IDENTIFYING POLLUTION SOURCES IN THE KELANI RIVER BASIN, SRI LANKA: AN INTEGRATION OF MACHINE LEARNING AND PROCESS-BASED MODELLING APPROACHES

Prasan Nalintha Wijayaweera

238046N

Degree of Master of Science

Department of Civil Engineering

Faculty of Engineering

University of Moratuwa
Sri Lanka

December 2024

IDENTIFYING POLLUTION SOURCES IN THE KELANI RIVER BASIN, SRI LANKA: AN INTEGRATION OF MACHINE LEARNING AND PROCESS-BASED MODELLING APPROACHES

Prasan Nalintha Wijayaweera

238046N

Supervised by

Dr H. G. L. N. Gunawardhana

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science

Department of Civil Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

December 2024

DECLARATION OF THE CANDIDATE AND SUPERVISOR

"I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text".

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

23/12/2024

Prasan Nalintha Wijayaweera Date

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science-Major Component of Research

The above candidate has carried out research for the Master's Thesis/Dissertation under my supervision.

23-12-2024

Dr H. G. L. N. Gunawardhana Date

IDENTIFYING POLLUTION SOURCES IN THE KELANI RIVER BASIN, SRI LANKA: AN INTEGRATION OF MACHINE LEARNING AND PROCESS-BASED MODELLING APPROACHES

ABSTRACT

Water pollution poses a significant challenge globally, especially in developing nations like Sri Lanka. The Central Environmental Authority has identified the Kelani River as the most polluted river in the country. This study utilized 17 water quality parameters varying across 17 water quality monitoring stations along the Kelani River for the 2016–2020 period to assess the pollution level and identify the pollution sources in the Kelani River Basin (KRB). Initially, unsupervised machine learning methods were used to identify spatial and seasonal water quality variations. Factor analysis was used to generate Latent Variables (LV) considering the total variance in the dataset. Then two pollution indices were developed: the Kelani River Basin-Industrial Pollution Index (KRB-IPI) and the Kelani River Basin-Sewage Pollution Index (KRB-SPI) to quantify the pollution and to identify pollution sources in the basin. The KRB-IPI and KRB-SPI were validated through a Long Short-Term Memory Artificial Neural Network (LSTM ANN). Finally, a pollutant transport model was developed to assess the impact of these sources on the main river by simulating the Iron (Fe) transport in the Lower Kelani River Basin (LKRB). The development of the pollutant transport model began with the creation of an HEC-HMS model to generate sub-catchment hydrographs, which were calibrated (2016–2017) and validated (2018–2020) against observed discharge at Hanwella and Nagalagam Street streamflow stations. Hybrid modelling techniques further reduced errors in the generated hydrographs. The error-corrected hydrographs and observed streamflow were used to develop an HEC-RAS model for the Kelani River, which was calibrated and validated with observed data. The flow, velocity, and water depth parameters from the HEC-RAS model were used to develop a Water Quality Analysis Simulation Program (WASP) model, calibrated and validated with measured Fe records.

The FA generated five LVs, accounting for 77% of the total variance in the dataset. The spatial variation analysis highlighted that the stations near industrial zones and urbanized areas showed higher water quality variations. Seasonal variation analysis revealed strong relationships between water quality and monsoonal patterns. The KRB-IPI identified Cadmium, Iron (Fe), and Zinc as the main industrial pollutants, while the KRB-SPI identified Dissolved Oxygen, Total Coliform, and Chemical Oxygen Demand as the main parameters influencing sewage pollution. The developed KRB-IPI and KRB-SPI were validated through an LSTM ANN model with a Nash–Sutcliffe Efficiency (NSE) value of 0.98 and a Root Mean Squared Error (RMSE) value of 0.81 for KRB-IPI and an NSE value of 0.99 and an RMSE value of 1.83 for KRB-SPI. The KRB-IPI identified Kollonnawa Ela, Raggahawatte Ela, Maha Ela, Pusseli Oya, and Pugoda Ela as potential industrial pollution sources. Similarly, the KRB-SPI identified Victoria, Kollonnawa, Raggahawatte Ela, Maha Ela, Pugoda Ela, and Seethawake as potential sewage pollution sources. The WASP model was calibrated and validated with observed Fe records at Victoria station (Root squared error (R^2) = 0.893 in calibration and R^2 = 0.768 in validation) and Hanwella station (R^2 = 0.851 in calibration and R^2 = 0.757 in validation). The pollutant transport model identified Pugoda Ela station as having the highest impact on the water pollution of the Kelani River with a Mean Absolute Percentage Error (MAPE) of 16.21%. Maha Ela and Pusseli Oya were also confirmed as pollutant sources with MAPE values of 10.09% and 8.27%, respectively. The results suggest that the proposed method can effectively identify and quantify pollution in source catchments, offering a scalable methodology for other river basins to ensure sustainable water resource management.

Keywords: HEC-HMS, HEC-RAS, Hybrid Modelling, WASP, Water Pollution

ACKNOWLEDGEMENT

I would like to take the opportunity to thank my research supervisor, Dr. H.G.L.N. Gunawardhana, whose unwavering support and guidance have been crucial to the success of this research. His patience, motivation, and encouragement have played a vital role in completing this research, and without his dedicated supervision, this thesis would not have been possible.

I am grateful to Professor R.L.H.L. Rajapakse, the UMCSAWM Centre Chairman, for giving me the opportunity to pursue a Master of Science degree under the SAF-Madanjeet Singh Scholarship Scheme. His vision, kindness, and guidance were pivotal in my completion of this research.

I extend my thanks to the Late Shri Madanjeet Singh, the Founder of the SAF-Madanjeet Singh Scholarship Scheme, the South Asia Foundation (SAF), and the UNESCO Madanjeet Singh Centre for South Asia Water Management, University of Moratuwa for enabling me to pursue a master’s degree in civil engineering.

I would like to express my sincere appreciation to the Central Environmental Authority of Sri Lanka for providing valuable data which was the backbone of my research. I want to extend my appreciation to the Irrigation Department of Sri Lanka and the Department of Meteorology, Sri Lanka for providing data for the study.

I would also like to extend my heartfelt gratitude to Dr. S.P. Chaminda, Dr. Janaka Bamunawela and Dr. Jeewanthi Sirisena for the support and guidance provided during the research.

I express my sincere gratitude to the Head of the Department of Civil Engineering, University of Moratuwa and the staff for the support given throughout my research.

Finally, I would like to extend my gratitude to Mr. Wajira Kumarasinghe, Mr. Samantha Ranaweera, Mrs. K.A.V. Kalanika, Ms. L.J.N. Silva, and Mr. W.R. Wijeweera for their support and encouragement throughout my studies at the UMCSAWM.

TABLE OF CONTENT

1 INTRODUCTION **1**

1.1 GENERAL 1

1.2 PROBLEM STATEMENT 3

1.3 AIM 4

1.4 OBJECTIVES 4

1.5 SCOPE AND LIMITATIONS 4

1.6 SIGNIFICANCE OF THE RESEARCH 5

1.7 STRUCTURE OF THE THESIS/ DISSERTATION 5

2 LITERATURE REVIEW **6**

2.1 WATER QUALITY 6

2.1.1 Introduction to Water Quality 6

2.1.2 Water Pollution in Rivers 7

2.2 ASSESSMENT OF WATER QUALITY VARIATIONS 9

2.3 WATER QUALITY AND POLLUTION INDEXING 10

2.3.1 Application of Water Quality and Pollution Indices 10

2.3.2 Major Steps in Index Development 12

2.3.3 Applications of Machine Learning in Indexing 13

2.4 RIVER WATER QUALITY MODELLING 14

2.4.1 Water Quality Simulation Models 14

2.4.2 Applications of WASP in Water Quality Modelling 16

2.4.3 Hydrological Engineering Centre-River Analysis System (HEC-RAS) 17

2.4.4 Hydrological Engineering Centre-Hydrologic Modelling System (HEC-HMS) 18

2.4.5 Hybrid Modelling for Error Correction in Process-based Models 18

2.5 SUMMARY OF THE LITERATURE REVIEW 19

3 MATERIALS AND METHODS **21**

3.1 STUDY AREA 21

3.2 RESEARCH METHODOLOGY 23

3.3 DATA COLLECTION 26

3.3.1 Water Quality Data 26

3.3.2 Rainfall and Streamflow Data 27

3.3.3 Single Mass Curve 28

3.3.4 Double Mass Curve 28

3.3.5 Visual Checking 30

3.4 FACTOR ANALYSIS 32

3.5 CLUSTER ANALYSIS 33
3.5.1 Hierarchical Cluster Analysis 33
3.5.2 Self-Organizing Maps 34
3.6 WATER POLLUTION INDEXING 34
3.6.1 Parameter Selection and Sub-indexing 34
3.6.2 Parameter Weighting 35
3.6.3 Parameter Aggregation and Classification 36
3.7 RANDOM FORREST MODEL 37
3.8 RANK ORDER CENTROID (ROC) METHOD 38
3.9 LONG SHORT-TERM MEMORY (LSTM) ARTIFICIAL NEURAL NETWORK (ANN) 38
3.10 RAINFALL-RUNOFF MODELLING WITH HEC-HMS 38
3.11 1D RIVER MODELLING IN HEC-RAS 39
3.12 WASP MODEL DEVELOPMENT 40
3.13 OBJECTIVE FUNCTIONS 40
3.14 CENTRAL ENVIRONMENTAL AUTHORITY GUIDELINES 42

4 RESULTS 44
4.1 ANALYSING WATER QUALITY VARIATIONS 44
4.1.1 Factor Analysis and Latent Variables 44
4.1.2 Spatial Variations of Water Quality 46
4.1.3 Seasonal Variations of Water Quality 48
4.2 WATER POLLUTION INDEXING 50
4.2.1 Selection of Parameters and Sub-index Development 51
4.2.2 Parameter Weighting 52
4.2.3 Sub-index Aggregation 53
4.3 RIVER WATER QUALITY MODELLING 58
4.3.1 HEC-HMS Modelling 58
4.3.2 Error Correction Using Hybrid Modelling 64
4.3.3 HEC-RAS Modelling 67
4.3.4 WASP Modelling 68

5 DISCUSSION 72
5.1 SPATIAL AND SEASONAL WATER QUALITY VARIATIONS 72
5.2 WATER POLLUTION INDEXING 79
5.3 WATER QUALITY MODELLING 83

6 CONCLUSIONS 88

7 RECOMMENDATIONS 90

LIST OF FIGURES

Figure 3.1 Kelani River Basin and the locations of the rainfall and streamflow monitoring stations 22
Figure 3.2 Thiessen averaged monsoon rainfall variations in the Kelani River Basin during the 2016–2020 period. Average rainfall is shown in red, while the whiskers extending from the top and bottom represent the maximum and minimum rainfall recorded. Additionally, upper, middle, and lower lines within the box illustrate the quartile ranges of the 75th, 50th, and 25th percentiles of rainfall distribution, respectively (NEM-Northeast monsoon, IM1-First intermonsoon, SWM-Southwest monsoon, and IM2-Second intermonsoon)........ 22
Figure 3.3 Monthly average rainfall and streamflow at Hanwella and Glencourse streamflow monitoring stations in the Kelani River Basin for the period 2016–2020 23
Figure 3.4 Research Methodology 25
Figure 3.5 Locations of the water quality monitoring stations in the Kelani River Basin 26
Figure 3.6 Single mass curve for Colombo, Hanwella, and Dunedin rainfall stations 28
Figure 3.7 The double mass curve for the Colombo rainfall station 29
Figure 3.8 The double mass curve for the Hanwella rainfall station 29
Figure 3.9 Water level time series for Nagalagam Street monitoring station 30
Figure 3.10 Streamflow time series for Hanwella monitoring station 31
Figure 3.11 Streamflow time series for Glencourse monitoring station 31
Figure 4.1 Correlation coefficients identified in Pearson's Correlation test 45
Figure 4.2 Scree plot used to identify the ideal number of factors to be retained along with the cumulative variance 45
Figure 4.3 Different clusters identified by hierarchical cluster analysis for the five latent variables 47
Figure 4.4 Different clusters identified by self-organizing maps for the five latent variables 48
Figure 4.5 Variation in correlation coefficients between rainfall and Latent Variables (LV) during different monsoon seasons 49
Figure 4.6 Correlation of Latent Variables (LVs) with river discharge in different monsoon seasons 50
Figure 4.7 Sub-index functions for the Kelani River Basin-Industrial Pollution Index (KRB-IPI) 51
Figure 4.8 Sub-index functions for Kelani River Basin-Sewage Pollution Index (KRB-SPI). The concentrations for BOD, COD, DO, and Phosphate were measured in *mg/l* while the concentration of T. Coli was measured in *MPN/ 1000 ml* 52
Figure 4.9 Results of the sensitivity analysis for Fe 54
Figure 4.10 Results of the sensitivity analysis for COD 54
Figure 4.11 Performance of the LSTM ANN model for KRB-IPI. The distribution of the calculated and predicted index values are shown in bar charts 55
Figure 4.12 Performance of the LSTM ANN model for KRB-SPI. The distribution of the calculated and predicted index values are shown in bar charts 55
Figure 4.13 Classification of the KRB-IPI into five categories based on percentile values 56

Figure 4.14 Classification of the KRB-SPI into five categories based on percentile values ... 56
Figure 4.15 Seasonal and annual average water pollution scores for the 17 water quality monitoring stations in the KRB calculated from the KRB-IPI ... 57
Figure 4.16 Seasonal and annual average water pollution scores for the 17 water quality monitoring stations in the KRB calculated from the KRB-SPI ... 57
Figure 4.17 Spatial variation of Iron across the stations in the Lower Kelani River Basin (LKRB) ... 58
Figure 4.18 Generated sub-basins for the HEC-HMS modelling ... 59
Figure 4.19 Time series plot of simulated and observed streamflow for the calibration period at Hanwella ... 61
Figure 4.20 Time series plot of simulated and observed streamflow for the validation period at Hanwella ... 61
Figure 4.21 Performance of the HEC-HMS model for the calibration and validation periods at Hanwella ... 62
Figure 4.22 Generated hydrograph for Kollonnawa sub-catchment ... 63
Figure 4.23 Generated hydrograph for Pueesli Oya sub-catchment ... 63
Figure 4.24 Performance of the LSTM ANN model for the training and validation periods ... 65
Figure 4.25 Error corrected hydrograph for Kollonnawa sub-catchment ... 66
Figure 4.26 Error corrected hydrograph for Pusseli Oya sub-catchment ... 66
Figure 4.27 Performance of the developed HEC-RAS model for the calibration and validation periods ... 67
Figure 4.28 Distributions of segments used to develop the WASP model ... 68
Figure 4.29 Performance of the developed WASP model for the calibration and validation periods at Segment 58 (Victoria Station) ... 69
Figure 4.30 Performance of the developed WASP model for the calibration and validation periods at Segment 24 (Hanwella Station) ... 69
Figure 4.31 Time series plot for the calibration period at Segment 58 (Victoria Station) ... 70
Figure 4.32 Time series plot for the validation period at Segment 58 (Victoria Station) ... 70
Figure 4.33 Time series plot for the calibration period at Segment 24 (Hanwella Station) ... 71
Figure 4.34 Time series plot for the validation period at Segment 24 (Hanwella Station) ... 71
Figure 5.1 Chloride level variation along the Kelani River ... 74
Figure 5.2 The population density (persons per square kilometre) distribution in the Kelani River Basin is illustrated using data from the Department of Census and Statistics Sri Lanka for 2011 ... 75
Figure 5.3 Spatial distribution of industrial pollution sources in the Kelani River Basin ... 81
Figure 5.4 Spatial distribution of sewage pollution sources in the Kelani River Basin ... 82
Figure 5.5 The impact of Pugoda Ela (S11) station on the pollution of the Kelani River. Segment 9 and Segment 11 represent the segments before and after the tributary is joined to the main river respectively. Segment 10 represents the Pugoda Ela station located in the tributary ... 85

Figure 5.6 The impact of Maha Ela (S6) station on the pollution of the Kelani River. Segment 41 and Segment 43 represent the segments before and after the tributary is joined to the main river respectively. Segment 42 represents the Maha Ela station located in the tributary 86

Figure 5.7 The impact of Pusseli Oya (S7) station on the pollution of the Kelani River. Segment 26 and Segment 28 represent the segments before and after the tributary is joined to the main river respectively. Segment 27 represents the Pusseli Oya station located in the tributary 86

Figure 5.8 The impact of Raggahawatte Ela (S4) station on the pollution of the Kelani River. Segment 43 and Segment 45 represent the segments before and after the tributary is joined to the main river respectively. Segment 44 represents the Raggahawatte Ela station located in the tributary 87

Figure 5.9 The impact of Kollonnawa (S2) station on the pollution of the Kelani River. Segment 56 and Segment 58 represent the segments before and after the tributary is joined to the main river respectively. Segment 57 represents the Kollonnawa Ela station located in the tributary 87

LIST OF TABLES

TABLE 2.1 DIFFERENT CATEGORIES OF WATER QUALITY 7
TABLE 2.2 DIFFERENT TYPES OF POLLUTION IN RIVERS 8
TABLE 2.3 APPLICATIONS OF WATER QUALITY INDICES 10
TABLE 2.4 APPLICATIONS OF WATER POLLUTION INDICES 11
TABLE 2.5 APPLICATIONS OF MACHINE LEARNING IN INDEXING 13
TABLE 2.6 OVERVIEW OF WATER QUALITY MODELLING SOFTWARE 14
TABLE 2.7 APPLICATIONS OF WATER QUALITY ANALYSIS SIMULATION PROGRAM (WASP) 17
TABLE 3.1 MEASURED WATER QUALITY PARAMETERS 26
TABLE 3.2 DETAILS OF THE RAINFALL AND STREAMFLOW DATA COLLECTION 27
TABLE 3.3 SUB-INDEX (Q_I) VALUES FOR OBSERVATIONS LOWER AND HIGHER THAN THE STANDARD LIMIT FOR EACH QUANTILE 35
TABLE 3.4 CLASSIFICATION OF THE WATER POLLUTION INDICES 37
TABLE 3.5 SELECTED MATHAMATICAL MODELS FOR THE HEC-HMS MODEL DEVELOPMENT 39
TABLE 3.6 CENTRAL ENVIRONMENTAL AUTHORITY (CEA) GUIDELINES FOR AMBIENT WATER QUALITY IN INLAND SURFACE WATERWAYS IN SRI LANKA 42
TABLE 4.1 FACTOR LOADINGS FOR THE ROTATED LATENT VARIABLES IDENTIFIED BY THE FACTOR ANALYSIS 46
TABLE 4.2 CALCULATED PARAMETER IMPORTANCE SCORES AND THE PARAMETER WEIGHTS 53
TABLE 4.3 CALIBRATED PARAMETERS FOR THE FIVE SUB-CATCHMENTS WITH WATER QUALITY MONITORING STATIONS 60
TABLE 4.4 PERFORMANCE OF THE HEC-HMS MODEL 62
TABLE 4.5 RESULTS OF THE HYPERPARAMETER TUNING FOR THE LSTM ANN MODEL 65
TABLE 4.6 PERFORMANCE OF THE LSTM ANN MODEL DEVELOPED FOR THE MAIN RIVER 65
TABLE 4.7 PERFORMANCE OF THE HEC-RAS MODEL 67
TABLE 5.1 CALCULATED MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) VALUES FOR THE FIVE TRIBUTARIES IDENTIFIED AS POLLUTION SOURCES IN THE KRB BY KRB-IPI 84

LIST OF ABBREVIATIONS

CA – Cluster Analysis

CEA – Central Environmental Authority

FA – Factor Analysis

HCA – Hierarchical Cluster Analysis

HEC-HMS – Hydrologic Engineering Centre-Hydrologic Modelling System

HEC-RAS – Hydrologic Engineering Centre-River Analysis System

IM1 – First Intermonsoon

IM2 – Second Intermonsoon

KRB – Kelani River Basin

KRB-IPI – Kelani River Basin-Industrial Pollution Index

KRB-SPI – Kelani River Basin-Sewage Pollution Index

LKRB – Lower Kelani River Basin

MICE - Multiple Imputation by Chained Equations

NEM – Northeast Monsoon

SOM – Self-Organizing Maps

SWM – Southwest Monsoon

WASP – Water Quality Analysis Simulation Program

WHO – World Health Organization

CHAPTER 1

1 INTRODUCTION

1.1 General

Water quality refers to the chemical, physical, and biological characteristics of water that determine its suitability for various applications, including drinking, agriculture, and industrial processes (Adjovu et al., 2023; Wagle et al., 2020). It is an essential aspect when it comes to sustaining life and supporting the health of ecosystems. However, the growing pressures of climate change, urbanisation, and population growth have intensified challenges related to water quality and availability of clean water in source catchments. These factors result in pollution, habitat loss, and altered hydrological patterns that complicate the supply and quality of usable water for various purposes (Chunju, 2018; Kumar et al., 2022). The United Nations 2015 Development Agenda highlights the importance of water resources management in sustainable development. Sustainable Development Goal number six, 'Ensure availability and sustainable management of water and sanitation for all', focuses on addressing the critical need for access to clean water and adequate sanitation facilities worldwide, aiming to promote hygiene, reduce water scarcity, and enhance water resource management for present and future generations (Ait-Kadi, 2016).

However, water pollution in rivers remains a major concern in the present century (WHO, 2018). Rivers are increasingly contaminated by various pollutants including domestic and industrial sewage, agricultural runoff with fertilisers and pesticides, and urban stormwater (Wear et al., 2021). Alarmingly, around 80% of domestic, municipal, and industrial wastewater worldwide is discharged into water bodies without any prior treatment, contributing to the pollution of rivers and other freshwater systems (du Plessis, 2022). It has significant concerns for human health and negative impacts on ecosystem health and biodiversity. However, water quality in many rivers remains poorly monitored, potentially exposing more than 3 billion people globally to waterborne diseases (UN-Water, 2021). Addressing the issues related to pollution of river bodies is important to promote the sustainability of the water resource for use by aquatic organisms and humans.

There are two methods in the literature to assess issues related to water quality and pollution in rivers. Water quality assessment of rivers is usually performed to establish an overall idea of the status of a river (Lukhabi et al., 2023; Uddin et al., 2021). This involves analysis of the chemical, physical, and biological characteristics to determine the suitability of water for various uses such as drinking, irrigation, and recreational activities. On the other hand, assessing river pollution usually involves determining certain kinds of pollution and the identification of pollution sources (Karaouzas et al., 2021). This approach goes beyond general water quality assessments, examining the pollutants discharged into the water system through industrial effluent, agricultural drainages, or untreated domestic sewage (Prasad et al., 2014). Pollution source identification often employs advanced modelling techniques, to detect spatial contamination patterns and trace them back to their origins (Moghaddam et al., 2021; Fan et al., 2022). The objective is not only to understand the extent of pollution but also to identify the sources of contamination, allowing for more effective pollution control measures. Although both strategies are intended to preserve the rivers and the health of people, they have different objectives and methods. Water quality assessments give a general picture of the health of the river while pollution assessment gives a detailed picture of the pollutants and their contribution, giving an understanding of the contamination processes in the river. Both methods involve analysing data that vary across multiple dimensions, such as geographic locations and periods. This complexity arises from the need to interpret extensive datasets containing numerous variables measured at different spatial and temporal scales.

The traditional approaches for water quality assessment used in water quality-related studies include stochastic, empirical, and mechanistic methods for quantifying and forecasting water quality and pollution trends, with their strengths and weaknesses. Stochastic methods use probability distributions to explain the variability and randomness inherent in water quality data. They are therefore ideal for predicting future changes in water quality under various risk conditions. However, stochastic methods require large data sets, intricate mathematical formulas, and high computational resources, which makes them slow and computationally costly and, thus are sometimes impractical in many scenarios (Wagle et al., 2020). On the other hand, empirical methods are derived from actual data and the correlation between the observed water quality parameters (Lindenschmidt et al., 2007). These methods include water quality indices and regression models, which are easy to use as they directly use the measured data to classify water bodies or forecast trends. Empirical techniques are easier and less costly than stochastic techniques, but they do not account for the underlying processes driving water quality changes. However, they are easy to implement and use, especially when data is easily accessible (Aslam et al., 2022). Mechanistic methods on the other hand capture the physical, chemical, and biological behaviour that influences the quality of water (Sharma & Kaushal, 2021). These methods simulate the behaviour of pollutants in a water body and give a clear picture of the factors that affect water quality. Mechanistic models are exact when sufficient input data is available, but they are often complicated and computationally intensive (Ambrose et al., 2011).

The application of machine learning has improved the efficiency of these methods and provided a more profound understanding of the water quality data with the capability to analyse datasets with multiple variables (Abba et al., 2020; Badillo et al., 2020). The methods of regression, classification, pattern recognition, and clustering are commonly used in water quality research. Regression models are widely applied to predict water quality parameters based on the correlation between environmental variables and water quality variables (Fan et al., 2022). While classification algorithms can classify water bodies by quality classes based on parameter limits, they provide a rapid evaluation of water quality and pollution levels (Xu et al., 2022). Many pattern identification and clustering techniques are applied to identify the concealed trends or spatial and temporal clusters of pollution sources and to determine the pollution hot spots and changes in the water quality with different seasons or regions (Xia et al., 2018). One of the key advantages of machine learning is its flexibility in handling datasets with diverse units, quantities of observation, and measurement scales (Uddin et al., 2023; Wu et al., 2023). This makes it particularly useful when working with multiple datasets like hydrometeorological and water quality data for instance. Moreover, hybrid models, which combine machine learning algorithms with process-based models, can be used for water quality modelling (Konapala et al., 2020). These models leverage the strengths of both approaches: whereas process-based models are used to model the changes in water quality, machine learning can enhance the model accuracy by tuning the parameters and enhancing the prediction capability (Aslam et al., 2022). The combination of the machine learning methods with the traditional approaches in water quality-related research provides a strong foundation for evaluating the intricate water pollution problems, especially in situations where data is scarce or ambiguous.

1.2 Problem Statement

Rivers and other freshwater sources are gradually becoming threatened by pollution from industrial wastes, agriculture drainages, and domestic sewage (Akhtar et al., 2021). This pollution is a great danger to human health, aquatic life, and the general availability of fresh water in the world (Fida et al., 2023). This is especially a problem in developing countries where there are inadequate rules and regulations to prevent pollutants from flowing into rivers, lakes, and groundwater systems (Hasan et al., 2019). In Sri Lanka, water pollution has become a significant environmental issue, especially in densely populated and industrial areas (Central Environmental Authority, 2017). Rivers and lakes in the country are facing considerable pressure from both urban and rural activities. Domestic sewage and industrial effluents in urban areas are released without treatment into rivers, while chemical fertilizers and other pollutants in agricultural runoff from rural areas are also released into water bodies.

The Kelani River Basin (KRB) is considered the most polluted watershed in the country, as identified by the Central Environmental Authority (Central Environmental Authority, 2017). Industrial waste from two major industrial zones, Biyagama and Seethawake, as well as pollution from domestic sewage, has led to the contamination of this critical waterway, which serves as the primary water source for over 80% of the population of Colombo, which is close to a quarter of the population in Sri Lanka (Narangoda et al., 2023). The water quality of the river is deteriorating, posing risks not only to public health but also to biodiversity and economic activities reliant on the river, such as fisheries, irrigation, and transportation. The KRB is highly susceptible to pollution due to the intense rainfall patterns influenced by monsoon seasons, which lead to high river discharge and the spread of contaminants across the river system (Mahagamage & Manage, 2018). Given the critical role of the Kelani River in providing drinking water, supporting agriculture, and sustaining the local economy, addressing water pollution in the basin is a pressing issue. Current efforts to manage water quality in the KRB are hindered by challenges in accurately identifying pollution sources and assessing the severity of contamination. This study seeks to address these challenges by developing a comprehensive methodology that uses empirical and mechanistic approaches to assess pollution patterns and identify sources, ultimately supporting more effective water management strategies in the Kelani River Basin.

1.3 Aim

This study aims to evaluate the spatial and temporal variability of water quality in the Kelani River Basin (KRB) and develop a water quality index for the management of surface water resources in the future.

1.4 Objectives

The objectives of this research are,

- to examine the seasonal and spatial variation of the surface water quality in the Kelani River Basin,
- to develop a water quality index for efficient water resources management in the Kelani River Basin,
- to develop an integrated modelling tool for simulating river water quality, and
- to use machine learning techniques to improve the accuracy of the model simulations.

1.5 Scope and Limitations

This study focuses on assessing water pollution and identifying the sources of pollution in the Kelani River Basin. The scope includes empirical and mechanistic methods in conjunction with machine learning. The primary data requirements for the study are measured water quality data, streamflow data, river cross-section data, and rainfall data from the catchment. However, due to time and resource constraints, real-time data collection will not be feasible. Instead, the necessary data will be obtained from existing datasets from government monitoring agencies. The length and type of data and the used models will be selected based on recommendations from the literature to ensure the accuracy and reliability of the model outputs. The study is limited to assessing the present status of water quality and pollution sources in the Kelani River Basin. Future projections, such as the impacts of climate change on water quality and pollutant transport, were not considered. This approach was chosen to provide a focused assessment of current conditions and to address specific pollution-related issues in the basin. The pollutant transport model developed as part of this research is limited to one parameter due to time and resource constraints. The selected parameter was chosen based on the findings of previous studies, ensuring its relevance and significance to the observed pollution issues.

1.6 Significance of the Research

The identification of pollution sources in rivers of source catchments has become a critical area of research due to the increasing pressures from population growth and socio-economic activities. Over time, methods to identify pollution sources have evolved from trial-and-error techniques to more sophisticated approaches based on probabilistic models and mathematical frameworks (Moghaddam et al., 2021). This shift has made it easier to identify pollutants and their concentrations in water which is vital in water resource management. As water becomes more and more contaminated by anthropogenic activities, assessment of the level of pollution in water sources has become crucial not only in raising awareness of the public but also in the formulation of strategies and policies for the sustainable utilization and management of water resources (Banda & Kumarasamy, 2020). Thus, this research proposes to integrate empirical and mechanistic approaches with the help of machine learning for more accurate identification of pollution sources in rivers and, therefore, more effective remediation. The results of this study will help to improve the understanding of the processes of water pollution and will contribute to the formation of more effective measures for the prevention of pollution and preservation of water resources under the conditions of the growth of anthropogenic and environmental impacts.

1.7 Structure of the Thesis/ Dissertation

The thesis is organized into seven chapters. Chapter 1 provides an introduction to the research, outlining the background and significance of the study. Chapter 2 presents a comprehensive literature review, which forms the basis for the research methodology. Chapter 3 details the data collection process and the specific research methods employed in the study. Chapters 4 and 5 cover the study's findings and provide an in-depth discussion of the results. Finally, Chapter 6 and Chapter 7 offer the conclusions drawn from the research and provide recommendations for future work and practical applications.

CHAPTER 2

2 LITERATURE REVIEW

This literature review was conducted to develop the methodology for achieving the aforementioned objectives of this study and identify the appropriate models to be used. It focused on examining relevant research on water quality, water pollution, and the various methods employed to assess pollution over the years. The reviewed studies covered a range of approaches to analyse water pollution issues globally, highlighting different techniques and models used to understand and address water quality challenges.

2.1 Water Quality

Water is an essential resource to sustain life hence the quality of water is also an important factor to consider. Water quality varies along a river and there are a lot of sources which affect the quality of water. However, water quality in many rivers remains poorly monitored, potentially exposing more than 3 billion people globally to waterborne diseases (UN-Water, 2021). Water pollution can have serious impacts on both the flora and fauna of the area hence knowledge of water quality-related areas is essential.

2.1.1 Introduction to Water Quality

Water quality refers to the chemical, physical, and biological properties of water that determine its suitability for various applications, such as drinking, irrigation, and recreation (WHO, 2018). These properties are used to measure the overall condition of water and its capacity to sustain diverse ecosystems and meet human needs. Numerous factors determine the purity of water, including the presence of contaminants such as bacteria, viruses, chemicals, and pollutants (Boyd, 2015). These contaminants can alter the flavour, odour, and appearance of water, as well as pose potential health risks. National and international organisations, such as the Central Environmental Authority (CEA), and World Health Organisation (WHO) establish water quality standards and regulations to guarantee the safety and suitability of water for various uses. These standards establish acceptable levels of contaminants in drinking water, as well as provide recommendations for monitoring and remediation. Overall, water quality plays an essential role in protecting human health, sustaining ecosystems, and promoting sustainable development. The assessment of water quality in rivers often involves the measurement of various parameters, including temperature, pH, dissolved oxygen (DO), turbidity, total suspended solids (TSS), nutrients (such as Nitrogen and Phosphorus), heavy metals, organic contaminants, and microbiological indicators. These indicators offer valuable insights into the water quality and pollution levels of a river (Ewaid et al., 2018). These water quality parameters are divided into four categories as shown in Table 2.1.

Table 2.1 Different categories of water quality

Category	Parameters	Reference
Physical Parameters	pH, Total Suspended Solids (TSS), Electrical Conductivity (EC), Salinity, Total Dissolved Solids (TDS)	Adjovu et al. (2023) Sharma et al. (2016)
Chemical Parameters	Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Bio-chemical oxygen demand (BOD) Nitrate, Nitrite, Phosphorus, Sulphate, Chloride	
Microbiological Parameters	Algae, Total coliform (T. Coli), Faecal coliform (E. Coli)	
Heavy Metals	Cadmium (Cd), Iron (Fe), Zinc(Zn), Chromium (Cr), lead (Pb)	

2.1.2 Water Pollution in Rivers

Water pollution in rivers is caused by three major mechanisms, namely, point source pollution, line source pollution, and diffuse source pollution (Akhtar et al., 2021; Shen et al., 2012). Point sources of pollution refer to sources that discharge pollutants directly into a body of water from a single, identifiable source such as municipal or industrial wastewater discharges (Schaffner et al., 2009). Line sources of pollution refer to pollution originating from a continuous or a linear source such as a pipeline, channel, or road (Xu et al., 2019). Diffuse sources of pollution refer to a broad, non-point origin pollution that cannot be traced back to a specific location. This type of pollution occurs from runoff from agricultural lands and urban runoff (Varekar et al., 2015). Depending on the type, there are different methods of pollution contributing to river pollution (Table 2.2).

These different types of pollution have adverse effects on waterbodies affecting both flora and fauna. In addition, water pollution can have devastating effects on public health (Fida et al., 2023; Hasan et al., 2019). The proliferation of waterborne diseases like cholera, typhoid, and hepatitis can result from contaminated water sources. These diseases can cause illness, hospitalisation, and even mortality, especially in regions where access to clean water and sanitation facilities is limited. Pollution of water can destabilise aquatic ecosystems and damage the plants and animals that rely on them (Uddin et al., 2023). Heavy metals, pesticides, and industrial refuse can accumulate in bodies of water, threatening the health and survival of aquatic organisms. This can result in a loss of biodiversity, disruption of food chains, and devastation of habitats. Water pollution in rivers can impact wildlife and natural habitats (Wear et al., 2021). The growth of toxins and pathogens in plants and animals living in or near contaminated water can hurt their health and survival. In aquatic ecosystems, water pollution can contribute to dangerously low oxygen levels. In polluted water, bacteria decompose organic matter, depleting available oxygen and creating dead zones. This can lead to stress, disease, and mortality of aquatic life (Preisner, 2020).

Table 2.2 Different types of pollution in rivers

Method	**Contributing pollutants**	**References**
Industrial Pollution	Nitrate, sulphate, chloride, Pb, Cr, Cd, Fe, Zn, Organic pollutants	Azam et al. (2019) Fořt et al. (2022) Kishor et al. (2020)
Sewage Pollution	Phosphate, DO, BOD, COD, Sulphate, T. Coli, E. Coli	Figueroa-Nieves et al. (2014) Preisner (2020)
Agricultural runoff	Phosphate, Nitrate	Ighalo and Adeniyi (2020)
Salinity Intrusion	EC, Chloride	Ashrafuzzaman et al. (2022) Xia et al. (2021)
Urban Runoff	TDS, TSS, Turbidity	Neupane et al. (2015) Wang et al. (2017)

In summary, water pollution is a concerning issue because it can lead to several negative impacts on humans, animals, and plants. Modern anthropogenic activities have resulted in water quality degradation of rivers leading not only impacting human health but also to flora and fauna relying on aquatic ecosystems. It is important to assess water quality to mitigate and control these impacts and to develop sustainable water resource management techniques. Studies for quantifying and monitoring water quality along a river are important because it is essential to develop water quality monitoring and mitigation plans.

2.2 Assessment of Water Quality Variations

Identifying spatial-temporal variations is important for formulating countermeasures for water quality-related problems. The assessment of spatial-temporal trends of water quality involves analysing data that vary across multiple dimensions, such as geographic locations and periods. This complexity arises from the need to interpret extensive datasets containing numerous variables measured at different spatial and temporal scales. Existing research studies have used a variety of multivariate statistical techniques for this purpose, including principal component analysis (PCA) (Ouyang et al., 2006; Zeinalzadeh & Rezaei, 2017), factor analysis (FA) (Aydin et al., 2021; Islam et al., 2017), discriminant analysis (DA) (Garizi et al., 2011), and cluster analysis (CA) (Mei et al., 2014; Ustaoğlu et al., 2020; Wu et al., 2023). These multivariate statistical techniques were used in these studies to reduce the dimensionality of a water quality data set and then identify the parameters contributing to the seasonal and spatial variations of water quality.

While traditional statistical methods can be computationally intensive and challenging to apply to large, complex datasets, machine learning techniques, offer a more efficient and often user-friendly alternative for uncovering patterns and trends in water quality data (Badillo et al., 2020). Clustering and dimensionality reduction techniques in unsupervised machine learning utilising Python have become popular and user-friendly and have been used in many applications and research including in water quality-related studies (Alhassan & Wan Zainon, 2021; Govender & Sivakumar, 2020; Vadyala et al., 2022). Unsupervised clustering in machine learning can simplify and interpret data meaningfully by unveiling hidden patterns and structures within large, complex datasets. Unsupervised learning is one of the two main branches of machine learning which mainly focuses on identifying latent yet unknown patterns without pre-existing knowledge from a complex dataset (Marsland, 2020). The two main objectives of unsupervised learning are pattern discovery and dimensionality reduction. Clustering is the most used pattern identification method, while the most used dimensionality reduction techniques include FA and PCA (Badillo et al., 2020; Dike et al., 2018). Unsupervised learning is often used as a part of exploratory data analysis to assess trends and patterns of a complex dataset (Abukmeil et al., 2021).

In summary, the water quality variations have been assessed using several multivariate statistical techniques including PCA, FA, clustering, etc. PCA is the most used method, but recent studies have focused on a more novel approach, which is the clustering analysis technique along with FA and PCA to identify spatial and seasonal variations of water quality. According to past studies, traditional clustering techniques are being transitioned into more sophisticated clustering methods which include Self-Organizing Map (SOM). Yet, there aren’t studies to compare and evaluate the performance of these clustering methods.

2.3 Water Quality and Pollution Indexing

In this section, literature regarding water quality and water pollution indexing methods are reviewed. The purpose of these two methods is different as the water quality indexing generally targets the overall water quality assessment of rivers while the water pollution indexing assesses the specific types of pollution in rivers.

2.3.1 Application of Water Quality and Pollution Indices

Water Quality Indices (WQI) are extensively employed for evaluating the overall status of water bodies, by combining several water quality data into a singular, clearly comprehensible number. Several WQI studies are presented in Table 2.3.

Table 2.3 Applications of water quality indices

Study	Parameters Used	Purpose	Key Insights
Hallock (2002)	Temperature, DO, pH, E. Coli, TN, TP, TSS, Turbidity	Evaluate river water quality and eutrophication	Emphasis on parameters influencing aquatic life and nutrient pollution
Akhtar et al. (2021)	Temperature, EC, TDS, DO, BOD, COD, Coliforms	Assess water quality for recreational, irrigation, and industrial uses	Calibrated with WHO and USEPA standards
Hurley et al. (2012)	pH, TSS, TOC, Nitrates, E. Coli	Evaluate drinking water quality	Versatility of WQI for multiple applications
Lukhabi et al. (2023)	DO, pH, Nitrates, TSS, E. Coli	Surface and groundwater evaluation in Africa	Advocates inclusion of more biological parameters
Uddin et al. (2021)	DO, BOD, pH, Nitrates, TSS, E. Coli	Simplify data for public and policymaker understanding	Region-specific modifications needed to improve precision

According to Table 2.3, WQI studies are conducted to assess the overall water quality of a water body. It is also used to assess the suitability of water for human use. Conversely, the Water Pollution Index (WPI) is extensively utilized to evaluate water pollution across many ecosystems. Table 2.4 presents some WPI studies.

Table 2.4 Applications of water pollution indices

Study	Parameters Used	Focus of WPI	Key Insights
Brady et al. (2015)	Pb, Cd, Zn, Ni	Heavy metal pollution in sediments	Modification needed to enhance sediment handling
Hoseinzadeh et al. (2015)	DO, BOD, SS, NH3-N	Agricultural waste, livestock, industrial pollution	Recommended improved weighting for regional relevance
Hossain and Patra (2020)	pH, EC, TDS, Na, Ca, Mg, Cd, Pb, Zn	Groundwater contamination from heavy metals and runoff	Suggested GIS for more detailed pollution mapping
Karaouzas et al. (2021)	Cd, Pb, Hg, Ni, Zn	Heavy metal pollution and human health risks	Advocates region-specific WPI adaptations
Kumar et al. (2019)	Pb, Cd, Zn, Cu, Ni	Heavy metal contamination and human health	Recommends multivariate analysis for source identification
Prasad et al. (2014)	Fe, Mn, Pb, Cu, Cd, Cr, Zn	Groundwater pollution from mining activities	Weightings should reflect local environmental conditions
Syeed et al. (2023)	BOD, COD, TDS, Pb, Hg	Industrial, agricultural, and domestic pollution	Recommends machine learning and advanced statistics

According to Table 2.4, the WPI is used to identify or quantify different types of pollution in water bodies by considering the specific parameters related to the specific type of pollution. The available literature suggests that the WQI and WPI methods should be enhanced through region-specific modifications and sophisticated statistical methodologies to augment their accuracy and flexibility. Subsequently, these studies suggest that future studies on WPI models could benefit from integrating more advanced statistical methods and machine learning techniques to improve accuracy and better account for regional variations in water quality and should focus on developing methodologies quantifying specific types of pollution in river basins (Karaouzas et al., 2021; Syeed et al., 2023; Uddin et al., 2023).

In summary, the literature suggests two methods to quantify pollution in rivers: WPI and WQI. Although both methods are used to assess water quality, the WPI is more inclined towards the estimation of certain quantities of pollutants such as heavy metals or organic compounds, in a given area, making it more appropriate for identifying certain pollutants in river systems. On the other hand, the WQI offers a more general picture of the water quality since it combines several physical, chemical, and biological variables into one number that indicates the water quality. The WQI is used more frequently to determine the quality of water for different purposes such as drinking, agricultural, or recreational purposes but it is more generalized than the WPI in terms of types of pollution

2.3.2 Major Steps in Index Development

The general index including the WPI is developed in four stages: selection of parameters, sub-indexing of parameters, weighting of parameters, and the aggregation of the sub-index values (Zakaria et al., 2014). For a pollution assessment, parameters are selected to represent the specific type of targeted pollution, such as industrial pollution, considered in this study. Sub-index functions are developed to convert the selected parameter concentrations into a unitless value (Chidiac et al., 2023). The sub-index can be developed based on rating curves or interpolation techniques using the water quality standards recommended by the local authorities. The sub-index values convert the parameter concentrations into a dimensionless value. The sub-index value ranges from zero (0) (minimum pollution) to 100 (maximum pollution). Parameter weighting is performed to find the relative importance and the impact of the selected water quality parameters concerning the river's pollution status (Uddin et al., 2023). Most studies have used the Delphi technique to assign the parameter weights (Lukhabi et al., 2023; Uddin et al., 2023), but more recent studies have started using more data-driven approaches to assign the parameter weights (Gaya et al., 2020; Xiong et al., 2023). Data-driven approaches tend to use more statistical approaches in determining the weights while considering the water quality standards imposed by the authorities. Aggregation combines the weighted sub-index value into a single water pollution value, which determines the overall pollution status of the water body (Syeed et al., 2023). Several weighted and unweighted sub-index aggregation functions are presented in the literature (Akhtar et al., 2021; Lukhabi et al., 2023; Uddin et al., 2022). The aggregation functions follow either an additive form, a multiplicative form, or a maximum/minimum operator form (Kumar & Alappat, 2004). The suitable aggregation function for the developed index is selected based on four criteria (Khouri & Al-Moufti, 2022). The selected aggregation function should minimise the ambiguity problem, which arises when the final index exceeds the critical level without any parameters exceeding the critical level. Secondly, the selected aggregation function should minimise the eclipsing problem, which arises when the aggregated index value is low, although the parameter sub-index values are high. Thirdly, the selected aggregation function should be sensitive to all the changes in individual sub-index values and, finally, the aggregation function should be easy to apply.

2.3.3 Applications of Machine Learning in Indexing

In recent years, the Water Quality Index (WQI) has increasingly integrated machine learning techniques to enhance the accuracy and effectiveness of water quality assessments. Table 2.5 shows some machine learning applications in index development.

Table 2.5 Applications of machine learning in indexing

Study	Machine Learning Models	Key Insights
Aslam et al. (2022)	Decision Trees, Artificial Neural Networks (ANN), Random Forest (RF)	Improved accuracy parameter weighting by reducing traditional method limitations
Gaya et al. (2020)	ANN	ANN and ANFIS were more effective than conventional regression techniques
Gupta et al. (2019)	ANN	ANN models improved accuracy in weighting, particularly cascade forward networks
Makubura et al. (2022)	Regression models	Used machine learning for the aggregation process
Uddin et al. (2023)	XGBoost	XGBoost improved seasonal trend forecasting and reduced uncertainty in parameter weighting.
Xiong et al. (2023)	Random Forest	Minimized subjectivity in parameter weight assignment

In these studies, machine learning has consistently shown the ability to improve WQI models by enhancing predictive accuracy, managing complex data, and enabling real-time monitoring, while also recommending the integration of advanced algorithms and larger datasets for broader geographic applicability. The use of machine learning in developing the index addressed several limitations of traditional methods, such as addressing uncertainties in the aggregation process and potential errors during sub-index weighting. Subsequently, these studies conclude that machine learning algorithms, especially hybrid methods, increase the accuracy of water quality estimates and should be developed further to incorporate real-time monitoring and adaptive learning to increase predictive accuracy and relevance for different geographic areas (Aslam et al., 2022; Uddin et al., 2023).

In summary, the use of machine learning in the development of indices has gradually been integrated into recent developments as a way of enhancing the ability of these indices to give accurate results. At the parameter weighting, sub-index generation, and sub-index aggregation stages, machine-learning approaches have been employed. For instance, Artificial Neural Networks (ANN) and Random Forest

(RF) methods have been employed to identify the key water quality parameters and enhance the accuracy of sub-indices of individual pollution indicators. This approach reduces the influence of human factors, improves the accuracy of predictions, and decreases the level of uncertainty that is inherent in the development of conventional indices. The literature strongly suggests that water pollution indices should be developed based on the characteristics of the catchment area and considering the spatial and seasonal variations unique to that area. This approach makes sure that the indices are more suitable and useful for local environmental management while providing the ability to be tailored to assess other catchments as well. Also, there is increasing interest in applying machine learning methods to improve the accuracy of these indices as well.

