HAYATI
Journal of Biosciences

Vol. 33 No. 1, January 2026 204-218
DOI:10.4308/hjb.33.1.204-218
ISSN: 1978-3019
EISSN: 2086-4094

Research Article

Assessing Climate Factors and Cyanobacterial Abundance on Microcystins Prediction Using Artificial Neural Network: A Case Study in Malaysia's Drinking Water Reservoir

Nurul Awatif Ahmad[1], Som Cit Sinang[1*], Nurul Hila Zainuddin[2], Noorazrin Abdul Rajak[2]

[1]Department of Biology, Faculty of Science and Mathematics, Sultan Idris Education University, 35900 Tanjung Malim, Perak, Malaysia
[2]Department of Mathematics, Faculty of Science and Mathematics, Sultan Idris Education University, 35900 Tanjung Malim, Perak, Malaysia

ARTICLE INFO

Article history:
Received March 6, 2025
Received in revised form August 19, 2025
Accepted October 8, 2025
Available Online October 21, 2025

KEYWORDS:
ELISA,
machine learning,
predictive model,
toxic cyanobacteria,
water security,
water supply systems

ABSTRACT

Toxic cyanobacterial blooms often lead to contamination with cyanotoxins, particularly microcystins. This study aims to examine microcystins persistence in a selected public water supply system and predict their concentration at various points based on climate factors and cyanobacterial abundance. Using the Enzyme-Linked Immunosorbent Assay (ELISA) method, microcystins concentrations were quantified at various points of the water supply system, including the raw water intake, reservoir, water treatment plant outlet, and distribution system. The highest microcystins concentration was detected at the reservoir with a mean concentration of 2.63 µg/L. An artificial neural network (ANN) model was developed to predict microcystins concentration. Rainfall, temperature, chlorophyll-a, phycocyanin (BGA-PC), and *mcyE* gene copy numbers were used as inputs, while microcystins concentrations at various water sampling points served as outputs of the multilayer perceptron ANN. Using the Statistical Package for the Social Sciences (SPSS, ver. 29), three networks with scaled conjugate gradient, sigmoid functions, and one hidden layer with 4 to 13 neurons were trained and validated to determine the best configuration that fits the observed data. The result shows a satisfactory prediction at the reservoir (Point 2) with low values of error (root mean square error = 0.065) and high coefficient values (R^2 = 0.894) between experimental and predicted values, which are below the maximum value of the actual concentrations. Phycocyanin (BGA-PC) and chlorophyll-a had the most positive effects in predicting microcystins concentrations. These results indicate that ANN modelling can be a reliable tool for predicting microcystins contamination in drinking water reservoir.

1. Introduction

The rapid proliferation of cyanobacteria is a sign of eutrophication in stagnant or slow-moving water bodies, such as lakes and reservoirs. This phenomenon is primarily caused by an increase in the availability of one or more growth factors required for photosynthesis (Schindler 2006). Among anthropogenic causes, agro-based industries are considered the primary contributors to eutrophication by discharging nutrients, such as nitrogen and phosphorus, into water bodies, which are the most crucial factors (Khan *et al.* 2018; Ebrahimzadeh *et al.* 2021). Other factors, such as light intensity, water chemistry, and climate-related variables like temperature and rainfall, as well as hydrological conditions, also contribute to the formation of cyanobacterial blooms (Muhetaer *et al.* 2020; Huang *et al.* 2021; Pham *et al.* 2021). This phenomenon leads to microcystin contamination in lakes and reservoirs, which serve as sources of drinking water, ultimately deteriorating water quality and safety.

*Corresponding Author
E-mail Address: somcit@fsmt.upsi.edu.my

Microcystins are the most concerning cyanobacterial toxin due of their hepatotoxic effects and persistence in freshwater ecosystems (Sivonen and Jones 1999; Massey and Yang 2020). By far, microcystin-LR (MC-LR) is the most potent in terms of acute toxicity (Baliu-Rodriguez *et al.* 2022). Hepatotoxic microcystins are commonly associated with liver-related conditions. Microcystins are also known as liver tumor promoters, which result in liver failure in humans due to the inhibitory activity of protein phosphatases 1 and 2A (He *et al.* 2018; Jia *et al.* 2019; Shi *et al.* 2021).

The escalating events of harmful cyanobacterial blooms, along with the associated production of microcystins in freshwater systems, pose a critical challenge to public health. It is necessary to develop a reliable model for predicting cyanobacterial blooms and microcystin, particularly for informing early intervention plans and effective management of water resources. Artificial Neural Networks (ANN) provide an appealing alternative to standard statistical models, which have long been used to forecast cyanobacterial blooms and microcystin concentrations. ANN is capable of resolving complex, nonlinear relationships within environmental data (Millie *et al.* 2014; Lu *et al.* 2016; Marrone *et al.* 2023).

The ANN model is a data-driven modeling approach for modeling ecosystems, and it has been used over the past few decades. This model is a powerful tool that can be used for machine learning and prediction based on observational data. ANN is able to learn from previous data and has been used to predict future values. In the field of environmental modelling, ANN has been used primarily for "prediction and forecasting" in areas such as environmental science and ecology (Bui *et al.* 2017; Harris and Graham 2017; Srisuksomwong and Pekkoh 2020; Maier *et al.* 2023). As described by Kumar *et al.* (2020), the ANN model builds a network analogous to the human brain's neuron system. This data-driven model will relate the input to the desired output mathematically. The ANN model can predict upcoming data by using the training parameters and previous data of the desired output. However, several internal parameters, such as the number of hidden layers, the number of nodes in each hidden layer, spread values, maximum epochs, and others, should be considered to achieve a highly accurate result (Kumar *et al.* 2020).

In a recent report by Ahmad and Sinang (2024), cyanobacterial blooms have been recorded in Malaysia's water supply source, with the presence of toxic cyanobacterial species. Various environmental factors might contribute to this phenomenon, including climate factors. However, the role of environmental factors tends to be site-specific (Sinang *et al.* 2013). The variability of local temperature and rainfall might influence the formation of cyanobacterial blooms in water supply sources. Currently, there is limited information available on cyanobacterial blooms and microcystin contamination in tropical water supply systems, specifically in Malaysia.

A lack of information on microcystin detection in drinking water sourced from freshwater ecosystems was identified in the reviewed studies, most of which focused on raw water sources (Ahmad *et al.* 2024). Despite cyanobacterial blooms occurring in raw water sources, only a few reviewed studies have discussed microcystin contamination in treated drinking water, particularly at the last stage of the water treatment process (Domingos *et al.* 1999; Kabziński *et al.* 2000; Vieira *et al.* 2005; Lehman 2007; Kaloudis *et al.* 2013; Mohamed 2016; Mchau *et al.* 2021; Hurtado *et al.* 2022). This knowledge gap poses a challenge in elucidating microcystin exposure in drinking water.

Based on current studies, hazardous exposure to microcystin contamination through the human consumption of contaminated drinking water is a possible common situation (Jochimsen *et al.* 1998; Qin *et al.* 2010). Therefore, this study is the first to report on the detection of microcystin contamination in Malaysia's drinking water supply system. This study also aims to develop an ANN model to predict microcystin concentrations at various water supply systems based on local climate and cyanobacteria abundance. This study will provide researchers and water authorities with a better understanding of microcystin persistence resulting from toxic cyanobacterial blooms, as well as the implementation of a compulsory microcystin parameter to be regularly monitored.

2. Materials and Methods

2.1. Study Area

Water sampling was conducted at Malaysia's water supply system as described in Ahmad and Sinang (2024). The reservoir is used for both domestic water supply and flood mitigation. The study reservoir has a 52-hectare catchment area with a total storage of 10.6 million cubic meters and a height of 37 meters. The study reservoir also recorded daily rainfall, with the minimum volume recorded being below 0.1 mm, while the maximum value recorded was 87.2 mm. Meanwhile, the minimum value of air temperature recorded was 24.4°C, while

the maximum was 29.7°C. The daily air temperature and rainfall data of the study reservoir from June 2022 to June 2023 were obtained from the National Climate Centre, Malaysian Meteorological Department in Petaling Jaya, Selangor.

