T.R.
ONDOKUZ MAYIS UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
ENVIRONMENTAL ENGINEERING DEPARTMENT

PREDICTION OF WATER QUALITY INDEX USING ARTIFICIAL NEURON NETWORK (ANN)

Master's Thesis

Hasanain Ali Banwan ZAMILI

Supervisor

Prof. Dr. Gülfem BAKAN

SAMSUN
2023

DECLARATION OF COMPLIANCE WITH SCIENTIFIC ETHIC

I hereby declare and undertake that I complied with scientific ethics and academic rules in all stages of my Master's Thesis , that I have referred to each quotation that I use directly or indirectly in the study and that the works I have used consist of those shown in the sources, that it was written in accordance with the institute writing guide and that the situations stated in the article 3, section 9 of the Regulation for TÜBİTAK Research and Publication Ethics Board were not violated.

Is Ethics Committee Necessary?

Yes ☐ (If it necessary, please add appendices.)

No ☒

16 /05 / 2023
Hasanain Ali Banwan ZAMILI

DECLARATION OF THE THESIS STUDY ORIGINALITY REPORT

Thesis Title : PREDICTION OF WATER QUALITY INDEX USING ARTIFICIAL NEURON NETWORK(ANN)

As a result of the originality report taken by me from the plagiarism detection program on 02 /05/2023 for the thesis titled above;

Similarity ratio : % 29

Single resource rate : % 5 has been released.

16 / 05 / 2023
Prof. Dr. Gülfem BAKAN

ÖZET

YAPAY NÖRON AĞI (ANN) KULLANARAK SU KALİTESİ ENDEKSİ TAHMİNİ

Hasanain Ali Banwan ZAMILI
Ondokuz Mayıs Üniversitesi
Lisansüstü Eğitim Enstitüsü
Çevre Mühendisliği Ana Bilim Dalı
Yüksek Lisans, Mayıs/2023
Danışman: Prof. Dr. Gülfem BAKAN

Tahmin, su kalitesi yönetimi için hayati önem taşır ve bu kirli çağda su gibi kıt bir kaynakla çalışırken esastır. Bu nedenle, önemli su kalitesi parametrelerine ve iklim faktörlerine dayalı olarak su kalitesi indeksini (WQI) tahmin etmek için tutarlı ve hassas makine öğrenimi modelleri oluşturmak, çevresel felaketleri önlemeye yönelik erken uyarılar sağlamak için çok önemlidir. Bunun ışığında, Türkiye'nin Çorum il sınırları içindeki Yeşilırmak nehri, su kalitesi indeksini on bir fizikokimyasal parametreyi ((biyokimyasal oksijen ihtiyacı (BOI), pH, elektriksel iletkenlik (EC), çözünmüş oksijen (ÇO), klorür (Cl^{-1}), kalsiyum (Ca^{+2}), magnezyum (Mg^{+2}), nitrit (NO_2^{-1})), sodyum (Na^{+1}), sülfat (SO_4^{-2}) ve toplam çözünmüş katılar (TDS)) su kalitesi indeksini (WQI) hesaplamak için seçilmiştir Sonuç, içme için uygun olmayan bir nokta (Haziran 2000) dışında iyiden kötüye doğru değişiyordu. Ayrıca, bu belirlemeyi gerçekleştirmek için üç metasezgisel optimizasyon algoritması; kısıtlama katsayısına dayalı parçacık sürüsü optimizasyonu ve kaotik yerçekimi arama algoritması (CPSOCGSA), deniz avcısının optimizasyon algoritması (MPA) ve parçacık sürüsü optimizasyonu (PSO), Levenberg-Marquardt üç katmanlı bir yapay sinir ağı (YSA) ile birleştirildi. Ayrıca, veri ön işleme teknikleri, doğal logaritma tekniği, tekil spektrum analizi (SSA) ve toleransın üç özelliği, orijinal veri setini normalleştirmek ve iyileştirmek ve en uygun model girdi senaryosunu belirlemek için sırasıyla kullanılmıştır. Ayrıca, bu modellerin doğruluğu çeşitli istatistiksel kriterlere (MAE, RMSE ve R^2) dayalı olarak doğrulanmıştır. Sonuç, CPSOCGSA-ANN modelinin performansının en yüksek doğruluğu (R^2 = 0.965, MAE = 0.01627 ve RMSE = 0.0187) elde ederek diğer modellerden (MPA-ANN ve PSO-ANN) üstün olduğunu ortaya koymuştur.

Anahtar Sözcükler: Tahmin, Su kalitesi indeksi WQI, ANN, CPSOCGSA, SMA, MPA

ABSTRACT

PREDICTION OF WATER QUALITY INDEX USING ARTIFICIAL NEURON NETWORK(ANN)

Hasanain Ali Banwan ZAMILI
Ondokuz Mayıs University
Institute of Graduate Studies
Department of Environmental Engineering
Master, May/2023
Supervisor: Prof. Dr. Gülfem BAKAN

Forecasting is vital to water quality management and is essential when working with a scarce resource like water in this polluted era. Therefore, generating consistent and precise machine learning models to predict the water quality index (WQI) based on significant water quality parameters and climate factors is essential for providing early warnings to prevent environmental disasters In view of this, Yesilirmak river, within the provincial borders of Çorum, turkey, was selected as a case study to predict the water quality index eleven physicochemical parameters (biochemical oxygen demand (BOD), PH, electrical conductivity (EC), dissolved oxygen (DO), chlorine (Cl^{-1}), calcium(Ca^{+2}), magnesium (Mg^{+2}), nitrite (NO_2^{-1}), sodium (Na^{+1}), sulfate (SO_4^{-2}) and total dissolved solids (TDS)) were assigned to calculate the water quality index (WQI). The result varied from good to poor, except for one point (June in 2000) that was unsuitable for drinking. Also, to accomplish this determination, three metaheuristic optimization algorithms (the constraint coefficient-based particle swarm optimization and chaotic gravitational search algorithm CPSOCGSA, the marine predator's optimization algorithm MPA, and particle swarm optimization PSO) were combined with an artificial neural network (ANN) with a Levenberg–Marquardt three-layer backpropagation algorithm architecture to build three hybrid models (CPSOCGSA-ANN, MPA-ANN, and PSO-ANN).In addition, three features of data pre-processing techniques (natural logarithm technique, singular spectrum analysis (SSA), and tolerance) have been used to normalize, enhance the original dataset and identify the optimal model input scenario (respectively). Furthermore, these models' accuracy was verified based on various statistical criteria (MAE, RMSE, and R^2). The result revealed that the performance of the CPSOCGSA-ANN model was superior to other models (MPA-ANN and PSO-ANN) by achieving the highest accuracy (R^2 = 0.965, MAE = 0.01627, and RMSE = 0.0187this window of the thesis that opens to the outside.

Keywords: Prediction, Water Quality index WQI, ANN, CPSOCGSA, SMA, MPA

ACKNOWLEDGMENT

Praise be to "ALLAH", who gave me the strength, health and safe to work and think to finish this work.

Master is a rewarding but challenging journey, which would not be possible without the help of many people.

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Gülfem BAKAN, for his continuous support, advice, and guidance. He has built and directed an environment that grants me an opportunity to learn and practise research skills. Special thanks go to all the staff members, the head of the environmental engineering department,

I would like to thank MDSI General Directorate of State Hydraulic Works of Turkey for helping me to get the data was used in my research.

Finally, I would like to thank my family. For my dear father who encouraged me to love science and supported me in chasing my dreams. For my beloved mother for all her love and encouragement.

Hasanain Ali Banwan ZAMILI

CONTENTS

SYMBOLS AND ABBREVIATIONS

WQI,W.Q.I	: Water Quality Index
WQ	: Water Quality
ANN	: Artificial Neuron Network
ARIMA	: Auto-Regressive Integrated Moving Averages
MLR,M.L.R	: Multiple Linear Regression
OECD	:Organisation for Economic Coopreation &Development
UN	: United Nations
BOD	: Biochemical oxygen demand
CPSOCGSA	: Hybridised Constriction Coefficient-Based Particle Swarm Optimisation And Chaotic Gravitational Search Algorithm
CGSA	: Chaotic Gravitational Search Algorithm
SMA	: Slime Mould Algorithm
MPA	: Marine Predators Algorithm
PSO	: Particle Swarm Optimisation
DO	: Dissolved Oxygen
DOC	: Dissolved Organic Carbon
COD	: Chemical Oxygen Demand
NOM	: Natural Organic Matter
NSF-WQI	: National Sanitation Foundation-Water Quality Index
BCWQI	: the British Columbia Ministry for Water Quality Index
IOU Index	: Intersection Over Union Index
CCME,C.C.M.E	: Canadian Council of Ministers of The Environment
F.I.S	: Fuzzy Inference System
M.W.Q.I.	: The Microbiological Water Quality Index
E.Q.I.	: The Ecofunctional Quality Index
B.C.W.Q.I.	: British Columbia Water Quality Index
W.J.W.Q.I.	: West Java Water Quality Index
MAPE	: Mean Absolute Percentage Error Criteria
RMSE	: Mean Absolute Percentage Error Criteria
NSE	: Nash Sutcliffe Model Efficiency Coefficient Criteria

LMI	: Logistics Managers' Index Criteria
CC	: Carpenter and Coustan Criteria
R^2	: Determination Coefficient Criteria
ANFIS,A.N.F.I.S.	: Adaptive Neuro Fuzzy İnference System
W.A.N.N.	: Weighted Nuclear Norm Minimization
W.N.F.	: Wavelet Neuro Fuzzy
G.P.	: Goal-Programming Model
S.V.R.	: Support Vector Regression
S.V.M.	: Support Vector Machines
W.S.V.R.	: Wavelet Support Vector Regression
GA-ANN	: Genetic Algorithms- Artificial Neuron Network
PBH	: The preprocessing-based hybrid model
OBH	: The parameter optimisation-based hybrid model
CBH	: The components combination-based Hybrid model
SBH	: The postprocessing-based hybrid model
HPH	: Hybridisation with preprocessing-based hybrid models
HOH	: Hybridisation with parameters optimisation-based hybrid
HCH	: Hybridisation with components combination-based hybrid
HSH	: Hybridisation with postprocessing-based hybrid models
RFC	: Rolling Contact Fatigue
CVPV	: Change Of Variable Perturbation Stochastic
XGB-ELM	: Extreme Gradient Boosting- Extreme Learning Machine
GPELM	: Genetic Programming Extreme Learning Machine
Mas	: Machines Algorithm
SSA	: Singular Spectrum Analysis

FIGURES LEGENDS

TABLES LEGENDS

1. INTRODUCTION

1.1. Overview

Rivers' water pollution is a significant concern for communities worldwide, requiring more attention from environmental researchers (Hameed Mohammed et al., 2016). In addition, the quality of water in watersheds is affected by various factors, including natural components such as temperature or precipitation and anthropogenic factors such as manufacturing practices, urbanization, and agriculture (Asadollah Seyed Babak Haji Seyed et al., 2021). Therefore, the implications of contaminated surface water are deterioration of water quality, direct threats to human health, disruption of the aquatic ecosystem's balance, and economic and social failure. Most ambient water bodies have specific quality standards that indicate their quality. Furthermore, water requirements for various applications and usages possess their own standards. For example, irrigation water should be neither too saline nor contain toxic materials that might be transmitted to plants or soil and thus damage ecosystems; industrial usage needs different properties depending on the specific industry applications. Hence, rapid industrial development has driven the depreciation of water quality at an alarming rate; furthermore, infrastructures, with a lack of public awareness and less hygienic features, significantly impact potable water quality (Aldhyani T. H. H. et al., 2020).

The expression "water quality" refers to the state or condition of water, considering the physical, chemical, and biological properties. Furthermore, assessment and prediction of the water quality are required to effectively manage river systems so that sufficient actions can be implemented to ensure that the pollution levels stay within permissible limits and are mandatory for developing various remediation strategies (Najah A et al., 2013).

A water quality index (WQI) is a composite indicator and mathematical tool that gives information to end-users based on specified water content factors, translated into a single unitless value (Kadam AK et al., 2019). (WQI) has been widely used "as a classification indicator" to evaluate and categorise the quality of water bodies based on a wide range of physiochemical and organic variables, and it plays an important role in water management. However, water quality management requires collecting and analysing large water quality datasets that can be difficult to evaluate and synthesise. Hence, various techniques have been created to evaluate water quality; the Water Quality Index (WQI) model is one such tool(Uddin Md Galal et al., 2021).

Forecasting is vital to water quality management and is essential when working with a scarce resource like water. Over the past few decades, many traditional water quality prediction methods have been used, such as auto-regressive integrated moving averages (ARIMA) (Araghinejad Shahab, 2013), multiple linear regression (MLR) (Rajaee Taher & Boroumand Amir, 2015), etc. However, with the expansion in the volume of data, traditional methods cannot effectively suit the requirements of researchers Because of the increase in computing power and the inability to capture non-stationarity and nonlinear water quality due to its sophisticated and complicated nature. Hence, artificial intelligence approaches such as artificial neural networks are becoming more popular in recent years because they can surmount the limitations or drawbacks of deterministic models. In addition, due to their similarities with the brain's nervous system, ANNs are suitable for analysing nonlinear and unpredictable problems and have become a hotspot in environmental quality research (Chen Yingyi et al., 2020). Various hybrid models have been created and used to enhance prediction accuracy. In general, the well-known hybrid models could be classified into four categories (1) preprocessing-based, (2) component combination-based, (3) parameter optimisation-based and (4) postprocessing-based hybrid models (Hajirahimi Zahra & Khashei Mehdi, 2022a).

1.2. Research Problem

In the last decade, rapid industrial expansion without managing outflows, heavy fertiliser usage in agriculture, and overexploitation of freshwater have degraded river ecosystems and had various adversely affected human health (Sakaa Bachir et al., 2020). Furthermore, the climate change of our world has begun greatly warming. One of the systems that will be much affected by this situation is the water quality system. There is no doubt that climate will exacerbate the hazards and vulnerabilities; because Turkey is located in one of the world's vulnerable areas faced by climate change (Ustaoğlu Fikret et al., 2020).

On the other hand, the environmental expectations of the Organisation for Economic Cooperation and Development (OECD) to 2050 indicate that global water demand is anticipated to increase by 55%, relative to 2000 as a baseline. Additionally, large proportion of the global population may be below the acute water stress threshold. The world population has insufficient access to potable water, as shown in Figure 1. It can be seen that the worldwide population percentage that lacks access to drinking water will increase as we go into the future, and it is likely to be 50% by 2050 (Fogden Josephine & Wood Geoffrey, 2009). The consequences of contaminated potable water are dangerous and may badly impact health, the environment, and infrastructure. As the United Nations (UN) announced, approximately 1.5 million people die annually due to diseases driven by polluted water. Also, it is reported that 80% of health troubles in developing countries are brought on by polluted water (Aldhyani T. H. H. et al., 2020).

In Turkey, the useable fresh water is around 112 billion m^3 each year; it uses about 40% of this quantity (i.e. 44 billion m^3). Most of this percentage has been used for irrigation (33 billion m^3). However, while this amount of freshwater appeared to be relatively large, it did not make Turkey a water-rich country. In reality, it is considered a "water-stressed" country due to the per capita freshwater amount being around 1500 m^3 which is an insufficient quantity compared with numerous European countries. It is projected that this quantity will decrease whether climate change is considered or not. Recently, river-based water management has become quite important because of the development of freshwater requirements (Ustaoğlu Fikret et al., 2020).

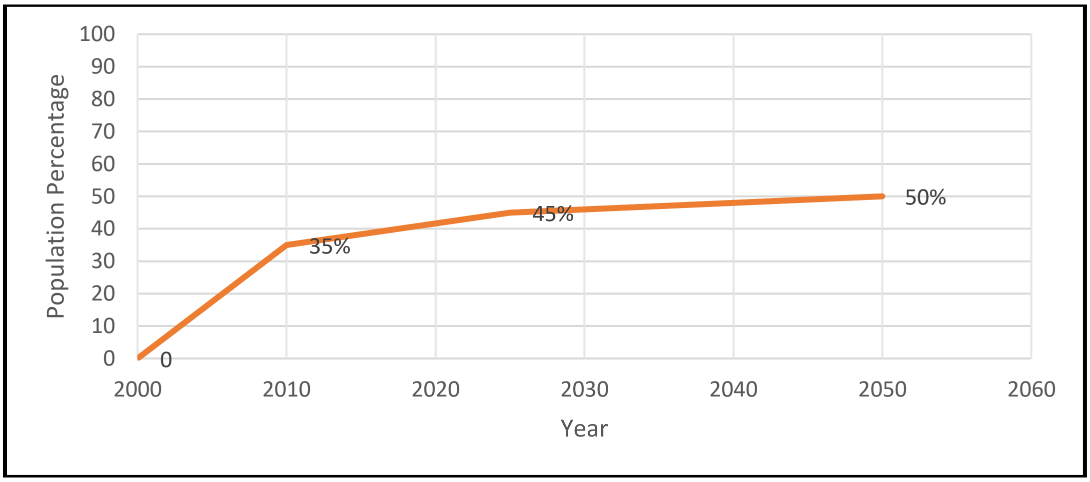


Figure 1.1. The global population has lacked access to safe drinking water (Fogden Josephine & Wood Geoffrey, 2009)

Water quality management requires collecting and analysing ample water parameters datasets that could be hard to assess and synthesise. Hence, various techniques have been created and developed to evaluate the quality of water, such as the Water Quality Index (WQI) technique. Furthermore, various parameters, such as the BOD test, necessitate highly controlled laboratory conditions and a lengthy period of time, making it costly (Das Kangabam Rajiv et al., 2017; Judran Naser Hussein & Kumar Ajeet, 2020).

Although previous studies have identified weather aspects, research has yet to investigate and determine their impact in terms of using sufficient data preprocessing to remove the effect of socio-economic variables that are insensitive to climate (Yuan Xiao-Chen et al., 2016). Additionally, using a capable and practical prediction approach based on a systematic method instead of trial and error. However, researchers have not determined the number of climate factors that are influenced by water quality. Overall, previous studies of water quality prediction have suffered from inadequate sample sizes, methodological limitation and a lack of a robust theoretical framework. So, the debate continues on the best strategies for managing water quality under the effect of climate change with a low level of certainty.

1.3. Aim and Objectives of this research

This research aims to predict the water quality index (WQI) for the long-term by considering the climatic variable (rainfall) and adopting a combination of approaches that have various kinds of hybrid artificial neural networks (ANN) models to provide the environmental organisation with the correct strategic decision in planning and controlling for the water quality index. Additionally, optimal circumstances for running parameters of an algorithm have been used to identify the number of neurons within the hidden layers and the combination of the dataset partition to provide the highest forecasting accuracy for the proposed Artificial Neural Network Water Quality Index (ANNWQI).

The specific objectives for realising the aim of the thesis, as stated above are:

1. Assess the water quality of the Yesilirmak river within the provincial borders of Çorum during the monitoring periods 1995-2014.
2. Show the SSA and tolerance approaches' benefits for enhancing data quality and picking the optimised scenario for input parameters.
3. A combination metaheuristic algorithm (CPSOCGSA) was hybridised with the ANN model to determine the ANN hyperparameters.
4. Extensive comparison of the performance accuracy of the CPSOCGSA-ANN with the recent models (SMA-ANN, PSO-ANN, MPA-ANN) by employing the same training and validation datasets to identify the optimal model.

5. Identifying and selecting the parameters related to the watershed's quality depends on impacts and their relevance to the water quality, Hence low prediction cost with high efficiency.
6. Predicting water quality will provide early timely warnings, especially when considering the implications of climate factors.

2. LITERATURE SURVEY

2.1. Impacts of Climate Change on Surface

The fundamental contributor influencing the quality of the aquatic system is anthropogenic activities (point-source contamination), which have been dominant to some extent. Non-point-source contamination has become a significant concern that must be solved expeditiously (Zhou Yun et al., 2014). In particular, climate change is the underlying parameter influencing nonpoint source contamination. With the change in climate, the frequency and intensity of extreme climatologies, like drought expansion and heavy precipitation, will exacerbate the potential for pollution incidents (Michalak Anna M., 2016; Shanley K., 2017). Typically, changes in water's temperature will impact the water's surface tension, density, viscosity, and other properties. Dissolved oxygen content will also fluctuate, upsetting the aquatic ecosystem's equilibrium. Additionally, the change in temperature will have an influence on photosynthesis, the average rate of chemical reactions, the toxicity of different contaminants, and the ability of aquatic creatures' microbes to degrade toxins (Bi Wuxia et al., 2018; Hammond D & Pryce AR, 2007). Scientists observed that surface water temperature was raised since the 1960s in North America, Europe and Asia (0.2–2 °C) due to atmospheric warming with solar radiation exceeding (Wang Kaicun & Dickinson Robert E, 2013). Observed an average rise in water temperature of around 2 °C respectively, in the Meuse and Rhine rivers during the severe drought of 2003, with a pH increase (reflecting a decrease in CO_2 concentration) and a minimised in dissolved oxygen (DO) solubility reflecting a lower DO solubility under higher water temperatures. A DO decrease can also be associated with increased DO assimilation of biodegradable organic matter by microorganisms (linked to an increase in dissolved organic carbon (DOC) (Prathumratana Lunchakorn et al., 2008). Additionally, in the same research dealing with surface-water quality in the Lower Mekong River, significant negative correlations were broadly found between precipitation (or discharge flow) and DO, pH, and conductivity (from 0.2 to 0.9). In many lakes in Northern America and Europe, the stratified period has lengthened by two or three weeks, and water temperatures have risen from 0.2 to 1.5 °C, which effectively promotes thermal stratification. Artificial intelligence predicts a 2°C increase by 2070 (in Europe lakes), though this rise will also be influenced by lake characteristics and seasons (Malmaeus J Mikael et al., 2006).In another context,(Prathumratana Lunchakorn et al., 2008) mentions that COD (chemical oxygen demand), used as an indicator of natural organic matter (NOM), has fair to weak correlations with precipitations (0.295– 0.426) and discharge flows (0.312–0.324) (Prathumratana Lunchakorn et al., 2008).

Overall, More study has been conducted on the influence of climate change on the number of aquatic resources. However, the water quality (WQ) response to climate change requires additional investigation**.**Table 2.1 displays the impacts of climate change on water quality parameters.

2.2. Water Quality Index

Water is a fundamental natural resource for living organisms on Earth, but the quality of the aquatic system has been degrading for decades due to both natural and human activity. Natural elements influencing freshwater are atmospheric, hydrological, topographical, climate, and lithological parameters (Magesh N. S. et al., 2013). While the anthropogenic activities that negatively influence are mining, livestock farming, production and disposal of waste (agricultural, industrial and municipal), and the rise of sediment run-off or soil erosion due to land-utilise change (Lobato TC et al., 2015). There are various categories for water quality parameters, including:

- Physiochemical parameters (e.g., pH, turbidity, SO_4^{-2}, temperature, and nutrients).
- Biological parameters with pathogens, microorganisms, cyanobacteria, and water quality proxies.
- Micropollutants parameters (inorganic and organic), such as metals, pharmaceuticals, and pesticides.

(Kummu Matti et al., 2010) showed in their study a nearly sixteen-fold increase in population due to water scarcity since the 1900s. In 2005, over 35% of the global population resided in countries with moderate to extreme water stress. This percentage will rise to around 52% by 2050 (Schlosser C Adam et al., 2014). However, the published academic literature between 2010-2019 was collected from the Google Scholar database. Over the past decade, a continuous increase in WQI publications has been reported (Figure 2.1). According to the increasing number of publications, it is clear that study of WQI approaches has gained popularity among freshwater resources, researchers, engineers, and scientists. (Gupta Suyog & Gupta Sunil Kumar, 2021).