2.4 River Water Quality Modelling

Modelling river water quality is a mechanistic method of assessing water quality and pollution in rivers including modelling pollutant transport using process-based models which consider the mass balance and mass transport in rivers. The section below discusses the water transport models that can be used while converging into the best model that can be used in the context of this study, available data, and the scope of the study.

2.4.1 Water Quality Simulation Models

Different water quality models have been used in the literature for simulating different pollutants in rivers. These models have their strengths and weaknesses, and by evaluating these weaknesses we can select the best water quality simulation model for our study. Table 2.6 presents the different water quality models and their strengths and weaknesses (Ambrose et al., 2011; Mawat & Hamdan, 2023; Sharma & Kaushal, 2021).

Table 2.6 Overview of water quality modelling software

Model	Uses	Strengths	Weaknesses
CE-QUAL-R1	Reservoirs, estuaries	Simple 1D model	1D only
CE-QUAL-W2	Lakes, estuaries	Detailed water quality predictions, 2D	Requires much data and computational power
HEC-RAS	Water temperature, sediment transport, unsteady flow	1D and 2D models for temperature and flow	Cannot model all water quality parameters
EFDC	Aquatic systems (coastal regions, estuaries, lakes)	1D, 2D, 3D modelling	High computational requirements

Model	Uses	Strengths	Weaknesses
HSPF	Hydrologic watershed model for nutrients and pesticides	Detailed hydrologic predictions	Complex calibration required
QUAL2E	Steady-state simulations of streams	Waste load evaluations	Limited to 1D
BATHTUB	Nutrient mass budgets and eutrophication	Steady-state nutrient assessment	Steady-state only
CCHE3D	Nutrient and DO variations	2D/3D nutrient and DO simulations	Restricted to 8 state variables
CE-QUAL-RIV1	Hydraulic and water quality changes in rivers	1D river model for steady and unsteady flow	1D only, not suitable for complex rivers
CE-QUAL-ICM	Physical and biogeochemical processes	Flexible multi-dimensional simulations	Requires external processors
EPD-RIV1	River waste load allocation	1D model for waste load	Does not model vertical distribution
MIKE-11	Nutrient transport and eutrophication	Applicable for river eutrophication	Limited spatial sensitivity
QUAL2KW	Contaminant transport in streams and rivers	Simple contaminant transport in 1D	Cannot model sediment movement
WASP	Surface water quality (nutrient cycling, toxicants)	1D, 2D, 3D models for nutrient and toxicant transport	Requires external hydrodynamic data
AQUATOX	Aquatic systems and pollutants	Bioaccumulation, toxicity modelling	Cannot simulate metals

In summary, there have been numerous water quality models developed and used to assess contaminant transport in rivers. According to the discussed papers, there are multiple choices for selecting a suitable water quality model for the study. Based on Table 2-6 EFDC, HSPF, WASP, and AQUATOX models emerge as the best models for modelling pollutant transport in complex river systems. However, the applications of EFDC and HSPF models are limited by the requirements of high computational power and extensive calibration. WASP and AQUATOX models can be used in complex river systems for modelling pollutant transport but the limitation of AQUATOX is the inability to model heavy metals in river systems. However, WASP software emerges as the best selection as it can be used in 1D, 2D, and 3D environments with the model's ability to simulate toxicants in river systems which include heavy metals.

WASP is an integrated environmental system developed to model the fate and transport of pollutants in surface water systems. Originally developed in the 1970s by Hydro Science, Inc., WASP was later supported and enhanced through funding from the U.S. Environmental Protection Agency (USEPA) after becoming publicly available in 1981. Over the past fifty years, WASP has been refined and updated to address evolving environmental issues and advancements in technology (Martin et al., 2013). This model has been used worldwide for water quality analysis because it is very flexible and can simulate many types of contaminants as well as environmental conditions. WASP is structured to simulate both conventional pollutants, such as nutrients, dissolved oxygen, and organic matter, as well as toxic substances, including metals, pesticides, and emerging contaminants like nanoparticles (Knightes et al., 2019). The model consists of a mass balance explaining the transport and transformation processes from different sources to pollutants in water bodies, modelled with differential equations. WASP is developed to model 1D, 2D, and 3D systems which makes it applicable to a wide range of water bodies like rivers, lakes reservoirs estuaries. It consists of two primary modules: the Advanced Eutrophication Module and the Advanced Toxicant Module (Wool et al., 2020). The Advanced Eutrophication Module simulates nutrients and related processes such as algal growth, dissolved oxygen dynamics, and sediment interactions. The Advanced Toxicant Module, recently redesigned in WASP, allows for the simulation of a wider range of chemical pollutants, including emerging contaminants like nanomaterials. This module can model chemical partitioning, particle attachment, and transformation processes, providing users with capabilities for assessing contaminant behaviour in aquatic environments.

2.4.2 Applications of WASP in Water Quality Modelling

Applications of WASP in water quality monitoring studies are discussed in Table 2.7. There have been quite a few studies using WASP to model pollutant transport in rivers for various applications. Most of the studies have used external hydrodynamic models to simulate the river and later were coupled with the water quality model for the model development. There have been several hydrodynamic models used and the most used model was the Hydrological Engineering Centre-River Modelling System (HEC-RAS). Mostly, 1D hydrodynamic models were developed for simulating pollutant transport in rivers (Ernst & Owens, 2009; Mahalakshmi et al., 2018; Prjapathi et al., 2023).

Table 2.7 Applications of Water Quality Analysis Simulation Program (WASP)

Study	**Water Quality Parameters**	**Hydrodynamic Linkage**
Chueh et al. (2021)	Copper concentrations (water column and sediment phases)	SWAT for non-point source pollution (sediment movement)
Shabani et al. (2021)	Arsenic transport during flood events	HEC-RAS 2D for hydrodynamics (flow velocity, shear stress, water depth)
Ernst and Owens (2009)	DO, BOD, inorganic phosphorus, organic phosphorus, ammonia, nitrate-nitrite, organic nitrogen, chlorophyll-a	HEC-RAS for hydrodynamics (river segmentation into 22 zones)
Gordillo et al. (2020)	DO, BOD, nitrogenous compounds, ammonia nitrogen (NH3), nitrate-nitrite nitrogen (NO3-NO2)	1D hydrodynamic model (Saint Venant equations for unsteady flow)
Prjapathi et al. (2023)	Copper and nickel transport (water and sediments)	HEC-RAS 1D for hydrodynamics (flow velocity, water depth)
Mbuh et al. (2019)	Nitrogen, Phosphorus, DO, Chlorophyll-a	DYNHYD for hydrodynamics, WASP used with BASINS for eutrophication and sensitivity analysis
Oldham et al. (2023)	Water quality variables including DO, Nutrients, Chlorophyll	Not specified, focus on WASP simulation of water quality
Mahalakshmi et al. (2018)	Various water pollutants including DO, Nutrients	HEC-RAS coupled with WASP for hydrodynamic and contaminant simulation

2.4.3 Hydrological Engineering Centre-River Analysis System (HEC-RAS)

Studies by the Dysarz (2020) and Zainal and Talib (2024) were used to develop this section. The HEC-RAS model is a powerful hydraulic simulation tool developed by the U.S. Army Corps of Engineers. It is primarily used to simulate river flows and the associated transport of water and sediment. The software supports both 1D and 2D modelling of rivers, offering flexibility depending on the complexity of the system being studied. In 1D modelling, HEC-RAS computes the longitudinal flow in rivers by assuming the flow is uniform across cross-sections, which simplifies the simulation of water surface profiles. This is effective for steady or unsteady flow conditions where lateral variations are not significant, such as narrow rivers or straightforward channels. HEC-RAS uses a variety of calculation methods to simulate river flow, sediment transport, and water quality. Steady flow relies on the energy equation to compute water surface profiles, while unsteady flow, solves the Saint-Venant equations, which account for changes in flow over time. Sediment transport is modelled using the Exner equation and empirical formulas like Meyerpeter müller to predict erosion and deposition. These methods allow HEC-RAS to provide detailed simulations of river dynamics based on inputs like flow rates, cross-sectional geometry, and roughness coefficients. One limitation of HEC-RAS is its reliance on observed streamflow data. It requires precise discharge and flow measurements at various points to accurately simulate river systems, making it difficult to model ungauged catchments. In regions without these measurements, it is challenging to apply the model, limiting its utility in some cases. Hydrological Engineering Centre has developed a rainfall-runoff model which can be used in these scenarios to simulate rainfall-runoff modelling to generate hydrographs for ungauged catchments and sub-catchments, which can assist river modelling (U.S. Army Corps of Engineers, 2023).

2.4.4 Hydrological Engineering Centre-Hydrologic Modelling System (HEC-HMS)

The Hydrologic Modelling System (HEC-HMS) is a powerful tool developed by the U.S. Army Corps of Engineers to simulate the hydrological processes within watersheds (HEC-HMS User's Manual, 2024). It is widely used for a variety of applications, such as flood forecasting, water resource management, and environmental studies. At its core, HEC-HMS models the rainfall-runoff process, which is crucial for understanding how precipitation translates into surface runoff in both urban and rural watersheds. HEC-HMS includes several methods to represent key hydrological components, such as precipitation, evaporation, infiltration, direct runoff, baseflow, and channel routing. The program allows users to simulate these processes under different scenarios, including single storm events or continuous simulations over long periods. For precipitation, HEC-HMS can use observed hydrographs, synthetic storms, or gridded precipitation data. However, when ungauged sub-catchments are modelled using HEC-HMS, there is potential for uncertainty or errors in the resulting hydrographs due to the lack of observed streamflow data for calibration. This reliance on estimated parameters or regionalized data for ungauged areas can lead to discrepancies between simulated and actual runoff behaviour, resulting in less accurate predictions of peak flow, and runoff volume.

2.4.5 Hybrid Modelling for Error Correction in Process-based Models

The study by Konapala et al. (2020) explores the use of hybrid modelling by integrating process-based (PB) hydrologic models, such as the Sacramento Soil Moisture Accounting (SAC-SMA) model, with machine learning techniques, specifically long short-term memory (LSTM) networks. This hybrid approach aims to improve the accuracy of streamflow simulation across diverse catchments in the contiguous United States. Traditional PB models are limited by structural errors due to incomplete representation of physical processes, while ML models, despite their high predictive accuracy, lack physical consistency and can lead to unreliable predictions under changing conditions. Hybrid models use the strengths of both approaches, using PB models to maintain physical consistency while utilizing machine learning techniques to reduce prediction errors by correcting residual biases in PB model outputs. The study tests two hybrid approaches: one where LSTM is trained to simulate streamflow based on both meteorological inputs and PB model outputs and another where LSTM is trained to correct the residuals between the PB model outputs and observed streamflow. Both approaches show significant improvement in the Nash-Sutcliffe Efficiency (NSE), especially in catchments where the PB models perform poorly. This paper suggested future research should be focussed on using hybrid models in regions with sparse data (e.g., ungauged basins) and periods of low flow.

In summary, most mechanistic methods involved in water quality modelling include coupled models where the hydrodynamic modelling is conducted using a hydrological or hydrodynamic model and then coupled with a water quality transport model for water quality monitoring. The WASP model is suggested in the literature as one of the best models for water quality modelling based on the applicability of the software for various pollution scenarios. Similarly, the HEC-RAS 1D modelling approach emerged as the hydrodynamic modelling model which was mostly coupled with the WASP model. Similarly, the applicability of machine learning models for error correction in hydrodynamic models is also highlighted as hybrid models can be used for better-performing models.

2.5 Summary of the Literature Review

In the review of water quality studies, a significant research gap has been identified in the use of empirical methods for identifying pollution sources across different catchments. While traditional multivariate statistical techniques have been used to assess water quality variations, there is limited research on the application of unsupervised learning methods, particularly clustering and pattern recognition, to identify pollution sources. This gap presents an opportunity for further exploration, as these methods could be highly effective for catchment-level pollution identification. Additionally, the development of water pollution indices with the aid of machine learning is an emerging area of interest. Machine learning techniques like ANN and RF are increasingly being integrated into pollution indexing to improve the accuracy and reliability of sub-index generation and parameter weighting. These advancements are targeted at reducing human error and uncertainty, making pollution indices more tailored to local environmental conditions. Similarly, there is a significant research gap in the use of mechanistic methods for pollution source identification. While these mechanistic methods can avoid the limitations of empirical methods, they can be employed in conjunction with machine learning to improve the accuracy and performance of the model. Combining mechanistic models with machine learning for the development of hybrid models is a hot research area in hydrological studies. This study will focus on the application of machine learning with empirical and mechanistic water pollution assessment methods for the identification of potential pollutant sources in source catchments. The application of data-driven machine learning techniques has been used in water quality-related studies and has become an emerging area of interest as the ability of machine learning methods to interpret large data sets often found in water quality studies in more reliable and accurate ways makes them invaluable tools for predicting and managing water quality. Machine learning techniques can efficiently handle complex, nonlinear relationships in water quality data, enabling the development of models that can forecast pollutant levels, track contamination sources, and assess the impacts of various environmental factors on water bodies. This data-driven approach not only enhances the reliability of predictions but also supports informed decision-making in water resource management and pollution control. The proposed research methodology provides a scalable methodology can be used in any catchment for effective pollutant source identification.

CHAPTER 3

3 MATERIALS AND METHODS

In this chapter, the general methodology followed to achieve the main and specific objectives are explained. In addition, the study area and the data used for the analysis were discussed.

3.1 Study Area

The Kelani River Basin (KRB) is located in the western part of Sri Lanka and belongs to the wet zone of the country (Figure 3.1). The Kelani River, stretching 144 km in length, originates from the central highlands of the country. The geographical area of the KRB is 2,230 km^2 and is home to more than 25% of the Sri Lankan population (Mahagamage & Manage, 2014). The annual total rainfall, averaged using the Thiessen Polygon method across six rain gauge stations, amounts to 3,720 mm over the period 2016–2020. Nearly 53% of this total occurs during the Southwest Monsoon (SWM) period from May to September, while the northeast monsoon period contributes the least, accounting for only 10% of the annual total rainfall from December to February. The first inter-monsoon (IM1) from March to April and the second inter-monsoon (IM2) from October to November also contribute 15% and 23% to the annual total rainfall. Regarding monthly rainfall intensity, the IM2 dominates the other three seasons, followed by the SWM season (Figure 3.2). Figure 3.3 shows that the river discharge is highly influenced by the rainfall pattern throughout the year, with peak discharge occurring in May, corresponding to the highest rainfall.

Two export processing industrial zones are located in the KRB, namely, the Biyagama Industrial Zone and the Seethawake Industrial Zone. These two major industrial zones consist of rubber manufacturing industries, textile and apparel industries, dairy industries, and pharmaceutical-related industries, while steel manufacturing industries, plywood factories, and soap/detergent factories are also located (Hemachandra & Pathiratne, 2018; Mahagamage & Manage, 2014). A total number of 2,842 industries as of 2020 are located inside the KRB (Manage et al., 2020). The Central Environmental Authority (Central Environmental Authority, 2021) has categorized these industries mainly into three groups by considering the attributable environmental impacts: Type A, Type B, and Type C with Type A industries being more precarious. About 30% of the industries located in the KRB are Type A, while 43% and 27% are considered Type B and Type C, respectively. Other sources of pollution in the KRB are identified as improper and unplanned land use practices, sewage-related pollution, sand mining/gem mining, and pollution due to fertilizers and pesticides.

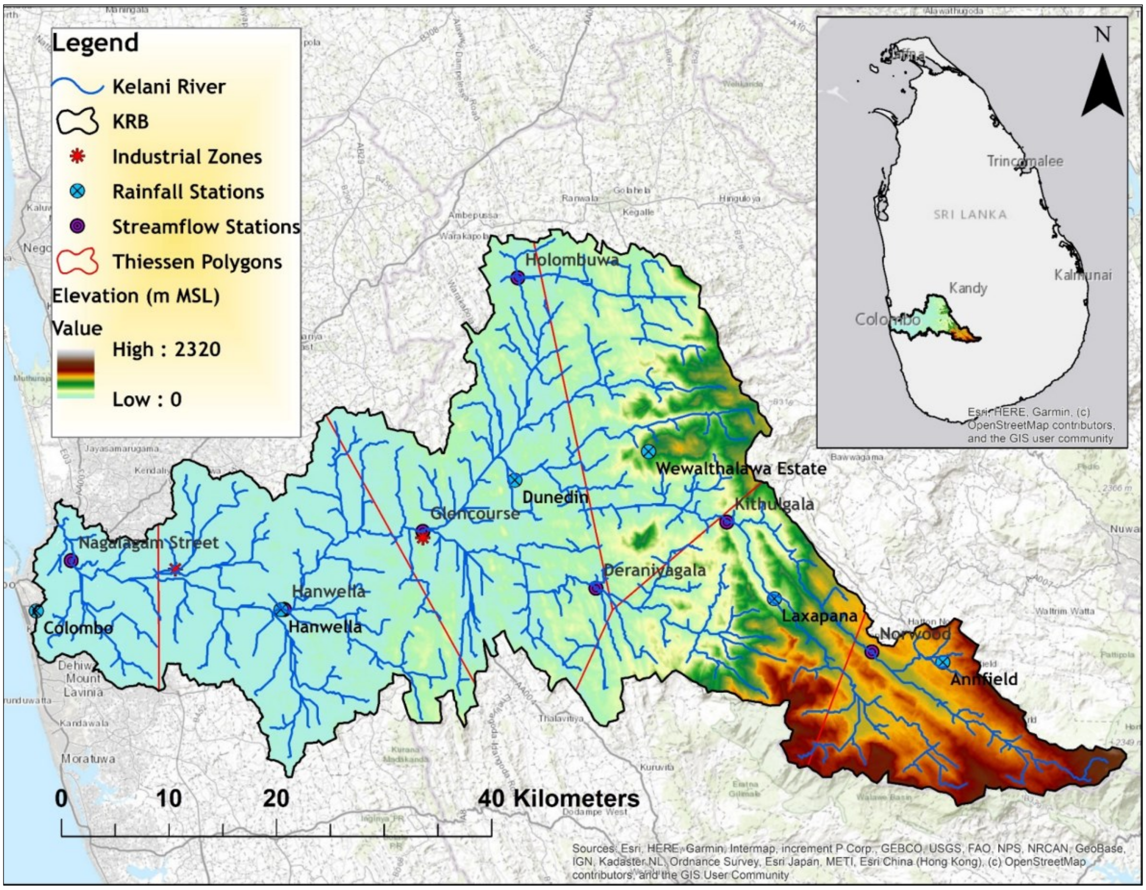


Figure 3.1 Kelani River Basin and the locations of the rainfall and streamflow monitoring stations

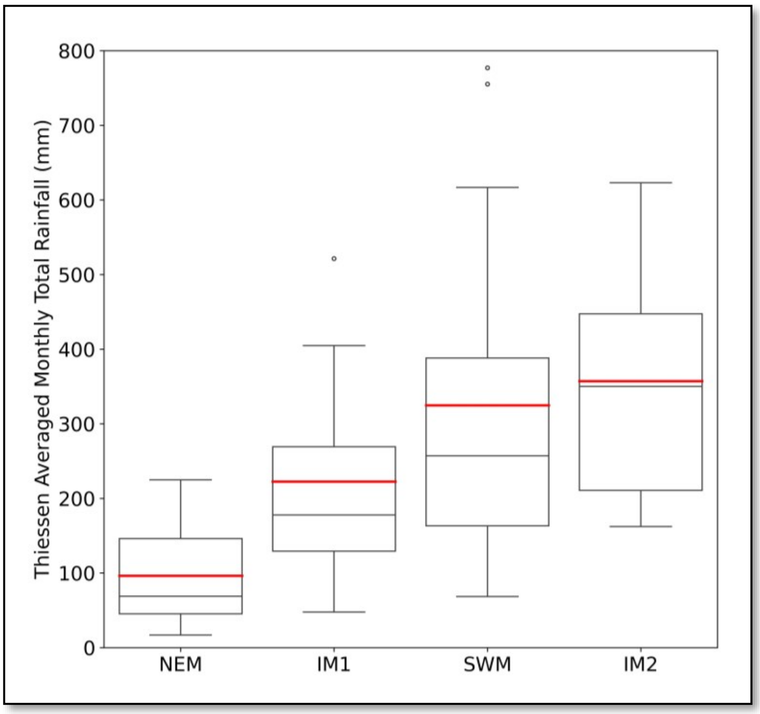


Figure 3.2 Thiessen averaged monsoon rainfall variations in the Kelani River Basin during the 2016–2020 period. Average rainfall is shown in red, while the whiskers extending from the top and bottom represent the maximum and minimum rainfall recorded. Additionally, upper, middle, and lower lines

within the box illustrate the quartile ranges of the 75th, 50th, and 25th percentiles of rainfall distribution, respectively (NEM-Northeast monsoon, IM1-First intermonsoon, SWM-Southwest monsoon, and IM2-Second intermonsoon)

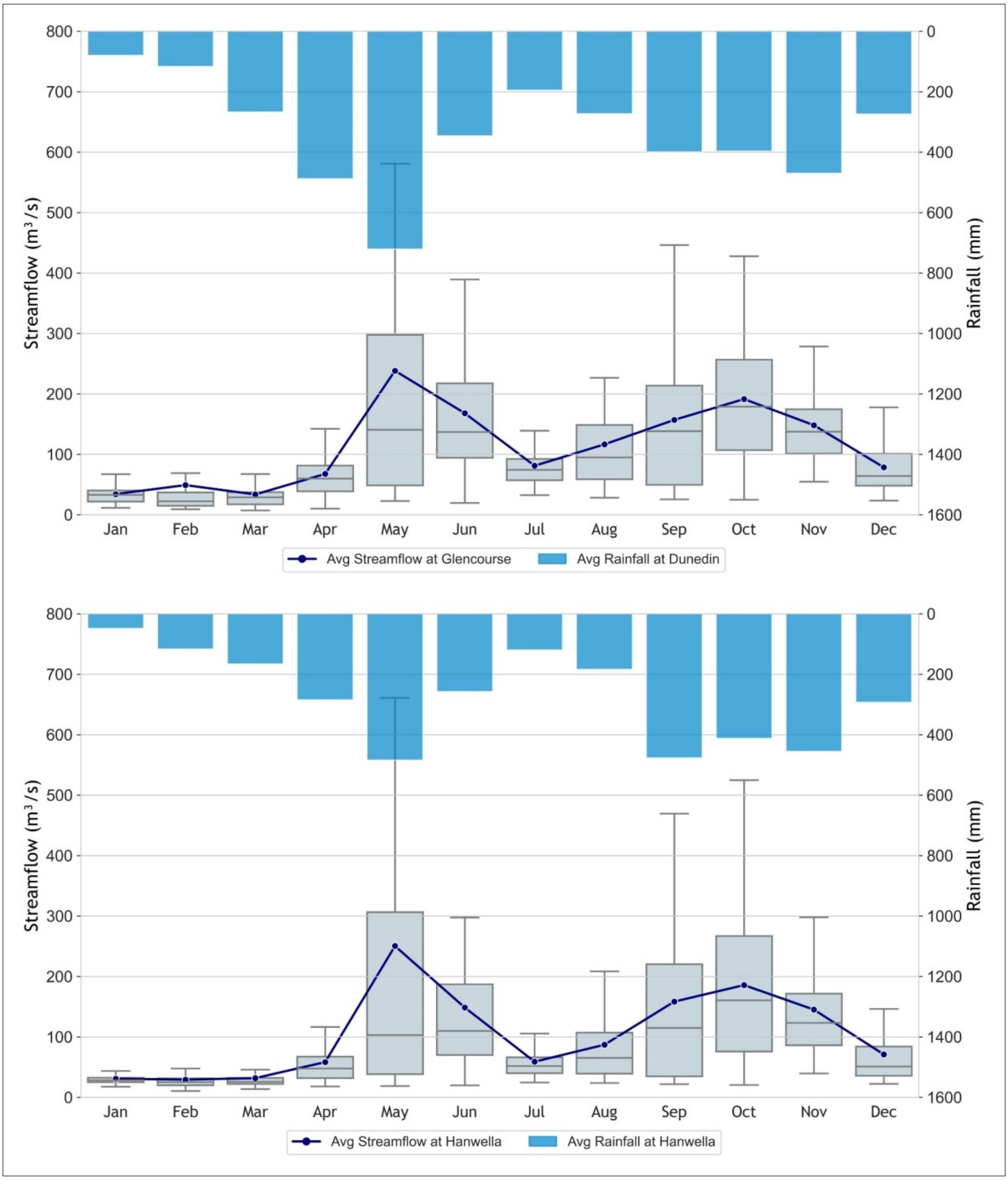


Figure 3.3 Monthly average rainfall and streamflow at Hanwella and Glencourse streamflow monitoring stations in the Kelani River Basin for the period 2016–2020

3.2 Research Methodology

Based on the literature review, a comprehensive methodology was developed to assess the pollution in a river. The flowchart developed for the study is shown in Figure 3.4. The methodology for assessing the river water quality and determining the sources of pollution was developed through empirical and mechanistic approaches. This combined approach guarantees analysis by integrating machine learning techniques and process-based modelling. Required data includes the measured water quality, observed stream flows, river cross-sections at various places along the Kelani River, and rainfall data of the KRB. These datasets are necessary for statistical analysis of pollution and for model development and validation of the hydrological and water quality models.

The first part of the analysis involved the determination of temporal and spatial variations in water quality in the catchment. Factor analysis (FA) and cluster analysis (CA) were used to analyze these variations. Factor analysis was used to identify factors that influence the level of pollution in the river and the cluster analysis grouped the various water quality monitoring stations according to the degree of pollution similarity. These analyses are important when searching for spatial patterns. The water pollution indexing technique was used to determine the pollution level in the Kelani River at different points. This method used the conventional pollution indices that integrate several water quality parameters into a single value and machine learning to enhance the performance and reliability of the indices. The parameters were weighed using machine learning to identify the most influential water quality parameters in the basin. The index generated offered an unambiguous quantitative assessment of water quality that was straightforward and was applied to determine probable sources of pollution in the river.

Then a water quality simulation model was developed for the study area to identify the potential impacts of the identified pollution sources on the Kelani River. WASP is a widely used and recognized tool for simulating the transport of pollutants in surface water bodies, including rivers. It simulated the physical processes that control the transport and fate of contaminants, including advection and dispersion. The hydrodynamic component of the WASP model was developed using HEC-RAS. HEC-RAS simulates the river flow with the help of cross-sectional geometry, stream flow data, and other physical attributes of the river. The sub-catchment hydrographs required for the HEC-RAS model were developed from the HEC-HMS model. HEC-HMS is used in the catchment to model rainfall-runoff processes, giving an estimation of the amount of water that is entering the river system from rainfall. To enhance the accuracy of the runoff predictions in this study, HEC-HMS was integrated with machine learning, to determine the potential sources of pollution in the river. The results from the WASP model were used to evaluate the impact of pollutant dynamics on the Kelani River. Based on the comparison of the model-predicted pollutant concentrations with the observed water quality data and the pollution index values, the pollution characteristics of the river were assessed. The integration of empirical data analysis, machine learning, and mechanistic modelling enabled the identification of pollution sources in the river.

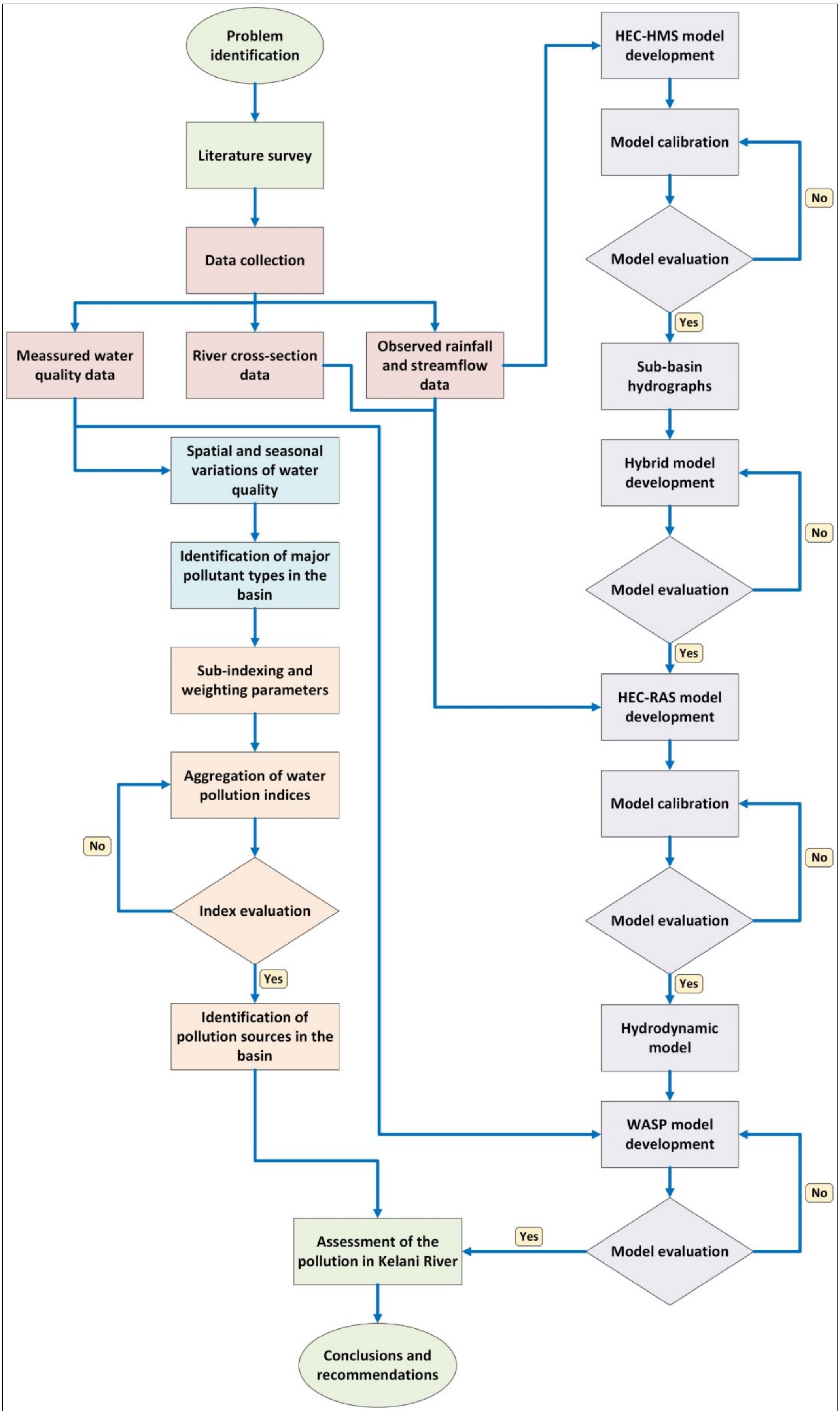


Figure 3.4 Research Methodology

3.3 Data Collection

3.3.1 Water Quality Data

Monthly water quality monitoring data at 17 stations along the Kelani River (Figure 3.5) were obtained for five years (2016–2020) from the Central Environmental Authority (CEA) of Sri Lanka. The 17 measured parameters are pH, electrical conductivity (EC), turbidity, temperature, dissolved oxygen (DO), five-day biological oxygen demand (BOD), chemical oxygen demand (COD), nitrate, phosphate, chloride (Cl^-), cadmium (Cd), lead (Pb), iron (Fe), chromium (Cr), zinc (Zn), total coliform (T. Coli), and fecal coliform (E. Coli). These parameters can be categorized into four water quality groups in Table 3.1.

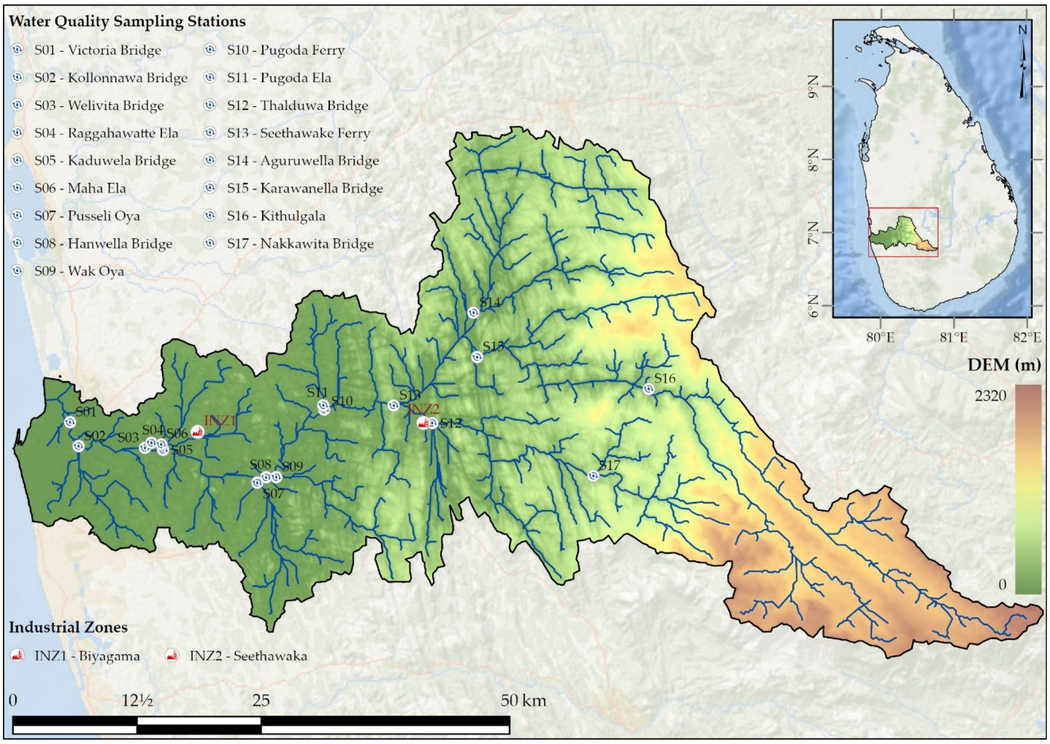


Figure 3.5 Locations of the water quality monitoring stations in the Kelani River Basin

Table 3.1 Measured water quality parameters

Category	Parameter
Physical Parameters	pH, EC, turbidity, temperature
Chemical Parameters	DO, BOD, COD, nitrate, phosphate, Cl^-
Microbiological Parameters	T. Coli, E. Coli
Heavy metals	Cd, Pb, Fe, Cr, Zn

3.3.2 Rainfall and Streamflow Data

A summary of the collected rainfall and streamflow data is shown in Table 3.2.

Table 3.2 Details of the rainfall and streamflow data collection

Data type	Station	Temporal resolution	Period	Source	Missing value percentage (%)
Rainfall	Colombo	Daily measurements	2016–2020	Department of Meteorology	0.05
	Hanwella		2016–2020		3.34
	Dunedin		2016–2020		13.48
	Laxapana		2016–2020		0.00
	Annfield		2016–2020		10.07
Streamflow	Hanwella	Daily measurements	2016–2020	Irrigation Department	-
	Glencourse		2016–2020		-
	Kithulgala		2016–2020		-
	Deraniyagala		2016–2020		-
	Norwood		2016–2020		-
Water level	Nagalagam Street	Daily measurements	2016–2020	Irrigation Department	-

Data checking was carried out mainly based on visual observation, single mass curve, and double mass curve. Single mass curve and double mass curve were used to check the consistency of the rainfall data while visual observation was carried out to check the consistency of the streamflow data. For the model development of hydrological models, a complete data set of hydrometeorological parameters is needed. Only the Colombo, Hanwella, and Dunedin rainfall stations were used for the model development hence missing values were filled only for these three stations. The maximum acceptable missing data percentage of any station is 10% of the total available data for the station (World Meteorological Organization, 2009), and the three stations had a missing percentage of less than 10%. The Multiple Imputation by Chained Equations (MICE) method was used to fill in the missing values in the rainfall data set (Jayaminda et al., 2023; van Buuren & Groothuis-Oudshoorn, 2011).

3.3.3 Single Mass Curve

The trends of different rainfall stations over time can be examined using single-mass curves (Jayaminda et al., 2023). Similarly, the consistency can be checked of the rainfall values through the single mass curve. The cumulative precipitation time series value for each station is plotted to identify trends and consistency. The developed single-mass curves are shown in Figure 3.6. According to the figure, Dunedin station has the highest cumulative rainfall for the period 2016–2020, Hanwella is the second highest, and Colombo with the lowest cumulative rainfall suggesting that higher rainfall occurs in the highland areas inland, like Dunedin, compared to coastal areas like Colombo, indicating that rainfall intensity increases toward the central highlands of Sri Lanka.

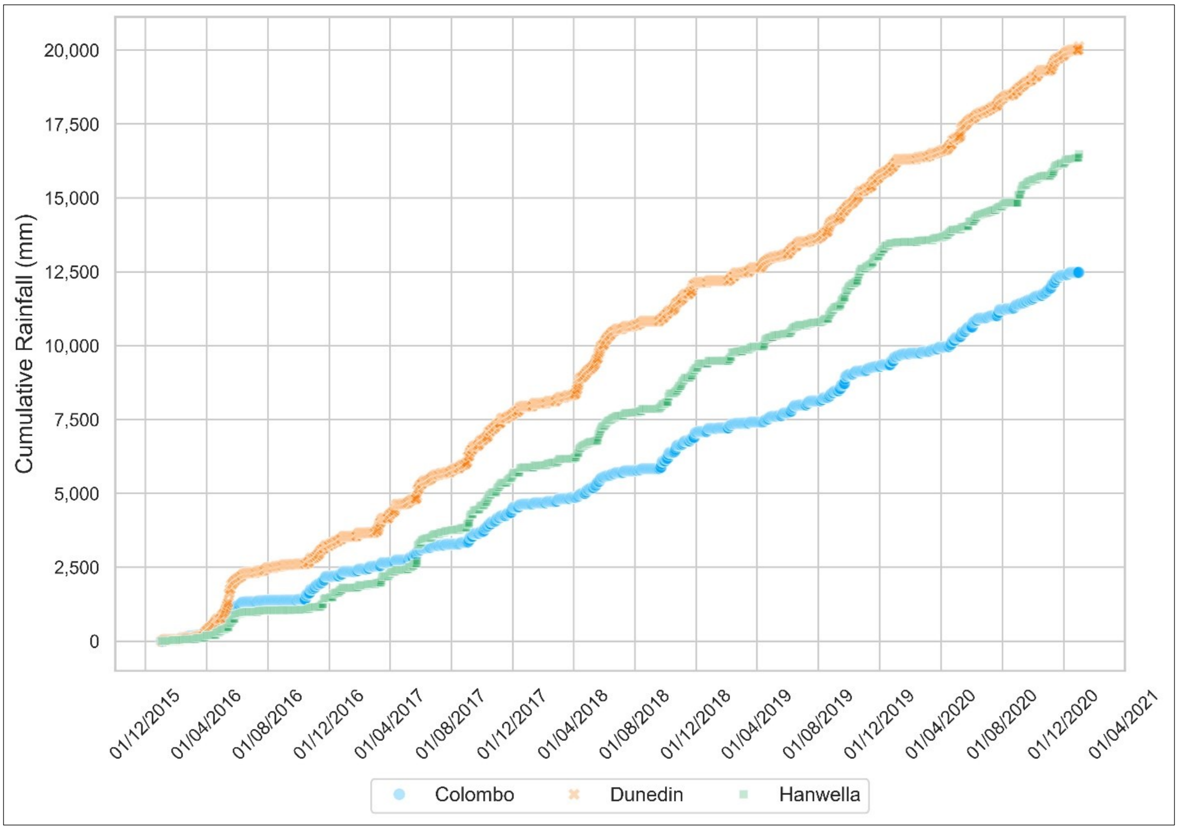


Figure 3.6 Single mass curve for Colombo, Hanwella, and Dunedin rainfall stations

3.3.4 Double Mass Curve

The double mass curve is used to compare the data from a single rainfall station with a pattern comprised of data from several other stations in the surrounding to analyze the behaviour of records of hydrometeorological data obtained at multiple locations (Guoshuai et al., 2023). It can be used to check the trends of rainfall data for long-term durations along with the consistency of the rainfall data. The Colombo, Hanwella, and Dunedin rainfall station data were used to develop the double mass curves for each station. The cumulative rainfall of the station that was assessed was plotted with the average cumulative rainfall of other rainfall stations to develop the double mass curve. The developed double mass curves for Colombo and Hanwella stations are shown in Figure 3.7 and Figure 3.8. The figures demonstrate the consistency of the observed rainfall data compared to the observed station data around the selected station.

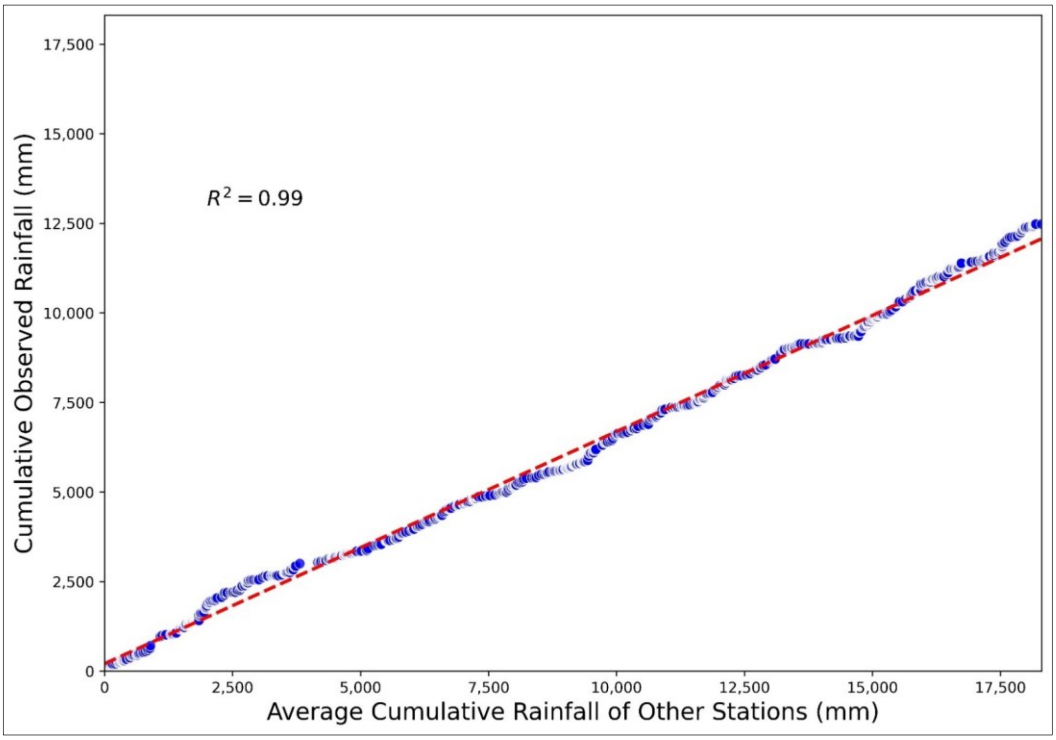


Figure 3.7 The double mass curve for the Colombo rainfall station

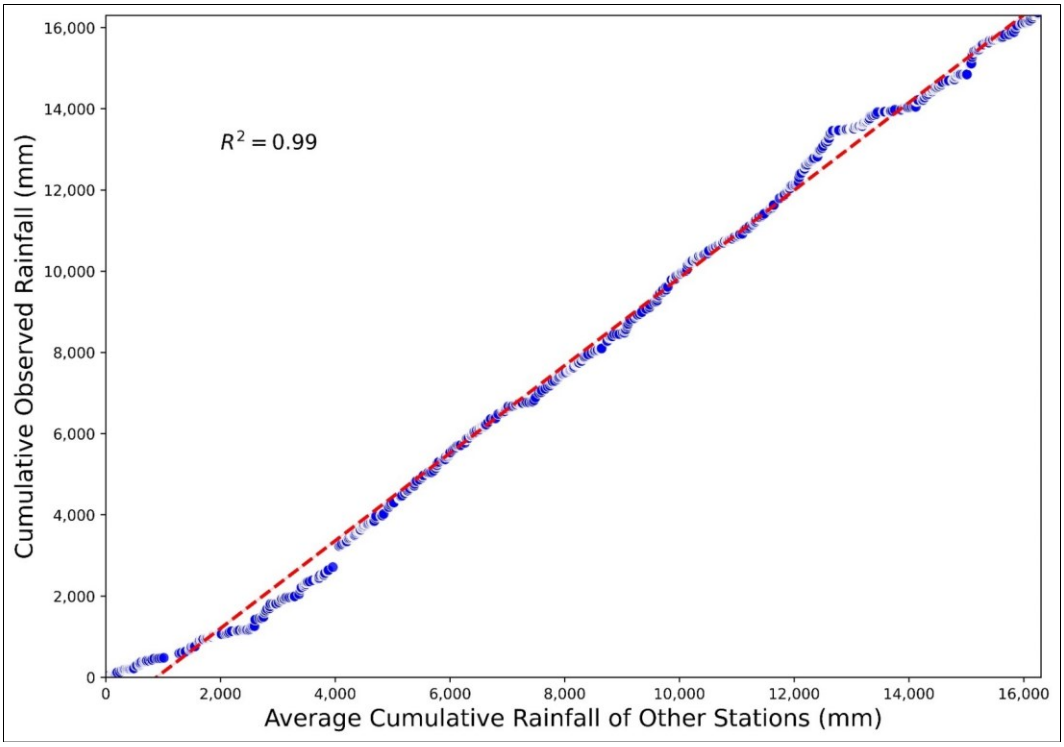


Figure 3.8 The double mass curve for the Hanwella rainfall station

3.3.5 Visual Checking

Visual checking of data is the most straightforward way to check the consistency of the observed streamflow data (Guoshuai et al., 2023). This method is conducted by plotting the streamflow time series values and the rainfall values to identify the correspondence between the rainfall and streamflow. Visual checking was performed for the Nagalagam street water level (Figure 3.9) and Hanwella (Figure 3.10) and Glencourse (Figure 3.11) streamflow stations, between 2016 and 2020. The Nagalagam Street water level was checked using Colombo rainfall data and for Hanwella and Glencourse stations the Hanwella and Dunedin stations were used respectively. According to the figures, observed water level and streamflow data are consistent with the observed rainfall data.