The water sampling to determine chlorophyll-a concentrations, phycocyanin biomass (BGA-PC), *mcyE* gene copy number, and microcystins was conducted at Point 1 (raw water intake), Point 2 (reservoir), Point 3 (water treatment plant outlet), and Point 4 (distribution system) as shown in Figure 1. At Point 2, three sampling sites were designated for water sampling, which were identified as Point 2 (A), Point 2 (B), and Point 2 (C). All the parameters were recorded bimonthly from June 2022 to June 2023. Water samples were analysed to determine cyanobacteria biomass parameters using methodologies published in Ahmad and Sinang (2024).

2.2. Microcystins Quantification

Water samples were collected at each sampling point and stored in cleaned amber glass bottles. Only treated water samples were treated with sodium thiosulfate pentahydrate (Emsure®, Merck) at a final concentration of 100 mg/L (< 1 mg/mL) immediately upon collection to remove residual chlorine, as chlorine will degrade microcystins, which potentially leads to test inaccuracy. Three cycles of the freeze/thaw method were used to lyse the cell-bound and free microcystins (total microcystins) in the water samples. Then, the water samples were filtered through a glass fibre filter into a clean glass vial. The ELISA procedure was performed using the Microcystins/Nodularins (ADDA) ELISA (EPA Method 546) (Abraxis 520011) according to the kit manufacturer's instructions. In brief, the ELISA method involves recognition of microcystins and their congeners by specific antibodies. The microplate wells were pre-coated with antibodies. The presence of a toxin in water samples will lead to the formation of a microcystin-protein analogue immobilized on the plate, competing for the binding sites of the anti-microcystin antibodies in the solution. Then, the addition of the second antibody-HRP was applied, which targets the ADDA moiety —a specific amino acid sequence present in all microcystin variants (Fischer *et al.* 2001).

2.3. Artificial Neural Network

The multilayer perceptron ANN model was designed to investigate the roles of climatic factors and cyanobacterial abundance in determining microcystin concentrations and their persistence at various points of the water sampling locations. The inputs consist of data from five parameters, including air temperature, rainfall, chlorophyll-a concentrations, phycocyanin biomass (BGA-PC), and *mcyE* gene copy number. Microcystin concentrations were introduced as the output using IBM SPSS software version 29 for Windows. Data normality was performed before all data was passed into the input, hidden, and output layers.

The development of the ANN model involved the use of the sigmoid activation function at both the hidden

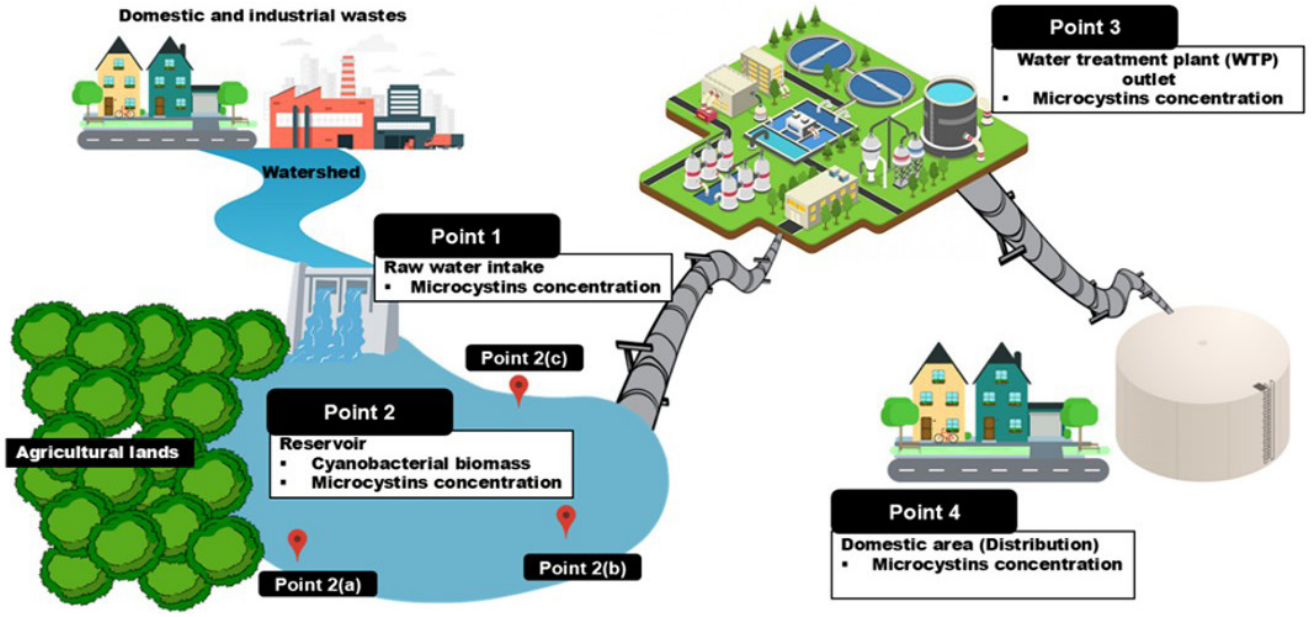

Figure 1. The outline of the process involved at each water sampling point of the water supply system

and output layers. Several indicators were considered in selecting the optimal ANN model, including the variability of network configurations, such as the number of hidden layers and the number of neurons per layer. Besides, the statistical indicators, including the coefficient of determination (R^2) and root mean square error (RMSE), were assessed, with smaller error values (closer to 0) and higher R^2 values (closer to 1) being selected for the model's performance. The final model was chosen based on its configuration, which offered excellent predictive accuracy and computational efficiency.

Two stages were involved in the modeling process: determining the ANN configuration and verifying the trained ANN model with the chosen configuration. At the configuration-determining stage, this ANN model was trained using a dataset of 24 experimental observations, which were randomly divided into 70% of the training data and 30% testing data. Then, at the verification stage, another dataset of experimental data was used to evaluate the resulting ANN model.

To improve network convergence and prediction accuracy, the number of neurons in the hidden layer should be optimized (Machón *et al.* 2007). A number of hidden nodes that is too large or too small will lead to overfitting (for instance, the model cannot predict untrained data) and underfitting (for instance, the model has high error tolerance), respectively. According to Wang *et al.* (2011), an appropriate number of neurons in one hidden layer should be determined as follows:

$$h = \frac{n + m}{2} + c$$

Where:
h : the number of neurons in the hidden layer
n : the number of input neurons
m : the number of output neurons
c : a constant between 1 and 10

In this study, n = 5 and m = 1; therefore, the number of neurons in the hidden layer should be varied between 4 and 13.

To avoid random correlation due to the random initialisation of the weights and bias, each configuration was repeated 10 times, and each run was trained for 100 epochs. An optimal ANN model was selected from the configurations based on the minimum error (root mean square error (RMSE)) and a maximum correlation coefficient (R^2) in both the total and testing set. These values will determine the performance of the formulated ANN model. Higher values of R^2 (close to 1) and smaller errors (close to 0) indicated better-trained models.

3. Results

In this study, microcystins were quantified throughout the water supply system, including the raw water intake point (Point 1), reservoirs (Points 2A, 2B, and 2C), water treatment plant outlet (Point 3), and distribution point (Point 4) (Figure 1). As shown in Figure 2, microcystins were detected at all sampling points. Microcystins were consistently detected at Point 2 (average of concentrations in 2A, 2B, and 2C), with the highest concentration of 2.629 µg/L recorded in August 2022, and the lowest concentration of 0.073 µg/L in February 2023. Meanwhile, at Points 1, 3, and 4, microcystins were either present at low concentrations or undetectable. At Point 1, the highest microcystin concentration was recorded at 0.221 µg/L in August 2022. In contrast, microcystins were not detected in early June 2022, late July 2022, mid-October 2022, late November 2022, early December 2022, mid-February 2023, and mid-May 2023. Additionally, Point 3 recorded the highest microcystins concentration at 0.697 µg/L in mid-October 2022. At this sampling point, microcystin concentration was also not detected at certain sampling events, including late July 2022, late September 2022, November 2022, February 2023, and mid-May 2023. At sampling Point 4, the highest microcystins reported were 0.188 µg/L in early March 2023. Microcystins were not detected in late July 2022, late September 2022, mid-October 2022, early December 2022, and mid-February 2023.