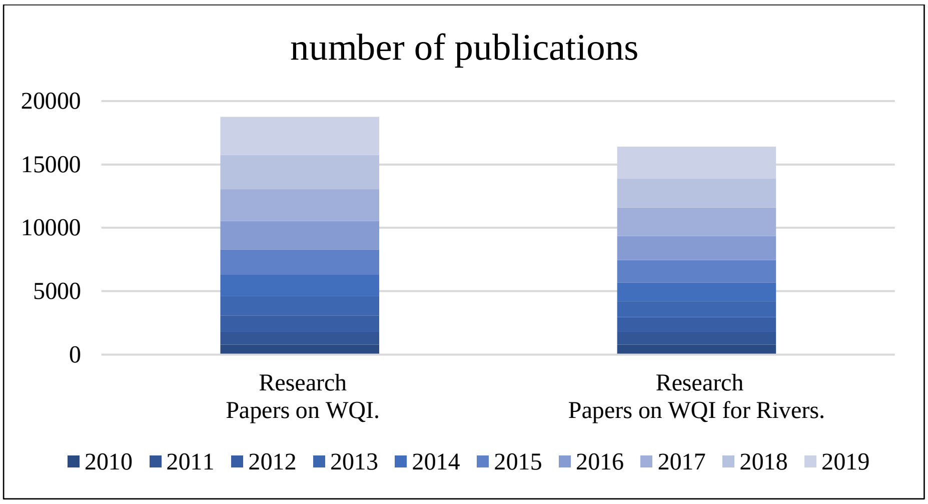


Figure 2.1. The research trend on WQI from 2010–2019 (Gupta Suyog & Gupta Sunil Kumar, 2021)

Horton proposed the contemporary approach of WQI in 1965 to declare the river's WQ status based on ten water quality characteristics considered significant in water bodies(Horton Robert K, 1965). Moreover, Brown collaborated with the National Sanitation Foundation to develop a more stringent version of Horton's WQI model. Although, 142 water quality experts participated in a panel that provided advice on parameter selection and weighting for the NSF-WQ (Uddin Md Galal et al., 2021). As a result, numerous alternative WQI models have since relied on the NSF-WQI (Abbasi Tasneem & Abbasi Shahid A, 2012). Also, The British Columbia WQI (BCWQI), created by the British Columbia Ministry, was another significant development (Saffran K et al., 2001). Additionally, models like the Malaysian, Almeida Index, IOU Index, and CCME WQI have been created.

Table 2.1. Impacts of climate change on water quality parameters (Gupta Suyog & Gupta Sunil Kumar, 2021)

CC factors affecting WQ	**W.Q parameter**	**Waterbody**	**Comments**
		Physiochemical Basic parameter	
Droughts	pH	River	(Meuse River) Greater than acceptable values.
Temperature increase	-	Lake	pH increasing.
Rainfalls	DO	River	(Mekong) adversely correlation.

CC factors affecting WQ	W.Q parameter	Waterbody	Comments
Droughts and temperature increase	-	River and lake	Reduce the solubility and concentration of dissolved oxygen.
Rainfalls	-	River	(Mekong) adversely correlation.
Droughts and temperature increase	Temperature	Rivers	The water column is warming.
-	-	Lakes	Increased epilimnion and hypolimnion, particularly in shallow lakes.
Droughts and temperature increase	Nutrients	Streams and lakes(peatland)	(UK and Mekong basin) DOC flux rising during the storm is happenings.
Droughts	-	River	(UK and Mekong basin) DOC flux rising during the storm is happenings.
Temperature increase	-	Surface water and groundwater	Rises the mineralisation and releases of N, C and P from soil OM.
Temperature and rainfall increase	-	River basins, lakes, fjords and groundwater	-Warm period hypolimnion phosphate and ammonium contents were elevated.-Increase N and P loading (Sweden) and nutrition loading (Canada). -Changes in nitrogen loading in shallow lakes are significant. Increase N River (Norway) and nitrate loading (UK river and Seine groundwater). -The impacts of climate change may be comparable for small and large lakes. -Nutrients from sediments are released and enter the euphotic zone. -
Rainfall	-	Stream and lake	Australia, Mekong basin. It increased P loading of Dutch lakes.
Micropollutants, inorganics			
Droughts	Metal	Forest watershed river	-Cadmium, Selenium, mercury, zinc, Barium, lead and nickel increased. -Nitrogen (TKN) loading increases (USA).

CC factors affecting WQ	W.Q parameter	Waterbody	Comments
Temperature increasing	-	High alpine lakes	(European Alps) Heavy snowmelt led to a rise in micropollutants.
Heavy rainfall increasing	-	Streams	correlations between the amounts of DOC and pharmaceuticals in freshwater.
Temperature and rainfall increasing	-	Streams	-The release of trace elements is accompanied by a rise in organic carbon content and a decrease in redox potential (Brittany). -Various trace elements were mobilised by humic and fulvic acids (Brittany).
Temperature and rainfall increase drying and rewetting cycles	Organic, pesticides	Surface and groundwater	-Changes in seasonality's rainfall intensity and rising temperatures are the primary climate variables influencing the fate and behaviour of pesticides. -Drying the soil boosted the binding of herbicides.
Rainfall increase	-	Streams	Pesticide flux rises as a result of increased precipitation.
Temperature's increase	Pharmaceuticals	Groundwater	(Deutschland) Pond for artificial recharge. Redox conditions (anaerobic conditions) change as temperature rises.
Rainfall increase	-	Streams	-Pharmacological concentrations in water and DOC concentrations. -Due to their high mobility, loperamide and clofibric acid have the potential to contaminate groundwater through rivers.
		Biological	
Temperature and rainfalls Increase	Pathogens	Surface water	-The seasonal and interannual variability in local rainfall accounts for over 70% of the variability in the coliform record. -Extreme rainfall was followed by outbreaks of half the water-borne diseases that occurred in the United States over the past 50 years.
rainfalls and Temperature increasing	Cyanobacteria	Lakes	(Nederland's and Sweden's) Heat waves during the summer encourage the growth of cyanobacterial blooms.

CC factors affecting WQ	W.Q parameter	Waterbody	Comments
temperature increase	Cyanotoxins	lakes	Microcystin flourishes in the summertime heat.

Furthermore, 35 WQI approaches have been introduced by different countries and/or agencies to assess surface water quality worldwide (Abbasi Tasneem & Abbasi Shahid A, 2012; Dadolahi-Sohrab A et al., 2012; Poonam Tirkey et al., 2013). Error! Reference source not found. shows that, although WQI models have been implemented in all major types of watersheds, only 82% of applications have evaluated river water quality. Moreover, the table shows that the NSF models and CCME have been used in 50% of the reviewed studies.

Parameters of the WQI model were typically chosen based on data availability, expert opinion, or the environmental significance of a water quality parameter (Debels Patrick et al., 2005); reported that many WQI models employed just the basic water parameters' due to lacking availability of other parameters measured data (Cude Curtis G, 2001). In addition, numerous researchers edited the model parameter lists depending on data accessibility and obtainability, and occasionally it is not possible to add the defining water parameter into the model for this reason (Ma Zhen et al., 2020). Because of the high analytical cost and scarcity of modern analytical laboratory facilities, several WQI models did not include microbiological, suspended solids, toxic compounds, and contamination (Kannel Prakash Raj et al., 2007).

Table 2.2. Summary list for WQI model uses reported in published literature between 1960 and 2019 (Uddin Md Galal et al., 2021)

WQI. models	NO. Applications	Type of Study Area		
		Rivers	Lakes	Marine,coastal or sea
C.C.M.E.	36	28	5	3
F.I.S.	12	10	1	1
N.S.F.	18	17	1	…
Horton	7	6	…	1
S.R.D.D.	6	6	…	…
M.W.Q.I.	8	6	1	1
Bascaron	4	3	…	1
E.Q.I.	2	1	1	…
Smith	2	2	…	…

WQI. models	NO. Applications	Type of Study Area		
		Rivers	Lakes	Marine,coastal or sea
Oregon	2	2	…	…
Almedia	1	1	…	…
B.C.W.Q.I.	1	1	…	…
Dalmatian	1	…	…	1
Dojildo	1	1	…	…
Liou index	1	1	…	…
Dinius	1	1	…	…
Hanh index	1	1	…	…
W.J.W.Q.I.	1	…	…	1
House index	1	1	…	…
Said	1	…	…	1

2.3. Data Pre-processing

Despite the high reliability of various models in forecasting water quality and hydroclimate variables, these models may be limited in their ability to deal with time-series that are frequently non-stationary and span a wide range of scales from a few minutes to several months. Pre-processing data will be required to overcome similar flaws and challenges (Nourani Vahid et al., 2019; Zubaidi Salah L., Dooley Jayne, et al., 2018). The dataset must be appropriately processed before being used in an ANN. These techniques are implemented to ensure that the learning model handles all data equally Maier Holger R. and Dandy Graeme C. (2000). According to (Unnikrishnan Poornima & Jothiprakash V., 2018), data preparation is essential for most water quality time series to achieve high prediction performance. It has been used successfully in a variety of academic fields, such as monthly precipitation prediction (Ouyang Qi & Lu Wenxi, 2017), water level prediction (Mohammed Sarah J et al., 2022), water parameter forecasting (Khudhair Zahraa S et al., 2022), and forecasting the drought (Alawsi Mustafa A et al., 2022). Pre-processing is separated into three sections:

2.4. Normalisation

Various approaches and evaluation metrics used in practical research frequently have different sizes and units, leading to various data analysis outcomes. So data normalisation is considered an essential part of data mining in learning machines (Wu Junhao & Wang Zhaocai, 2022). Data curation is a crucial preprocessing step in machine learning techniques, starting with data normalisation to reduce the input value range (Sofos F. & Karakasidis T. E., 2021). (Zubaidi Salah L., Gharghan Sadik K., et al., 2018) confirm that continuous variables must be normalised to produce time series that have a normal or near-normal distribution. Consequently, dataset normalisation aims to obtain the same range of values for all of the model's parameters and generate a time-series with a normal distribution (Zubaidi Salah L. et al., 2020).

2.5. Cleaning

Some parameter has an extremely high value, which is described as an outlier and will lead to distortion of the statistics. Therefore, techniques for cleaning data include eliminating outliers and signal's pretreatment (Zubaidi Salah L., Gharghan Sadik K., et al., 2018). The box-whisker approach was used to discover extremist data, which was subsequently processed. This strategy considerably enhances the predictive accuracy of the proposed model (Zubaidi Salah L. et al., 2022). Additionally, There are numerous noise elements within every time-series, and the pretreatment signal techniques (i.e. Empirical Mode Decomposition (Le Justin A. et al., 2020), Wavelet (Zhu S. et al., 2020) and Singular Spectrum Analysis (Le H. D. et al., 2021), etc.) are most reliable techniques for denoising the primary time-series through decomposet of its many components. (Zubaidi Salah L. et al., 2020).

2.6. Selecting the Best Model Inputs

Selecting the best model is the most critical part of pre-processing data and creating a reliable prediction model's input parameters (Zubaidi Salah L., Gharghan Sadik K., et al., 2018).

Error! Reference source not found. provides a summary of data preparation b ased on prior research analysed for this study. Most of these studies employed one or\and two data pre-processing procedures. Moreover, several did not utilise the optimal model input. In addition, few studies utilised all data preprocessing processes.

Table 2.3. Data pre-processing Summaries

NO.	Literature	Normalisation data	Cleaning data	Optimal model's input
1.	(Alawsi Mustafa A et al., 2022)	yes	yes	yes

NO.	Literature	Normalisation data	Cleaning data	Optimal model's input
2.	Kisi Ozgur et al. (2019)	Yes	No	Yes
3.	(Mohammed Sarah J et al., 2022)	yes	yes	yes
4.	Ali Mumtaz et al. (2019)	Yes	Yes	Yes
5.	Pham Quoc Bao et al. (2021)	Yes	Yes	Yes
6.	(Koranga Manisha et al., 2022)	Yes	No	No
7.	Danandeh Mehr Ali et al. (2020)	Yes	No	Yes
8.	Safavi Hamid R. et al. (2018)	Yes	Yes	No
9.	(Wu Junhao & Wang Zhaocai, 2022)	Yes	No	No
10.	Fung Kit Fai et al. (2019)	Yes	Yes	No
11.	(Khudhair Zahraa S et al., 2022)	yes	yes	yes
12.	Zhang Yuhu et al. (2019)	No	Yes	Yes
13.	Aghelpour Pouya et al. (2021)	Yes	No	No
14.	Khan Md Munir Hayet et al. (2020)	No	Yes	No
15.	(Abdul Kareem Baydaa et al., 2022)	yes	yes	yes
16.	Nabipour Narjes et al. (2020)	Yes	No	No
17.	Wu Xianghua et al. (2021)	Yes	Yes	No
18.	Djerbouai Salim and Souag-Gamane Doudja (2016)	Yes	Yes	No
19.	Başakın Eyyup Ensar et al. (2020)	Yes	Yes	No
20.	Das Prabal et al. (2020)	Yes	Yes	No
21.	Adnan Rana Muhammad et al. (2021)	Yes	No	No

NO.	Literature	Normalisation data	Cleaning data	Optimal model’s input
22.	Belayneh A. et al. (2015)	Yes	Yes	No
23.	Soh Y. W. et al. (2018)	Yes	Yes	No

2.7. Artificial Neural Network

Neural network (NN) models are utilized as extremely potent machine learning techniques for time-series forecasting in various engineering applications. ANN simulates the human brain's functionality by attempting the same connections and operations as biological neurons. Also, the ANN could capture the non-stationary and nonlinear relations by using input variables and output training (Zubaidi Salah L., Dooley Jayne, et al., 2018). Moreover, the ANN is designed methodically to improve a performance criterion or monitor's criterion of implied internal constraints, which is commonly referred to as the process or learning imperative. Additionally, the learning process includes adjusting connection weights and updating network architecture so that a network can efficiently realise particular perception tasks. The ANN's creator selects the network topology, performance function, learning rule, training algorithm, and criterion for terminating the training phase, but the system changes the parameters (Zaqoot Hossam et al., 2017). According to (Zhang Liang et al., 2021), ANNs incorporate the two fundamental structures of biological neural nets, as shown in **Error! Reference source not found.**.

1. Neurones (Nodes).
2. Synapses (Weights).

Several previous studies have compared and validated traditional models with ANN to forecast water quality. It is discovered that the ANN technique has the potential to forecast water quality significantly more accurately than the conventional models:

(Kadam AK et al., 2019) compare MLR with the ANN approach, and he observes that the accuracy is higher in the ANN model. Thus, the ANN technique is advantageous because of its iterative’s procedure, which yields more accurate findings in both seasons. Consequently, the ANN method will become more efficient for water quality forecasting in the future.

(Patki Vinayak K et al., 2021) applied the developed ANN and Multiple Linear

Regression (M.L.R.) approach, which is a popular utilised statistical method for simulating water quality parameters. The research disclosed that the ANN model is superior to the (M.L.R.) approach for forecasting the distribution system water quality and is a reliable machine for comprehension of the weakly defined relationship between water quality parameters and WQI.

(Singh Balraj et al., 2021) evaluated soft computing methods by employing several criteria: MAPE, RMSE, NSE, LMI, CC, and MAE. Conclusions clarified that ANN outperforms the ANFIS model in WQI prediction.

According to (Wu Junhao & Wang Zhaocai, 2022), the ANN-WT-LSTM model utilised in this paper outperformed the conventional approach in different fitness criteria. Hence, the models could supply practical reference and technical support for monitoring and managing water quality.

According to the results of (Kulisz and Kujawska 2021), the ANN models successfully predict the value of the watershed quality. Therefore, they may be considered sufficient for simulation by units monitoring the status of the aquatic system.

Forecasting water quality is an essential subject for policy-makers in the water industry. However, achieving the expected prediction accuracy for the forecasting trends remains challenging. As a result, several optimization approaches could be used to address the implementation issues. The goal of optimisation algorithms is to select the optimum values for a system's variables under various conditions. Recently, many hybrid models have been utilised to forecast water quality accurately b(Zubaidi, Ortega-Martorell et al. 2020).

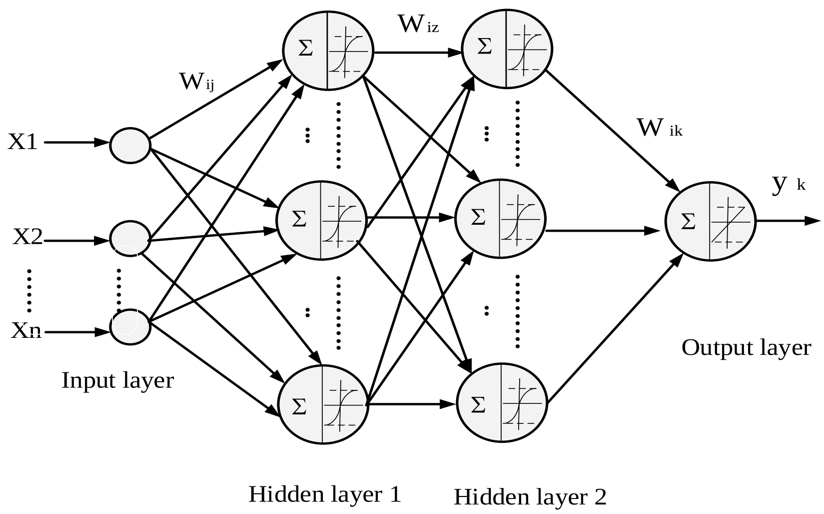


Figure 2.2. Structure of the ANN (Zubaidi Salah L et al., 2018)

2.8. Hybrid Model

The imperative for increased reliability, accuracy and capability around data-driven techniques has encouraged scientists to develop creative models. The imperative for increased reliability, accuracy and capability around data-driven techniques has encouraged scientists to develop creative models. Models are being developed to achieve new requirements; their head objective is to combine the advantages of many techniques in such a way as to enhance the potential of single models. These hybrid models generally integrate techniques in a process where traditionally, one technology is considered the basic technique, the others functioning as preprocessing or postprocessing methods (Modaresi Fereshteh & Araghinejad Shahab, 2014).

(Hajirahimi Zahra & Khashei Mehdi, 2022b) proposed that time series forecasting is divided into three main approaches:

- Hybrid models.
- Hybrid of hybrid models
- Individual models.

Hybrid models have attracted significant attention from numerous scientists. Until now, fourth varieties of a pure hybrid approach that adopt individual prediction methods are reported

- The preprocessing-based hybrid model (PBH)(Guermoui Mawloud et al., 2020),
- The parameter optimisation-based hybrid model (OBH)(Mehta Meenu et al., 2019),
- The components combination-based Hybrid model (CBH)(Vannitsem Stéphane et al., 2021; Zhang Liang et al., 2021).
- The postprocessing-based hybrid model (SBH)(Vannitsem Stéphane et al., 2021).

In the same context, researchers proved that hybridisation guarantees enhanced prediction accuracy instead of individual methods because of their apparent ability to deal with complicated patterns. So, according to various hybridisation kinds of pure-hybrid categories, four groups of hybridisation of hybrid models could be constituted:

Hybridisation with preprocessing-based hybrid models(HPH)(Wang Pengfei et al., 2021).

- Hybridisation with parameters optimisation-based hybrid models(HOH)(Liu Ruilin et al., 2020).
- Hybridisation with components combination-based hybrid models (HCH) (Colombo Nicoletta et al., 2019; Ruiz-Aguilar Juan Jesús et al., 2020).
- Hybridisation with postprocessing-based hybrid models HSH groups (Colombo Nicoletta et al., 2019).

Each of them is divided into four main branches, as shown in Figure 2.3.In humid weather, Iran (Bui Duie Tien et al., 2020) evaluated the effectiveness of four (12 hybrids model based on optimization algorithms, hybrid of the bagging with standalone, RFC, and CVPS) and standalone (M5P, RF, RT, and REPT) for monthly prediction W.Q.I.the Hybrid algorithms optimised the forecasting power of various standalone models.

The hybrid XGBELM and GPELM models suggested by (Abba SI et al., 2020) have improved prediction accuracy with fewer input parameters, hence considered reliable WQI's forecasting methods.

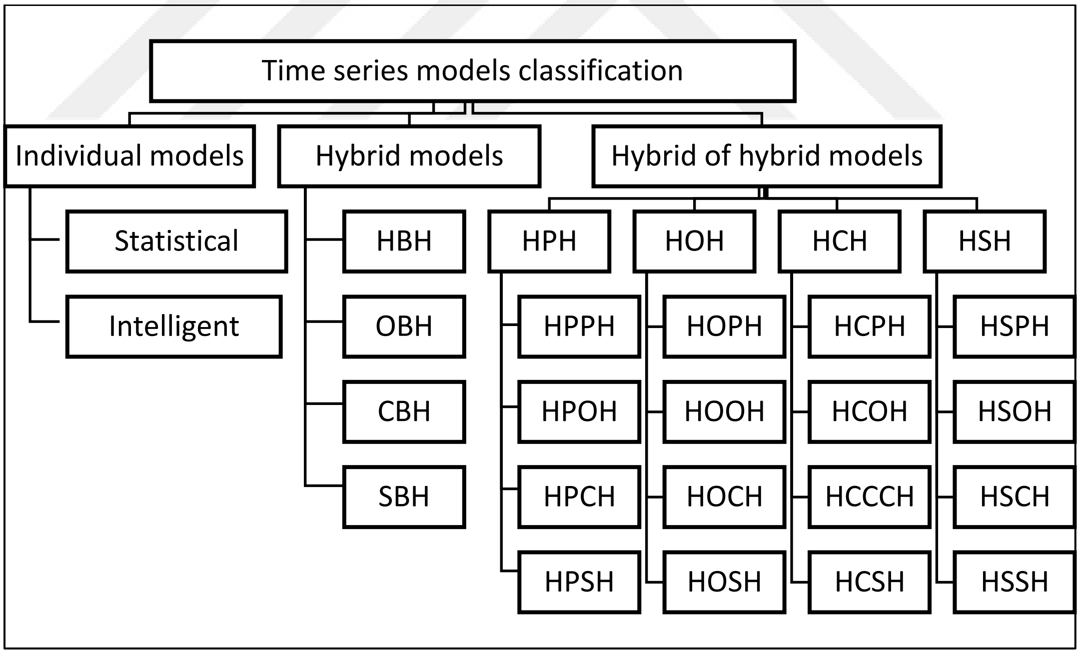


Figure 2.3. categorization of a prediction model in time series models for predicting (Hajirahimi Zahra & Khashei Mehdi, 2022b)

(Song Chenguang et al., 2021)Suggest a hybrid model utilised as an alternative approach for WQ forecasting, which could support the decision maker with a fundamental for water monitoring and contamination control in the basin.

From the comparison results, (Yaseen Zaher Mundher et al., 2018) says the three hybrid models gave a reasonable estimate for the water quality index at the river basin. Depending on the mean of the applied statistics, cluster-based ANFIS-FCM and ANFIS-SC performed with superior accuracy to the ANFIS-G.

From (51) journals' papers disseminated during the period (2000-2016) and interacting with the utilizing the hybrid and single artificial intelligence models for aquatic system quality forecasting were chosen, (Rajaee Taher et al., 2020)recommended the ANN Model with haybird. Moreover, he refers to improving their WQ modelling accuracy when using it.

2.9. Metaheuristic Optimisation Algorithms

Metaheuristic algorithms are computational intelligence paradigms mainly used for sophisticated solving optimisation problems(Abdel-Basset Mohamed et al., 2018). Due to MA's superior performance and shorter computation times and resources than deterministic algorithms in various optimisation specified, (MAs) have recently become popular in many practical disciplines(Wang Mingjing et al., 2017).

Numerous optimisation methods can be adapted to handle various issues for different applications. Recently, the (PSO) Particle Swarm Optimisation algorithm has been utilised in various scientific areas, such as estimating biological treatment(Ethaib Saleem et al., 2016).

In 2020, Li Shimin et al. (2020) developed the slime mould algorithm (SMA), which has been applied to solve computational problems based on the natural slime mould oscillation mode. SMA has a number of novel features, including a unique mathematical model that simulates the production of positive and negative feedback of the slime mould propagation wave based on a bio-oscillator to create the ideal path for connecting food with excellent exploratory ability and exploitation propensity. The proposed SMA is compared with up-to-date metaheuristics using an extensive set of benchmarks to verify its efficiency. Also, According to various statistical criteria, the hybrid model SMA-ANN performs better than ANN (stand-alone) in forecasting urban stochastic water demand (Zubaidi Salah L et al., 2020).

Moreover, A comprehensive comparison was applied by (Mostafa, Rezk et al. 2020) between existing photovoltaic cell parameter extraction methods and the proposed slime mould algorithm. The outcomes demonstrate that the suggested slime mould algorithm outperforms and is more accurate than the existing prevalent techniques.

