and high-dimensional datasets, such as multi-parameter water quality data. The key characteristic of deep MLAs is their composition of multiple layers of interconnected neurons, allowing them to model complex relationships between input and output data (Lee et al., 2017).

There are several other documented advantages to ANNs, especially multilayer perceptron (MLP)-ANN. With a specific comparison to linear models, the MLP-ANN can model complex and non-linear relationships between input and output data because of the presence of multiple layers and non-linear activation functions such as the Rectified Linear Unit (ReLU). This allows them to capture intricate patterns in the data that linear models cannot capture. In addition, this method facilitates the automatic learning of features from the input data, thus eliminating the need for manual feature engineering and saving time and effort. The MLP-ANN are also versatile for a wide range of tasks because they can approximate any continuous function to a desired level of accuracy, given enough hidden neurons in the hidden layer. Furthermore, MLP-ANNs are adaptable to various scenarios and can be applied to tasks such as classification, regression, and time series. This can be achieved by adjusting the model architecture, activation functions, and training parameters. The MLP-ANNs can also be updated with new data incrementally, which is very useful for scenarios in which data arrive gradually or when adapting the model to changing patterns. All ANN models are data-driven; thus, the addition of data also improves the predictive power, accuracy, and reliability of the model (Awadallah et al., 2023; Najah Ahmed et al., 2019; Yang, 2019).

To address the challenges and limitations of WQI models and embrace the advantages of artificial intelligence, this study aimed to develop a model for predicting water quality independent of the development stages of WQI and solely based on AI tools such as the ANN. Through this study, we aim to develop a desktop-based tool that can be employed at any given time to forecast any water quality parameter, given that there is historical data. It should be noted, however, that the nature of ANN-based models is that they are data-dependent and thus

require a lot of data for good performance (Vu & Do, 2021). Fig. 6.1 illustrates the model development framework used in this study.

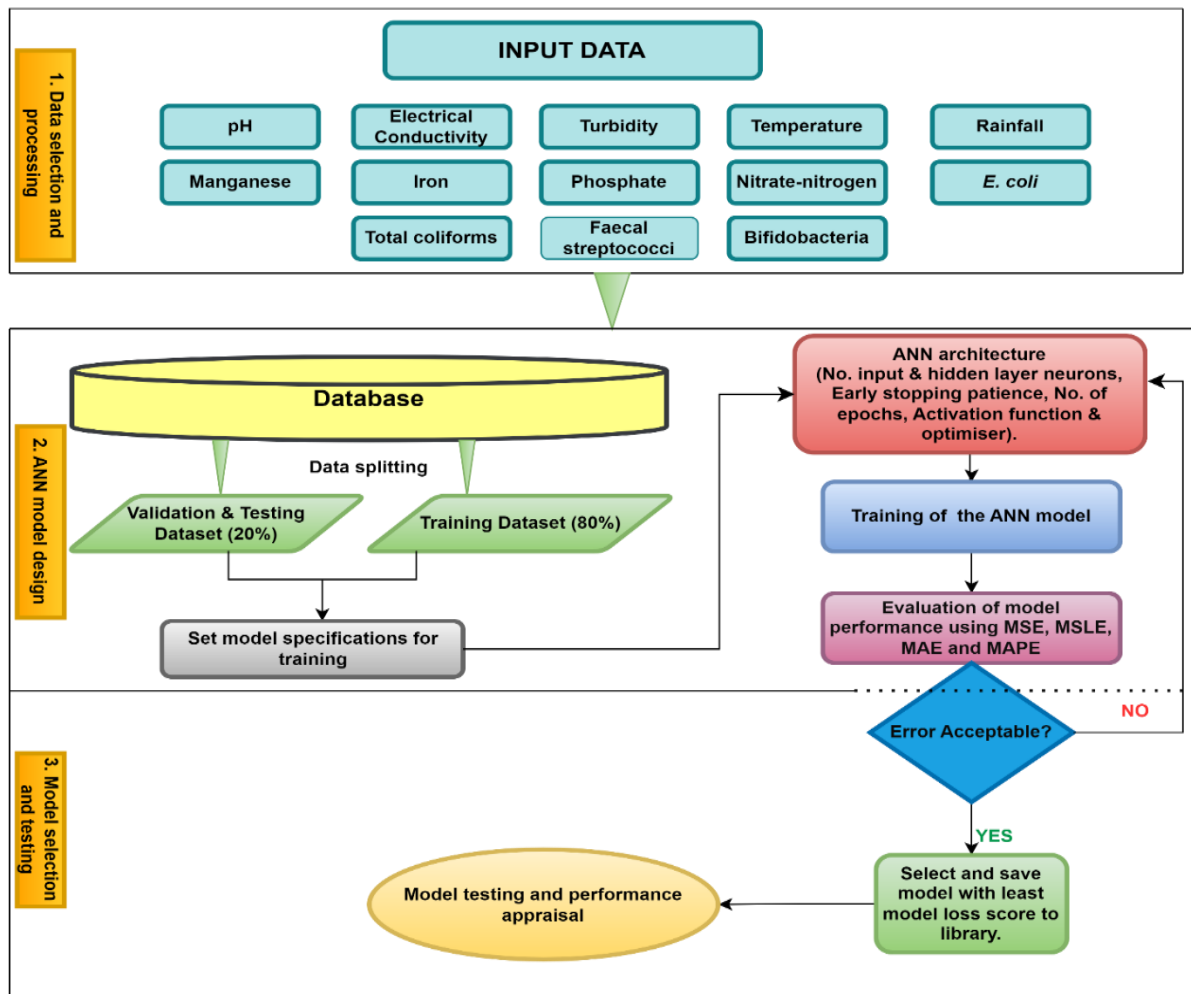


Fig. 6.1 Water quality prediction model development framework

6.4. Materials and methods

6.4.1. Study site

The Crocodile River catchment is located entirely within the Mpumalanga province in South Africa (Fig. 6.2). The river originates from areas that lie in the vicinity of Dullstroom, where it flows into the Kwenya Dam, passing through Mbombela. The catchment receives rainfall in the summer and has a subtropical climate. In addition, the Crocodile River bears with it some considerable ecological significance as it passes along the southern boundary of the Kruger National Park (KNP). There are notable activities in the surrounding areas of the catchment that may affect water quality. Closer to the sampling areas of the study are a dumping site (Fig.

S1) and a wastewater treatment plant that discharges its effluent into the river. Fig. 6.2 shows the sampling site with coordinates (25° 29' 18''S 31° 10' 23''E)

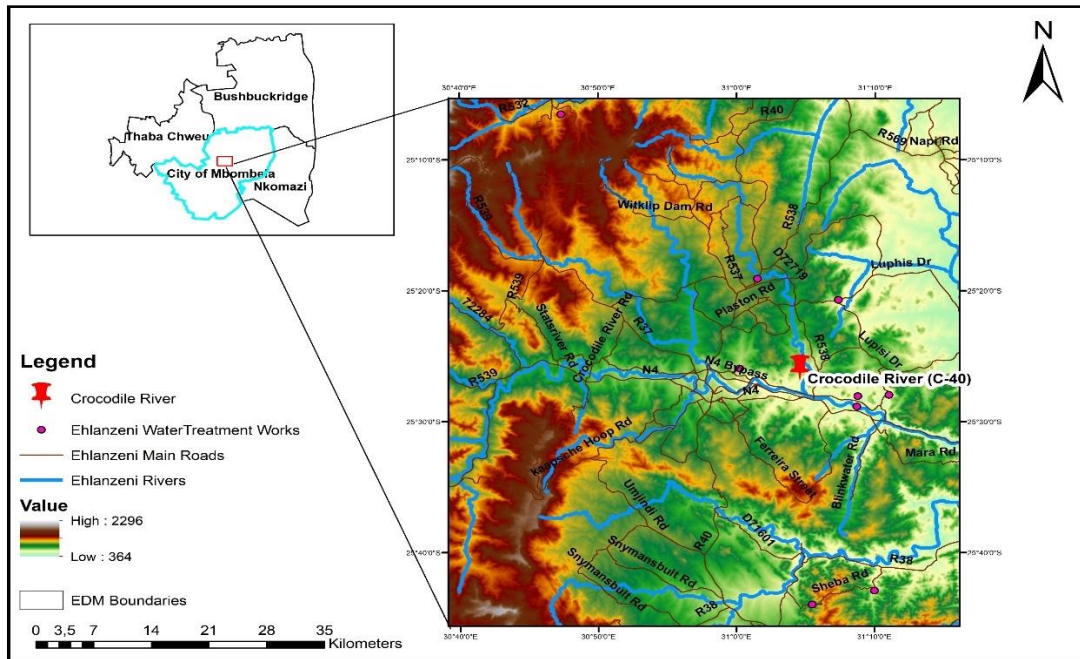


Fig. 6.2 Map of the study area in the Crocodile River situated within the northeastern parts of South Africa

6.4.2. Data generation from water quality parameters

A total of 13 water quality parameters (WQP) including pH, turbidity, electrical conductivity (EC), temperature, Manganese (Mn), Iron (Fe), Nitrate-nitrogen ($\text{NO}_3\text{-N}$), Phosphate (PO_4^{3-}), total coliforms (TC), Bifidobacteria, Faecal streptococci (FS), and *E. coli*, were selected. Additional data for certain WQP were obtained from the Inkomati-Usuthu catchment management agency (IUCMA), the custodian of watershed management in most parts of the Mpumalanga province in South Africa. All rainfall data were obtained with consent from the South African Weather Services (SAWS).

6.4.3. Data Standardisation

As in traditional WQI development, data from WQP must be transformed to a single scale prior to model training. Additionally, according to (Rahman, 2019, 2020), transforming all parameters to a common scale reduces model training errors. Furthermore, several recent studies have used the z-score data normalisation to standardise WQP data (Pras & Mamane, 2023; M. G. Uddin et al., 2022)). The z score is a measure of the number of standard deviations

between a data point and the mean. This study also employed the technique used by (M. G. Uddin et al., 2022) to standardise the WQP data using Equation 1:

$$z = \frac{x_i - \bar{x}}{\sigma} \quad (6.1)$$

Here, z = is the standardised value/z-score; x_i = is the i th parameter value; $\bar{x}$ = mean of all parameter values; σ = the standard deviation of the dataset.

6.4.4. Data Preparation, Training, Validation, and Testing

The final stage of data preparation involved partitioning the data into three sets: Training, Validation and Testing. The training data comprised the highest percentage (80%) of the data set to provide sufficient data for the algorithm to learn the data patterns and avoid model underfitting. Validation and testing were allocated to the remaining data partition (20%). The validation data set was used along with the training data set to evaluate the performance of the training process using cost/loss functions such as the mean squared error (MSE), the mean average percent error (MAPE), the mean squared logarithmic error (MSLE), and the mean absolute error (MAE). The testing dataset provides the final performance evaluation of the training and validation exercise to avoid biases in the model performance appraisal (Pras & Mamane, 2023).

6.4.5. ANN Model network architecture

6.4.5.1. The feedforward multilayer perceptron

The multilayer perceptron (MLP) forms a category of deep ANNs that has been recently used to classify complex relationships. In environmental water samples, ANNs have been used for nonlinear model calibration (Mohseni-Dargah et al., 2022). This study used the MLP-ANN feedforward algorithm, which uses a backpropagation learning algorithm. The network consists of an input layer, a hidden layer, and an output layer (Fig. 6.3).

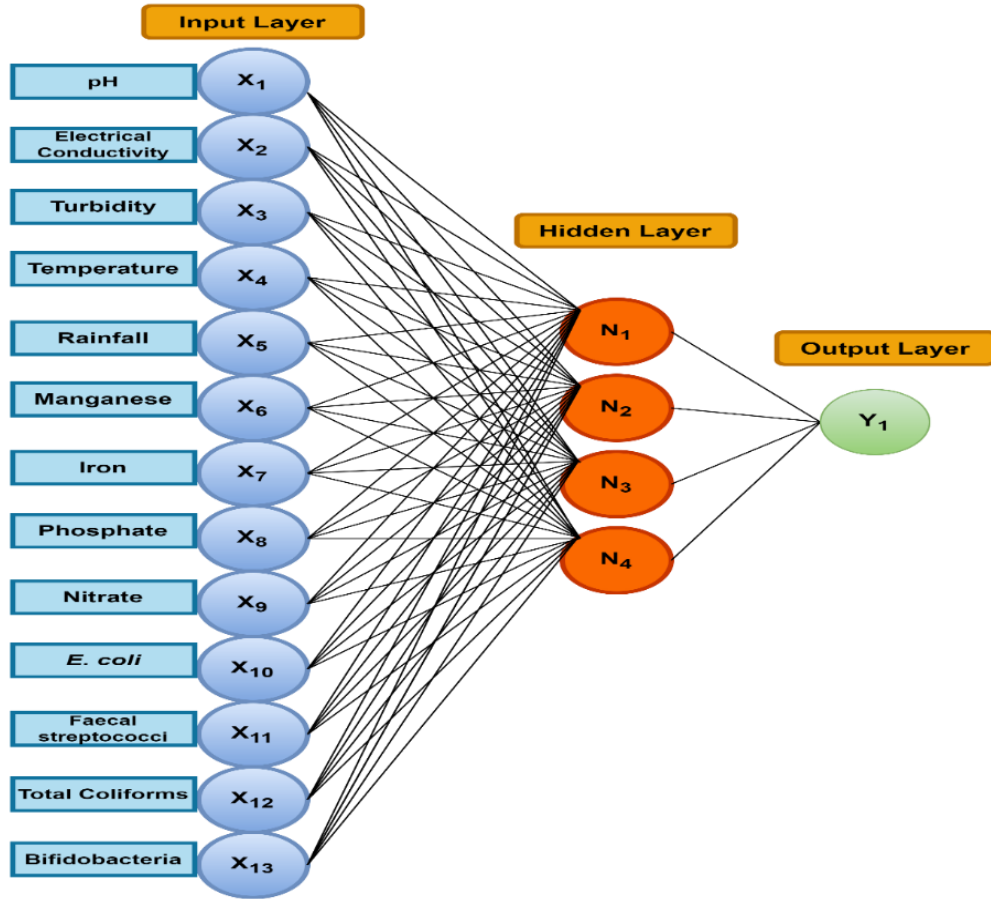


Fig. 6.3 One output fully connected MLP-ANN which can be employed by the developed model for selecting of one desired output parameter

The network can consist of one or more outputs. The addition of output parameters increases the neurons in the hidden layer. The hidden layer can be calculated manually using Equation 1:

$$\text{Hidden layer neurons} = \sqrt{\sum y_n * x_n} \quad (6.2)$$

Where Y_n = the nth number of output variables; X_n = nth number of input variables.

The network structure of the multiple output parameters is illustrated in Fig. 6.4. This study employed a trial-and-error method to select the most effective number of neurons in the hidden layer.

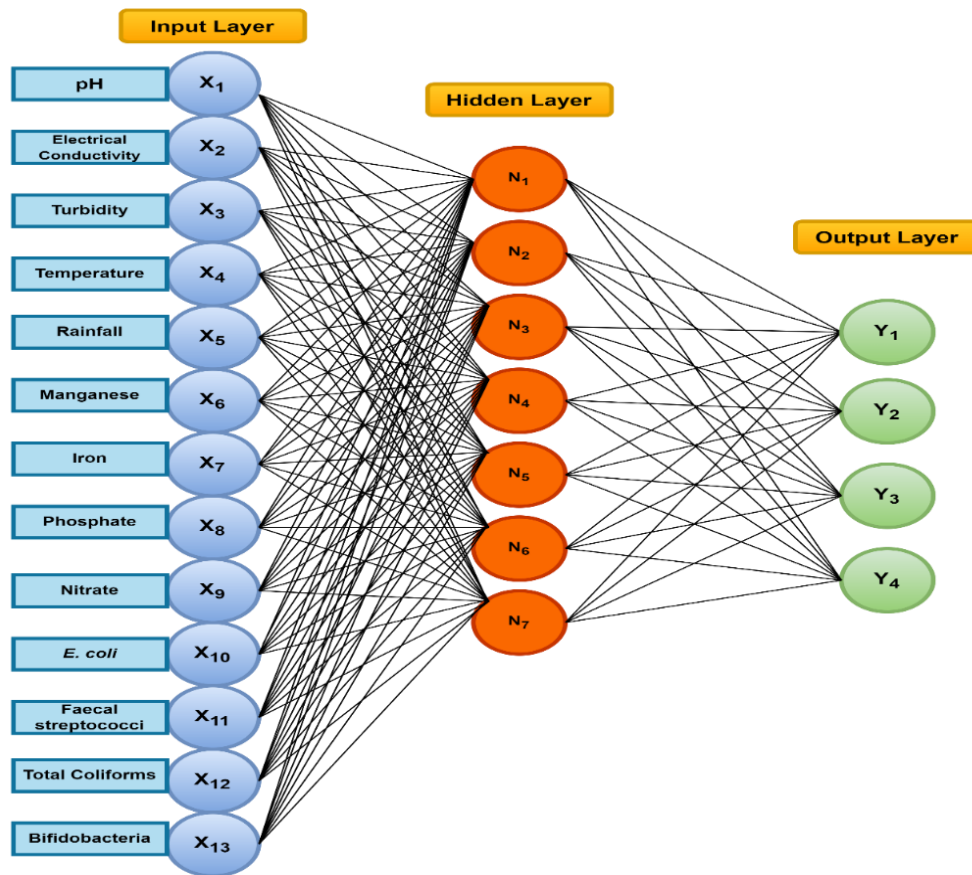


Fig. 6.4 Basic multiple outputs fully connected feedforward MLP-ANN with error backpropagation. The number of nodes in the hidden layer increased as the number of output nodes

6.4.5.2. Activation functions

Neural network models without the use of activation functions are like simple linear regression models with restricted performance. By design, neural networks are supposed to learn and compute a linear function but also perform tasks more complicated like modelling complicated data types. The activation functions introduce non-linearity to the network, which allows the neural network to learn complex relationships and patterns in the data (Dubey et al., 2022a; Szandała, 2020). This study used two activation functions to train the deep learning model.

i. Rectified linear unit (ReLU)

The Rectified Linear Unit activation function is a popular activation function in deep learning models. The rectified linear activation function returns the input directly if it is positive, and zero otherwise. The Rectified Linear Unit activation function, also known as RLU or ReLU, is a popular choice in deep learning models due to its simplicity and ability to address the

vanishing gradient problem. It is a nonlinear activation function that improves the efficiency of gradient-based optimisation algorithms, making it a suitable choice for deep neural networks. In deep machine learning, the Rectified Linear Unit activation function is widely used (Türk, 2021). It is a faster learning non-saturating activation function that provides better performance than traditional activation functions like tanh and sigmoid. However, although much preferred, the ReLU has its issues such as the dying ReLU problem where neurons can become inactive and stop learning (S. Sharma et al., 2020). The ReLU function is expressed as follows:

$$\text{ReLU}(x) = \max(0, x) \quad (6.3)$$

Where 0 and x represent the gradient input

ii. Sigmoid activation function

The sigmoid function squashes the input values into the range (0,1), making it suitable for binary classification. However, it suffers from the vanishing gradient problem, where the gradients become insignificant for extremely low or large input values. This phenomenon leads to slow convergence during model training. The sigmoid function is expressed as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (6.4)$$

Where: e represents Euler's number =2.718

6.4.5.3. Optimiser functions

In machine learning, the optimiser function is a critical component of the training process. The role of the optimiser function is to minimise the loss or cost functions by adjusting the weights and biases of the neural network. There are several types of optimiser functions commonly used in deep ML, including stochastic gradient descent, Adam, Adagrad, RMSprop, and Momentum. Each optimiser function has its way of updating the weights and biases, in consideration of factors such as the current gradient, learning rate, and momentum. Stochastic Gradient Descent (SGD) is a popular optimiser function in deep learning. SGD takes a random subset of data for each iteration and updates the weights and biases based on the estimated

gradient. Nadam optimiser was used throughout the study to standardise the training parameters (Diehl et al., 2016; Nagpal & Dubey, 2019).