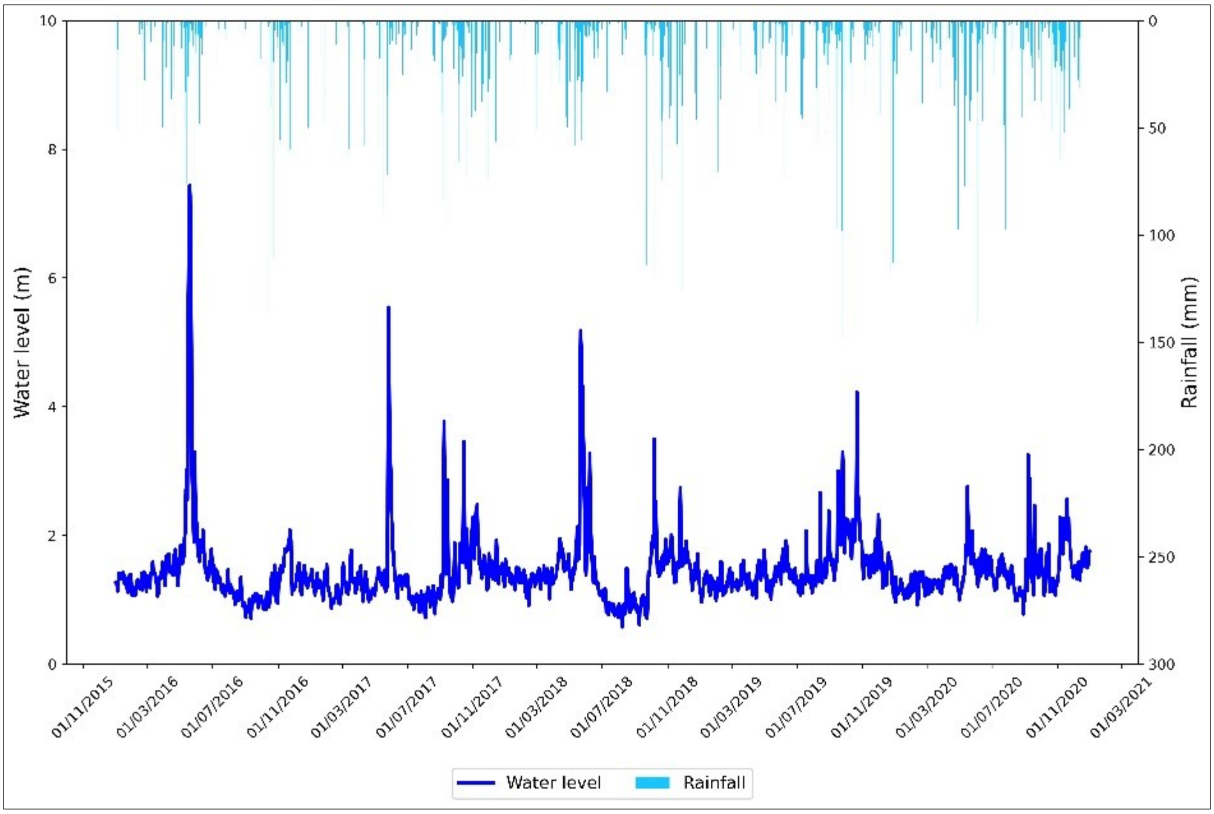


Figure 3.9 Water level time series for Nagalagam Street monitoring station

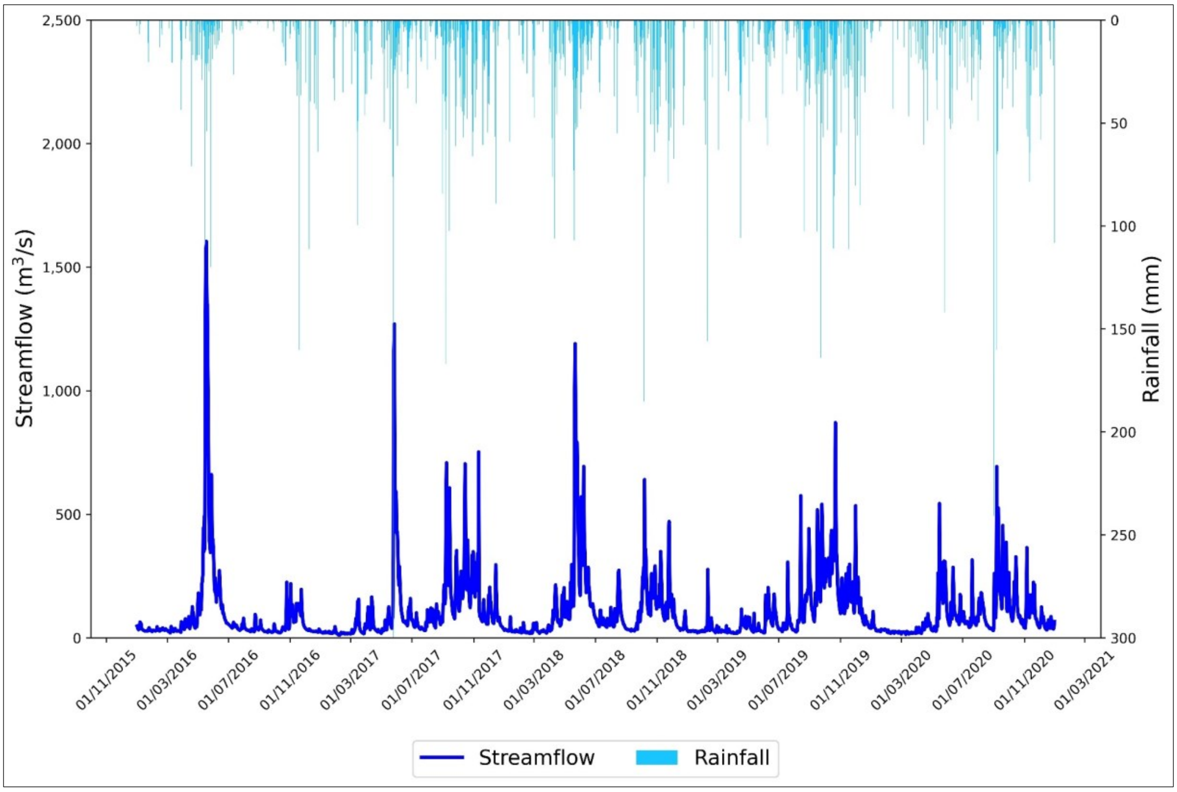


Figure 3.10 Streamflow time series for Hanwella monitoring station

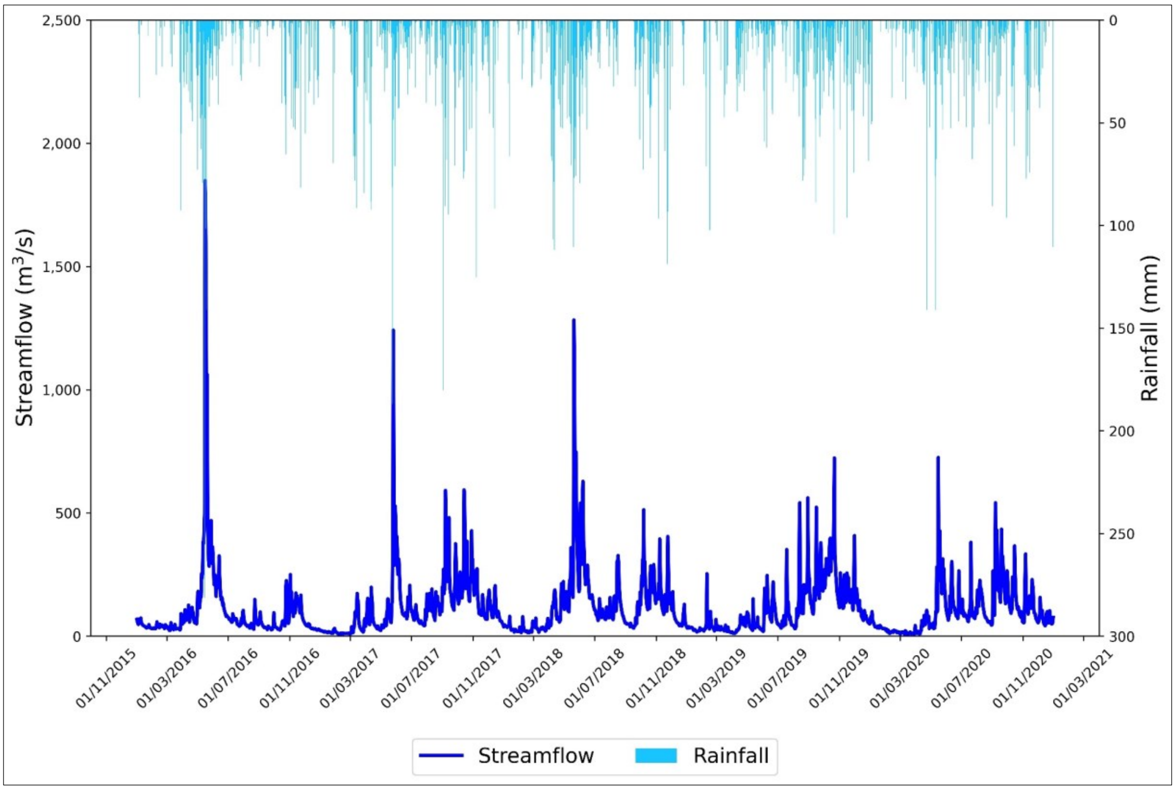


Figure 3.11 Streamflow time series for Glencourse monitoring station

3.4 Factor Analysis

Factor Analysis (FA) plays a crucial role in reducing the dimensionality of complex water quality data. When dealing with numerous parameters (17 parameters as in this study), FA helps identify underlying patterns and relationships among them, effectively reducing the data to a smaller set of factors that capture most of the variation (Aydin et al., 2021; Islam et al., 2017). It is an unsupervised learning technique performed by identifying latent patterns or relationships. It is divided into two sections, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The EFA is performed to find the underlying yet unknown factors in a complex dataset while the CFA is performed to confirm the observed latent patterns. In EFA, this objective is achieved by grouping variables into latent variables (LV) based on their shared variance. These LVs are unobservable, hypothetical constructs that represent the common trends behind the observed variables (Yong & Pearce, 2013).

In EFA, p denotes the number of water quality parameters ($X_1, X_2, \ldots, X_p$) and m denotes the number of underlying latent variables (LV_1, LV_2,…, LV_m) while X_j is a water quality parameter represented in the LV. Hence, the FA model assumes that there is m number of LVs, and each observed water quality parameter is a linear function of these LVs along with a residual variate. This FA model intends to reproduce the maximum correlations. The factor loadings are $a_{j1}, a_{j2}, \ldots, a_{jm}$ which denotes that a_{j1} is the factor loading of the j^{th} water quality parameter on the 1st LV. The residual variate is denoted by e_j as follows,

$$X_j = a_{j1}LV_1 + a_{j2}LV_2 + \ldots + a_{jm}LV_m + e_j \qquad \text{(Equation 1)}$$

where $j = 1, 2, \ldots, m$

The factor loadings indicate how much the LV has contributed to the particular water quality parameter; the larger the factor loading the more the LV has contributed to that parameter (Asuero et al., 2006). Extracting too many LVs may produce unsuitable error variance, while inadequate LVs may exclude important common variance. The eigenvalues and scree plots determine how many LVs are to be retained. Kaiser's criterion is a rule of thumb for determining the number of LVs to retain. This criterion suggests retaining all LVs with eigenvalues greater than one. The applicability of a dataset in FA can be assessed using the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test (Shrestha, 2021).

In EFA, the factor loadings are rotated after the LVs are estimated from the model to produce a simpler and more interpretable structure that can be used to identify which water quality parameters are influencing a specific LV. Varimax rotation and Promax rotation are the most common rotation methods (Asuero et al., 2006). Accordingly, the Factor-Analyzer Module in the Python environment was used to perform the EFA to identify the hidden correlations between the water quality parameters (Ahlgren et al., 2003). Pearson's correlation matrix can be used to identify the correlating parameters to be used in the FA. It is a mathematical representation of the pairwise correlations of two water quality parameters in the dataset. Pearson's correlation coefficient (*r*) measures the linear relationship between two parameters by considering the covariance of the two parameters.

3.5 Cluster Analysis

Hierarchical Cluster Analysis and Self-Organizing Map techniques were used in this study for clustering. The two methods are presented in this section.

3.5.1 Hierarchical Cluster Analysis

Hierarchical Cluster Analysis (HCA) is a traditional clustering technique used to classify large numbers of data into meaningful subsets or groups according to their characteristics. In this study, the HCA was used to identify the different spatial characteristics of the LVs identified in FA and to group them according to the variations of these LVs. Distance methods in HCA are widely utilized for assessing similarity between groups, with the Euclidean distance (denoted as *d*) being the most frequently used approach for measuring distance. Euclidean distance is calculated as follows,

$$d(z_u, z_w) = \sqrt{\sum_{j=1}^{N_d} (z_{u,j} - z_{w,j})^2} = ||(z_u - z_w|| \qquad \text{(Equation 2)}$$

where:

- $d(z_u, z_w)$ represents the Euclidean distance between two points, Z_u and Z_w,
- N_d represents the number of dimensions or features in the data, and
- $Z_{u,j}$ and $Z_{w,j}$ represent the values of the j^{th} feature of the data sample.

The result of the HCA method is a dendrogram that represents the clustered groupings of the objects and the similarity levels at which the clusters change. The data objects are clustered by cutting the dendrogram at the desired similarity level. Clusters are merged or divided based on the measure of similarity. The Silhouette score is a metric used to calculate the effectiveness of the clustering results. The value ranges from -1 to 1 and a value in the range of 0 to 1 is considered justifiable (Omran et al., 2007; Zakaria et al., 2014). The Silhouette score for a single data point *i* is calculated using (Mulyani et al., 2023):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad \text{(Equation 3)}$$

where, $a(i)$ is the average distance from *i* to other data points within the same cluster, and $b(i)$ is the average distance from *i* to data points in a different cluster, minimized over clusters.

The Silhouette score for the entire dataset is the average of the Silhouette scores for all data points. The HCA in this study was conducted in the Python environment using the Sci-kit Learn module (Pedregosa et al., 2011).

3.5.2 Self-Organizing Maps

Self-Organizing Map (SOM) is an unsupervised neural network method used for clustering. The SOM technique employs an iterative approximation algorithm to map a set of vectors from a high-dimensional space to a low-dimensional space while maintaining the original spatial relationships between them. This method is used to map and visualize data in a nonlinear manner (Zakaria et al., 2014). The SOM was also implemented to analyse the spatial variations along with the seasonal distribution of the LVs. Unlike a traditional feedforward network, a SOM is organised in n-dimensional grids. A node's location in the grid relates to a clustered region of the higher-dimensional state space. The node is activated when the network receives a vector that "originates from" its corresponding region in state space. This mapping is random yet consistent. In the training process, vectors are randomly distributed across the grid. This method guarantees that the winning node has the highest activation while adjacent nodes' activations are reduced. Eventually, only one output node is activated for each set of vectors. All vectors that activate the same output node can be regarded as part of the same cluster (Omran et al., 2007; Treshansky & McGraw, 2001).

In this study, the vectors refer to a data point of the LVs while n-dimensional grids represent the ‘n’ number of data entries which are the number of water quality measuring stations. The SOM algorithm will reduce the dimensionality of these data to identify the clusters. In this study, the SOM grouped 17 water quality measuring stations into different clusters based on the variation of LVs. Self-organising maps can be implemented in the Python environment using the MiniSom module (Kiran et al., 2021).

3.6 Water Pollution Indexing

This section discusses the methods used to develop the water pollution index used in this study. The different steps of index development are discussed in this section.

3.6.1 Parameter Selection and Sub-indexing

For a pollution assessment, parameters are selected to represent the specific type of targeted pollution, such as industrial pollution or sewage pollution. Sub-index functions are developed to convert the selected parameter concentrations into a unitless value (Omran et al., 2007). The sub-index can be developed based on rating curves or interpolation techniques using the water quality standards recommended by the local authorities. The sub-index value ranges from 0 (minimum pollution) to 100 (maximum pollution). In this study, the methods proposed by Hallock (2002) and Casillas-García et al. (2021) were used to develop rating curves accommodating the local standard limits. The first step was to compare the observations with the particular water quality standard limits (V_{perm}) to determine the number of observations that exceeded the standards within the analysis period. Subsequently, the data set was divided into two subsets: the observations less than or equal to and higher than the permissible values. A sub-index value (Q_i) was assigned to the permissible limit ($Q_{i,\ perm}$) depending on the percentage of the observations that exceeded the standard limit,

$$Q_{i,\ perm} = 50 + 15P_e \quad \text{(Equation 4)}$$

where $0 \leq P_e \leq 1$ is the proportion of observations that exceeded the standard limit of each parameter.

Accordingly, the sub-index took the minimum values of 50 and 65 when all the observations were lower than and exceeded the standard limit, respectively. For the development of rating curves, the quantiles 0, 0.1, 0.8, 0.95, 0.99, and 1 for the subset of observations lower than the standard limit were calculated, and the sub-index (Q_i) was assigned using the following equations presented in Table 3.3.

Table 3.3 Sub-index (Q_i) values for observations lower and higher than the standard limit for each quantile

Quantiles	**For a subset of observations lower than the standard limit**	**For a subset of observations higher than the standard limit**
0	0	$Q_{i,perm}$
0.1	$0.1Q_{i,perm}$	$1.1Q_{i,perm}$
0.8	$0.2Q_{i,perm}$	$1.2Q_{i,perm}$
0.95	$0.6Q_{i,perm}$	$1.4Q_{i,perm}$
0.99	$0.8Q_{i,perm}$	$1.5Q_{i,perm}$
1	$Q_{i,perm}$	100

3.6.2 Parameter Weighting

The parameter weighting is performed to find the relative importance and the impact of the selected water quality parameters concerning the pollution status of the river (Chidiac et al., 2023). Most studies have used the Delphi technique to assign the parameter weights (Chidiac et al., 2023; Zakaria et al., 2014), but more recent studies have started using more data-driven approaches to assign the parameter weights (Uddin et al., 2021). Data-driven approaches tend to use more statistical approaches in determining the weights while considering the water quality standards imposed by the authorities. Initially, the water pollution status was assessed by comparing the measured values with the standard limit using a simple binary function to obtain parameter weight values. If all water quality parameters met the guideline limit, the status was assigned a binary value of 0, indicating 'unpolluted.' Conversely, if any parameter exceeded the standard limit, a binary value of 1 was assigned, indicating 'polluted' (Uddin et al., 2023). When a parameter's value exceeded the standard limit, the measured concentration was replaced with the difference between the measured and standard limits. If the parameter value did not exceed the limit, the concentration was replaced with a value of 0. This method allows for incorporating the frequency and magnitude of standard limit exceedances to determine the river's initial pollution status. A Random Forest (RF) model was then employed to assess the parameters and identify their relative importance to the initial pollution status. Further, the Rank Order Centroid (ROC) approach was used to assign weights to the parameters, and these weights were validated by comparing them with the relative importance derived from the RF model.

3.6.3 Parameter Aggregation and Classification

Aggregation combines the weighted sub-index value into a single water pollution value, which determines the overall pollution status of the water body (Ahmed Salman et al., 2024). Several weighted and unweighted sub-index aggregation functions are presented in the literature (Biau & Scornet, 2016; Varshney et al., 2024; Zakaria et al., 2014). The aggregation functions follow either an additive form, a multiplicative form, or a maximum/minimum operator form (Treshansky & McGraw, 2001). The suitable aggregation function for the developed index is selected based on four criteria (Syeed et al., 2023). The selected aggregation function should minimise the ambiguity problem, which arises when the final index exceeds the critical level without any parameters exceeding the critical level. Secondly, the selected aggregation function should minimise the eclipsing problem, which arises when the aggregated index value is low, although the parameter sub-index values are high. Thirdly, the selected aggregation function should be sensitive to all the changes in individual sub-index values, and finally, the aggregation function should be easy to apply.

Based on the literature, five mostly used aggregation functions were selected for this study. For N number of water quality parameters, ranked from i=1 to N where the sub-index value of the ith parameter is S_i, and the weight of the ith parameter is W_i:

$$Weighted\ Arithmetic\ Mean\ (WAM) = \sum_{i=1}^{n} S_i W_i \qquad \text{(Equation 5)}$$

$$Weighted\ Multiplicative\ Function\ (WMF) = \prod_{i=1}^{n} {S_i}^{W_i} \qquad \text{(Equation 6)}$$

$$Root\ Sum\ Power\ Funtion\ (r = 2)(RSPF2) = \left(\sum_{i=1}^{n} W_i S_i^2\right)^{\frac{1}{2}} \qquad \text{(Equation 7)}$$

$$Root\ Sum\ Power\ Funtion\ (r = 4)(RSPF4) = \left(\sum_{i=1}^{n} W_i S_i^4\right)^{\frac{1}{4}} \quad \text{(Equation 8)}$$

$$Weighted\ Ambiguity\ and\ Eclipsity\ Free\ Function\ (WAEF) = \left(\sum_{i=1}^{n} W_i S_i^{2.5}\right)^{0.4} \quad \text{(Equation 9)}$$

Subsequently, a sensitivity analysis was conducted to determine the most appropriate aggregation function for the developed indices. The aggregation function that exhibited the highest sensitivity to changes in parameter values was selected as the optimal choice for each pollution index. A Long Short-Term Memory (LSTM) Artificial Neural Network (ANN) model was developed to validate the results of the selected aggregation function. The percentile method was used to classify the water pollution values into five categories (Bayissa et al., 2019; Shen, 2018). The lowest aggregated value was classified into the lowest pollution type, and the highest aggregated value was categorised into the maximum polluted status (Moghaddam et al., 2021; Prasad et al., 2014) (Table 3.4).

Table 3.4 Classification of the water pollution indices

Percentile Range	Pollution Type
0–20	Very Low Pollution
20–40	Low Pollution
40–60	Moderate Pollution
60–80	High Pollution
80–100	Extreme Pollution

3.7 Random Forrest Model

In applying the Random Forest (RF) model to assess water quality, the model constructs an ensemble of decision trees to enhance the accuracy and stability of predictions related to water quality parameters. The RF algorithm introduces two layers of randomisation to improve model robustness. First, it randomly selects samples from the water quality dataset to form the root node samples for each decision tree, ensuring a diverse representation of the data (Gaya et al., 2020). Additionally, a random subset of potential water quality parameters is chosen at each node of the tree. The model selects the most appropriate parameter from this subset to create a split, driving the decision process. This method of random resampling and parameter selection leads to the creation of multiple decision trees, each contributing to the final model. The collective prediction is obtained by aggregating the outcomes from all trees through majority voting or averaging. This approach allows the RF model to evaluate the relative importance of each water quality parameter in the weighting process of index development, ultimately providing a comprehensive understanding of how these parameters influence the overall water quality assessment. In-depth discussions on the RF model are presented by Xiong et al. (2023) and Uddin et al. (2021).

3.8 Rank Order Centroid (ROC) Method

In a water quality study, the Rank Order Centroid (ROC) approach provides a straightforward method for assigning weights to various water quality parameters (Xu et al., 2022). In this approach, rankings are produced based on the data distribution of the water quality parameters, and these ranks are then converted into weights. The ROC method minimises the maximum error by identifying the centroid of all possible weight distributions while preserving the rank order derived from the data. The ROC weights (w_i) of a set of N variables, ranked from i =1 to N, are calculated according to:

$$w_i = \frac{1}{N}\sum_{i-1}^{N}\frac{1}{i} \qquad \text{(Equation 10)}$$

3.9 Long Short-Term Memory (LSTM) Artificial Neural Network (ANN)

LSTM networks, a specialised artificial neural network (ANN), are designed to store and manage information over extended periods, making them particularly suitable for modelling temporal dependencies in meteorological data (Akhtar et al., 2021). The LSTM network achieves this by utilising cells and gates to handle memory effects. The cells serve as memory units, while the gates control the flow of information, effectively addressing the challenges of exploding and vanishing gradients that can affect traditional ANNs. This capability allows the LSTM network to establish robust connections for historical meteorological inputs and water quality indices over long periods. An in-depth discussion on LSTM ANN is presented by Uddin et al. (2021) and Khouri and Al-Moufti (2022).

3.10 Rainfall-Runoff Modelling with HEC-HMS

Rainfall-runoff modelling in HEC-HMS is conducted by assigning a loss method, a transformation method, and a baseflow method (US Army Corps of Engineers, 1998). The runoff is calculated from meteorological data by deducting losses, transforming excess rainfall, and adding baseflow. In addition, a canopy component to incorporate evapotranspiration and intercept, and a surface component to account for the water trapped in surface depressional storage were added to the model (Sahu, et al., 2020). In addition to the above-mentioned methods, the Muskingum method was selected as the flow routing method (Table 3.5).

Table 3.5 Selected mathematical models for the HEC-HMS model development

Component	**Selected method**
Loss model	Deficit and constant loss model
Transformation model	SCS curve number transformation model
Baseflow model	Recession baseflow model
Canopy model	Simple canopy model
Surface model	Simple surface model
River routing model	Muskingum model

3.11 1D River Modelling in HEC-RAS

HEC-RAS 1D unsteady-state modelling is a powerful hydrodynamic simulation that uses mass and momentum conservation equations to simulate the flow in open channels under a non-steady state (Brunner, 2024). HEC-RAS 1D unsteady flow is based on the Saint-Venant equations which include the continuity and momentum equations to predict the flow rate variation, depth and velocity along a river reach. In this regard, the continuity equation (or mass conservation equation) guarantees that the storage variation in the channel section corresponds to the inflow and outflow rates in a segment (Arash & Yasi, 2023; Pathan & Agnihotri, 2021). The basic form of the continuity equation used in HEC-RAS is:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q \qquad \text{(Equation 11)}$$

where A represents the cross-sectional area, Q is the flow rate, and q is the lateral inflow per unit length of the channel.

The momentum equation, on the other hand, represents the conservation of momentum, including terms for gravity, pressure forces, and friction. In its simplified form, the momentum equation for 1D unsteady flow in HEC-RAS is given by:

$$\frac{\partial Q}{\partial t} + \frac{\partial Qv}{\partial x} + gA\frac{\partial h}{\partial x} + S_f = 0 \qquad \text{(Equation 12)}$$

where v is the cross-sectional velocity, h is the water surface elevation, g is gravitational acceleration, and S_f represents the friction slope, calculated using Manning's equation.

To solve these equations HEC-RAS uses either the implicit finite difference or finite volume schemes. Preissmann's scheme used in the finite difference method offers stability and convergence in gradually varied flows and is ideal in channels with subcritical flows (Arash & Yasi, 2023; Pathan & Agnihotri, 2021). This method simplifies the nonlinear Saint-Venant equations to ensure numerical stability in the computational process. Also, there is a semi-implicit finite volume method that deals with rapidly varying flow more efficiently and can be important in transitional flow conditions when both subcritical and supercritical flows are observed.

3.12 WASP Model Development

WASP is a versatile mass-balance modelling system frequently used to model the transport of contaminants in aquatic systems. WASP is designed in a modular fashion that allows it to simulate various environmental processes in different compartments of surface water, sediment, and the water column by advective and dispersive transport mechanisms (Ambrose et al., 2011). This makes WASP useful for modelling different water bodies such as rivers, streams, lakes, and estuaries and modelling a wide range of pollutants from nutrients to toxicants such as heavy metals (Wool et al., 2020). WASP is particularly useful in assessing pollutant dynamics under different scenarios by simulating how pollutants disperse, transform, and interact with other environmental factors.

WASP can predict contaminant concentration trends and their potential impacts on aquatic life and human health. Heavy metal simulation, for instance, is one of the key applications in WASP, particularly through its advanced toxicant modelling (Wimordi et al., 2021). The WASP framework employs mass balance equations that are the basis of the model's transport and transformation computations. Advection and dispersion are the main transport mechanisms while kinetic reactions are the various transformation processes. The advanced toxicant model can simulate the conditions for both dissolved and particulate-bound heavy metals. In aquatic systems, metals interact with suspended particles or sediments, altering their bioavailability, toxicity, and transport.

The WASP uses a segmentation network to solve the conservation of momentum, energy, and mass equations and simulate the transport of contaminants and sediment. The river is divided into segments to simulate the contaminant transport (Prajapathi et al., 2023). The water flow through these segments can be fed into the model from external hydrodynamic models like HEC RAS. The Stream Routing method was used as the transport mode in WASP to simulate the water flow in the WASP model. The stream routing method uses specified flows for each segment to simulate the water flow. The volume, depth, and velocity are adjusted for variable flow based on the hydro geometry (Wool et al., 2020).

3.13 Objective Functions

In the case of classification tasks, F1-score and Area Under the Curve (AUC) functions can be used to evaluate the results (Han & Wang, 2021). A True Positive (TP) occurs when the RF model correctly identifies an instance of pollution when it is present. Conversely, a True Negative (TN) is recorded when the model accurately determines the absence of pollution when no pollution is present. A False Positive

(FP) is identified when the model incorrectly predicts pollution in an instance where no pollution exists, while a False Negative (FN) is noted when the model fails to detect pollution despite its actual presence.

F1-score statistics were used to capture the RF model's precision rate and sensitivity.

$$Precision = \frac{TP}{(TP + FP)} \qquad \text{(Equation 13)}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \qquad \text{(Equation 14)}$$

$$F1 - score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \qquad \text{(Equation 15)}$$

A Receiver Operating Characteristic (ROC) curve is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate (TPR) and False Positive Rate (FPR). AUC is the area under the ROC curve. It provides an aggregate measure of performance across all possible classification thresholds. An AUC of 1 represents perfect classification, while an AUC of 0.5 represents an entirely random classification.

$$TPR = \frac{TP}{(TP + FN)} \qquad \text{(Equation 16)}$$

$$FPR = \frac{FP}{(FP + TN)} \qquad \text{(Equation 17)}$$

The Nash–Sutcliffe efficiency (NSE), the Root Mean Squared Error (RMSE), and the Root Mean Absolute Error were used to evaluate the performance of the LSTM ANN models and the process-based models in this study. The NSE value assesses how well the predicted data series matches the observed data (Jaffar et al., 2022). When O_i and P_i represent the sample (of size N) containing the calculated and predicted values from the model, respectively, and $\hat{o}$ is the mean of the calculated values, the NSE value is given by:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \hat{o})^2} \qquad \text{(Equation 18)}$$

Root Mean Squared Error (RMSE) assesses the average difference between predicted and actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \qquad \text{(Equation 19)}$$

Root Mean Absolute Error (RMAE) is useful for analysing error magnitude while being sensitive to the scale of errors in the same way as RMSE.

$$RMAE = \sqrt{\frac{\sum_{i=1}^{N}|O_i - P_i|}{N}} \qquad \text{(Equation 20)}$$

The Mean Absolute Percentage Error (MAPE) can be used to assess the performance of two data series and to calculate the mean deviation of the two data series (Chicco et al., 2021). When the sample size is N, Y_i represents the reference value and X_i is the value which is compared, The MAPE is given by,

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - X_i}{Y_i}\right| \qquad \text{(Equation 21)}$$

3.14 Central Environmental Authority Guidelines

The National Environmental Act No. 47 of 1980 is the legislation established by the CEA for protecting and managing the environment in Sri Lanka. The 19th amendment of the act was issued in 2019, and the guidelines for Ambient Water Quality suggested in the guidelines are used as the local guideline to assess inland waterways in Sri Lanka (Central Environmental Authority, 2019). The guidelines suggest six categories to assess the water quality (Table 3.6).

Table 3.6 Central Environmental Authority (CEA) guidelines for ambient water quality in inland surface waterways in Sri Lanka

Parameter	**Unit**	**Category A**	**Category B**	**Category C**	**Category D**	**Category E**	**Category F**
EC	µS/cm, max	-	-	-	-	700	-
Turbidity	NTU, max	5	-	-	-	-	-
DO	mg/l, min	6	5	5	4	3	3
Phosphate	mg/l, max	0.7	0.7	0.4	0.7	-	-
Nitrate	mg/l, max	10	10	10	10	-	10
Pb	mg/l, max	0.05	-	0.002	0.05	-	-
Cd	mg/l, max	0.005	-	0.005	0.005	-	5
Fe	mg/l, max	1	-	-	2	-	-
T. Coli	MPN/100ml, max	10,000	10,000	-	10,000	-	-

- Category A shall be water that requires simple treatment, for drinking.
- Category B shall be bathing and contact recreational water.
- Category C shall be water suitable for aquatic life.

- Category D shall be the water source that is required to undergo a general treatment process, for drinking.
- Category E shall be water suitable for irrigation and agricultural activities.
- Category F shall be water with minimum quality but does not fall into categories A to E

CHAPTER 4

4 RESULTS

This chapter presents the results obtained under each objective. Statistical analysis of water quality parameters including Factor Analysis, Cluster Analysis, and Pollution Indexing are presented first. The findings are then linked with the river flow and pollutant transport modelling to identify the potential pollutant sources.

4.1 Analysing Water Quality Variations

The seasonal and spatial variations of water quality in the Kelani River were analyzed using unsupervised learning methods. This section presents the results of this analysis.

4.1.1 Factor Analysis and Latent Variables

Reducing the dimensionality of the dataset with 17 water quality parameters at 17 measuring points is crucial for simplifying analysis and easier interpretation while retaining the essential information in the original dataset. The intercorrelation analysis among the water quality parameters identifies groups of parameters that tend to vary together. These groups can then be represented by LVs, reducing the dimensionality of the dataset. The Pearson correlation matrix was utilised to examine the primary parameters driving spatiotemporal variation of the entire dataset. From this analysis, 10 parameters were identified as significant: EC, Turbidity, Temperature, DO, Phosphate, Nitrate, Pb, Fe, T. Coli, and Cd (Figure 4.1).

The 10 correlated water quality parameters were then used to find the LVs using the FA. The results of the FA identified five LVs contributing to the spatiotemporal variation of water quality from the scree plot (Table 4.1, Figure 4.1, Figure 4.2). A KMO value of 0.61 and Bartlett's test p-value of 0.01 ($p < 0.05$) were obtained. Varimax rotation (Islam et al., 2017; Mei et al., 2014) was used to rotate the factor loading to interpret data meaningfully. Accordingly, the obtained five LVs with eigenvalues larger than 1 contributed to 77% of the total variance of the water quality dataset.

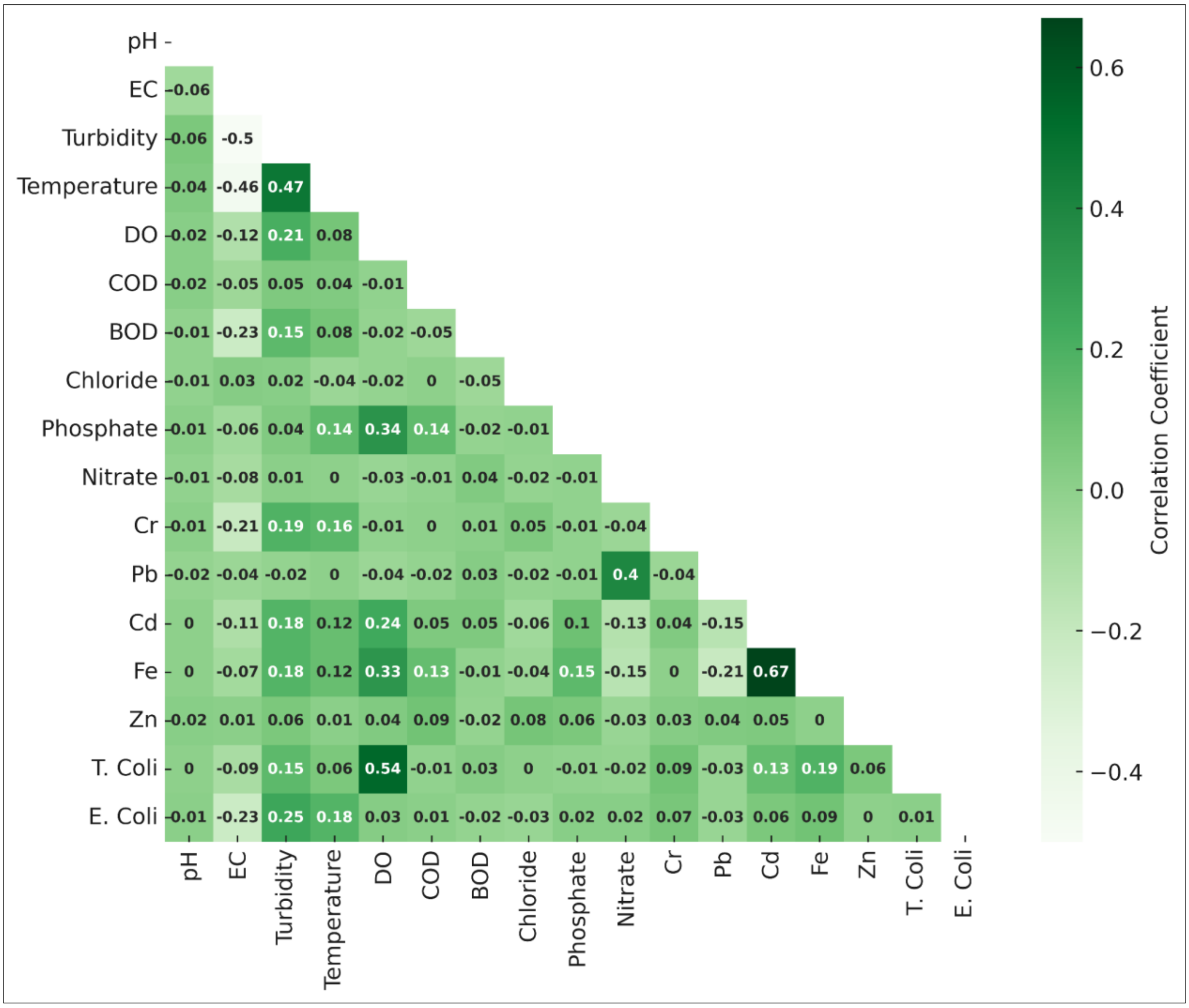


Figure 4.1 Correlation coefficients identified in Pearson's Correlation test

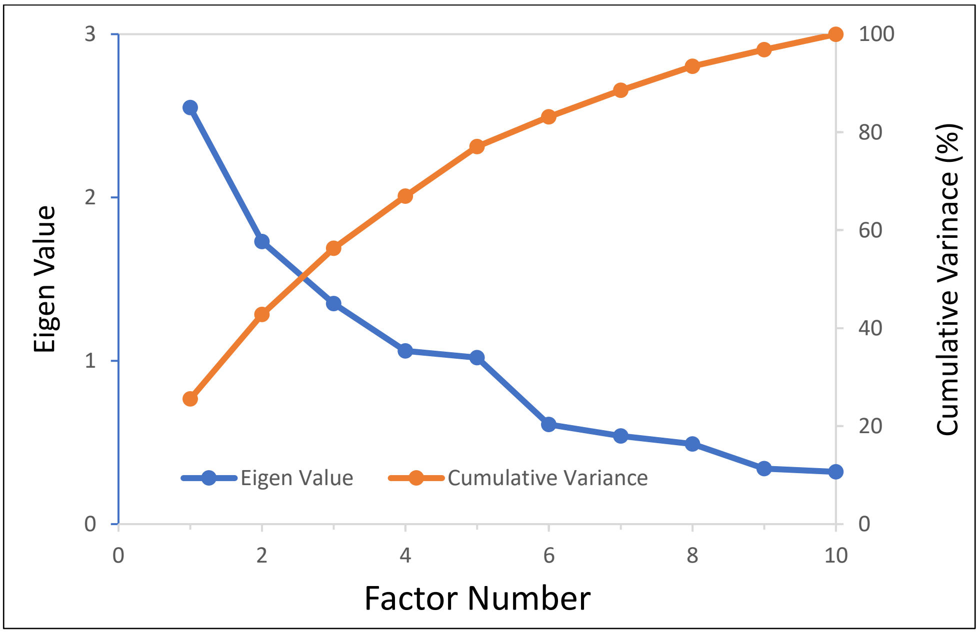


Figure 4.2 Scree plot used to identify the ideal number of factors to be retained along with the cumulative variance

Table 4.1 Factor loadings for the rotated latent variables identified by the factor analysis

Parameter	Latent Variable 1	Latent Variable 2	Latent Variable 3	Latent Variable 4	Latent Variable 5
EC	-0.69	-0.03	-0.06	-0.07	-0.01
TURB	0.71	0.12	0.14	-0.01	-0.03
TEMP	0.66	0.06	0.00	-0.01	0.09
DO	0.08	0.16	0.95	-0.01	0.25
PO4-3	0.06	0.08	0.08	0.00	0.99
NO-3	0.03	-0.03	-0.02	0.92	-0.01
Pb	0.01	-0.15	-0.01	0.43	0.00
Cd	0.13	0.64	0.12	-0.13	0.04
Fe	0.07	0.97	0.16	-0.14	0.05
T. Coli	0.08	0.10	0.56	-0.02	-0.07
Variance (%)	25.55	17.25	13.54	10.58	10.17
Cumulative Variance (%)	25.55	42.80	56.34	66.92	77.08
Eigen Value	2.55	1.73	1.35	1.06	1.02

4.1.2 Spatial Variations of Water Quality

Exploring the spatial variation of water quality parameters can unveil potential pollutant sources within the river basin. This analysis helps identify regions where certain parameters vary significantly, indicating potential pollution hotspots, which may correlate with changes in land use types, such as built-up areas or agricultural land, further influencing water quality dynamics (Cheng et al., 2022; Pak et al., 2021). The five LVs obtained from the FA were clustered individually using HCA and SOM techniques to assess the spatial similarities between the 17 water quality monitoring stations having different land use practices. The results show that the clustering from both methods grouped the stations almost similarly. Hierarchical clustering identified three clusters for LV-1 while two clusters each were identified for the other four LVs (Figure 4.5). The SOM technique also identified three clusters for LV-1 and two clusters each for the other LVs (Figure 4.6).

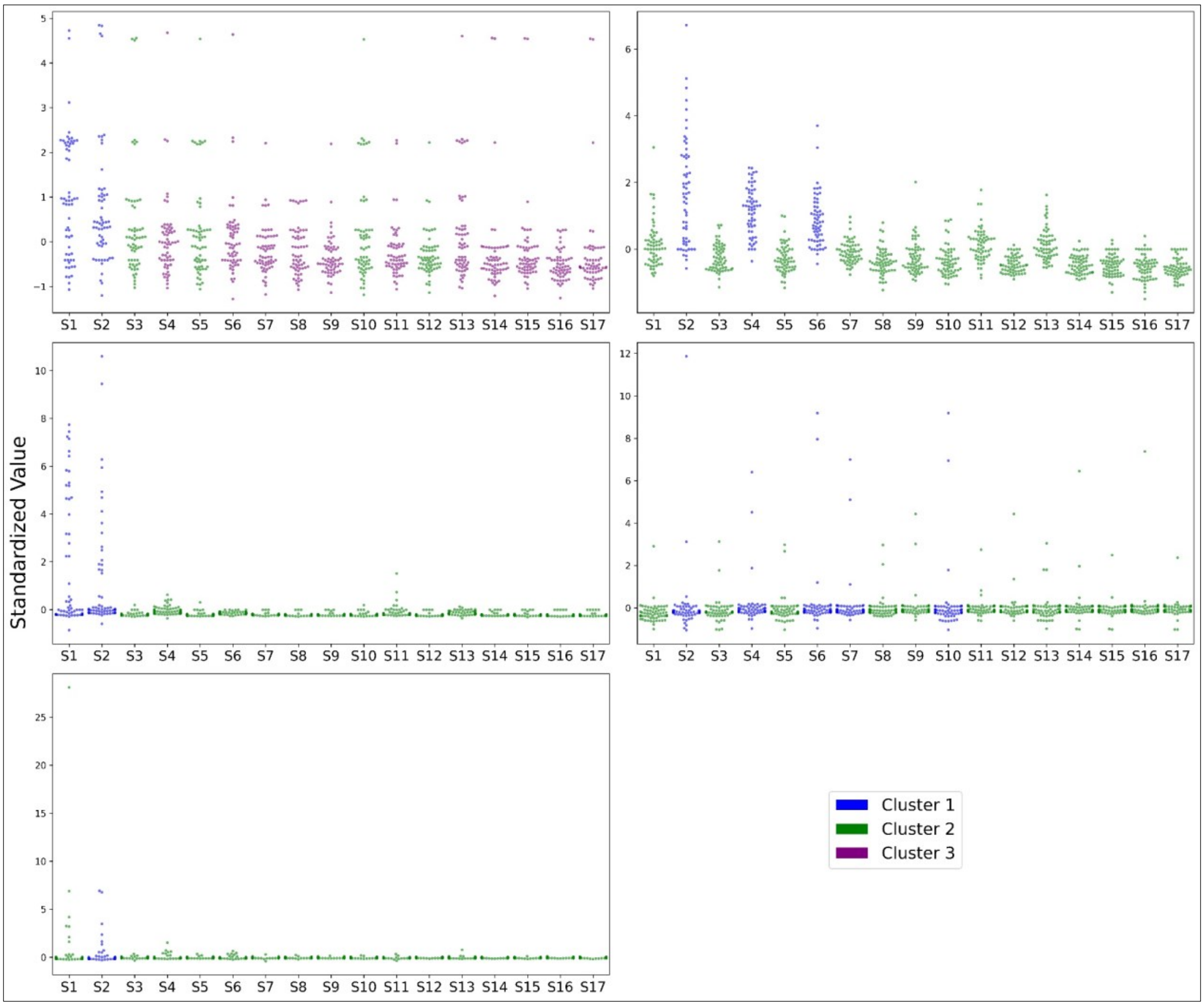


Figure 4.3 Different clusters identified by hierarchical cluster analysis for the five latent variables

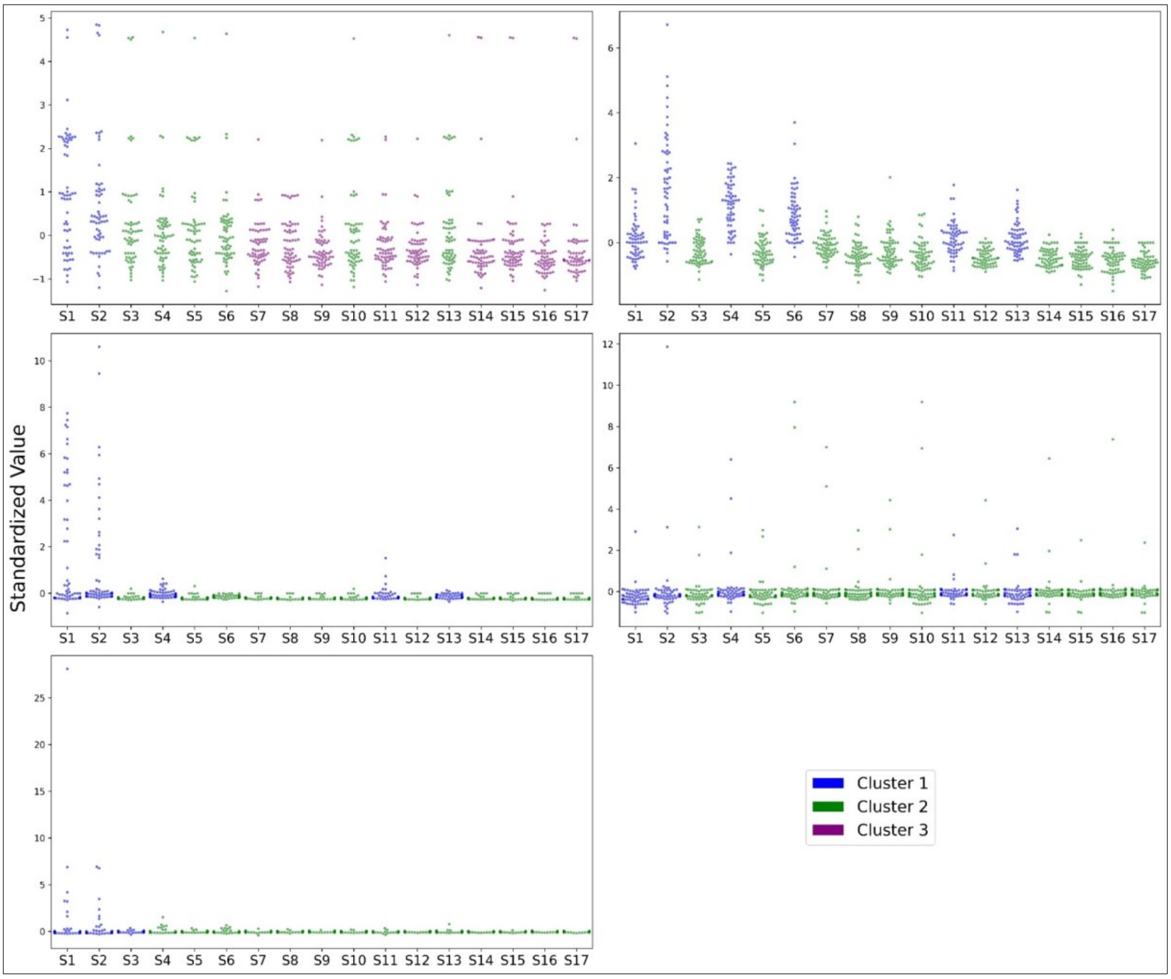


Figure 4.4 Different clusters identified by self-organizing maps for the five latent variables

4.1.3 Seasonal Variations of Water Quality

The CA results show that water quality parameters of certain clusters displayed notable variation during the study period, frequently surpassing national water quality standards. Consequently, it is imperative to investigate the seasonal fluctuations of the selected LVs concerning hydrometeorological variables such as rainfall and river discharge. There are six rain gauge stations in the KRB. Thiessen Polygon method was used to identify the water quality monitoring stations that correspond to each rain gauge station (Figure 3.1). Pearson's Correlation Matrix was then used to correlate the LVs with the measured rainfall (Prathumratna et al., 2008). To assess the correlation of streamflow with LVs, three water quality measuring stations located near the streamflow measuring stations and along the same streams were selected. A correlation analysis was conducted to explore the relationship between LVs and rainfall dependency. Out of six rain gauge stations, the results for the Colombo and Hanwella stations are shown in Figure 4.5, where the latent variables exhibit significant correlations with rainfall variability.