Additionally, according to (Liu, Heidari et al. 2021), the simulation findings show that a new SMA-based approach could accurately extract the anonymous variables of the unknown photovoltaic solar cells and obtain superior convergence stable and rapidity performance.

In 2020, (Faramarzi Afshin et al., 2020) created Marine Predators Algorithm (MPA), a metaheuristic with a naturalistic inspiration. MPA is primarily motivated by the widespread foraging strategies of ocean predators, especially Lévy and Brownian movements, and the best encounter rate policy in biological interactions between predators and prey. As a result, MPA was adopted in several scientific applications.

According to (Al-Qaness Mohammed AA et al., 2020), the results of COVID-19 Forecasting Confirmed Cases show that MPA-ANFIS outperforms all compared methods in almost all performance measures, including (Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Root Mean Squared Relative Error (RMSRE), Coefficient of Determination (R^2)and Mean Absolute Error (MAE)).

(Zubaidi S et al.) Using MPA-ANN algorithm to predict water demand, The findings indicate that the MPA enhances the accuracy of the ANN model.

In 2019, (Rather Sajad Ahmad & Bala P Shanthi, 2019b) introduced a novel hybridization algorithm called CPSOCGSA, which combines the gravitational search algorithm with particle swarm optimization based on constriction coefficients. In order to achieve the optimal outcome, it combines the exploration and exploitation capabilities of PSO and GSA, respectively.

(Abdul Kareem Baydaa et al., 2022)The hybrid CPSOCGSA-ANN outperformed and produced reliable outcomes in streamflow forecasting algorithms based on various statistical criteria such as RMSE, MAE, and R^2=0.91.

(Alawsi Mustafa A et al., 2022) Compare different Metaheuristic algorithms for Drought prediction. The finding was that The performance of CPSOCGSA-ANN was superior to other algorithms with a coefficient of determination R^2=0.93

Depending on these findings, four optimisation methods are used in this thesis to estimate the number of neurons in each hidden layer coefficient and the optimal number for the learning rate of the ANN model to forecasting long term water quality considering the climate factor.

3. METHODOLOGY

The framework of the proposed methodology comprises five stages: study area and data collection, calculating the WQI, preprocessing for data, metaheuristic optimisation CPSOCGSA, MPA,SMA and PSO algorithm, artificial neural network(ANN) and evaluation of the model's performance, as shown in Figure 3.1. In this work, Matlab Toolbox was running to implement the ANN technique.

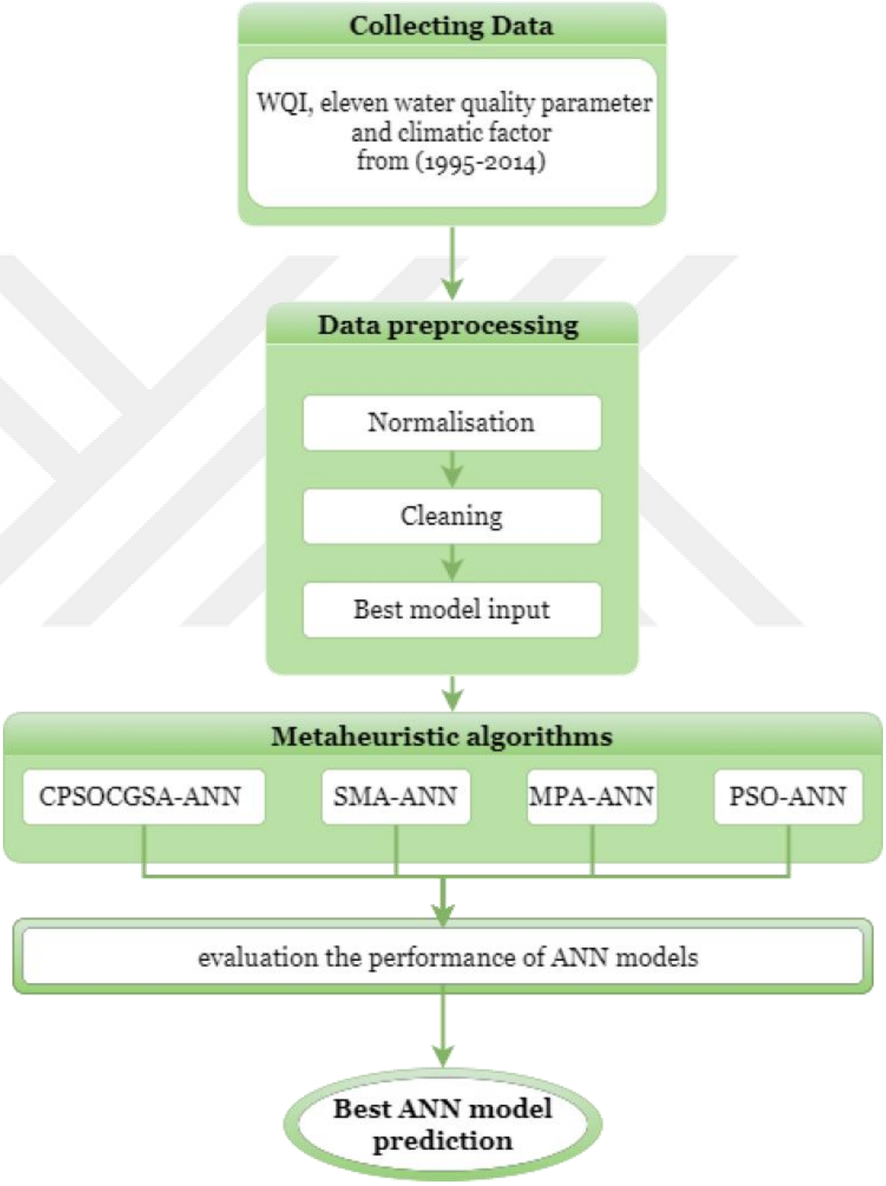


Figure 3.1. Flowchart of the architecture of ANN to predict WQI

3.1. Study Area and Data Collection

The Yesilirmak River is considered one of the longest rivers in Turkey and located between approximately 29′ and 40′ N. 31′ 40′ N and 44′ 71′ E. 44′ 61′ E, with

39,628 km^2 covering's area and discharges into the Black Sea in Samsun as shown in Figure 3.2 (b). The basin has transported partially and fully treated municipal wastewater and pollution from various diffuse sources, including farmyards and agricultural areas. Additionally, different delta parts have dried up in recent decades (Dinc, Çelebi et al. 2021).

As (see Figure 3.2) (a) The maps of Yesilirmak River catchment in Northern Turkey. The River catchment including tributaries Kelkit, Cekeret, Çorum and Tersakan rivers and major cities (population exceeds 50,000).

The station case study was located within the provincial borders of Çorum, Eastern Black Sea Basin. Therefore, the coastal sea climate is dominant in this basin zone. The Annual precipitation was 443 mm.

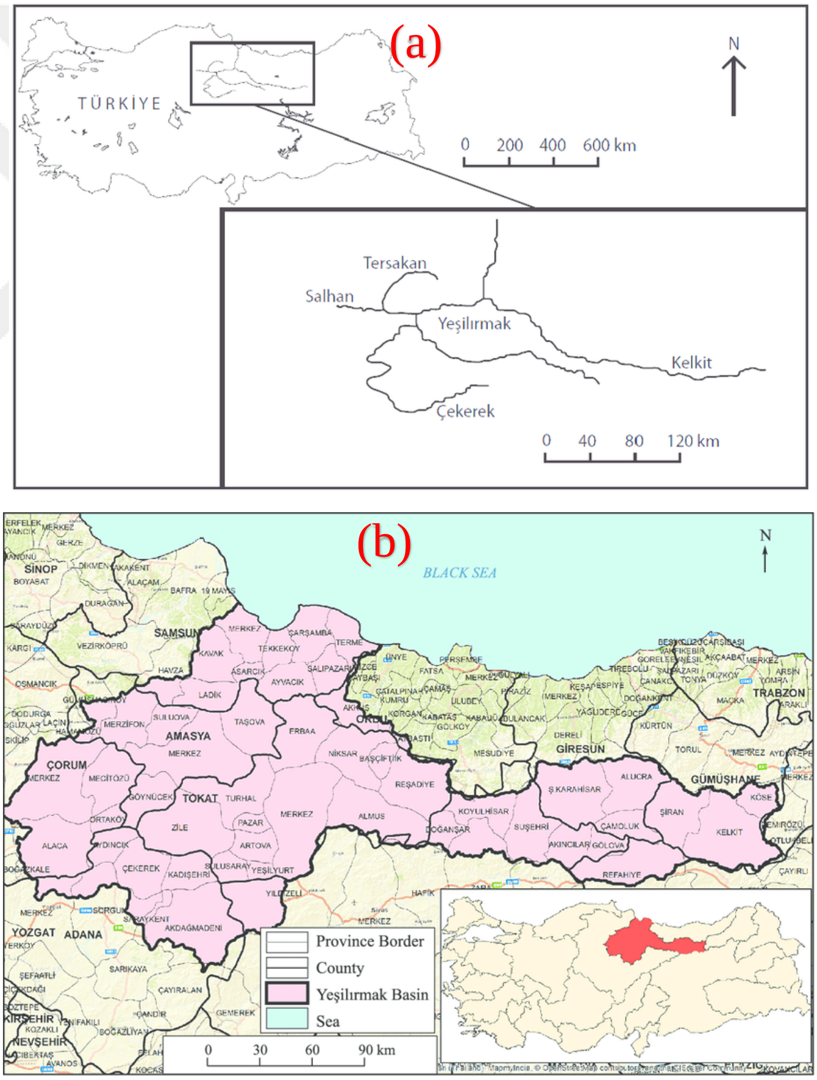


Figure 3.2 Yesilirmak River,and their basin (Başörens Özge Ertunç & Kazanci Nilgün, 2012; Özden S & Ayan S, 2016)

In several countries, the essential obstructive faced by researchers is the lack of data. Therefore, these stations were sufficiently selected for assessing and building prediction models regarding the number of data and availability of parameters. Additionally, National Oceanic and Atmospheric Administration (NOAA) (NOAA 2021) was adopted for obtaining the climate variables. The dataset was collected for twenty years, from 1995 to 2014(Table 1) . It contained eleven water quality parameters with one influential climatic factor (Rainfall), where the values are distributed regularly along the monitoring periods. Table 3.1 Presents the variables for the water quality,the water standards, assigned weights, and relative weight for the WQI's Equation.

3.2. Calculation of Water Quality Index

The Water Quality Index(WQI) is an evaluation method which reflects the influence of individual water quality variables on the overall quality of aquatic systems (Ramakrishnaiah CR et al., 2009). For calculating the WQI, 11 significant water quality parameters were specified regarding the World Health Organization (Organization World Health et al., 2004). The physicochemical parameters were (biological oxygen demand (BOD), pH, electrical conductivity (EC), dissolved oxygen (DO), chlorine (Cl^{-1}), calcium(Ca^{+2}), magnesium (Mg^{+2}), nitrite(NO_2^{-1} **),** sodium (Na^{+1}), sulfate (SO_4) and total dissolved solids (TDS) (Ewaid Salam Hussein et al., 2018; Kulisz Monika et al., 2021).In this research, WQI calculation can be categorised in three stages:

Firstly, weights (W_i) were assigned to all water quality parameters with a scale ranging from 1 to 5, reflecting their significance in affecting water bodies. **Table 3.1** The water quality standards, assigned weights, and relative weight of (the WQI's Equation).

Table 3.1. The water standards, assigned weights, and relative weight for the WQI's Equation (Sharma Prerna et al., 2014), (Kulisz Monika et al., 2021), (Kadam AK et al., 2019) and (Şener Şehnaz et al., 2017)

Parameters	**Water quality standard**	**Assigned weight (AW.)**	**Relative weight (RW.)**
BOD_5 mg/l	5	4	0.102564
DO mg/l	5	5	0.128205
PH	6.5–8.5	4	0.102564

Parameters	Water quality standard	Assigned weight (AW.)	Relative weight (RW.)
Ca^{+2} mg/L	300	2	0.051282
Cl^{-1} mg/l	250	3	0.076923
EC µS/cm	250	5	0.128205
Mg^{+2} mg/L	30	2	0.051282
Na^{+1} mg/l	200	2	0.051282
NO_2^{-1} mg/l	3	4	0.102564
SO_4^{-2} mg/l	250	4	0.102564
TDS mg/l	1000	4	0.102564

Secondly, calculate the relative weights (W_i) by Equation (1):

$$Wi = \frac{wi}{\sum_{i=1}^{n} wi} \quad \textbf{(3.1)}$$

Wi is the relative weight, wi is the weight of every parameter, and n is the number of parameters.

$$qi = (ci/si) * 100 \quad \textbf{(3.2)}$$

$$qi = ((ci - vi)/(si - vi)) * 100 \quad \textbf{(3.3)}$$

Thirdly, from Equation (2), a quality rating scale(q_i) was assigned by dividing every parameter concentration in the water samples by its respective standard in WHO guidelines. In contrast, The quality rating for pH and DO was calculated by Equation(3).

Where C_i = the measured concentration of parameters in each simple, S_i = the drinking water standard for each parameter, Vi = Ideal value for pH = 7.0 and DO = 14.6, and SIi = the subindex of parameters.

Finally, WQI was derived from formulas (4) and (5).

$$SIi = qi * Wi \quad \textbf{(3.4)}$$

$$WQI = \sum SIi \quad \textbf{(3.5)}$$

Table 3.2 presents the classification of WQI rgarding the range proposed by previous studies (Aldhyani Theyazn HH et al., 2020; Ewaid Salam Hussein et al., 2018).

Table 3.2. water quality classification(WQC) (Ramakrishnaiah CR et al., 2009)

WQI range	0-25	26-50	51-75	76-100	>100
classified	Excellent	Good	Poor	Very poor	Unsuitable for drinking

3.3. Data Pre-processing

Three strategies were adopted in this study as preprocessing data techniques, which are normalisation, cleaning, and select the optimal model's input:

1. Normalisation

Data normalisation is a crucial step for mining datasets in soft computing. In various scientific research, techniques and assessment criteria mostly have various scale sizes and units, producing various analysis data results. Therefore, normalisation will be a fundamental task to achieve comparability among the variables and realise the expectations of optimisation data. Moreover, it minimises the dimension's impact of different time series and eliminates the influence of outliers and multicollinearity of model parameters (Wu Junhao & Wang Zhaocai, 2022). In this work, the SPSS 26 statistics package was employed to normalise the time series by applying the natural logarithm technique.

2. Cleaning

Data cleaning methods comprise locating outliers and noise and then treating outliers and eliminating noise to optimize the data analysis outcomes (Tabachnick Barbara G et al., 2007). So, the box whisker technique was adopted in this study to identify the outliers were existed outside of the period in the formula below (Kossieris Panagiotis & Makropoulos Christos, 2018):

$$\mp 1.5 * CQR\ (CQR\ =\ third\ quartile\ (Q3) - first\ quartile\ (Q1)) \qquad \textbf{(3.6)}$$

Also, the technique of singular spectrum analysis (SSA) is used to detect and remove noise from time series by decompose the original time series into different components after that eliminate the noise signal . It is an efficient method for analyzing the initial time series into various components (Golyandina Nina et al., 2018; Karami Farzane & Dariane Alireza B, 2022). This approach has shown to be effective in diverse speciality areas, such as rainfall prediction(Reddy Pundru Chandra Shaker et al., 2022), hydrology (Ouyang Qi & Lu Wenxi, 2018), drought forecasting (Pham Quoc Bao et al., 2021), and groundwater prediction (Polomčić Dušan et al., 2017).

3. Identifying The Best Model Input

Determining the appropriate predictors is fundamental in establishing a forecast model's structure and enhancing the performance of model. Therefore, the tolerance approach was applied to identify the optimal scenario for the predictors to avoid multicollinearity (Calì Davide et al., 2016; Pallant Julie, 2020). It recommended a tolerance coefficient value equal to or higher than 0.2 for selecting the model's predictors.

3.4. Optimisation Algorithms

4. Hybridised Constriction Coefficient-Based Particle Swarm Optimisation and Chaotic Gravitational Search Algorithm (CPSOCGSA)

CPSOGSA is a hybrid heuristic optimization that utilizes the intensification potential of the CPSO algorithm with the diversification capability of GSA's algorithm. The components of this hybridization technique will be clarified in the subsections below.

A. Constriction Coefficient based Particle Swarm Optimization (CCPSO.)

The PSO algorithm is a popular optimization approach inspired by natural swarm behaviour for birds or fish. The PSO architecture consists of three principal parameters: abest, bbest and inertia weight. Where the abest and bbest assist the finding of the search-space region. The inertia weight has a significant impact on the process of global exploration. In (Equations 7,8), the Particle Swarm mathematical formulation explains the updating of the process for the particle's location and velocity during the alteration of the particle values.

$$v_x^d(t+1) = w(t)v_x^d(t) + c_1 r_{x1}(abest - x_x^d(t)) + c_2 r_{x2}(bbest - x_x^d(t)) \quad \textbf{(3.7)}$$

$$x_x^d(t+1) = x_x^d(t) + v_x^d(t+1) \quad \textbf{(3.8)}$$

Where the c_1,c_2 are the learning constants, while r_{x1} and r_{x2} are the numbers range from 0 to 1.

Constriction coefficients were employed to improve Particle Swarm Optimisation (PSO) exploitation stage by minimizing the impact of particle movements outside the solution space and hastening convergence during the optimization stage. The coefficients are described as follows:

$$\varphi_1 = 2.05, \varphi_2 = 2.05, \ \varphi = \varphi_1 + \varphi_2 \quad \textbf{(3.9)}$$

$$K = 2/\left(\varphi - 2 + \sqrt{(\varphi^2 - 4)}\right) \quad \textbf{(3.10)}$$

Substituting the inertia weight by the notation K, (Equation 11) can be rewritten as in below:

$$v_x^d(t+1) = Kv_x^d(t) + K\varphi_1 r_{x1}\Big(abest_x(t) - x_x^d(t)\Big) + K\varphi_2 r_{x2}\Big(bbest - x_x^d(t)\Big) \quad \textbf{(3.11)}$$

Where $K\varphi_1 = c_1$, $K\varphi_2 = c_2$

b. Chaotic Gravitational Search Algorithm CGSA.

Gravity and motion Newton's law is considered fundamental to the configuration of the GSA algorithm. Newton's law states that "the gravitational force between two masses is directly proportional to the product of their masses and inversely proportional to the square of the distance between them". Accordingly, the gravitational force (F_ij) between masses (i.e., searching agents) (x) and (y) at time (t) can be represented as in the following (Equation 12):

$$F_{xy} = G(t)\frac{m_{px}(t)m_{ay}(t)}{R_{xy}(t)+\in}\Big(x_x^d(t) + x_y^d(t)\Big) \quad \textbf{(3.12)}$$

Where m_{px} = the attractive mass, $and\ m_{ay}$ = passive masses. While $R_{xy}(t)$ = the Euclidian distance between the two masses at the time (t), while$\in=$ is a small coefficient, and G = the constant help for controlling the solution space to secure a feasible region.

The (G) constant can be represented by (Equation 13):

$$G(t) = G(t_o)e^{\left(-\alpha\frac{CI}{MI}\right)} \quad \textbf{(3.13)}$$

Where $G(t_o)\&G(t) =$ the initial and final values of G, respectively. CI = the current iteration, α = the small constant, and MI = the maximum value of iterations. The behaviour of G over time is proposed using a chaotic normalization (Rather Sajad Ahmad & Bala P Shanthi, 2019a, 2019b). Hence, (Equation 14) descript the final formula of the gravitational constant's:

$$G^c(t) = C_i^{norm}(t) + G(t_o)e^{\left(-\alpha\frac{CI}{MI}\right)} \quad \textbf{(3.14)}$$

The aggregate force exerted by the masses (i.e., searching agents) could be found as shown in (Equation 15):

$$F_x^d(t) = \sum_{y=1, y\neq x}^{m} \gamma_y F_{xy} \quad \textbf{(3.15)}$$

Where γ value ranges from 0 to 1, it is essential to calculate the position and velocity to find the global optimum, which could be represented according to

(Equations 16, 17):

$$v_x^d(t+1) = \gamma_y v_x^d(t) + a_x^d(t) \quad \textbf{(3.16)}$$

$$x_x^d(t+1) = x_x^d(t) + v_x^d(t+1) \quad \textbf{(3.17)}$$

Where $a_x^d(t) =$ the mass acceleration.

C. the Combination algorithm (CPSOCGSA)

Combining the two heuristic techniques (CPSO and CGSA) could assist to exceeded each technique's imperfections. The hybridization equation formula can be described in (Equation 18):

$$v_x^d(t+1) = \left(2/\left(\varphi - 2 + \sqrt{\varphi^2 - 4}\right)\right) v_x^d(t) + K\varphi_1 r_{x1}\left(a_x^d(t) - x_x^d(t)\right) + K\varphi_2 r_{x2}\left(bbest - x_x^d(t)\right) \quad \textbf{(3.18)}$$

The particle's location is indicated by (Equation 19):

$$x_x^d(t+1) = x_x^d(t) + v_x^d(t+1) \quad \textbf{(3.19)}$$

PSOCGSA is a hybrid heuristic optimisation that utilises the intensification potential of the CPSO algorithm with the diversification capability of GSA's algorithm. The components of this hybridisation technique will be clarified in the subsections below.

5. Marine Predators Algorithm (MPA)

Faramarzi Afshin et al. (2020) suggested MPA as one of the most recent nature-inspired optimisation algorithms. The idea of the MPA algorithm is derived from the movement of marine predators such as sunfish and sharks.

Step One: Prey’s Population Initialisation

The MPA starts by setting an initial random solution group X_0 according to Equation (20) This random group is generated within the search space:

$$X_0 = X_{lb} + r * (X_{ub} - X_{lb}) \quad \textbf{(3.20)}$$

Where: X_{lb} is the lower bond of variables $and\ X_{ub}$ is the upper bond of variables.

r is a random vector in a range (0-1).

Step Two: Predator Matrix Creation

In the MPA, both prey and predators are regarded as search agents because they seek their own food. The top predator in the search agents, who is naturally more skilled than other search agents, is called the Elite. Based on the information on prey locations, the Elite matrix is mathematically updated. The Elite and prey matrix is constructed as follows:

$$Elite = \begin{bmatrix} X_{11}^1 & X_{12}^1 & & X_{1d}^1 \\ X_{21}^1 & X_{22}^1 & \vdots & X_{2d}^1 \\ . & . & . & . \\ . & . & \vdots & . \\ X_{n1}^1 & X_{n2}^1 & & X_{nd}^1 \end{bmatrix} \tag{3.21}$$

$$X = \begin{bmatrix} X_{11} & X_{12} & . & X_{1d} \\ X_{21} & X_{22} & . & X_{2d} \\ . & . & . & . \\ X_{n1} & X_{n2} & . & X_{nd} \end{bmatrix} \tag{3.22}$$

Step Three: MPA Optimisation Process

After creating the prey and Elite matrices, the locations of the prey and predators are updated in three phases. The velocity ratio between the prey and the predator determines these phases. The three phases consist of low-velocity ratio, unit velocity ratio, and high-velocity ratio.

Phase One: High-Velocity Ratio

In this phase, the predator moves more quickly than the prey. Additionally, the step size of prey movement is updated as shown in Equations (23) and (24).

$$S_i = R_B \otimes (Elite_i - R_B \otimes X_i), \qquad i = 1,2,\ldots,n \tag{3.23}$$

$$X_i = X_i + P.R \otimes S_i \tag{3.24}$$

Where R: random vector where the values of its elements are in a range (0-1) value., *P*: constant number.

R_B: random vector refereeing to the Brownian motion.

The $\otimes$ symbol represents element-wise multiplication process.

This phase happens in one-third of the total number of iterations (i.e., $\frac{1}{3} t_{max}$).

Phase Two: Unit Velocity Ratio

This phase simulates the process of looking for prey or food. Levy flight represents the movement of the prey, while Brownian motion represents the predator's movement.