The Nesterov Accelerated Adaptive Moment Estimation (Nadam)

The adaptive moment estimation (Adam) is an adaptive optimiser that adjusts the learning rate dynamically for each parameter based on past gradients. The Adam is a widely recommended optimisation algorithm in deep learning. The Nesterov-accelerated Adaptive Moment Estimation is an extension of the Adam algorithm with Nesterov's Accelerated Gradient (NAG). This allows for faster convergence and better performance on prediction problems (Khan et al., 2019).

6.4.5.4. Loss/cost functions used to appraise the performance of ANN models.

To evaluate the performance of a neural network in various tasks, such as classification or regression, loss functions are employed. These loss functions quantify the discrepancy between the predicted outputs of the neural network and the actual ground truth values. By measuring this discrepancy, the loss function can be used to measure how well the neural network is performing in its specific task. Loss functions are crucial for assessing the accuracy and effectiveness of neural networks by quantifying the difference between predicted outputs and actual values. They adjust the weights of the neural network during the training process, thereby allowing the network to minimise this discrepancy and improve its performance (Chen et al., 2022; Wang et al., 2017; Zhang, Lou, et al., 2021)). The following section provides detailed insights of the loss or cost functions used in this study.

i. Mean Squared error

The Mean Squared Error (MSE) is commonly used as a loss function in regression problems. The MSE measures the average squared difference between the predicted values and true values. Squaring the differences penalises larger errors more heavily than smaller errors, which makes it particularly sensitive to outliers. The MSE is as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6.5)$$

Where:

n is the number of samples in the dataset.
 y_i is the true target of the i -th sample; and
 $\hat{y}_i$ is the predicted target of the i -th sample.

ii. Mean Absolute Error

The Mean Absolute Error (MAE) is a commonly used loss function in regression problems. The MAE measures the average absolute difference between predicted and true values. Unlike the MSE, the MAE does not square the errors, which makes it less sensitive to outliers. It provides a more straightforward interpretation of the average magnitude of the prediction errors. The MAE is expressed as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6.6)$$

Where:

n is the number of samples in the dataset.

y_i is the true target of the i -th sample; and

$\hat{y}_i$ is the predicted target of the i -th sample.

iii. Mean Squared Logarithmic Error (MSLE)

The Mean Squared Logarithmic Error (MSLE) is a loss function commonly used in regression problems where the prediction targets have exponential growth. Using logarithms in MSLE makes it particularly useful when the target values exhibit exponential growth, as it penalises underestimates and overestimates on a logarithmic scale. The MSLE equation expressed is as follows:

$$\text{MSLE} = \frac{1}{n} \sum_{i=1}^n (\log(1 + y_i) - \log(1 + \hat{y}_i))^2 \quad (6.7)$$

Where:

n is the number of samples in the dataset.

y_i is the true target for the i -th sample; and

$\hat{y}_i$ is the predicted target for the i -th sample.

iv. Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is a commonly used loss function for regression problems, especially in contexts where it's important to measure the relative error percentage. MAPE expresses the prediction error as a percentage of the true value, making it easy to interpret. However, note that MAPE has a drawback: it becomes undefined when true values (y_i) are zero. In practise, you might need to handle or avoid cases where the true values are zero or close to zero to prevent issues with the MAPE calculation. The MAPE equation is as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (6.8)$$

Where:

n is the number of samples in the dataset.

y_i is the true target of the i -th sample; and

$\hat{y}_i$ is the predicted target of the i -th sample.

Several other cost functions are used to determine the performance of prediction models such as the Huber loss, Kullback-Leibler Divergence, Hinge loss, and binary cross-entropy loss. To select the best loss/cost function, this study compared the performance of MSE, MAPE, MSLE, and MAE on the study dataset.

6.5. Validity and Reliability

Validation is the most important step in model development because it accounts for the reliability and fit of the model (Tsiptsias et al., 2016). There are a variety of validation methods, such as k-fold cross-validation, train and test split, bootstrapping, Monte Carlo cross-validation and leave one out cross-validation. This study used cross-validation of the train and test split. The entire process of training and test split cross-validation is illustrated in Fig. 6.5.

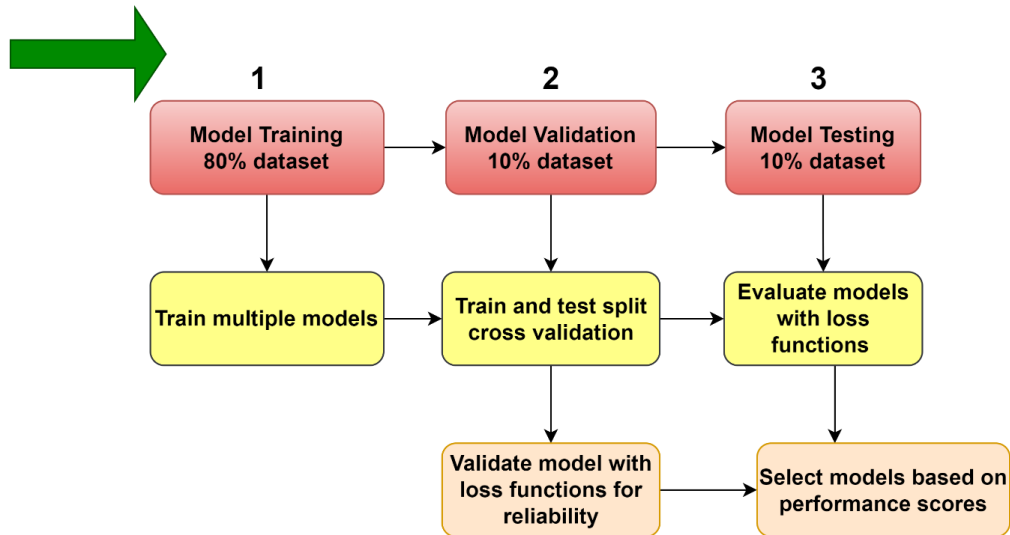


Fig. 6.5 Model validation procedure to ensure the reliability of the selected model (Source: own elaboration)

Loss or cost functions, such as MSE, MSLE, MAPE, and MAE were applied to validate the models and ensure the reliability of the selected model. A validation test was run simultaneously with model training. The loss functions of the model were used to quantify the mismatch between the training and validation data. Furthermore, the best model was selected based on good loss values of less than 0.5.

6.6. Results and Discussion

The obtained results demonstrate that the ANN did not perform uniformly across all water quality parameters. Instead, performance was strongly parameter dependent and was influenced by the activation function, number of hidden layer neurons, and loss function used. Reliable models were obtained for pH, turbidity, temperature, phosphate, nitrate-nitrogen, iron, and manganese, whereas electrical conductivity and all microbiological parameters presented demonstrated substantial prediction difficulties. This difference in performance is a significant finding rather than merely a limitation of the model because it shows that the predictability of a water quality variable depends on the strength of the relationships represented in the training dataset.

6.6.1. Model training and performance evaluation for physical parameters.

i. pH

The model was trained with different deep learning parameters, such as different hidden layer neurons, activation functions, and model loss evaluation. Model loss or cost was evaluated using functions such as MSE, MAPE, MSLE, and MAE. These loss functions were evaluated under two changing conditions: (i) two activation functions, ReLU and Sigmoid; (ii) hidden layer nodes at 4 and 16. The training batch, early stopping patience, number of epochs, and optimiser functions were kept constant. Table 6.1 shows the results of pH model training using different training specifications for all possible model output parameters.

Table 6.1: Model training specifications and training evaluation results for pH prediction

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
pH	8	50	ReLU	Nadam	4	MSE	80.54	<0.5
pH	8	50	ReLU	Nadam	4	MAPE	3.58%	<50%
pH	8	50	ReLU	Nadam	4	MSLE	4.67	<0.5
pH	8	50	ReLU	Nadam	4	MAE	2.78	<0.5
PH	8	50	Sigmoid	Nadam	4	MSE	0.16	<0.5
PH	8	50	Sigmoid	Nadam	4	MAPE	3.36%	<50%
PH	8	50	Sigmoid	Nadam	4	MSLE	0.0	<0.5
PH	8	50	Sigmoid	Nadam	4	MAE	0.26	<0.5
pH	8	50	ReLU	Nadam	16	MSE	3045.91	<0.5
pH	8	50	ReLU	Nadam	16	MAPE	926.83%	<50%
pH	8	50	ReLU	Nadam	16	MSLE	5.56	<0.5
pH	8	50	ReLU	Nadam	16	MAE	314.25	<0.5
PH	8	50	Sigmoid	Nadam	16	MSE	0.11	<0.5
PH	8	50	Sigmoid	Nadam	16	MAPE	3.36%	<50%
PH	8	50	Sigmoid	Nadam	16	MSLE	4.6	<0.5
PH	8	50	Sigmoid	Nadam	16	MAE	0.26	<0.5

Green colour = reliable model; Red colour = unreliable model

The pH model demonstrated one of the clearest differences between the two activation functions evaluated. When ReLU was used with four hidden layer nodes, none of the four loss functions produced an acceptable model. MSE, MSLE, and ME values were 80.54, 4.67, and 2.78 respectively, while MAPE was 3.58%. Increasing the hidden layer neurons to 16 resulted in further deterioration for the ReLU models with MSE increasing to 3045.91, and MAPE to 926.83%.

In contrast sigmoid activation produced substantially better results. With four hidden layer nodes, MSE was 0.16, MSLE was 0.00, and MAE was 0.26, all within the predefined acceptable range. With 16 neurons, MSE remained low at 0.11 and MAE at 0.26, while MSLE increased

substantially. MAPE remained unchanged at 3.36% for sigmoid models. Previous studies have demonstrated the comfortability of the MAPE loss function when combined with sigmoid activation function in reliable predictions of pH (Khoshvaght et al., 2025; Wu & Liu, 2019). In addition, MAPE was only unsuccessful when using ReLU with higher hidden neurons (Table 6.1). Figure 6.6 shows the training and validation data loss functions presented in Table 6.1.

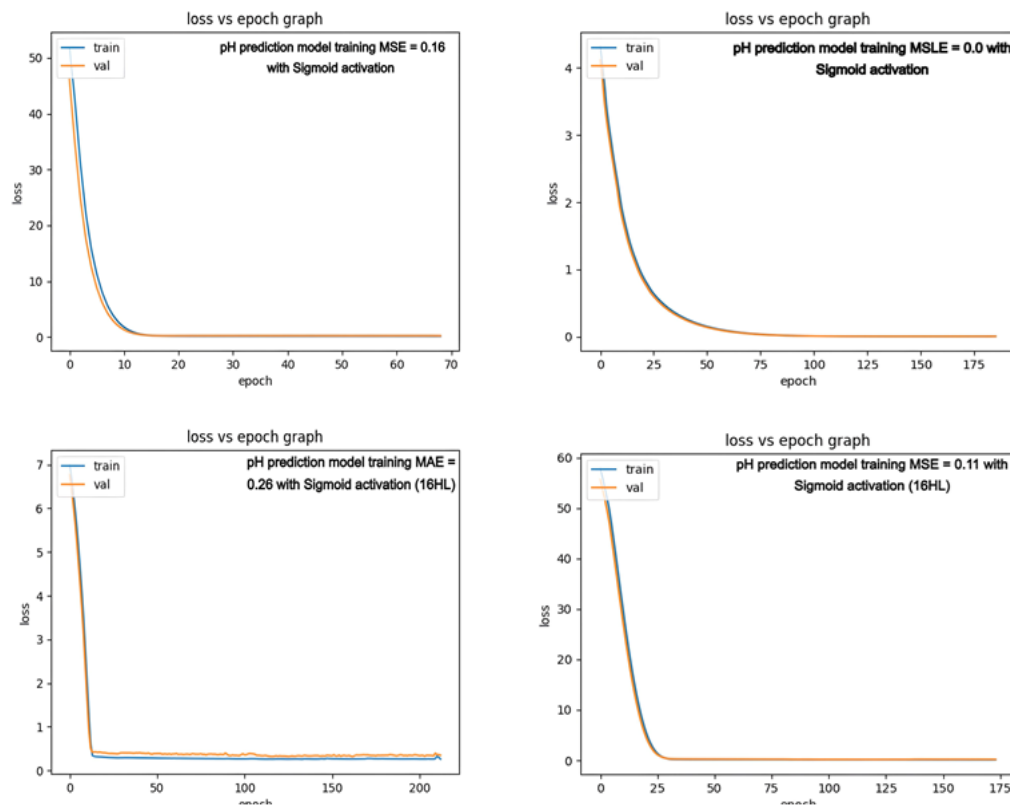


Fig. 6.6 Illustration of the successful cost functions for pH prediction training. The graphs denote the closeness of the validation and training experiments, the closer the graphs are to each other, the more reliable the trained model

ii. Electrical conductivity (EC)

The parameters used to train the model could not produce a reliable prediction model for electrical conductivity with all loss functions above one for all activation functions when using a small or higher number of hidden layer neurons. However, we observed that the MSLE loss value increases with increasing hidden layers (Appendix 5, Table S6.1). This is in contrast with the trend observed by Deng, (2023). Nonetheless, the result is particularly significant because EC was identified as an influential water quality parameter in (Mogane et al., 2024), especially in the Crocodile River in summer. As such, the inability to predict EC reliably does not imply

that EC is unimportant. Rather, it indicates that its temporal behaviour was not sufficiently represented by the available dataset for the ANN to learn a stable relationship (Mateo et al., 2024). The value of EC is highly influenced by the concentration and movement of dissolved ions and can respond to several other processes such as runoff, wastewater inputs, and seasonal hydrological conditions (Belanha et al., 2026; Motamedi et al., 2026). As a result, the poor model performance highlights the distinction between environmental importance and predictive docility.

iii. Turbidity

The MSLE loss function is reliable when used with the ReLU and sigmoid action functions and a lower or higher number of hidden layer neurons. The MAPE reported a good model (36.54%; 37.61%) only when using the sigmoid activation (Appendix 5, Table S6.2). With MAPE, a 20 – 50% loss value is regarded as reasonably reliable. Other studies have also reported MAPE loss values below 50% loss values (Palabıyık & Akkan, 2024) when predicting turbidity. Fig. 6.7 shows the MSLE training and validation cost plots corresponding to Table S6.2. The multiple graphs in Fig. 6.7 demonstrate that MSLE may be the best performance assessment tool for the generated turbidity models.