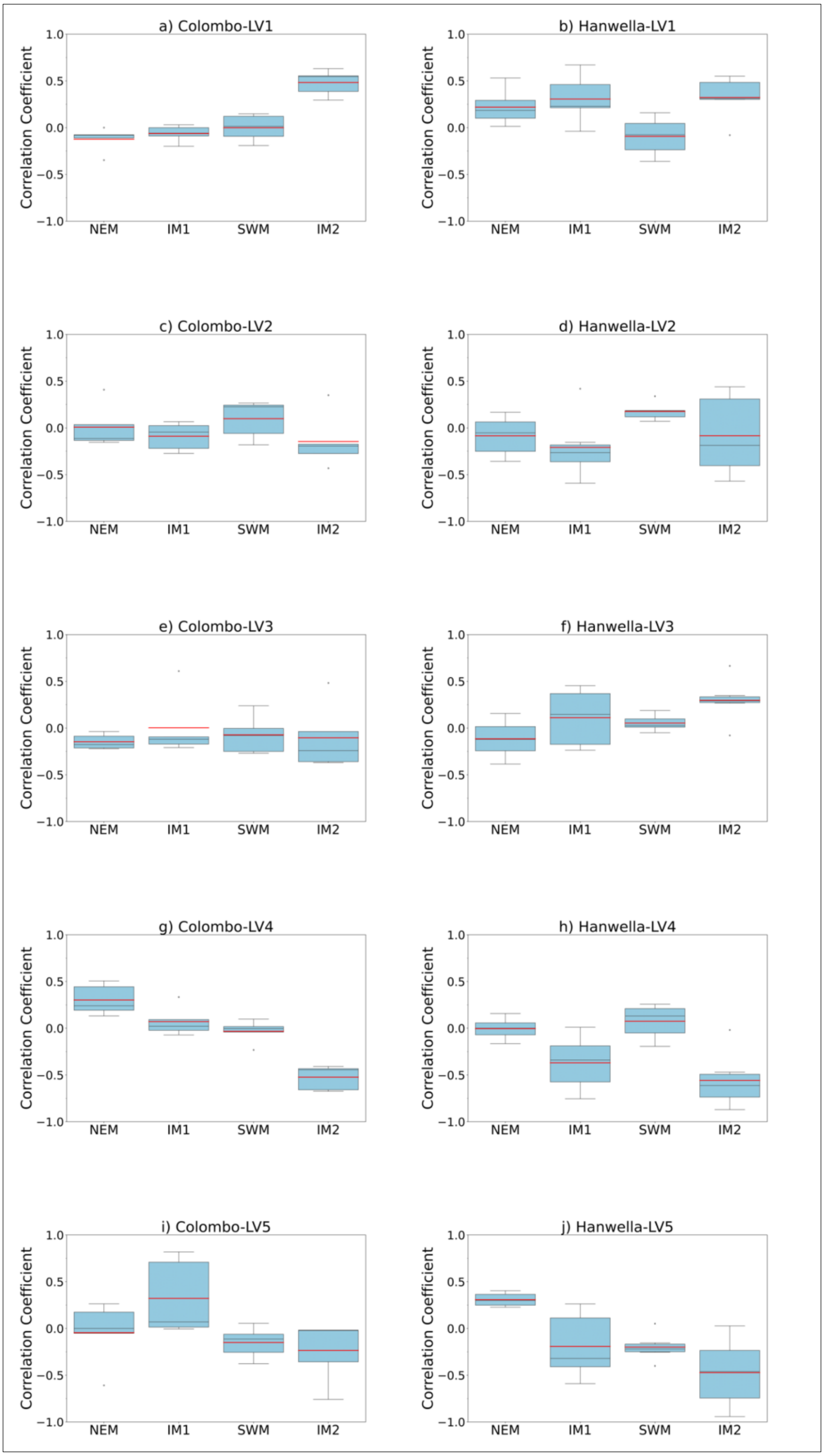


Figure 4.5 Variation in correlation coefficients between rainfall and Latent Variables (LV) during different monsoon seasons

While the correlations between rainfall and LVs were investigated at each water quality monitoring station, a selective analysis approach was employed for streamflow stations. The analysis only included water quality monitoring stations close to the streamflow gauging station. Kithulgala station was analysed with the Kithulgala streamflow station to obtain the correlation coefficients. Similarly, the Seethawake water quality monitoring station was analysed with the streamflow data at Glencourse station and the Victoria Bridge station was analysed with the Nagalagam Street streamflow station to obtain the correlation coefficients.

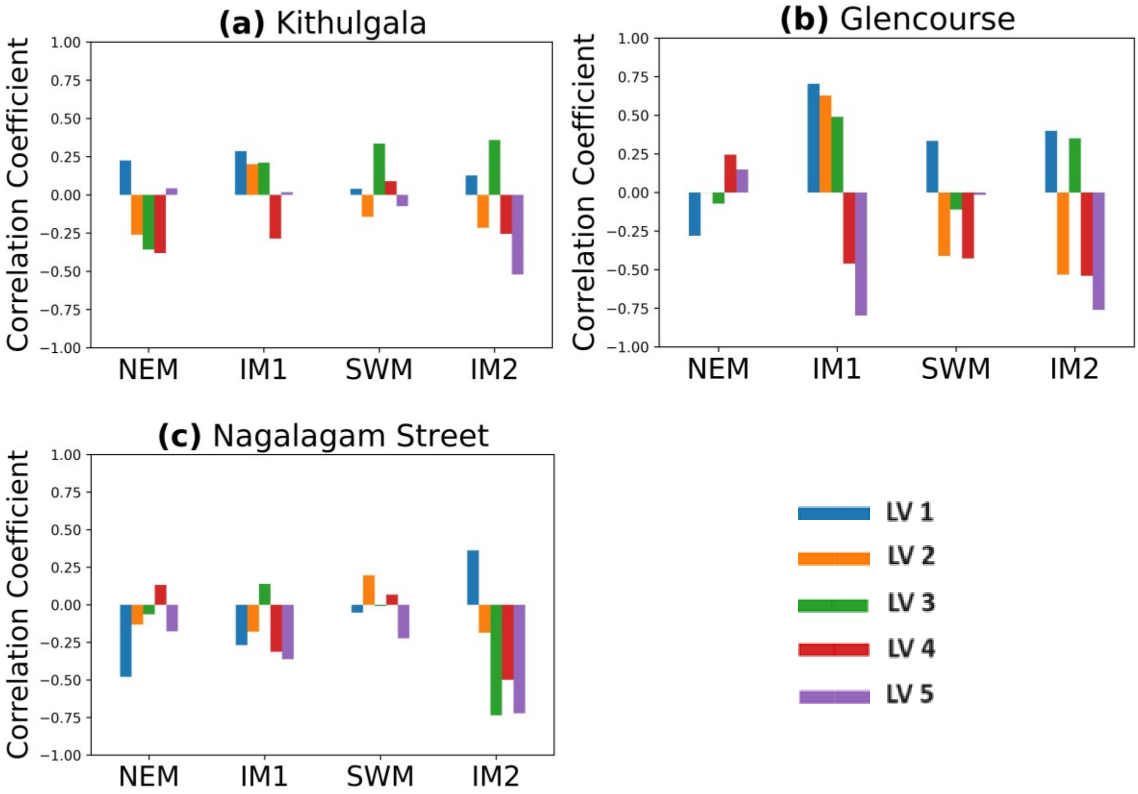


Figure 4.6 Correlation of Latent Variables (LVs) with river discharge in different monsoon seasons

4.2 Water Pollution Indexing

During the seasonal and spatial analysis of water quality variation in the Kelani River, it was identified that pollution due to industrial effluents and sewage-related pollutants were the main sources of pollution in the KRB. Hence two water pollution indices were developed to assess these two types of pollution. The Kelani River Basin-Industrial Pollution Index (KRB-IPI) and the Kelani River Basin-Sewage Pollution Index (KRB-SPI) were developed to assess the industrial and sewage pollution in the Kelani River and to identify the pollution sources in the KRB. Both indices were developed through the basic steps of index development and this section presents the results obtained in the water pollution indexing method.

4.2.1 Selection of Parameters and Sub-index Development

The selection of parameters to develop the indices was based on the FA and literature findings. From the 17 measured water quality parameters Nitrate, Cd, Fe, Cr, Zn, and Pb were selected to develop the KRB-IPI (Azam et al., 2019; Fořt et al., 2022; Kishor et al., 2020). Similarly, for the KRB-SPI DO, COD, BOD, T. Coli, and Phosphate were selected (Ballesté et al., 2020; Mohammed & Al-Obaidi, 2021). The rating curves for these parameters were developed using the CEA water quality standards. The rating curves developed for the six parameters for the KRB-IPI (Figure 4.7) and the five parameters for the KRB-SPI (Figure 4.8).

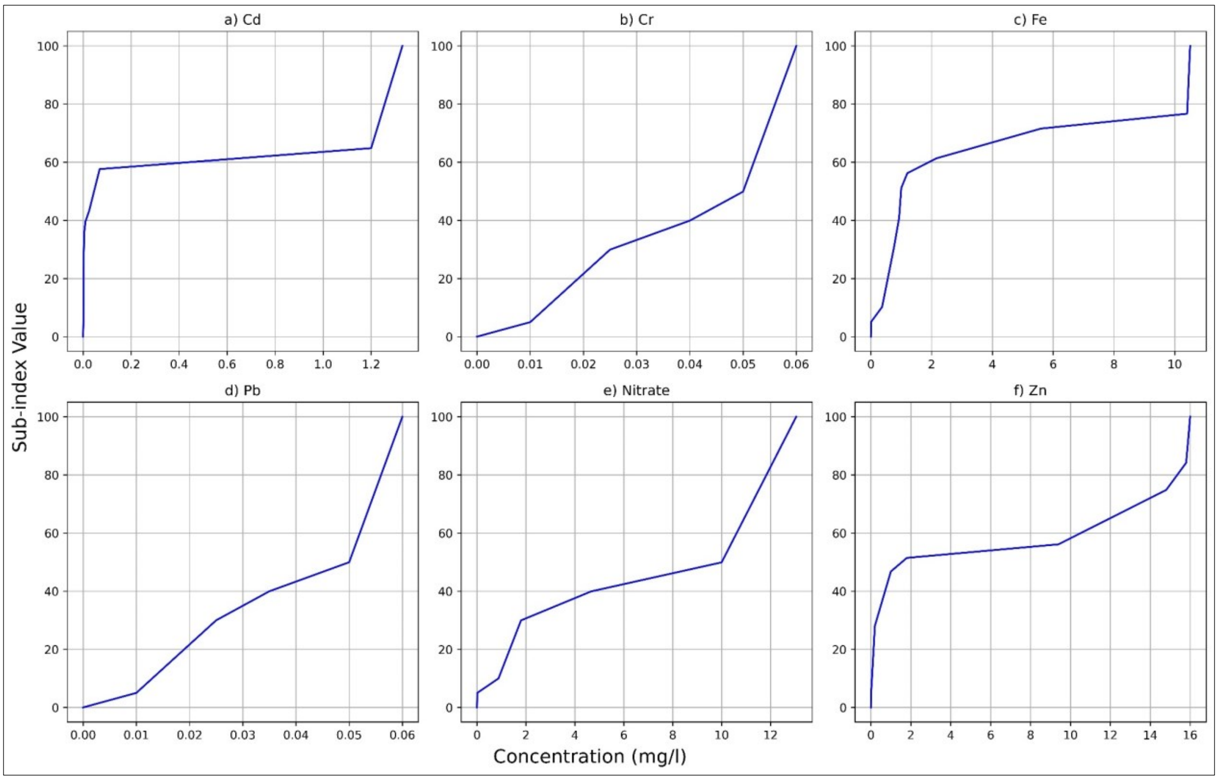


Figure 4.7 Sub-index functions for the Kelani River Basin-Industrial Pollution Index (KRB-IPI)

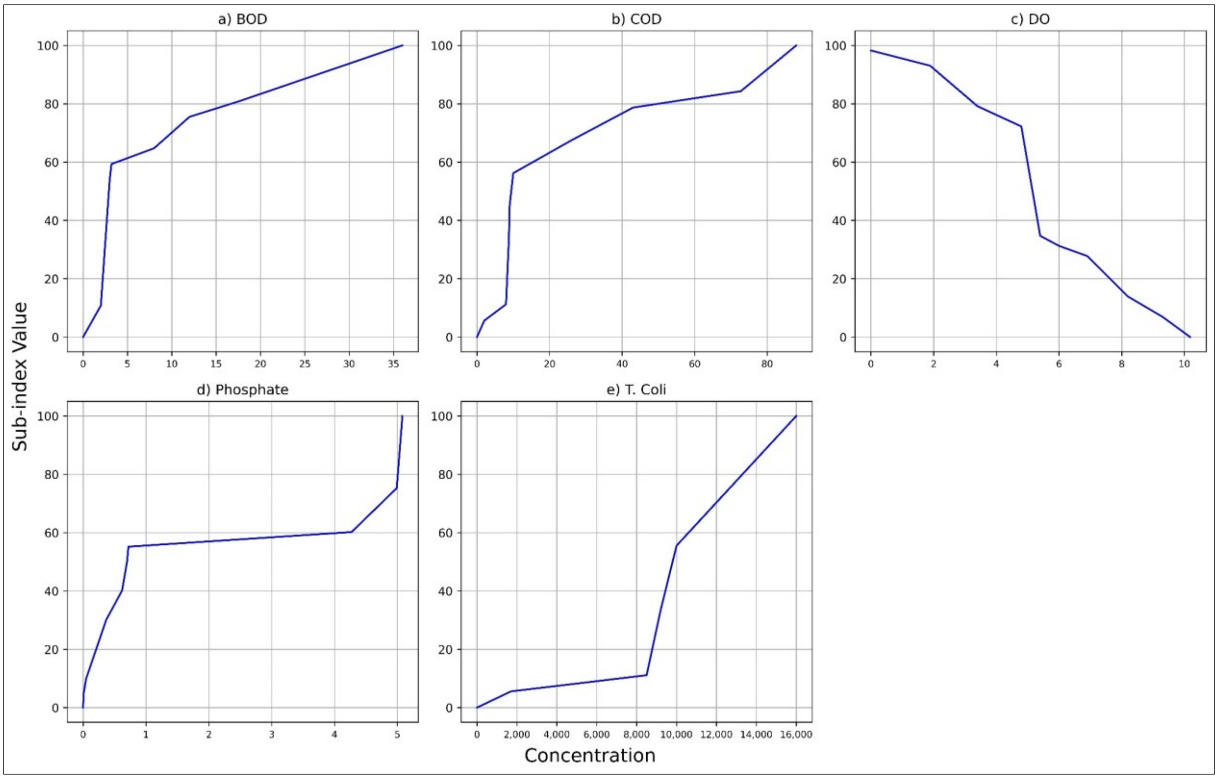


Figure 4.8 Sub-index functions for Kelani River Basin-Sewage Pollution Index (KRB-SPI). The concentrations for BOD, COD, DO, and Phosphate were measured in *mg/l* while the concentration of T. Coli was measured in *MPN/ 1000 ml*

4.2.2 Parameter Weighting

The parameter weighting process was conducted to identify the most influential water quality parameters affecting water pollution in the Kelani River. The Random Forest (RF) algorithm was employed to identify the most significant contributing parameters to KRB-IPI and KRB-SPI, providing a deeper understanding of the parameters under investigation. In the case of KRB-IPI, the RF model achieved an AUC value of 0.98 and an F1-score of 0.99, identifying Cd, Zn, and Fe as the primary parameters influencing river pollution status. Similarly, the RF model which obtained an AUC value of 0.94 and an F1-score of 0.93 identified DO as the most influential parameter affecting the initial pollution status, followed closely by COD and T. Coli. Consistent with these findings, the ROC method also assigned parameter weights in the same order of importance with higher weight values for Cd, Zn, Fe, DO, COD, and T. Coli (Table 4.2).

Table 4.2 Calculated parameter importance scores and the parameter weights

Index	Parameter	Parameter Importance (RF model)	Parameter Weight (ROC)
KRB-IPI	Cd	0.525	0.245
	Fe	0.148	0.232
	Zn	0.154	0.218
	Cr	0.073	0.109
	Pb	0.055	0.107
	Nitrate	0.045	0.089
KRB-SPI	DO	0.395	0.356
	COD	0.337	0.309
	T. Coli	0.258	0.187
	BOD	0.008	0.080
	Phosphate	0.002	0.068

4.2.3 Sub-index Aggregation

The final water pollution values were calculated using the five weighted aggregation functions for both indices. Then, sensitivity analysis were conducted for both indices, considering the parameters to identify the best-fitting aggregation function for the changes in sub-index values. In the case of the KRB-IPI the RSPF 2 (Equation 7) and RSPF 4 (Equation 8) calculated index values exceeding the 0–100 range. Hence, these two functions were not used in the sensitivity analysis for KRB-IPI. Hence the WAA function (Equation 5), the WMF function (Equation 6), and the WAEFF function (Equation 9) were used in the sensitivity analysis. For the KRB-SPI, the calculated index values were within the 0–100 range hence all five functions were used for the sensitivity analysis. The results of the sensitivity analysis for Fe and COD are shown in Figure 4.9 and Figure 4.10, respectively.

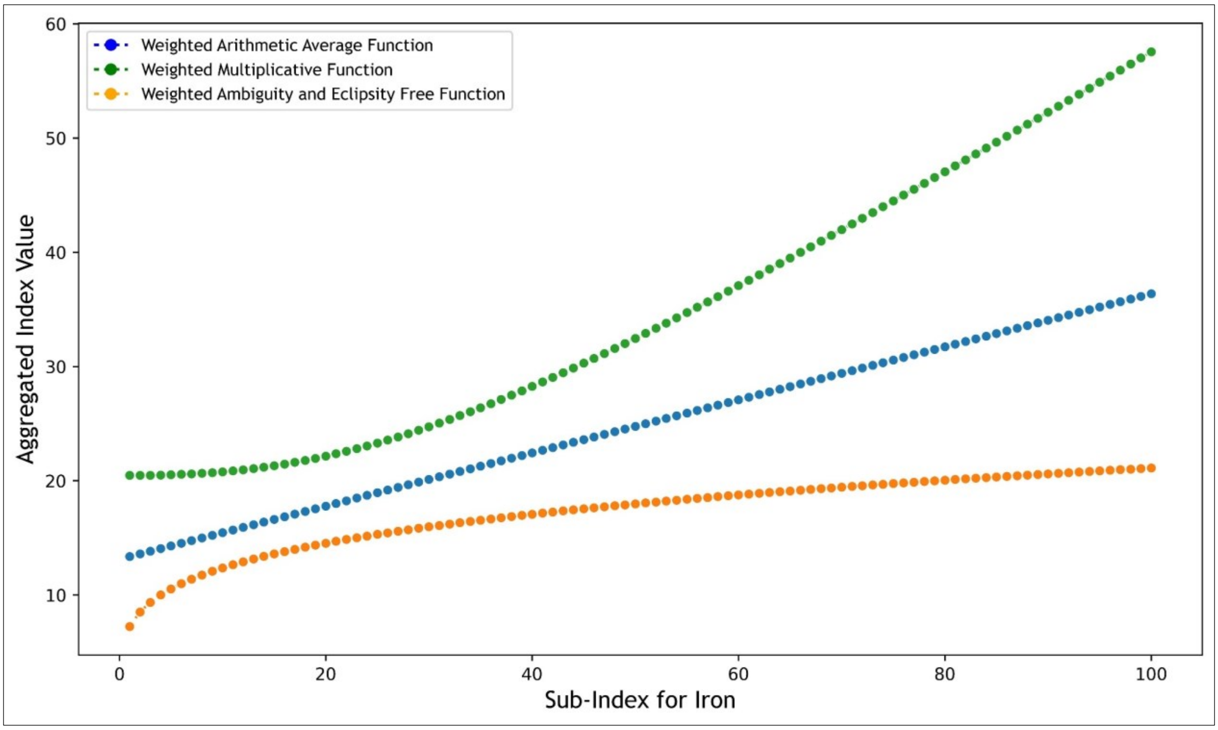


Figure 4.9 Results of the sensitivity analysis for Fe

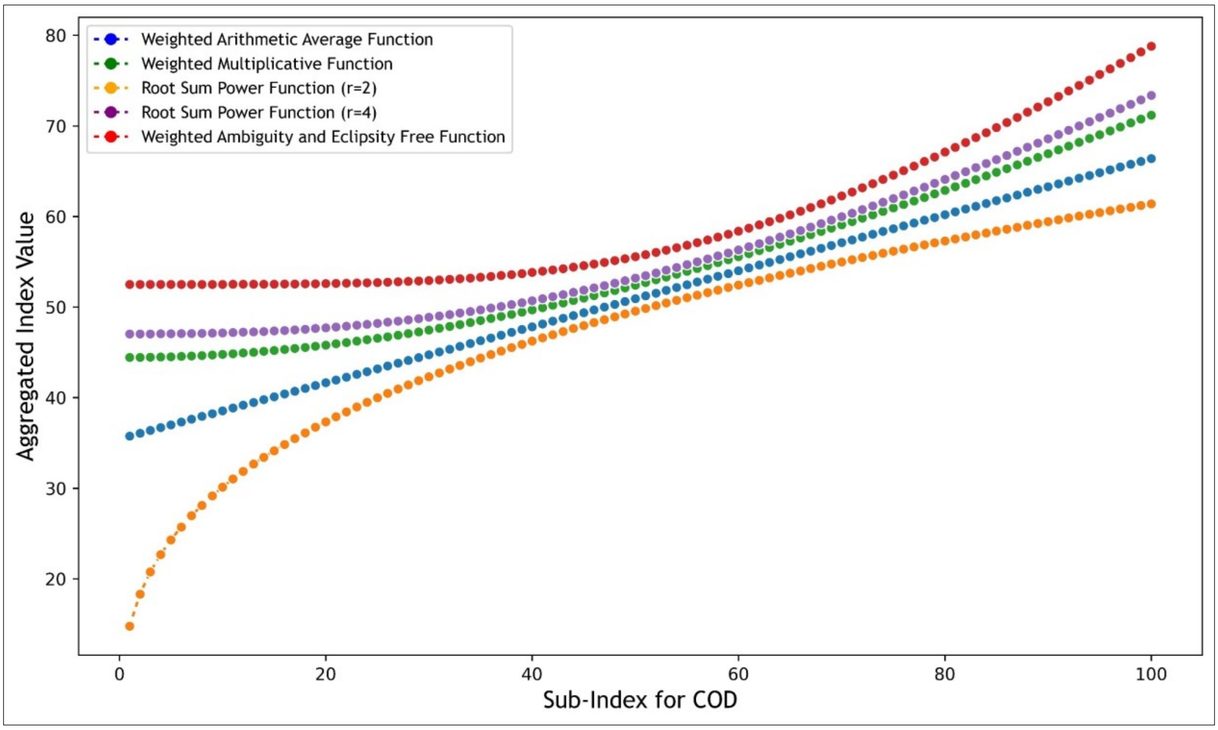


Figure 4.10 Results of the sensitivity analysis for COD

LSTM ANN models for the aggregated KRB-IPI and KRB-SPI were developed to check the consistency of the aggregated index values. About 70% of the calculated index values were used to train the LSTM ANN model, and the rest were used for validation. The input variables (X) for the ANN model were the concentrations of the parameters considered in each index, while the target variable (Y) was the calculated index value. This method checked the consistency of the calculated index with the initial parameter concentrations. The developed LSTM ANN model for the KRB-IPI achieved an NSE value of 0.98 and an RMSE value of 0.81, indicating excellent performance of the developed index. Similarly, the LSTM ANN model for the KRB-SPI achieved an NSE value of 0.99 and an RMSE value of 1.83. The performance of the KRB-IPI model and the KRB-SPI model are shown in detail in Figure 4.11 and Figure 4.12.

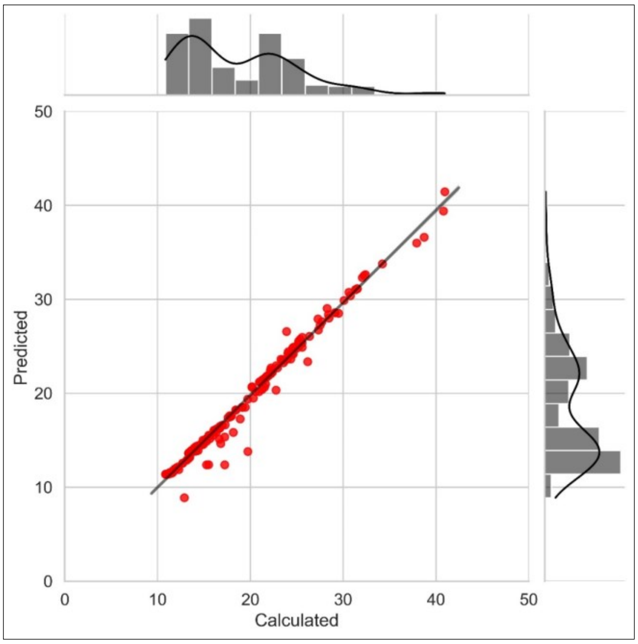


Figure 4.11 Performance of the LSTM ANN model for KRB-IPI. The distribution of the calculated and predicted index values are shown in bar charts

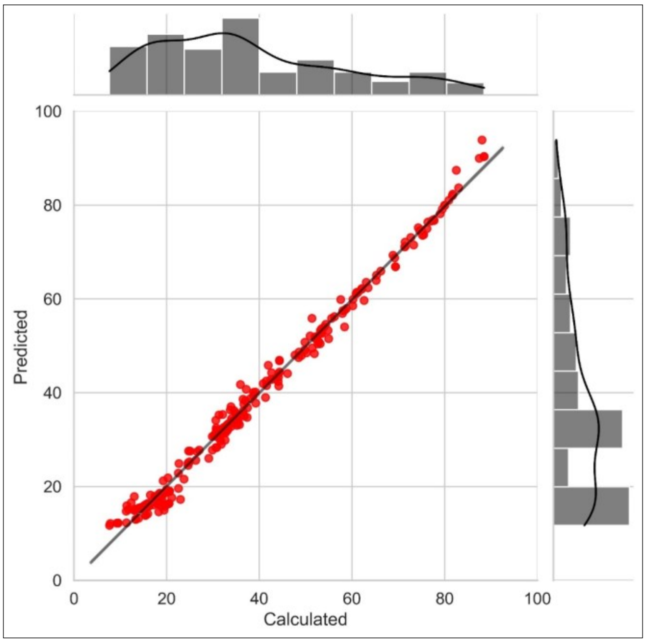


Figure 4.12 Performance of the LSTM ANN model for KRB-SPI. The distribution of the calculated and predicted index values are shown in bar charts

The newly developed indices were classified into five categories using the percentile method (Figure 4.13, Figure 4.14).

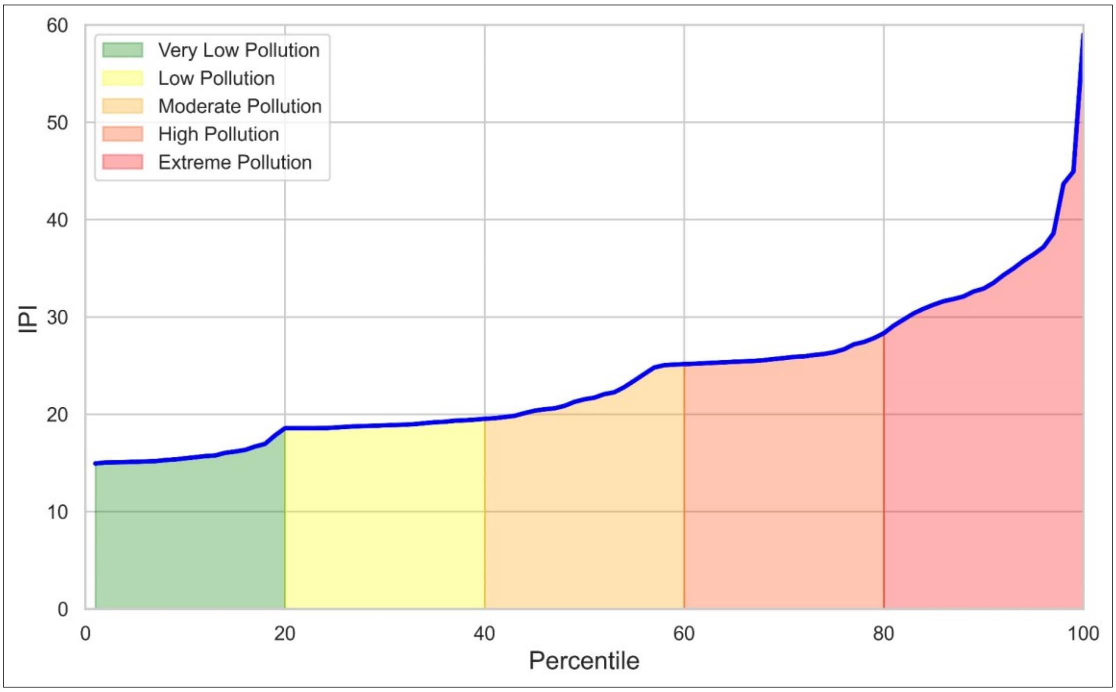


Figure 4.13 Classification of the KRB-IPI into five categories based on percentile values

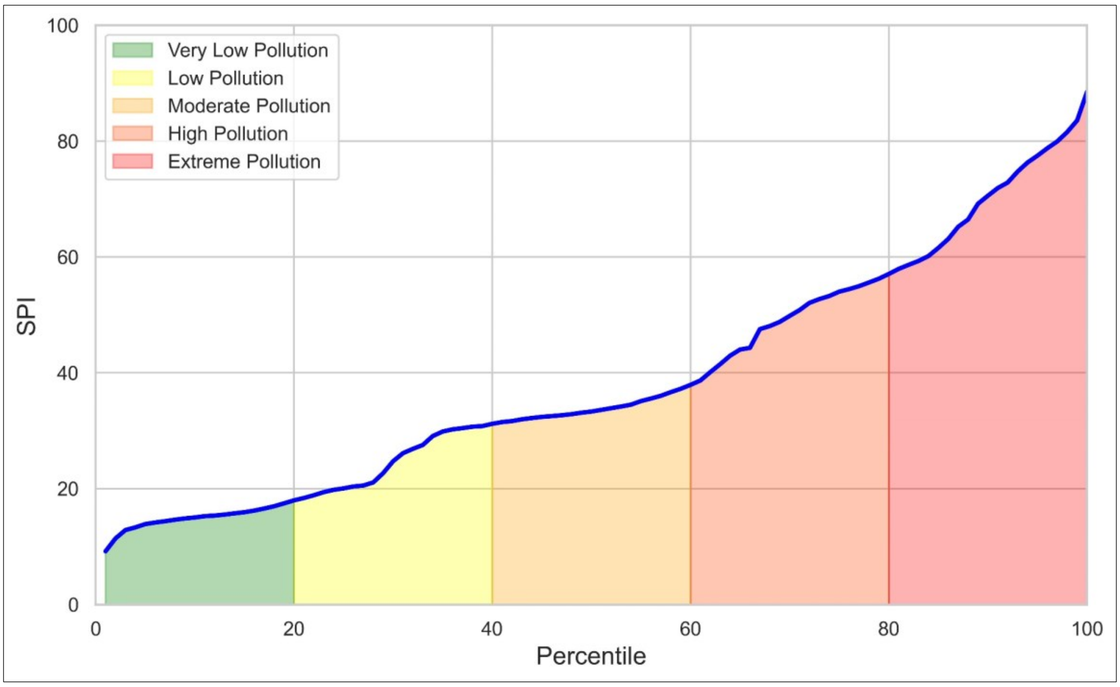


Figure 4.14 Classification of the KRB-SPI into five categories based on percentile values

Then the indices were applied to assess the pollution status of the Kelani River across the four monsoon seasons, providing a clearer understanding of the river's condition (Figure 4.15, Figure 4.16). The calculated pollution values helped in the identification of the pollution sources in the KRB.

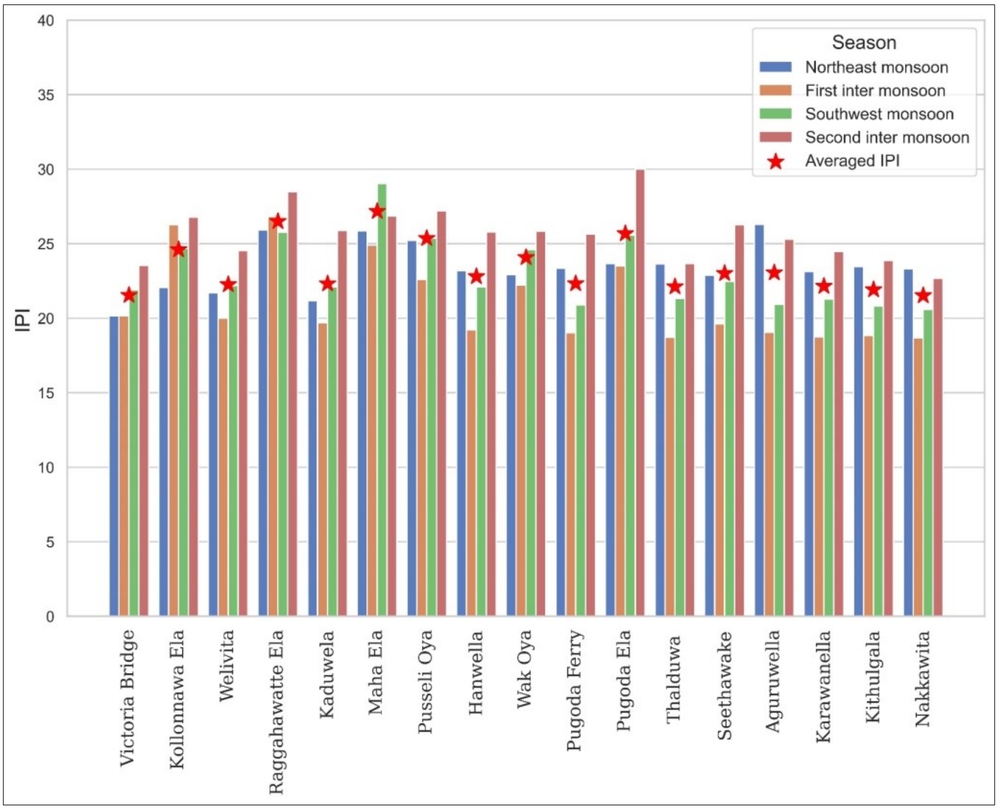


Figure 4.15 Seasonal and annual average water pollution scores for the 17 water quality monitoring stations in the KRB calculated from the KRB-IPI

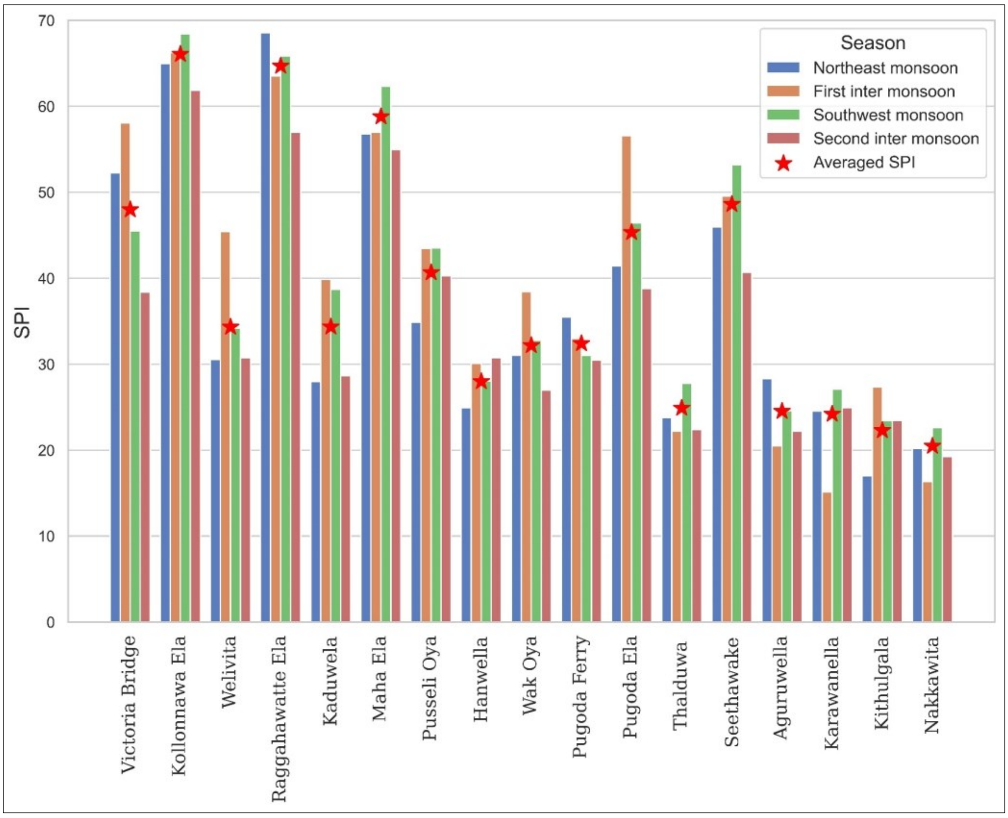


Figure 4.16 Seasonal and annual average water pollution scores for the 17 water quality monitoring stations in the KRB calculated from the KRB-SPI

4.3 River Water Quality Modelling

This section covers the results obtained from the mechanistic modelling approach to assess river pollution in the Kelani River. It was identified that in the spatial and seasonal analysis, heavy metal pollution was present in the basin. Similarly, Fe was identified as a predominant parameter in the KRB affecting river pollution, and a considerable spatial variance was observed for Fe (Figure 4.17). Hence Fe was chosen as the parameter used in the river water quality modelling approach. Subsequently, the spatial and seasonal variation analysis and the pollution indexing methods identified that the stations upstream from Seethawake station (S13) had relatively lesser pollution hence Kelani River from the Seethawake station to the downstream was used in the modelling approach which is referred to as the Lower Kelani River Basin (LKRB) from here onwards.

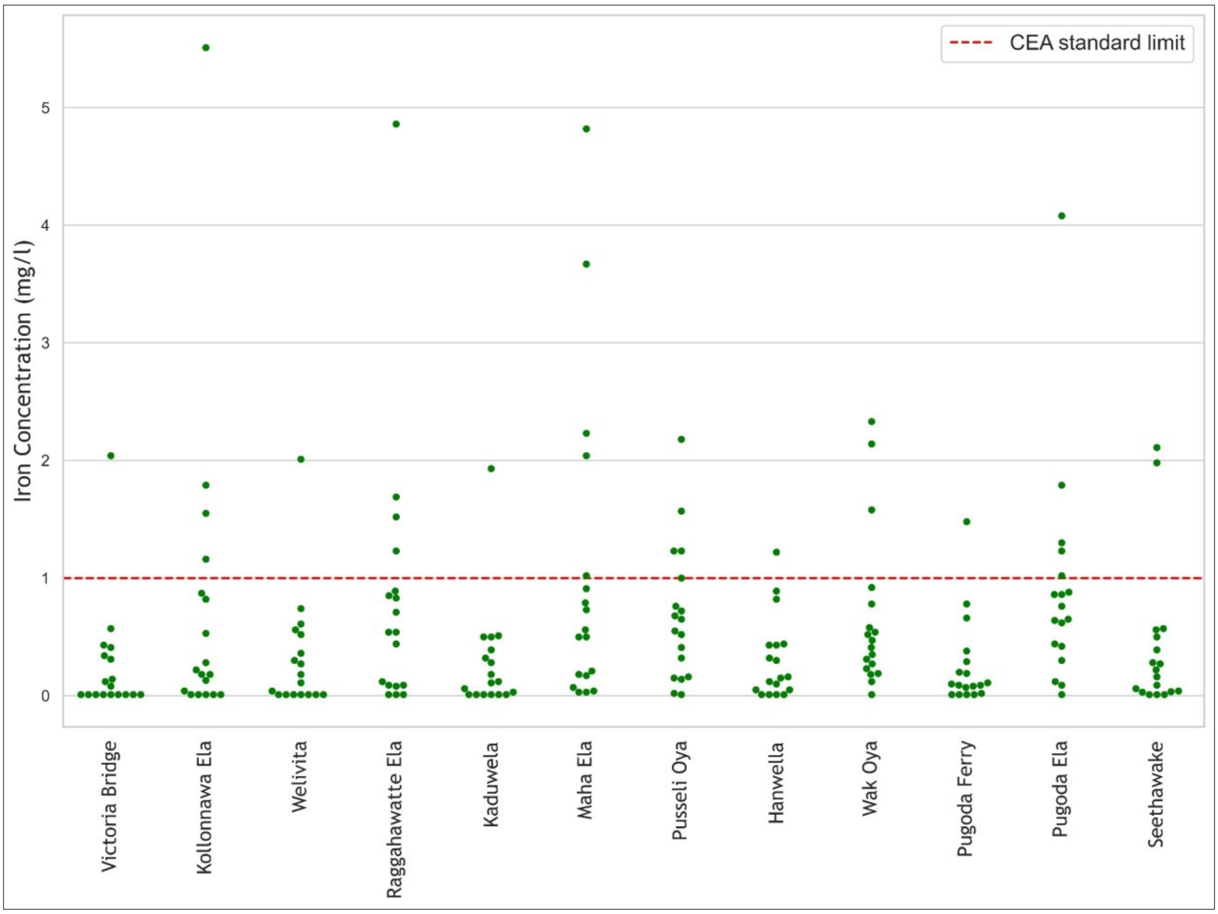


Figure 4.17 Spatial variation of Iron across the stations in the Lower Kelani River Basin (LKRB)

4.3.1 HEC-HMS Modelling

The HEC-HMS model was used to generate the sub-catchment hydrographs necessary for the river modelling. HEC-HMS version 4.10 was used for this study. The KRB from the Glencourse River Station was used for this modelling approach. The observed streamflow measurements at the Glencourse station were assigned as the upstream boundary condition for the model. The Lower Kelani River Basin (LKRB) was divided into 10 sub-catchments for modelling purposes (Figure 4.18).

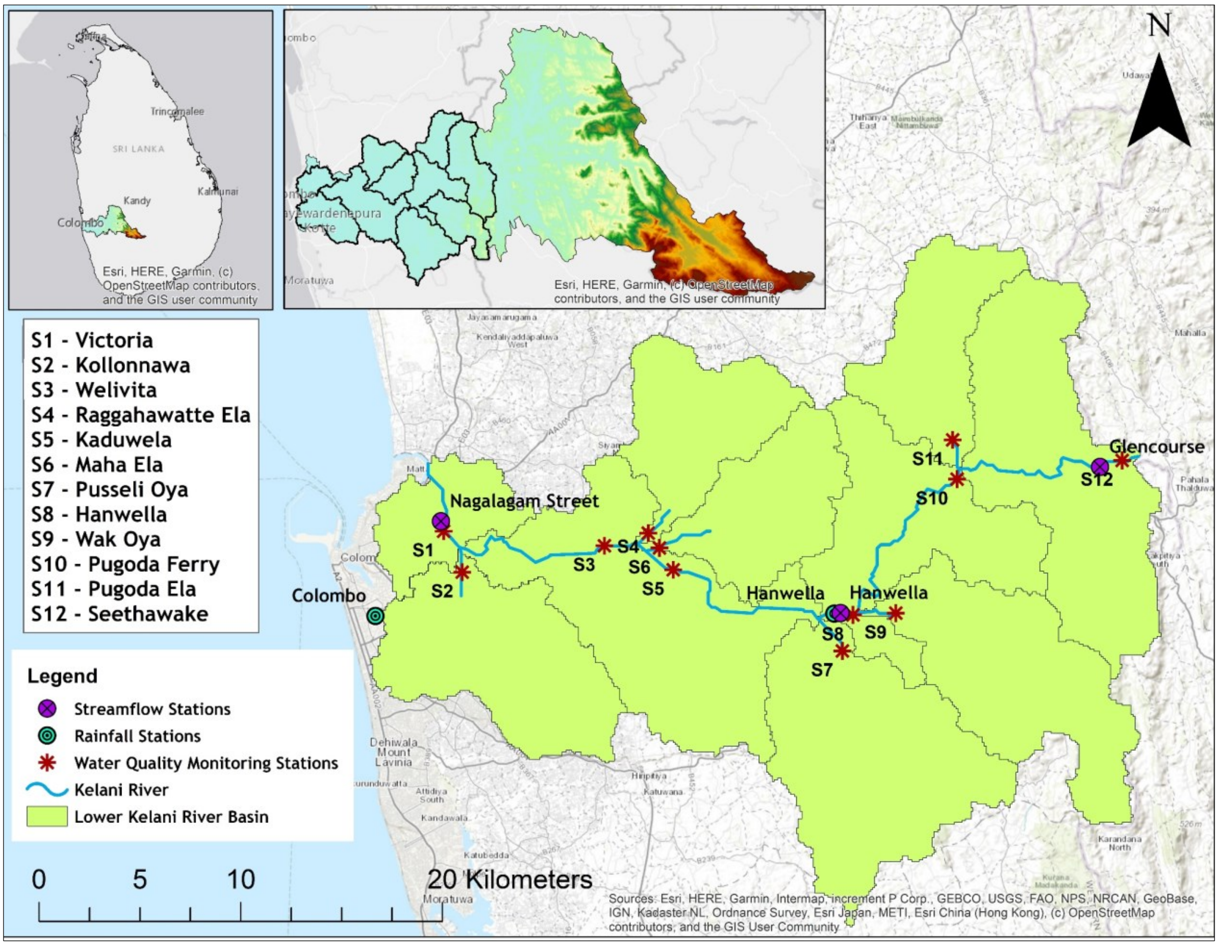


Figure 4.18 Generated sub-basins for the HEC-HMS modelling

The Colombo and Hanwella Rainfall station data were used to develop the model. Similarly, the Glencourse streamflow monitoring station was assigned the upstream boundary condition. The Nagalagam Street monitoring station only monitored water level data hence the Area Ratio method was used to transfer the streamflow observations at Hanwella station to the Nagalagam Street monitoring station (USACE & US Army Corps of Engineers, 1998). The model was calibrated using Hanwella and Nagalagam Street streamflow data observations. The period 2016–2017 was selected as the calibration period and the 2018–2020 period was used as the validation period. Table 4.3 presents the calibrated parameters for the five sub-catchments with water quality monitoring stations. The curve numbers for each sub-catchment were calculated from the Global Curve Number dataset (GCN250) global database (Jaafar et al., 2019).