The concentration of microcystins was predicted at each sampling point based on cyanobacterial abundance, air temperature, and rainfall. Prediction of microcystins in the water supply system is very important to ensure the quality and safety of the treated water before being distributed to the end users. Using the ANN model as shown in Figure 3 (a-c), the best-fitted ANN managed to provide a composition with N (5-5-1), N (5-4-1), and N (5-10-1). This consists of one input layer with four to ten neurons inside, including bias, one hidden layer, and one final output of an artificial neural network.

The grey and blue arrows in Figure 3A-C represent the connection between the input layer and the hidden layer, and between the hidden layer and the output layer, with synaptic weights more than zero and less than zero, respectively. Six input neurons were abbreviated

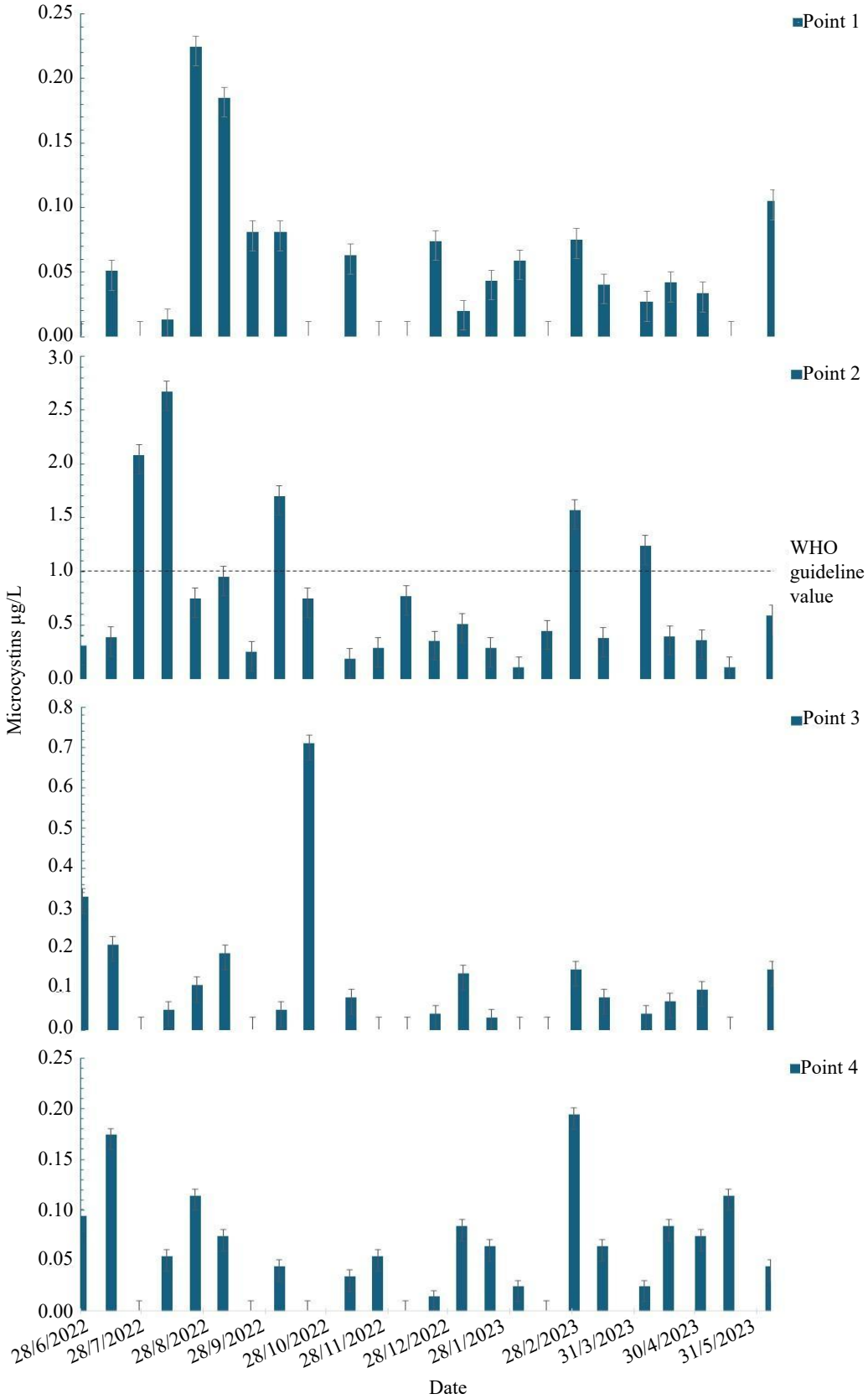

Figure 2. Mean values of microcystin concentrations at various points of the water supply systems. Point 1 (raw water intake), Point 2 (reservoir), Point 3 (water treatment plant outlet), Point 4 (distribution system)