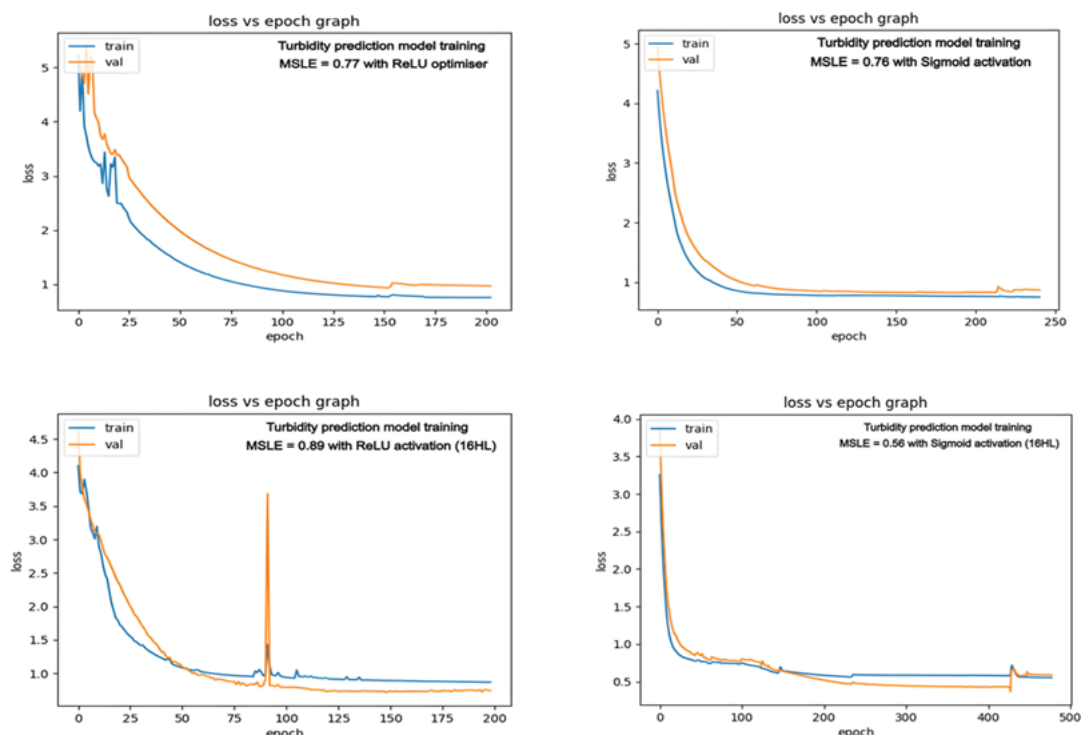


Fig. 6.7 MSLE cost functions for the training and validation of the turbidity prediction model

iv. Temperature

For temperature, the model was only reliable with the MAPE and MSLE cost functions when using both the sigmoid and ReLU activation functions and a low number of hidden layer neurons (Appendix 5, [Table S6.3](#)). Similar studies have shown reliable loss values when using regression functions such as MSLE (0.16) (Nithin et al., 2024) and unreliable models when using other regression models such as MAE (1.140) (Graf et al., 2019).

Fig. 6.8 is a depiction of the best-fit training and validation graphs for some of the training and validation results using the MSLE loss function. Fig. 6.8 illustrates the good alignment of the training and validation datasets, which supported the loss score of $MSLE = 0.02$.

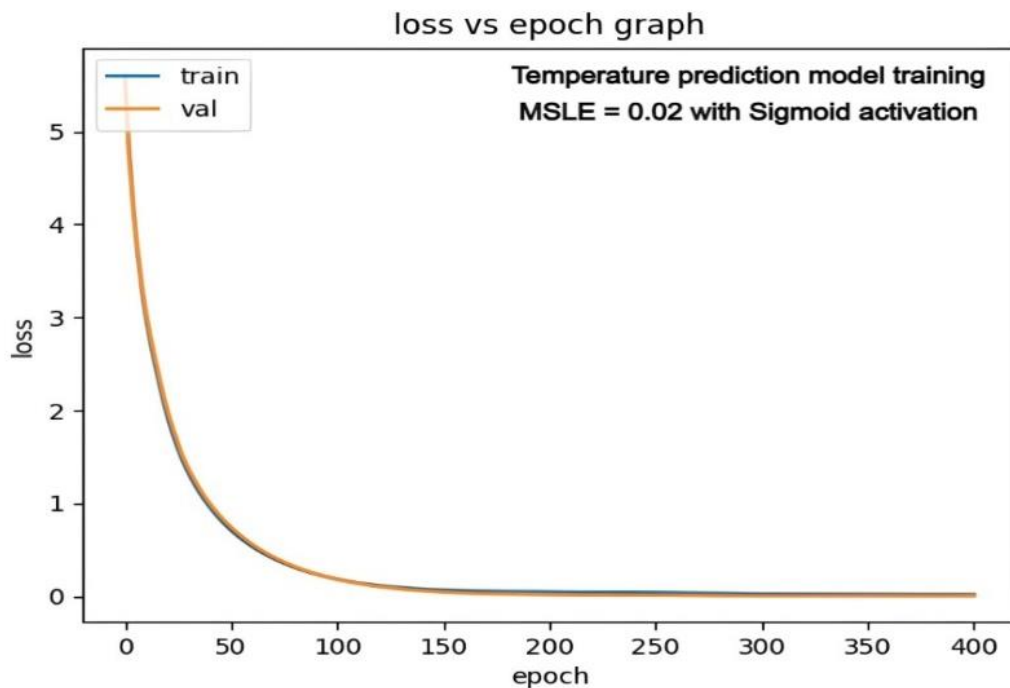


Fig. 6.8 MSLE cost functions for temperature prediction model training and validation

6.6.2. Model training and performance evaluation for chemical parameters.

i. Phosphate

All the trained chemical parameters were successful in either of the loss functions used to evaluate the performance of each model. However, MSLE was more responsible for the most successful models when using either the ReLU or Sigmoid activation functions. MSE failed to produce reliable models with the ReLU activation functions and decreased in accuracy as hidden layers increased. This is in agreement with other phosphate prediction studies where MSE produced unreliable models (1.439) while producing reliable models with other loss functions (Latif et al., 2022). Table 6.2 presents the successful and failed prediction models for phosphate.

Table 6.2: Model training specifications and training evaluation results for phosphate prediction

Output Parameter	Training Batch	Early Stopping patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
PO ₄ ³⁻	8	50	ReLU	Nadam	4	MSE	58.78	<0.5
PO ₄ ³⁻	8	50	ReLU	Nadam	4	MAPE	64.7%	<50%
PO ₄ ³⁻	8	50	ReLU	Nadam	4	MSLE	0.04	<0.5
PO ₄ ³⁻	8	50	ReLU	Nadam	4	MAE	0.1	<0.5
PO ₄ ³⁻	8	50	Sigmoid	Nadam	4	MSE	0.04	<0.5
PO ₄ ³⁻	8	50	Sigmoid	Nadam	4	MAPE	64.09%	<50%
PO ₄ ³⁻	8	50	Sigmoid	Nadam	4	MSLE	0.04	<0.5
PO ₄ ³⁻	8	50	Sigmoid	Nadam	4	MAE	0.09	<0.5
PO ₄ ³⁻	8	50	ReLU	Nadam	16	MSE	1465.19	<0.5
PO ₄ ³⁻	8	50	ReLU	Nadam	16	MAPE	61.6%	<50%
PO ₄ ³⁻	8	50	ReLU	Nadam	16	MSLE	1.29	<0.5
PO ₄ ³⁻	8	50	ReLU	Nadam	16	MAE	4.62	<0.5
PO ₄ ³⁻	8	50	Sigmoid	Nadam	16	MSE	0.03	<0.5
PO ₄ ³⁻	8	50	Sigmoid	Nadam	16	MAPE	61.16%	<50%
PO ₄ ³⁻	8	50	Sigmoid	Nadam	16	MSLE	0.02	<0.5
PO ₄ ³⁻	8	50	Sigmoid	Nadam	16	MAE	0.09	<0.5

Green colour = reliable model; Red colour = unreliable model

For PO_4^{3-} prediction training, only the MAPE loss function produced an unreliable model with both ReLU and Sigmoid activation functions at both low and high hidden layers (Table 6.2). Fig. 6.9 illustrates some successful training-validation graphs obtained from MSLE.

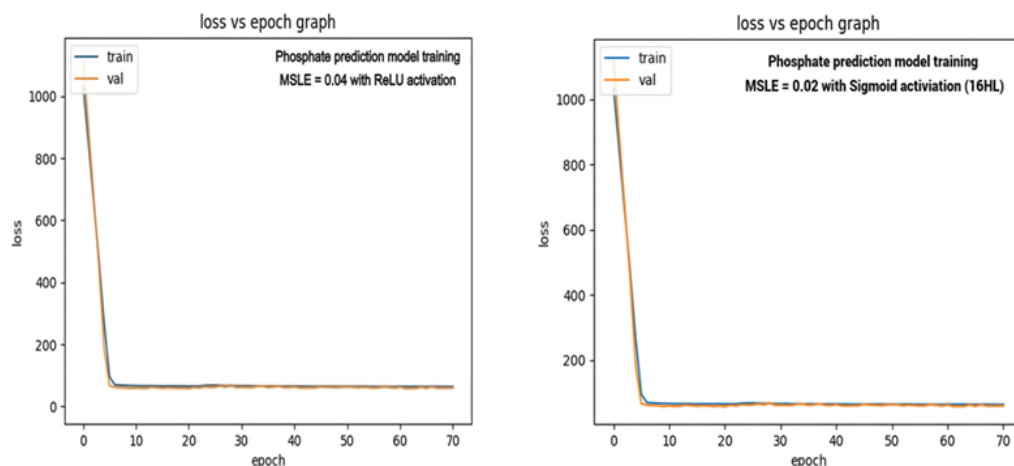


Fig. 6.9 Loss functions for phosphate prediction model training and validation show similar learning patterns exhibited by different models

ii. Nitrate-Nitrogen

The model training results for $\text{NO}_3\text{-N}$ were only positive twice with MSLE (0.23; 0.21) (Appendix 5, Table S6.4). The results show that the sigmoid activation function is more suited to train the nitrate prediction model with or without a change in the number of hidden layer neurons. The detailed training information is presented in Table S6.4. Fig. 6.10 illustrates some of the training-validation plots that were successful for nitrate-nitrogen forecasting. Fig. 6.10 also shows that training accuracy increases (0.23-0.21) when the hidden layer neurons are increased, when the model is evaluated with the same loss function. The ability to predict nitrate-nitrogen may therefore contribute to early identification of nutrient enrichment (Mahlknecht et al., 2023). Nonetheless, the narrow range of successful model configurations indicates that the nitrate model remains sensitive to model architecture and may require further validation before operational application.

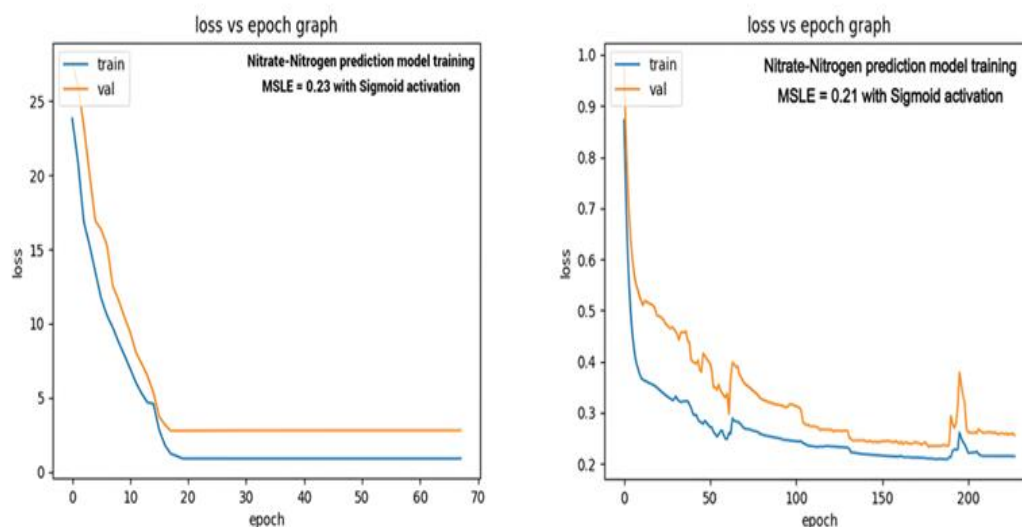


Fig. 6.10 Illustration of loss function results from training of the nitrate-nitrogen prediction model

iii. Iron

The Iron (Fe) prediction model obtained one of the best results among the chemical parameters. [Table S6.5](#) lists the reliable and failed models. The Fe-trained models were successful when evaluated against three different loss functions (MSLE, MAPE, and MAE) using flexible activation functions and hidden layer nodes. Fig. 6.11 presents training-validation loss graphs computed using the MAPE loss function. This is the only time during which the MAPE loss function produced a reliable model. It is also important to note that this function only worked when the sigmoid activation function was utilised.

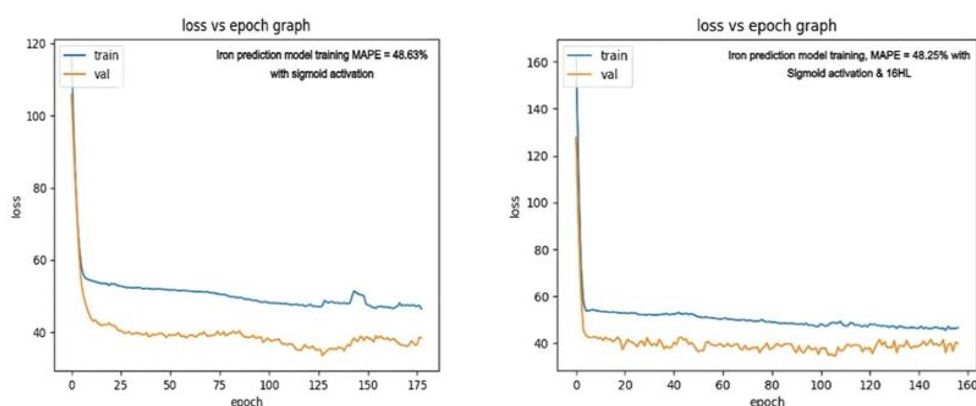


Fig. 6.11 Illustration of the loss function results from the training of the Iron prediction model

iv. *Manganese*

The ReLU and Sigmoid activation functions produced good loss functions, except for MAPE (Appendix 5, [Table S6.6](#)), with a lower number of hidden layers. However, the same did not produce reliable models when the number of hidden layers was increased. The failure to produce reliable models with a higher number of hidden layers can also be attributed to loss functions and their data handling characteristics (Deng, 2023). Fig. 6.12 illustrates some of the successful loss measures for Mn prediction model training.

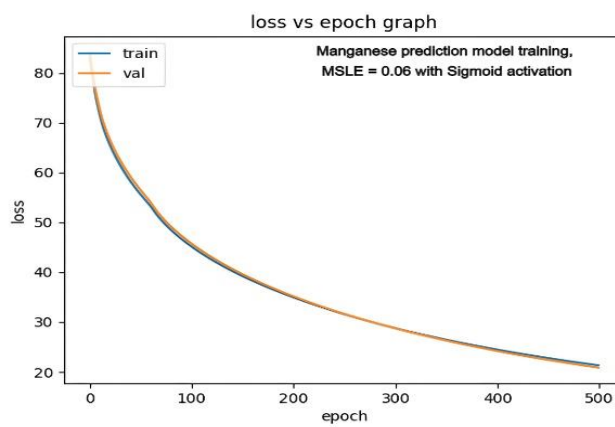


Fig. 6.12 Loss function results from training the Manganese prediction model

6.6.3. Model training and performance evaluation for microbiological parameters.

i. *E. coli*

A total of 16 models were trained for each microbiological parameter. All models trained using the microbiological parameters failed to produce a reliable model. This result can be attributed to the amount of data used to train these parameters. The training method was trial and error, perhaps more time was required to tweak the model architecture when training the microbiological data for forecasting. One of the significant observations is that the MSLE performance appraisal produced the best models for all the microbiological parameters. Table 6.3 presents the performance of the loss functions for *E. coli* model training.

Table 6.3: Model training specifications and training evaluation results for *E. coli* prediction

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
<i>E. coli</i>	8	50	ReLU	Nadam	4	MSE	39126282240	<0.5
<i>E. coli</i>	8	50	ReLU	Nadam	4	MAPE	1221.23%	<50%
<i>E. coli</i>	8	50	ReLU	Nadam	4	MSLE	14.43	<0.5
<i>E. coli</i>	8	50	ReLU	Nadam	4	MAE	34351.47	<0.5
<i>E. coli</i>	8	50	Sigmoid	Nadam	4	MSE	47127052288	<0.5
<i>E. coli</i>	8	50	Sigmoid	Nadam	4	MAPE	3737.69%	<50%
<i>E. coli</i>	8	50	Sigmoid	Nadam	4	MSLE	11.54	<0.5
<i>E. coli</i>	8	50	Sigmoid	Nadam	4	MAE	47257.41	<0.5
<i>E. coli</i>	8	50	ReLU	Nadam	16	MSE	40598679552	<0.5
<i>E. coli</i>	8	50	ReLU	Nadam	16	MAPE	6775258%	<50%
<i>E. coli</i>	8	50	ReLU	Nadam	16	MSLE	2.14	<0.5
<i>E. coli</i>	8	50	ReLU	Nadam	16	MAE	33500,3	<0.5
<i>E. coli</i>	8	50	Sigmoid	Nadam	16	MSE	47086170112	<0.5
<i>E. coli</i>	8	50	Sigmoid	Nadam	16	MAPE	438191.66%	<50%
<i>E. coli</i>	8	50	Sigmoid	Nadam	16	MSLE	46.43	<0.5
<i>E. coli</i>	8	50	Sigmoid	Nadam	16	MAE	47268.93	<0.5

Red colour = unreliable model

Although *E. coli* is an important indicator of pollution, the 16 trained models could not produce a reliable model. However, there was a promising model that produced a cost value of MSLE = 2.14 (Table 6.3). Further training with increased data and hidden layers may yield improved results. As a result, the inability to predict *E. coli* should not be interpreted as evidence that environmental variables are unrelated to microbial contamination. Rather, the result suggests that the available temporal resolution and dataset size were insufficient to capture the complex dynamics of faecal pollution.

ii. *Total coliforms*

Similar results were obtained for training the total coliform prediction model; only MSLE = 3.19, MSLE = 2.15, and MSLE = 16.54 came close to producing reliable forecast models (Appendix 5, [Table S6.7](#)). This result is consistent with the difficulty observed for *E. coli*.

iii. *Bifidobacteria*

The Bifidobacteria models were consistent with the trend of being close to reliability only when subjected to the MSLE loss functions, irrespective of the activation function or hidden layer nodes used. From [Table S6.8](#), the loss functions constantly range between MSLE = 4.78 and 6.79 with a quarter of the values ranging around 6. This indicates that the model does not have sufficient data, and the training parameters should remain unchanged.

iv. *Faecal streptococci*

The observations from the trained faecal streptococci models suggested that two loss functions (MSLE and MAE) favour the models' performance (Appendix 5, [Table S6.9](#)). However, no acceptable models are available to continue testing with reliable results. Again, this is likely a phenomenon of insufficient training data.