This phase happens during the second-third of all iterations (i.e., $\frac{1}{3}t_{max} < t < \frac{2}{3}t_{max}$). The first half of the population is represented by using the following Equations.

$$\mathrm{S_i} = \mathrm{R_L} \otimes (\mathrm{Elite_i} - \mathrm{R_L} \otimes \mathrm{X_i}), \qquad \mathrm{i} = 1,2,\dots,\mathrm{n} \tag{3.25}$$

$$\mathrm{X_i} = \mathrm{X_i} + \mathrm{P.R} \otimes \mathrm{S_i} \tag{3.26}$$

Where R_L: numbers following Levey distribution. The Equations (27) and (28) are applied to the remaining half of the population:

$$S_i = R_B \otimes (R_B \otimes Elite_i - X_i), \qquad i = 1,2,\dots,n \tag{3.27}$$

$$X_i = X_i + P.CF \otimes S_i \,, CF = \left(1 - \frac{t}{t_{max}}\right)^{2\left(\frac{t}{t_{max}}\right)} \tag{3.28}$$

CF: The parameter that controls the movement of predator step size.

Phase Three: Low-Velocity Ratio

This phase is the last one in the optimisation process and simulates the predator's movements when it is quicker than the prey. It happens in the latter third of all iterations (i.e., $\frac{2}{3}\mathrm{t_{max}}$).

$$S_i = R_L \otimes (R_L \otimes Elite_i - X_i), \qquad i = 1,2,\dots,n \tag{3.29}$$

$$X_i = X_i + P.CF \otimes S_i \,, CF = \left(1 - \frac{t}{t_{max}}\right)^{2\left(\frac{t}{t_{max}}\right)} \tag{3.30}$$

Step Four: Eddy Formation and FADs

It is also possible to include environmental parameters in the simulation, such as the fish aggregation device and eddy formation. The FAD's impact as:

$$X_i = \begin{cases} X_i + CF[X_{min} + R \otimes (X_{max} - X_{min})] \otimes \overline{U} & if\ r < FAD \\ X_i + [FADS(1-r) + r](X_{r1} - X_{r2}) & if\ r > FAD \end{cases} \tag{3.31}$$

Where *r*: random value in a range (0-1)., r1 and r2 refer to the random indices from the pre matrix., FADS: the FAD's probability., and $\overline{U}$: binary vector.

Step Five: Marine Memory

Marine predators have a strong memory for successful foraging positions. This was simulated by allowing the MPA to save the fitness values of the solution after each iteration and compare them to fitness values from successive iterations.

6. Slime Mould Algorithm (SMA)

The slime mould algorithm (SMA), which was developed by Li Shimin et al. (2020), has been utilized to resolve optimization issues such as the Prediction of water consumption (Zubaidi Salah L et al., 2020), estimation of solar photovoltaic cell parameters(Kumar C et al., 2020), and for optimal economic emission dispatch(Hassan Mohamed H et al., 2021). SMA is suggested based on the nature of the oscillation mode of slime mould. The algorithm primarily employs the weights to simulate the negative and positive feedback of the slime mould's wave of propagation based on a bio-oscillator to create the best path for connecting food with exploitation propensity and good exploratory ability. The scenario is described in the following.

Firstly, in the food searching process, slim mould has utilized the odor in the air reaches the foods. Then, based on the slime mould's behaviour, the procedure is formulated below to simulate the mode contraction (Nakagaki Toshiyuki et al., 2000).

$$\overrightarrow{S}(t+1) = \begin{cases} \vec{S}_b(t) + \overrightarrow{vb} * \left(\overrightarrow{W} + \vec{S}_A(t) - \vec{S}_B(t)\right), & r < p \\ \overrightarrow{vc} * \vec{S}(t), & r \geq p \end{cases} \quad \textbf{(3.32)}$$

$\overrightarrow{vb}$ *is randomly generated within* $[a, -a]$ *as*:

$$\overrightarrow{\mathrm{vb}} = [-\mathrm{a}, \mathrm{a}] \quad \textbf{(3.33)}$$

$$\mathrm{a} = \operatorname{arctanh}\left(-\left(\frac{t}{t_{max}}\right) + 1\right) \quad \textbf{(3.34)}$$

$\overrightarrow{S}(t+1)$ refers to the location that the current slime mould (S.M.) will occupy next, $\vec{S}_b$ a vector that includes the place with the greatest concentration of odours so far discovered, $\vec{S}(t)$ indicates the existing location of the (SM), and $\vec{S}_A$ and $\vec{S}_B$ are two vectors representing the positions of two randomly chosen individuals from the population, $\overrightarrow{vc}$ Linearly decreases from (1 to 0), tmax refers to the max iteration, t points to the current iteration, the parameter r is a random number between 0 and 1, and finally $\overrightarrow{W}$ Clarify the weight of slime mould and calculate as follows below:

$$\overrightarrow{\mathrm{W}(\text{smell ındex}(l))} = \begin{cases} 1 + r * \log\left(\dfrac{bF - S(i)}{bF - wF} + 1\right), & condition \\ 1 - r * \log\left(\dfrac{bF - S(i)}{bF - wF} + 1\right), & other \end{cases} \quad \textbf{(3.35)}$$

Under the current iterations, bF refers to the optimal fitness value. At the same time, wF stands for the worst one, (the smell index) points to the fitness values that are sorted, the parameter r is a random number between (0 and 1), and the S(i) is a condition that indicates the rank in the first half of the population. Relative to the variable (p) in Equation (36) is modelled as follows:

$$p = tanh(|f\ (i) - DF\ |) \quad \textbf{(3.36)}$$

Where i ∈ 1, 2, 3, . . ., n, f (i) is the fitness of the current $\vec{X}$, DF is the best fitness obtained so far.

Secondly, This stage (the wrapped) mimics the contraction mode of the slime mould's venous structure, which adjusts its placements based on the food quality; a zone's weight increases when the food concentration is high. Otherwise, the zone's weight is transferred to research other locations. The SM must determine when to leave the present region for a new one to discover a range of food sources concurrently rather than the previous one. In general, the mathematical model of updating the (S.M.) location might be remodelled, as described in Equation (37), to imitate the SM's methodology to identify numerous foods sources when foraging different locations.

$$\vec{S}(t+1) = \begin{cases} \vec{S}_b(t) + \overrightarrow{vb} * \left(\vec{W} + \vec{S}_A(t) - \vec{S}_B(t)\right), & r < p \\ \overrightarrow{vc} * \vec{S}(t), & r \geq p \\ rand * (UB - LB) + LB, rand & < z \end{cases} \quad \textbf{(3.37)}$$

The search space's upper and lower boundaries are denoted by UB and LB, respectively. (rand, and r) are randomly generated, ranging (from 0 to 1). The probability z determines whether the SMA will explore other food sources or stay close to the best available ones. In relative to $\overrightarrow{vb}$, $\vec{W}$, and $\overrightarrow{vc}$, They are applied to simulate the change in venous width. The phases of SMA are described in the script below:

SMA, similar to other metaheuristic algorithms, requires a swarm to begin the search. More details on the SMA algorithm can be found in Li Shimin et al. (2020).

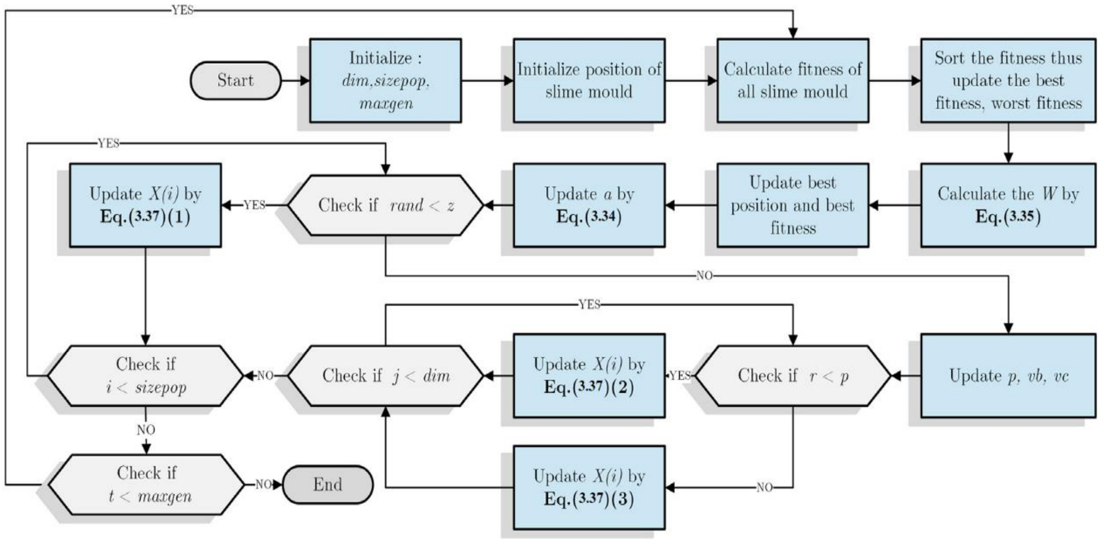


Figure 3.3. Flowchart of SMA Li Shimin et al. (2020)

3.5. Artificial Neural Network (ANN)

ANNs are an information-processing approach designed to simulate the functioning of the human brain by attempting the same connectivity and processes as biological neurons (Kouadri Saber et al., 2022). In this research, the ANN structure was composed of four layers of neurons: an input layer which has WQ parameters and rainfall data, two hidden layers to address the nonlinearity relationship, and an output layer which has the target (WQI)(see Figure 3.4). The multilayer perceptron (MLP) is performed with feed-forward backpropagation (FFBB). The efficiency with a low error rate and speed of the learning algorithm (Levenberg-Marquardt, LM) was an underlying factor employed with FFBB for training the ANN model (Payal Ashish et al., 2015). Matlab Toolbox was running to implement the ANN technique.

The data were separated into three groups: training, testing, and validation, with 70%, 15%, and 15% of values utilised for each set, respectively, such as (Kulisz Monika et al., 2021; Zubaidi Salah L., Gharghan Sadik K., et al., 2018).

However, the traditional method (trial and error) was time-consuming and did not always introduce the optimal solution. As a result, combining metaheuristic algorithms with the ANN model is regarded as a superior technique for determining the optimal neurons’ number in the hidden layers (N1, N2) and selecting the best learning rate coefficient (Lr). These hyperparameters are responsible for choosing the best predictors and target mapping to avoid underfitting or overfitting the model.

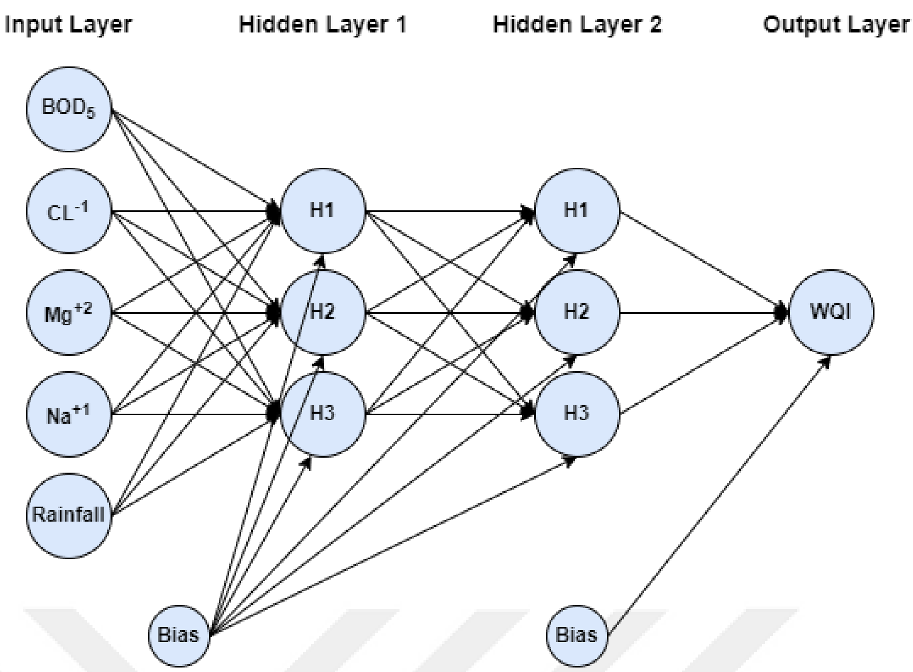


Figure 3.4. ANN structure for CPSOCGSA-ANN

3.6. Model Performance Assessment

Three statistical criteria are used in this study to evaluate the WQI prediction model's performance. The criteria are the root means squared error (RMSE, (Equation 38)), mean absolute error (MAE, (Equation 39)), and determination coefficient (R^2, (Equation 40)) (Mohammadi Babak et al., 2020; Mohammadi Babak & Mehdizadeh Saeid, 2020):

$$MAE = \frac{\sum_{i=1}^{N} |O_i - F_i|}{N} \quad \textbf{(3.38)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (O_i - F_i)^2}{N}} \quad \textbf{(3.39)}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N} (O_i - \bar{O}_i)(F_i - F_i)}{\sqrt{\sum (O_i - \bar{O}_i)^2 \sum (F_i - \bar{F}_i)^2}}\right]^2 \quad \textbf{(3.40)}$$

where O_i= represents observed WQI, F_i= predicted WQI, N = sample size, $\bar{F}_i$ = mean of predicted WQI, and $\bar{O}_i$= mean of observed WQI.

Additionally, the Taylor diagram was selected to evaluate the modelling results by presenting pattern statistics to create a visual comprehension of performance by showing diverse points on a polar plot for multisets of modelling outcomes (Taylor Karl E., 2001).

4. RESULTS AND DISCUSSION

4.1. Water Quality Index Assessment

The assigned weight index approach was juxtaposed with the WHO standards for drinking water quality to calculate The Water Quality Index for The Yesilirmak river. Table 4.1 presents the eleven parameters that were combined for the WQI prediction and the descriptive statistics values for them. As can be observed, the WQI varies from 37.41 to 181.06, and the mean was 51.36; this variation reflects the unsuitability for drinking, as shown in Table 3.2 .calcium(Ca^{+2}), nitrite(NO_2), chloride(Cl^{-1}), and sodium(Na) contents did not exceed the acceptable range in any tested values. While pH, biological oxygen demand (BOD), electrical conductivity (EC), magnesium (Mg^{+2}), sulfate (SO_4^{-2}) and total dissolved solids (TDS) were higher than the specified value in the WHO guidelines. In the same context, the dissolved oxygen exceeded the minimum acceptable value in six samples during the monitoring periods.find Table 1 (full calculation for WQI of Yesilirmak river).

Table 4.1. Descriptive statistics for the Yesilirmak river (1995-2014).

Variables	Mean	Min.	Max.	Std. Error	standard
WQI	51.36	37.41	181.06	1.51	-
BOD_5 (mg/l)	4.06	0.0	40.00	0.46	5
DO (mg/l)	7.60	0.04	12.70	0.18	5
pH	8.25	6.90	9.00	0.03	6.5–8.5
Ca^{+2} (mg/L)	55.59	19.00	129.00	1.55	300
Cl^{-1} (mg/l)	22.78	1.28	102.90	1.04	250
EC (µS/cm)	710.03	399.00	1903.00	14.66	250
Mg^{+2} (mg/L)	39.81	17.60	105.20	0.94	30
Na^{+1} (mg/l)	37.83	10.16	73.80	1.02	200
NO_2^{-1} (mg/l)	0.08	0.0	0.80	0.01	3
SO_4^{-2}(mg/l)	54.50	16.80	352.00	3.20	250
TDS (mg/l)	461.16	241.00	1217.90	9.49	1000

4.2. Preparation of Input and Target Variables

The input data (i.e., WQ parameters and rainfall) and target data (WQI)) were normalized by applying anatural logarithm (Ln) to minimize the adverse impact of extreme values and achieve the normal time series distribution. Then, if any outliers remained after the normalization, it was adjusted. The box plot technique displayed in (Figure 4.1) decreased the variation scale of normalization data compared to the raw data. It also showed how the data had been cleaned from outliers.

After that, the SSA was applied to decompose the normalized and cleaned time series into three different components to eliminate the noise. Also, the data pre-processing technique enhanced the correlation coefficient (R) between the dependent and independent factors, such as improving the correlation between WQI and chlorine from 0.609 to 0.7106.

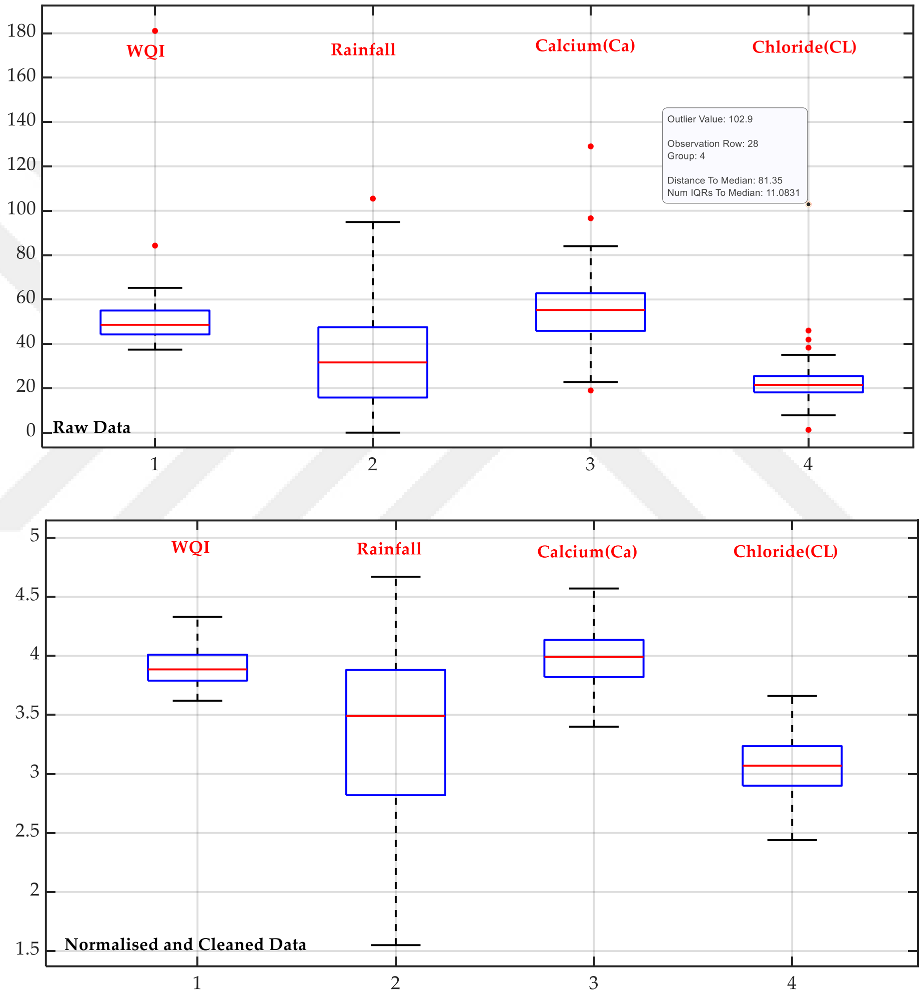


Figure 4.1. Box plot drawings of WQI with some of the selected parameters before and after Normalised and cleaned data

After that, the singular spectrum analysis technique (SSA) was applied to decompose the original time series into three different components to eliminate the noise. The initial time series (after normalised and cleaned) for WQI (1st shape), the denoise time series (2nd shape) and two noise signals (3rd and 4th shapes) are shown in Figure 4.2.

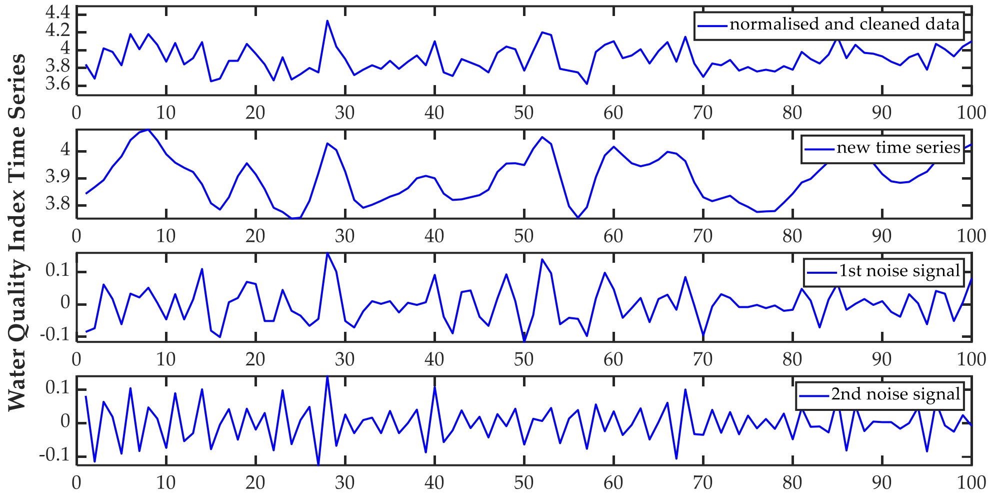


Figure 4.2. the WQI initial time series, the denoise time series and two noise signals

Also, We observed that the (SSA) technique enhanced the correlation coefficient (R) between the water quality parameters and the target (WQI), such as improving the correlation of the water quality index with Chlorine and magnesium (from 0.609 to 0.7106),(from 0.527 to 0.616) respectively.

Then, A tolerance approach was applied to select the optimal model input scenario to simulate the WQI with high accuracy and prevent multicollinearity by clearing superfluous parameters. According to (Pallant Julie, 2020), the tolerance coefficient for the nominated input variables must be higher than 0.2. As shown in Table 4.2 below, five variables were chosen; biological oxygen demand, Chlorine, magnesium, potassium and rainfall as the optimal input data model.

Table 4.2. Collinearity statistics for the specified predictors

Predictors	**Tolerance value**
BOD_5	0.645
CL^{-1}	0.352
Mg^{+2}	0.543
Na^{+1}	0.520
Rainfall	0.729

In the last step, according to (Gharghan, Nordin and Ismail, 2016; M Kulisz and Kujawska, 2021), the preprocessing data were categorised into 70% training, 15% testing, and 15% validation to create and evaluate the forecasting model.

4.3. Configuring the Hybrid Model

The artificial neural network was hybridized with a metaheuristic approach to locate the optimum LR, N1, and N2. The Matlab program was used for running the hybrid models (i.e., MPA-ANN, CPSOCGSA-ANN,SMA-ANN and PSO-ANN) to identify the optimum ANN's hyperparameters (LR, N1, and N2). This research utilized swarm sizes (10, 20, 30, 40, and 50 pop) for every swarm repeated five times with (200 iterations) to achieve the minimum fitness function (MSE). For example, Figure 4.3 depicts the CPSOCGSA-ANN performance and exposes the optimum fitness function for all swarms of WQI.