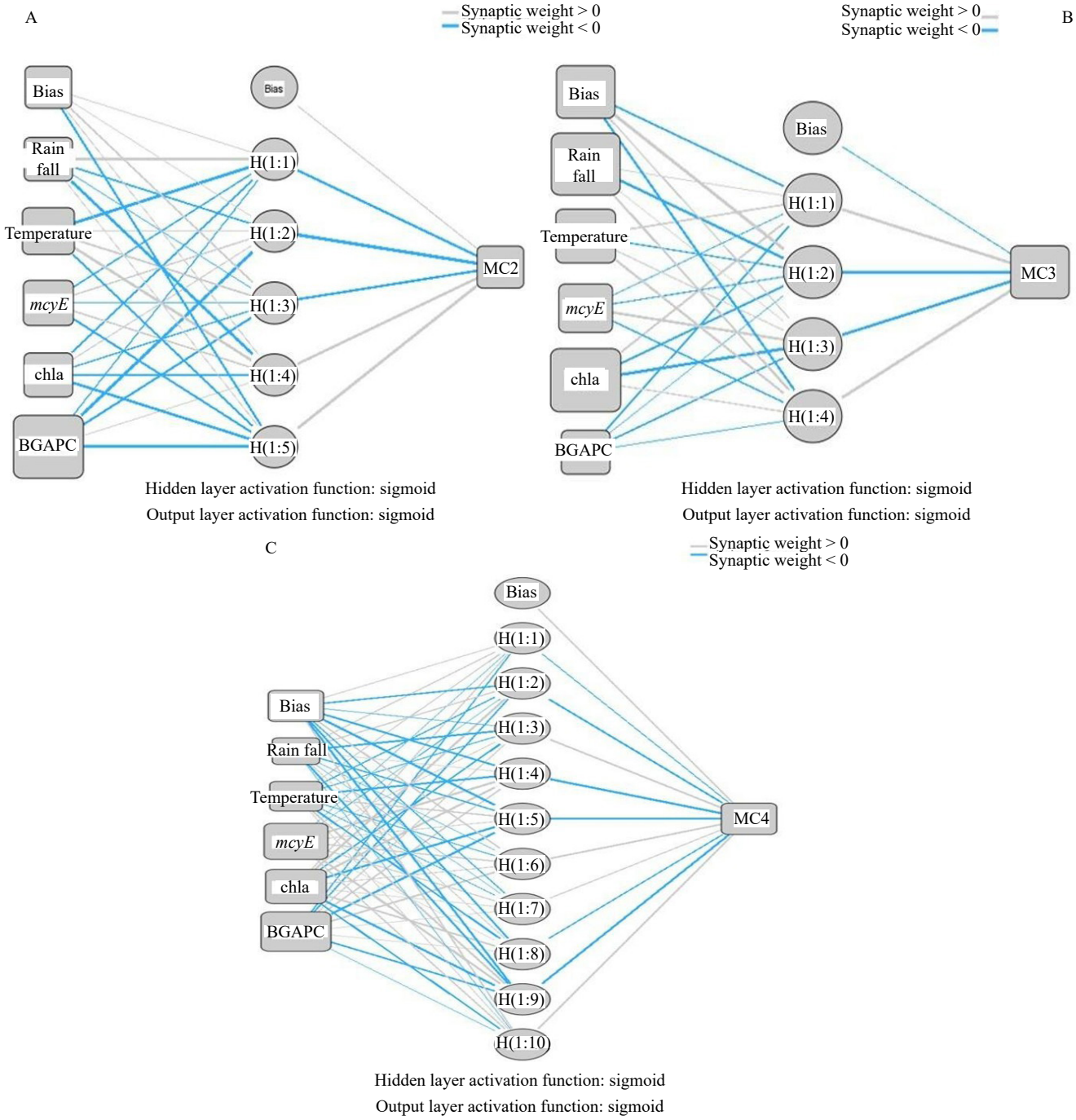

Figure 3. (A) $N^{(5\text{-}5\text{-}1)}$ of artificial neural network for sampling point 2 (reservoir), (B) $N^{(5\text{-}4\text{-}1)}$ of an artificial neural network for sampling point 3 (water treatment plant outlet), (C) $N^{(5\text{-}10\text{-}1)}$ of artificial neural network for sampling point 4 (distribution system)

as "Bias" for bias, "mcyE" for *mcyE* gene copy number, “Chla” for chlorophyll-a, and "BGAPC" for BGA-PC. They included the rainfall and temperature sample data. Meanwhile, the categorical output neuron abbreviations were "MC2" for microcystin concentration at Point 2 of the studied reservoir, "MC3" for microcystin concentration at Point 3 of the water treatment plant outlet, and "MC4" for microcystin concentration at Point 4 of the distribution system. The model was evaluated using three performance metrics: root mean squared error (RMSE), relative error (RE), and coefficient of determination (R^2), with a sigmoid as the activation function.

In determining the microcystin concentration at three different sampling points, the 24 sample data included in this study were divided into 70% of the

training data and 30% testing data for ANN analysis. This technique allows regression analysis mapping to significant sampling points. The efficiency of the regression ANN was assessed for every sampling point, resulting in finding the smaller root mean squared error (RMSE) and relative error (RE) for both train and test sample data as described in Table 1. ANN models of three sampling points had <1.0 RMSE values for both train and test sample data for microcystins prediction. Besides, the difference of relative error values between training and testing, which ranges from 0.01 to 0.10 (1% to 10%), suggests that these models generalize well to unseen data, exhibiting slight overfitting without overfitting.

Additionally, Figure 4A-C illustrates the R^2 values between the predicted and observed microcystin concentrations of these three models. All the R^2 values recorded were greater than 0.700, with the highest R^2 value recorded at 0.894 for sampling Point 2 of the ANN model. These values indicate the proportion of variance in the target variable explained by the model.

Figure 5A-C shows the residual analysis of the ANN model for each sampling point. This analysis indicates the specificity of a developed model. The residual analysis between residuals and predicted values of these three ANN models shows that the residual plots are randomly scattered around zero with no discernible pattern or clear systematic trend, indicating that the models are a good fit. These results suggest that these models capture the overall relationship between the independent and the dependent variable fairly well.

The most important variables for predicting microcystins at sampling points 2, 3, and 4 were described in Figure 6A-C. At sampling points 2 and 4, BGA-PC concentration is the most important variable in predicting microcystin concentration, with 100.0% normalized importance, while rainfall is the least important, with 53.6% and 22.4% normalized importance, respectively. Meanwhile, the most important variable at Sampling Point 3 was chlorophyll-a concentration, which accounted for 100.0% of normalized importance, while BGA-PC was the least important, accounting for 8.0%.

Furthermore, the actual and predicted microcystin concentrations of sampling Points 2, 3, and 4 were plotted, as shown in Figure 7A-C, using the regression fit line formula shown in Figure 4A-C. The data was predicted using the actual data collected throughout the sampling events. The predicted microcystin values were plotted in an orange solid line while the actual microcystins were plotted in a blue solid line. Overall, the microcystin prediction model at Sampling Point 2 outperformed the models of Sampling Points 3 and 4, with the prediction closely replicating the actual concentration values. However, none of the models predicted microcystin concentration more than the maximum value of the actual concentrations.

4. Discussion

Microcystin contamination was persistently detected from raw water sources to treated water samples in selected drinking water reservoirs in Malaysia. Microcystin concentrations quantified at Point 2 (reservoir) surpassed the safe levels of guideline value, which is 1.0 µg/L during five sampling events in which recorded at 2.036 µg/L (late July 2022), 2.629 µg/L (early August 2022), 1.659 µg/L (early October 2022), 1.532 µg/L (early March 2023), and 1.195 µg/L (early April 2023). At sampling Points 3 and 4, microcystins were persistently detected at concentrations below 1.0 µg/L to undetectable levels; however, these concentrations were acceptable based on the recommended guideline value for safe microcystin levels in drinking water (WHO 2022). A few studies have reported similar findings, in which the microcystin levels surpassed the safe level at the reservoir (raw water), but in the treated water, the microcystin concentration remained below 1.0 µg/L (Robert *et al.* 2005; Haddix *et al.* 2007). Even so, consuming water with microcystin contamination for a long time may cause chronic liver damage and inflammation, liver cancer, intestinal cancer, colorectal cancer, tumour promotion, and may produce genotoxicity (Falconer 1991; Svirčev *et al.* 2010; Rastogi *et al.* 2015; Massey *et al.* 2018; Ge *et al.* 2021).

Table 1. Comparison estimation of error for every sampling point

Estimation value	Point 2 (reservoir)		Point 3 (water treatment plant outlet)		Point 4 (distribution system)	
	RMSE	RE	RMSE	RE	RMSE	RE
Training	0.065	0.113	0.079	0.293	0.070	0.181
Testing	0.026	0.073	0.053	0.197	0.039	0.101
R^2	0.894		0.758		0.844	