6.6.4. Model testing and performance appraisal.

i. Physical parameters testing time series

The three physical parameters pH, temperature, and turbidity produced good and acceptable training models. These models were then tested to predict known data and forecast three datasets in the future. Fig. 6.13 presents the experimental results. However, because we used an unreliable model to estimate the electrical conductivity, the prediction was also unreliable.

Part C of Fig. 6.13 illustrates how low the predicted values are even though the model should first match the actual data points before forecasting. The of the test of the performance evaluation for Part A (pH); B (Temperature), C (Electrical conductivity), and D (Turbidity) was $MSE = 0.02$; $MSLE = 0.19$; $MSLE = 1.93$; and $MSLE = 0.0$, respectively.

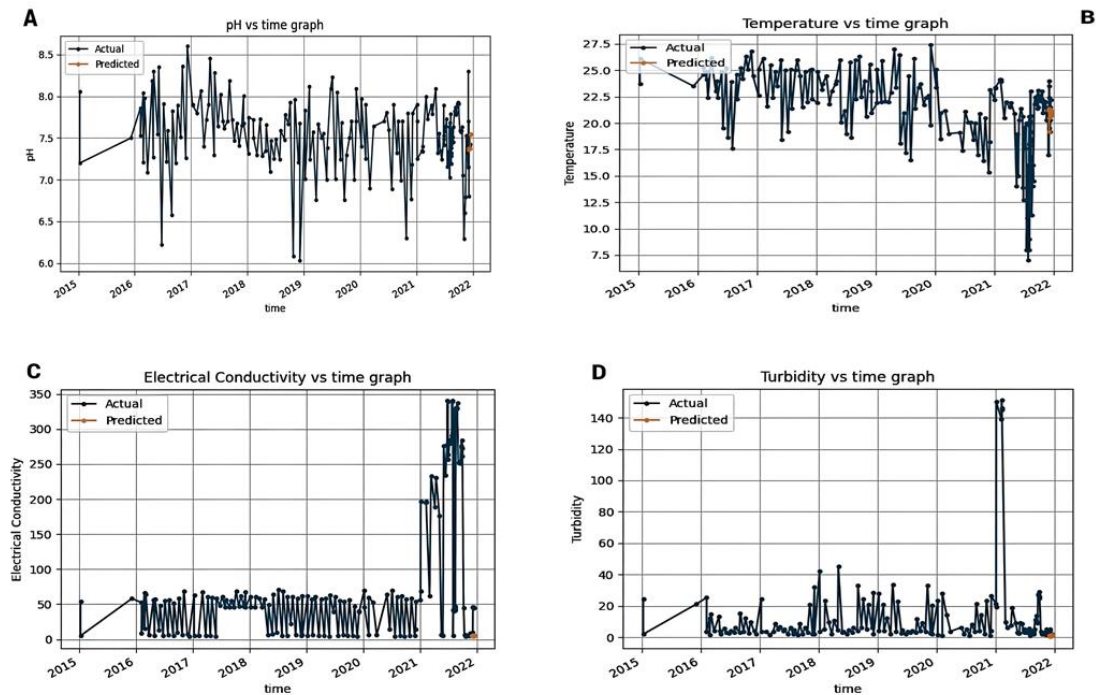


Fig. 6.13 Model prediction testing time series graphs for all physical parameters under study

ii. Chemical parameters testing time series.

All successfully trained models were used to test the prediction power of the models. Like the physical parameters, the models were experimented with by allowing them to predict known data and then predict three datasets in the future. Parts A, B, C, and D, in Fig. 6.13 represent Mn, PO_4 , Fe, and NO_3-N respectively. The predictions presented in Fig. 6.13 provide a satisfactory evaluation of the test performance. The Mn and PO_4 prediction models were evaluated with $MSE = 0.04$, $MSE = 0.0$, $MSE = 0.88$ respectively, whereas Fe was evaluated using $MAPE = 9.21\%$.

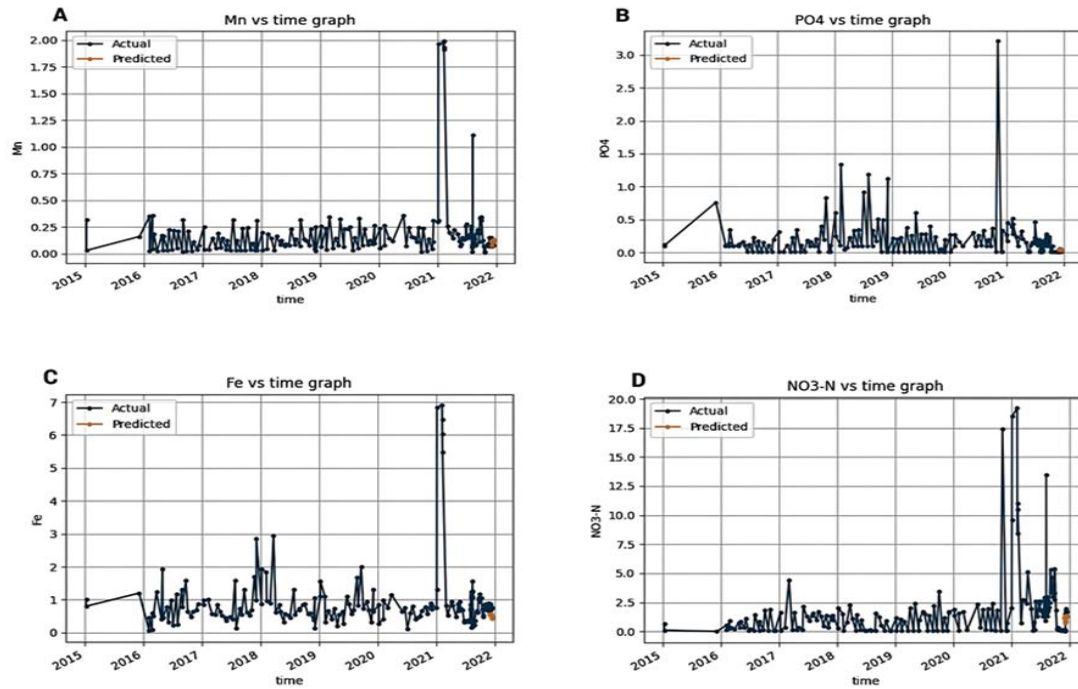


Fig. 6.14 Model prediction testing time series graphs for all chemical parameters under study

iii. Microbiological parameters testing graphs over time.

All the trained models predicting the microbiological quality of water could not produce any reliable models. However, in the interest of curiosity, the study conducted a testing experiment using the models that produced the lowest loss values. The same procedure of first predicting known data and forecasting future data was applied to these models. However, the results observed only confirm that the models are not reliable. Fig. 6.10 shows the testing experiments conducted with actual and predicted data.

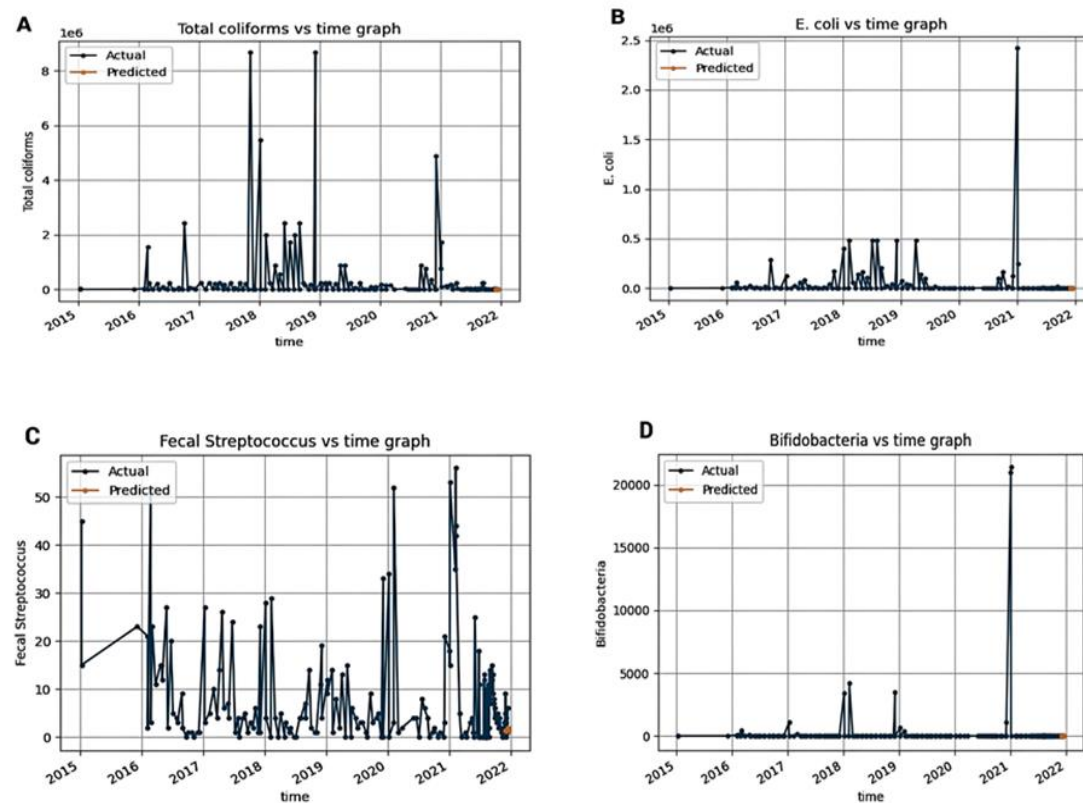


Fig. 6.15 Model prediction testing time series graphs for all microbiological parameters under study. Parts A, B, C, and D, specifically demonstrate the prediction of known data points prior to future predictions

6.6.5. Inherent aspects that influence model performance

i. Data splitting

Data splitting in machine learning assists in assigning data for different purposes such as training and evaluation. Different models are designed with different data splitting ratios ranging from 90% and 10% to 70% training and 30% validation (Nguyen et al., 2021). This study used an 80:20 ratio to standardise the training environment and avoid underfitting, which may occur when the data are insufficient for training. Data-splitting approaches have always yielded different results in relation to model performance. For example, Pham et al. (2020) reported an improvement in training when the training data size was increased from 30% to 80%. However, an opposite trend was observed in model performance when training data size was increased from 80% to 90%. Qi et al. (2018), also reported that training data size generally has a significant impact on model performance. The results of this study have demonstrated that data quantity remained a fundamental limitation to model performance. This finding has

implications beyond the ANN itself. It suggests that future water quality monitoring programmes should be designed with two purposes in mind: first, to establish compliance with water quality objectives; and second, to generate sufficient and representative datasets for predictive modelling.

ii. Activation functions

The primary function of the activation function is to activate the input neurons during the training of a model. According to Dubey et al. (2022), typical activation functions have specific characteristics such as i). “Introduction of non-linearity to improve model training, ii). Simplify or reduce the computational complexity of the model, iii). They should not interfere with gradient flow during training, and iv) they should maintain the data distribution to facilitate better network training”. The current prediction model was designed with 14 activation functions to select the most effective but used only two for comparison. In a study conducted by Dubey et al. (2022), 18 activation functions were compared, and it was concluded that the Sigmoid activation function should be avoided for Convolutional Neural Networks (CNN) as it leads to poor convergence. The authors further concluded that although ReLU and its derivatives are popular, other activation functions such as Swish and Mish should be used to solve different problems. Moreover, the parametric activation function such as Sigmoid, generally demonstrates better convergence because it adapts to data faster by learning the parameters from the data (Dubey et al., 2022). In this study, the most successful models were calculated using the sigmoid activation function, which achieved an overall performance of 30.21% compared to 10.42% for ReLU. When examining models based on physical parameters, sigmoid scored 50.0% while ReLU achieved 8.33%. Improved performance was observed for both activation functions when applied to chemical parameter data, with ReLU scoring 33.33% and sigmoid achieving 70.83%. These results indicate that the sigmoid activation function is better suited for the dataset used in this study. Additionally, the results show that activation functions performance is dataset specific. The superior performance of sigmoid in this study should therefore not be interpreted as evidence that sigmoid is universally superior to ReLU. Instead, it shows that the structure and distribution of the water quality dataset were more compatible with sigmoid activation.

iii. Loss/cost functions employed

Each machine learning model must meet specific prediction requirements to be considered reliable. Loss functions play a significant role in assessing and improving the performance of machine learning algorithms (Wang et al., 2022). The MSE is one of the simplest loss functions (C. , Sammut & Webb, 2011) with MSLE as its closest variable (Acharya et al., 2019). The present study experimented with four different loss functions (MSE, MSLE, MAPE, and MAE) to produce reliable models with the MSLE producing the most reliable models. According to Shan et al. (2023), the MSE and MAE loss functions have strong data-denoising properties. Thus, it is important to consider the type of data used when selecting loss functions to evaluate the performance of machine learning models. Considering this, the stronger performance of MSLE can be interpreted in relation to the characteristics of the water quality data.

6.6.6. Limitations of the developed models.

The performance of the ANN model is heavily dependent on the quantity and quality of the available data. Limited data for certain parameters, such as electrical conductivity and microbiological indicators, led to a reduced predictive precision for these variables. In addition, the model reliability was not consistent across all water quality parameters, indicating potential challenges in generalising the tool for different contexts without additional data and model tuning. Moreover, the application of this modelling tool in this study served as a pilot to determine applicability. Although the tool was able to generate 62% reliable models, the current version of the model is primarily a reference tool and may not be fully reliable for real-time or critical decision-making without further enhancements and validations. Furthermore, the development and implementation of the ANN-based tool require a significant level of expertise in machine learning and data handling, which may limit its accessibility for broader use, especially in resource-limited settings.

6.7. Conclusions and prospects

This study focused on the development of a water quality prediction tool using an artificial neural network (ANN) algorithm, designed to improve water quality monitoring and management. Traditional methods are often time-consuming and expensive, making them less suitable for rural areas with limited resources. The proposed tool integrates 13 water quality parameters, including pH, turbidity, electrical conductivity, and various chemical and microbiological indicators, to provide rapid and reliable predictions. The model was trained

using 80% of the data and validated with the remaining 20%, employing various metrics such as MSE, MAPE, MAE, and MSLE to evaluate performance. Although the tool generally showed reliability, it struggled to predict electrical conductivity and microbiological parameters, likely due to insufficient data.

Nonetheless, this study successfully developed a predictive tool using an Artificial Neural Network (ANN) to forecast water quality parameters, demonstrating the potential of machine learning in improving water resource management. Despite its promising results, the model's effectiveness is constrained by the availability and diversity of data. The tool represents a significant step forward in the prediction of water quality, offering a more efficient and potentially more accurate alternative to traditional methods. However, to fully realise its potential, more research is needed to address its limitations, particularly to expand the data set and refine the model to improve its accuracy and applicability in various scenarios. Future work should focus on continuous data collection and model optimisation to ensure the tool's robustness and reliability, making it an asset for water quality management in diverse environments.

6.8. Statements and Declarations

Competing interests: The authors declared no conflicting interests regarding publishing this paper.

Data availability: The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Funding: No funding was obtained for this study

Author contributions

Lazarus Katlego Mogane: was responsible for conceptualisation, data collection and curation, coding of the software, writing of the first draft and final revision. **Esper J. Ncube:** was responsible for the conceptualisation, editing, and reviews. **Titus A.M. Msagati:** was responsible for the conceptualisation, editing and reviews. **Tracy M. Masebe:** was responsible for the conceptualisation, Methodology, editing, and reviews. All the authors read and approved the final draft of the manuscript.