Table 4.3 Calibrated parameters for the five sub-catchments with water quality monitoring stations

Model	**Parameter**	**Pugoda Ela**	**Pusseli Oya**	**Raggahawatte Ela**	**Maha Ela**	**Kollonnawa Ela**
Canopy	Initial Storage %	0	30	30	30	30
	Max Storage (mm)	20	20	20	20	20
Surface	Initial Storage %	0	0	0	0	0
	Max Storage (mm)	438	386	61	166	144
Deficit and Constant Loss	Initial Deficit (mm)	1	1	1	1	1
	Maximum Deficit (mm)	200	200	200	200	200
	Constant Rate (mm/hr)	7.00	2.64	2.20	1.22	0.77
	Impervious %	7	7	3	7	3
SCS Transformation	Peak Rate Factor	PRF 600	PRF 600	PRF 550	PRF 500	PRF 550
	Lag Time (min)	44.40	5.69	58.23	43.36	44.35
Recession Base Flow	Initial Discharge (m^3/s)	0	35	35	35	35
	Recession Constant	0.960	0.945	0.950	0.955	0.950
	Ratio to Peak	0.55	0.50	0.55	0.55	0.50

The results of the calibrated and validated HEC-HMS model for the Hanwella streamflow station are presented in Figure 4.19, Figure 4.20, and Figure 4.21. The performance of the HEC-HMS model is shown in Table 4.4.

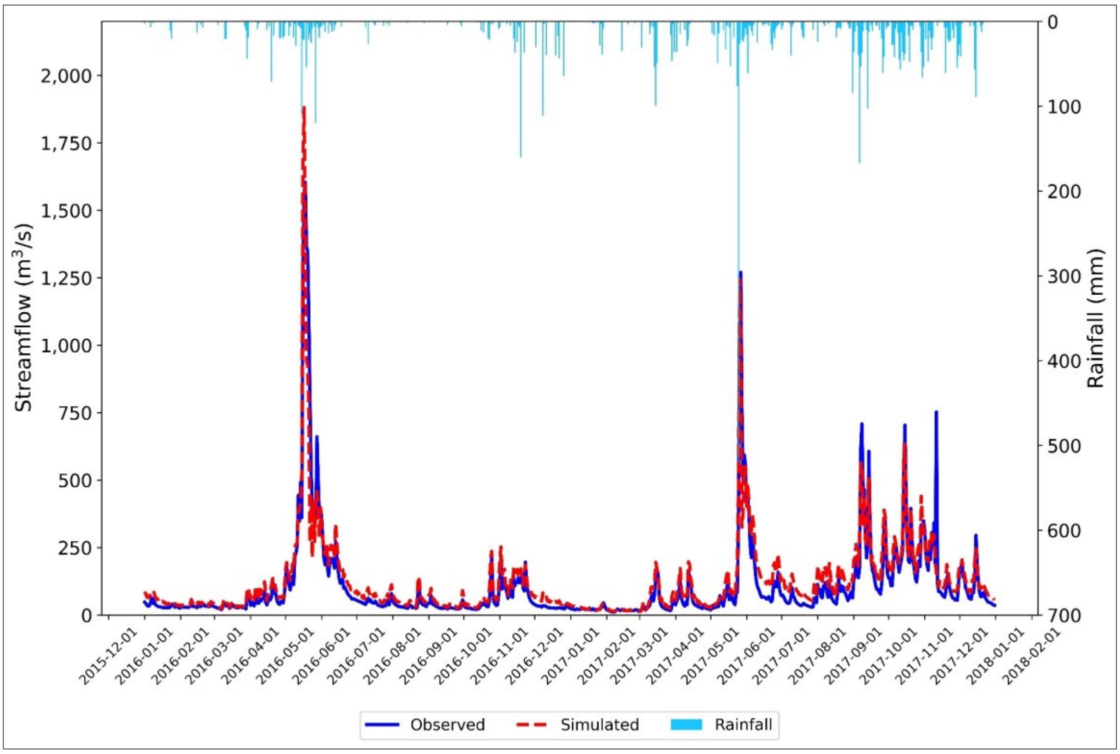


Figure 4.19 Time series plot of simulated and observed streamflow for the calibration period at Hanwella

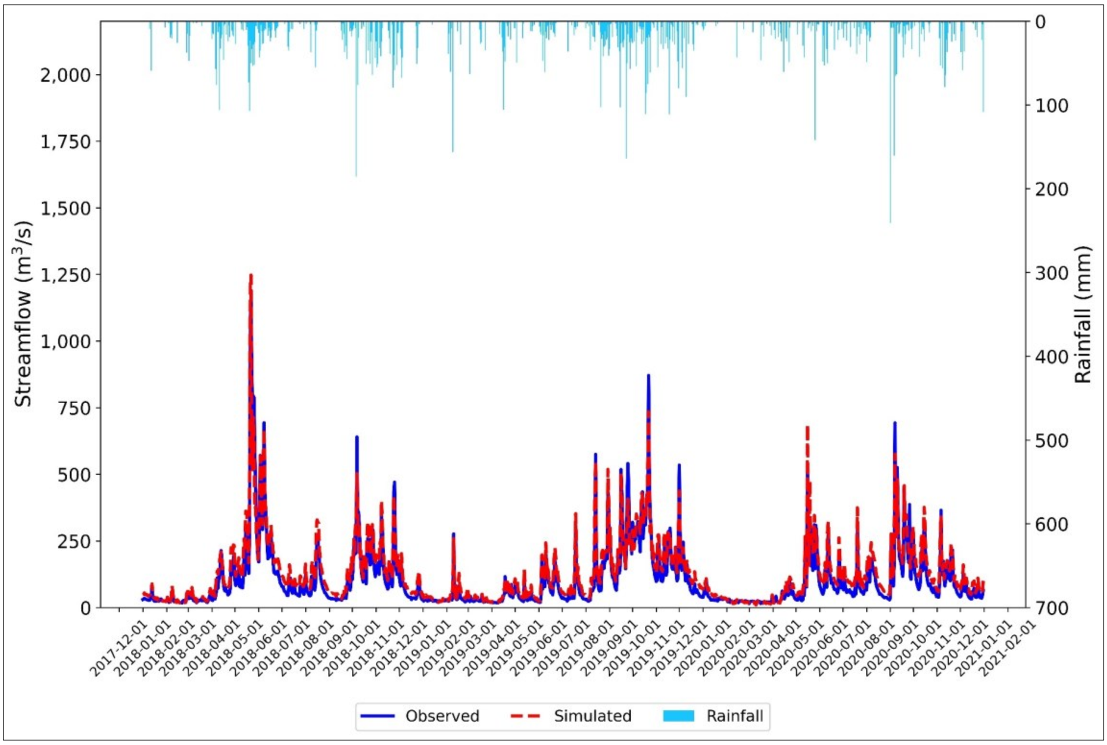


Figure 4.20 Time series plot of simulated and observed streamflow for the validation period at Hanwella

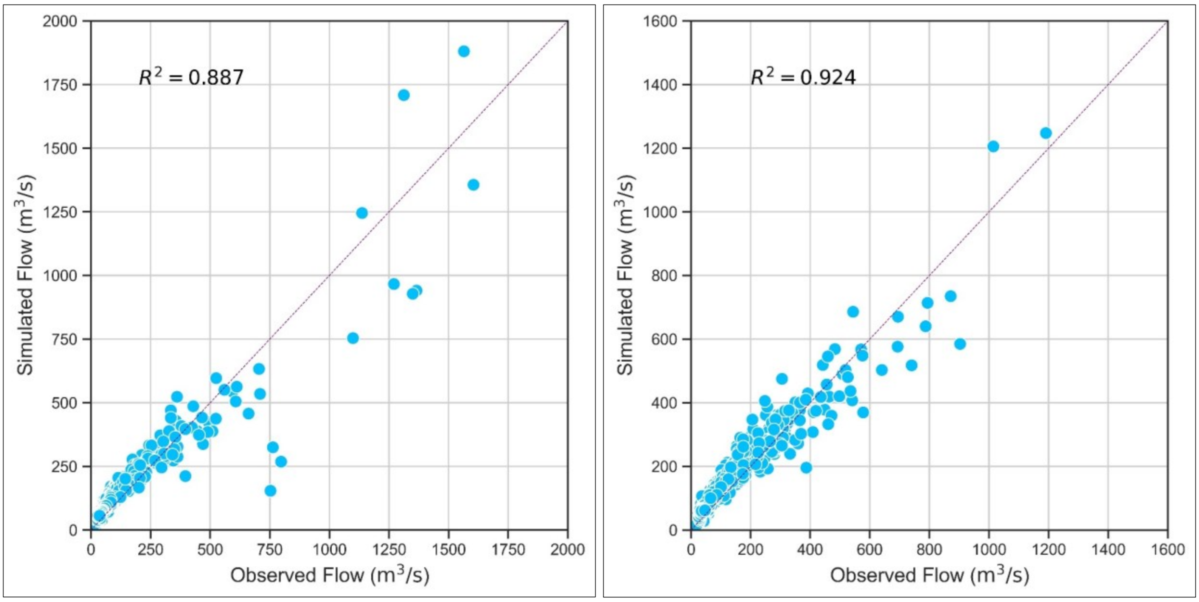


Figure 4.21 Performance of the HEC-HMS model for the calibration and validation periods at Hanwella

Table 4.4 Performance of the HEC-HMS model

Station	**Period**	**NSE**	**RMAE**	**R^2 value**
Hanwella	Calibration period	0.87	5.79	0.89
	Validation period	0.89	5.39	0.92
Nagalagam Street	Calibration period	0.86	5.98	0.91
	Validation period	0.82	5.51	0.89

The calibrated and validated HEC-HMS model generated the sub-catchment hydrographs for the five sub-catchments. The generated sub-catchment hydrographs for Kollonnawa and Pusseli Oya are shown in Figure 4.22 and Figure 4.23.

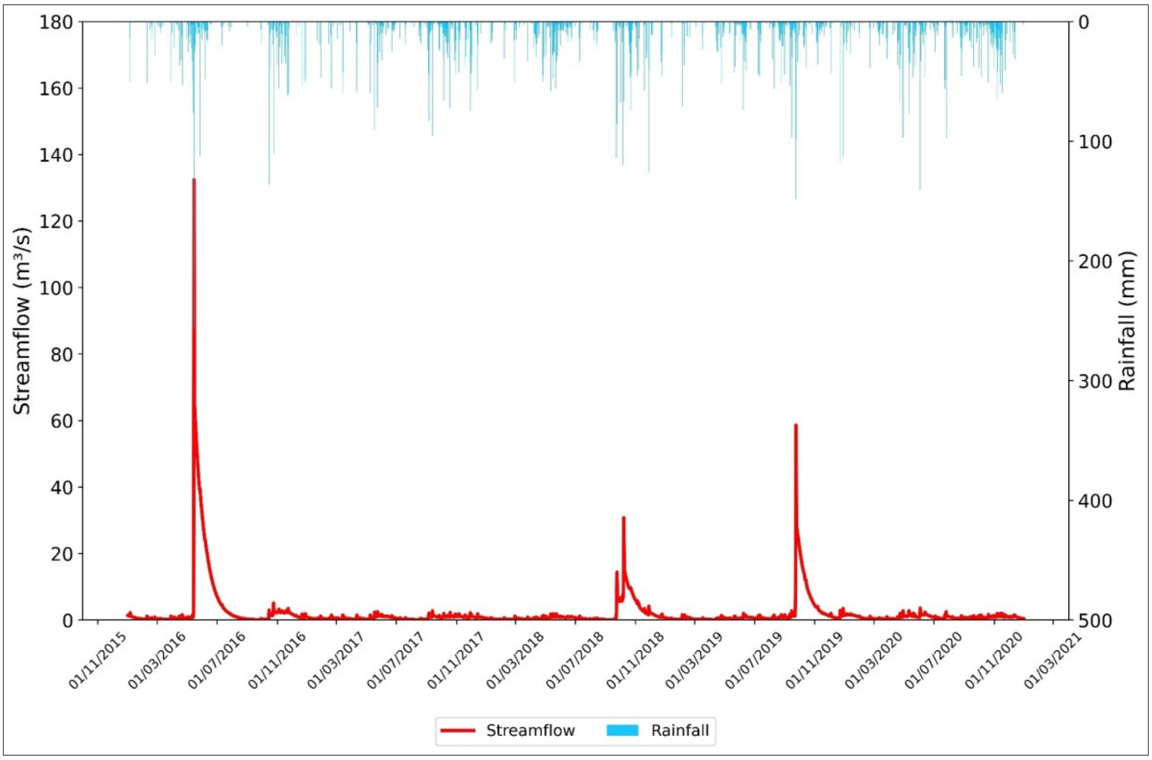


Figure 4.22 Generated hydrograph for Kollonnawa sub-catchment

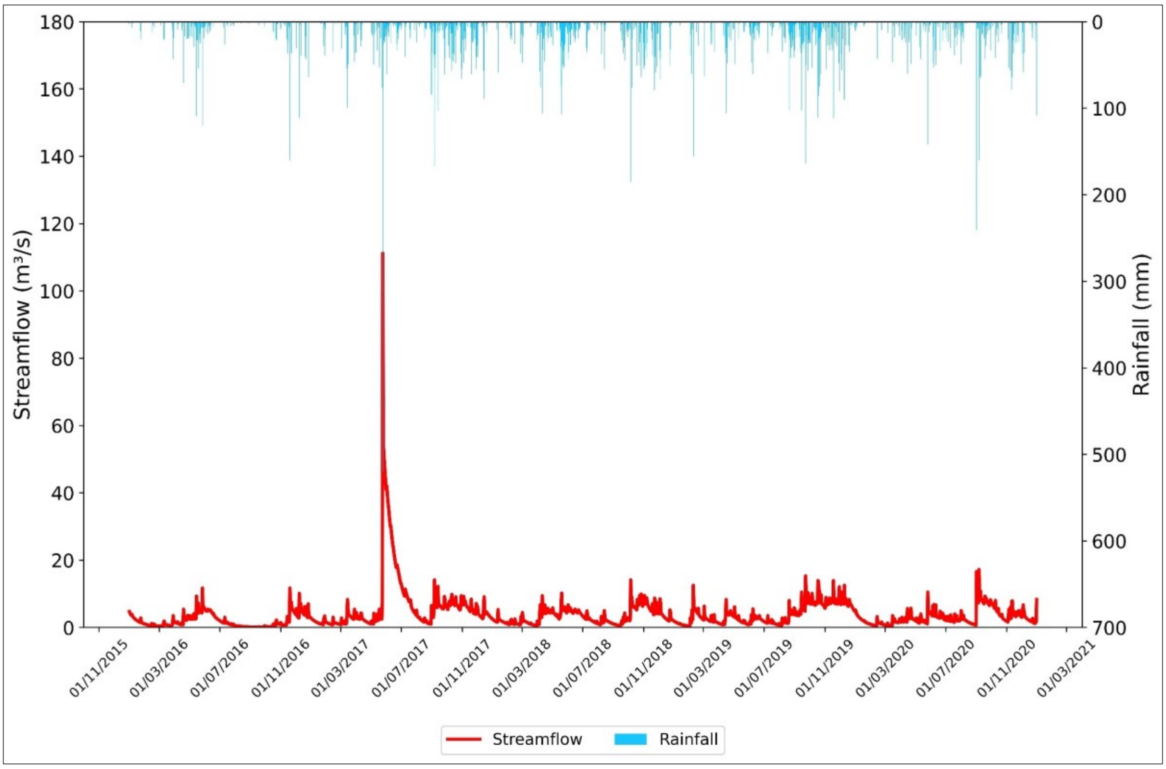


Figure 4.23 Generated hydrograph for Pusseli Oya sub-catchment

The generated hydrographs demonstrated the variations of flows following the observed streamflow values in the main river. However, the absence of observed streamflow values for the sub-catchments posed an issue for validating the simulated hydrographs. Subsequently, the simulated streamflow hydrographs illustrated an error in streamflow with extreme values in streamflow.

4.3.2 Error Correction Using Hybrid Modelling

The generated hydrographs for the five sub-catchments demonstrated extreme values which can be errors. To mitigate these errors the hybrid modelling approach was used. An LSTM ANN model was used for the error correction along with the results of the HEC-HMS model. The LSTM ANN model was selected for this approach as it is a widely used model for streamflow prediction (Akiner et al., 2024; Khan et al., 2023; Nakhaei et al., 2024; Sahoo et al., 2019). In this approach, first, an LSTM ANN model was trained to predict the observed streamflow based on the inputs used to develop the HEC-HMS model and the residual error between the observed streamflow at Nagalagam Street streamflow station and the HEC-HMS simulated streamflow at the same station (Konapala et al., 2020). Hence the inputs used to develop the LSTM ANN model were rainfall at Colombo and Hanwella stations, monthly evaporation data, streamflow at Glencourse monitoring station and the residual between the observed and simulated streamflow at Nagalagam Station. A hyperparameter tuning for the model was conducted to identify the best parameters for the LSTM ANN model. The results of the hyperparameter tuning are shown in Table 4.5.

Then an LSTM ANN model was developed for the main river to assess the performance of the model. The developed model was trained for the period 2016–2017 and was validated for the period 2018–2020. The developed LSTM ANN model demonstrated great accuracy in predicting the river flow (Table 4.6, Figure 4.24) hence it was used to predict the sub-catchment flow. A similar approach was used for the main river to predict the sub-catchment streamflow. The LSTM ANN model was developed to predict the main river for the sub-catchments. The inputs for the LSTM ANN model were the rainfall station assigned for the sub-catchment in the HEC-HMS model, monthly evaporation data, and the residual error of the simulated and observed values of the main river at Nagalagam Street Station. The streamflow for the five sub-catchments was generated using this LSTM ANN model. The generated hydrographs for the Kollonnawa and Pusseli Oya sub-catchments are shown in Figure 4.25 and Figure 4.26.

Table 4.5 Results of the hyperparameter tuning for the LSTM ANN model

Hyperparameter	**Description**
Layer 1	LSTM with 170 units
Layer 2	Dropout with rate of 0.3
Layer 3	LSTM with 60 units
Layer 4	Dropout with rate of 0.3
Layer 5	Dense (output layer) with 1 unit, ReLU activation
Optimizer	RMSprop
Learning rate	0.00063
Epochs	50

Table 4.6 Performance of the LSTM ANN model developed for the main river

Station	**Period**	**NSE**	**RMAE**	**R^2 error**
Nagalagam Street	Training Period	0.91	5.04	0.92
	Validation Period	0.95	4.65	0.95

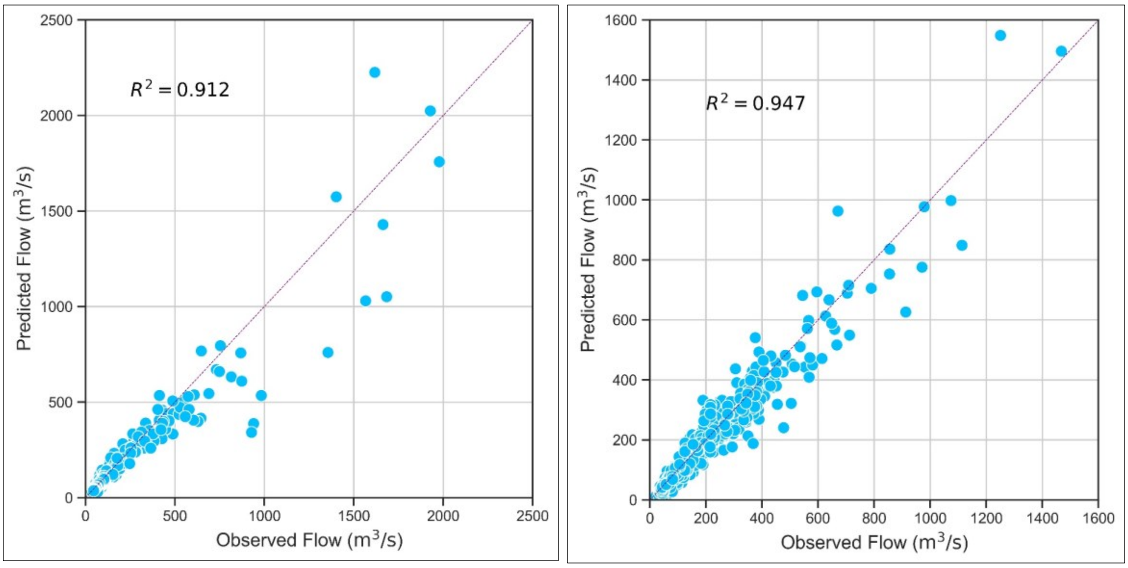


Figure 4.24 Performance of the LSTM ANN model for the training and validation periods

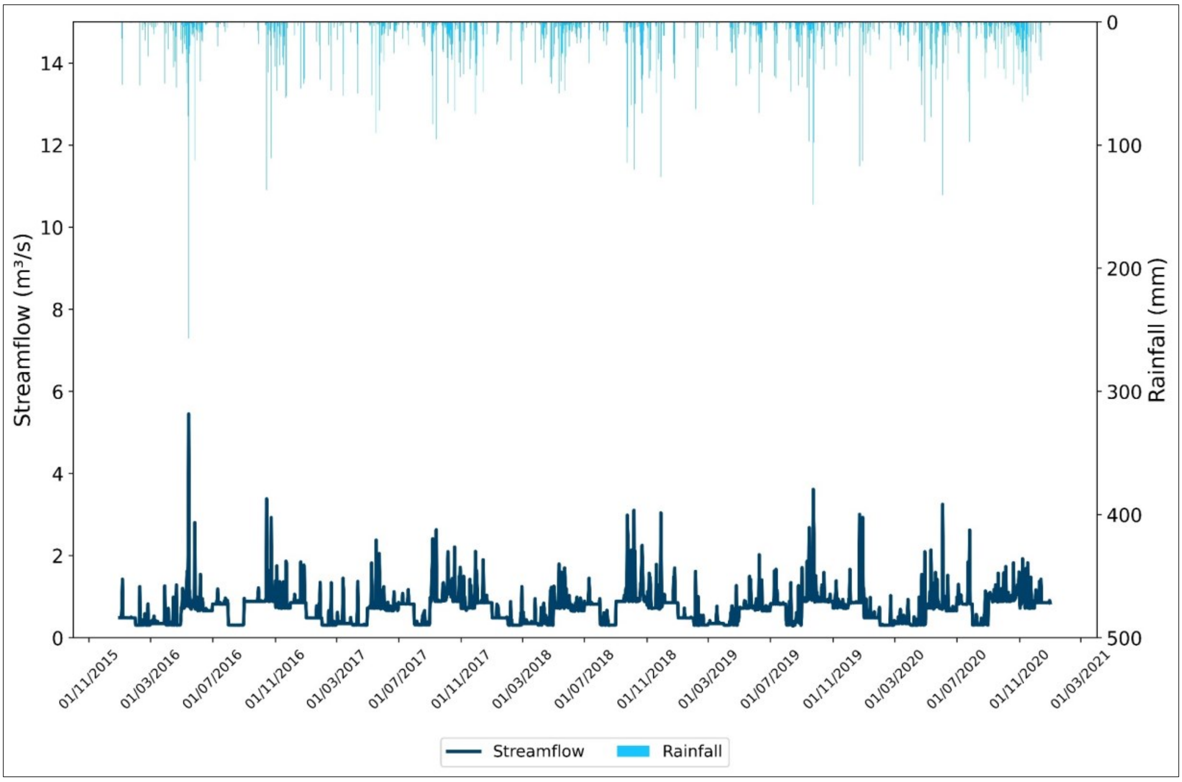


Figure 4.25 Error corrected hydrograph for Kollonnawa sub-catchment

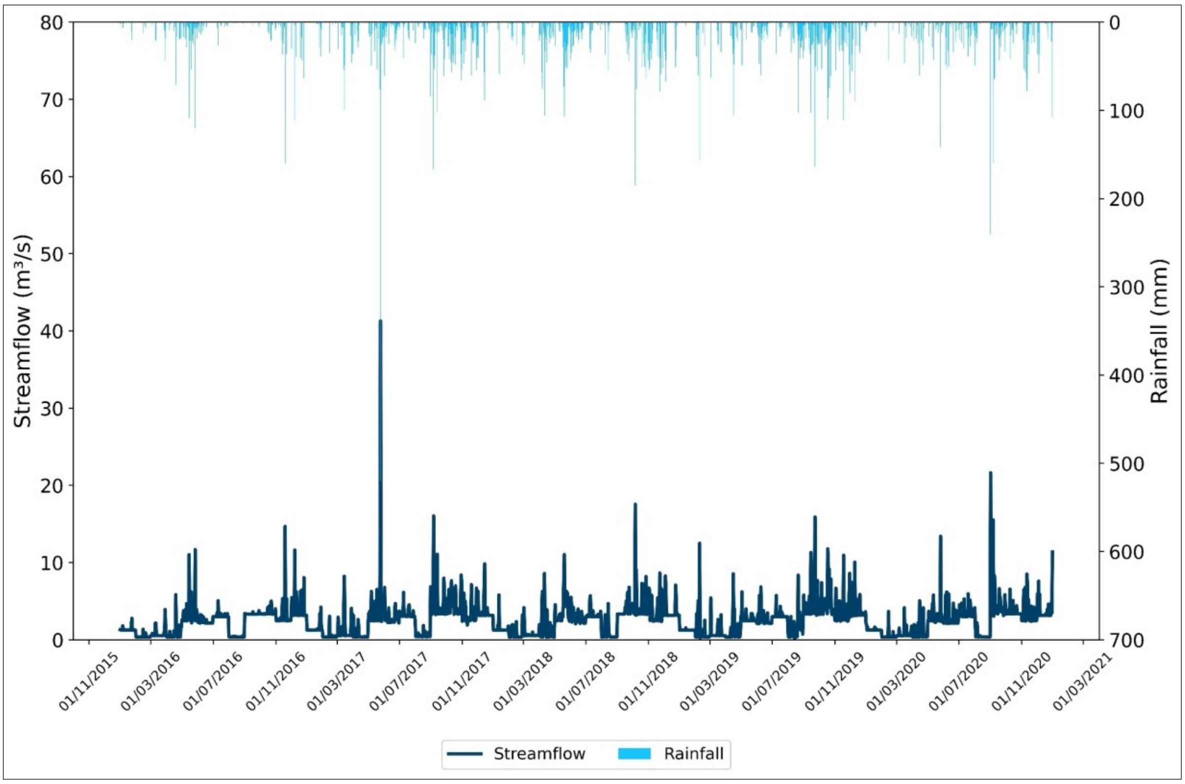


Figure 4.26 Error corrected hydrograph for Pusseli Oya sub-catchment

4.3.3 HEC-RAS Modelling

The HEC-RAS model was then developed using the observed data of the main river and the generated sub-catchment hydrographs. A 1D unsteady state model was developed for the period 2016–2020. The calibration period was selected as 2016–2017 and the validation period was 2018–2020. The model was calibrated by changing Manning's roughness coefficient values. The HEC-RAS model was developed using the obtained river cross-section data. The observed daily discharge at the Glencourse monitoring station was assigned as the upstream boundary condition for the model while the observed water level at the Nagalagam Street monitoring station was used as the downstream boundary condition. The generated hydrographs for the five tributaries of the Kelani River were used as the inflow boundary conditions from the sub-catchments. The model was calibrated using the observed streamflow data at the Hanwella monitoring station. The Manning's roughness coefficient ranged between 0.04 and 0.025 for the river and between 0.05 and 0.03 for the riverbanks. The performance of the HEC-RAS model is shown in Table 4.7 and Figure 4.27.

Table 4.7 Performance of the HEC-RAS model

Station	**Period**	**NSE**	**RMAE**	**R^2 error**
Hanwella	Calibration Period	0.87	5.86	0.88
	Validation Period	0.85	6.39	0.93

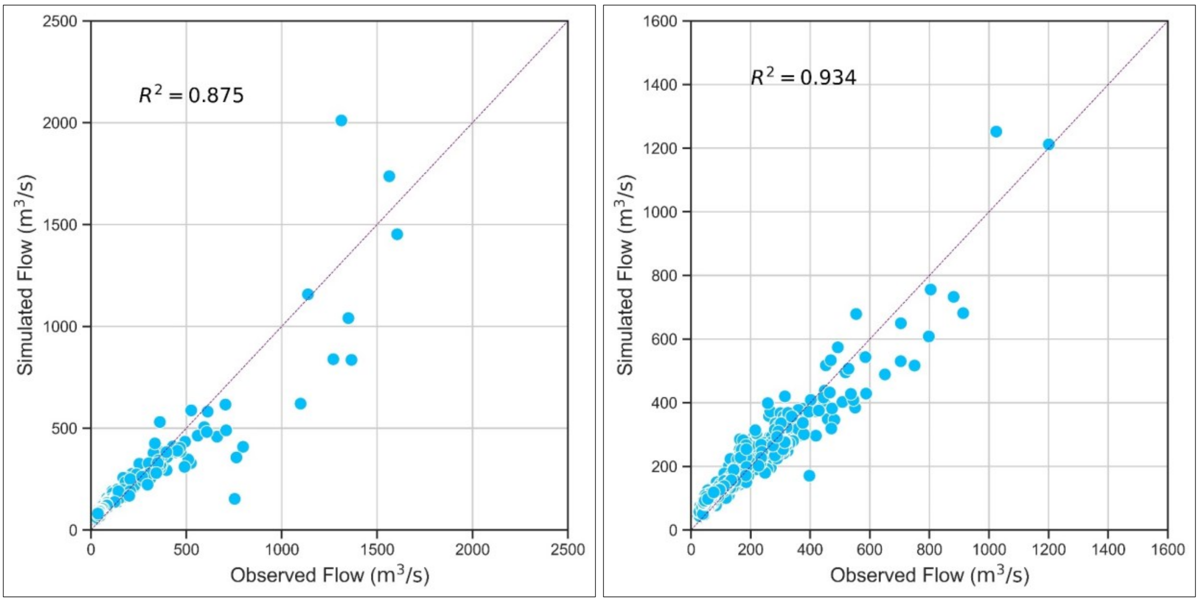


Figure 4.27 Performance of the developed HEC-RAS model for the calibration and validation periods

4.3.4 WASP Modelling

The WASP model was developed using the flows simulated using the HEC-RAS hydrodynamic model. The Kelani River and the sub-catchments were divided into 58 segments. The segments were divided using the river-cross sections used to develop the HEC-RAS model. The upstream segment was numbered Segment 1 and the most downstream station was Segment 58 (Figure 4.28). Segment lengths were assigned the distance between each cross-section and the flow rates, velocities, and flow depths for each cross-section were averaged to calculate the flow rates, velocities, and flow depths for each segment (Prjapathi et al., 2023).

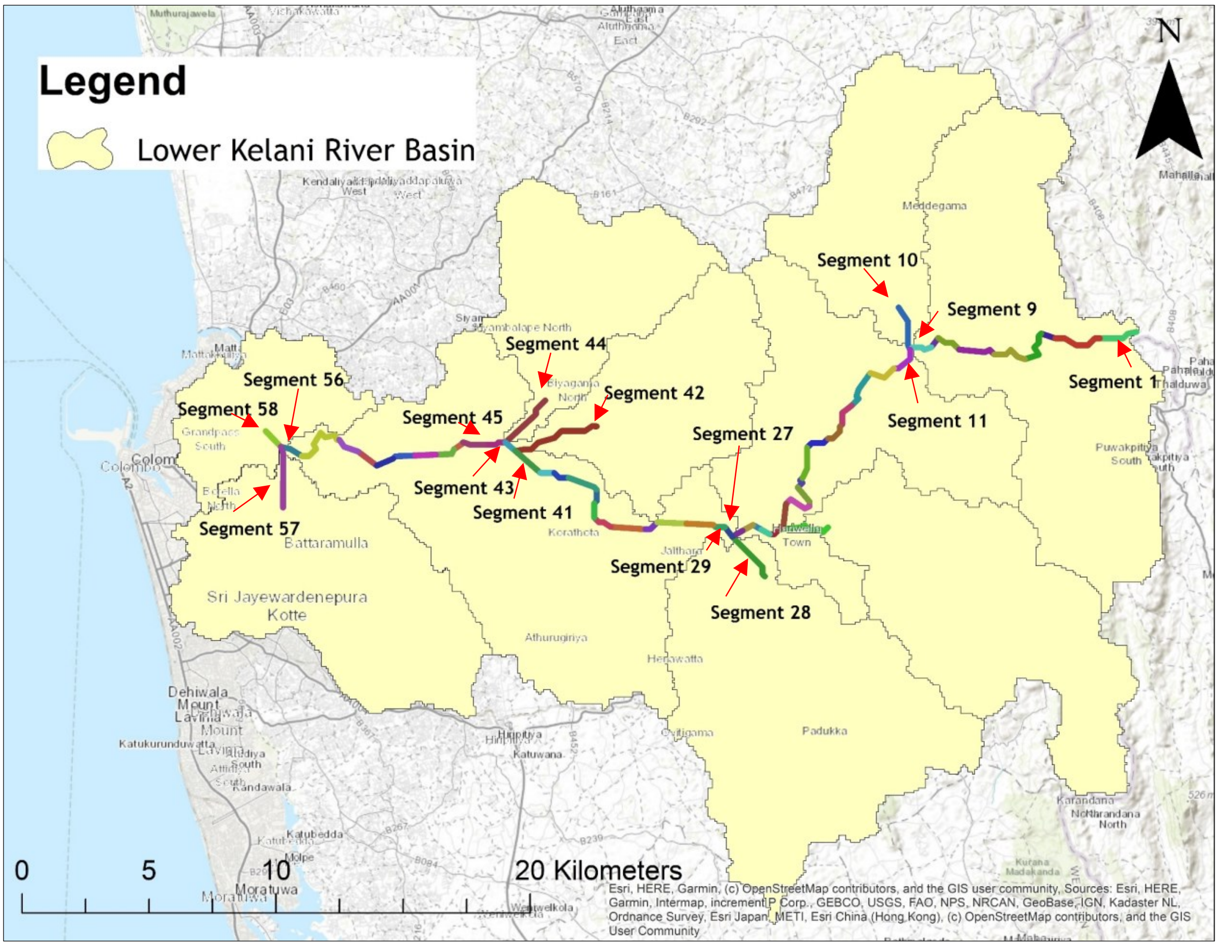


Figure 4.28 Distributions of segments used to develop the WASP model

The measured concentrations for Fe were instantaneous measurements in monthly intervals. For the WASP model development, these values were considered daily measurements, and the same value was used for the entire month. Then these measurements were considered as calibration parameters and were changed in a range from -50% to +50% during the model calibration and validation. Similarly, the solid settling velocities and solid resuspension velocities for each segment were also used as calibration parameters. The considered calibration period was 2016–2017 and the validation period was 2018–2020. The performance of the model was evaluated at two stations. Segment 58 was chosen as one of the stations and the Fe concentrations at Victoria station (S1) were used for this. Measurements at the Hanwella monitoring station (Segment 24) were also used to evaluate the model performance. The performance of the developed WASP model is shown in Figure 4.29 and Figure 4.30. Time series plots for calibration and validation periods are shown in Figure 4.31, Figure 4.32, Figure 4.33, and Figure 4.34.

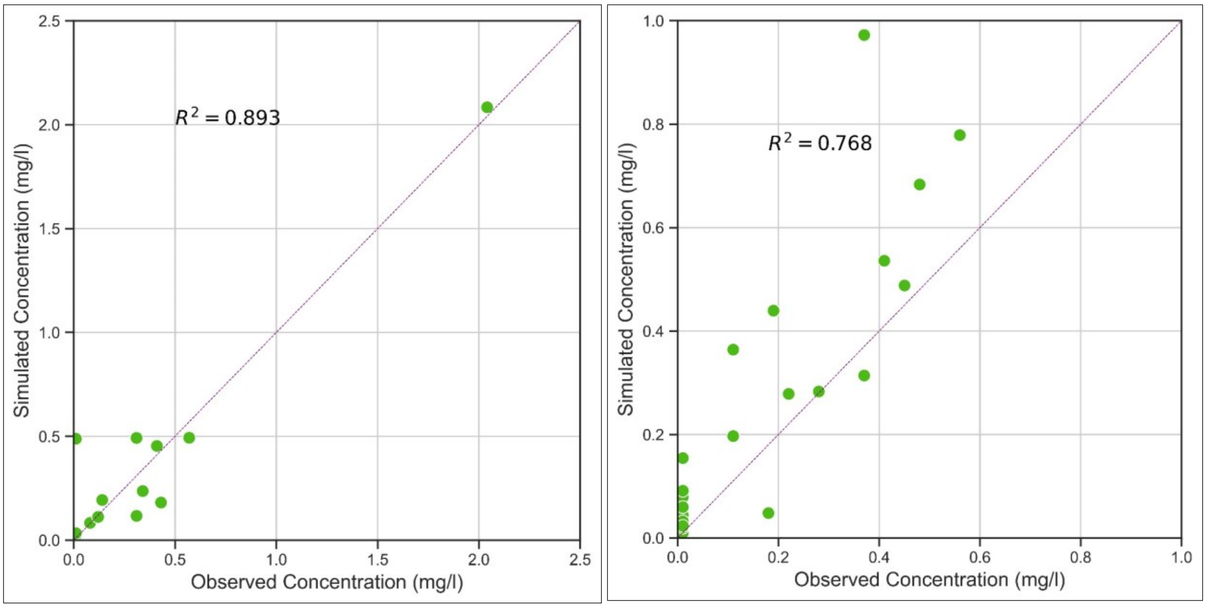


Figure 4.29 Performance of the developed WASP model for the calibration and validation periods at Segment 58 (Victoria Station)

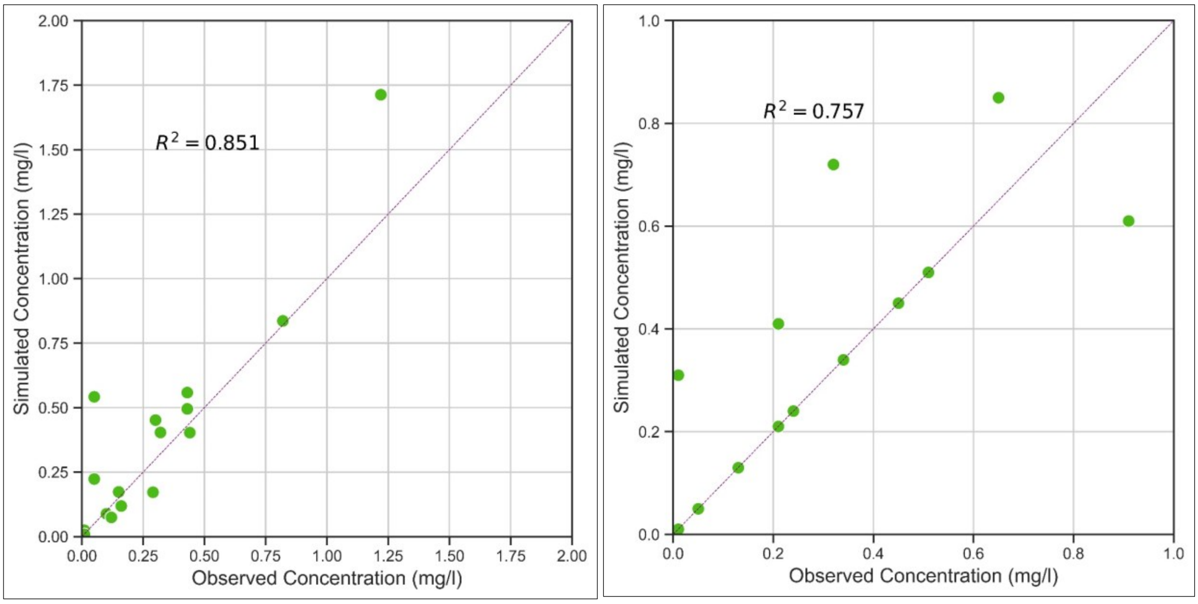


Figure 4.30 Performance of the developed WASP model for the calibration and validation periods at Segment 24 (Hanwella Station)

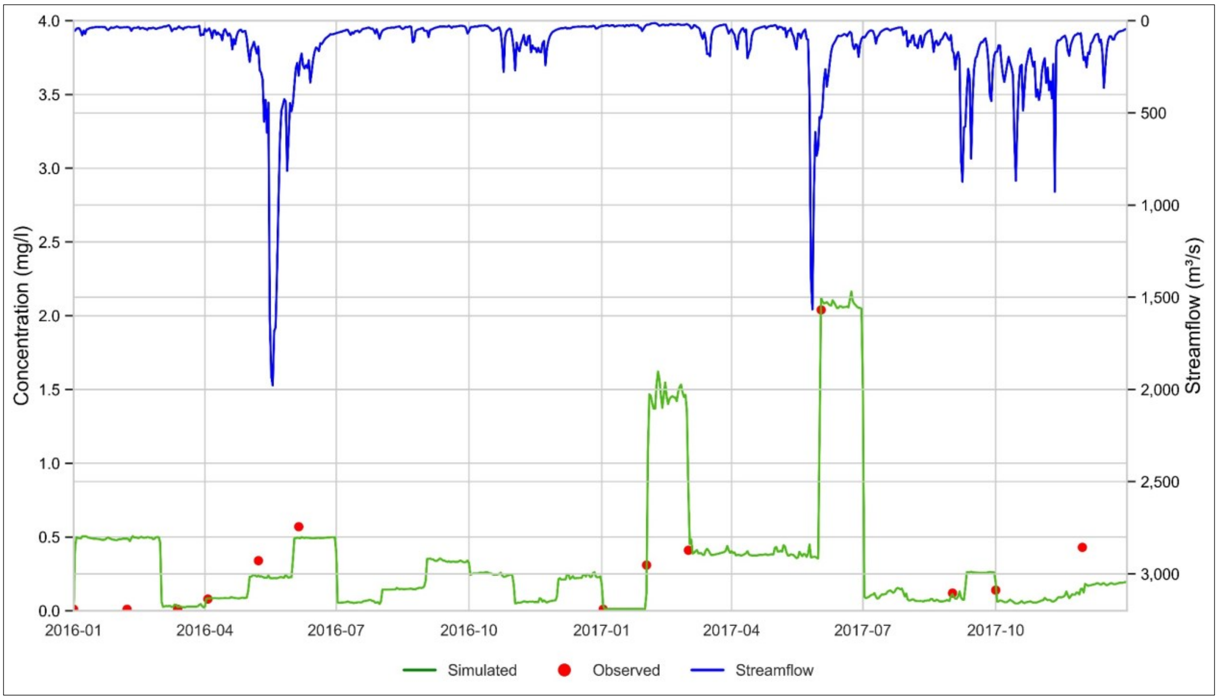


Figure 4.31 Time series plot for the calibration period at Segment 58 (Victoria Station)

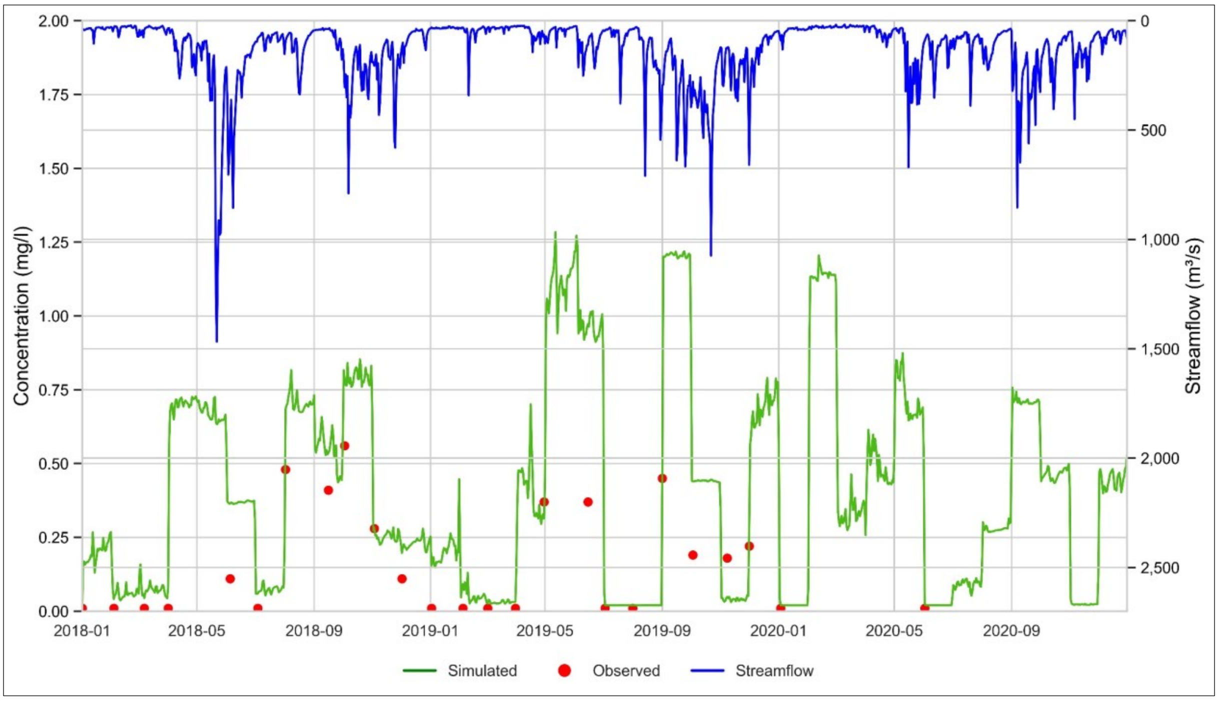


Figure 4.32 Time series plot for the validation period at Segment 58 (Victoria Station)

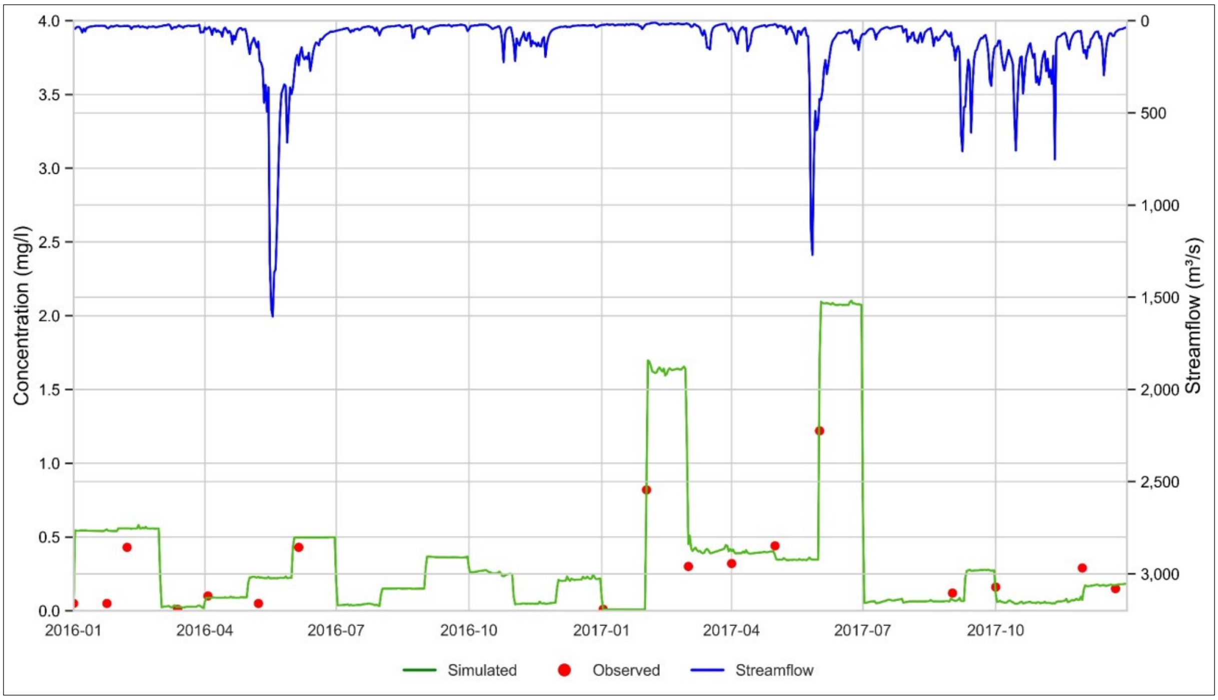


Figure 4.33 Time series plot for the calibration period at Segment 24 (Hanwella Station)

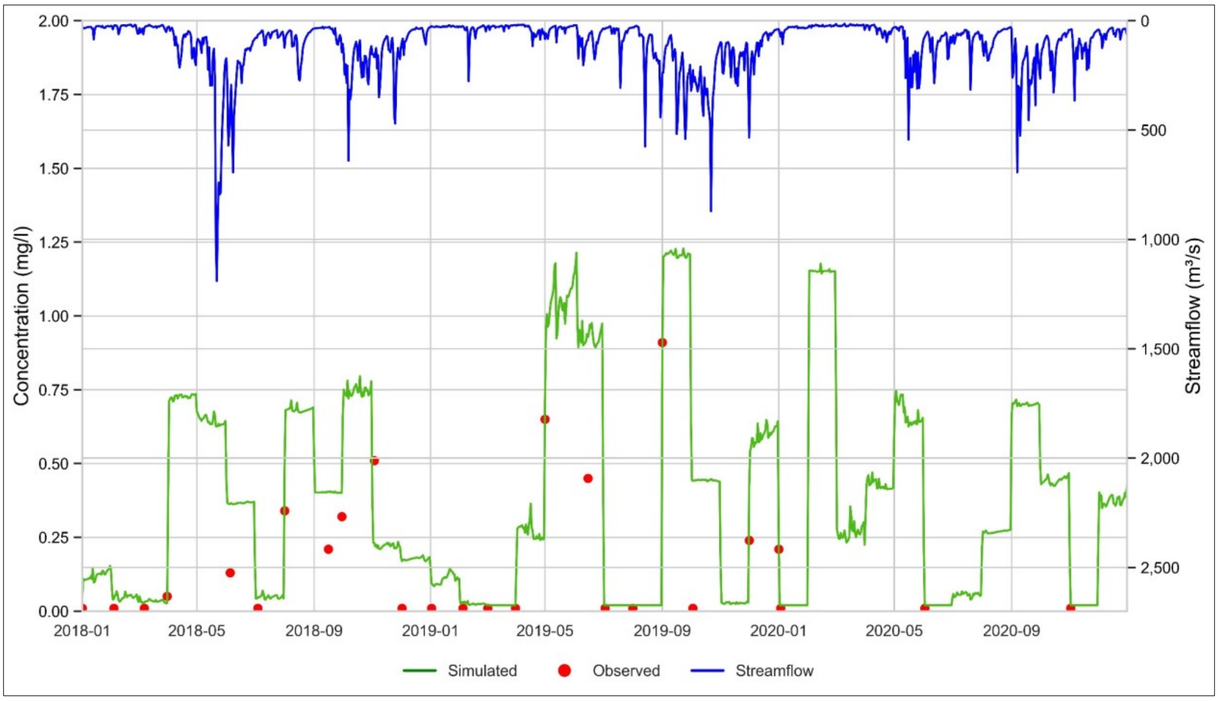


Figure 4.34 Time series plot for the validation period at Segment 24 (Hanwella Station)

CHAPTER 5

5 DISCUSSION

This section presents a detailed discussion of the results obtained and interprets findings in the context of the study objectives.