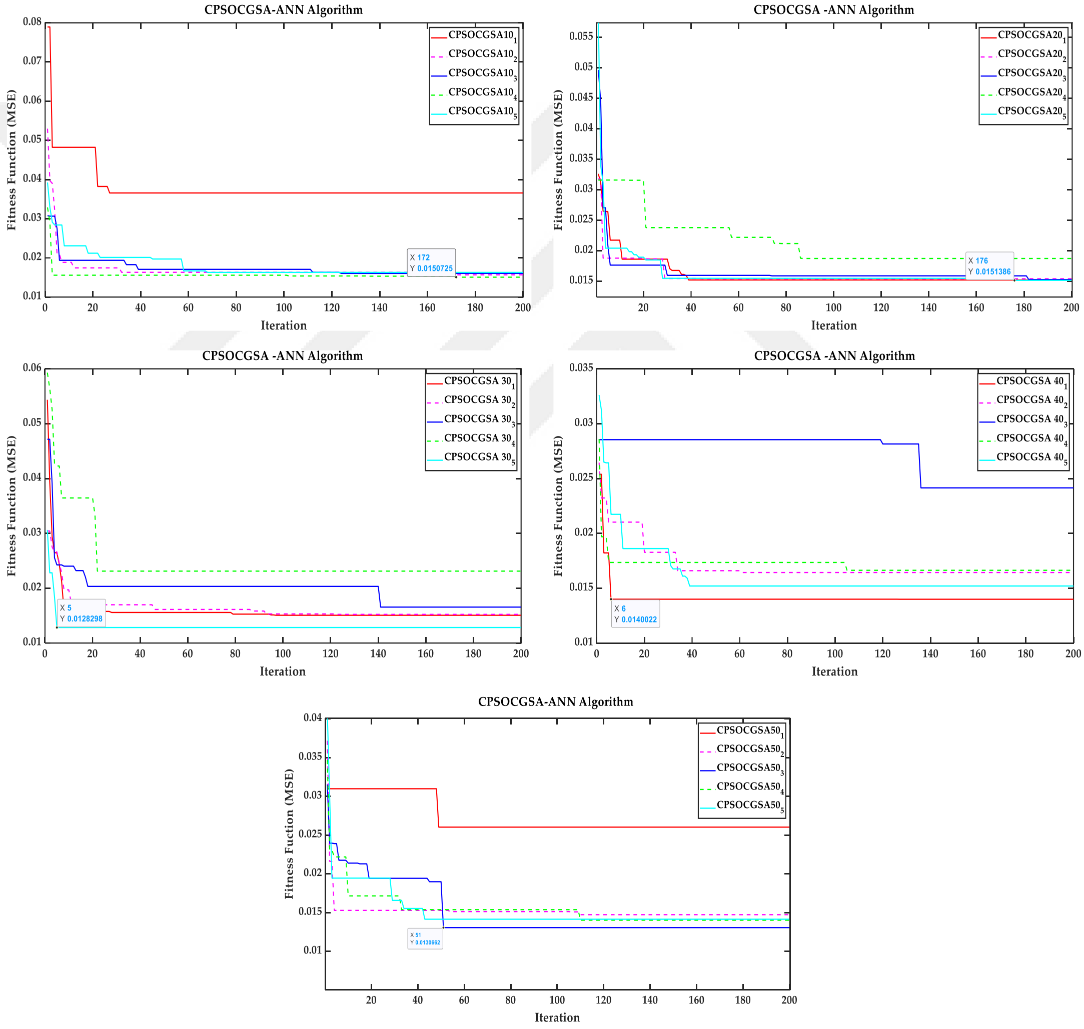


Figure 4.3. CPSOCGSA algorithm Performance for WQI

In addition, Figure 4.4 depicts the effectiveness of the MPA-ANN algorithm and the optimal fitness function for each swarm in WQI.Also,Figure 4.5 displays the performance of the SMA-ANN algorithm and shows the best fitness function for each swarm in WQI.

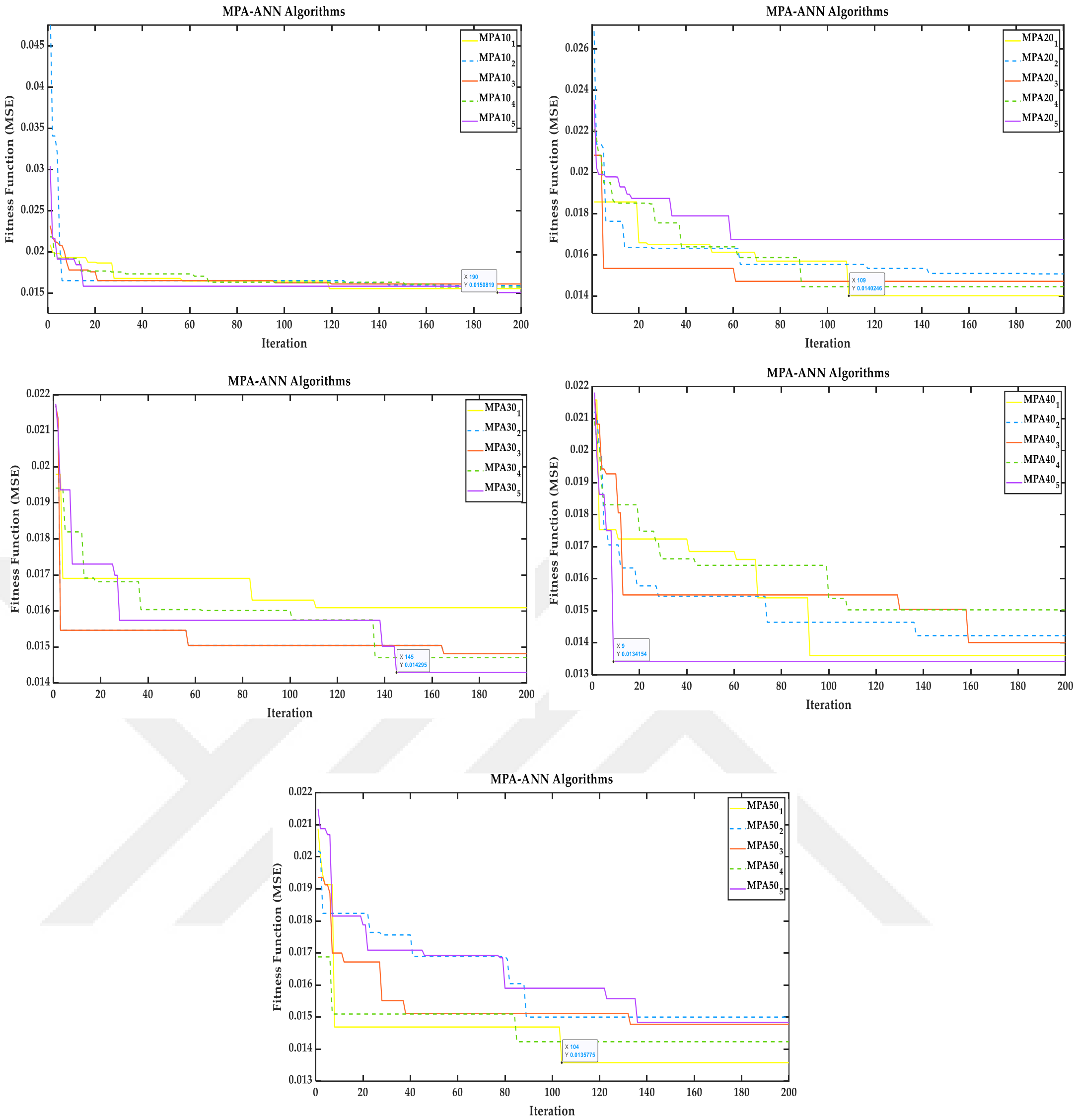
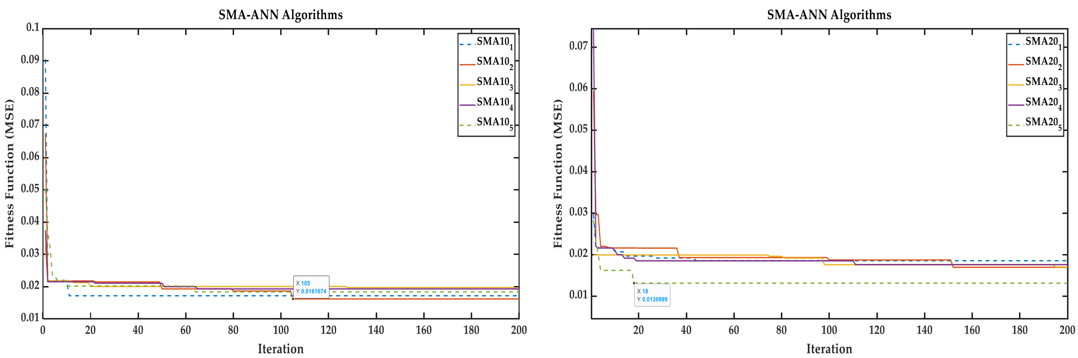


Figure 4.4. Performance of the MPA algorithm for WQI



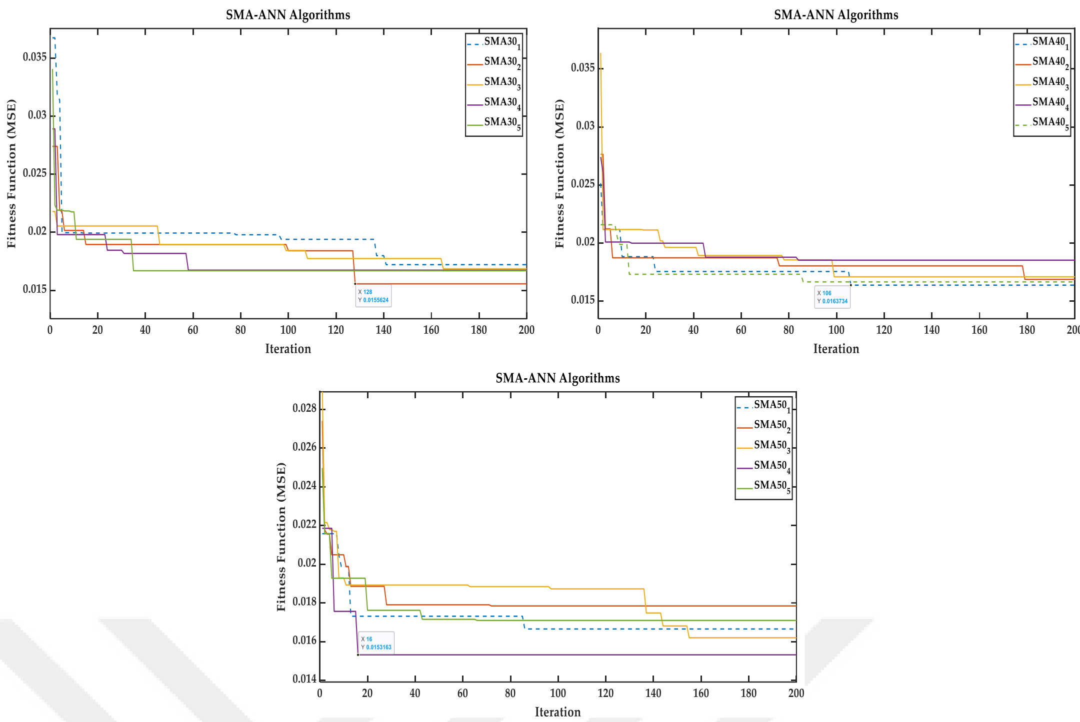
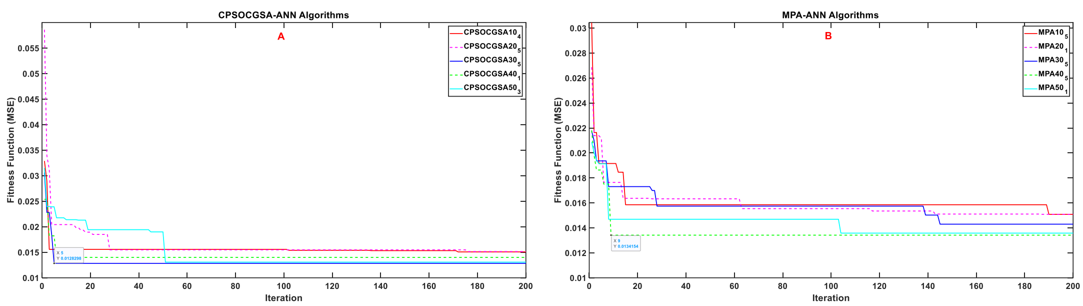


Figure 4.5. Performance of the SMA algorithm for WQI

Hwoever, Figure 4.6(A) shows that the swarm size (30_5) was superior in the CPSOCGSA algorithm and produced the minimum fitness function (MSE = 0.001283, with five Iterations). While, swarm size (40_5) for the MPA algorithm introduces the best solution within the fitness function (MSE= 0.01341, with nine Iterations), as shown in Figure 4.6(B). About the SMA algorithm Figure 4.6(C), the size (40_1) outperformed other swarms and achieved the lowest fitness function (MSE = 0.0131, with 18 Iterations). Finally, the PSO algorithm within swarm size (40_4) was superior to other PSO swarms and gave less error (MSE= 0.01479, with 94 Iterations), as shown in Figure 4.6(D).



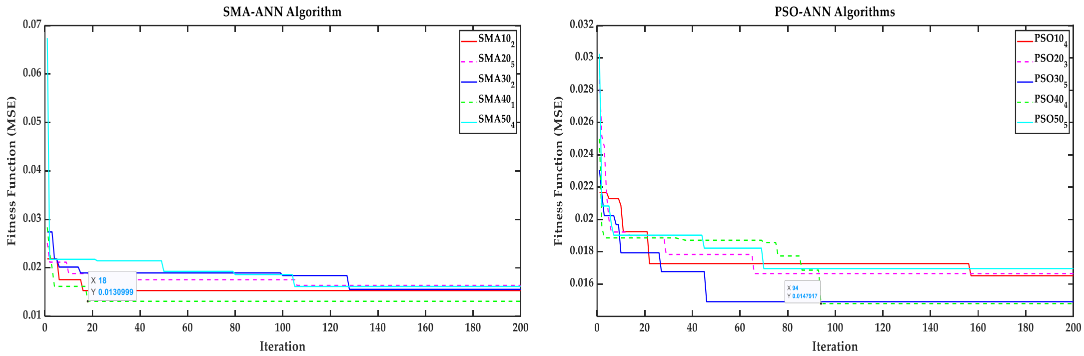


Figure 4.6. Best performed of (CPSOCGSA, PSO, SMA, and MPA) algorithms for WQI modelling

Table 4.3 displays the hyperparameters for ANN models and optimal swarm performance for each metaheuristic algorithm. Where Lr is the rate of learning and N1, N2 is the number of nodes in the hidden layers. Also, For more clarification, find Figure Figure 3.4 above.

Table 4.3. ANN hyperparameters of all algorithms

Model	**Lr**	**N1**	**N2**
CPSOCGSA-ANN	0.6622	3	3
MPA-ANN	0.013	4	1
SMA-ANN	0.0491	4	18
PSO-ANN	0.2321	1	4

4.4. Forecasting Results of Various Algorithms

After determining the optimal hyperparameters for the ANN approach, each ANN method was run several times to identify the optimal network that delivers an accurate result. Three statistical metrics (RMSE, MAE, and R2) were employed to assess the model's effectiveness in predicting WQI data. The results of the performance criteria indices for the CPSOCGSA-ANN, PSO-ANN, and MPA-ANN strategy in the validation section are depicted in Table 4.4 (Dawson Christian W et al., 2007). The CPSOCGSA-ANN was superior to other models because it realized the minimum values of MAE, RMSE criteria and the highest value of R^2.

Table 4.4. Performance assessment for validation data stage.

Model	**MAE**	**RMSE**	**R^2**
CPSOCGSA-ANN	0.01627	0.0187	0.965

Model	MAE	RMSE	R^2
MPA-ANN	0.0308	0.0391	0.937
SMA-ANN	0.0186	0.0210	0.921
PSO-ANN	0.0324	0.0399	0.924

The Taylor diagram (Fig. 4.7) was utilized to evaluate the hybrid models' performance in the validation phase. The measured WQI time series is constituted by the red character (Reference) on the Taylor diagram's X-axis. Taylor's graphical diagram compares three statistics, including correlation coefficient (R), standard deviation (SD), and root mean square error difference (RMSD). It thus delivers a dependable evaluation of the relative performance of various strategy (Ghorbani Mohammad Ali et al., 2018; Tao Hai et al., 2021). Regarding the Taylor diagram, the CPSOCGSA-ANN model was indicated to be the best forecast model compared with the PSO-ANN and MPA-ANN models because it is the nearest to the reference point.

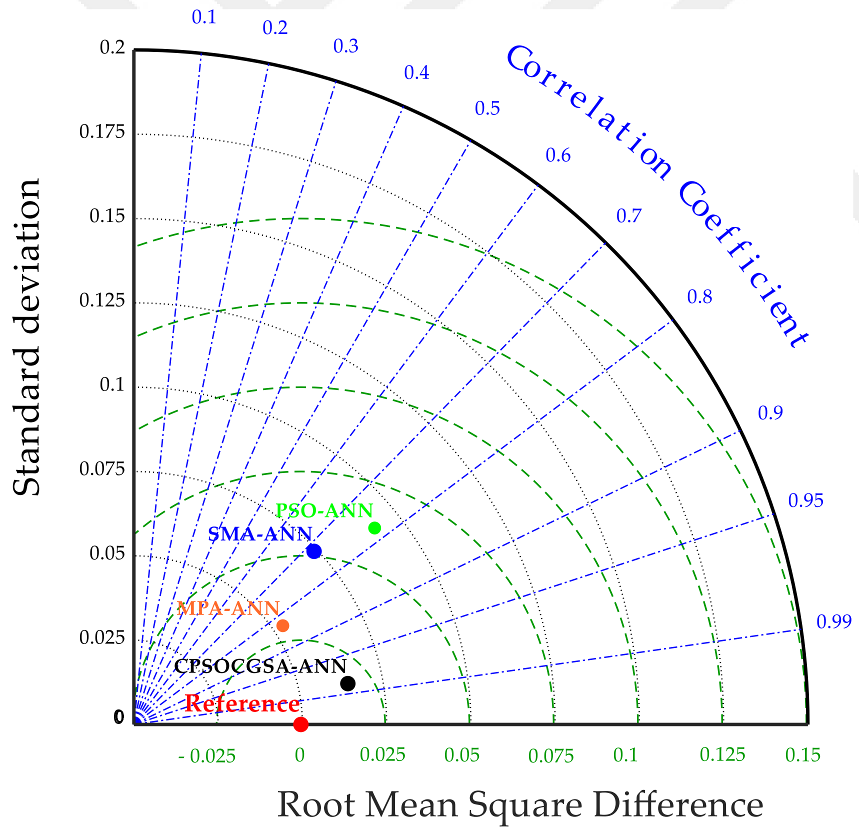


Figure 4.7. Taylor diagrams for different forecasting models

Moreover, to increase the reliability of the CPSOCGSA-ANN strategy, the Kolmogorov_Smirnov (K_S) and Shapiro_Wilk (S_W) tests were adopted to test the normality of the error data. The results showed that the (p-values) of K_S and S_W tests were more than 0.05, indicating that the errors have a normal distribution according to Valentini Marlon et al. (2021) (see Table 4.5).

Table 4.5. Tests of Normality

Target	Kolmogorov-Smirnova (K_S)	Shapiro-Wilk (S_W)
WQI	0.200	0.224

Based on the above results, it is possible to conclude that:

- The assessment result of Yesilrmak river varied from good to poor, except for one point (June in 2000) that was unsuitable for drinking.
- The SSA and tolerance approach has high benefits for enhancing data quality and picking the optimized scenario for input parameters.
- Comparing the performance accuracy of the hybrid models, including CPSOCGSA-ANN, PSO-ANN,SMA-ANN and MPA-ANN, showed that the CPSOCGSA-ANN was a reliable and superior model to predict WQI data.

5. CONCLUSION AND RECOMMENDATIONS

The present research has assessed the WQ during the monitoring period (1995-2014) for the Yesilrmak river within the provincial borders of Çorum, Turkey. To this end, eleven physicochemical parameters were assigned to calculate the WQI. The result varied from good to poor, except for one point (June 2000) that was unsuitable for drinking, as shown in Table 1. Additionally, this study employed novel hybridization models combining pre-processing data and ANN approach optimized by multiple metaheuristic methods (i.e., CPSOCGSA, MPA,SMA and PSO) to predict the WQI. Also, to raise the accuracy of models, the implications of climate factors cannot be neglected; hence, adopting rainfall beside the WQ parameters as predictors. The results for the models presented two crucial aspects. The first aspect confirms that pre-processing methods (i.e., SSA and tolerance) enhanced the dataset quality and yielded appropriate scenario of predictors. The second aspect, the performance of CPSOCGSA-ANN, was superior to other models (MPA-ANN,SMA-ANN and PSO-ANN) that yielded $R^2 = 0.965$, MAE = 0.01627, and RMSE = 0.0187.

For future research, it is recommended to use these metaheuristic algorithms to integrate other machine learning models. Also, assess the impact of additional predictors such as an expansion of urbanization, the volume of wastewater discharges, and water consumption.

REFERENCES

Abba, S., Hadi, S. J., Sammen, S. S., Salih, S. Q., Abdulkadir, R., Pham, Q. B., & Yaseen, Z. M. (2020). Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *Journal of Hydrology, 587*, 124974.

Abbasi, T., & Abbasi, S. A. (2012). *Water quality indices*: Elsevier.

Abdel-Basset, M., Abdel-Fatah, L., & Sangaiah, A. K. (2018). Metaheuristic algorithms: A comprehensive review. *Computational intelligence for multimedia big data on the cloud with engineering applications*, 185-231.

Abdul Kareem, B., Zubaidi, S. L., Ridha, H. M., Al-Ansari, N., & Al-Bdairi, N. S. S. (2022). Applicability of ANN Model and CPSOCGSA Algorithm for Multi-Time Step Ahead River Streamflow Forecasting. *Hydrology, 9*(10), 171.

Adnan, R. M., Mostafa, R. R., Islam, A. R. M. T., Gorgij, A. D., Kuriqi, A., & Kisi, O. (2021). Improving Drought Modeling Using Hybrid Random Vector Functional Link Methods. *Water, 13*(23), 3379. doi:10.3390/w13233379

Aghelpour, P., Mohammadi, B., Mehdizadeh, S., Bahrami-Pichaghchi, H., & Duan, Z. (2021). A novel hybrid dragonfly optimization algorithm for agricultural drought prediction. *Stochastic Environmental Research and Risk Assessment, 35*(12), 2459-2477. doi:10.1007/s00477-021-02011-2

Al-Qaness, M. A., Ewees, A. A., Fan, H., Abualigah, L., & Abd Elaziz, M. (2020). Marine predators algorithm for forecasting confirmed cases of COVID-19 in Italy, USA, Iran and Korea. *International journal of environmental research and public health, 17*(10), 3520.

Alawsi, M. A., Zubaidi, S. L., Al-Ansari, N., Al-Bugharbee, H., & Ridha, H. M. (2022). Tuning ANN Hyperparameters by CPSOCGSA, MPA, and SMA for Short-Term SPI Drought Forecasting. *Atmosphere, 13*(9), 1436.

Aldhyani, T. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water quality prediction using artificial intelligence algorithms. *Applied Bionics and Biomechanics, 2020*.

Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water Quality Prediction Using Artificial Intelligence Algorithms. *Appl Bionics Biomech, 2020*, 6659314. doi:10.1155/2020/6659314

Ali, M., Deo, R. C., Maraseni, T., & Downs, N. J. (2019). Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *Journal of Hydrology, 576*, 164-184. doi:10.1016/j.jhydrol.2019.06.032

Araghinejad, S. (2013). *Data-driven modeling: using MATLAB® in water resources and environmental engineering* (Vol. 67): Springer Science & Business Media.

Asadollah, S. B. H. S., Sharafati, A., Motta, D., & Yaseen, Z. M. (2021). River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering, 9*(1). doi:10.1016/j.jece.2020.104599

Başakın, E. E., Ekmekcioğlu, Ö., & Özger, M. (2020). Drought prediction using hybrid soft-computing methods for semi-arid region. *Modeling Earth Systems and Environment, 7*(4), 2363-2371. doi:10.1007/s40808-020-01010-6

Başörens, Ö. E., & Kazanci, N. (2012). Clustering of Simuliidae (Insecta, Diptera) species and sampling sites according to their similarities by using the UPGMA method. *Review of Hydrobiology, 5*(1).

Belayneh, A., Adamowski, J., & Khalil, B. (2015). Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. *Sustainable Water Resources Management, 2*(1), 87-101. doi:10.1007/s40899-015-0040-5

Bi, W., Weng, B., Yuan, Z., Ye, M., Zhang, C., Zhao, Y., . . . Xu, T. (2018). Evolution characteristics of surface water quality due to climate change and LUCC under scenario simulations: a case study in the Luanhe River Basin. *International journal of environmental research and public health, 15*(8), 1724.

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment, 721*, 137612.

Calì, D., Osterhage, T., Streblow, R., & Müller, D. (2016). Energy performance gap in refurbished German dwellings: Lesson learned from a field test. *Energy and buildings, 127,* 1146-1158.

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

Colombo, N., Sessa, C., du Bois, A., Ledermann, J., McCluggage, W. G., McNeish, I., . . . Vergote, I. (2019). ESMO–ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Annals of Oncology, 30*(5), 672-705.

Cude, C. G. (2001). Oregon water quality index a tool for evaluating water quality management effectiveness 1. *JAWRA Journal of the American Water Resources Association, 37*(1), 125-137.