6.9. References

- Acharya, M. S., Armaan, A., & Antony, A. S. (2019, February 1). A comparison of regression models for prediction of graduate admissions. *ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings*. <https://doi.org/10.1109/ICCIDS.2019.8862140>
- Aldreess, A., Awan, H. H., Javed, M. F., & Mohamed, A. M. (2022). Prediction of water quality indexes with ensemble learners: Bagging and boosting. *Process Safety and Environmental Protection*, 168, 344–361. <https://doi.org/10.1016/j.psep.2022.10.005>
- Alnuwaiser, M. A., Javed, M. F., Khan, M. I., Ahmed, M. W., & Galal, A. M. (2022). Support vector regression and ANN approach for predicting the ground water quality. *Journal of the Indian Chemical Society*, 99(7). <https://doi.org/10.1016/j.jics.2022.100538>
- Awadallah, M. A., Abu-Doush, I., Al-Betar, M. A., & Braik, M. S. (2023). Metaheuristics for optimizing weights in neural networks. *Comprehensive Metaheuristics: Algorithms and Applications*, 359–377. <https://doi.org/10.1016/B978-0-323-91781-0.00005-3>
- Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment*, 721. <https://doi.org/10.1016/j.scitotenv.2020.137612>
- Chen, X., Wang, Z., Liu, J., Gong, C., & Pang, Y. (2022). A Neural Network-Based Mesh Quality Indicator for Three-Dimensional Cylinder Modelling. *Entropy*, 24(9). <https://doi.org/10.3390/e24091245>
- Chounlamany, V., Tanchuling, M. A., & Inoue, T. (2017). Spatial and temporal variation of water quality of a segment of Marikina River using multivariate statistical methods. *Water Science and Technology*, 76(6), 1510–1522. <https://doi.org/10.2166/wst.2017.279>
- da S. Belanha, L., Paz, O. L. S., Padial, A. A., de Azevedo, J. C. R., Hughes, R. M., & Ruaro, R. (2026). Water pollution drives environmental degradation in a seasonally influenced Neotropical coastal river. *Environmental Monitoring and Assessment*, 198(4), 322. <https://doi.org/10.1007/s10661-025-14956-w>
- Deng, T. (2023). Effect of the Number of Hidden Layer Neurons on the Accuracy of the Back Propagation Neural Network. In *Highlights in Science, Engineering and Technology MISBP* (Vol. 2023).
- Diehl, P. U., Pedroni, B. U., Cassidy, A., Merolla, P., Neftci, E., & Zarrella, G. (2016). TrueHappiness: Neuromorphic Emotion Recognition on TrueNorth. *International Joint Conference on Neural Networks*. <https://doi.org/http://dx.doi.org/10.1109/IJCNN.2016.7727758>
- Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022a). Activation functions in deep learning: A comprehensive survey and benchmark. In *Neurocomputing* (Vol. 503, pp. 92–108). Elsevier B.V. <https://doi.org/10.1016/j.neucom.2022.06.111>

- Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022b). Activation functions in deep learning: A comprehensive survey and benchmark. In *Neurocomputing* (Vol. 503, pp. 92–108). Elsevier B.V. <https://doi.org/10.1016/j.neucom.2022.06.111>
- Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, 64(11), 2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Graf, R., Zhu, S., & Sivakumar, B. (2019). Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *Journal of Hydrology*, 578. <https://doi.org/10.1016/j.jhydrol.2019.124115>
- Gupta, S., & Gupta, S. K. (2021). A critical review on water quality index tool: Genesis, evolution and future directions. *Ecological Informatics*, 63. <https://doi.org/10.1016/j.ecoinf.2021.101299>
- Izhar Shah, M., Alaloul, W. S., Alqahtani, A., Aldrees, A., Ali Musarat, M., & Javed, M. F. (2021). Predictive modeling approach for surface water quality: Development and comparison of machine learning models. *Sustainability (Switzerland)*, 13(14). <https://doi.org/10.3390/su13147515>
- Khan, M., Liu, T., & Ullah, F. (2019). A new hybrid approach to forecast wind power for large scale wind turbine data using deep learning with tensorflow framework and principal component analysis. *Energies*, 12(11). <https://doi.org/10.3390/en12122229>
- Khoshvaght, H., Permala, R. R., Razmjou, A., & Khiadani, M. (2025). Factors influencing the effectiveness of machine learning algorithms in predicting raw wastewater characteristics: An analysis based on weekly field data. *Journal of Water Process Engineering*, 78, 108702. <https://doi.org/10.1016/J.JWPE.2025.108702>
- Latif, S. D., Birima, A. H., Ahmed, A. N., Hatem, D. M., Al-Ansari, N., Fai, C. M., & El-Shafie, A. (2022). Development of prediction model for phosphate in reservoir water system based machine learning algorithms. *Ain Shams Engineering Journal*, 13(1). <https://doi.org/10.1016/j.asej.2021.06.009>
- Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging: General overview. In *Korean Journal of Radiology* (Vol. 18, Issue 4, pp. 570–584). Korean Radiological Society. <https://doi.org/10.3348/kjr.2017.18.4.570>
- Madni, H. A., Umer, M., Ishaq, A., Abuzinadah, N., Saidani, O., Alsubai, S., Hamdi, M., & Ashraf, I. (2023). Water-Quality Prediction Based on H2O AutoML and Explainable AI Techniques. *Water (Switzerland)*, 15(3). <https://doi.org/10.3390/w15030475>
- Mahlknecht, J., Torres-Martínez, J. A., Kumar, M., Mora, A., Kaown, D., & Loge, F. J. (2023). Nitrate prediction in groundwater of data scarce regions: The futuristic fresh-water management outlook. *Science of The Total Environment*, 905, 166863. <https://doi.org/10.1016/J.SCITOTENV.2023.166863>

- Malek, N. H. A., Yaacob, W. F. W., Nasir, S. A. M., & Shaadan, N. (2022). Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. *Water (Switzerland)*, 14(7). <https://doi.org/10.3390/w14071067>
- Mateo, L. F., Más-López, M. I., García-del-Toro, E. M., García-Salgado, S., & Quijano, M. Á. (2024). Artificial Neural Networks to Predict Electrical Conductivity of Groundwater for Irrigation Management: Case of Campo de Cartagena (Murcia, Spain). *Agronomy*, 14(3). <https://doi.org/10.3390/agronomy14030524>
- Mogane, L. K., Masebe, T., Msagati, T. A. M., & Ncube, E. (2023). A comprehensive review of water quality indices for lotic and lentic ecosystems. In *Environmental Monitoring and Assessment* (Vol. 195, Issue 8). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s10661-023-11512-2>
- Mogane, L. K., Ncube, E. J., Msagati, T. A. M., & Masebe, T. M. (2024). Impact of parametric seasonal variations on water quality in the Crocodile River and Inyaka Dam in the Mpumalanga Province, South Africa. *Heliyon*, 10(19). <https://doi.org/10.1016/j.heliyon.2024.e38246>
- Mohseni-Dargah, M., Falahati, Z., Dabirmanesh, B., Nasrollahi, P., & Khajeh, K. (2022). Machine learning in surface plasmon resonance for environmental monitoring. *Artificial Intelligence and Data Science in Environmental Sensing*, 269–298. <https://doi.org/10.1016/B978-0-323-90508-4.00012-5>
- Motamedi, B., Heiderscheidt, E., Tesfamariam, A., & Rossi, P. M. (2026). Electrical conductivity and groundwater elevation as parameters for continuous infiltration and inflow detection in wastewater networks. *Journal of Hydrology*, 677, 135936. <https://doi.org/10.1016/J.JHYDROL.2026.135936>
- Nagpal, C., & Dubey, S. R. (2019). A Performance Evaluation of Convolutional Neural Networks for Face Anti Spoofing. *International Joint Conference on Neural Networks*. <http://arxiv.org/abs/1805.04176>
- Najafzadeh, M., Ghaemi, A., & Emamgholizadeh, S. (2019). Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science and Technology*, 16(10), 6377–6396. <https://doi.org/10.1007/s13762-018-2049-4>
- Najah Ahmed, A., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., Ehteram, M., & Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*, 578, 124084. <https://doi.org/10.1016/J.JHYDROL.2019.124084>
- Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Van Le, H., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021. <https://doi.org/10.1155/2021/4832864>
- Nithin, G., Sai Trilochan, M., Sangamesh, G., Reddy, M. V., Shaik, S., & Students, I. Y. (2024). Regression techniques based on whether forecasting prediction (Vol. 11). JETIR. www.jetir.orgk390
- Palabıyık, S., & Akkan, T. (2024). Evaluation of water quality based on artificial intelligence: performance of multilayer perceptron neural networks and multiple linear regression versus

water quality indexes. *Environment, Development and Sustainability*.
<https://doi.org/10.1007/s10668-024-05075-6>

- Pham, B. T., Qi, C., Ho, L. S., Nguyen-Thoi, T., Al-Ansari, N., Nguyen, M. D., Nguyen, H. D., Ly, H. B., Van Le, H., & Prakash, I. (2020). A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil. *Sustainability (Switzerland)*, 12(6), 1–16. <https://doi.org/10.3390/su12062218>
- Pras, A., & Mamane, H. (2023). Nowcasting of fecal coliform presence using an artificial neural network. *Environmental Pollution*, 326, 121484. <https://doi.org/10.1016/j.envpol.2023.121484>
- Qi, C., Fourie, A., Chen, Q., & Zhang, Q. (2018). A strength prediction model using artificial intelligence for recycling waste tailings as cemented paste backfill. *Journal of Cleaner Production*, 183, 566–578. <https://doi.org/10.1016/j.jclepro.2018.02.154>
- Rahman, A. (2019). Based Data Preprocessing Methods and Machine Learning Algorithms for Big Data Analysis. In *International Journal of Artificial Intelligence* (Vol. 17, Issue 2).
- Rahman, A. (2020). Statistics for Data Science and Policy Analysis. In *Statistics for Data Science and Policy Analysis*. Springer Singapore. <https://doi.org/10.1007/978-981-15-1735-8>
- Rustam, F., Ishaq, A., Kokab, S. T., de la Torre Diez, I., Mazón, J. L. V., Rodríguez, C. L., & Ashraf, I. (2022). An Artificial Neural Network Model for Water Quality and Water Consumption Prediction. *Water (Switzerland)*, 14(21). <https://doi.org/10.3390/w14213359>
- Sammur, C., & Webb, G. I. (2011). Mean Squared Error. In C., W. G. I. Sammut (Ed.), *Encyclopedia of Machine Learning* (pp. 653–653). Springer US. https://doi.org/10.1007/978-0-387-30164-8_528
- Sattari, M. T., Joudi, A. R., & Kusiak, A. (2016). Estimation of water quality parameters with data-driven model. *Journal - American Water Works Association*, 108(4), E232–E239. <https://doi.org/10.5942/jawwa.2016.108.0012>
- Shah, M. I., Abunama, T., Javed, M. F., Bux, F., Aldrees, A., Rehman Tariq, M. A. U., & Mosavi, A. (2021). Modeling surface water quality using the adaptive neuro-fuzzy inference system aided by input optimization. *Sustainability (Switzerland)*, 13(8). <https://doi.org/10.3390/su13084576>
- Shah, M. I., Javed, M. F., & Abunama, T. (2021). Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environmental Science and Pollution Research*, 28, 13202–13220. <https://doi.org/10.1007/s11356-020-11490-9>/Published
- Shan, H., Vimieiro, R. B., Borges, L. R., Vieira, M. A. C., & Wang, G. (2023). Impact of loss functions on the performance of a deep neural network designed to restore low-dose digital mammography. *Artificial Intelligence in Medicine*, 142. <https://doi.org/10.1016/j.artmed.2023.102555>
- Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation functions in neural networks. In *International Journal of Engineering Applied Sciences and Technology* (Vol. 4). <http://www.ijeast.com>

- Sutadian, A. D., Muttill, N., Yilmaz, A. G., & Perera, B. J. C. (2016). Development of river water quality indices—a review. *Environmental Monitoring and Assessment*, 188(1), 1–29. <https://doi.org/10.1007/s10661-015-5050-0>
- Sutadian, A. D., Muttill, N., Yilmaz, A. G., & Perera, B. J. C. (2018). Development of a water quality index for rivers in West Java Province, Indonesia. *Ecological Indicators*, 85(April 2017), 966–982. <https://doi.org/10.1016/j.ecolind.2017.11.049>
- Szandała, T. (2020). Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. *Springer Indonesia*, 903. https://doi.org/https://doi.org/10.1007/978-981-15-5495-7_11
- Tsioptsias, N., Tako, A., & Robinson, S. (2016). Model validation and testing in simulation: A literature review. *OpenAccess Series in Informatics*, 50, 6.1-6.11. <https://doi.org/10.4230/OASIs.SCOR.2016.6>
- Türk, Ö. (2021). Classification of electroencephalogram records related to cursor movements with a hybrid method based on deep learning. *International Journal of Imaging Systems and Technology*, 31(4), 2322–2333. <https://doi.org/10.1002/ima.22643>
- Uddin, M. G., Nash, S., Mahammad Diganta, M. T., Rahman, A., & Olbert, A. I. (2022). Robust machine learning algorithms for predicting coastal water quality index. *Journal of Environmental Management*, 321. <https://doi.org/10.1016/j.jenvman.2022.115923>
- Uddin, M. G., Nash, S., & Olbert, A. I. (2021a). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. <https://doi.org/10.1016/j.ecolind.2020.107218>
- Uddin, M. G., Nash, S., & Olbert, A. I. (2021b). A review of water quality index models and their use for assessing surface water quality. In *Ecological Indicators* (Vol. 122). Elsevier B.V. <https://doi.org/10.1016/j.ecolind.2020.107218>
- Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023a). A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. *Water Research*, 229. <https://doi.org/10.1016/j.watres.2022.119422>
- Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023b). A sophisticated model for rating water quality. *Science of the Total Environment*, 868. <https://doi.org/10.1016/j.scitotenv.2023.161614>
- Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023c). Assessing optimization techniques for improving water quality model. *Journal of Cleaner Production*, 385. <https://doi.org/10.1016/j.jclepro.2022.135671>
- Vu, N. T., & Do, K. U. (2021). Prediction of Ammonium Removal by Biochar Produced From Agricultural Wastes Using Artificial Neural Networks: Prospects and Bottlenecks. *Soft Computing Techniques in Solid Waste and Wastewater Management*, 455–467. <https://doi.org/10.1016/B978-0-12-824463-0.00012-4>
- Wang, H., Wang, Y., Zhang, Q., Xiang, S., & Pan, C. (2017). Gated convolutional neural network for semantic segmentation in high-resolution images. *Remote Sensing*, 9(5). <https://doi.org/10.3390/rs9050446>

- Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2022). A Comprehensive Survey of Loss Functions in Machine Learning. *Annals of Data Science*, 9(2), 187–212. <https://doi.org/10.1007/s40745-020-00253-5>
- Wu, W., & Liu, H.-B. (2019). Estimation of soil pH with geochemical indices in forest soils. *PLOS ONE*, 14(10), e0223764-. <https://doi.org/10.1371/journal.pone.0223764>
- Yang, X.-S. (2019). Neural networks and deep learning. *Introduction to Algorithms for Data Mining and Machine Learning*, 139–161. <https://doi.org/10.1016/B978-0-12-817216-2.00015-6>
- Zhang, X., Lou, Y., Hu, S., & Zhu, D. (2021). Backpropagation Neural Network Algorithm-Based Color Doppler Ultrasound Detection of Gestational Diabetes Mellitus and Perinatal Outcomes. *Scientific Programming*, 2021. <https://doi.org/10.1155/2021/4258360>

7: Synthesis of chapter findings, Conclusions and Recommendations

7.1. Preamble

This chapter presents a synthesis of the findings and draws together the key aspects explored throughout the study. It captures a summary of the research, articulates conclusive insights, and proposes actionable recommendations for stakeholders and decision-makers. Through diligent investigation and thoughtful consideration, the study has endeavoured to provide valuable insights into the water resource quality management field.

Through this chapter, we intend to offer a cogent overview of the research, enabling readers to grasp the significance of our findings and the subsequent implications for future actions. The culmination of this effort is reflected in the synthesis of information, the identification of trends, and the formulation of actionable recommendations. This chapter should serve as a compass guiding future initiatives. We hope that the insights contained herein will inform strategic decision-making, inspire further exploration, and contribute to the collective knowledge in the field.

7.2. Synthesis of chapter findings

This study sought to develop a water quality assessment tool for surface water resources in the Mpumalanga province of South Africa. This study was motivated by the need to improve the assessment and management of surface water resources in a context characterised by water scarcity, variable climatic conditions, anthropogenic pressures, and the financial and logistical constraints associated with conventional water quality monitoring. This study, therefore, sort not merely to describe the quality of the two water resources, but to develop an integrated understanding of the factors influencing water quality and to translate this understanding into a predictive assessment capability. This chapter synthesis is meant to demonstrate the interconnectedness of the four substantive chapters of the study, Chapter 2: which reviewed existing water quality index models (WQI); Chapter 4: which investigated seasonal physicochemical and microbiological variability; Chapter 5: which examined bacterial community structure and its physicochemical drivers; and Chapter 6: which developed an artificial neural network (ANN)-based prediction tool. Although each chapter addressed a different research question, their findings converged around a common proposition that water quality in freshwater systems cannot be adequately represented by a static or a generic assessment approach. Rather, meaningful water quality assessment requires a consideration of the ecosystem type, temporal variability, physicochemical conditions. and microbiological responses, and the capacity to anticipate future changes.

Chapter two involved the synthesis of literature to appraise existing water quality indices towards the possibility of applying one WQI model to assess both lotic and lentic systems. Water quality monitoring being central to the continued use of WQI models, chapter 2 reported that despite the wide use and adoption of WQI models throughout the world, these models presented some shortcomings relating to uncertainty, ambiguity and eclipsing, lack of a universal WQI model, and the “metaphoring” classification schemes. Statistical methods such as factor analysis, cluster analysis, and analytical hierarchy process have been introduced, but fell short because natural ecosystems tend to be too complex for these statistical methods. However, the introduction of machine learning classifier techniques such as extreme gradient boosting, support vector machine, Gaussian Naïve Bayes, random forest, and decision tree to correct these shortcomings has been a major development for the continued use of WQI models.

In chapter four, we explored the effect of climatic changes on the general water quality in the Crocodile River and Inyaka Dam. Temperature and rainfall were discovered to be the main contributors to variability in water quality across observed seasons. These highlighted threats of climate change in the areas where the water resources are situated with further implications for the sustainability of agricultural activities, recreation, and general livelihoods. Experimental measurements of twelve water quality parameters suggested that the seasonal climatic variability affects the two systems differently. This is the case because one is a lentic system and thus, a subject of stratification, while the other is a lotic system, and a subject of stream flow and run-off. As such, this chapter highlighted the need to discriminate between the two aquatic systems when designing water pollution mitigation strategies or policy frameworks. Moreover, the findings of this chapter encourage the separation of water quality management strategies based on the challenges faced by each water resource. Furthermore, the findings suggest the need for advocacy to implement climate change-resilient water quality management policies to control water pollution and have a firm grasp on water security.