5.1 Spatial and Seasonal Water Quality Variations

Reducing the dimensionality of the dataset with 17 water quality parameters at 17 measuring points is crucial for simplifying analysis and easier interpretation while retaining the essential information in the original dataset. The Pearson correlation matrix was utilised to examine the primary parameters driving spatiotemporal variation of the entire dataset. From this analysis, 10 parameters were identified as significant: EC, turbidity, temperature, DO, phosphate, nitrate, Pb, Fe, T. Coli, and Cd (Figure 4.1). The correlation analysis results presented in Table 4 revealed strong associations among turbidity, EC, and temperature. Specifically, EC exhibited correlations with turbidity and temperature, with correlation coefficients (r values) of −0.50 and −0.46, respectively. These negative correlations indicate an inverse relationship between the respective parameters. In addition, relatively higher correlations were observed between phosphate and DO ($r = 0.34$), Fe and DO ($r = 0.33$), T. Coli and DO ($r = 0.54$), Pb and nitrate ($r = 0.40$), and Fe and Cd ($r = 0.67$). There were no obvious correlations ($r < 0.30$) between the other water quality parameters.

According to the FA (Table 4.1), latent variable 1 (LV-1) accounted for 25.55% of the total variance of the entire water quality dataset and was correlated primarily with EC, turbidity, and temperature. This factor implies a typical mixed-type pollution with the physical water quality parameters being the prominent parameters contributing to LV-1. Latent variable 2 (LV-2) accounted for 17.25% of the total variance. The most important water quality parameters contributing to the LV were Cd and Fe, indicating heavy metal pollution in the river. Latent variable 3 (LV-3) was mainly associated with DO and T. Coli. The total variance accounted for by LV-3 was 13.54%. Latent Variable 4 (LV-4) was correlated primarily with nitrate and Pb accounting for 10.58% of the variance. Latent variable 5 (LV-5) accounted for 10.17% of the total variance and was correlated with phosphate. The identified LVs suggest different pollution types in the KRB. The LV-1 suggests a mixed-type pollution with EC and turbidity being the prominent parameters contributing to the LV. Similarly, these parameters suggest pollution due to salinity intrusion (Ashrafuzzaman et al., 2022; Xia et al., 2021) and pollution caused by urban runoff (Neupane et al., 2015; Wang et al., 2017). Similarly, LV-2 and LV-4 suggest pollution due to industrial effluents as the correlating parameters for these two LVs are pollutants discharged from industries (Azam et al., 2019; Fořt et al., 2022; Kishor et al., 2020). The LV-3 and LV-5 suggest sewage pollution as the correlating parameters for this LV are DO, T. Coli, and phosphate (Figueroa-Nieves et al., 2014; Preisner, 2020; Raj, 2013). The five LVs identified by the FA therefore suggest the presence of physical, chemical, heavy metal, and micro-biological pollution in the Kelani River.

The spatial variation analysis helps identify regions where certain parameters vary significantly, indicating potential pollution hotspots, which may correlate with changes in land use types, such as built-up areas or agricultural land, further influencing water quality dynamics (Cheng et al., 2022; Pak et al., 2021). The five LVs obtained from the FA were clustered individually using HCA and SOM techniques to assess the spatial similarities between the 17 water quality monitoring stations having different land use practices. The results show that the clustering from both methods grouped the stations almost similarly (Figure 4.3, Figure 4.4). Hierarchical clustering identified three clusters for LV-1, while two clusters each were identified for the other four LVs. The SOM technique also identified three clusters for LV-1 and two clusters each for the other LVs. The clustering process is based on the similarities of the data, and the different clustering methods identify groups differently. In this study, both HCA and SOM identified almost identical clusters for the LVs. Water quality monitoring stations like Seethawake Ferry (S13), Pugoda Ela (S11), and Maha Ela (S6) demonstrate different results compared to the other stations. Both HCA and SOM have identified three clusters for LV-1 which is mainly influenced by the physical properties (EC, turbidity, and temperature) of water quality. Both methods grouped the stations closer to the sea outfall of the Kelani River into one separate cluster. The EC is a correlating parameter in LV-1 which changes with the distance from the coast towards the inland area reflecting the salinity intrusion. Salinity intrusion is reportedly high in the lower reach of the KRB resulting in variations in EC and Chloride levels (Gunawardena et al., 2017). Figure 5.1 shows the variation of the Chloride level across various stations with the most significant variations observed in those closest to the sea (S1 and S2). The average EC levels at Victoria Bridge (S1) and Kollonnawa Bridge (S2) stations were 528 µS/cm and 622 µS/cm, respectively, during the study period (2016–2020) compared to the national water quality standard of 700 µS/cm for EC. During the study period, it was observed that the EC levels exceeded the national standards approximately 7% and 9% of the time (mainly during the NWM period) at the Kollonnawa Bridge and Victoria Bridge stations, respectively.

The turbidity level in river water closely relates to the population density of the KRB (Figure 5.2). The average turbidity levels at five stations identified in the CA, namely, Maha Ela Canal (S6), Raggahawatte Canal (S4), Kaduwela Bridge (S5), Welivita Bridge (S3), Seethawake Ferry (S13), and Pugoda Ela (S11) stations, were between 15 and 69 NTU for the study period (2016–2020) exceeding the national water quality standard of 5 NTU. Hence, the CA has identified the station correlated with LV-1 having heightened variation. The turbidity level in the Kelani River is influenced by the land use practices of the region and the sand mining activities in the river (Athukorala et al., 2013). According to the results of the CA of LV-1, the stations upstream with a lesser population density are grouped into Cluster 3. In contrast, the stations in a more urbanized area are clustered into Cluster 2 and the stations near the river mouth are clustered into Cluster 3.

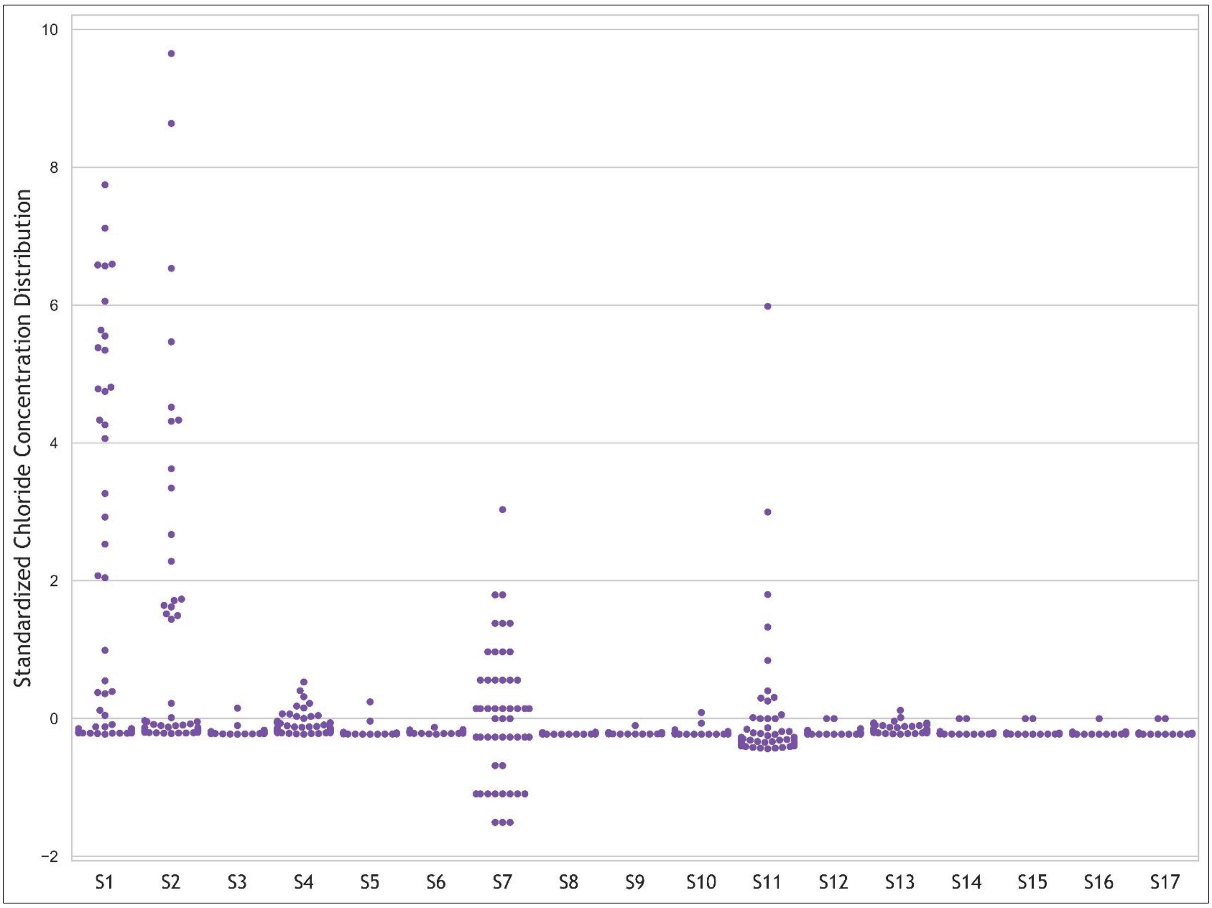


Figure 5.1 Chloride level variation along the Kelani River

Heavy metals risk human health and can lead to diseases such as chronic kidney failure (Banarjee, 2012; Kodikara et al., 2022). According to the FA, LV-2 is mainly correlated with Fe and Cd which are discharged primarily from industries and factories. Both clustering techniques have grouped Maha Ela (S6), Raggahawatte Ela (S4), and Kollonnawa Bridge (S2) into a separate cluster (Cluster 2). The SOM clustered Pugoda Ela (S11) and Victora Bridge (S1) into Cluster 2. The water quality of Pugoda Ela (S11) has degraded over the years due to the combined effect of effluents discharged from industries and agricultural practices in the area (Makubura et al., 2022). Wastewater generated at INZ1 discharged into Raggahawatte Ela (S6) (Gunawardena et al., 2017) and Maha Ela (S4) (Ruvinda & Pathiratne, 2018). Heavy metal pollution due to municipal waste is also considered a concerning factor in the drainage network in Colombo City which includes the Victoria Bridge (S1) and Kollonnawa Bridge (S2) (Ranasinghe et al., 2006). The average Fe and Cd levels at these stations ranged between 0.32–0.16 mg/l and 0.01–0.04 mg/l, respectively, during the study period (2016–2020) compared to the national water quality standard of 1 mg/l and 0.005 mg/l for Fe and Cd. The five stations identified in Cluster 2 have an exceedance percentage of Category A of CEA guidelines between 17 and 34% of the time for

Fe and a range of 30% to 32% of the time for Cd. The clustering has successfully identified the stations where the effluents from the industrial zones are discharged and grouped them into separate clusters.

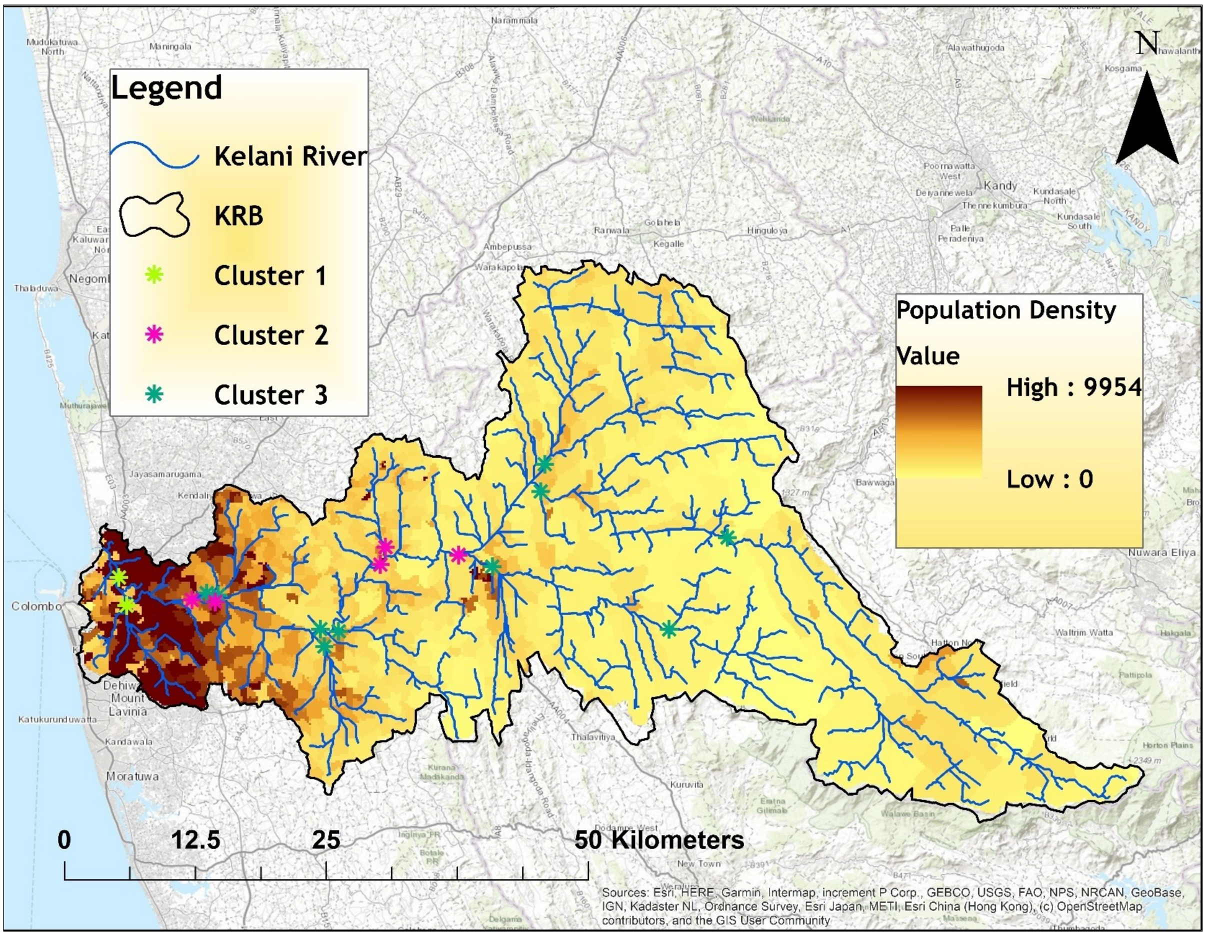


Figure 5.2 The population density (persons per square kilometre) distribution in the Kelani River Basin is illustrated using data from the Department of Census and Statistics Sri Lanka for 2011

According to the results of the FA, DO, and T. Coli are the most influential factors for LV-3, which represents microbiological pollution. Microbiological pollution mainly occurs due to the discharge of domestic and municipal waste into the river ecosystem. Wastewater discharged from the dairy industries located inside the KRB may also be contributing to the micro-biological pollutants in the river system. Victoria Bridge and Kollonnawa Bridge were clustered into one group by both clustering techniques while SOM grouped Seethawake Ferry (S13), Pugoda Ela (S11), and Raggahawatte Canal (S4) into the same group. Victoria Bridge (S17) and Kollonnawa Bridge (S16) are considered the most polluted sampling stations for microbiological pollution (Ranasinghe et al., 2006). The average DO and T. Coli levels at these stations ranged between 3.4–6.4 mg/l and 6,475–8,300 MPN/100 ml, respectively, during the study period (2016–2020) compared to the national water quality standard of 3 mg/l and 10,000 MPN/100 ml for DO and T. Coli, respectively. The stations identified in Cluster 2 showed an exceedance of CEA guidelines with a 5% to 44% range for DO and a 9% to 22% range for T. Coli. Victoria Bridge, Kollonnawa Bridge, and Raggahawatte Ela were the stations having the highest exceedance among the five stations. According to the CA results, Cluster 2, which includes Victoria Bridge and Kollonnawa Bridge by both clustering techniques, can be considered as the stations with the highest variation in microbiological pollution in the river followed by the stations located near the two industrial zones.

The main parameters contributing to LV-4 are nitrate and Pb, which indicate pollution due to industrial effluents. The cluster analysis identified two distinct clusters according to the spatial distribution of these parameters. The clustering has identified the stations where the effluents from INZ1 and INZ2 are discharged (Makubura et al., 2022) and clustered them into a separate cluster the same for LV-2. Nitrate and Pb are common effluents in wastewater discharged from textile and apparel factories, while Nitrate is also a pollutant due to agricultural runoff. The stations identified with higher variance for LV-2 are also pinpointed for LV-4, underscoring the industrial pollution present in those areas. The average levels at these stations ranged around 0.01 mg/l for Pb and between 0.55 and 1.77 mg/l for nitrate during the study period (2016–2020) compared to the national water quality standard of 0.05 mg/l and 10 mg/l for Pb and nitrate, respectively. Subsequently, the identified station had an exceedance of Category A of the CEA guidelines within a range of 2 to 6% for nitrate and 1 to 3% for Pb. The highest exceedances were identified at Victoria Bridge and Maha Ela stations.

The main parameter contributing to LV-5 is phosphate. The cluster analysis identified two distinct clusters according to the spatial distribution of the parameter. Victoria Bridge and Kollonnawa Bridge were grouped into Cluster 2 by both clustering methods, while the other stations were grouped into the other cluster. Phosphate indicates pollution due to agricultural runoff and municipal and domestic waste. Victoria Bridge is located closest to the sea outfall of the Kelani River, and it accumulates the water from the drainage network of Colombo city and its suburbs. Similarly, the Kollonnawa Canal is highly polluted due to municipal and domestic waste. The average phosphate levels at these two stations were 0.1 mg/l and 0.15 mg/l for Victoria Bridge and Kollonnawa Bridge during the study period (2016–2020) compared to the national water quality standard of 0.7 mg/l. In contrast to the other water quality parameters, phosphate concentration remained below the national water quality standard for almost all of the monitoring period. The analysis of spatial variations of the five latent variables thus validates the presence of different types of pollution types in the KRB. The results suggested that industrial pollutants have higher variations around the industrial zones, and the effect of salinity intrusion is concentrated near the sea outfall of the Kelani River. Similarly, the parameters contributing to sewage pollution also had higher variations near industrial zones and urbanized areas.

The CA results suggest that water quality parameters of certain clusters displayed notable variation during the study period, frequently surpassing national water quality standards. Consequently, it is imperative to investigate the seasonal fluctuations of the selected LVs concerning hydrometeorological variables such as rainfall and river discharge. There are six rain gauge stations in the KRB. Thiessen Polygon method was used to identify the water quality monitoring stations that correspond to each rain gauge station (Figure 3.1). Pearsons correlation matrix was then used to correlate the LVs with the measured rainfall (Prathumratana et al., 2008). To assess the correlation of streamflow with LVs, three water quality measuring stations located near the streamflow measuring stations and along the same streams were selected.

A correlation analysis was conducted to explore the relationship between LVs and rainfall dependency. Out of six rain gauge stations, the results for the Colombo and Hanwella stations are shown in Figure 4.5, where the latent variables exhibit significant correlations with rainfall variability. There are five water quality monitoring stations corresponding to the Colombo rainfall station, namely, Victoria Bridge, Kollonnawa Bridge, Kaduwela, Welivita Bridge, and Raggahawatte Ela. The CA identified Kollonnawa Bridge, Victoria Bridge, and Raggahawatte Ela as stations exhibiting elevated variability in industrial effluent levels (LV-2 and LV-4). Furthermore, the Kollonnawa Bridge and Victoria Bridge stations were identified for exhibiting a greater variation in LV-1, attributed to the impact of salinity intrusion. Similarly, the Welivita Bridge and Kaduwela stations were recognized for experiencing a heightened variation in LV-1, attributed to turbidity. The KRB experienced higher rainfall during the IM2 and the SWM periods. The highest correlation values can be observed during the IM2 and SWM, while negative correlation values are observed during the IM1 and NEM for LV-1. The heightened rainfall results in increased surface runoff generating higher turbidity levels, hence explaining the high correlation values during the IM2 and the SWM seasons. In contrast, salinity intrusion occurs during the low flow periods, hence the EC levels in the river increase with the decrease of flow explaining the negative relationships between the LV-1 and rainfall during IM1 and NEM seasons (Figure 4.5). In contrast, LV-2 and LV-4 which indicated pollution due to industrial effluents had lower average correlation (negative correlation) coefficient values during the SWM and IM2, while the correlations were comparatively higher with the heightened precipitation. It implies that the heightened rainfall dilutes the pollutant concentration indicating year-long pollution due to industrial effluents in the region. Similarly, LV-3, consisting of microbiological parameters (DO and T. Coli), exhibited negative correlations across all four seasons, with a tendency for dilution during months characterized by increased rainfall. The latent variable consisting of only phosphate (LV-5) had negative correlations with precipitation except during the IM1. It generally means that the pollutant accumulates in the river during low-flow periods and dilutes its concentration during rainy seasons.

There are six water quality monitoring stations corresponding to the Hanwella rainfall station, namely, Wak Oya, Pugoda Ferry, Pugoda Ela, Pusseli Oya, Maha Ela, and Hanwella stations. Maha Ela and Pugoda Ela stations were identified in the CA as stations with heightened variation for LVs due to industrial pollution. The effect of LV-1 was significant at Pugoda Ela, Seethawake Ferry, and Pugoda Ferry stations due to higher variation of turbidity levels. According to the spatial variation analysis results, four of the six stations were clustered into the group with minimal variation of water quality for LV-1, which expresses the low correlation coefficient values for LV-1 with the precipitation. Subsequently, for LV-2 and LV-4, correlation values closer to zero were predominantly observed, whereas higher variation in correlation values was noted during IM2 and SWM. The highest negative correlation values were observed during the IM2, and the stations correlated with the values were Pugoda Ela and Maha Ela stations. Hence, the dilution of pollutants due to heightened rainfall is observed for LV-2 and LV-4. The stations considered were clustered into the group with minimal variations for LV-3 in the spatial variation analysis. The mean correlations were close to zero for LV-3. LV-5 negatively correlated with precipitation except during the NEM period. LV-5 is primarily associated with phosphorus, a pollutant connected with agricultural activities. Agriculture dominates the KRB during the IM2 and SWM seasons, which are characterized by heavy rainfall (Mahagamage & Manage, 2018). Thus, the observed negative correlations can be attributed to the diluting effect of heavy rainfall on agricultural runoff, which results in strong negative correlations.

Four water quality monitoring stations were associated with the Dunedin rainfall station, namely, Karawanella, Aguruwella, Thalduwa, and Seethawake Ferry. The Seethawake Ferry station was identified as having heightened variations due to industrial effluents and turbidity, whereas the remaining three stations demonstrated minimal variations across LVs. Subsequently, during the IM2 and SWM periods, LV-2 and LV-4 showed negative correlations with Seethawake Ferry, while the other three stations showed correlations close to zero. Regarding LV-1, Seethawake Ferry had the highest positive correlation values, though correlation coefficients varied more widely across all four stations during IM1. The three water quality monitoring stations associated with the Wewalthalawa Estate rainfall station were the Karawanella, Nakkawaita, and Kithulgala stations. All three stations were clustered into clusters with minimal water quality variations for all the LVs. Consequently, all the LVs showed minimal correlation with the precipitation indicating the minimal water quality variation in the three stations as identified in the spatial variation analysis.

In summary, downstream stations exhibiting greater variations in water quality showed strong correlations with rainfall patterns, indicating strong seasonal variations. The observed dilution effect of pollutants during increased precipitation suggests that pollution in the region continues throughout the year with high concentrations during the non-rainy seasons. While the correlations between rainfall and LVs were investigated at each water quality monitoring station, a selective analysis approach was employed for streamflow stations (Figure 4.6). The analysis only included water quality monitoring stations close to the streamflow gauging station. Kithulgala station was analyzed with the Kithulgala streamflow station to obtain the correlation coefficients.

Kithulgala station was categorized into a group with minimal variations of water quality for all the LVs. Subsequently, the correlation coefficients fail to identify a strong correlation with the LVs validating the results of the spatial variation analysis for the Kithulgala station. However, strong negative correlations can be observed during the IM2 season for LV-5 which can be due to the agricultural activities in the region resulting in agricultural runoff. The variations of water quality parameters for the Seethawake Ferry station were analyzed with the streamflow data of the Glencourse stream flow station.

Seethawake Ferry station was identified as having heightened variations due to industrial effluents along with turbidity levels. Subsequently, during the IM1, SWM, and IM2 seasons, LV-1 showed strong positive correlations, indicating an association with increased runoff and elevated turbidity levels. The LV-4 showed significant negative correlations during the IM1, IM2, and SWM seasons, indicating a dilution effect on pollutants due to increased river flow. Furthermore, strong negative correlations were observed for LV-5 during the IM1, IM2, and SWM seasons, owing to agricultural runoff caused by seasonal agricultural activities, affecting water phosphate levels. Victoria Bridge station was analyzed with the Nagalagam Street streamflow station to obtain the correlation coefficients. Victoria Bridge station was identified as having heightened variation in water quality for all the LVs. Subsequently, for LV-1, negative correlations were observed during NEM and IM1 demonstrating the effect of salinity intrusion in the area. Subsequently, LV-2, LV-3, and LV-5 indicated negative correlations during the IM1, IM2, and SWM seasons, indicating decreased pollutant concentrations due to heightened rainfall leading to increased river flow. In contrast, during the IM2 season, LV-1 showed a positive correlation, indicating turbidity effects at the water quality monitoring station located near a densely urbanized area.

Seasonal analysis demonstrates that latent LVs derived from the FA effectively capture variations throughout the seasons, in conjunction with hydrometeorological parameters. These LVs exhibit noticeable fluctuations in response to distinct land use and industrial practices highlighting the influence of monsoonal patterns on water pollution in the KRB. The capacity of LVs to reflect such variations underscores their significance in water quality assessment studies. The seasonal and spatial analysis of water quality suggested that there is significant pollution in the KRB with pollution due to industrial effluents and pollution due to municipal and domestic sewage being the prominent methods of river pollution. The strong correlation with hydrometeorological data suggested that these pollution methods have distinct patterns correlating with the monsoon climate of Sri Lanka. Hence to evaluate the pollution of the Kelani River, industrial pollution and sewage pollution were considered and the pollutant source identification was conducted for these two prominent pollution methods.

5.2 Water Pollution Indexing

Pollution due to industrial effluents and sewage was identified as the prominent method of pollution in the KRB. Hence water pollution indices were developed to quantify the level of pollution and to identify the potential pollution sources for each method of pollution. A total number of six water quality parameters were selected for the Kelani River Basin-Industrial Pollution Index (KRB-IPI) and five parameters were selected to develop the Kelani River Basin-Sewage Pollution Index (KRB-SPI). The rating curves were developed for the KRB-IPI and KRB-SPI (Figure 4.7, Figure 4.8). It was observed that Cd, Fe, and Zn had a significant number of values exceeding the permissible limits specified by the Central Environmental Authority (CEA). As a result, the curve is spiralled toward the left side of the graphs for these parameters, indicating a higher rate of exceedance. In contrast, nitrate, Pb, and Cr showed lower percentages of exceedance relative to the standard limits, with the curve spiralling further to the right of the curve. A similar trend was observed for BOD, COD, and phosphate as these parameters exhibited a higher rate of exceedance of the CEA standard limits. These trends illustrate the pattern of exceedance for each parameter and demonstrate how the developed rating curve effectively identifies the rates of standard limit exceedances for different pollutants. The Random Forest (RF) algorithm was employed to identify the most significant contributing parameters to KRB-IPI and KRB-SPI, providing a deeper understanding of the six parameters under investigation. The RF model for the KRB-IPI achieved an AUC of 0.98 and an F1-score of 0.99, identifying Cd, Zn, and Fe as the primary parameters influencing river pollution status. Similarly, the RF model for KRB-SPI obtained an AUC value of 0.94 and an F1-score of 0.93 identifying DO, COD, and T. Coli as the most influential parameters affecting the sewage pollution in the KRB (Table 4.2). The ROC weights also followed the same hierarchical importance in assigning the weights for the parameters. The sub-index aggregation was carried out finally to develop the water pollution indices (Figure 4.13, Figure 4.14).

Then the indices were applied to assess the pollution status of the Kelani River across the four monsoon seasons, providing a clearer understanding of the condition of the river. The results of the KRB-IPI (Figure 4.15) suggested that the Kollonnawa Bridge (S2), Raggahawatte Ela (S4), Pusseli Oya (S7), and Pugoda Ela (S11) stations experienced high pollution, while the Maha Ela (S6) station was identified with extreme pollution. Nevertheless, comparing water pollution values following the monsoonal patterns of Sri Lanka suggests that the pollution varies during each season. The Raggahawatte Ela (S4), Maha Ela (S6), and Pusseli Oya (S7) stations experienced high and extreme pollution levels year-round, indicating persistent pollution. The Pugoda Ela (S11) and Seethawake Ferry (S13) stations showed significant pollution during the SWM and IM2 periods. Similarly, the results of the KRB-SPI (Figure 4.16) indicated that pollution from industrial, municipal, and domestic sewage is prevalent throughout the year across the entire KRB. Specifically, Seethawake Ferry (S13), Pugoda Ela (S11), Maha Ela (S6), Raggahawatte Ela (S4), Kollonnawa Bridge (S2), and Victoria Bridge (S1) were identified as the most polluted stations within the basin. The findings suggest that industrial and municipal sewage pollution levels are particularly high. Stations Raggahawatte Ela (S4), Maha Ela (S6), and Seethawake (S13) exhibit severe industrial sewage pollution. Similarly, stations Victoria (S1) and Kollonnawa (S2) are significantly impacted by municipal sewage, corroborating literature findings. The results highlight the concern of domestic sewage pollution within the basin. Stations such as Pugoda Ferry (S10), Wak Oya (S9), Kaduwela (S5), and Welivita (S3), located in less urbanized areas, are also polluted year-round.

These identified stations are consistent with findings in the literature. For instance, Pugoda Ela (S11) has experienced significant water quality degradation over the years due to industrial effluents and agricultural runoff (Makubura et al., 2022). Wastewater generated at INZ1 is discharged into Raggahawatte Canal (S4) (Gunawardana et al., 2022) and Maha Ela Canal (S6) (Ruvinda & Pathiratne, 2018), while the Seethawake station (S13) downstream of the INZ2. The Pusseli Oya (S7) station, situated in a sub-catchment of the KRB, is notably affected by heavy metal pollution (Makubura et al., 2022). Municipal waste, particularly in the drainage network of Colombo city, also contributes to heavy metal contamination, affecting sites like Kollonnawa Bridge (S2) and Victoria Bridge (S1) (Ranasinghe et al., 2006). According to previous studies, the Kollonnawa Bridge (S2) station is primarily polluted by municipal and domestic sewage. Thus, the key industrial pollution sources identified in this study include the stations at Maha Ela (S6), Raggahawatte Ela (S4), Pusseli Oya (S7), Pugoda Ferry (S10), Pugoda Ela (S11), and Seethawake Ferry (S13) (Figure 5.3). Similarly, Victoria (S1) and Kollonnawa (S2) stations are identified as polluted by municipal sewage. Stations Raggahawatte Ela (S4), Maha Ela (S6), and Seethawake (S13), located near industrial zones, exhibit severe industrial sewage pollution. Furthermore, the results highlight the concern of domestic sewage pollution within the basin. Stations Pugoda Ferry (S10), Wak Oya (S9), Kaduwela (S5), and Welivita (S3), located in less urbanized areas, were also identified as potential pollutant sources (Figure 5.4).

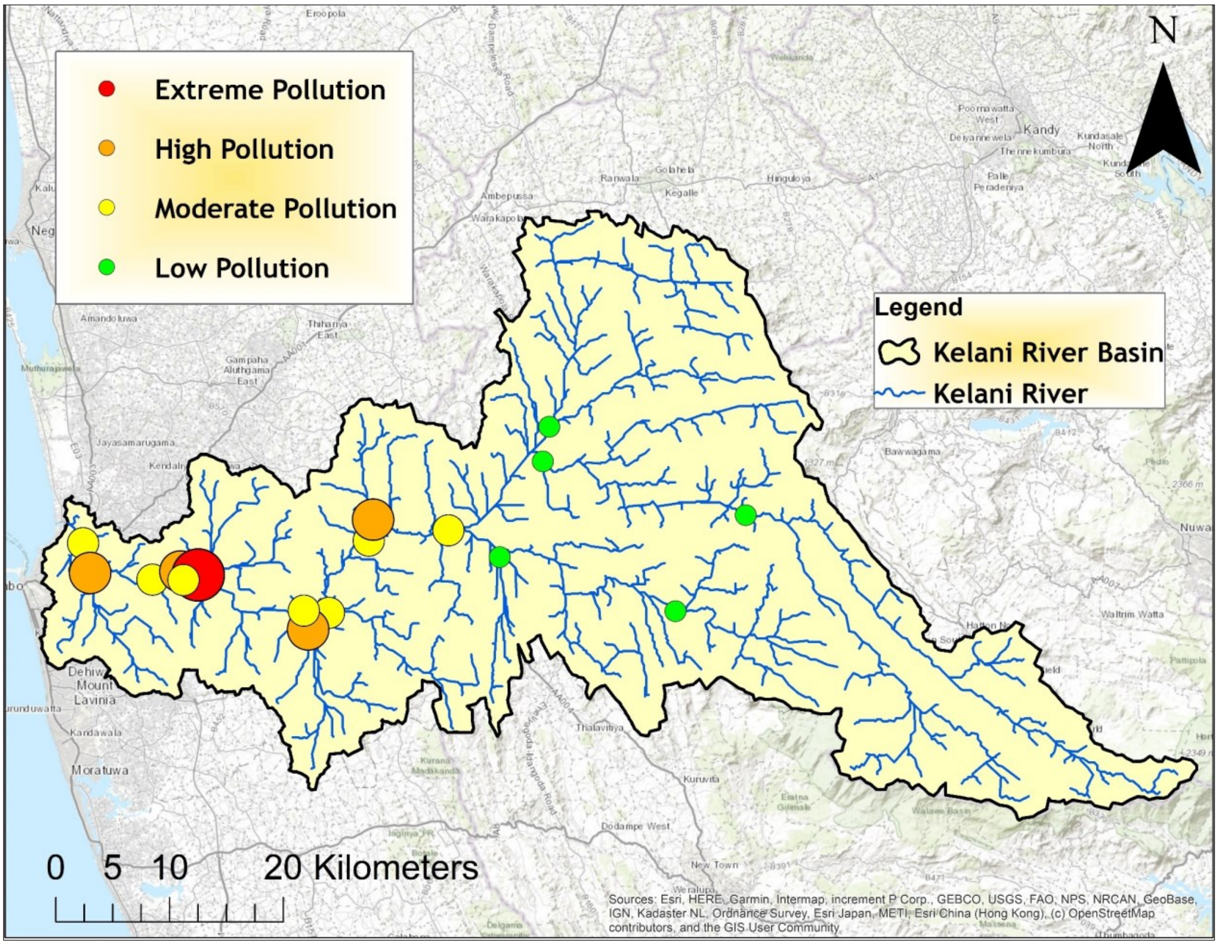


Figure 5.3 Spatial distribution of industrial pollution sources in the Kelani River Basin

The results of both KRB-IPI and KRB-SPI also highlighted that the water quality monitoring stations near the industrial zones show elevated pollution levels for both industrial effluents and sewage. Hence highlighting that pollution near the industrial zones is high all year around. Similarly, the S1 and S2 stations also demonstrate high levels of both industrial and sewage pollution. According to the results pollution due to domestic sewage is also a concerning issue in the KRB. Additionally, results of both

KRB-IPI and KRB-SPI revealed that the pollution levels increase during periods of heavy rainfall, particularly during the Southwest Monsoon (SWM) and second Inter-monsoon (IM2) seasons (Figure 4.15, Figure 4.16). Typically, higher precipitation and streamflow lead to dilution, resulting in lower pollutant concentrations (Xu et al., 2019). However, in this case, pollutant levels exceeded standard guidelines more frequently, likely due to surface runoff introducing additional contaminants into the water system. This suggests that significant pollutant loads may be released into the water during heavy rainfall, or that runoff is mobilizing pollutants from surrounding areas. A similar phenomenon of increasing pollutant levels during rainfall events has been observed in other basins within Sri Lanka (Diwyanjalee & Premarathne, 2024) as well as in countries with monsoon climates (Tuan & Babel, 2024). These findings imply that pollutant discharges into surrounding areas are not adequately regulated by the Central Environmental Authority's guidelines or the Environmental Protection License scheme, leading to a potential increase in contamination during rainfall events.

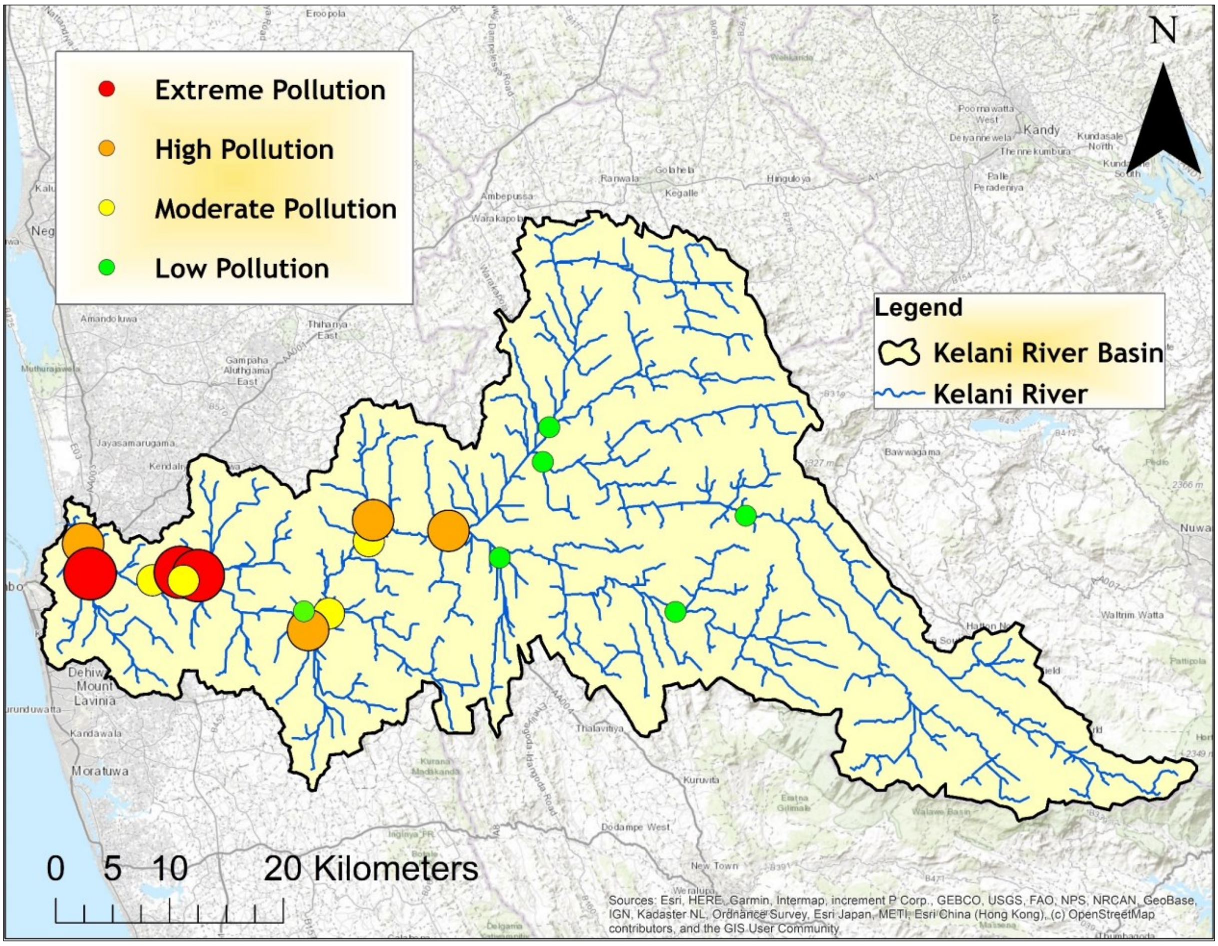


Figure 5.4 Spatial distribution of sewage pollution sources in the Kelani River Basin

Another key result identified in the KRB-IPI was that the highest exceedance of standard limits was observed for cadmium (Cd) and iron (Fe). Similarly, Cd and Fe had the highest weighted values in the KRB-IPI analysis. This indicates that Cd and Fe are the two primary contributors to the overall heavy metal pollution status of the Kelani River. Throughout the study period, Cd concentrations exceeded standard limits by 29% to 32% across all 17 stations, while Fe concentrations exceeded limits by 2% to 37%, which explains the moderate pollution scores for most stations, except those identified as key pollutant sources. The frequent exceedance of Cd limits underscores the need for immediate countermeasures, particularly due to the presence of water treatment plants along the Kelani River and the variation of Fe concentrations in KRB highlights the variation of industrial pollution in the KRB.