Dadolahi-Sohrab, A., Arjomand, F., & Fadaei-Nasab, M. (2012). Water quality index as a simple indicator of watersheds pollution in southwestern part of I ran. *Water and Environment Journal, 26*(4), 445-454.

Danandeh Mehr, A., Vaheddoost, B., & Mohammadi, B. (2020). ENN-SA: A novel neuro-annealing model for multi-station drought prediction. *Computers & Geosciences, 145*, 104622. doi:10.1016/j.cageo.2020.104622

Das Kangabam, R., Bhoominathan, S. D., Kanagaraj, S., & Govindaraju, M. (2017). Development of a water quality index (WQI) for the Loktak Lake in India. *Applied Water Science, 7*(6), 2907-2918.

Das, P., Naganna, S. R., Deka, P. C., & Pushparaj, J. (2020). Hybrid wavelet packet machine learning approaches for drought modeling. *Environmental Earth Sciences, 79*(10), 221. doi:10.1007/s12665-020-08971-y

Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software, 22*(7), 1034-1052.

Debels, P., Figueroa, R., Urrutia, R., Barra, R., & Niell, X. (2005). Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified water quality index. *Environmental monitoring and assessment, 110*(1), 301-322.

Djerbouai, S., & Souag-Gamane, D. (2016). Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Stochastic Models: Case of the Algerois Basin in North Algeria. *Water Resources Management, 30*(7), 2445-2464. doi:10.1007/s11269-016-1298-6

Ethaib, S., Omar, R., Mazlina, M. K. S., Radiah, A. B. D., & Syafiie, S. (2016). Microwave-assisted dilute acid pretreatment and enzymatic hydrolysis of sago palm bark. *BioResources, 11*(3), 5687-5702.

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 & Innovation, 11*, 390-398.

Faramarzi, A., Heidarinejad, M., Mirjalili, S., & Gandomi, A. H. (2020). Marine Predators Algorithm: A nature-inspired metaheuristic. *Expert systems with applications, 152*, 113377.

Fogden, J., & Wood, G. (2009). Access to safe drinking water and its impact on global economic growth. *HaloSource Inc*.

Fung, K. F., Huang, Y. F., & Koo, C. H. (2019). Coupling fuzzy–SVR and boosting–SVR models with wavelet decomposition for meteorological drought prediction. *Environmental Earth Sciences, 78*(24), 693. doi:10.1007/s12665-019-8700-7

Ghorbani, M. A., Deo, R. C., Karimi, V., Yaseen, Z. M., & Terzi, O. (2018). Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stochastic Environmental Research and Risk Assessment, 32*(6), 1683-1697.

Golyandina, N., Korobeynikov, A., & Zhigljavsky, A. (2018). *Singular spectrum analysis with R*: Springer.

Guermoui, M., Melgani, F., Gairaa, K., & Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production, 258*, 120357.

Gupta, S., & Gupta, S. K. (2021). A critical review on water quality index tool: Genesis, evolution and future directions. *Ecological Informatics, 63*. doi:10.1016/j.ecoinf.2021.101299

Hajirahimi, Z., & Khashei, M. (2022a). Hybridization of hybrid structures for time series forecasting: a review. *Artificial Intelligence Review*. doi:10.1007/s10462-022-10199-0

Hajirahimi, Z., & Khashei, M. (2022b). Hybridization of hybrid structures for time series forecasting: A review. *Artificial Intelligence Review*, 1-61.

Hameed, M., Sharqi, S. S., Yaseen, Z. M., Afan, H. A., Hussain, A., & Elshafie, A. (2016). Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications, 28*(S1), 893-905. doi:10.1007/s00521-016-2404-7

Hammond, D., & Pryce, A. (2007). Climate change impacts and water temperature. *Climate change impacts and water temperature.*

Hassan, M. H., Kamel, S., Abualigah, L., & Eid, A. (2021). Development and application of slime mould algorithm for optimal economic emission dispatch. *Expert systems with applications, 182*, 115205.

Horton, R. K. (1965). An index number system for rating water quality. *J Water Pollut Control Fed, 37*(3), 300-306.

Judran, N. H., & Kumar, A. (2020). Evaluation of water quality of Al-Gharraf River using the water quality index (WQI). *Modeling Earth Systems and Environment, 6*(3), 1581-1588. doi:10.1007/s40808-020-00775-0

Kadam, A., Wagh, V., Muley, A., Umrikar, B., & Sankhua, R. (2019). Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment, 5*(3), 951-962.

Kannel, P. R., Lee, S., Lee, Y.-S., Kanel, S. R., & Khan, S. P. (2007). Application of Water Quality Indices and Dissolved Oxygen as Indicators for River Water Classification and Urban Impact Assessment. *Environmental monitoring and assessment, 132*(1), 93-110. doi:10.1007/s10661-006-9505-1

Karami, F., & Dariane, A. B. (2022). Melody Search Algorithm Using Online Evolving Artificial Neural Network Coupled with Singular Spectrum Analysis for Multireservoir System Management. *Iranian Journal of Science and Technology, Transactions of Civil Engineering, 46*(2), 1445-1457.

Khan, M. M. H., Muhammad, N. S., & El-Shafie, A. (2020). Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *Journal of Hydrology, 590*, 125380. doi:10.1016/j.jhydrol.2020.125380

Khudhair, Z. S., Zubaidi, S. L., Al-Bugharbee, H., Al-Ansari, N., & Ridha, H. M. (2022). A CPSOCGSA-tuned neural processor for forecasting river water salinity: Euphrates river, Iraq. *Cogent Engineering, 9*(1), 2150121.

Kisi, O., Docheshmeh Gorgij, A., Zounemat-Kermani, M., Mahdavi-Meymand, A., & Kim, S. (2019). Drought forecasting using novel heuristic methods in a semi-arid environment. *Journal of Hydrology, 578*, 124053. doi:10.1016/j.jhydrol.2019.124053

Koranga, M., Pant, P., Kumar, T., Pant, D., Bhatt, A. K., & Pant, R. P. (2022). Efficient water quality prediction models based on machine learning algorithms for Nainital Lake, Uttarakhand. *Materials Today: Proceedings, 57*, 1706-1712. doi:10.1016/j.matpr.2021.12.334

Kossieris, P., & Makropoulos, C. (2018). Exploring the statistical and distributional properties of residential water demand at fine time scales. *Water, 10*(10), 1481.

Kouadri, S., Pande, C. B., Panneerselvam, B., Moharir, K. N., & Elbeltagi, A. (2022). Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. *Environmental Science and Pollution Research, 29*(14), 21067-21091.

Kulisz, M., Kujawska, J., Przysucha, B., & Cel, W. (2021). Forecasting water quality index in groundwater using artificial neural network. *Energies, 14*(18), 5875.

Kumar, C., Raj, T. D., Premkumar, M., & Raj, T. D. (2020). A new stochastic slime mould optimization algorithm for the estimation of solar photovoltaic cell parameters. *Optik, 223*, 165277.

Kummu, M., Ward, P. J., de Moel, H., & Varis, O. (2010). Is physical water scarcity a new phenomenon? Global assessment of water shortage over the last two millennia. *Environmental Research Letters, 5*(3), 034006.

Le, H. D., Le, T. N., Wang, J. W., & Liang, Y. S. (2021). Singular Spectrum Analysis for Background Initialization with Spatio-Temporal RGB Color Channel Data. *Entropy (Basel), 23*(12), 1644. doi:10.3390/e23121644

Le, J. A., El-Askary, H. M., Allali, M., Sayed, E., Sweliem, H., Piechota, T. C., & Struppa, D. C. (2020). Characterizing El Niño-Southern Oscillation Effects on the Blue Nile Yield and the Nile River Basin Precipitation using Empirical Mode Decomposition. *Earth Systems and Environment, 4*(4), 699-711. doi:10.1007/s41748-020-00192-4

Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer Systems, 111*, 300-323. doi:10.1016/j.future.2020.03.055

Liu, R., Ding, H., Zhang, Z., Yang, C., & Zhou, G. (2020). Study on prediction model of diesel engine with regulated two-stage turbocharging system based on hybrid genetic algorithm-particle swarm optimization method at different altitudes. *Energy sources, part A: Recovery, utilization, and environmental effects*, 1-15.

Lobato, T., Hauser-Davis, R., Oliveira, T., Silveira, A., Silva, H., Tavares, M., & Saraiva, A. (2015). Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. *Journal of Hydrology, 522*, 674-683.

Ma, Z., Li, H., Ye, Z., Wen, J., Hu, Y., & Liu, Y. (2020). Application of modified water quality index (WQI) in the assessment of coastal water quality in main aquaculture areas of Dalian, China. *Marine Pollution Bulletin, 157*, 111285.

Magesh, N. S., Krishnakumar, S., Chandrasekar, N., & Soundranayagam, J. P. (2013). Groundwater quality assessment using WQI and GIS techniques, Dindigul district, Tamil Nadu, India. *Arabian Journal of Geosciences, 6*(11), 4179-4189. doi:10.1007/s12517-012-0673-8

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software, 15*, 101–124.

Malmaeus, J. M., Blenckner, T., Markensten, H., & Persson, I. (2006). Lake phosphorus dynamics and climate warming: A mechanistic model approach. *Ecological Modelling, 190*(1-2), 1-14.

Mehta, M., Tewari, D., Gupta, G., Awasthi, R., Singh, H., Pandey, P., . . . Hansbro, P. M. (2019). Oligonucleotide therapy: An emerging focus area for drug delivery in chronic inflammatory respiratory diseases. *Chemico-biological interactions, 308*, 206-215.

Michalak, A. M. (2016). Study role of climate change in extreme threats to water quality. *Nature, 535*(7612), 349-350. doi:10.1038/535349a

Modaresi, F., & Araghinejad, S. (2014). A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. *Water Resources Management, 28*(12), 4095-4111.

Mohammadi, B., Linh, N. T. T., Pham, Q. B., Ahmed, A. N., Vojteková, J., Guan, Y., . . . El-Shafie, A. (2020). Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series. *Hydrological Sciences Journal, 65*(10), 1738-1751. doi:10.1080/02626667.2020.1758703

Mohammadi, B., & Mehdizadeh, S. (2020). Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agricultural Water Management, 237*, 106145. doi:10.1016/j.agwat.2020.106145

Mohammed, S. J., Zubaidi, S. L., Al-Ansari, N., Ridha, H. M., & Al-Bdairi, N. S. S. (2022). Hybrid Technique to Improve the River Water Level Forecasting Using Artificial Neural Network-Based Marine Predators Algorithm. *Advances in Civil Engineering, 2022*.

Nabipour, N., Dehghani, M., Mosavi, A., & Shamshirband, S. (2020). Short-Term Hydrological Drought Forecasting Based on Different Nature-Inspired Optimization Algorithms Hybridized With Artificial Neural Networks. *IEEE Access, 8*, 15210-15222. doi:10.1109/access.2020.2964584

Najah, A., El-Shafie, A., Karim, O. A., & El-Shafie, A. H. (2013). Application of artificial neural networks for water quality prediction. *Neural Computing and Applications, 22*(1), 187-201.

Nakagaki, T., Yamada, H., & Ueda, T. (2000). Interaction between cell shape and contraction pattern in the Physarum plasmodium. *Biophysical chemistry, 84*(3), 195-204.

Nourani, V., Molajou, A., Uzelaltinbulat, S., & Sadikoglu, F. (2019). Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: northern Cyprus. *Theoretical and Applied Climatology, 138*(3-4), 1419-1434. doi:10.1007/s00704-019-02904-x

Organization, W. H., WHO., & Staff, W. H. O. (2004). *Guidelines for drinking-water quality* (Vol. 1): World Health Organization.

Ouyang, Q., & Lu, W. (2017). Monthly Rainfall Forecasting Using Echo State Networks Coupled with Data Preprocessing Methods. *Water Resources Management, 32*(2), 659-674. doi:10.1007/s11269-017-1832-1

Ouyang, Q., & Lu, W. (2018). Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. *Water Resources Management, 32*(2), 659-674.

Özden, S., & Ayan, S. (2016). Forest crimes as a threat to sustainable forest management. *Sibirskij Lesnoj Zurnal\Siberian Journal of Forest Science, 4*, 49-55.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*: Routledge.

Patki, V. K., Jahagirdar, S., Patil, Y., Karale, R., & Nadagouda, A. (2021). Prediction of water quality in municipal distribution system. *Materials Today: Proceedings*.

Payal, A., Rai, C. S., & Reddy, B. R. (2015). Analysis of some feedforward artificial neural network training algorithms for developing localization framework in wireless sensor networks. *Wireless Personal Communications, 82*(4), 2519-2536.

Pham, Q. B., Yang, T.-C., Kuo, C.-M., Tseng, H.-W., & Yu, P.-S. (2021). Coupling singular spectrum analysis with least square support vector machine to improve accuracy of SPI drought forecasting. *Water Resources Management, 35*(3), 847-868.

Polomčić, D., Gligorić, Z., Bajić, D., & Cvijović, Č. (2017). A hybrid model for forecasting groundwater levels based on fuzzy C-mean clustering and singular spectrum analysis. *Water, 9*(7), 541.

Poonam, T., Tanushree, B., & Sukalyan, C. (2013). Water quality indices-important tools for water quality assessment: a review. *International Journal of Advances in chemistry, 1*(1), 15-28.

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.

Rajaee, T., & Boroumand, A. (2015). Forecasting of chlorophyll-a concentrations in South San Francisco Bay using five different models. *Applied Ocean Research, 53*, 208-217.

Rajaee, T., Khani, S., & Ravansalar, M. (2020). Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems, 200*, 103978.

Ramakrishnaiah, C., Sadashivaiah, C., & Ranganna, G. (2009). Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. *E-Journal of chemistry, 6*(2), 523-530.

Rather, S. A., & Bala, P. S. (2019a). *Hybridization of constriction coefficient-based particle swarm optimization and chaotic gravitational search algorithm for solving engineering design problems.* Paper presented at the International conference on advanced communication and networking.

Rather, S. A., & Bala, P. S. (2019b). *Hybridization of constriction coefficient based particle swarm optimization and gravitational search algorithm for function optimization.* Paper presented at the Proceedings of the International Conference on Advances in Electronics, Electrical & Computational Intelligence (ICAEEC).

Reddy, P. C. S., Yadala, S., & Goddumarri, S. N. (2022). Development of rainfall forecasting model using machine learning with singular spectrum analysis. *IIUM Engineering Journal, 23*(1), 172-186.

Ruiz-Aguilar, J. J., Urda, D., Moscoso-López, J. A., González-Enrique, J., & Turias, I. J. (2020). *Container Demand Forecasting at Border Posts of Ports: A Hybrid SARIMA-SOM-SVR Approach.* Paper presented at the International Conference on Optimization and Learning.

Safavi, H. R., Sharghi, S., & Komasi, M. (2018). Wavelet and cuckoo search-support vector machine conjugation for drought forecasting using Standardized Precipitation Index (case study: Urmia Lake, Iran). *Journal of Hydroinformatics, 20*(4), 975-988. doi:10.2166/hydro.2018.115

Saffran, K., Cash, K., Hallard, K., & WRIGHT, R. (2001). Canadian water quality guidelines for the protection of aquatic life, CCME water quality Index 1, 0, Users manual. *Excerpt from Publication, 1299*.

Sakaa, B., Brahmia, N., Chaffai, H., & Hani, A. (2020). Assessment of water quality index in unmonitored river basin using multilayer perceptron neural networks and principal component analysis. *Desalination and Water Treatment, 200*, 42-54. doi:10.5004/dwt.2020.26108

Schlosser, C. A., Strzepek, K., Gao, X., Fant, C., Blanc, É., Paltsev, S., . . . Gueneau, A. (2014). The future of global water stress: An integrated assessment. *Earth's Future, 2*(8), 341-361.

Şener, Ş., Şener, E., & Davraz, A. (2017). Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Science of the Total Environment, 584*, 131-144.

Shanley, K. (2017). Climate Change and Water Quality: Keeping a Finger on the Pulse. *Am J Public Health, 107*(1), e10. doi:10.2105/ajph.2016.303504

Sharma, P., Meher, P. K., Kumar, A., Gautam, Y. P., & Mishra, K. P. (2014). Changes in water quality index of Ganges river at different locations in Allahabad. *Sustainability of Water Quality and Ecology, 3*, 67-76.

Singh, B., Sihag, P., Singh, V. P., Sepahvand, A., & Singh, K. (2021). Soft computing technique-based prediction of water quality index. *Water Supply, 21*(8), 4015-4029.

Sofos, F., & Karakasidis, T. E. (2021). Nanoscale slip length prediction with machine learning tools. *Sci Rep, 11*(1), 12520. doi:10.1038/s41598-021-91885-x

Soh, Y. W., Koo, C. H., Huang, Y. F., & Fung, K. F. (2018). Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia. *Computers and Electronics in Agriculture, 144*, 164-173. doi:10.1016/j.compag.2017.12.002

Song, C., Yao, L., Hua, C., & Ni, Q. (2021). A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. *Journal of Hydrology, 603*, 126879.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5): pearson Boston, MA.

Tao, H., Al-Bedyry, N. K., Khedher, K. M., Shahid, S., & Yaseen, Z. M. (2021). River water level prediction in coastal catchment using hybridized relevance vector machine model with improved grasshopper optimization. *Journal of Hydrology, 598*, 126477.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres, 106*(D7), 7183-7192. doi:10.1029/2000jd900719

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.

Unnikrishnan, P., & Jothiprakash, V. (2018). Daily rainfall forecasting for one year in a single run using Singular Spectrum Analysis. *Journal of Hydrology, 561*, 609-621. doi:10.1016/j.jhydrol.2018.04.032

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*. doi:10.1016/j.ecolind.2019.105815

Valentini, M., dos Santos, G. B., & Muller Vieira, B. (2021). Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul—Brazil. *SN Applied Sciences, 3*(1), 70. doi:10.1007/s42452-020-04005-1

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., . . . Atencia, A. (2021). Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society, 102*(3), E681-E699.

Wang, K., & Dickinson, R. E. (2013). Contribution of solar radiation to decadal temperature variability over land. *Proceedings of the National Academy of Sciences, 110*(37), 14877-14882.

Wang, M., Chen, H., Yang, B., Zhao, X., Hu, L., Cai, Z., . . . Tong, C. (2017). Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses. *Neurocomputing, 267*, 69-84.

Wang, P., Casner, R. G., Nair, M. S., Wang, M., Yu, J., Cerutti, G., . . . Shapiro, L. (2021). Increased resistance of SARS-CoV-2 variant P. 1 to antibody neutralization. *Cell host & microbe, 29*(5), 747-751. e744.

Wu, J., & Wang, Z. (2022). A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water, 14*(4), 610.

Wu, X., Zhou, J., Yu, H., Liu, D., Xie, K., Chen, Y., . . . Xing, F. (2021). The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis. *Atmosphere, 12*(1), 74. doi:10.3390/atmos12010074

Yaseen, Z. M., Ramal, M. M., Diop, L., Jaafar, O., Demir, V., & Kisi, O. (2018). Hybrid adaptive neuro-fuzzy models for water quality index estimation. *Water Resources Management, 32*(7), 2227-2245.

Yuan, X.-C., Sun, X., Lall, U., Mi, Z.-F., He, J., & Wei, Y.-M. (2016). China's socioeconomic risk from extreme events in a changing climate: a hierarchical Bayesian model. *Climatic Change, 139*, 169-181.

Zaqoot, H., Aish, A., & Abdeljawad, S. (2017). *Application of Artificial Neural networks for Predicting Water Quality*.

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., & Livingood, W. (2021). A review of machine learning in building load prediction. *Applied Energy, 285*, 116452.

Zhang, Y., Yang, H., Cui, H., & Chen, Q. (2019). Comparison of the Ability of ARIMA, WNN and SVM Models for Drought Forecasting in the Sanjiang Plain, China. *Natural Resources Research, 29*(2), 1447-1464. doi:10.1007/s11053-019-09512-6

Zhou, Y., Khu, S.-T., Xi, B., Su, J., Hao, F., Wu, J., & Huo, S. (2014). Status and challenges of water pollution problems in China: learning from the European experience. *Environmental Earth Sciences, 72*(4), 1243-1254. doi:10.1007/s12665-013-3042-3

Zhu, S., Lu, H., Ptak, M., Dai, J., & Ji, Q. (2020). Lake water-level fluctuation forecasting using machine learning models: a systematic review. *Environ Sci Pollut Res Int, 27*(36), 44807-44819. doi:10.1007/s11356-020-10917-7

Zubaidi, S., Al-Bdairi, N., Ortega Martorell, S., Ridha, H., Al-Bugharbee, H., Hashim, K., & Gharghan, S. Assessing the Benefits of Nature-Inspired Algorithms for the Parameterisation of ANN in the Prediction of Water Demand. *Journal of Water Resources Planning and Management*.

Zubaidi, S. L., Abdulkareem, I. H., Hashim, K. S., Al-Bugharbee, H., Ridha, H. M., Gharghan, S. K., . . . Al-Khaddar, R. (2020). Hybridised artificial neural network model with slime mould algorithm: a novel methodology for prediction of urban stochastic water demand. *Water, 12*(10), 2692.

Zubaidi, S. L., Dooley, J., Alkhaddar, R. M., Abdellatif, M., Al-Bugharbee, H., & Ortega-Martorell, S. (2018). A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. *Journal of Hydrology, 561*, 136-145. doi:10.1016/j.jhydrol.2018.03.047

Zubaidi, S. L., Gharghan, S. K., Dooley, J., Alkhaddar, R. M., & Abdellatif, M. (2018). Short-term urban water demand prediction considering weather factors. *Water resources management, 32*, 4527-4542.