Chapter 5 investigated the influence of variations in physicochemical parameters on community structure and bacterial diversity in the Inyaka Dam and Crocodile River using 16S rRNA sequencing.

Relative abundance results revealed that the phylum *Proteobacteria* and class *Gammaproteobacteria* are dominant in both sites across all observed seasons. Moreover, Alpha diversity revealed significantly higher bacterial richness and diversity in the Crocodile River compared to the Inyaka Dam, due to nutrient dynamics such as Mn, Fe, PO_4^{3-} , and electrical conductivity. The Inyaka Dam maintained a homogeneous water column with compliant physicochemical concentrations accounting for low species abundance. Moreover, Beta diversity analysis demonstrated distinct microbial community assemblages between the two sites, thus indicating strong environmental filtering. This data highlights the significance of hydrological ecosystem context in structuring freshwater microbiomes. It further underscores the critical role of bacteria as sensitive indicators of environmental variations. As such, the present these findings support the notion of developing a microbial metrics system for ecosystem monitoring and management.

Chapter 6 provides a final distillation of the study as articulated in the pilot chapter. The aim was to develop a water quality assessment tool using machine learning techniques. The tool was coded, developed, tested, and launched on the Python programming platform. The

architecture of the tool takes the shape of a fully connected feedforward multi-layer perceptron that uses the backpropagation algorithm to forecast specified water quality indicators. Although with some limitations, the tool has demonstrated the capacity and reliability to predict both known and future datasets. Moreover, the tool is designed to predict one or more parameters at a time. However, every increase in the number of output parameters causes accuracy problems because the exercise requires a lot of input data. As such, the perfect functioning of the tool requires enormous data inputs. Notwithstanding, the tool produced reliable models for all chemical parameters, pH, turbidity, and temperature. However, it could not produce any reliable models for all microbiological parameters when trained with known input data. This is attributed primarily to model underfitting because of insufficient data. However, the testing phase of these failed models produced promising results. This warranted the need for additional data generation to enhance the predictive efficacy of all the failed models.

Combined, these chapters contribute to the overall aim of developing a more responsive and context-sensitive water quality assessment framework.

7.2.1. Original Contribution to Knowledge

The thesis makes an original contribution to knowledge by developing and empirically demonstrating an integrated framework for water quality assessment that moves beyond the limitations of WQIs. This it does by linking conventional water quality measurements and statistical characterisation, microbial community dynamics analysis and ANN based prediction of water quality. The following section breaks down the substantive parts of the contribution:

i. A multi-dimensional water quality assessment perspective:

The study has demonstrated that water quality in complex aquatic systems cannot be fully represented by a single aggregated index alone. This thesis has successfully integrated physicochemical, microbiological, and ecological information to provide a more comprehensive representation of water quality.

ii. A comprehensive characterization of contrasting aquatic ecosystems

By assessing lotic and lentic systems, the thesis has demonstrated that water quality behaviour is ecosystem specific. Seasonal variability, rainfall, anthropogenic activities, and internal processes influence the two systems differently. This provided empirical evidence for the significance of ecosystem context when interpreting and predicting water quality.

iii. Integration of environmental and microbial dimensions

The incorporation of 16S microbial community analysis extends beyond the conventional water quality assessment of chemical and microbiological indicator measurements. This demonstrates that environmental gradients are associated with changes in microbial community structure. This provides an ecological dimension to the interpretation of water quality deterioration and variability, and a new research direction to the field.

iv. Importance of data characteristics for ANN prediction

In the context of this study, the thesis has further demonstrated that ecological importance and predictive power are not equivalent. Parameters that are important indicators of water quality may, nonetheless, be difficult to predict when the available observations are limited, highly variable for machine learning, or insufficiently representative. This provides even an important methodological qualification for applying ANN approaches in water quality monitoring.

Collectively, these findings contribute to a shift from the predominantly retrospective water quality assessment towards an integrated and potentially predictive assessment framework. This framework was demonstrated to link the measurements of water quality with the identification of controlling relationships, ecological interpretation, and prediction to provide grounds for more informed and proactive decision making. As such, the principal contribution is therefore not the application of the ANN or indicator analysis in isolation, but their integration into a coherent assessment framework.

7.3. Conclusions and Recommendations

Specific objective 1: To investigate the microbial and physicochemical water quality characteristics of the Crocodile River and Inyaka Dam across all four seasons.

H₀: Seasonal weather variations have no impact on the water quality of surface water.

H_a: Seasonal weather variations have an impact on the water quality of surface water.

Research question: How does water quality vary across seasons, and what are the main influencers?

Twelve water quality parameters were considered based on the risk profile of the study area and the gazetted resource quality objectives. The experimental measurements of seasonal data suggested that there is great variability of parameter measurements, mainly owing to rainfall

and temperature. Despite observations of variability across the seasons, the Inyaka Dam indicators of water quality were always within the stipulated requirements. The process of stratification seems to play a much significant role in the concentration of parameters in the Inyaka Dam. For example, Manganese recorded the highest concentrations in winter, when rainfall measurements were lowest. This was because of temperature equalisation between the top water layer and the bottom layers, leading to mixing and subsequent sediment disturbance. The preceding trend manifests differently in the Crocodile River, as high rainfall and runoff determine the concentration of water quality parameters. As such, this study rejects the null hypothesis which claims that seasonal weather variations have no impact on the water quality of surface water. To answer the research question, this study found that there is great variability in the measurements of water quality parameters for both aquatic systems. While water quality is influenced by different aspects, rainfall and temperature affect both ecosystems.

Recommendations and future work

Over and above this study objective, the findings have advanced our understanding of the impact of climate change on water quality and security. As such, the study found it imperative that any policy that seeks to advance better water quality management should avoid the use of similar strategies or policies when mitigating pollution or when developing climate change resilience frameworks in these distinct water resources. Regarding the revealed pollution influencers, it is recommended that more resources be directed towards contingency plans for dealing with agricultural run-off into rivers while also preparing for the treatment of increased chemical concentrations at the water treatment plants. Future research in these water resources should include parameters such as BOD5 and COD due to the proximity of these water resources to wastewater treatment plants.

Specific objective 2: To investigate the bacterial community structure and the dominant bacterial taxa in the Crocodile River and Inyaka Dam across seasons.

H₀: The bacterial community structure and dominant bacterial taxa would not be uniform in different seasons.

H_a: The bacterial community structure and dominant bacterial taxa would be uniform in different seasons.

Research Question: What are the effects of significantly variable water quality on freshwater ecosystems and ecological function?

This study revealed clear and significant differences in bacterial compositions between the lotic (Crocodile River) and lentic (Inyaka Dam) systems. Key physicochemical drivers in the lotic system, such as flow rate, nutrient concentrations of PO_4^{3-} , $\text{NO}_3\text{-N}$, Mn, Fe, and EC, were influencers of microbial distribution and diversity, irrespective of season. This favoured a more diverse and spatially variable microbial assemblage. The lentic system, however, supported a more homogeneous community structure across all seasons, largely due to water column stability and nutrient stratification. Despite the observed dissimilarities between the systems, the study revealed that the phylum *Proteobacteria* dominated both systems across all observed seasons. Moreover, *Gammaproteobacteria* were more abundant across all seasons in both ecosystems, suggesting the ubiquity and adaptability of *Proteobacteria*. This leads the study to reject the null hypothesis which states that the bacterial community structure and dominant bacterial taxa would not be uniform in different seasons. To answer the research question, “what are the effects of significantly variable water quality on freshwater ecosystems and ecological function?”, the dominance of one phylum *Proteobacteria* across seasons, demonstrates that there was not much water quality difference to affect the normal functioning of freshwater ecosystems. Moreover, this could suggest that *Proteobacteria* have a broad tolerance range.

Recommendations and future work

These findings underscore the importance of hydrological context in shaping bacterial communities and highlight the need for ecosystem-specific monitoring frameworks. In addition, it enhances our comprehension of microbial biogeography in freshwater habitats and offers valuable insights for biodiversity conservation, water quality assessment, and ecological restoration in freshwater ecosystems. Furthermore, the findings of this study support the development of an ecosystem-specific microbial metrics system for water quality management. However, to advance progress on this subject, future work should be prioritised towards incorporating temporal sampling and functional metagenomics to unravel pathways through which environmental gradients mediate microbial communities.

Specific objective 3: To develop an inclusive water quality assessment model for the Crocodile River and the Inyaka Dam.

Research question: Would an artificial intelligence (AI)-based water quality prediction model help to identify trends in water quality to obtain necessary information on pollution sources?

A mix of historical and present data from thirteen parameters, including pH, turbidity, temperature, electrical conductivity, rainfall, Manganese, Iron, phosphate, nitrate, *E. coli*, total coliforms, Bifidobacteria, and faecal streptococci. We developed a functional water quality forecasting tool that is propelled by artificial intelligence tools such as ANN. The reliability of the tool is backed by cost functions such as MSE, MSLE, MAPE, and MAE that evaluate the performance of the trained model. The tool was able to produce 62% reliable models from all trained parameters. It should be noted that the tool is data-dependent, and its accuracy improves with the continuous addition of data. Consequently, the tool can be used to predict pH, turbidity, temperature, Manganese, Iron, nitrate-nitrogen, and phosphate. All models trained for the prediction of biological parameters did not yield reliable models. To answer the pertinent question of the study, the ability of the model to predict multiple outputs at a time allows for the identification of potential pollution trends and therefore has the potential as an early warning and decision support mechanism.

Recommendations and future work

More work is needed to sharpen the tool for application in other aquatic systems. This proposed work should be directed towards acquiring more datasets, primarily for all the biological parameters under study. Moreover, there is a need for the revision of the current microbial monitoring framework to include parameters such as bifidobacteria and faecal streptococci, which should serve as early warning systems for faecal contamination in drinking water and ecosystem health in surface water.

Appendices

Appendix 1: University of South Africa Research Ethics Clearance Letter (2020/CAES_HREC/047)



UNISA-CAES HEALTH RESEARCH ETHICS COMMITTEE

Date: 04/07/2023

Dear Mr Mogane

NHREC Registration # : REC-170616-051
REC Reference # : 2020/CAES_HREC/047
Name : Mr LK Mogane
Student # : 61804649

**Decision: Ethics Approval
Confirmation after First Review
from 05/11/2020 to 31/10/2025**

Researcher(s): Mr LK Mogane
moganelk@gmail.com

Supervisor (s): Dr T Masebe
masebtm@unisa.ac.za; 011-471-2268

Dr EJ Ncube
encube@randwater.co.za; 016-430-8460

Prof TAM Msagati
msagatam@unisa.ac.za; 073-528-9143

Working title of research:

A water quality assessment model for the Crocodile River and the Inyaka Dam

Qualification: PhD Life Science

Thank you for the submission of your yearly progress report to the Unisa-CAES Health Research Ethics Committee for the above mentioned research. Ethics approval is confirmed to continue for the originally approved period, subject to submission of yearly progress reports. **Failure to submit the progress report will lead to withdrawal of the ethics clearance until the report has been submitted.**

The researcher is cautioned to adhere to the Unisa protocols for research during Covid-19.

Due date for next progress report: 30 June 2024

The progress report form can be downloaded from the college ethics webpage:



University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

*The **low risk application** was **reviewed** by the UNISA-CAES Health Research Ethics Committee on 05 November 2020 in compliance with the Unisa Policy on Research Ethics and the Standard Operating Procedure on Research Ethics Risk Assessment.*

The proposed research may now commence with the provisions that:

1. The researcher will ensure that the research project adheres to the relevant guidelines set out in the Unisa Covid-19 position statement on research ethics attached.
2. The researcher(s) will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.
3. Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study should be communicated in writing to the Committee.
4. The researcher(s) will conduct the study according to the methods and procedures set out in the approved application.
5. Any changes that can affect the study-related risks for the research participants, particularly in terms of assurances made with regards to the protection of participants' privacy and the confidentiality of the data, should be reported to the Committee in writing, accompanied by a progress report.
6. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study. Adherence to the following South African legislation is important, if applicable: Protection of Personal Information Act, no 4 of 2013; Children's act no 38 of 2005 and the National Health Act, no 61 of 2003.
7. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data require additional ethics clearance.
8. No field work activities may continue after the expiry date. Submission of a completed research ethics progress report will constitute an application for renewal of Ethics Research Committee approval.

Note:

*The reference number **2020/CAES_HREC/047** should be clearly indicated on all forms of communication with the intended research participants, as well as with the Committee.*

Yours sincerely,



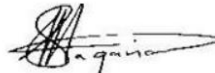
URERC 25.04.17 - Decision template (V2) - Approve

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za



Prof MA Antwi
Chair of UNISA-CAES Health REC

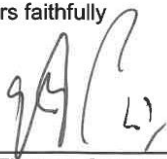
E-mail: antwima@unisa.ac.za
Tel: (011) 670-9391



Prof SR Magano
Acting Executive Dean : CAES

E-mail: magansr@unisa.ac.za
Tel: (011) 471-3649

Appendix 2: Gatekeeper letter from Inkomati-Usuthu Catchment Management Agency (IUCMA) to conduct research on catchments under IUCMA

		
Suite 801, 8th Floor The MAXSA Building 13 Streak Street Mbombela	Private Bag X11214 Mbombela 1200	Tel 013 753 9000 Fax 013 753 2786
		
Enquiries: Mr. Thabo Rasiuba Mobile: 078 451 0140 E-mail : rasiubat@iucma.co.za Reference No: N/A Date : 17 September 2019		
PO Box 19 469 The Village 1218 Attention: Katlego Mogane Dear Sir RE: REQUEST FOR WRITTEN CONSENT TO CONDUCT RESEARCH ON CATCHMENTS UNDER INKOMATI-USUTHU CMA Your request to undertake an academic research within Inkomati-Usuthu Catchment Management Agency (IUCMA) jurisdiction catchments dated 23 May 2019 refers. The IUCMA received your request to undertake an academic research within its jurisdiction area with University of South Africa (UNISA) as the academic institution. The IUCMA does not object to your request to conduct a study within its area of jurisdiction. Conducting a research in any catchments of the Inkomati- Usuthu Water Management Area will enrich the area with information which can be used to solve both present and future water challenges the Agency is facing, and it is for such reasons that an initiative like this is highly appreciated and welcomed. Should you need any information or have any enquiries, please do not hesitate to contact Mr. Thabo Rasiuba at the contact details indicated above. Yours faithfully  _____ Dr. Thomas Gyedu-Ababio Chief Executive Officer		
Ms TP Nyakane-Maluka (Chairperson) Mr MS Mthembu (Deputy Chairperson) Mr M Gangazhe Dr TM Kelly Adv MB Shabangu Ms LM Sikhakhane Mr PA Tshabangu Ms SD Wiggins Ms LC Zulu Dr TK Gyedu-Ababio (Ex-Officio)		

Appendix 3: Gatekeeper letter from Bushbuckridge Local Municipality to conduct research experiments in the municipality's premises



**BUSHBUCKRIDGE
Local Municipality**

Private Bag x 9308
Bushbuckridge
1280
R533 Graskop Road opp. Mapulaneng DLTC
Bushbuckridge
Cell: 081 24 35 710 E
04 00 24 35 710 S

Tel: 013 004 0291/92/95

Email: info@bushbuckridge.gov.za

Website: www.bushbuckridge.gov.za

Enquiries: Mnisi M.M
Ref No: 4/4/2

13 March 2020

Mr. L.K Mogane
P.O Box 19469
The Village
1218

Tel : 060 567 4556
Email : moganelk@gmail.com

Dear Mr Mogane

RE: PERMISSION TO CONDUCT RESEARCH

The above is in respect of your request to conduct research under the topic: **A microbial Assessment Model for the Crocodile River and Inyaka Dam.**

The municipality appreciates your choice in associating your research studies towards a microbial assessment model for the crocodile river and inyaka dam.

As a result, the municipality has no objection to the effect of your request. You are also advised that your interaction with the municipality shall be subjected to the policies as approved by the Council.

The municipality wishes you best of luck in your endeavor and the Skills Development Facilitator Ms. Mnisi M.M is the contact person 081 044 2030.