5.3 Water Quality Modelling

The KRB-IPI identified five stations having high to extreme levels of pollution in the KRB. Stations Kollonnawa Bridge (S2), Raggahawatte Ela (S4), Pusseli Oya (S7), and Pugoda Ela (S11) were evaluated as stations with high pollution, and Maha Ela (S6) station was identified as having extreme pollution. All these stations were located in sub-catchments of the KRB hence the developed WASP model was used to assess the impact of each of these sources of pollution to the main river. The KRB-IPI identified Fe as a prominent pollutant in the basin and has a spatial variation across the KRB, hence Fe was selected as the pollutant for modelling.

The developed HEC-HMS model generated hydrographs for the six ungauged sub-catchments of the KRB. The applied hybrid modelling approach increased the accuracy of the model predictions. The hybrid model predicted the six sub-catchment hydrographs with a reduced error. The hybrid modelling approach was effective in reducing the error of the process-based model and increasing the accuracy of the simulated river flow. The error-corrected sub-catchment hydrographs were used in conjunction with the streamflow observations of the main river to develop the HEC-RAS model. The calibrated and validated HEC-RAS model was used to generate the flow rates, velocities, and water depths for each of the cross-sections used. These data were used to divide the segments in the WASP model and to calculate the flow characteristics for each segment. The segment characteristics were used to develop the WASP model in conjunction with the measured water quality concentrations.

The developed WASP model was then used to assess the impact of the five identified industrial pollutant sources on the water pollution status of the Kelani River. The impact of these sources was investigated by comparing the Fe concentrations of the main river before the tributary joined the main river and after the tributary joined the main river. The Mean Absolute Percentage Error (MAPE) value was used to quantify the difference in the pollution level of the two instances. The Fe concentrations of the main river before the tributary was connected were used as the reference and were assessed with the Fe concentrations after the tributary connected to the main river. The MAPE value suggests the mean relative deviation between the two data series as a percentage. The calculated MAPE values for each of the five tributaries are shown in Table 5.1.

The results demonstrate that the highest contribution to the pollution status of the river is caused by Pugoda Ela station and Maha Ela station. Pusseli Oya station also demonstrates a significant contribution to the pollution of the Kelani River compared with the other stations. The lowest contribution to the pollution of the Kelani River is caused by Kollonnawa Ela and Raggahawatte Ela.

Pugoda Ela station (S11) was identified as having high pollution in the KRB. It was also observed that increased pollution was present in the station during the SWM and IM2. The MAPE values suggest that the Pugoda Ela station has the highest impact on the pollution status of the Kelani River as the highest MAPE value is calculated for the Pugoda Ela station. It can be observed in Figure 5.5 that the higher pollution concentrations at Pugoda Ela station were observed during times of increased streamflow, necessarily during SWM and IM2 periods. Although high concentrations were observed during these periods, it can be observed that when joining the main river these concentrations were diluted due to the increased streamflow. Similarly, it can be observed that the difference in concentrations in the main river upstream and downstream of Pugoda Ela show significant deviations during the periods of increased river flow.

Table 5.1 Calculated Mean Absolute Percentage Error (MAPE) values for the five tributaries identified as pollution sources in the KRB by KRB-IPI

Tributary	**MAPE value**
Pugoda Ela (S11)	16.21%
Pusseli Oya (S7)	8.27%
Maha Ela (S6)	10.09%
Raggahawatte Ela (S4)	1.41%
Kollonnawa (S2)	1.26%

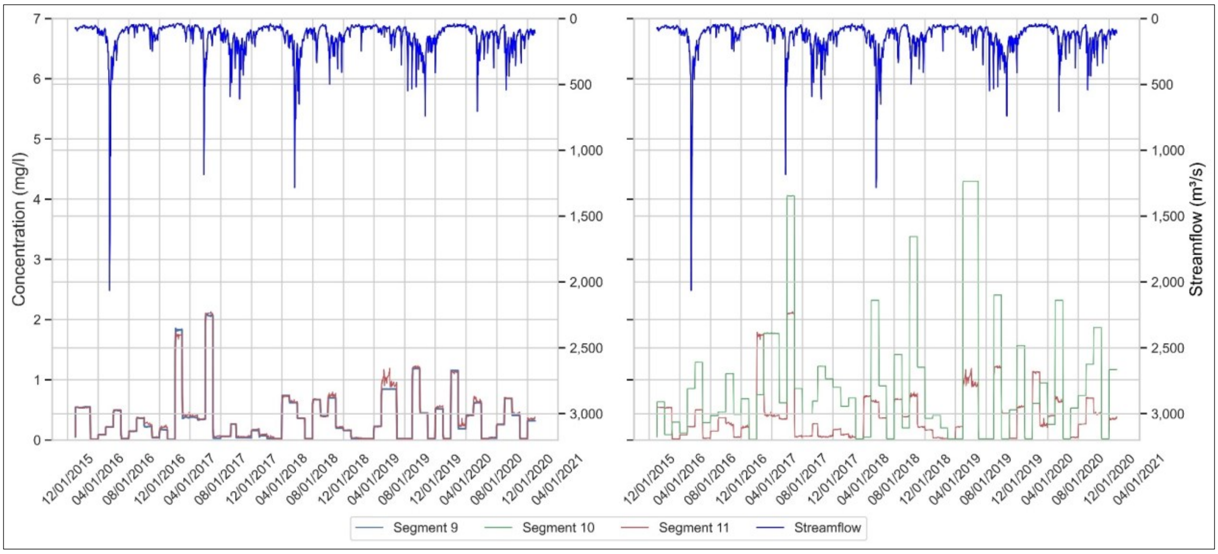


Figure 5.5 The impact of Pugoda Ela (S11) station on the pollution of the Kelani River. Segment 9 and Segment 11 represent the segments before and after the tributary is joined to the main river respectively. Segment 10 represents the Pugoda Ela station located in the tributary

Maha Ela station (S6) was categorized as having extreme pollution by KRB-IPI and it is evident that the contribution of Maha Ela station to the Kelani River pollution is comparatively higher. Similarly, a MAPE value of 10.09% demonstrates that the Maha Ela station is one of the significant pollution sources in the Kelani River. Similar to the result identified at Pugoda Ela station, the Maha Ela station also demonstrates significant pollution during periods of increased river flow as shown in Figure 5.6. It can be observed that although higher concentrations of Fe can be observed at Maha Ela station (Segment 42), the concentrations are reduced when connected to the main river (Segment 43) during periods of increased river flow. Similar to Pugoda Ela, significant deviations can be observed between the Fe concentrations at segments upstream and downstream of the Maha Ela station during periods of increased river flow.

Similar to these two stations Pusseli Oya (Figure 5.7), Raggahawatte Ela (Figure 5.8), and Kollonnawa (Figure 5.9) stations also demonstrate similar trends suggesting increased pollution in the Kelani River during periods of increased river flow and rainfall. Similarly, these three stations also demonstrate the deviations of Fe concentrations at segments upstream and downstream of the tributaries during periods of increased river flow.

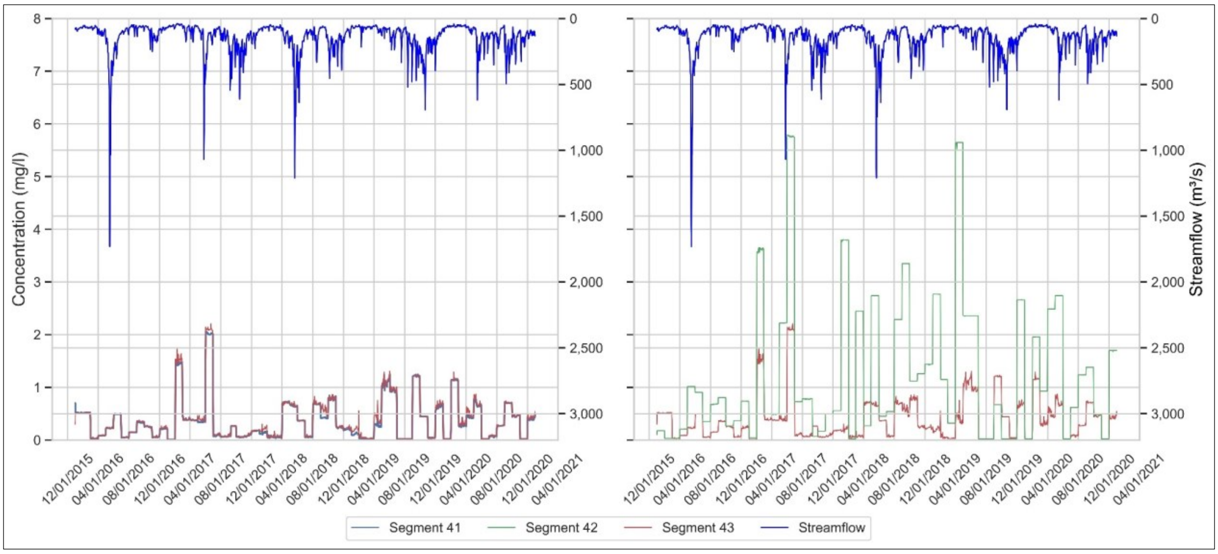


Figure 5.6 The impact of Maha Ela (S6) station on the pollution of the Kelani River. Segment 41 and Segment 43 represent the segments before and after the tributary is joined to the main river respectively. Segment 42 represents the Maha Ela station located in the tributary

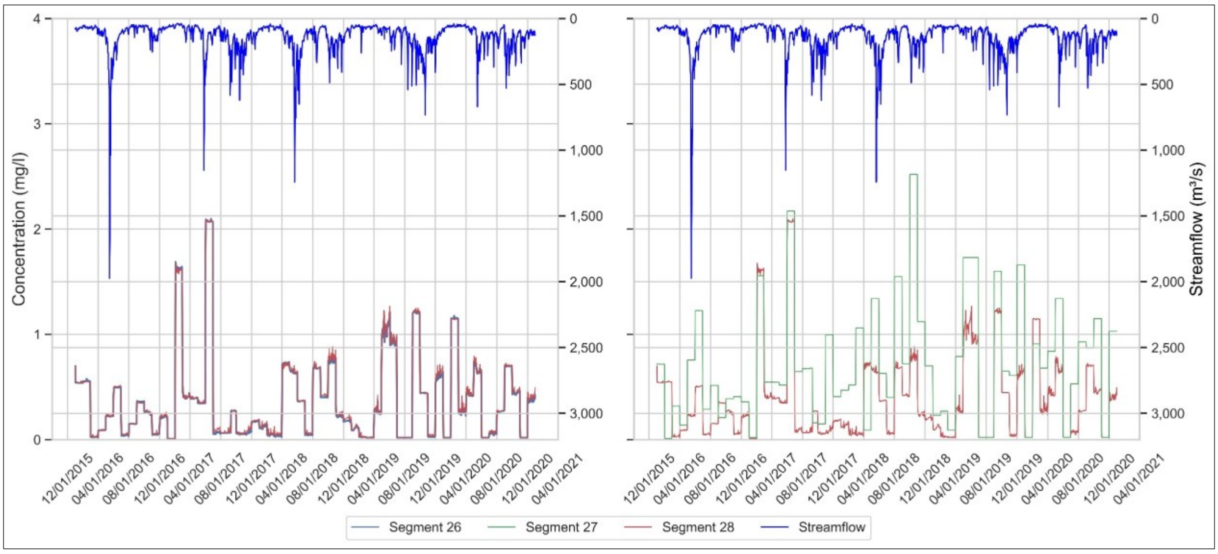


Figure 5.7 The impact of Pusseli Oya (S7) station on the pollution of the Kelani River. Segment 26 and Segment 28 represent the segments before and after the tributary is joined to the main river respectively. Segment 27 represents the Pusseli Oya station located in the tributary

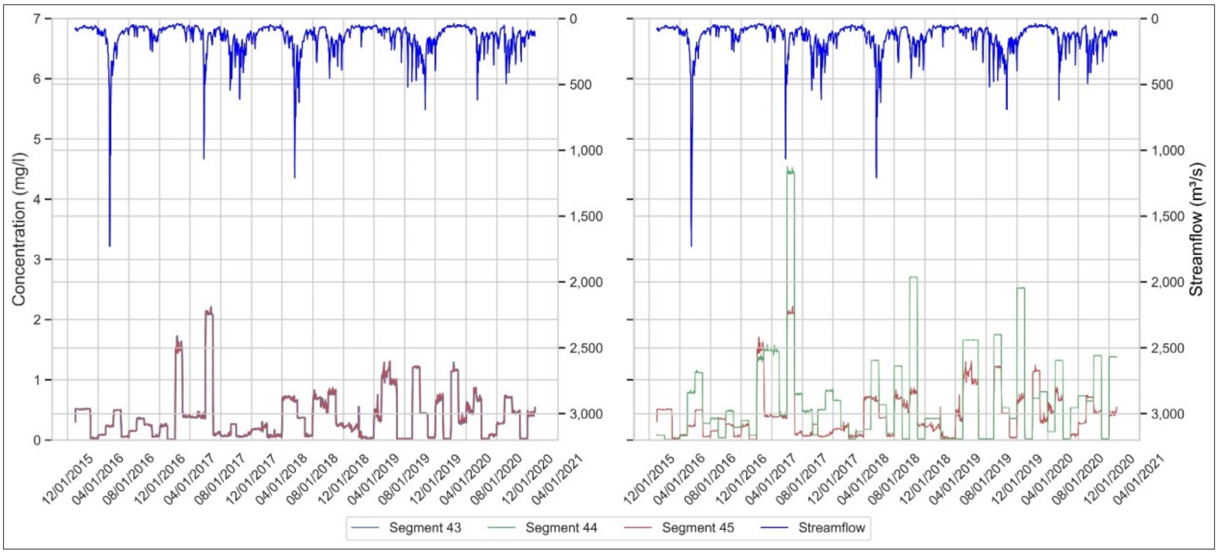


Figure 5.8 The impact of Raggahawatte Ela (S4) station on the pollution of the Kelani River. Segment 43 and Segment 45 represent the segments before and after the tributary is joined to the main river respectively. Segment 44 represents the Raggahawatte Ela station located in the tributary

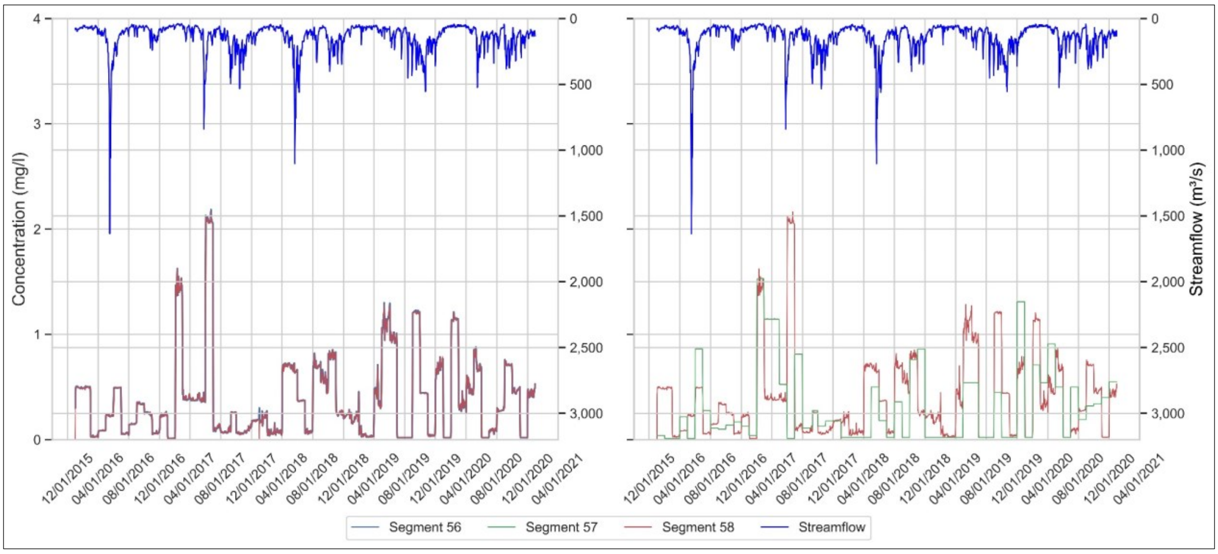


Figure 5.9 The impact of Kollonnawa (S2) station on the pollution of the Kelani River. Segment 56 and Segment 58 represent the segments before and after the tributary is joined to the main river respectively. Segment 57 represents the Kollonnawa Ela station located in the tributary

The results suggest that although high to extreme pollution can be observed in the tributaries, the effect of the pollution is reduced due to the comparatively higher flow in the main river. However, significant deviations of Fe concentrations can be observed in the Pugoda Ela, Pusseli Oya, and Maha Ela stations as the higher pollutant concentrations affect the pollution of the main river. Similarly, it can be observed that comparatively higher pollution can be observed during periods of increased river flow as suggested by the KRB-IPI. The dilution of pollutant concentrations is critical in managing the pollution in river flow discussed above. According to the results, the Pugoda Ela (S11) station is the main source of pollution in terms of industrial pollution in the KRB followed by the Maha Ela (S6) station.

CHAPTER 6

6 CONCLUSIONS

The conclusions of the study can be summarized as shown below.

1. Pollution due to industrial effluents and pollution due to municipal, domestic, and industrial sewage are the main pollution methods in the Kelani River Basin (KRB). Out of five latent variables (LVs) accounting for 77% of the variation of 17 water quality parameters, four latent variables correlated with industrial pollution and sewage pollution.
2. The spatial variation analysis of the LVs suggested that higher variations in water quality occurred near industrial zones and urbanized areas. Similarly, the seasonal variation analysis suggested a strong correlation between monsoonal patterns and water quality dynamics in the KRB.
3. The developed Kelani River Basin-Industrial Pollution Index (KRB-IPI) and Kelani River Basin-Sewage Pollution Index (KRB-SPI) identified potential pollutant sources in the KRB related to industrial pollution and sewage pollution. The Kollonnawa, Raggahawatte Ela, Maha Ela, Pusseli Oya, and Pugoda Ela stations were identified by KRB-IPI as major industrial pollutant sources. Similarly, Victoria, Kollonnawa, Raggahawatte Ela, Maha Ela, Pugoda Ela, and Seethawake stations were identified by KRB-SPI as major sewage pollution sources.
4. The results of both the KRB-IPI and KRB-SPI revealed that increased pollution was observed during periods of increased rainfall and streamflow. This suggested that significant increase in pollutant loads during heavy rainfall, or that the runoff is mobilizing pollutants from the surrounding areas.
5. It was observed that Cd and Fe are the most prominent industrial pollutants in the basin with Cd concentrations exceeding standard limits by 29% to 32% across all 17 stations, while Fe concentrations exceeding limits by 2% to 37% across all the stations.
6. Results of the water quality simulation model revealed that Pugoda Ela and Maha Ela stations had the highest impact out of the identified industrial pollution sources, on the river pollution of the Kelani River with Mean Absolute Percentage Error values of 16.21% and 10.09% respectively.
7. It was observed that although high concentrations were observed during periods of increased river flow, when joining the main river these concentrations were diluted due to the increased

streamflow. Subsequently, all five stations demonstrate the deviations of Fe concentrations at segments upstream and downstream of the tributaries during periods of increased river flow.

8. The integration of machine learning and process-based models proved highly beneficial in this study, significantly improving the results and overall accuracy. The combination of HEC-HMS and LSTM ANN effectively corrected errors in simulated data using residual errors, even in the absence of observations. Additionally, machine learning facilitated the development of tailored pollution indices based on local data distributions, while unsupervised learning methods uncovered patterns in water pollution without requiring prior knowledge.
9. The developed method can be applied to any river basin to identify the major pollutant types in the basin, to identify the pollution hotspots in the basin, and to quantify the effect of major pollution sources on the main river.

CHAPTER 7

7 RECOMMENDATIONS

Based on the findings of this study, the following recommendations are proposed for the sustainable management of water quality in the Kelani River.

1. The results of this study revealed that the pollution of the Kelani River correlated with the monsoonal patterns in the country. Further studies can be aimed at assessing the pollution of the Kelani River with the changes in monsoonal patterns due to climate change impacts.
2. The impact of sewage pollution can be further assessed by following the same methodology used in this study, thereby identifying the key segments in Kelani River affected by sewage pollution.
3. It is revealed that increased pollution can be observed during periods of increased rainfall and streamflow. Effective remediation and management plans should be implemented to identify the potential pollutant discharge into surrounding areas.
4. The stations Pugoda Ela, Maha Ela, and Pusseli Oya should be monitored regularly to reduce the industrial pollutants discharged into the river ecosystem and for the sustainable management of the river.
5. It was identified that more frequent monitoring of the water quality parameters would help more accurate results in pollution modelling and identifying pollution sources in the basin.
6. Further studies on the applicability of other machine learning techniques in water pollution assessment should be carried out. The effective application of machine learning-based frameworks combining process-based modelling should be implemented for more effective modelling and source identification.
7. The public should be informed that the water pollution status of the Kelani River worsens during periods of increased rainfall. Additionally, pollution levels near industrial zones are notably high, while sewage pollution is distributed across the lower Kelani River Basin. This highlights the risk of illnesses associated with using water without proper sanitation facilities.
8. The water pollution near the existing water intakes for water treatment plants should be monitored regularly, as heavy metal pollution poses a serious threat to public health and safety.

BIBLIOGRAPHY

Abba, S. I., Hadi, S. J., Sammen, S. S., Salih, S. Q., Abdulkadir, R. A., Pham, Q. B., & Yaseen, Z. M. (2020). Evolutionary computational intelligence algorithm coupled with a self-tuning predictive model for water quality index determination. *Journal of Hydrology*, *587*(April), 124974. https://doi.org/10.1016/j.jhydrol.2020.124974

Abukmeil, M., Ferrari, S., Genovese, A., Piuri, V., & Scotti, F. (2021). A Survey of Unsupervised Generative Models for Exploratory Data Analysis and Representation Learning. *ACM Computing Surveys*, *54*(5), 1–37. https://doi.org/10.1145/3450963

Adjovu, G. E., Stephen, H., James, D., & Ahmad, S. (2023). Overview of the Application of Remote Sensing in Effective Monitoring of Water Quality Parameters. *Remote Sensing*, *15*(7). https://doi.org/10.3390/rs15071938

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, *54*(6), 550–560. https://doi.org/10.1002/asi.10242

Ahmed Salman, H., Kalakech, A., Steiti, A., & History, A. (2024). *Random Forest Algorithm Overview A R T I C L E I N F O. 2024*, 69–79. https://doi.org/10.58496/BJML/2024/007;

Ait-Kadi, M. (2016). Water for Development and Development for Water: Realizing the Sustainable Development Goals (SDGs) Vision. *Aquatic Procedia*, *6*, 106–110. https://doi.org/10.1016/j.aqpro.2016.06.013

Akhtar, N., Ishak, M. I. S., Ahmad, M. I., Umar, K., Md Yusuff, M. S., Anees, M. T., Qadir, A., & Almanasir, Y. K. A. (2021). Modification of the water quality index (Wqi) process for simple calculation using the multi-criteria decision-making (mcdm) method: A review. *Water (Switzerland)*, *13*(7). https://doi.org/10.3390/w13070905

Akhtar, N., Syakir Ishak, M. I., Bhawani, S. A., & Umar, K. (2021). Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water (Switzerland)*, *13*(19). https://doi.org/10.3390/w13192660

Akiner, M. E., Kartal, V., Guzeler, A. C., & Karakoyun, E. (2024). Exploring the applicability of the experiment-based ANN and LSTM models for streamflow estimation. *Earth Science Informatics*, 3111–3135. https://doi.org/10.1007/s12145-024-01332-4

Alhassan, A. M., & Wan Zainon, W. M. N. (2021). Review of Feature Selection, Dimensionality Reduction and Classification for Chronic Disease Diagnosis. *IEEE Access*, *9*, 87310–87317. https://doi.org/10.1109/ACCESS.2021.3088613

Ambrose, R. B., Wool, T. A., & Barnwell, T. O. (2011). Development of water quality modelling in the

United States. *Environmental Engineering Research*, *16*(4), 200–210. https://doi.org/10.4491/eer.2009.14.4.200

Arash, A. M., & Yasi, M. (2023). The assessment for selection and correction of RS-based DEMs and 1D and 2D HEC-RAS models for flood mapping in different river types. *Journal of Flood Risk Management*, *16*(1), 1–15. https://doi.org/10.1111/jfr3.12871

Ashrafuzzaman, M., Artemi, C., Santos, F. D., & Schmidt, L. (2022). Current and Future Salinity Intrusion in the South-Western Coastal Region of Bangladesh. *Spanish Journal of Soil Science*, *12*(March), 1–20. https://doi.org/10.3389/sjss.2022.10017

Aslam, B., Maqsoom, A., Cheema, A. H., Ullah, F., Alharbi, A., & Imran, M. (2022). Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach. *IEEE Access*, *10*(September), 119692–119705. https://doi.org/10.1109/ACCESS.2022.3221430

Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry*, *36*(1), 41–59. https://doi.org/10.1080/10408340500526766

Athukorala, S. W., Weerasinghe, L. S., Jayasooria, M., Rajapakshe, D., Fernando, L., Raffeeze, M., & Miguntanna, N. P. (2013). *Analysis of Water Quality Variation in Kelani River, Sri Lanka Using Principal Component Analysis*. *2013*, 129–135.

Aydin, H., Ustaoğlu, F., Tepe, Y., & Soylu, E. N. (2021). Assessment of water quality of streams in northeast Turkey by water quality index and multiple statistical methods. *Environmental Forensics*, *22*(1–2), 270–287. https://doi.org/10.1080/15275922.2020.1836074

Azam, H. M., Alam, S. T., Hasan, M., Yameogo, D. D. S., Kannan, A. D., Rahman, A., & Kwon, M. J. (2019). Phosphorous in the environment: characteristics with distribution and effects, removal mechanisms, treatment technologies, and factors affecting recovery as minerals in natural and engineered systems. *Environmental Science and Pollution Research*, *26*(20), 20183–20207. https://doi.org/10.1007/s11356-019-04732-y

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning. *Clinical Pharmacology and Therapeutics*, *107*(4), 871–885. https://doi.org/10.1002/cpt.1796

Ballesté, E., Demeter, K., Masterson, B., Timoneda, N., Sala-Comorera, L., & Meijer, W. G. (2020). Implementation and integration of microbial source tracking in a river watershed monitoring plan. *Science of the Total Environment*, *736*. https://doi.org/10.1016/j.scitotenv.2020.139573

Banda, T. D., & Kumarasamy, M. V. (2020). Development of water quality indices (WQIs): A review. *Polish Journal of Environmental Studies*, *29*(3), 2011–2021. https://doi.org/10.15244/pjoes/110526

Barati Moghaddam, M., Mazaheri, M., & Mohammad Vali Samani, J. (2021). Inverse modeling of contaminant transport for pollution source identification in surface and groundwaters: a review. *Groundwater for Sustainable Development*, *15*(August), 100651. https://doi.org/10.1016/j.gsd.2021.100651

Bayissa, Y. A., Tadesse, T., Svoboda, M., Wardlow, B., Poulsen, C., Swigart, J., & Van Andel, S. J. (2019). Developing a satellite-based combined drought indicator to monitor agricultural drought: a case study for Ethiopia. *GIScience and Remote Sensing*, *56*(5), 718–748. https://doi.org/10.1080/15481603.2018.1552508

Biau, G., & Scornet, E. (2016). A random forest-guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Boyd, C. E. (2015). Water quality: An introduction. In *Water Quality: An Introduction*. https://doi.org/10.1007/978-3-319-17446-4

Brady, J. P., Ayoko, G. A., Martens, W. N., & Goonetilleke, A. (2015). Development of a hybrid pollution index for heavy metals in marine and estuarine sediments. *Environmental Monitoring and Assessment*, *187*(5). https://doi.org/10.1007/s10661-015-4563-x

Brunner, G. W. (2024). HEC-RAS HEC-RAS Hydraulic Reference. *U.S. Army Corps of Engineers*, *6.5*(May), 1–477. https://www.hec.usace.army.mil/confluence/rasdocs/ras1dtechref/latest/introduction

Casillas-García, L. F., de Anda, J., Yebra-Montes, C., Shear, H., Díaz-Vázquez, D., & Gradilla-Hernández, M. S. (2021). Development of a specific water quality index for the protection of aquatic life of a highly polluted urban river. *Ecological Indicators*, *129*. https://doi.org/10.1016/j.ecolind.2021.107899

Central Environmental Authority (2017). *Ministry of Mahaweli Development and Environment ANNUAL REPORT - 2017*.

Central Environmental Authority (2019). Ambient water quality standards. *The Gazette of the Democratic Socialist Republic of Sri Lanka*, *2017*(I), 4. http://www.cea.lk/web/images/pdf/epc/2148-20_E-1.pdf

Central Environmental Authority (2021). *Annual Report 2021 Central Environmental Authority*.

Cheng, C., Zhang, F., Shi, J., & Kung, H. Te. (2022). What is the relationship between land use and surface water quality? A review and prospects from a remote sensing perspective. *Environmental Science and Pollution Research*, *29*(38), 56887–56907. https://doi.org/10.1007/s11356-022-21348-x

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623

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). Springer Netherlands. https://doi.org/10.1007/s11157-023-09650-7

Chueh, Y. Y., Fan, C., & Huang, Y. Z. (2021). Copper concentration simulation in a river by SWAT-WASP integration and its application to assessing the impacts of climate change and various remediation strategies. *Journal of Environmental Management*, *279*(October), 111613. https://doi.org/10.1016/j.jenvman.2020.111613

Chunju, Z. (2018). Overview of Prominent Problems in Huai River Basin, China. *International Journal of Hydrology*, *2*(1), 9–12. https://doi.org/10.15406/ijh.2018.02.00041

Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018). Unsupervised Learning Based On Artificial Neural Network: A Review. *2018 IEEE International Conference on Cyborg and Bionic Systems, CBS 2018*, *July 2021*, 322–327. https://doi.org/10.1109/CBS.2018.8612259

Diwyanjalee, G. R., & Premarathne, W. A. P. J. (2024). Impact of rainfall on the water quality of a tropical river: based on the Nilwala River in the southern province of Sri Lanka between March and October 2019. *Water Practice and Technology*, *19*(6), 2352–2363. https://doi.org/10.2166/wpt.2024.119

du Plessis, A. (2022). Persistent degradation: Global water quality challenges and required actions. *One Earth*, *5*(2), 129–131. https://doi.org/10.1016/j.oneear.2022.01.005

Dysarz, T. (2020). Development of methodology for assessment of long-term morphodynamic impact on flood hazard. *Journal of Flood Risk Management*, *13*(4), 1–24. https://doi.org/10.1111/jfr3.12654

Ernst, M. R., & Owens, J. (2009). Development and application of a WASP model on a large Texas reservoir to assess eutrophication control. *Lake and Reservoir Management*, *25*(2), 136–148. https://doi.org/10.1080/07438140902821389

Ewaid, S. H., Abed, S. A., & Kadhum, S. A. (2018). Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. *Environmental Technology and Innovation*, *11*, 390–398. https://doi.org/10.1016/j.eti.2018.06.013

Fan, X., Zhang, L., Yuan, L., Guo, B., Zhang, Q., & Huang, H. (2022). Urbanization and water quality dynamics and their spatial correlation in coastal margins of mainland China. *Ecological Indicators*, *138*(January), 108812. https://doi.org/10.1016/j.ecolind.2022.108812

Fida, M., Li, P., Wang, Y., Alam, S. M. K., & Nsabimana, A. (2023). Water Contamination and Human Health Risks in Pakistan: A Review. *Exposure and Health*, *15*(3), 619–639. https://doi.org/10.1007/s12403-022-00512-1

Figueroa-Nieves, D., McDowell, W. H., Potter, J. D., Martínez, G., & Ortiz-Zayas, J. R. (2014). Effects of Sewage Effluents on Water Quality in Tropical Streams. *Journal of Environmental Quality*, *43*(6), 2053–2063. https://doi.org/10.2134/jeq2014.03.0139

Fořt, J., Kobetičová, K., Böhm, M., Podlesný, J., Jelínková, V., Vachtlová, M., Bureš, F., & Černý, R. (2022). Environmental Consequences of Rubber Crumb Application: Soil and Water Pollution. *Polymers*, *14*(7). https://doi.org/10.3390/polym14071416

Mahalakshmi, G., M. Kumar, T. R. (2018). Assessment of Surface Water Quality of Noyyal River Using Wasp Model. *Asian Journal of Engineering and Applied Technology*, *7*(1), 37–40. https://doi.org/10.51983/ajeat-2018.7.1.865

Gaya, M. S., Abba, S. I., Abdu, A. M., Tukur, A. I., Saleh, M. A., Esmaili, P., & Wahab, N. A. (2020). Estimation of water quality index using artificial intelligence approaches and multi-linear regression. *IAES International Journal of Artificial Intelligence*, *9*(1), 126–134. https://doi.org/10.11591/ijai.v9.i1.pp126-134

Gordillo, G., Morales-Hernández, M., & García-Navarro, P. (2020). Finite volume model for the simulation of 1D unsteady river flow and water quality based on the WASP. *Journal of Hydroinformatics*, *22*(2), 327–345. https://doi.org/10.2166/hydro.2019.080

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, *11*(1), 40–56. https://doi.org/10.1016/j.apr.2019.09.009

Gunawardana, M. H. M. A. S. V., Sanjaya, K., Atapaththu, K. S. S., Yapa Mudiyanselage, A. L. W. Y., Masakorala, K., & Widana Gamage, S. M. K. (2022). Quantitative prediction of toxin-producing Aphanizomenon cyanobacteria in freshwaters using Sentinel-2 satellite imagery. *Journal of Water and Health*, *20*(9), 1364–1379. https://doi.org/10.2166/wh.2022.093

Gunawardena, A., Wijeratne, E. M. S., White, B., Hailu, A., & Pandit, R. (2017). Industrial pollution and the management of river water quality: a model of Kelani River, Sri Lanka. *Environmental Monitoring and Assessment*, *189*(9). https://doi.org/10.1007/s10661-017-6172-3

Guoshuai, Z., Anjun, D., Jianguo, C., Dangwei, W., Qin, L., & Yiqin, X. (2023). Improving the double mass curve method to analyze the variation trend of runoff and sediment load in watersheds. *LHB: Hydroscience Journal*, *109*(1). https://doi.org/10.1080/27678490.2023.2236971

Gupta, R., Singh, A. N., & Singhal, A. (2019). Application of ANN for water quality index. *International Journal of Machine Learning and Computing*, *9*(5), 688–693.

https://doi.org/10.18178/ijmlc.2019.9.5.859

Hallock, D. (n.d.). *A Water Quality Index for Ecology's Stream Monitoring Program Author : 02*.

Han, K., & Wang, Y. (2021). A review of artificial neural network techniques for environmental issues prediction. *Journal of Thermal Analysis and Calorimetry*, *145*(4), 2191–2207. https://doi.org/10.1007/s10973-021-10748-9

Hasan, M. 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

Hemachandra, C. K., & Pathiratne, A. (2018). Assessing toxicity of two industrial zone effluents reaching Kelani River, Sri Lanka. *Journal of the National Science Foundation of Sri Lanka*, *46*(4), 539–546. https://doi.org/10.4038/JNSFSR.V46I4.8629

Hoseinzadeh, E., Khorsandi, H., Wei, C., & 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

Hossain, M., & Patra, P. K. (2020). Water pollution index – A new integrated approach to rank water quality. *Ecological Indicators*, *117*(July), 106668. https://doi.org/10.1016/j.ecolind.2020.106668

Hurley, T., Sadiq, R., & Mazumder, A. (2012). Adaptation and evaluation of the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) for use as an effective tool to characterize drinking source water quality. *Water Research*, *46*(11), 3544–3552. https://doi.org/10.1016/j.watres.2012.03.061

Ighalo, J. O., & Adeniyi, A. G. (2020). A comprehensive review of water quality monitoring and assessment in Nigeria. *Chemosphere*, *260*, 127569. https://doi.org/10.1016/j.chemosphere.2020.127569

Islam, M. M., Lenz, O. K., Azad, A. K., Ara, M. H., Rahman, M., & Hassan, N. (2017). Assessment of Spatio-Temporal Variations in Water Quality of Shailmari River, Khulna (Bangladesh) Using Multivariate Statistical Techniques. *Journal of Geoscience and Environment Protection*, *05*(01), 1–26. https://doi.org/10.4236/gep.2017.51001

Jaafar, H. H., Ahmad, F. A., & El Beyrouthy, N. (2019). GCN250, new global gridded curve numbers for hydrologic modeling and design. *Scientific Data*, *6*(1), 1–9. https://doi.org/10.1038/s41597-019-0155-x

Jaffar, A., Thamrin, N. M., Ali, M. S. A. M., Misnan, M. F., & Yassin, A. I. M. (2022). Water Quality Prediction Using Lstm-Rnn: a Review. *Journal of Sustainability Science and Management*, *17*(7), 204–225. https://doi.org/10.46754/jssm.2022.07.015

Jayaminda, C., Gunawardhana, L., & Rajapakse, L. (2023). Rating Performances of Global Climate Models in Capturing Monsoon Rainfall Patterns in Sri Lanka. *Moratuwa Engineering Research Conference, MERCon*, 264–269. https://doi.org/10.1109/mERCon60487.2023.10355390

Karaouzas, I., Kapetanaki, N., Mentzafou, A., Kanellopoulos, T. D., & Skoulikidis, N. (2021). Heavy metal contamination status in Greek surface waters: A review with application and evaluation of pollution indices. *Chemosphere*, *263*. https://doi.org/10.1016/j.chemosphere.2020.128192

Khan, M., Khan, A. U., Khan, J., Khan, S., Haleem, K., & Khan, F. A. (2023). Streamflow forecasting for the Hunza river basin using ANN, RNN, and ANFIS models. *Water Practice and Technology*, *18*(5), 981–993. https://doi.org/10.2166/wpt.2023.060

Khouri, L., & Bashar Al-Moufti, M. (2022). Selection of suitable aggregation function for estimation of water quality index for the Orontes River. *Ecological Indicators*, *142*(August), 109290. https://doi.org/10.1016/j.ecolind.2022.109290

Kishor, R., Purchase, D., Ferreira, L. F. R., Mulla, S. I., Bilal, M., & Bharagava, R. N. (2020). Environmental and Health Hazards of Textile Industry Wastewater Pollutants and Its Treatment Approaches. *Handbook of Environmental Materials Management*, 1–24. https://doi.org/10.1007/978-3-319-58538-3_230-1

Knightes, C. D., Ambrose, R. B., Avant, B., Han, Y., Acrey, B., Bouchard, D. C., Zepp, R., & Wool, T. (2019). Modeling framework for simulating concentrations of solute chemicals, nanoparticles, and solids in surface waters and sediments: WASP8 Advanced Toxicant Module. *Environmental Modelling and Software*, *111*(November 2018), 444–458. https://doi.org/10.1016/j.envsoft.2018.10.012

Kodikara, K. A. S., Lewandowski, S., de Silva, P. M. C. S., Gunarathna, S. D., Madarasinghe, S. K., Ranasinghe, P., Jayatissa, L. P., & Dahdouh-Guebas, F. (2022). Heavy metal pollution in selected upland tributaries of Sri Lanka: comprehension towards the localization of sources of pollution. *Journal of Water and Health*, *20*(3), 505–517. https://doi.org/10.2166/WH.2022.250

Konapala, G., Kao, S. C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, *15*(10). https://doi.org/10.1088/1748-9326/aba927

Kumar, A. K., Sharma, S., Patel, A., Dixit, G., & Shah, E. (2019). Comprehensive evaluation of microalgal based dairy effluent treatment process for clean water generation and other value added products. *International Journal of Phytoremediation*, *21*(6), 519–530. https://doi.org/10.1080/15226514.2018.1537248

Kumar, A., Palmate, S. S., & Shukla, R. (2022). Water Quality Modelling, Monitoring, and Mitigation. *Applied Sciences (Switzerland)*, *12*(22), 1–7. https://doi.org/10.3390/app122211403

Kumar, D., & Alappat, B. J. (2004). Selection of the appropriate aggregation function for calculating Leachate Pollution Index. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, *8*(4), 253–264. https://doi.org/10.1061/(ASCE)1090-025X(2004)8:4(253)

Lindenschmidt, K. E., Fleischbein, K., & Baborowski, M. (2007). Structural uncertainty in a river water quality modelling system. *Ecological Modelling*, *204*(3–4), 289–300. https://doi.org/10.1016/j.ecolmodel.2007.01.004

Lukhabi, D. K., Mensah, P. K., Asare, N. K., Pulumuka-Kamanga, T., & Ouma, K. O. (2023). Adapted Water Quality Indices: Limitations and Potential for Water Quality Monitoring in Africa. *Water (Switzerland)*, *15*(9), 1–30. https://doi.org/10.3390/w15091736

Mahagamage, M. G. Y. L., & Manage, P. M. (2014). Water Quality Index (CCME-WQI) Based Assessment Study of Water Quality in Kelani River Basin, Sri Lanka. *International Journal of Environment and Natural Resources*, *1*(November 2014), 199–204.

Mahagamage, M. G. Y. L., & Manage, P. M. (2018). Socio-economic background of the head and transitional regions of the Kelani River Basin, Sri Lanka. *Asian Journal of Microbiology, Biotechnology and Environmental Sciences*, *20*(3), 744–756.

Makubura, R., Meddage, D. P. P., Azamathulla, H. M., Pandey, M., & Rathnayake, U. (2022). A Simplified Mathematical Formulation for Water Quality Index (WQI): A Case Study in the Kelani River Basin, Sri Lanka. *Fluids*, *7*(5). https://doi.org/10.3390/fluids7050147

Manage, P., Mahagamage, Y. L., Ajward, R., Amaratunge, S., & Amarathunga, V. I. (2020). The need for proper management leading to the sustainability of the Kelani River and its lower basin. *Journal of Water and Land Development*, *47*(1), 10–15. https://doi.org/10.24425/jwld.2020.135026

Marsland, S. (2020). Unsupervised Learning. *Machine Learning*, 211–236. https://doi.org/10.1201/9781420067194-13

Martin, J. L., Ambrose, R. B., & Wool, T. A. (2013). *WASP8 Macro Algae - Model Theory and User ' s Guide Supplement to Water Analysis Simulation Program (WASP) User Documentation*. 37.

Mawat, M., & Hamdan, A. N. (2023). Review of Mathematical Surface Water's Hydrodynamic/Water Quality Models with Their Application on the Shatt Al Arab River Southern Iraq. *European Journal of Engineering Science and Technology*, *6*(1), 30–49. https://doi.org/10.33422/ejest.v6i1.1057

Mbuh, M. J., Mbih, R., & Wendi, C. (2019). Water quality modeling and sensitivity analysis using Water Quality Analysis Simulation Program (WASP) in the Shenandoah River watershed. *Physical Geography*, *40*(2), 127–148. https://doi.org/10.1080/02723646.2018.1507339

Mei, K., Liao, L., Zhu, Y., Lu, P., Wang, Z., Dahlgren, R. A., & Zhang, M. (2014). Evaluation of spatial-temporal variations and trends in surface water quality across a rural-suburban-urban interface. *Environmental Science and Pollution Research*, *21*(13), 8036–8051. https://doi.org/10.1007/s11356-014-2716-z

Mohammed, R., & Al-Obaidi, B. (2021). Treatability influence of municipal sewage effluent on surface water quality assessment based on Nemerow pollution index using an artificial neural network. *IOP Conference Series: Earth and Environmental Science*, *877*(1). https://doi.org/10.1088/1755-1315/877/1/012008

Mulyani, H., Setiawan, R. A., & Fathi, H. (2023). Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data). *Journal of Information Technology and Its Utilization*, *6*(2), 45–50. https://doi.org/10.56873/jitu.6.2.5243

Nakhaei, M., Zanjanian, H., Nakhaei, P., Gheibi, M., Moezzi, R., Behzadian, K., & Campos, L. C. (2024). Comparative Evaluation of Deep Learning Techniques in Streamflow Monthly Prediction of the Zarrine River Basin. *Water (Switzerland)*, *16*(2). https://doi.org/10.3390/w16020208

Narangoda, C., Amarathunga, D., & Dangalle, C. D. (2023). Evaluation of water quality in the upper and lower catchments of the Kelani River Basin, Sri Lanka. *Water Practice and Technology*, *18*(3), 716–737. https://doi.org/10.2166/wpt.2023.034

Neupane, S., Vogel, J. R., Storm, D. E., Barfield, B. J., & Mittelstet, A. R. (2015). Development of a Turbidity Prediction Methodology for Runoff-Erosion Models. *Water, Air, and Soil Pollution*, *226*(12). https://doi.org/10.1007/s11270-015-2679-9

Oldham, L. D., Freer, J., Coxon, G., Howden, N., Bloomfield, J. P., & Jackson, C. (2023). Evidence-based requirements for perceptualising intercatchment groundwater flow in hydrological models. *Hydrology and Earth System Sciences*, *27*(3), 761–781. https://doi.org/10.5194/hess-27-761-2023

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, *11*(6), 583–605. https://doi.org/10.3233/ida-2007-11602

Organization, W. M. (2009). World Meteorological Organization GUIDELINES FOR CONVERTING. *Engineering*.

Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., & Huang, C. H. (2006). Assessment of seasonal variations in surface water quality. *Water Research*, *40*(20), 3800–3810. https://doi.org/10.1016/j.watres.2006.08.030

Pak, H. Y., Chuah, C. J., Yong, E. L., & Snyder, S. A. (2021). Effects of land use configuration, seasonality and point source on water quality in a tropical watershed: A case study of the Johor River Basin. *Science of the Total Environment*, *780*, 146661. https://doi.org/10.1016/j.scitotenv.2021.146661

Pathan, A. I., & Agnihotri, P. G. (2021). Application of new HEC-RAS version 5 for 1D hydrodynamic flood modeling with special reference through geospatial techniques: a case of River Purna at Navsari, Gujarat, India. *Modeling Earth Systems and Environment*, *7*(2), 1133–1144. https://doi.org/10.1007/s40808-020-00961-0

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prasad, B., Kumari, P., Bano, S., & Kumari, S. (2014). Ground water quality evaluation near mining area and development of heavy metal pollution index. *Applied Water Science*, *4*(1), 11–17. https://doi.org/10.1007/s13201-013-0126-x

Prathumratana, L., Sthiannopkao, S., & Kim, K. W. (2008). The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong River. *Environment International*, *34*(6), 860–866. https://doi.org/10.1016/j.envint.2007.10.011

Preisner, M. (2020). Surface Water Pollution by Untreated Municipal Wastewater Discharge Due to a Sewer Failure. *Environmental Processes*, *7*(3), 767–780. https://doi.org/10.1007/s40710-020-00452-5

Prjapathi, Saurabh.; Sabokruhie, Pouya.; Brinkmann, Markus.; and Lindenschmidt, K.-E. (2023). *South Saskatchewan River Using WASP — TOXI*. 1–14.