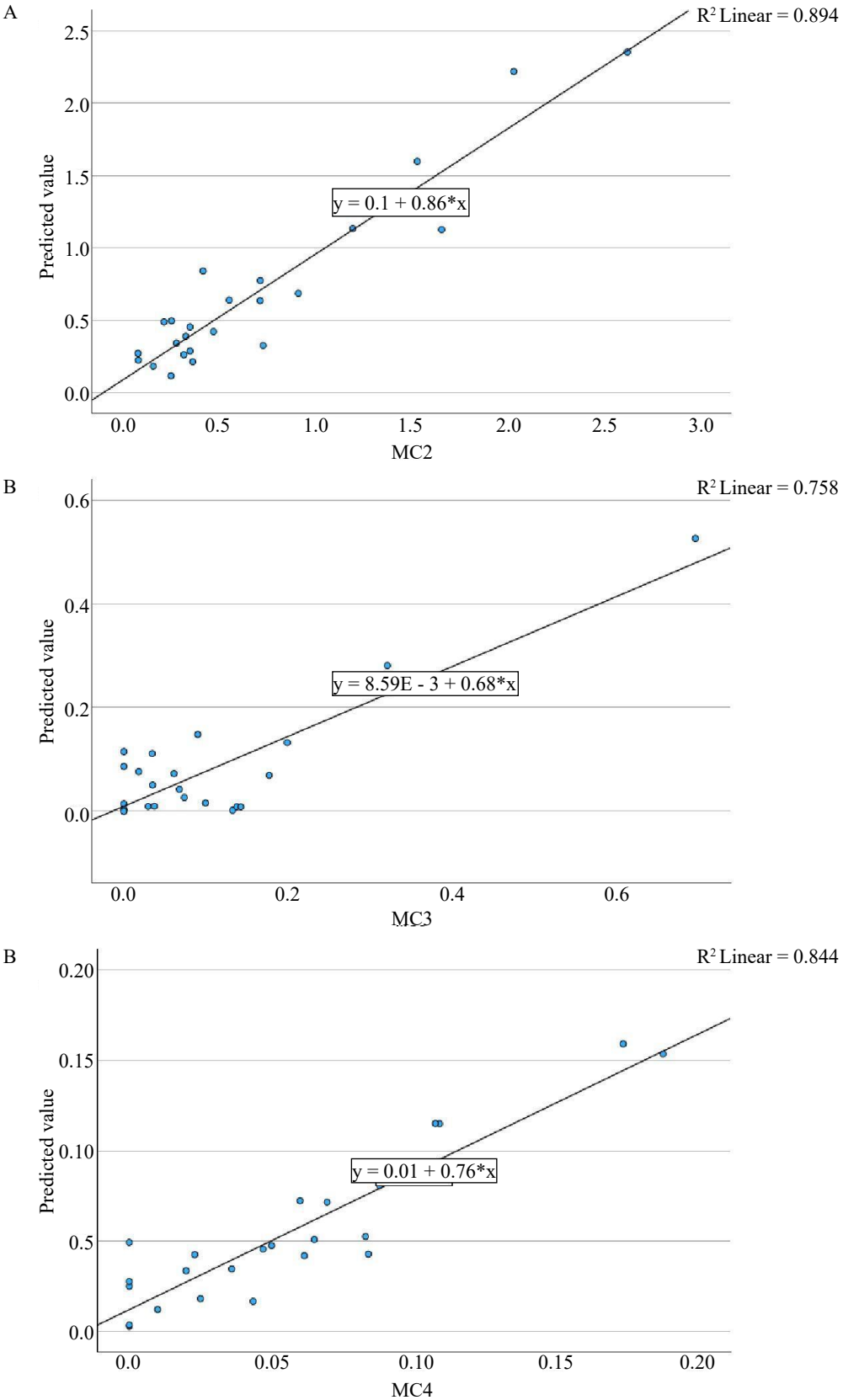

Figure 4. (A) Observed compared to predicted microcystins concentration at sampling point 2 (reservoir), (B) point 3 (water treatment plant outlet), and (C) point 4 (distribution system). The solid line represents the linear regression line

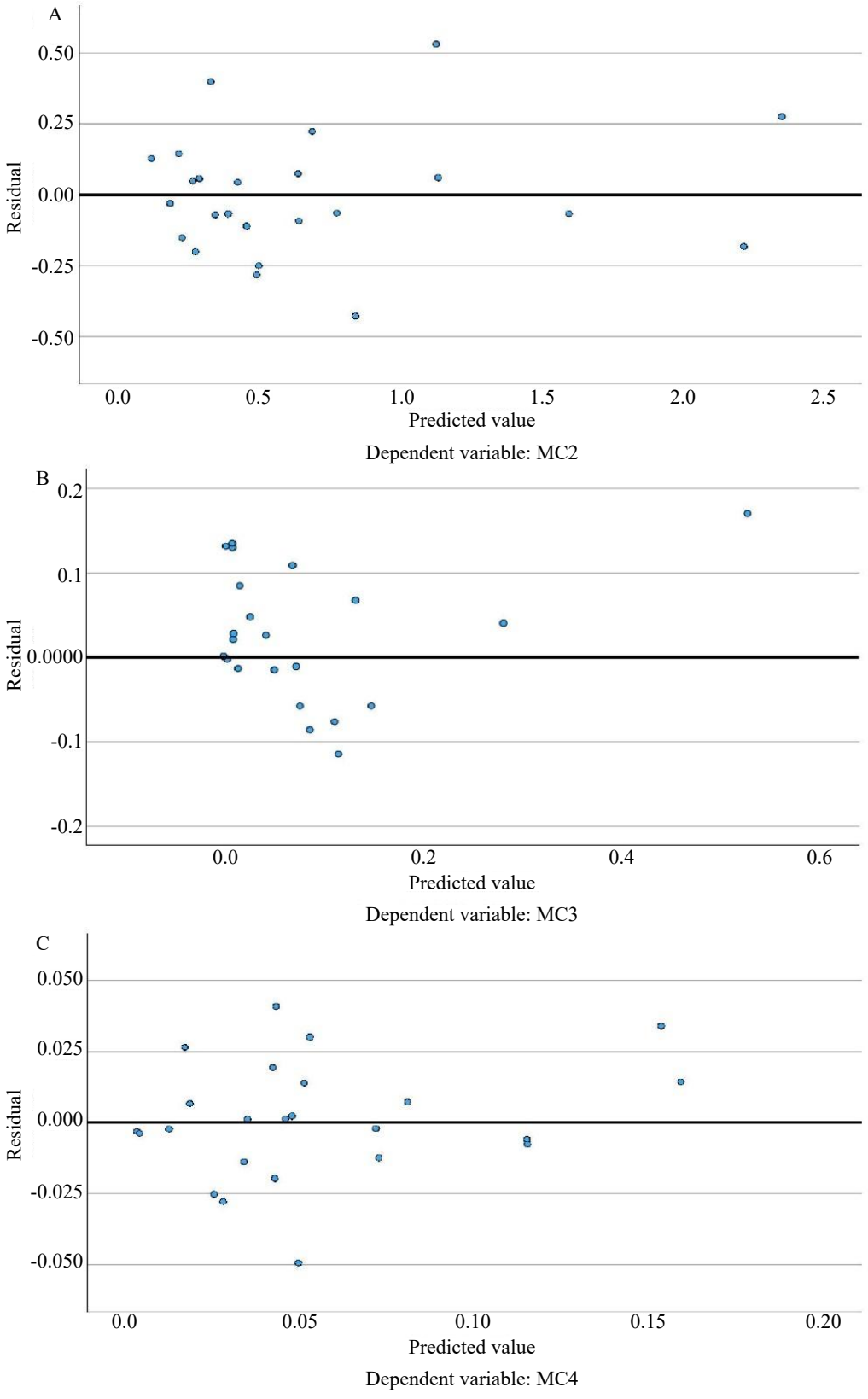

Figure 5 (A) Scattered plots of predicted values against residuals at sampling point 2 (reservoir), (B) point 3 (water treatment plant outlet), and (C) point 4 (distribution system)