Kind regards,

**C.A Nkuna
MUNICIPAL MANAGER**

Appendix 4: Chapter 4 supplementary tables and figures



Fig. S4.1 Illegal waste disposal site near the Crocodile River in the City of Mbombela (Photo by LK Mogane,

The Maviljan wastewater treatment plant is currently not fully functional and only serves as a reservoir for domestic sewage. Frame A in Fig S2 shows the sewage stored in the ponds to settle the water. Frame B is supposed to be the treated effluent sampling point, while frame C shows an overflowing manhole of the untreated effluent which pours directly into the Inyaka Dam. Frame D shows the ongoing progress towards completion of the Maviljan wastewater treatment works. This risk highlights the significance of water quality monitoring in the Inyaka dam to aid in the mitigation of pollution strategies and provide interim alternative solutions

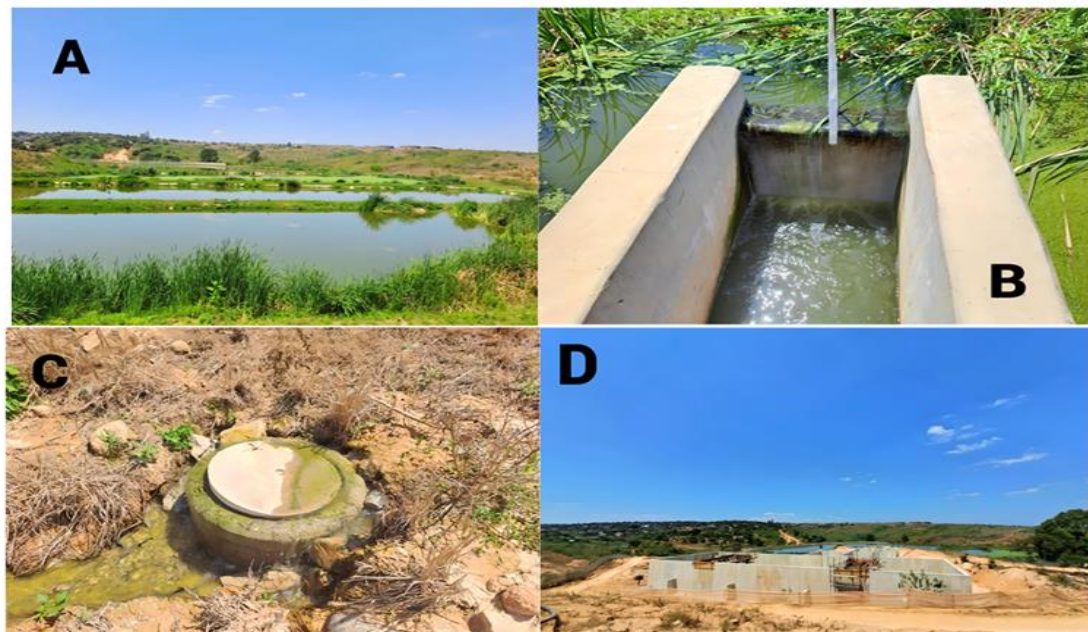
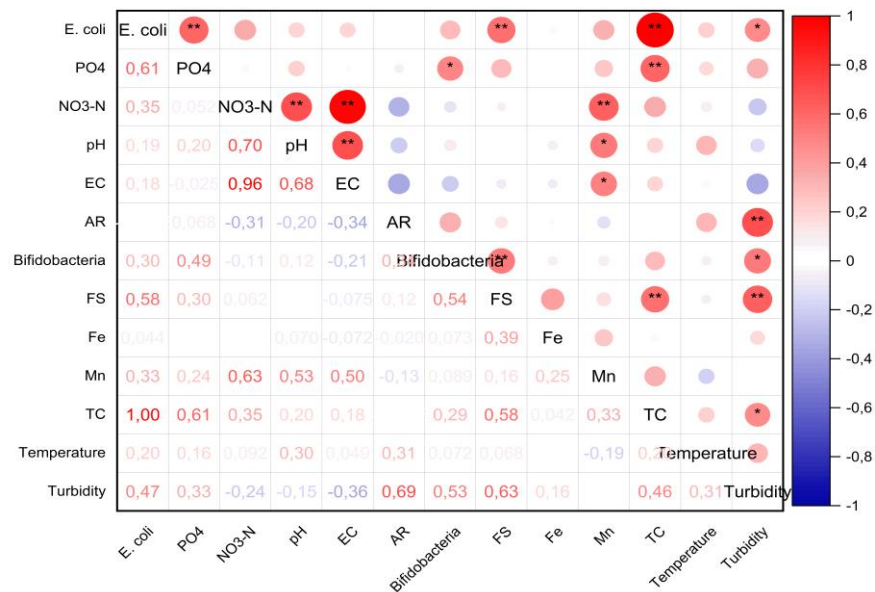
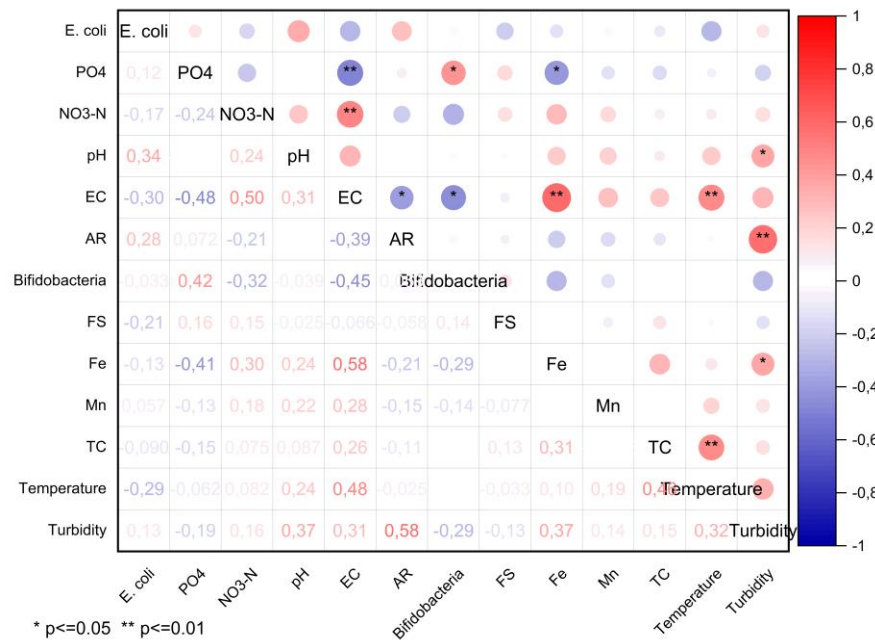


Fig. S4.2 The Maviljan wastewater treatment plant in Bushbuckridge, Mpumalanga, which discharges its effluent in the Inyaka Dam (Photo by LK Mogane, 2021).



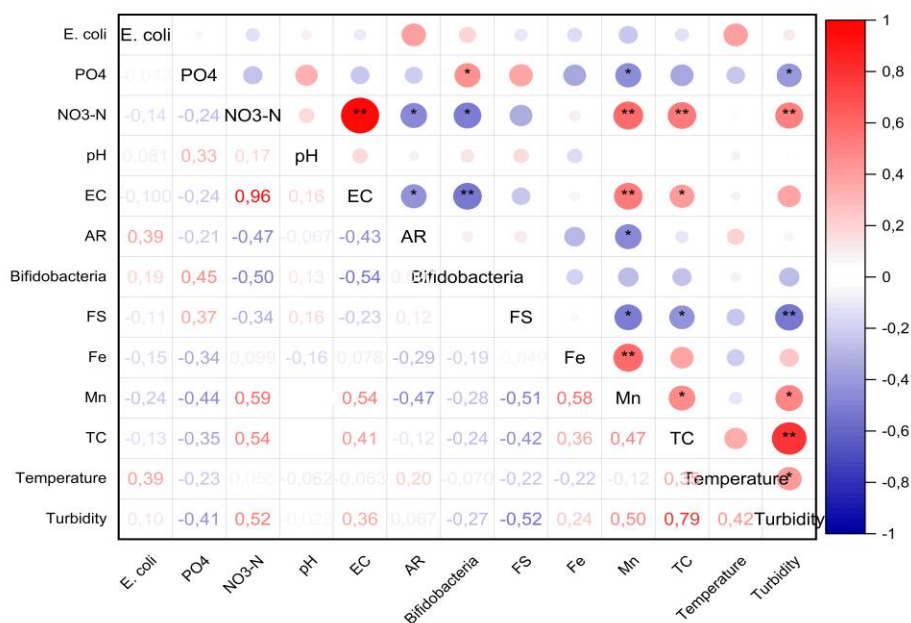
* p<=0.05 ** p<=0.01

Fig. S4.3 Correlation matrix for all measured parameters in the autumn season in the Inyaka Dam



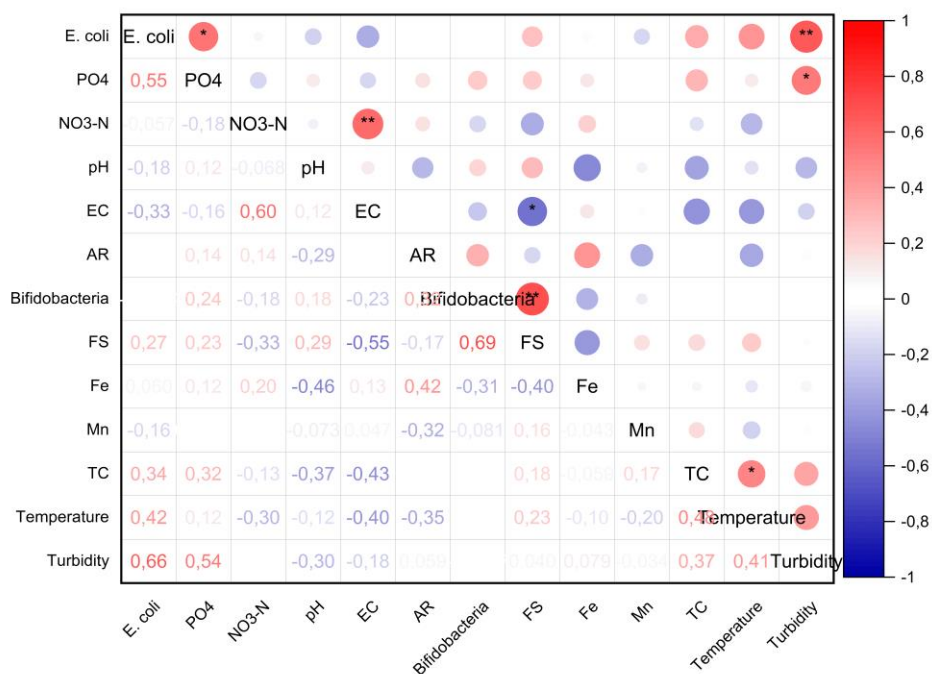
* p<=0.05 ** p<=0.01

Fig. S4.4 Correlation matrix for all measured parameters in the winter season in the Inyaka Dam



* p<=0.05 ** p<=0.01

Fig. S4.5 Correlation matrix for all measured parameters in the spring season in the Inyaka Dam



* p<=0.05 ** p<=0.01

Fig. S4.6 Correlation matrix for all measured parameters in the autumn season in the Crocodile River

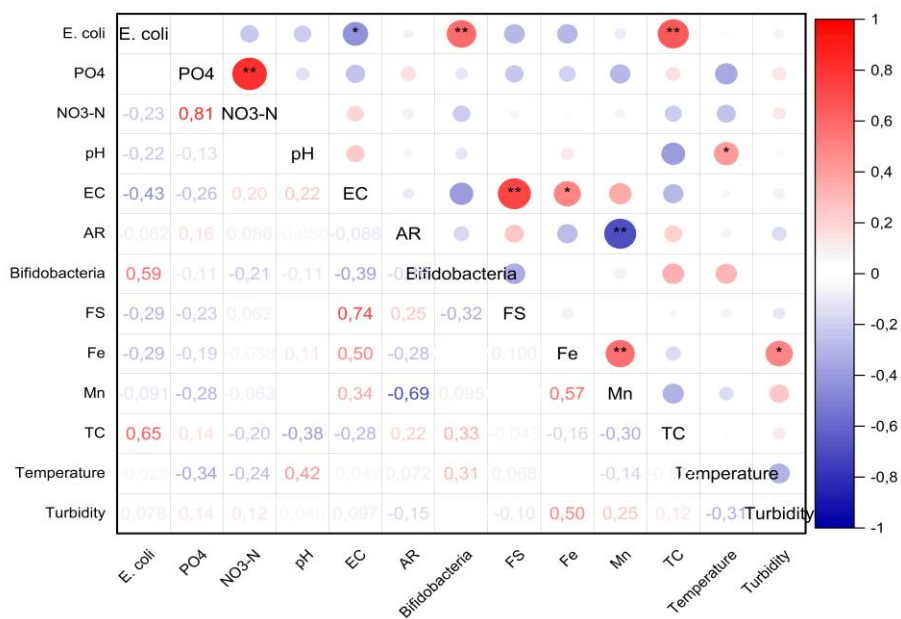


Fig. S4.7 Correlation matrix for all measured parameters in the spring season in the Crocodile River

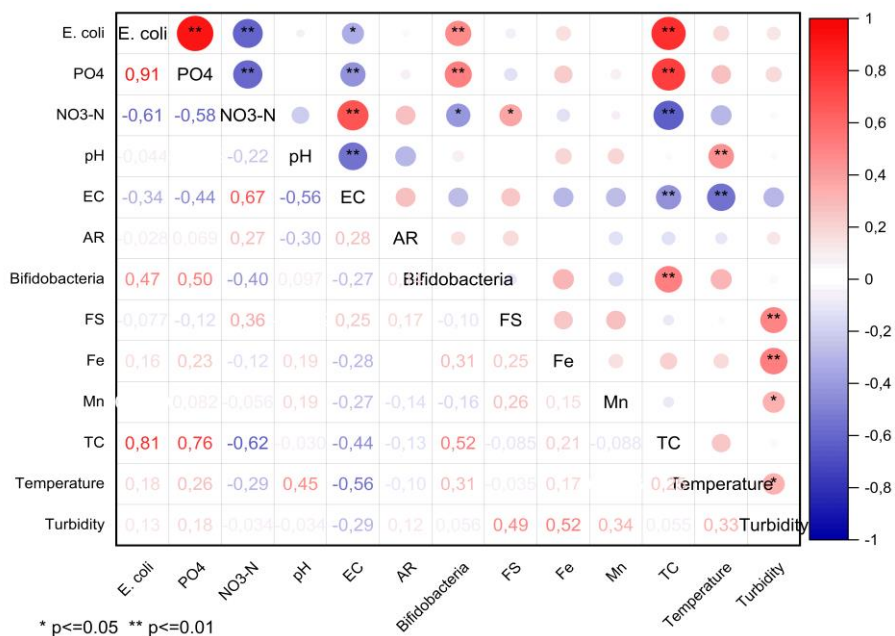


Fig. S4.8 Correlation matrix for all measured parameters in the winter season in the Crocodile River.

Table S4.1: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in summer

Parameters	Components		
	PC1	PC2	PC3
Average Rainfall (mm)	0.900	-0.069	-0.058
<i>E. coli</i>	0.844	-0.022	0.227
Turbidity	0.747	-0.029	-0.242
Faecal streptococci	0.651	-0.261	0.055
Bifidobacteria	0.566	-0.276	0.292
Total coliforms	0.536	0.030	-0.245
Fe	-0.086	0.902	-0.041
Temperature	0.020	-0.870	-0.030
NO ₃ -N	-0.215	0.736	0.570
Electrical conductivity	-0.265	0.626	0.609
PO ₄	0.102	0.028	0.748
Mn	0.020	-0.161	-0.026
pH	0.024	0.221	0.396
Initial eigenvalue	3.72	2.53	1.60
Percentage of variance (%)	28.59	19.53	12.28
Cumulative variance (%)	28.59	48.12	60.40

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.2: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in autumn

Parameters	Components		
	PC1	PC2	PC3
Total coliforms	0.934	0.219	0.022
<i>E. coli</i>	0.934	0.217	0.025
PO ₄	0.756	0.050	0.095
Faecal streptococci	0.629	-0.059	0.120
Bifidobacteria	0.491	-0.139	0.346
NO ₃ -N	0.147	0.938	-0.124
Electrical conductivity	-0.005	0.920	-0.149
pH	0.054	0.865	0.170
Mn	0.235	0.647	-0.204
Temperature	0.087	0.229	0.785
Aver. rainfall	0.033	-0.296	0.781
Turbidity	0.530	-0.285	0.615
Fe	-0.049	0.072	-0.010
Initial eigenvalue	4.00	3.36	1.37
Percentage of variance (%)	30.80	25.90	10.50
Cumulative variance (%)	30.80	56.60	67.20

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.3: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in winter

Parameters	Components		
	PC1	PC2	PC3
Bifidobacteria	-0.755	0.197	-0.064
PO ₄	-0.744	-0.080	0.001
Electrical conductivity	0.728	0.444	-0.232
Fe	0.690	0.247	0.078
NO ₃ -N	0.564	0.005	-0.188
Temperature	0.015	0.884	0.021
Total coliforms	0.082	0.718	0.056
Aver. rainfall	-0.175	-0.075	0.892
Turbidity	0.375	0.282	0.758
pH	0.096	0.124	0.218
Mn	0.107	0.104	-0.271
<i>E. coli</i>	-0.190	-0.443	0.408
Faecal streptococci	-0.101	0.062	-0.090
Initial eigenvalue	3.33	1.96	1.54
Percentage of variance (%)	25.60	15.05	11.84
Cumulative variance (%)	25.60	40.65	52.50

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.4: Correlation table of variables and main axes used to compute the PCA plots for the Inyaka Dam in spring

Parameters	Components		
	PC1	PC2	PC3
Turbidity	0.812	0.278	0.224
Total coliforms	0.777	0.288	0.018
Faecal streptococci	-0.770	0.020	-0.062
Mn	0.682	0.284	-0.521
Electrical conductivity	0.295	0.860	-0.192
NO ₃ -N	0.451	0.818	-0.167
Bifidobacteria	0.036	-0.816	0.038
Temperature	0.360	0.063	0.711
Aver. rainfall	-0.184	-0.235	0.696
<i>E. coli</i>	0.126	-0.163	0.667
Fe	0.398	-0.076	-0.576
pH	0.004	0.127	0.042
PO ₄	-0.390	-0.291	-0.113
Initial eigenvalue	4.39	2.24	1.59
Percentage of variance (%)	33.74	17.23	12.22
Cumulative variance (%)	33.74	50.98	63.21

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.5: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in summer

Parameters	Components (Summer)		
	PC1	PC2	PC3
EC	0.986	-0.113	-0.047
Turbidity	0.977	-0.133	-0.017
Fe	0.977	-0.092	-0.037
Mn	0.977	-0.112	-0.098
NO ₃ -N	0.900	0.094	-0.155
pH	-0.737	0.048	-0.450
FS	0.543	-0.438	-0.187
AR	0.527	-0.191	0.019
Temperature	0.358	-0.153	-0.012
Bifidobacteria	-0.096	0.938	-0.011
<i>E. coli</i>	-0.118	0.876	0.113
PO ₄	0.088	0.025	0.903
TC	-0.222	0.124	0.805
Initial eigenvalue	6.23	1.96	1.52
Variance explained (%)	39.90	10.80	7.50
Cumulative variance (%)	39.90	50.70	58.20

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.6: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in autumn

Parameters	Components (Autumn)		
	PC1	PC2	PC3
<i>E. coli</i>	0.778	0.045	0.178
PO ₄	0.542	0.193	0.397
NO ₃ -N	-0.260	-0.530	0.142
pH	-0.439	0.549	-0.104
Electrical conductivity	-0.544	-0.508	0.015
Average rainfall	-0.022	-0.196	0.890
Bifidobacteria	-0.033	0.716	0.519
Faecal streptococcus	0.210	0.859	0.052
Fe	0.136	-0.658	0.373
Mn	-0.028	0.008	-0.407
Total coliforms	0.719	0.046	-0.131
Temperature	0.682	0.181	-0.344
Turbidity	0.772	-0.128	0.151
Initial eigenvalue	6.23	1.96	1.52
Variance explained (%)	47.90	15.10	11.70
Cumulative variance (%)	47.90	63.00	74.70

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.7: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in winter

Parameters	Components (Winter)		
	PC1	PC2	PC3
<i>E. coli</i>	0.939	-0.040	0.063
PO ₄	0.917	0.055	0.126
Total coliforms	0.893	0.084	-0.007
NO ₃ -N	-0.738	-0.375	0.194
Bifidobacteria	0.564	0.294	0.065
pH	-0.110	0.820	-0.009
Temperature	0.171	0.767	0.164
Electrical conductivity	-0.444	-0.743	-0.070
Turbidity	0.106	0.155	0.854
Faecal streptococcus	-0.196	-0.198	0.766
Fe	0.190	0.305	0.627
Mn	0.016	0.037	0.465
Average rainfall	-0.056	-0.296	0.296
Initial eigenvalue	4.28	2.15	1.9
Variance explained (%)	32.90	16.60	14.70
Cumulative variance (%)	32.90	49.50	64.20

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation.