Ranasinghe, P. N., Siriwardana, Y. P. S., & Wanasinghe, V. R. (2006). Heavy metal pollution in drainage network of Colombo City and suburbs of Sri Lanka. *Chinese Journal of Geochemistry*, *25*(S1), 84–85. https://doi.org/10.1007/bf02839899

Ruvinda, K. M., & Pathiratne, A. (2018). Biomarker responses of Nile Tilapia (Oreochromis niloticus) exposed to polluted water from Kelani river basin, Sri Lanka: Implications for biomonitoring river pollution. *Sri Lanka Journal of Aquatic Sciences*, *23*(1), 105. https://doi.org/10.4038/sljas.v23i1.7551

Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica*, *67*(5), 1471–1481. https://doi.org/10.1007/s11600-019-00330-1

Salil Sahu, et al., S. S. et al. . (2020). Review on the HEC-HMS Rainfall-Runoff Simulation Model. *International Journal of Agricultural Science and Research*, *10*(4), 183–190. https://doi.org/10.24247/ijasraug202024

Schaffner, M., Bader, H. P., & Scheidegger, R. (2009). Modeling the contribution of point sources and non-point sources to Thachin River water pollution. *Science of the Total Environment*, *407*(17), 4902–4915. https://doi.org/10.1016/j.scitotenv.2009.05.007

Shabani, A., Woznicki, S. A., Mehaffey, M., Butcher, J., Wool, T. A., & Whung, P. Y. (2021). A coupled hydrodynamic (HEC-RAS 2D) and water quality model (WASP) for simulating flood-induced soil, sediment, and contaminant transport. *Journal of Flood Risk Management*, *14*(4), 1–17. https://doi.org/10.1111/jfr3.12747

Sharma, N., & Kaushal, A. (2021). Assessing Water Quality Models for Surface Water Resources: A Review. *Journal of Soil Salinity and Water Quality*, *13*(2), 135–146.

Sharma, V., Kumar, Y., & Kumar, A. (2016). Assessment of Physico Chemical Parameters for Analysing Water : A Review Assessment of Physico Chemical Parameters for Analysing Water : A Review. *J. Biol. Chem. Chron.*, *2*(1), 25–33.

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Shen, Z., Liao, Q., Hong, Q., & Gong, Y. (2012). An overview of research on agricultural non-point source pollution modelling in China. *Separation and Purification Technology*, *84*, 104–111. https://doi.org/10.1016/j.seppur.2011.01.018

Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4–11. https://doi.org/10.12691/ajams-9-1-2

Syeed, M. M. M., Hossain, M. S., Karim, M. R., Uddin, M. F., Hasan, M., & Khan, R. H. (2023). Surface water quality profiling using the water quality index, pollution index and statistical methods: A critical review. *Environmental and Sustainability Indicators*, *18*(January), 100247. https://doi.org/10.1016/j.indic.2023.100247

Treshansky, A., & McGraw, R. M. (2001). Overview of clustering algorithms. *Enabling Technology for Simulation Science V*, *4367*(September 2001), 41. https://doi.org/10.1117/12.440039

Tuan, A., & Babel, S. (2024). Science of the Total Environment Microplastics and heavy metals in a tropical river : Understanding spatial and seasonal trends and developing response strategies using DPSIR framework. *Science of the Total Environment*, *897*(March 2022), 165405. https://doi.org/10.1016/j.scitotenv.2023.165405

Uday Kiran, G., Vasumathi, D., & Roman, N. (2021). *Predicting Parkinson's Disease using Extreme Learning Measure and Principal Component Analysis based Mini SOM*. *25*(4), 16099–16111. http://annalsofrscb.ro

Udayashankar Banarjee. (2012). Source and distribution of Lead, Cadmium, Iron and Manganese in the river Damodar near Asansol Industrial Area, West Bengal, India. *International Journal of*

Environmental Sciences, *2*(3), 1531–1542. https://doi.org/10.6088/ijes.00202030038

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. G., Nash, S., Rahman, A., & Olbert, A. I. (2022). Development of a water quality index model - a comparative analysis of various weighting methods. *Mediterranean Geosciences Union Annual Meeting (MedGU-21)*, *December*, 1–6.

Uddin, M. G., Rahman, A., Nash, S., Diganta, M. T. M., Sajib, A. M., Moniruzzaman, M., & Olbert, A. I. (2023). Marine waters assessment using improved water quality model incorporating machine learning approaches. *Journal of Environmental Management*, *344*(June), 118368. https://doi.org/10.1016/j.jenvman.2023.118368

Uddin, M., Kormoker, T., Siddique, M. A. B., Billah, M. M., Rokonuzzaman, M., Al Ragib, A., Proshad, R., Hossain, M. Y., Haque, M. K., Ibrahim, K. A., & Idris, A. M. (2023). An overview on water quality, pollution sources, and associated ecological and human health concerns of the lake water of megacity: a case study on Dhaka city lakes in Bangladesh. *Urban Water Journal*, *20*(3), 261–277. https://doi.org/10.1080/1573062X.2023.2169171

UN-Water. (2021). Summary Progress Update 2021 : SDG 6 — water and sanitation for all. *UN-Water Integrated Monitoring Initiative*, 1–58. https://www.unwater.org/new-data-on-global-progress-towards-ensuring-water-and-sanitation-for-all-by-2030/

U.S. Army Corps of Engineers. (2024). *HEC-HMS user's manual (Version 4.10)*. Hydrologic Engineering Center. https://www.hec.usace.army.mil

US Army Corps of Engineers. (1998). HEC-HMS Technical Reference Manual Introduction. *Computer Manual*, 1–288.

USACE, & US Army Corps of Engineers. (1998). HEC-HMS Technical Reference Manual Introduction. *Computer Manual*, 1–288.

Ustaoğlu, F., Tepe, Y., & Taş, B. (2020). Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index. *Ecological Indicators*, *113*(October). https://doi.org/10.1016/j.ecolind.2019.105815

Vadyala, S. R., Betgeri, S. N., Matthews, J. C., & Matthews, E. (2022). A review of physics-based machine learning in civil engineering. *Results in Engineering*, *13*(October 2021), 100316. https://doi.org/10.1016/j.rineng.2021.100316

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Varekar, V., Karmakar, S., Jha, R., & Ghosh, N. C. (2015). Design of sampling locations for river water quality monitoring considering seasonal variation of point and diffuse pollution loads. *Environmental Monitoring and Assessment*, *187*(6). https://doi.org/10.1007/s10661-015-4583-6

Varshney, T., Waghmare, A. V., Singh, V. P., Ramu, M., Patnana, N., Meena, V. P., Azar, A. T., & Hameed, I. A. (2024). Investigation of rank order centroid method for optimal generation control. *Scientific Reports*, *14*(1), 1–21. https://doi.org/10.1038/s41598-024-61945-z

Wagle, N., Acharya, T. D., & Lee, D. H. (2020). Comprehensive review on application of machine learning algorithms for water quality parameter estimation using remote sensing data. *Sensors and Materials*, *32*(11), 3879–3892. https://doi.org/10.18494/SAM.2020.2953

Wang, Q., Zhang, Q., Wu, Y., & Wang, X. C. (2017). Physicochemical conditions and properties of particles in urban runoff and rivers: Implications for runoff pollution. *Chemosphere*, *173*, 318–325. https://doi.org/10.1016/j.chemosphere.2017.01.066

Wear, S. L., Acuña, V., McDonald, R., & Font, C. (2021). Sewage pollution, declining ecosystem health, and cross-sector collaboration. *Biological Conservation*, *255*. https://doi.org/10.1016/j.biocon.2021.109010

WHO. (2018). A global overview of national regulations and standards for drinking-water quality. In *Verordnung über die Qualitä t von Wasser für den menschlichen Gebrauch (Trinkwasserverordnung -TrinkwV 2001)*. http://apps.who.int/bookorders.

Wimordi, C., Yudianto, D., & Guan, Y. (2021). Application of WASP model to simulate water pollution control of Duriangkang Dam. *Lakes and Reservoirs: Science, Policy and Management for Sustainable Use*, *26*(1), 23–32. https://doi.org/10.1111/lre.12350

Wool, T., Ambrose, R. B., Martin, J. L., & Comer, A. (2020). WASP 8: The next generation in the 50-year evolution of USEPA's water quality model. *Water (Switzerland)*, *12*(5), 1–33. https://doi.org/10.3390/W12051398

Wu, J., Cheng, S. P., He, L. Y., Wang, Y. C., Yue, Y., Zeng, H., & Xu, N. (2023). Assessing water quality in the Pearl River for the last decade based on clustering: Characteristic, evolution and policy implications. *Water Research*, *244*(April), 120492. https://doi.org/10.1016/j.watres.2023.120492

Xia, C., Liu, G., Xia, H., Jiang, F., & Meng, Y. (2021). Influence of saline intrusion on the wetland ecosystem revealed by isotopic and hydrochemical indicators in the Yellow River Delta, China. *Ecological Indicators*, *133*, 108422. https://doi.org/10.1016/j.ecolind.2021.108422

Xia, F., Qu, L., Wang, T., Luo, L., Chen, H., Dahlgren, R. A., Zhang, M., Mei, K., & Huang, H. (2018). Distribution and source analysis of heavy metal pollutants in sediments of a rapid developing urban river system. *Chemosphere*, *207*, 218–228. https://doi.org/10.1016/j.chemosphere.2018.05.090

Xiong, Y., Zhang, T., Sun, X., Yuan, W., Gao, M., Wu, J., & Han, Z. (2023). Groundwater Quality Assessment Based on the Random Forest Water Quality Index—Taking Karamay City as an Example. *Sustainability (Switzerland)*, *15*(19). https://doi.org/10.3390/su151914477

Xu, G., Li, P., Lu, K., Tantai, Z., Zhang, J., Ren, Z., & Wang, X. (2019). Catena Seasonal changes in water quality and its main in fl uencing factors in the Dan River basin. *Catena*, *173*(April 2018), 131–140. https://doi.org/10.1016/j.catena.2018.10.014

Xu, H., Gao, Q., & Yuan, B. (2022). Analysis and identification of pollution sources of comprehensive river water quality: Evidence from two river basins in China. *Ecological Indicators*, *135*, 108561. https://doi.org/10.1016/j.ecolind.2022.108561

Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

Zainal, N. N., & Talib, S. H. A. (2024). Review paper on applications of the HEC-RAS model for flooding, agriculture, and water quality simulation. *Water Practice and Technology*, *19*(7), 2883–2900. https://doi.org/10.2166/wpt.2024.173

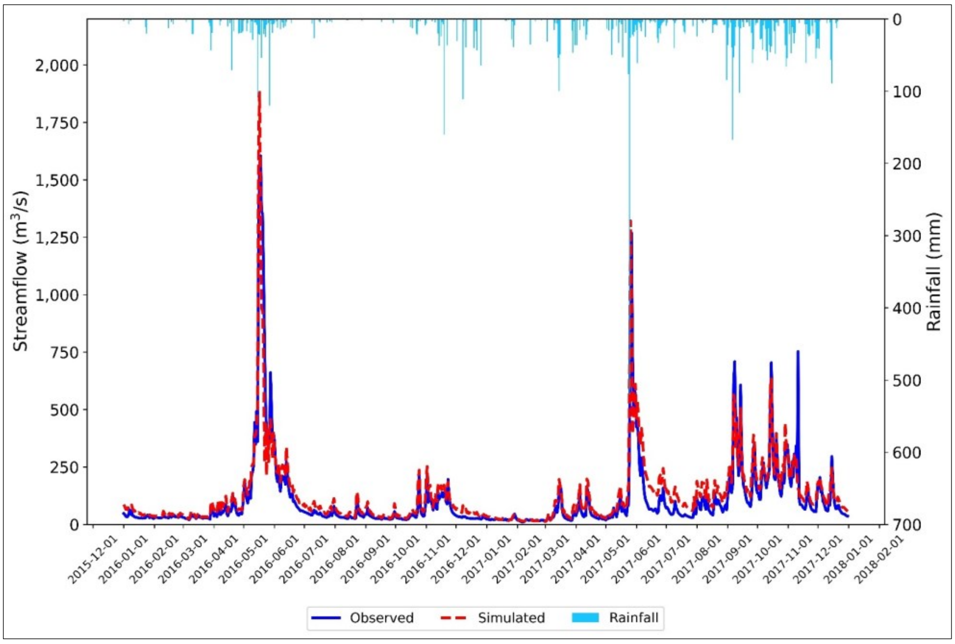
Zakaria, M., Al-Shebany, M., & Sarhan, S. (2014). Artificial Neural Network : A Brief Overview. *Journal of Engineering Research and Applications Www.Ijera.Com*, *4*(2), 7–12. www.ijera.com

Zare Garizi, A., Sheikh, V., & Sadoddin, A. (2011). Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods. *International Journal of Environmental Science and Technology*, *8*(3), 581–592. https://doi.org/10.1007/bf03326244

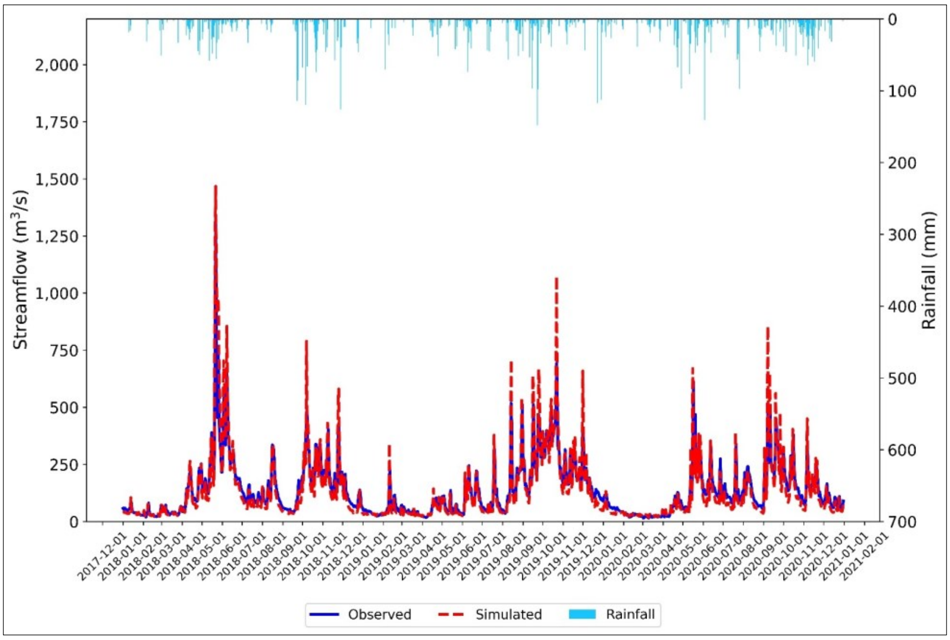
Zeinalzadeh, K., & Rezaei, E. (2017). Determining spatial and temporal changes of surface water quality using principal component analysis. *Journal of Hydrology: Regional Studies*, *13*(June), 1–10. https://doi.org/10.1016/j.ejrh.2017.07.002

APPENDIX A

TIME SERIES GRAPH FOR THE HEC-HMS MODEL AT NAGALAGAM STREET STATION



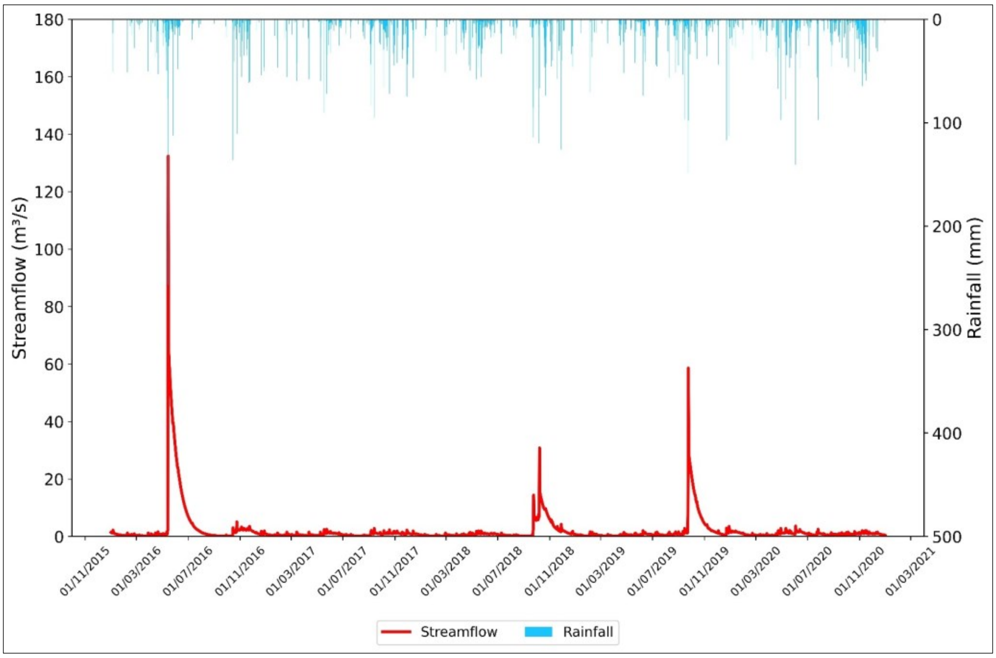
Appendix A–1: Time series graph for the calibration period at Nagalagam Street station



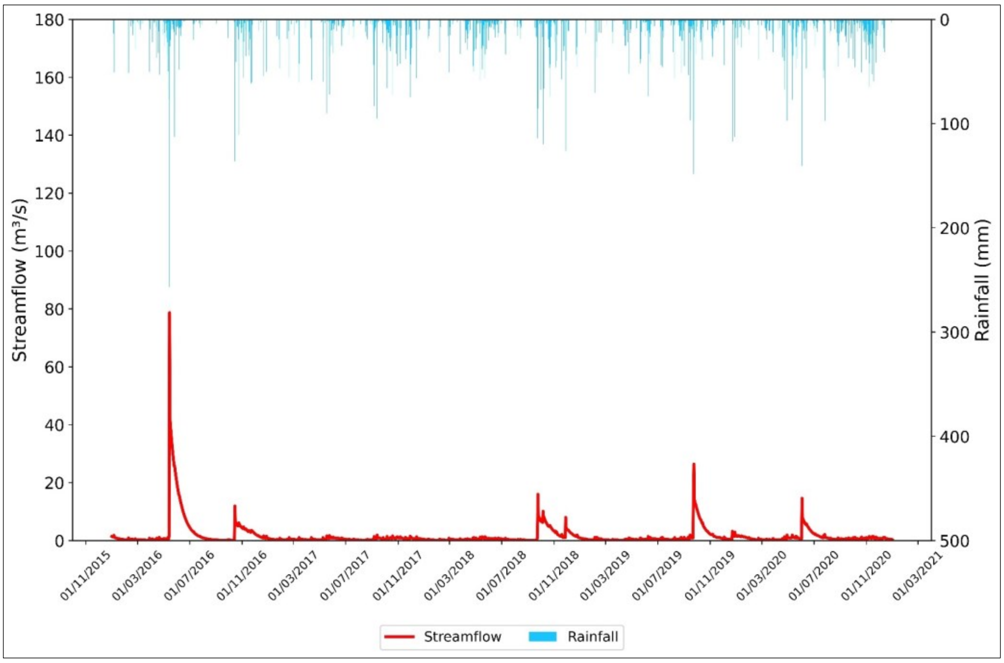
Appendix A–2: Time series graph for the validation period at Nagalagam Street station

APPENDIX B

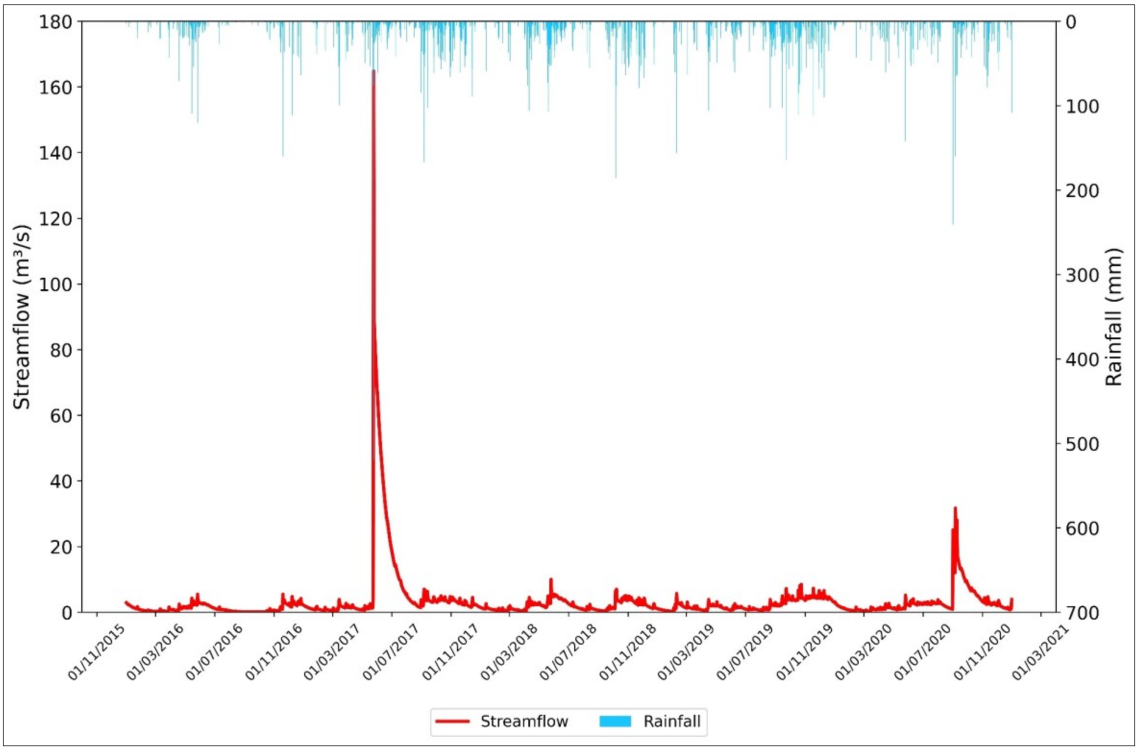
SUB-CATCHMENT HYDROGRAPHS GENERATED IN HEC-HMS



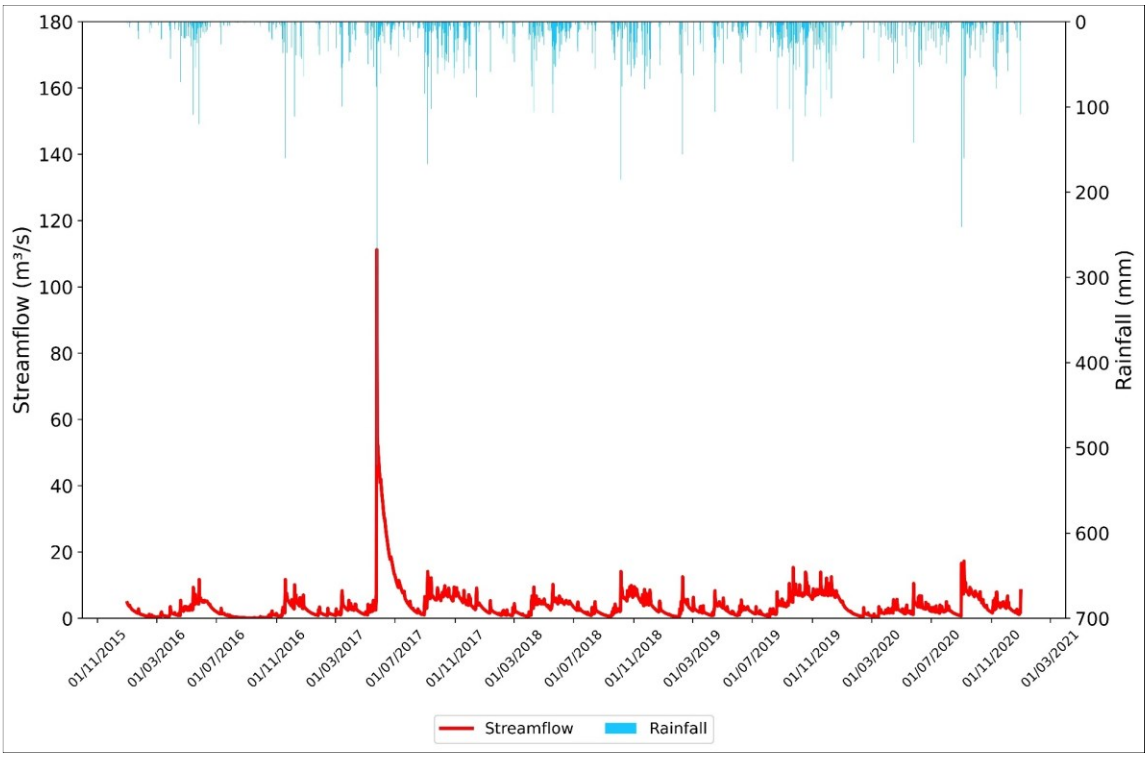
Appendix B–1: Generated sub-catchment hydrograph at Kollonnawa Ela



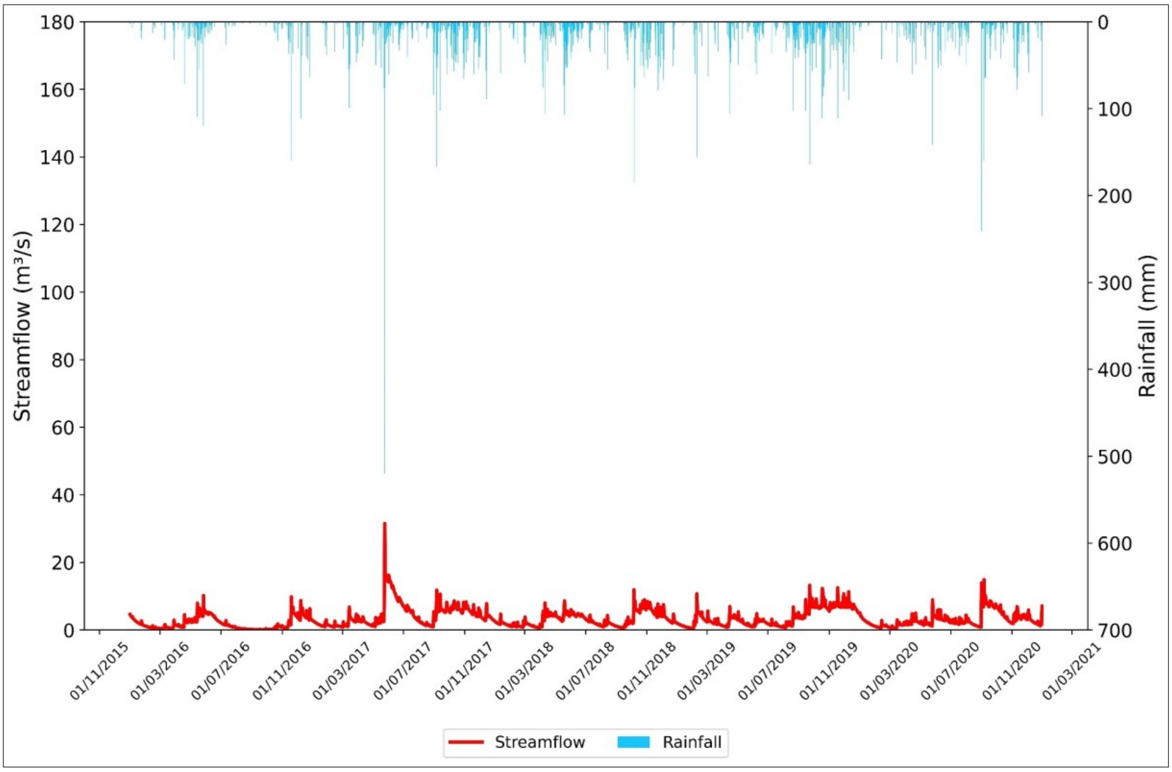
Appendix B–2: Generated sub-catchment hydrograph at Raggahawatte Ela



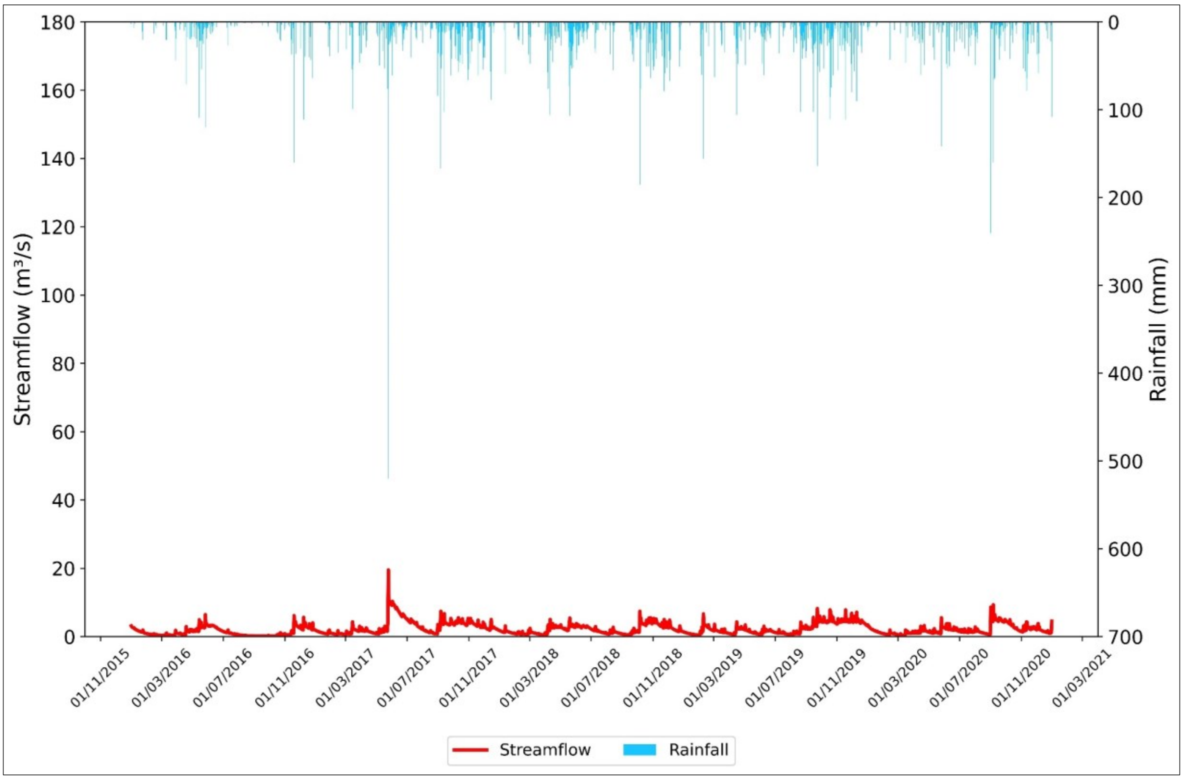
Appendix B–3: Generated sub-catchment hydrograph at Maha Ela



Appendix B–4: Generated sub-catchment hydrograph at Pusseli Oya



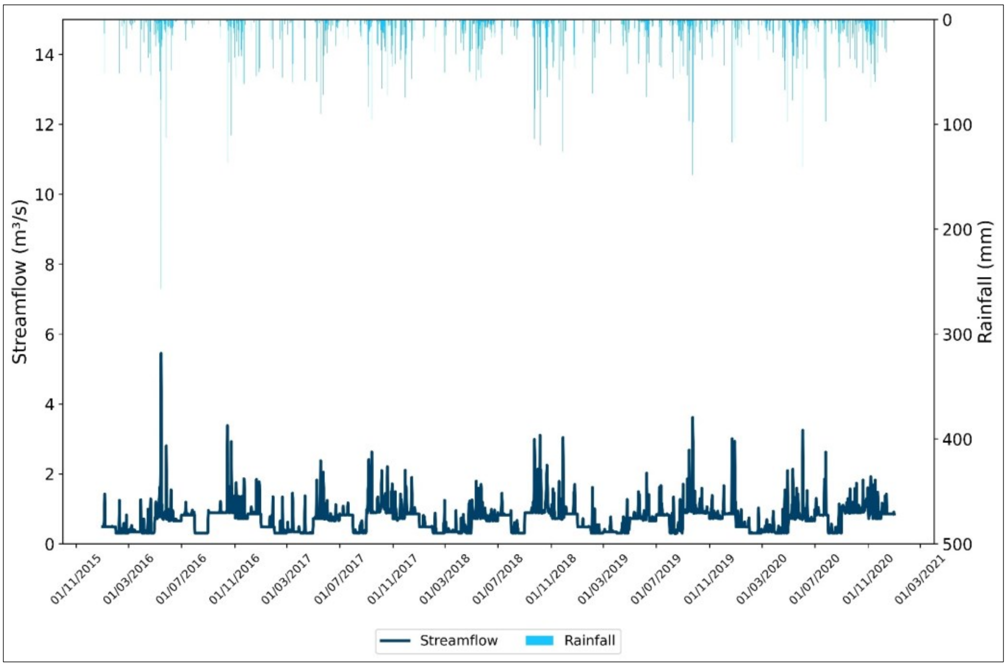
Appendix B–5: Generated sub-catchment hydrograph at Wak Oya



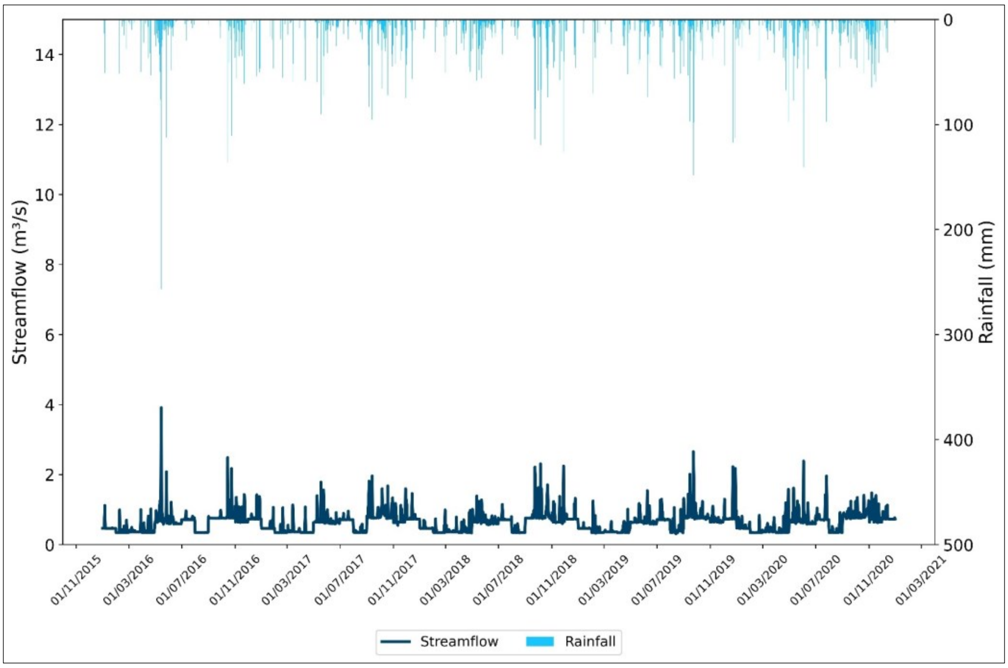
Appendix B–6: Generated sub-catchment hydrograph at Pugoda Ela

APPENDIX C

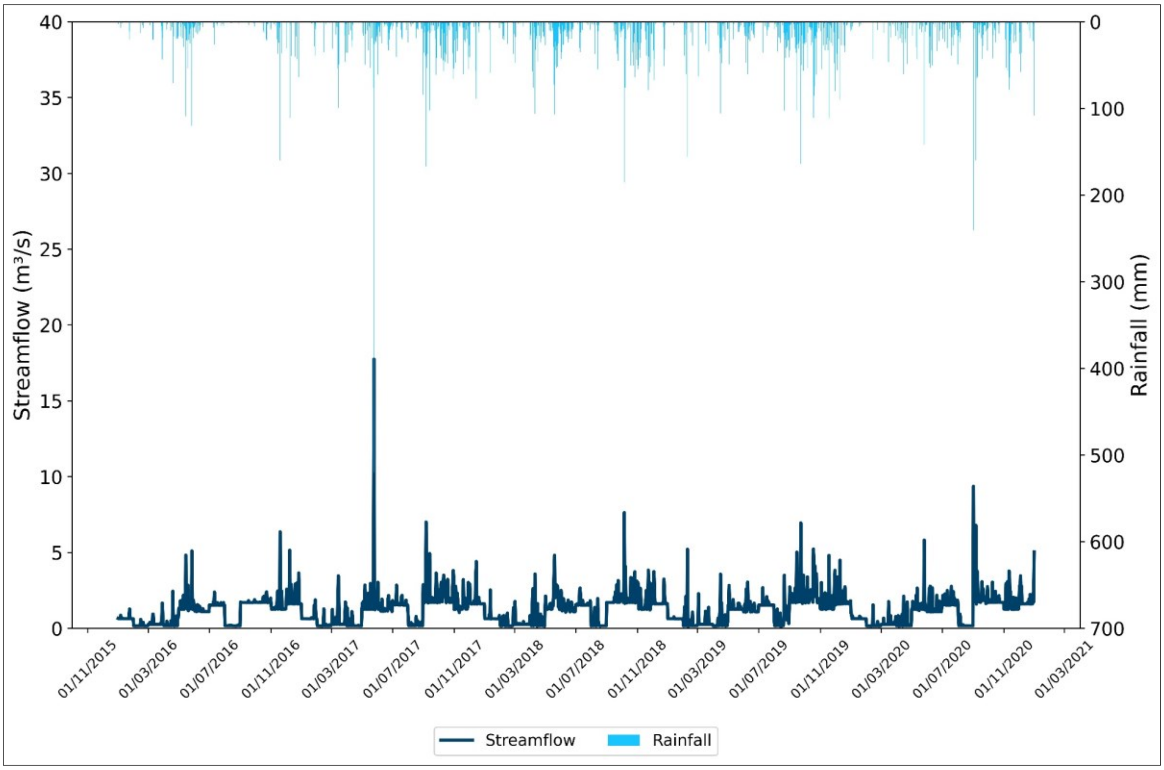
ERROR CORRECTED SUB-CATCHMENT HYDROGRAPHS



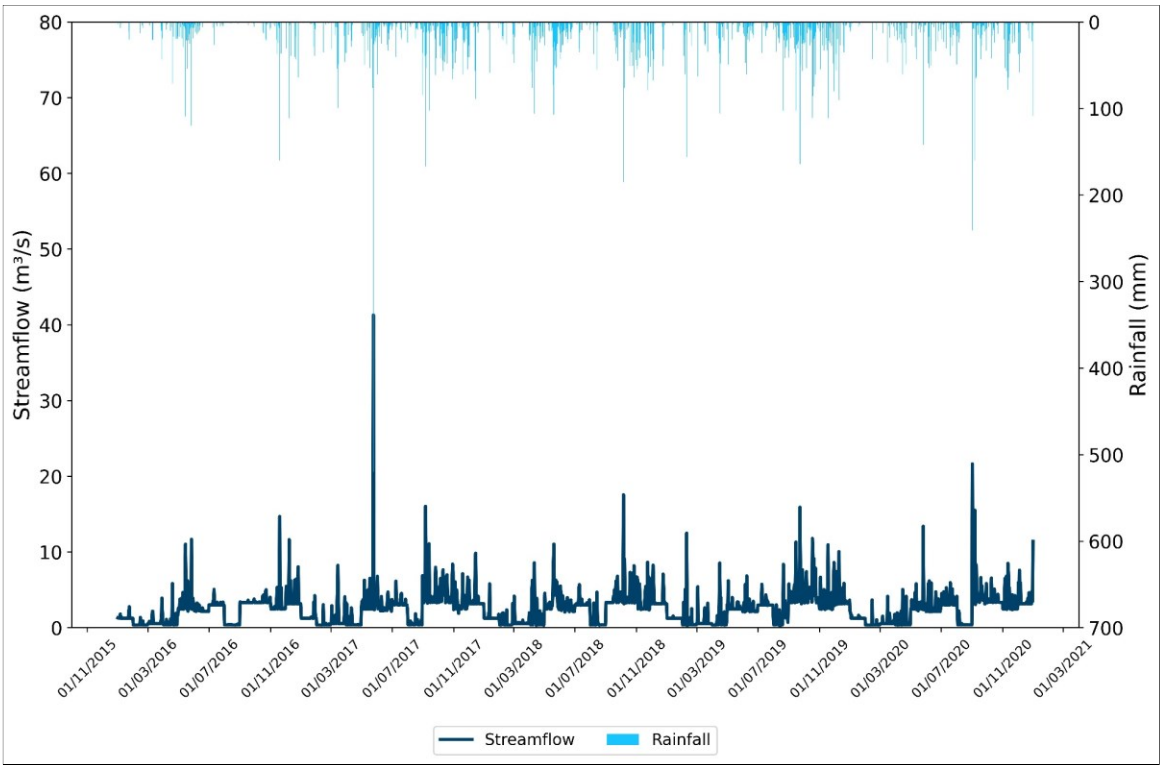
Appendix C–1: Error corrected sub-catchment hydrograph at Kollonnawa Ela



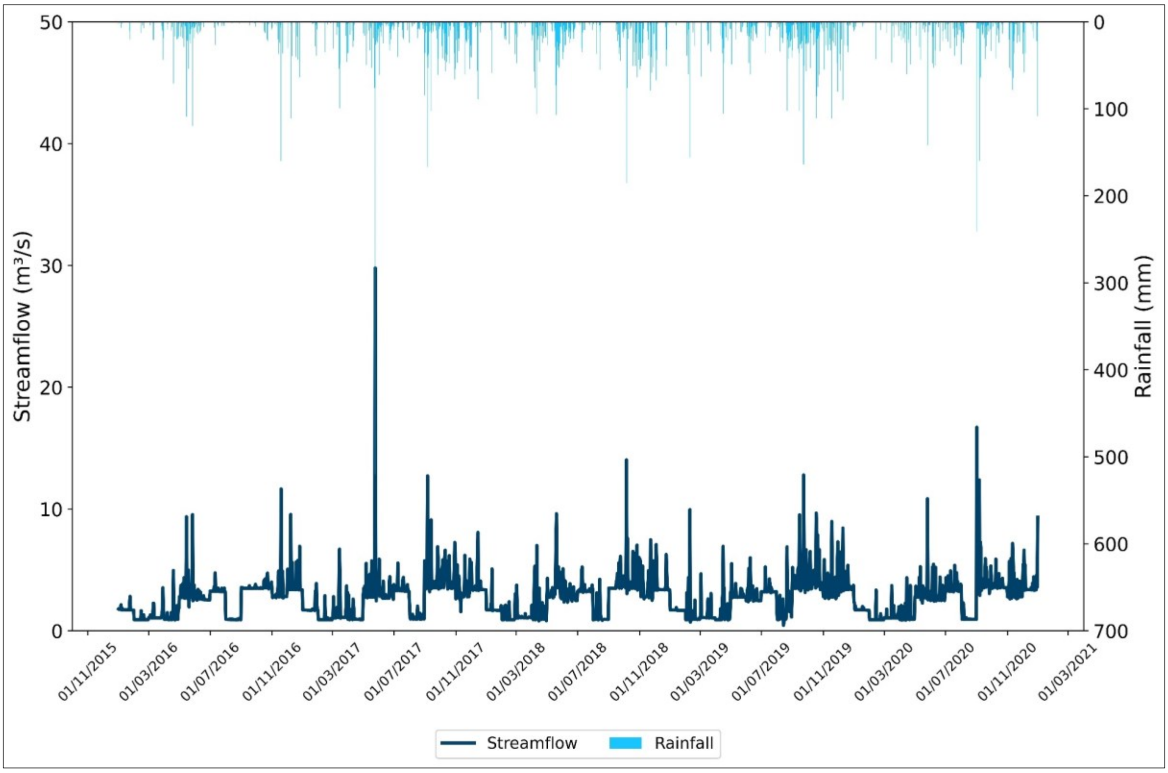
Appendix C–2: Error corrected sub-catchment hydrograph at Raggahawatte Ela



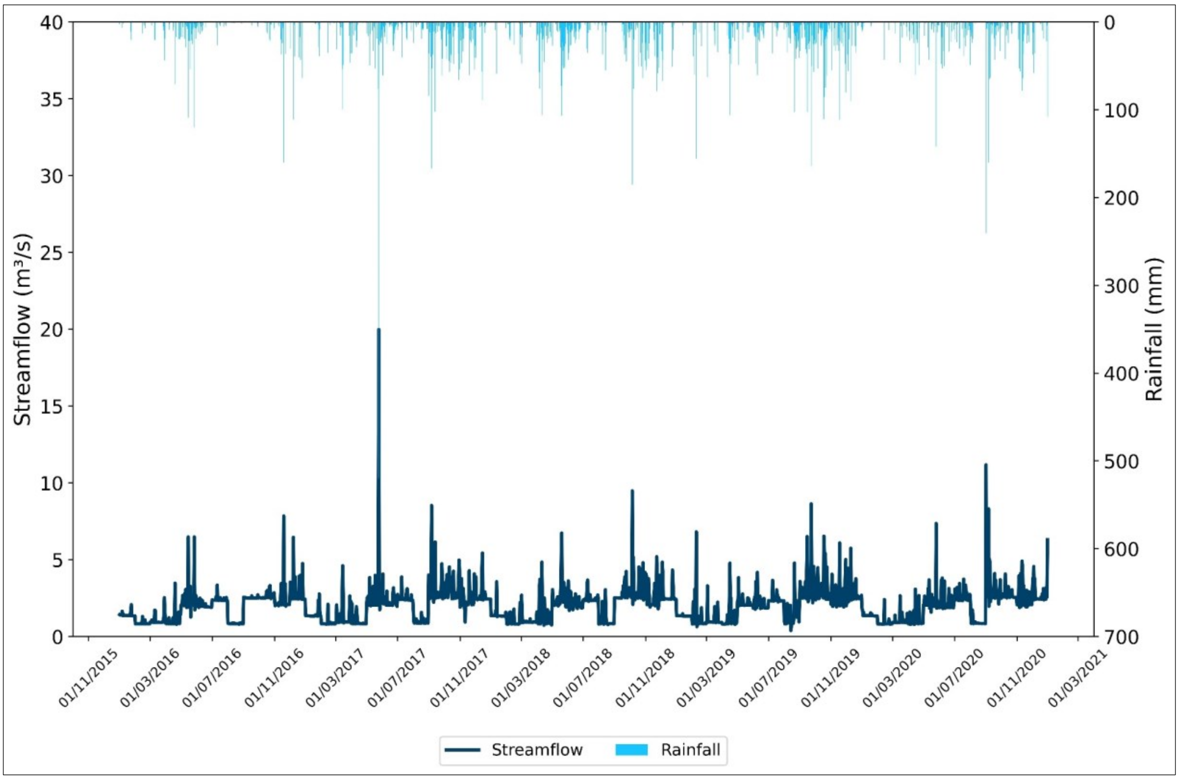
Appendix C–3: Error corrected sub-catchment hydrograph at Maha Ela



Appendix C–4: Error corrected sub-catchment hydrograph at Pusseli Oya



Appendix C–5: Error corrected sub-catchment hydrograph at Wak Oya



Appendix C–6: Error corrected sub-catchment hydrograph at Pugoda Ela