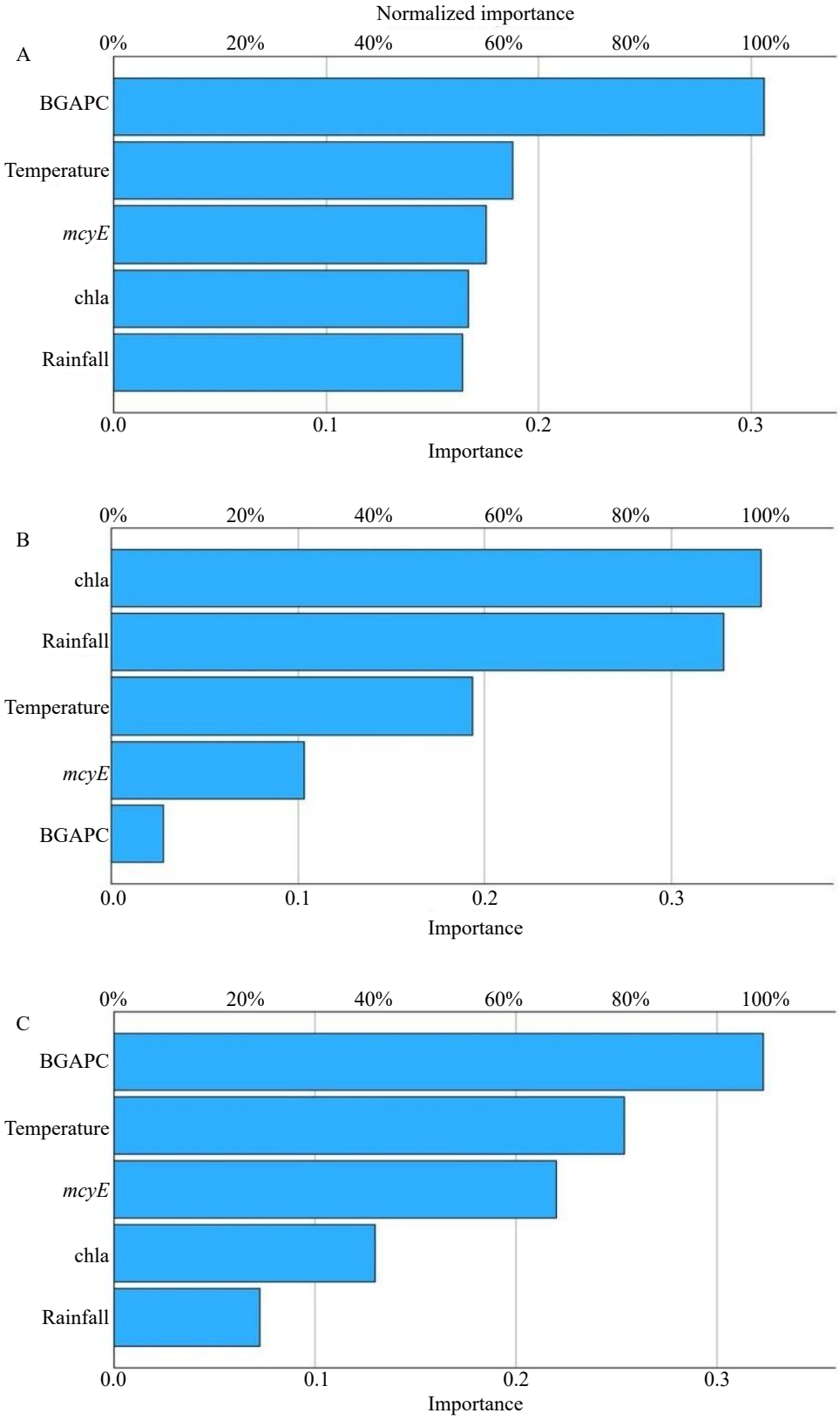

Figure 6. (A) The most important variables for predicting microcystins concentration are sampling point 2 (reservoir), (B) point 3 (water treatment plant outlet), and (C) point 4 (distribution system)

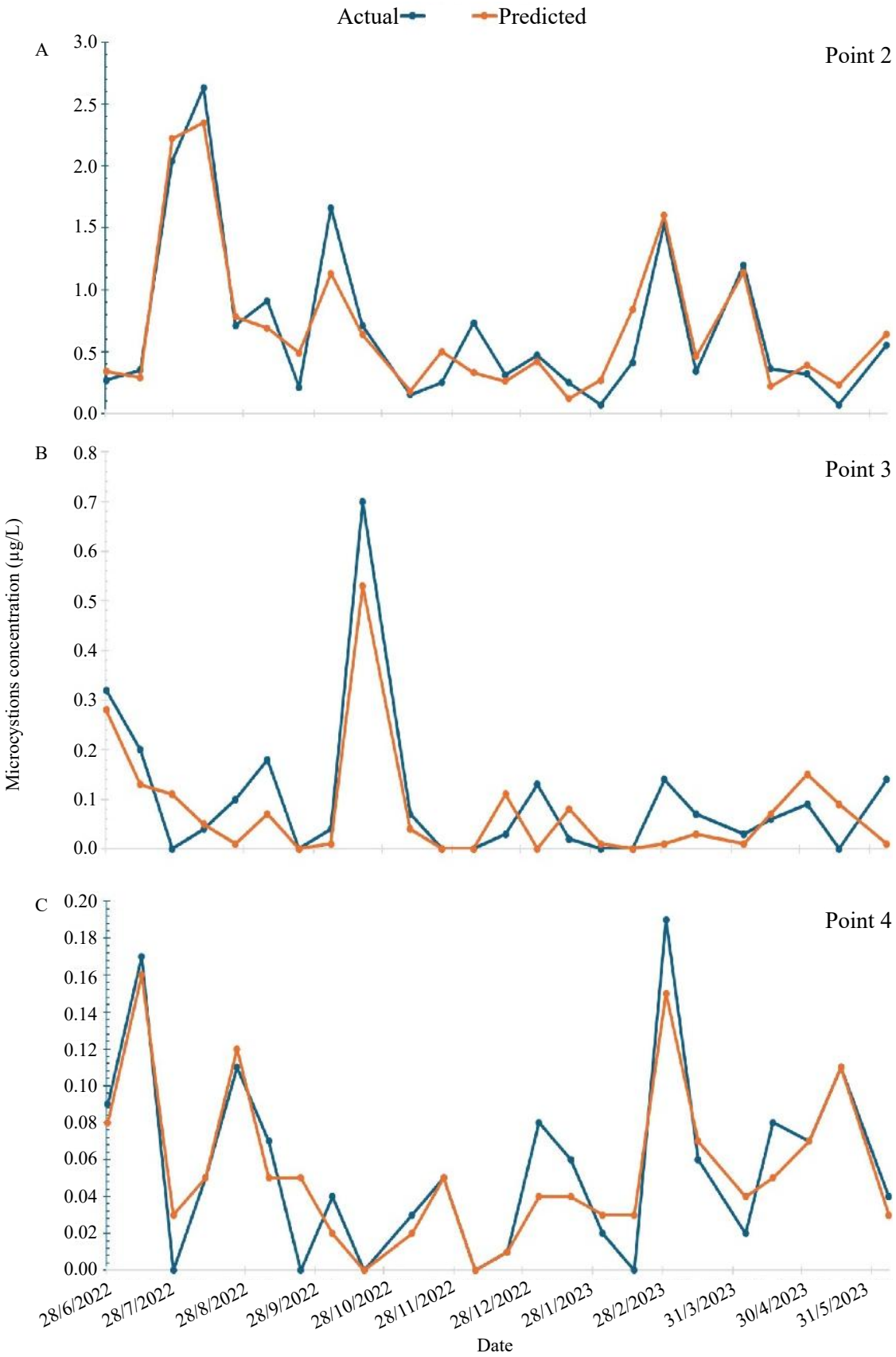

Figure 7. (A-C) Actual and predicted microcystin concentrations at sampling points 2, 3, and 4 from June 2022 to June 2023 in the study reservoir using a neural network model approach

In this study, an ANN model was developed to predict the concentration of microcystins at various water sampling points based on climate factors and cyanobacterial abundance. Several metrics, including R^2, RMSE, RE, and residual analysis, were evaluated to assess the fitness of ANN models. Overall, these ANN models are categorized as fit, and the residual analysis, combined with the evaluation of error metrics, indicates the model's validation (Kutner *et al.* 2005; Montgomery *et al.* 2021; Shanmugam *et al.* 2022). Meanwhile, the training and testing datasets were partitioned from the same dataset, which could lead to overfitting. However, using a multilayer perceptron ANN model in SPSS, this issue was resolved by implementing batch training with an early stopping rule, which can reduce the overfitting issues (Srivastava *et al.* 2014; Ying 2019). This suggests that the multilayer perceptron of the ANN can be utilized to develop a good mathematical model for predicting the microcystin concentration of the study reservoir. Even so, it is best to validate the model on an independent dataset to confirm the robustness. Due to the unavailability of the independent dataset of the study reservoir, further validation cannot be performed.

In developing an ANN prediction model, several factors were assessed to predict microcystin concentrations using meteorological, hydrological, and environmental parameters (Oliver *et al.* 2012; Rigosi *et al.* 2015; Nunes Carvalho *et al.* 2022). While these models offer a promising approach for forecasting microcystin levels, their predictive accuracy remains constrained by the complex and often nonlinear relationship between cyanobacterial cell abundance and toxin production. Existing studies suggest that an increase in cyanobacterial cell density does not necessarily correspond to elevated microcystin concentrations, thereby limiting the efficacy of ANN-based models for microcystin risk management (Turner *et al.* 2018; Hartnell *et al.* 2020). Future research should focus on integrating additional biochemical and genetic markers, beyond the *mcyE* gene, to enhance model reliability and improve predictive capabilities for effective water quality management.