Table S4.8: Correlation table of variables and main axes used to compute the PCA plots for the Crocodile River in spring

Parameters	Components (Spring)		
	PC1	PC2	PC3
<i>E. coli</i>	-0.044	0.771	-0.357
PO ₄	-0.184	0.031	-0.216
NO ₃ -N	-0.024	-0.244	0.055
pH	0.106	-0.271	0.024
Electrical conductivity	0.348	-0.294	0.806
Average rainfall	-0.708	0.198	0.381
Bifidobacteria	0.146	0.600	-0.437
Faecal streptococcus	-0.108	-0.099	0.884
Fe	0.786	0.028	0.352
Mn	0.854	-0.203	-0.037
Total coliforms	-0.206	0.863	0.042
Temperature	-0.191	0.078	-0.016
Turbidity	0.612	0.351	0.144
Initial eigenvalue	3.14	2.40	2.10
Variance explained (%)	24.10	18.50	16.2
Cumulative variance (%)	24.10	42.60	58.80

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser Normalisation

Appendix 5: Chapter 6 Tables

Table S6.1: Model training specifications and training evaluation results for predicting turbidity.

Output Parameter	Training batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden layer nodes	Loss Function	Loss value	Acceptable range
EC	8	50	ReLU	Nadam	4	MSE	72961.91	<0.5
EC	8	50	ReLU	Nadam	4	MAPE	70.55%	<50%
EC	8	50	ReLU	Nadam	4	MSLE	5.55	<0.5
EC	8	50	ReLU	Nadam	4	MAE	512.6	<0.5
EC	8	50	Sigmoid	Nadam	4	MSE	6995.34	<0.5
EC	8	50	Sigmoid	Nadam	4	MAPE	70.17%	<50%
EC	8	50	Sigmoid	Nadam	4	MSLE	15.35	<0.5
EC	8	50	Sigmoid	Nadam	4	MAE	51.47	<0.5
EC	8	50	ReLU	Nadam	16	MSE	17511.35	<0.5
EC	8	50	ReLU	Nadam	16	MAPE	88.96%	<50%
EC	8	50	ReLU	Nadam	16	MSLE	15.35	<0.5
EC	8	50	ReLU	Nadam	16	MAE	82.43	<0.5
EC	8	50	Sigmoid	Nadam	16	MSE	6913.51	<0.5
EC	8	50	Sigmoid	Nadam	16	MAPE	70.37%	<50%
EC	8	50	Sigmoid	Nadam	16	MSLE	15.35	<0.5
EC	8	50	Sigmoid	Nadam	16	MAE	54.94	<0.5

Red colour = unreliable model

Table S6.2: Model training specifications and training evaluation results for predicting turbidity.

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden layer nodes	Loss Function	Loss Value	Acceptable range
Turb	8	50	ReLU	Nadam	4	MSE	8049.42	<0.5
Turb	8	50	ReLU	Nadam	4	MAPE	79.52%	<50%
Turb	8	50	ReLU	Nadam	4	MSLE	0.77	<0.5
Turb	8	50	ReLU	Nadam	4	MAE	8.62	<0.5
Turb	8	50	Sigmoid	Nadam	4	MSE	640.06	<0.5
Turb	8	50	Sigmoid	Nadam	4	MAPE	36.54%	<50%
Turb	8	50	Sigmoid	Nadam	4	MSLE	0.76	<0.5
Turb	8	50	Sigmoid	Nadam	4	MAE	8.23	<0.5
Turb	8	50	ReLU	Nadam	16	MSE	829.68	<0.5
Turb	8	50	ReLU	Nadam	16	MAPE	131.24%	<50%
Turb	8	50	ReLU	Nadam	16	MSLE	0.89	<0.5
Turb	8	50	ReLU	Nadam	16	MAE	11.11	<0.5
Turb	8	50	Sigmoid	Nadam	16	MSE	592.1	<0.5
Turb	8	50	Sigmoid	Nadam	16	MAPE	37.61%	<50%
Turb	8	50	Sigmoid	Nadam	16	MSLE	0.56	<0.5
Turb	8	50	Sigmoid	Nadam	16	MAE	8.49	<0.5

Green colour = reliable model; Red colour = unreliable model

Table S6.3: Model training specifications and training evaluation results for predicting turbidity.

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
Temperature	8	50	ReLU	Nadam	4	MSE	1632.22	<0.5
Temperature	8	50	ReLU	Nadam	4	MAPE	111.11%	<50%
Temperature	8	50	ReLU	Nadam	4	MSLE	9.49	<0.5
Temperature	8	50	ReLU	Nadam	4	MAE	3.05	<0.5
Temperature	8	50	Sigmoid	Nadam	4	MSE	8.43	<0.5
Temperature	8	50	Sigmoid	Nadam	4	MAPE	12.9%	<50%
Temperature	8	50	Sigmoid	Nadam	4	MSLE	0.02	<0.5
Temperature	8	50	Sigmoid	Nadam	4	MAE	2.15	<0.5
Temperature	8	50	ReLU	Nadam	16	MSE	172853.98	<0.5
Temperature	8	50	ReLU	Nadam	16	MAPE	8.65%	<50%
Temperature	8	50	ReLU	Nadam	16	MSLE	1.18	<0.5
Temperature	8	50	ReLU	Nadam	16	MAE	2.94	<0.5
Temperature	8	50	Sigmoid	Nadam	16	MSE	8.35	<0.5
Temperature	8	50	Sigmoid	Nadam	16	MAPE	13.0%	<50%
Temperature	8	50	Sigmoid	Nadam	16	MSLE	9.49	<0.5
Temperature	8	50	Sigmoid	Nadam	16	MAE	1.91	<0.5

Green colour = reliable model; Red colour = unreliable model

Table S6.4: Model training specifications and training evaluation results for Nitrate-Nitrogen prediction

Output Parameter	Training batch	Early Stopping patience	Activation Function	Optimiser Function	Hidden layer nodes	Loss Function	Loss value	Acceptable range
NO ₃ -N	8	50	ReLU	Nadam	4	MSE	9.11	<0.5
NO ₃ -N	8	50	ReLU	Nadam	4	MAPE	10194%	<50%
NO ₃ -N	8	50	ReLU	Nadam	4	MSLE	1.1	<0.5
NO ₃ -N	8	50	ReLU	Nadam	4	MAE	534.05	<0.5
NO ₃ -N	8	50	Sigmoid	Nadam	4	MSE	7.41	<0.5
NO ₃ -N	8	50	Sigmoid	Nadam	4	MAPE	73.54%	<50%
NO ₃ -N	8	50	Sigmoid	Nadam	4	MSLE	0.23	<0.5
NO ₃ -N	8	50	Sigmoid	Nadam	4	MAE	1.05	<0.5
NO ₃ -N	8	50	ReLU	Nadam	16	MSE	17194046	<0.5
NO ₃ -N	8	50	ReLU	Nadam	16	MAPE	72.3%	<50%
NO ₃ -N	8	50	ReLU	Nadam	16	MSLE	1.53	<0.5
NO ₃ -N	8	50	ReLU	Nadam	16	MAE	1.28	<0.5
NO ₃ -N	8	50	Sigmoid	Nadam	16	MSE	7.21	<0.5
NO ₃ -N	8	50	Sigmoid	Nadam	16	MAPE	69.27%	<50%
NO ₃ -N	8	50	Sigmoid	Nadam	16	MSLE	0.21	<0.5
NO ₃ -N	8	50	Sigmoid	Nadam	16	MAE	0.95	<0.5

Green colour = reliable model; Red colour = unreliable model

Table S6.5: Model training specifications and training evaluation results for Iron (Fe) prediction

Output Parameter	Training batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
Fe	8	50	ReLU	Nadam	4	MSE	1.49	<0.5
Fe	8	50	ReLU	Nadam	4	MAPE	56.82%	<50%
Fe	8	50	ReLU	Nadam	4	MSLE	0.45	<0.5
Fe	8	50	ReLU	Nadam	4	MAE	0.47	<0.5
Fe	8	50	Sigmoid	Nadam	4	MSE	1.17	<0.5
Fe	8	50	Sigmoid	Nadam	4	MAPE	48.63%	<50%
Fe	8	50	Sigmoid	Nadam	4	MSLE	0.11	<0.5
Fe	8	50	Sigmoid	Nadam	4	MAE	0.47	<0.5
Fe	8	50	ReLU	Nadam	16	MSE	1314107.88	<0.5
Fe	8	50	ReLU	Nadam	16	MAPE	53.87%	<50%
Fe	8	50	ReLU	Nadam	16	MSLE	0.24	<0.5
Fe	8	50	ReLU	Nadam	16	MAE	72.09	<0.5
Fe	8	50	Sigmoid	Nadam	16	MSE	1.16	<0.5
Fe	8	50	Sigmoid	Nadam	16	MAPE	48.25%	<50%
Fe	8	50	Sigmoid	Nadam	16	MSLE	0.11	<0.5
Fe	8	50	Sigmoid	Nadam	16	MAE	0.45	<0.5

Green colour = reliable model; Red colour = unreliable model

Table S6.6: Training specifications of models and results of training evaluation for manganese (Mn) prediction

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
Mn	8	50	ReLU	Nadam	4	MSE	0.12	<0.5
Mn	8	50	ReLU	Nadam	4	MAPE	61.74%	<50%
Mn	8	50	ReLU	Nadam	4	MSLE	0.15	<0.5
Mn	8	50	ReLU	Nadam	4	MAE	0.13	<0.5
Mn	8	50	Sigmoid	Nadam	4	MSE	0.39	<0.5
Mn	8	50	Sigmoid	Nadam	4	MAPE	62.05%	<50%
Mn	8	50	Sigmoid	Nadam	4	MSLE	0.06	<0.5
Mn	8	50	Sigmoid	Nadam	4	MAE	0.13	<0.5
Mn	8	50	ReLU	Nadam	16	MSE	1,4863x10 ¹²	<0.5
Mn	8	50	ReLU	Nadam	16	MAPE	80.33%	<50%
Mn	8	50	ReLU	Nadam	16	MSLE	21.32	<0.5
Mn	8	50	ReLU	Nadam	16	MAE	326924.28	<0.5
Mn	8	50	Sigmoid	Nadam	16	MSE	1,48592x10 ¹²	<0.5
Mn	8	50	Sigmoid	Nadam	16	MAPE	69.74%	<50%
Mn	8	50	Sigmoid	Nadam	16	MSLE	89.74	<0.5
Mn	8	50	Sigmoid	Nadam	16	MAE	326356.19	<0.5

Green colour = reliable model; Red colour = unreliable model

Table S6.7: Model training specifications and evaluation results for Total Coliforms (TC) prediction

Output Parameter	Training batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
TC	8	50	ReLU	Nadam	4	MSE	9.83x10 ¹¹	<0.5
TC	8	50	ReLU	Nadam	4	MAPE	73.84%	<50%
TC	8	50	ReLU	Nadam	4	MSLE	3.19	<0.5
TC	8	50	ReLU	Nadam	4	MAE	256153.36	<0.5
TC	8	50	Sigmoid	Nadam	4	MSE	1.49x10 ¹²	<0.5
TC	8	50	Sigmoid	Nadam	4	MAPE	78.06%	<50%
TC	8	50	Sigmoid	Nadam	4	MSLE	22.43	<0.5
TC	8	50	Sigmoid	Nadam	4	MAE	326968.25	<0.5
TC	8	50	ReLU	Nadam	16	MSE	9.22x10 ¹¹	<0.5
TC	8	50	ReLU	Nadam	16	MAPE	101.8%	<50%
TC	8	50	ReLU	Nadam	16	MSLE	2.15	<0.5
TC	8	50	ReLU	Nadam	16	MAE	263617	<0.5
TC	8	50	Sigmoid	Nadam	16	MSE	1.49x10 ¹²	<0.5
TC	8	50	Sigmoid	Nadam	16	MAPE	71.64%	<50%
TC	8	50	Sigmoid	Nadam	16	MSLE	16.54	<0.5
TC	8	50	Sigmoid	Nadam	16	MAE	326316.53	<0.5

Red colour = unreliable model

Table S6.8: Model training specifications and training evaluation results for Bifidobacteria prediction

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
Bifidobacteria	8	50	ReLU	Nadam	4	MSE	5835004	<0.5
Bifidobacteria	8	50	ReLU	Nadam	4	MAPE	473654.38%	<50%
Bifidobacteria	8	50	ReLU	Nadam	4	MSLE	6.3	<0.5
Bifidobacteria	8	50	ReLU	Nadam	4	MAE	366.44	<0.5
Bifidobacteria	8	50	ReLU	Nadam	16	MSE	470373.94	<0.5
Bifidobacteria	8	50	ReLU	Nadam	16	MAPE	299259008%	<50%
Bifidobacteria	8	50	ReLU	Nadam	16	MSLE	6.77	<0.5
Bifidobacteria	8	50	ReLU	Nadam	16	MAE	370.58	<0.5
Bifidobacteria	8	50	Sigmoid	Nadam	4	MSE	5934472.5	<0.5
Bifidobacteria	8	50	Sigmoid	Nadam	4	MAPE	242111.91%	<50%
Bifidobacteria	8	50	Sigmoid	Nadam	4	MSLE	6.79	<0.5
Bifidobacteria	8	50	Sigmoid	Nadam	4	MAE	366.41	<0.5
Bifidobacteria	8	50	Sigmoid	Nadam	16	MSE	5934587	<0.5
Bifidobacteria	8	50	Sigmoid	Nadam	16	MAPE	2217247.5%	<50%
Bifidobacteria	8	50	Sigmoid	Nadam	16	MSLE	4.78	<0.5
Bifidobacteria	8	50	Sigmoid	Nadam	16	MAE	366.45	<0.5

Red colour = unreliable model

Table S6.9: Training specifications of the model and training evaluation results for the prediction of faecal streptococci (FS).

Output Parameter	Training Batch	Early Stopping Patience	Activation Function	Optimiser Function	Hidden Layer Nodes	Loss Function	Loss Value	Acceptable range
FS	8	50	ReLU	Nadam	4	MSE	163.27	<0.5
FS	8	50	ReLU	Nadam	4	MAPE	16322.12%	<50%
FS	8	50	ReLU	Nadam	4	MSLE	1.7	<0.5
FS	8	50	ReLU	Nadam	4	MAE	6.78	<0.5
FS	8	50	Sigmoid	Nadam	4	MSE	126.76	<0.5
FS	8	50	Sigmoid	Nadam	4	MAPE	69246.75%	<50%
FS	8	50	Sigmoid	Nadam	4	MSLE	1.17	<0.5
FS	8	50	Sigmoid	Nadam	4	MAE	5.92	<0.5
FS	8	50	ReLU	Nadam	16	MSE	135.17	<0.5
FS	8	50	ReLU	Nadam	16	MAPE	7444286976%	<50%
FS	8	50	ReLU	Nadam	16	MSLE	1.52	<0.5
FS	8	50	ReLU	Nadam	16	MAE	6.52	<0.5
FS	8	50	Sigmoid	Nadam	16	MSE	126.58	<0.5
FS	8	50	Sigmoid	Nadam	16	MAPE	265242.69%	<50%
FS	8	50	Sigmoid	Nadam	16	MSLE	3.44	<0.5
FS	8	50	Sigmoid	Nadam	16	MAE	6.29	<0.5

Red colour = unreliable model