Zubaidi, S. L., Gharghan, S. K., Dooley, J., Alkhaddar, R. M., & Abdellatif, M. (2018). Short-Term Urban Water Demand Prediction Considering Weather Factors. *Water Resources Management, 32*(14), 4527-4542. doi:10.1007/s11269-018-2061-y

Zubaidi, S. L., Hashim, K., Ethaib, S., Al-Bdairi, N. S. S., Al-Bugharbee, H., & Gharghan, S. K. (2022). A novel methodology to predict monthly municipal water demand based on weather variables scenario. *Journal of King Saud University - Engineering Sciences, 34*(3), 163-169. doi:10.1016/j.jksues.2020.09.011

Zubaidi, S. L., Ortega-Martorell, S., Al-Bugharbee, H., Olier, I., Hashim, K. S., Gharghan, S. K., . . . Al-Khaddar, R. (2020). Urban Water Demand Prediction for a City That Suffers from Climate Change and Population Growth: Gauteng Province Case Study. *Water, 12*(7), 1885. doi:10.3390/w12071885

APPENDIX

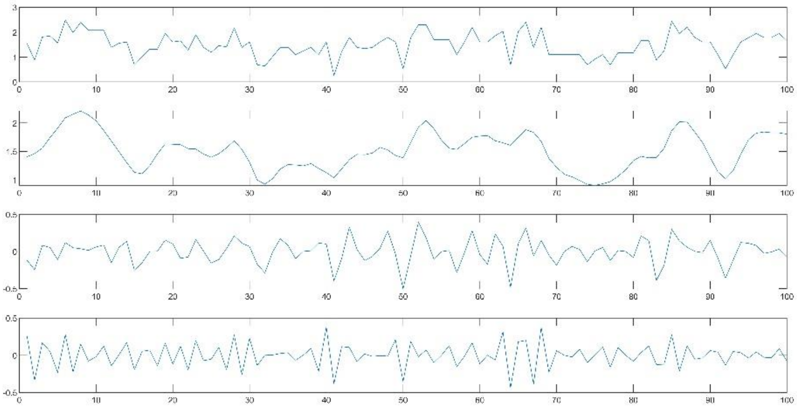


Figure 1. The BOD initial time series, the denoise time series and two noise signals

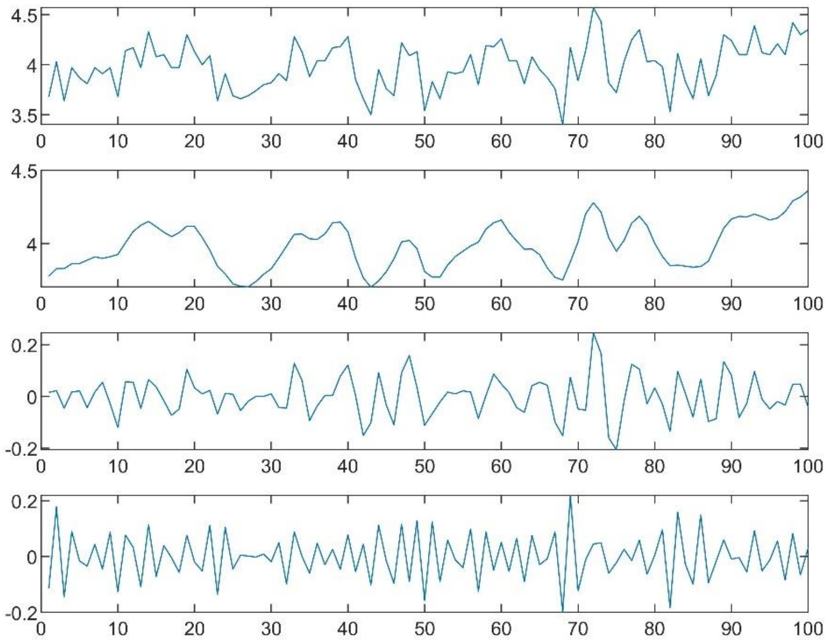


Figure 2. The Ca^{+2} initial time series, the denoise time series and two noise signals

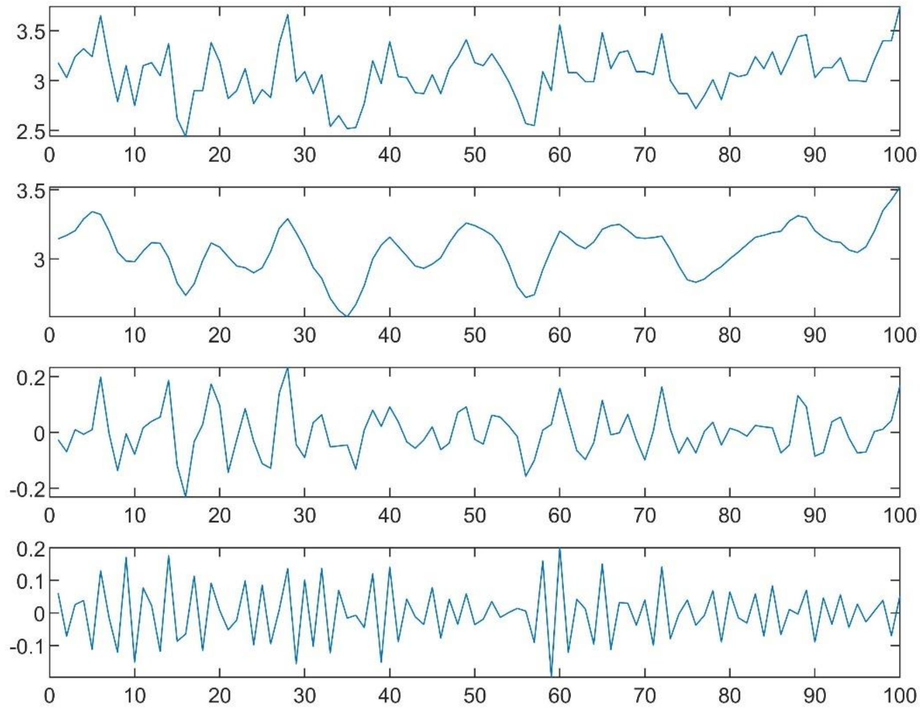


Figure 3. The Cl^{-1} initial time series, the denoise time series and two noise signals

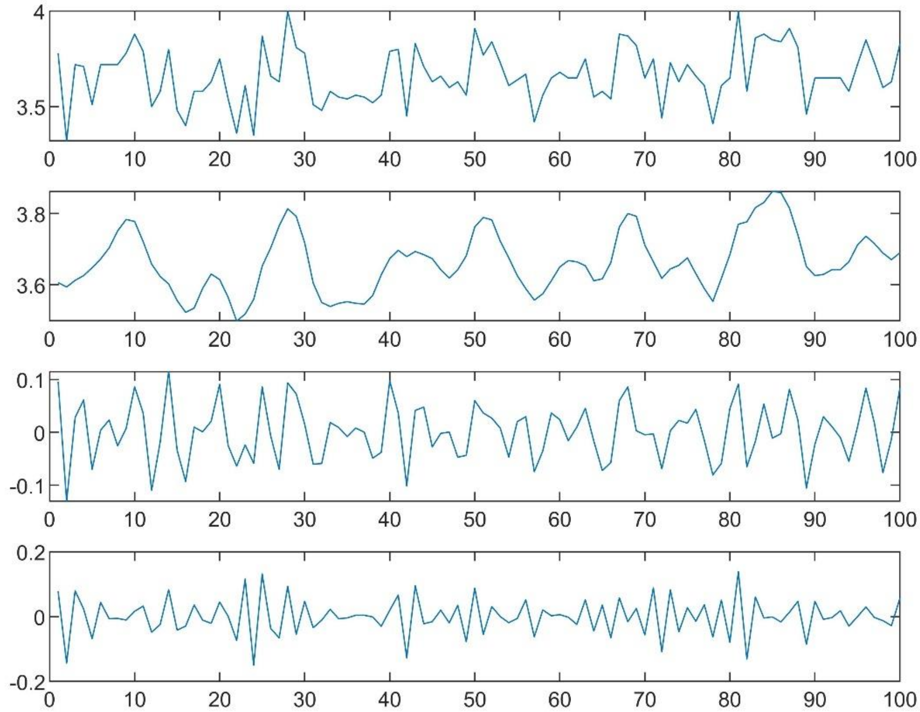


Figure 4. The mg^{+2} initial time series, the denoise time series and two noise signals

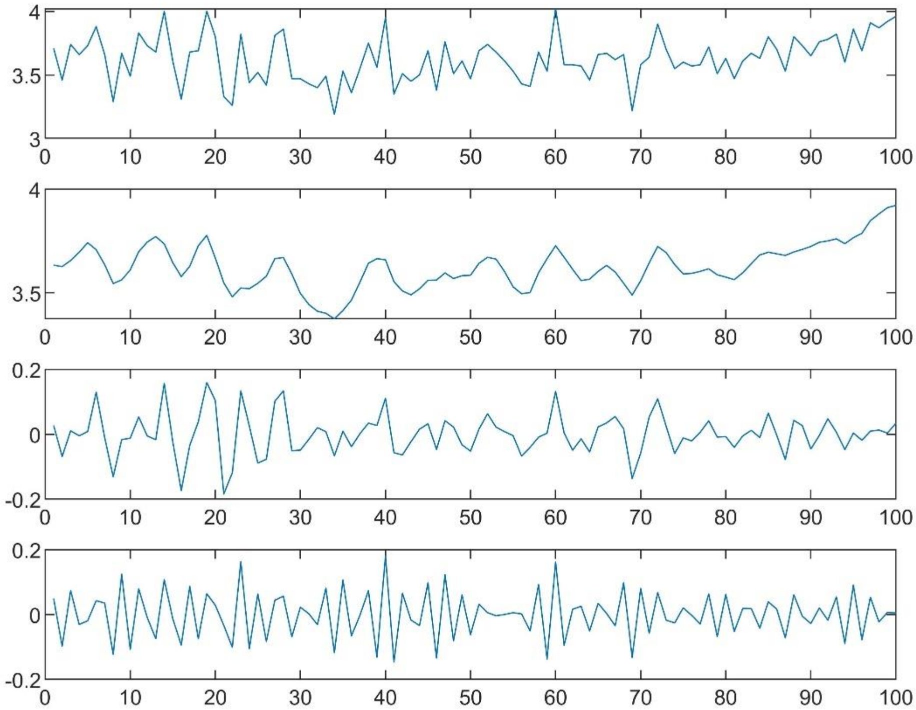


Figure 5. The Na^{+2} initial time series, the denoise time series and two noise signals

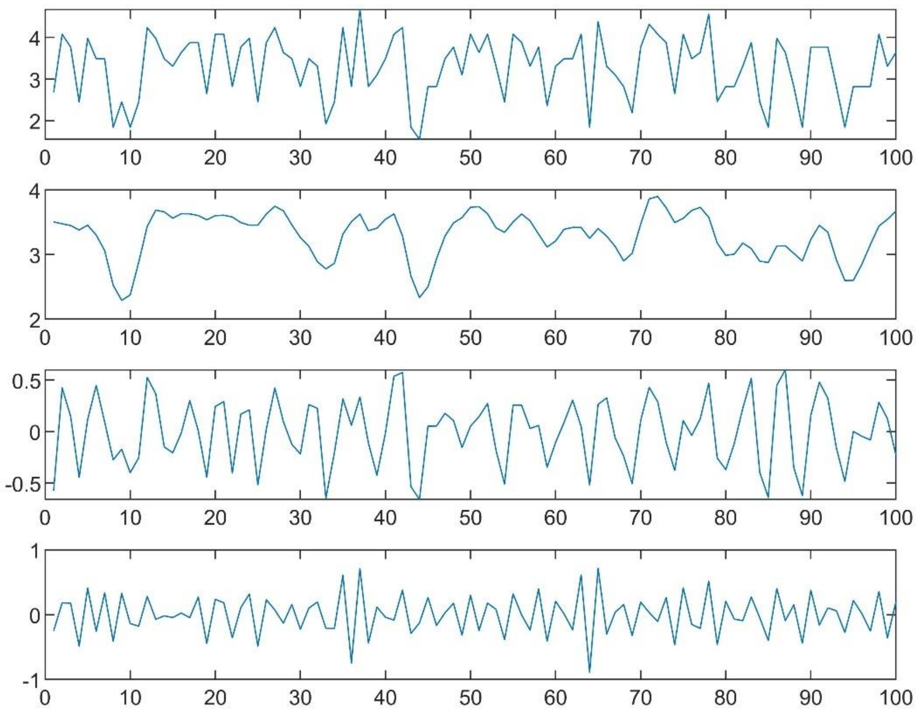


Figure 6. The rainfall initial time series, the denoise time series and two noise signal

Table 1. Water Quality Index calculation

Date	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl-1 (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
1995/2	46.540	3.472	11.624	7.590	0.390	0.742	7.718	7.487	1.049	0.137	2.384	3.949
1995/4	39.806	8.547	8.205	2.872	0.747	0.196	7.641	4.718	0.813	0.034	2.125	3.908
1995/6	55.609	10.150	7.521	10.462	0.079	0.359	8.705	10.598	1.085	0.137	2.051	4.462
1995/8	53.411	8.480	9.368	10.974	0.687	0.423	8.266	7.017	0.994	0.650	2.290	4.263
1995/11	46.185	10.150	7.521	7.795	0.600	0.359	7.090	5.709	1.069	0.103	2.158	3.631
996/2	65.277	4.674	7.521	22.769	0.546	0.774	10.590	7.043	1.815	0.034	4.094	5.415
1996/4	55.421	8.360	9.368	12.779	0.687	0.306	8.559	7.026	1.000	0.650	2.256	4.432
1996/6	65.114	13.889	8.889	20.513	0.636	0.056	8.026	7.060	0.333	0.000	1.600	4.113
1996/8	57.720	8.280	9.368	14.421	0.687	0.288	8.754	7.492	1.004	0.650	2.233	4.544
1996/11	47.818	11.485	9.573	0.000	0.453	0.033	8.564	8.291	0.503	2.735	1.793	4.390
1997/2	58.982	8.280	8.889	14.359	0.870	0.288	9.474	7.573	1.177	1.368	1.854	4.851
1997/4	46.656	8.948	9.573	6.154	0.902	0.310	8.744	3.863	1.069	0.137	2.200	4.757
1997/6	50.001	8.133	9.368	7.692	0.687	0.212	9.144	6.145	1.012	0.650	2.188	4.770
1997/8	59.892	8.013	11.624	8.205	1.097	0.473	11.538	7.624	1.398	0.786	2.888	6.246
1997/11	38.366	6.277	9.573	2.051	0.805	-0.222	7.692	5.521	0.522	0.684	1.268	4.195
1998/2	39.605	8.280	8.889	0.000	0.818	-0.101	8.013	5.128	0.705	2.017	1.497	4.359

Date	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl-1 (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
1998/4	48.232	7.973	9.368	5.641	0.687	0.119	9.535	6.103	1.021	0.650	2.143	4.995
1998/6	48.337	7.933	9.368	5.641	0.687	0.119	9.632	6.103	1.023	0.650	2.132	5.051
1998/8	58.407	6.544	10.940	12.308	1.060	0.484	10.769	6.479	1.407	0.068	2.954	5.395
1998/11	52.694	6.677	10.940	8.205	0.856	0.314	9.551	7.282	1.143	0.650	2.224	4.851
999/3	46.693	8.413	8.889	8.615	0.715	0.069	7.308	5.880	0.713	0.103	2.244	3.744
1999/4	38.685	7.746	6.838	5.333	0.809	0.114	7.090	4.906	0.667	0.034	1.518	3.631
1999/6	50.482	7.879	7.521	11.692	0.420	0.265	8.513	6.342	1.169	0.068	2.252	4.359
1999/8	39.232	8.948	6.838	6.154	0.636	-0.154	6.538	4.889	0.503	0.068	1.469	3.344
1999/11	41.576	6.811	6.838	4.718	0.456	0.124	7.526	8.222	0.867	0.000	2.158	3.856
2000/2	44.590	8.547	7.521	6.769	0.438	0.078	7.756	6.632	0.785	0.205	1.887	3.969
2000/4	42.422	7.212	5.470	6.359	0.456	0.474	8.923	6.444	1.162	0.000	1.358	4.564
2000/6	181.061	19.444	3.419	82.051	2.056	2.885	24.397	17.983	1.892	0.000	14.441	12.491
2000/8	56.782	14.557	12.308	6.154	0.539	0.173	8.359	7.726	0.823	0.205	1.657	4.280
2000/11	49.527	9.215	7.521	8.205	0.557	0.242	8.936	7.453	0.823	0.068	1.932	4.574
2001/2	41.097	9.615	8.205	2.051	0.632	0.101	7.718	5.744	0.795	0.137	1.899	4.199
2001/4	43.866	11.485	8.889	1.846	0.575	0.219	8.038	5.556	0.767	0.137	1.986	4.369
2001/6	46.186	12.153	10.940	0.000	1.034	-0.063	8.590	6.103	0.839	0.000	1.904	4.687
2001/8	44.236	4.541	11.282	6.154	0.853	-0.012	8.179	5.962	0.624	0.479	1.726	4.449
2001/11	48.521	10.817	8.889	6.154	0.608	-0.073	8.718	5.915	0.878	0.000	1.887	4.728

Date	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl-1 (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
200 2/4	44.173	9.615	8.205	4.103	0.759	-0.150	8.308	6.034	0.735	0.684	1.358	4.523
200 2/4	48.055	10.216	9.231	5.128	0.758	0.043	8.750	5.949	0.892	0.479	1.984	4.626
200 2/6	51.648	11.218	9.573	6.154	0.902	0.323	9.000	5.778	1.086	0.274	2.449	4.892
200 2/8	46.189	7.612	10.940	4.103	0.913	0.160	8.782	5.983	0.906	0.239	2.080	4.472
200 2/11	60.334	9.615	13.675	8.205	1.039	0.488	10.269	7.538	1.333	0.205	2.622	5.344
200 3/2	42.495	10.016	8.889	0.615	0.582	0.206	7.487	7.675	0.733	0.034	2.421	3.836
200 3/4	41.037	10.016	6.154	4.923	0.438	0.196	7.654	5.385	0.862	0.171	1.321	3.918
200 3/6	49.495	8.814	7.521	10.256	0.331	0.105	7.692	7.897	0.805	0.034	2.101	3.938
200 3/8	47.506	10.417	7.521	6.154	0.672	0.101	8.128	6.957	0.846	0.137	2.408	4.164
200 3/11	45.758	10.016	7.521	5.744	0.510	0.219	8.026	6.444	1.023	0.205	1.936	4.113
200 4/2	42.523	9.215	6.154	6.154	0.456	0.101	7.846	6.632	0.751	0.068	1.124	4.021
200 4/4	52.981	13.355	5.470	8.205	0.960	0.265	9.731	6.239	1.105	0.000	2.667	4.985
200 4/6	56.818	15.358	7.521	10.256	0.816	0.359	8.679	6.444	0.862	0.137	1.941	4.444
200 4/8	55.003	15.358	6.838	8.205	0.852	1.026	9.179	6.034	0.951	0.034	1.826	4.699
200 4/11	43.596	12.153	5.470	1.436	0.356	0.310	8.269	8.530	0.823	0.274	1.739	4.236
200 5/2	54.233	8.814	10.940	10.256	0.564	0.288	8.385	7.402	1.026	0.034	1.965	4.559
200 5/4	84.282	9.081	13.675	36.923	0.438	0.382	8.051	7.966	1.074	0.034	2.277	4.379
200 5/6	64.411	8.347	12.923	18.462	0.657	0.278	8.705	7.147	1.013	0.034	2.110	4.735
200 5/8	44.263	7.612	12.171	0.000	0.633	0.173	9.359	6.328	0.952	0.000	1.944	5.091

Date	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl-1 (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
2005/11	43.318	9.615	9.846	0.000	0.657	0.061	9.083	6.506	0.872	0.000	1.737	4.941
2006/2	42.518	11.619	7.521	0.000	0.826	-0.051	8.808	6.684	0.791	0.000	1.530	4.790
2006/4	37.411	10.150	10.940	4.103	0.544	-0.435	5.115	3.573	0.261	0.000	0.689	2.472
2006/6	53.321	9.215	9.573	8.205	0.924	0.242	10.744	5.983	1.021	0.103	2.400	4.913
2006/8	58.244	10.283	8.205	16.410	0.910	0.117	8.628	6.564	0.877	0.137	1.764	4.349
2006/11	60.060	11.619	10.940	8.205	1.006	0.669	10.667	6.786	1.429	0.615	2.708	5.415
2007/2	49.917	10.337	6.154	8.205	0.758	0.235	9.195	6.568	0.919	0.068	2.216	5.262
2007/4	51.667	10.337	6.222	11.077	0.758	0.235	9.187	6.568	0.922	0.000	2.207	4.154
2007/6	55.123	9.348	6.838	13.744	0.546	0.173	8.731	7.282	0.908	0.068	2.347	5.138
2007/8	47.180	11.351	8.889	2.051	0.798	0.173	10.244	5.932	0.815	0.000	2.219	4.708
2007/11	53.926	8.948	7.521	13.538	0.672	0.588	9.038	6.120	1.000	0.342	2.219	3.938
2008/6	59.805	10.283	6.838	20.513	0.600	0.265	7.731	5.915	1.003	0.068	1.760	4.831
2008/4	47.771	9.482	6.154	6.154	0.510	0.392	8.987	8.308	0.962	0.000	2.166	4.656
2008/6	63.171	13.221	8.205	16.410	0.277	0.405	8.372	8.205	0.995	0.068	1.977	5.035
2008/8	46.816	10.150	6.154	4.103	0.906	0.242	9.513	7.795	0.641	0.068	2.059	5.186
2008/11	40.324	11.752	-0.684	4.103	0.575	0.242	9.551	6.564	0.915	0.034	1.969	5.303
2009/2	46.933	10.016	6.838	4.103	0.862	0.219	9.308	7.282	0.977	0.034	2.211	5.084
2009/4	46.236	10.817	6.838	4.103	1.473	0.569	10.205	3.009	1.264	0.068	2.757	5.133
2009/6	48.708	10.417	6.838	4.103	1.247	0.178	9.744	7.145	1.042	0.342	2.470	5.183

Date	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl-1 (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
2009/8	43.563	9.348	7.111	2.051	0.561	0.099	9.641	6.474	0.890	0.000	2.081	5.308
2009/11	45.277	8.146	9.573	3.077	0.234	0.101	8.846	7.077	0.938	0.000	2.002	5.283
2010/2	42.883	10.150	4.786	4.103	0.750	0.019	8.885	6.667	0.909	0.000	1.846	4.769
2010/4	43.816	9.882	6.838	2.051	0.993	0.086	9.064	6.342	0.922	0.068	2.187	5.383
2010/6	43.052	9.749	8.889	0.000	1.132	0.184	9.295	5.197	1.056	0.034	1.957	5.559
2010/8	45.537	12.153	10.256	0.000	0.747	0.067	7.897	6.325	0.860	0.137	1.936	5.159
2010/11	43.746	10.283	7.658	0.000	0.758	0.235	9.054	6.568	0.972	0.068	2.048	6.103
2011/2	53.593	9.482	7.248	8.697	0.697	0.206	10.282	9.333	0.526	0.103	2.363	4.656
2011/4	49.470	10.150	7.521	8.697	0.349	0.219	8.103	6.137	0.946	0.137	2.539	4.672
2011/6	47.103	9.348	7.521	2.872	0.834	0.359	10.038	8.103	1.010	0.137	2.408	4.472
2011/8	52.080	9.749	10.256	5.128	0.564	0.265	9.205	8.308	0.972	0.000	2.392	5.241
2011/11	63.274	8.013	7.521	21.538	0.438	0.402	9.128	8.017	1.149	0.068	2.322	4.677
2012/2	49.763	3.606	8.205	12.308	0.780	0.219	9.449	7.983	1.041	0.068	2.146	3.959
2012/4	58.185	9.215	6.838	16.410	0.456	0.359	9.103	8.513	0.564	0.034	2.088	4.605
2012/6	53.183	8.681	8.205	10.256	0.618	0.542	9.833	7.692	1.149	0.000	1.920	4.286
2012/8	52.422	9.215	8.205	8.205	1.057	0.565	10.128	5.419	1.067	1.368	2.322	4.872
2012/11	50.843	7.212	9.573	8.205	0.981	0.196	10.346	6.547	0.985	0.068	1.838	4.892
2013/2	48.182	7.786	9.231	4.103	0.823	0.274	10.188	6.564	1.097	0.684	2.370	5.064
2013/4	45.943	7.626	9.436	1.436	0.823	0.274	10.030	6.564	1.128	0.684	2.391	5.552

Date	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl-1 (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
2013/6	50.526	8.681	9.573	4.103	1.189	0.351	9.872	6.581	1.171	1.299	2.425	5.282
2013/8	52.304	9.348	9.573	8.205	0.840	0.181	9.756	6.162	0.938	0.000	2.055	5.245
2013/11	43.882	7.171	9.983	0.000	0.823	0.178	9.821	7.082	1.221	0.308	2.454	4.841
2014/2	58.366	8.146	10.256	12.308	0.952	0.174	9.885	8.000	1.023	0.615	2.174	4.831
2014/4	55.057	6.864	10.393	10.256	0.823	0.336	10.192	7.137	1.283	0.342	2.497	4.933
2014/6	50.889	2.537	10.940	10.256	1.233	0.497	10.500	6.274	1.229	0.068	2.297	5.056
2014/8	56.575	6.143	10.256	12.308	1.061	0.498	10.474	6.479	1.298	1.436	2.581	4.041
2014/11	60.376	8.146	11.624	8.697	1.127	0.893	11.538	7.863	1.826	0.103	3.602	4.957

Table 2. Preprocessing Data

Date	Rainfall	WQI	DO (mg/l)	pH	BOD_5 (mg/l)	Ca^{+2} (mg/L)	Cl^{-1} (mg/l)	EC (µS/cm)	Mg^{+2} (mg/L)	Na^{+1} (mg/l)	NO_2^{-1} (mg/l)	SO_4^{-2} (mg/l)	TDS (mg/l)
1995/2	3.5	3.84	2.13	2.12	1.41	3.78	3.15	6.43	3.61	3.63	0.03	3.97	5.99
1995/4	3.47	3.87	2.09	2.11	1.46	3.83	3.17	6.45	3.59	3.63	0.05	3.97	6
1995/6	3.45	3.89	2.06	2.11	1.56	3.83	3.2	6.44	3.61	3.66	0.06	3.97	6
1995/8	3.38	3.94	2.07	2.1	1.74	3.86	3.29	6.47	3.63	3.7	0.07	3.99	6.02
1995/11	3.45	3.98	2.11	2.1	1.91	3.86	3.34	6.48	3.65	3.74	0.07	4.01	6.04
996/2	3.3	4.04	2.15	2.1	2.09	3.89	3.32	6.52	3.67	3.71	0.07	4	6.08
1996/4	3.06	4.07	2.12	2.11	2.15	3.91	3.2	6.52	3.7	3.64	0.08	3.95	6.08
1996/6	2.53	4.08	2.04	2.12	2.21	3.9	3.05	6.51	3.75	3.54	0.1	3.88	6.07
1996/8	2.29	4.04	2	2.12	2.14	3.91	2.98	6.51	3.78	3.56	0.14	3.86	6.07
1996/11	2.38	3.99	1.98	2.12	2.04	3.93	2.98	6.53	3.78	3.61	0.15	3.84	6.09
1997/2	2.88	3.96	2.04	2.12	1.87	4.01	3.06	6.55	3.72	3.7	0.15	3.88	6.12