Microcystis spp. was known to be the dominant species and microcystin producer in this study reservoir, and the variations of microcystin concentration are likely linked to the overall abundance of this species. However, the actual cyanobacterial biomass does not always accurately describe the microcystin concentration in the water sources. During cyanobacterial blooms, gene expression, the percentage of cells that can produce microcystins, and the amount of microcystins produced per cell, such as cell quota, may influence the variability of microcystins in the study reservoir (Chorus and Bartram 1999; Harris and Graham 2017; Bui *et al.* 2017). Thus, not all cyanobacterial blooms produce microcystins, which may limit the predictive model's ability to estimate maximum concentrations accurately.

The predictive ANN models outperformed in predicting microcystin concentration at various points of the water supply system. However, none of the models predicted the highest concentration of microcystins in the actual dataset. Several studies have shown similar outcomes in which the developed models were unable to predict the highest actual data of microcystins concentration (Recknagel *et al.* 2017; Harris and Graham 2017; Bui *et al.* 2017). Several probabilities may influence the underestimation of maximum concentration by the developed models, such as temporal variabilities of BGA-PC and chlorophyll-a concentrations at the study reservoir. The potential underprediction observed in this study highlights the importance of improving model performance by increasing temporal resolution (from daily to high-frequency intervals), as high temporal resolution can influence model performance and indicate variable sensitivities across water quality parameters (Shi *et al.* 2025).

The ANN predictive model can be enhanced by utilizing a long-term dataset to increase the temporal robustness of the predictive model (Harris and Graham 2017; Marrone *et al.* 2023; Park *et al.* 2024). The long-term dataset will capture the complete range of intra- and inter-annual variability within the reservoir, which indicates the multi-year variability in both interpretive and response variables when creating predictive models (Harris and Graham 2017; Li *et al.* 2021). However, this predictive model is site-specific, which can only be implemented at the study reservoir. Additionally, other factors, such as climate change, will gradually increase environmental variability in the tropical reservoir, potentially impacting model outcomes (Wiegand *et al.* 2021; Cottingham *et al.* 2021). These factors may be contributing to the temporal inconsistency of predictive model outcomes in the study reservoir. Hence, these two factors should be taken into consideration when constructing a predictive model.

Among the generated ANN models, the predictive ANN model at sampling Point 2 is more reliable compared to others. This is due to the predicted values being plotted in a manner similar to the actual values. However, the inefficiency of predictive models in

accurately predicting microcystin concentrations at maximum actual values poses a risk to public health. Microcystins exposure through the ingestion of drinking water poses a great threat to humans due to their potential toxicity. In 2021, Chorus and Welker identified cyanotoxins as an emerging health risk. Cyanotoxins, specifically microcystins, have been linked to various health issues, including gastrointestinal disorders (Drobac *et al.* 2017; Kubickova *et al.* 2019), skin irritation (Nielsen and Jiang 2020), and neurological impairments (Metcalf *et al.* 2021; Méresse *et al.* 2022). In addition, microcystin poisoning can cause symptoms like vomiting, diarrhoea, abdominal pain, and liver failure, which can be fatal in severe cases (Falconer *et al.* 1983; Svirčev *et al.* 2010; Chen *et al.* 2016; Spoof and Catherine 2017).

The WHO (2022) has established a guideline value for the cyanobacterial biomass indicator and microcystin concentration in drinking water. Alert level 1 is triggered when the cyanobacteria biovolume is observed at 0.3 mm^3/L, chlorophyll-a concentration is observed at 1 µg/L, and microcystins concentration is recorded at 1 µg/L. Meanwhile, Alert level 2 is triggered when the cyanobacteria biovolume is observed >4 mm^3/L, >12 µg/L of chlorophyll-a concentration, and microcystins concentration is recorded at 12 µg/L. Hence, a developed model must predict cyanobacterial biomass and microcystin concentrations more precisely and effectively to ensure that the concentrations of cyanobacterial biomass and microcystins do not surpass the guideline values, which in turn will impact human health.

In conclusion, microcystin concentration was persistently detected at one of Malaysia's water supply systems. Microcystin concentration was recorded at the reservoir, with the highest concentration detected above 1.0 µg/L. This value has surpassed the recommended guideline value for a safe microcystin level in drinking water, as established by the WHO. However, at the water treatment plant and distribution system, the concentrations detected were below 1.0 µg/L. This indicates the effectiveness of water treatment in the study reservoir before water is distributed to the users. Although the microcystin concentration was at a safe level, prolonged consumption of contaminated water can have an impact on human health, including liver damage and inflammation.

Additionally, three unique and specific ANN models were successfully developed in this study based on local climate conditions and cyanobacterial abundance. ANN models developed in this study can be used to predict the occurrence of microcystins at the study reservoir; however, the models are site-specific and cannot predict the maximum actual concentration, potentially due to a lack of temporal robustness. These significant findings suggest that BGA-PC and chlorophyll-a concentrations directly influence the persistence of microcystins in the water supply system. Even so, more data on environmental factors are required to thoroughly study the key environmental factors that influence cyanobacterial abundance and microcystin persistence in the study reservoir. This is crucial in developing a robust ANN model that can accurately predict cyanobacterial blooms and microcystin occurrence.

Acknowledgements

The Ministry of Higher Education Malaysia funded this research under the Fundamental Research Grant Scheme 2021-0193-103-02 (FRGS/1/2021/STG03/UPSI/02/1). We would like to thank Sultan Idris Education University for its assistance in managing the grant, as well as everyone directly or indirectly involved in this research.

References

Ahmad, N.A., Sinang, S.C., 2024. Using phycocyanin as cyanobacterial biomass indicator to determine potentially- toxic bloom: an example from a Malaysia reservoir. *EnvironmentAsia.* 17, 116-130. https://doi.org/10.14456/ea.2024.40

Ahmad, N.A., Sinang, S.C., Rahman, A.R.A.A., Song, H., 2024. Cyanobacteria and microcystins contamination in drinking water sources: a review of occurrence, exposure, and mitigation measures. *Toxicol. Environ. Health Sci.* 17, 1-22. https://doi.org/10.1007/s13530-024-00239-3

Baliu-Rodriguez, D., Peraino, N.J., Premathilaka, S.H., Birbeck, J.A., Baliu-Rodriguez, T., Westrick, J.A., Isailovic, D., 2022. Identification of novel microcystins using high-resolution MS and MS n with python code. *Environ. Sci. Technol.* 56, 1652-1663. https://pubs.acs.org/doi/10.1021/acs.est.1c04296

Bui, M.H., Pham, T.L., Dao, T.S., 2017. Prediction of cyanobacterial blooms in the Dau Tieng Reservoir using and artificial neural network. *Mar. Freshwater Res.* 68, 2070-2080. https://doi.org10.1071/MF16327

Chen, L., Chen, J., Zhang, X., Xie, P., 2016. A review of reproductive toxicity of microcystins. *J. Hazard. Mater.* 301, 381-399. https://doi.org/10.1016/j.jhazmat.2015.08.041

Chorus, I., Bartram, J., 1999. *Toxic Cyanobacteria in Water: a Guide to Their Public Health Consequences, Monitoring and Management*. Spon Press: London, UK.

Chorus, I., Welker, M., 2021. *Toxic Cyanobacteria in Water: a Guide to Their Public Health Consequences, Monitoring and Management*. Taylor & Francis, Boca Raton, Florida.