1997/4	3.43	3.94	2.08	2.13	1.68	4.08	3.12	6.58	3.66	3.74	0.14	3.95	6.17
1997/6	3.69	3.92	2.14	2.13	1.49	4.13	3.11	6.6	3.62	3.77	0.15	4	6.19
1997/8	3.66	3.88	2.17	2.14	1.3	4.15	3.01	6.59	3.6	3.74	0.17	3.99	6.19
1997/11	3.56	3.81	2.19	2.13	1.13	4.12	2.83	6.54	3.56	3.65	0.19	3.9	6.14
1998/2	3.63	3.79	2.18	2.13	1.11	4.08	2.74	6.53	3.52	3.58	0.19	3.84	6.13
1998/4	3.63	3.83	2.18	2.12	1.25	4.05	2.82	6.56	3.53	3.63	0.16	3.9	6.15
1998/6	3.6	3.91	2.19	2.13	1.46	4.08	2.99	6.62	3.59	3.73	0.14	3.99	6.19
1998/8	3.54	3.96	2.21	2.14	1.63	4.12	3.12	6.62	3.63	3.78	0.1	4.08	6.18
1998/11	3.6	3.92	2.21	2.13	1.63	4.12	3.09	6.55	3.61	3.67	0.09	4.01	6.1
999/3	3.61	3.86	2.18	2.12	1.62	4.04	3.02	6.46	3.56	3.55	0.05	3.94	6.01
1999/4	3.58	3.79	2.16	2.1	1.55	3.95	2.95	6.39	3.5	3.48	0.04	3.82	5.94
1999/6	3.49	3.78	2.15	2.09	1.54	3.85	2.94	6.37	3.52	3.52	0.02	3.82	5.92
1999/8	3.45	3.75	2.14	2.08	1.45	3.79	2.9	6.35	3.56	3.52	0.02	3.79	5.91
1999/11	3.45	3.76	2.16	2.08	1.4	3.73	2.94	6.39	3.65	3.55	0.02	3.83	5.94
2000/2	3.63	3.82	2.12	2.08	1.46	3.71	3.05	6.46	3.7	3.58	0.02	3.84	6.01
2000/4	3.74	3.92	2.03	2.08	1.56	3.71	3.22	6.56	3.77	3.66	0.02	3.85	6.11
2000/6	3.68	4.03	1.91	2.09	1.69	3.74	3.29	6.62	3.81	3.67	0.02	3.84	6.17
2000/8	3.46	4.01	1.86	2.11	1.54	3.79	3.19	6.58	3.79	3.59	0.03	3.82	6.13
2000/11	3.26	3.93	1.88	2.11	1.31	3.83	3.08	6.52	3.72	3.5	0.03	3.83	6.09
2001/2	3.13	3.82	1.89	2.12	1	3.9	2.94	6.46	3.6	3.44	0.03	3.84	6.05
2001/4	2.89	3.79	1.91	2.12	0.93	3.98	2.86	6.46	3.55	3.41	0.04	3.84	6.06
2001/6	2.78	3.8	1.93	2.14	1.02	4.06	2.71	6.47	3.54	3.4	0.04	3.83	6.08
2001/8	2.87	3.82	2.02	2.14	1.19	4.07	2.63	6.48	3.55	3.37	0.07	3.81	6.09
2001/11	3.32	3.83	2.01	2.13	1.27	4.03	2.58	6.5	3.55	3.41	0.08	3.82	6.1
2002/4	3.51	3.84	1.98	2.12	1.26	4.03	2.67	6.5	3.55	3.46	0.11	3.86	6.11
2002/4	3.62	3.86	1.95	2.12	1.25	4.07	2.81	6.52	3.55	3.55	0.11	3.92	6.12
2002/6	3.37	3.9	1.97	2.13	1.29	4.14	3	6.55	3.57	3.64	0.1	3.99	6.14
2002/8	3.41	3.91	2.01	2.14	1.2	4.15	3.1	6.55	3.63	3.66	0.07	4.03	6.13
2002/11	3.54	3.9	2.01	2.13	1.13	4.08	3.16	6.53	3.67	3.66	0.05	4.05	6.1
2003/2	3.63	3.84	2.01	2.12	1.04	3.9	3.09	6.47	3.7	3.55	0.04	4.01	6.02
2003/4	3.29	3.82	1.99	2.1	1.2	3.77	3.02	6.43	3.68	3.51	0.03	3.97	5.98

2003/6	2.67	3.82	1.99	2.09	1.35	3.7	2.95	6.41	3.69	3.49	0.03	3.95	5.96
2003/8	2.34	3.83	1.98	2.09	1.45	3.75	2.93	6.43	3.68	3.52	0.04	3.93	5.98
2003/11	2.5	3.84	1.95	2.08	1.44	3.81	2.96	6.45	3.67	3.56	0.03	3.92	6.01
2004/2	2.93	3.86	1.89	2.07	1.47	3.9	3.01	6.49	3.64	3.56	0.03	3.91	6.04
2004/4	3.29	3.92	1.79	2.07	1.57	4.01	3.12	6.53	3.62	3.6	0.02	3.93	6.09
2004/6	3.49	3.96	1.71	2.07	1.53	4.02	3.2	6.54	3.64	3.57	0.03	3.89	6.09
2004/8	3.57	3.96	1.72	2.08	1.43	3.97	3.26	6.53	3.68	3.58	0.03	3.85	6.1
2004/11	3.73	3.95	1.81	2.1	1.39	3.81	3.24	6.5	3.76	3.58	0.04	3.83	6.07
2005/2	3.74	4.01	1.95	2.13	1.66	3.77	3.21	6.48	3.79	3.64	0.03	3.88	6.08
2005/4	3.63	4.05	2.05	2.15	1.92	3.77	3.17	6.49	3.78	3.67	0.02	3.92	6.1
2005/6	3.41	4.03	2.1	2.17	2.04	3.85	3.1	6.52	3.72	3.66	0.01	3.91	6.13
2005/8	3.34	3.91	2.07	2.15	1.9	3.91	2.97	6.54	3.68	3.6	0	3.84	6.16
2005/11	3.5	3.8	2	2.14	1.7	3.95	2.8	6.54	3.63	3.53	0	3.76	6.15
2006/2	3.63	3.76	1.94	2.13	1.56	3.98	2.72	6.54	3.59	3.5	0	3.75	6.12
2006/4	3.52	3.79	1.93	2.12	1.54	4.01	2.74	6.54	3.56	3.5	0.01	3.79	6.07
2006/6	3.31	3.91	1.94	2.13	1.64	4.1	2.92	6.59	3.57	3.6	0.04	3.89	6.1
2006/8	3.12	3.99	1.92	2.12	1.75	4.14	3.07	6.6	3.61	3.66	0.06	3.95	6.13
2006/11	3.21	4.02	1.89	2.11	1.77	4.16	3.2	6.63	3.65	3.73	0.07	4.01	6.17
2007/2	3.39	3.99	1.91	2.09	1.78	4.08	3.16	6.59	3.67	3.67	0.05	4.02	6.17
2007/4	3.42	3.96	1.92	2.09	1.69	4.02	3.1	6.59	3.66	3.61	0.03	4.03	6.15
2007/6	3.42	3.95	1.95	2.09	1.65	3.96	3.07	6.58	3.65	3.56	0.02	4.01	6.12
2007/8	3.25	3.95	1.94	2.09	1.61	3.96	3.12	6.57	3.61	3.57	0.03	3.98	6.1
2007/11	3.4	3.97	1.97	2.09	1.74	3.92	3.21	6.55	3.62	3.6	0.04	3.94	6.08
2008/6	3.28	4	1.95	2.09	1.88	3.83	3.24	6.51	3.66	3.63	0.03	3.9	6.1
2008/4	3.13	3.99	1.95	2.08	1.83	3.77	3.25	6.51	3.76	3.6	0.02	3.9	6.14
2008/6	2.9	3.96	1.9	2.08	1.68	3.75	3.21	6.53	3.8	3.55	0.01	3.89	6.19
2008/8	3.02	3.89	1.89	2.08	1.38	3.88	3.15	6.57	3.79	3.49	0.01	3.91	6.21
2008/11	3.47	3.83	1.87	2.08	1.22	4.01	3.15	6.6	3.71	3.56	0.01	3.95	6.23
2009/2	3.86	3.82	1.89	2.08	1.1	4.2	3.15	6.62	3.66	3.64	0.02	4.02	6.23
2009/4	3.9	3.83	1.91	2.08	1.05	4.28	3.17	6.63	3.62	3.72	0.03	4.07	6.23
2009/6	3.72	3.84	1.96	2.09	0.99	4.21	3.07	6.62	3.64	3.69	0.04	4.05	6.23

2009/8	3.49	3.81	2	2.09	0.92	4.04	2.95	6.6	3.65	3.64	0.02	3.98	6.23
2009/11	3.56	3.8	2.03	2.09	0.9	3.95	2.85	6.57	3.68	3.59	0.01	3.91	6.22
2010/2	3.68	3.78	2.01	2.08	0.93	4.02	2.83	6.56	3.63	3.59	0.01	3.88	6.22
2010/4	3.73	3.78	1.96	2.09	0.97	4.14	2.85	6.54	3.59	3.6	0.01	3.89	6.24
2010/6	3.58	3.78	1.91	2.1	1.06	4.19	2.91	6.54	3.55	3.62	0.02	3.88	6.27
2010/8	3.18	3.81	1.88	2.11	1.18	4.12	2.94	6.53	3.62	3.59	0.03	3.9	6.27
2010/11	2.99	3.84	1.91	2.1	1.34	4	3	6.55	3.68	3.58	0.03	3.95	6.25
2011/2	3.01	3.89	1.95	2.1	1.41	3.91	3.05	6.57	3.77	3.56	0.03	4.02	6.19
2011/4	3.18	3.9	1.98	2.1	1.39	3.85	3.1	6.58	3.78	3.6	0.03	4.07	6.15
2011/6	3.09	3.93	2.01	2.1	1.4	3.85	3.16	6.59	3.82	3.64	0.03	4.08	6.13
2011/8	2.9	3.96	2.06	2.11	1.56	3.85	3.17	6.58	3.83	3.68	0.02	4.06	6.13
2011/11	2.88	4.01	2.1	2.1	1.87	3.84	3.19	6.59	3.86	3.7	0.02	4.02	6.11
2012/2	3.13	4.01	2.12	2.1	2.02	3.84	3.2	6.59	3.86	3.69	0.01	3.97	6.07
2012/4	3.13	4.01	2.1	2.09	2.01	3.88	3.27	6.6	3.82	3.68	0.03	3.94	6.07
2012/6	3.02	3.98	2.1	2.1	1.84	3.99	3.31	6.63	3.74	3.7	0.05	3.92	6.09
2012/8	2.9	3.96	2.11	2.11	1.66	4.11	3.3	6.66	3.65	3.71	0.09	3.93	6.13
2012/11	3.24	3.92	2.15	2.12	1.4	4.17	3.2	6.67	3.63	3.72	0.11	3.95	6.17
2013/2	3.45	3.89	2.16	2.12	1.14	4.19	3.16	6.67	3.63	3.74	0.14	4	6.21
2013/4	3.34	3.88	2.15	2.13	1.02	4.18	3.13	6.66	3.64	3.75	0.14	4.02	6.24
2013/6	2.92	3.89	2.12	2.13	1.16	4.2	3.12	6.65	3.64	3.76	0.12	4.04	6.24
2013/8	2.6	3.91	2.11	2.13	1.45	4.18	3.06	6.64	3.66	3.74	0.09	4.01	6.22
2013/11	2.6	3.93	2.15	2.14	1.71	4.16	3.05	6.65	3.71	3.77	0.1	4.02	6.19
2014/2	2.85	3.96	2.19	2.14	1.82	4.17	3.09	6.66	3.74	3.79	0.1	4.03	6.17
2014/4	3.15	3.98	2.25	2.14	1.84	4.22	3.2	6.68	3.72	3.85	0.11	4.06	6.15
2014/6	3.44	4.01	2.27	2.15	1.83	4.29	3.35	6.71	3.69	3.88	0.1	4.08	6.14
2014/8	3.54	4.01	2.29	2.15	1.83	4.32	3.43	6.72	3.67	3.91	0.1	4.11	6.12
2014/11	3.67	4.03	2.27	2.15	1.8	4.36	3.52	6.74	3.69	3.92	0.08	4.12	6.12

Table 3. The correlation of the WQI parameters and the rainfall after the SSA technique

		WQI	BOD_5	Ca^{+2}	Cl^{-1}	DO	EC	Mg^{+2}	Na^{-1}	NO_2^{-1}	pH	SO_4^{-2}	TDS	Rainfall
WQI	Pearson Correlation	1	.787**	-0.050	.716**	0.086	.398**	.616**	.518**	0.154	0.193	.348**	0.060	-.219*
	Sig. (2-tailed)		0.000	0.623	0.000	0.394	0.000	0.000	0.000	0.127	0.055	0.000	0.551	0.028
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
BOD	Pearson Correlation	.787**	1	-.226*	.480**	.233*	0.096	.456**	.329**	0.040	0.157	0.056	-.255*	-.298**
	Sig. (2-tailed)	0.000		0.024	0.000	0.020	0.341	0.000	0.001	0.692	0.119	0.579	0.011	0.003
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Ca^{+2}	Pearson Correlation	-0.050	-.226*	1	-0.021	.244*	.680**	-.407**	.488**	.491**	.436**	.445**	.585**	0.171
	Sig. (2-tailed)	0.623	0.024		0.834	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.090
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Cl^{-1}	Pearson Correlation	.716**	.480**	-0.021	1	0.122	.411**	.556**	.672**	-0.050	-0.122	.556**	0.034	0.064
	Sig. (2-tailed)	0.000	0.000	0.834		0.226	0.000	0.000	0.000	0.620	0.225	0.000	0.741	0.524
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
DO	Pearson Correlation	0.086	.233*	.244*	0.122	1	.208*	-0.096	.542**	.511**	.501**	.348**	-0.112	0.053
	Sig. (2-tailed)	0.394	0.020	0.014	0.226		0.038	0.343	0.000	0.000	0.000	0.000	0.267	0.600
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Mg^{+2}	Pearson Correlation	.398**	0.096	.680**	.411**	.208*	1	.255*	.722**	.321**	.275**	.574**	.741**	0.058

		WQI	BOD_5	Ca^{+2}	Cl^{-1}	DO	EC	Mg^{+2}	Na^{-1}	NO_2^{-1}	pH	SO_4^{-2}	TDS	Rainfall
	Sig. (2-tailed)	0.000	0.341	0.000	0.000	0.038		0.010	0.000	0.001	0.006	0.000	0.000	0.563
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Mg^{+2}	Pearson Correlation	.616**	.456**	-.407**	.556**	-0.096	.255*	1	.255*	-.239*	-0.159	.217*	0.109	-.326**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.343	0.010		0.010	0.017	0.115	0.030	0.281	0.001
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Na^{-1}	Pearson Correlation	.518**	.329**	.488**	.672**	.542**	.722**	.255*	1	.411**	.321**	.716**	.368**	0.158
	Sig. (2-tailed)	0.000	0.001	0.000	0.000	0.000	0.000	0.010		0.000	0.001	0.000	0.000	0.117
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
NO_2^{-1}	Pearson Correlation	0.154	0.040	.491**	-0.050	.511**	.321**	-.239*	.411**	1	.514**	.232*	0.183	-0.097
	Sig. (2-tailed)	0.127	0.692	0.000	0.620	0.000	0.001	0.017	0.000		0.000	0.020	0.069	0.337
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
pH	Pearson Correlation	0.193	0.157	.436**	-0.122	.501**	.275**	-0.159	.321**	.514**	1	0.129	0.160	-0.024
	Sig. (2-tailed)	0.055	0.119	0.000	0.225	0.000	0.006	0.115	0.001	0.000		0.201	0.111	0.816
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
SO_4^{-2}	Pearson Correlation	.348**	0.056	.445**	.556**	.348**	.574**	.217*	.716**	.232*	0.129	1	.320**	0.017
	Sig. (2-tailed)	0.000	0.579	0.000	0.000	0.000	0.000	0.030	0.000	0.020	0.201		0.001	0.867
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
TDS	Pearson	0.060	-.255*	.585**	0.034	-0.112	.741**	0.109	.368**	0.183	0.160	.320**	1	0.121

		WQI	BOD_5	Ca^{+2}	Cl^{-1}	DO	EC	Mg^{+2}	Na^{-1}	NO_2^{-1}	pH	SO_4^{-2}	TDS	Rainfall
	Correlation													
	Sig. (2-tailed)	0.551	0.011	0.000	0.741	0.267	0.000	0.281	0.000	0.069	0.111	0.001		0.230
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Rainfall	Pearson Correlation	-.219*	-.298**	0.171	0.064	0.053	0.058	-.326**	0.158	-0.097	-0.024	0.017	0.121	1
	Sig. (2-tailed)	0.028	0.003	0.090	0.524	0.600	0.563	0.001	0.117	0.337	0.816	0.867	0.230	
	N	100	100	100	100	100	100	100	100	100	100	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 4. The correlation of the WQI parameters and rainfall before the SSA technique

		WQI	BOD_5	Ca^{+2}	Cl^{-1}	DO	EC	Mg^{+2}	Na^{-1}	NO_2^{-1}	pH	SO_4^{-2}	TDS	Rainfall
WQI	Pearson Correlation	1	.754**	0.024	.609**	-0.023	.480**	.527**	.509**	0.074	.212*	.336**	.334**	-0.150
	Sig. (2-tailed)		0.000	0.811	0.000	0.822	0.000	0.000	0.000	0.462	0.034	0.001	0.001	0.136
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
BOD	Pearson Correlation	.754**	1	-0.155	.367**	0.163	0.165	.338**	.332**	-0.021	0.102	.427**	0.024	-0.166
	Sig. (2-tailed)	0.000		0.123	0.000	0.104	0.101	0.001	0.001	0.835	0.313	0.000	0.813	0.099
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Ca^{+2}	Pearson Correlation	0.024	-0.155	1	0.107	0.083	.502**	-.304**	.298**	.281**	.269**	-.279**	.365**	0.042
	Sig. (2-tailed)	0.811	0.123		0.291	0.412	0.000	0.002	0.003	0.005	0.007	0.005	0.000	0.675
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Cl^{-1}	Pearson Correlation	.609**	.367**	0.107	1	0.102	.542**	.333**	.719**	0.005	-0.012	.354**	.294**	0.095

		WQI	BOD_5	Ca^{+2}	Cl^{-1}	DO	EC	Mg^{+2}	Na^{-1}	NO_2^{-1}	pH	SO_4^{-2}	TDS	Rainfall
	Sig. (2-tailed)	0.000	0.000	0.291		0.314	0.000	0.001	0.000	0.958	0.905	0.000	0.003	0.349
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
DO	Pearson Correlation	-0.023	0.163	0.083	0.102	1	0.108	-0.009	.256*	.259**	.297**	.228*	-0.083	0.038
	Sig. (2-tailed)	0.822	0.104	0.412	0.314		0.283	0.928	0.010	0.009	0.003	0.022	0.412	0.708
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Mg^{+2}	Pearson Correlation	.480**	0.165	.502**	.542**	0.108	1	.325**	.654**	0.169	0.158	-0.006	.747**	0.016
	Sig. (2-tailed)	0.000	0.101	0.000	0.000	0.283		0.001	0.000	0.094	0.115	0.954	0.000	0.876
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Mg^{+2}	Pearson Correlation	.527**	.338**	-.304**	.333**	-0.009	.325**	1	.201*	-0.048	-0.003	.198*	.209*	-.205*
	Sig. (2-tailed)	0.000	0.001	0.002	0.001	0.928	0.001		0.045	0.635	0.979	0.048	0.037	0.041
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Na^{-1}	Pearson Correlation	.509**	.332**	.298**	.719**	.256*	.654**	.201*	1	.216*	.229*	0.165	.484**	0.164
	Sig. (2-tailed)	0.000	0.001	0.003	0.000	0.010	0.000	0.045		0.031	0.022	0.101	0.000	0.104
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
NO_2^{-1}	Pearson Correlation	0.074	-0.021	.281**	0.005	.259**	0.169	-0.048	.216*	1	.307**	0.105	0.075	-0.046
	Sig. (2-tailed)	0.462	0.835	0.005	0.958	0.009	0.094	0.635	0.031		0.002	0.300	0.460	0.652
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
pH	Pearson Correlation	.212*	0.102	.269**	-0.012	.297**	0.158	-0.003	.229*	.307**	1	0.008	0.127	-0.115
	Sig. (2-tailed)	0.034	0.313	0.007	0.905	0.003	0.115	0.979	0.022	0.002		0.934	0.207	0.253
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
SO_4^{-2}	Pearson Correlation	.399**	0.090	.310**	.601**	.212*	.572**	.261**	.672**	0.159	0.174	0.076	.401**	0.023
	Sig. (2-tailed)	0.000	0.372	0.002	0.000	0.034	0.000	0.009	0.000	0.113	0.083	0.451	0.000	0.823
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
TDS	Pearson Correlation	.334**	0.024	.365**	.294**	-0.083	.747**	.209*	.484**	0.075	0.127	-.234*	1	0.005

		WQI	BOD_5	Ca^{+2}	Cl^{-1}	DO	EC	Mg^{+2}	Na^{-1}	NO_2^{-1}	pH	SO_4^{-2}	TDS	Rainfall
	Sig. (2-tailed)	0.001	0.813	0.000	0.003	0.412	0.000	0.037	0.000	0.460	0.207	0.019		0.962
	N	100	100	100	100	100	100	100	100	100	100	100	100	100
Rainfall	Pearson Correlation	-0.150	-0.166	0.042	0.095	0.038	0.016	-.205*	0.164	-0.046	-0.115	-0.087	0.005	1
	Sig. (2-tailed)	0.136	0.099	0.675	0.349	0.708	0.876	0.041	0.104	0.652	0.253	0.391	0.962	
	N	100	100	100	100	100	100	100	100	100	100	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

CURRICULUM VITEA

Hasanain Ali Banwan ZAMILI was graduate from Dijlaa High School in 2011–2012,after that he got a bachelor's degree from Wasit University, Engineering College,Civil Department in 2015–2016. then, he has been accepted to attend Ondokuz Mayis University, Environmental Engineering Department, Master’s Programme (English) in 2020/Sep. On the other hand, he is employed at Antonoil Company for oil service. Furthermore, he communicates in two languages: English (ESL) and Arabic (native).

Contact Information :

ORCID ID : 0000-0002-1261-8591

Publications :

1. Zamili, H., Bakan, G., Zubaidi, S. L., & Alawsi, M. A. (2023). Water quality index forecast using artificial neural network techniques optimized with different metaheuristic algorithms. Modeling Earth Systems and Environment, 1-11. https://doi.org/10.1007/s40808-023-01750-1
2. Hasanain Zamili, Gulfem Bakan. Application Of Slime Mould Algorithm SMA And ANN Model To Predict Water Quality Index: The Yeşılırmak River, Turkey. Internatıonal Mediterranean Scientific Research Congress. https://www.izdas.org/akdeniz.
3. H A Gzar, H M H Al- Hachami, Y A Zakoor and H A B Zamili. Oily Wastewater Treatment Using Modified Rice Husks as a Cost-Effective Adsorbent. 2nd Al-Muthanna International Conference on Engineering Science and Technology MICEST-2022. https://aip.scitation.org/journal/apc