Cottingham, K.L., Weathers, K.C., Ewing, H.A., Greer, M.L., Carey, C.C., 2021. Predicting the effects of climate change on freshwater cyanobacterial blooms requires consideration of the complete cyanobacterial life cycle. *J. Plankton Res.* 43, 10-19. https://doi.org/10.1093/plankt/fbaa059

Domingos, P., Rubim, T.K., Molica, R.J.R., Azevedo, S.M.F.O., Carmichael, W.W., 1999. First report of microcystin production by picoplanktonic cyanobacteria isolated from a northeast Brazilian drinking water supply. *Environ. Toxicol.* 14, 31–35. https://doi.org/10.1002/(SICI)1522-7278(199902)14:1<31::AID-TOX6>3.0.CO;2-B

Drobac, D., Tokodi, N., Kiprovski, B., Malenčić, D., Važić, T., Nybom, S., *et al.* Svirčev, Z., 2017. Microcystin accumulation and potential effects on antioxidant capacity of leaves and fruits of *Capsicum annuum*. *J. Toxicol. Environ. Health Part A.* 80, 145-154. https://doi.org/10.1080/15287394.2016.1259527

Ebrahimzadeh, G., Alimohammadi, M., Kahkah, M.R.R., Mahvi, A.H., 2021. Relationship between algae diversity and water quality-a case study: Chah Niemeh reservoir Southeast of Iran. *J. Environ. Health Sci. Eng.* 19, 437-443. https://doi.org/10.1007/s40201-021-00616-x

Falconer, I.R., 1991. Tumor promotion and liver injury caused by oral consumption of cyanobacteria. *Environ. Toxicol. & Water Qual.* 6, 177-184. https://doi.org/10.1002/tox.2530060207

Falconer, I.R., Runnegar, M.T., Beresford, A.M., 1983. Evidence of liver damage by toxin from a bloom of the blue-green alga, *Microcystis aeruginosa*. *MJA.* 1, 511-514. https://doi.org/10.5694/j.1326-5377.1983.tb136192.x

Fischer, W.J., Garthwaite, I., Miles, C.O., Ross, K.M., Aggen, J.B., Chamberlin, A.R., Towers, N.R., Dietrich, D.R., 2001. Congener-independent immunoassay for microcystins and nodularins. *Environ. Sci. & Tech.* 35, 4849-4856. https://doi.org/10.1021/es011182f

Ge, S., Qiao, X., Zhao, X., Li, X., Liu, Y., 2021. Microcystin in source water: pollution characteristics and human health risk assessment. *RSC Adv.* 11, 6415-6422. https://doi.org/10.1039/D0RA08983D

Haddix, P.L., Hughley, C.J., Lechevallier, M.W., 2007. Occurrence of microcystins in 33 US water supplies. *J. AWWA.* 99, 118-125. https://doi.org/10.1002/j.1551-8833.2007.tb08033.x

Harris, T.D., Graham, J.L., 2017. Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset. *LRM.* 33, 32-48. https://doi.org/10.1080/10402381.2016.1263694

Hartnell, D.M., Chapman, I.J., Taylor, N.G., Esteban, G.F., Turner, A.D., Franklin, D.J., 2020. Cyanobacterial abundance and microcystin profiles in two Southern British Lakes: the importance of abiotic and biotic interactions. *Toxins*. 12, 503. https://doi.org/10.3390/toxins12080503

He, L., Huang, Y., Guo, Q., Zeng, H., Zheng, C., Wang, J., Chen, J., Wang, L., Shu, W., 2018. Chronic microcystin-LR exposure induces hepatocarcinogenesis via increased gankyrin *in vitro* and *in vivo*. *Cell. Physiol. Biochem.* 49, 1420-1430. https://doi.org/10.1159/000493446

Huang, J., Xu, Q., Wang, X., Ji, H., Quigley, E.J., Sharbatmaleki, M., Li, S., Xi, B., Sun, B., Li, C., 2021. Effects of hydrological and climatic variables on cyanobacterial blooms in four large shallow lakes fed by the Yangtze River. *ESE.* 5, 100069. https://doi.org/10.1016/j.ese.2020.100069

Hurtado, I., Pouget, L., Fernández, S., Cascales, P., 2022. Monitoring and forecasting cyanobacteria risk for a drinking water plant in Spain. *WSFU.* 22, 6296-6307. https://doi.org/10.2166/ws.2022.246

Jia, X., Guan, B., Liao, J., Hu, X., Fan, Y., Li, J., Zhao, H., Huang, Q., Ma, Z., Zhu, X., Fei, M., Lu, G., Nong, Q., 2019. Down-regulation of GCLC is involved in microcystin-LR-induced malignant transformation of human liver cells. *J. Toxicol.* 421, 49-58. https://doi.org/10.1016/j.tox.2019.03.010

Jochimsen, E.M., Carmichael, W.W., An, J., Cardo, D.M., Cookson, S.T., Holmes, C.E.M., Antunes, M.B., Filho, D.A.M., Lyra, T.M., Barreto, V.S.T., Azevedo, S.M.F.O., Jarvis, W. R., 1998. Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *NEJM* 338, 873-878. https://doi.org/10.1056/nejm199803263381304

Kabziński, A.K.M., Juszczak, R., Miękoś, E., Tarczyńska, M., Sivonen, K., Ra¬pala, J., 2000. The first report about the presence of cyanobacterial toxins in Polish lakes. *Pol. J. Environ. Stud.* 9, 171.

Kaloudis, T., Zervou, S.K., Tsimeli, K., Triantis, T.M., Fotiou, T., Hiskia, A., 2013. Determination of microcystins and nodularin (cyanobacterial toxins) in water by LC-MS/MS. Monitoring of Lake Marathonas, a water reservoir of Athens, Greece. *J. Hazard. Mater.* 263, 105–115. https://doi.org/10.1016/j.jhazmat.2013.07.036

Khan , M.N., Sial, T.A., Ali, A., Wahid, F., 2024. *Impact of Agricultural Wastes on Environment and Possible Management Strategies. In: Frontier Studies in Soil Science.* Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-50503-4_4

Kubickova, B., Babica, P., Hilscherová, K., Šindlerová, L., 2019. Effects of cyanobacterial toxins on the human gastrointestinal tract and the mucosal innate immune system. *Environ. Sci. Eur.* 31, 1-27. https://doi.org/10.1186/s12302-019-0212-2

Kumar, P., Lai, S.H., Mohd, N.S., Kamal., M.R., Afan, H.A., Ahmed, A.N., Sherif, M., Sefelnasr, A., El-shafie, A., 2020. Optimised neural network model for river-nitrogen prediction utilizing a new training approach. *PLoS ONE.* 15, 1-23. https://doi.org/10.1371/journal.pone.0239509

Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. *Applied Linear Statistical Models*. McGraw-hill, New York.

Lehman, E.M., 2007. Seasonal occurrence and toxicity of Microcystis in impoundments of the Huron River, Michigan, USA. *Water Res.* 41, 795-802. https://doi.org/10.1016/j.watres.2006.09.030

Li, H., Qin, C., He, W., Sun, F., Du, P., 2021. Improved predictive performance of cyanobacterial blooms using a hybrid statistical and deep-learning method. *Environ. Res. Lett.* 16, 124045. https://doi.org/10.1088/1748-9326/ac302d

Lu, F., Chen, Z., Liu, W., Shao, H., 2016. Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecol. Eng.* 95, 422-429. https://doi.org/10.1016/j.ecoleng.2016.06.072

Machón, I., López, H., Rodríguez-Iglesias, J., Marañón, E., Vázquez, I., 2007. Simulation of a coke wastewater nitrification process using a feed-forward neuronal net. *Environ. Modell. Software.* 22, 1382-1387. https://doi.org/10.1016/j.envsoft.2006.10.001

Maier, H.R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I.N., Sànchez-Marrè, M., Acutis, M., Wu, W., Humphrey, G.B., 2023. Exploding the myths: an introduction to artificial neural networks for prediction and forecasting. *Environ. Modell. Software.* 167, 105776. https://doi.org/10.1016/j.envsoft.2023.105776

Marrone, B.L., Banerjee, S., Talapatra, A., Gonzalez-Esquer, C.R., Pilania, G., 2023. Toward a predictive understanding of cyanobacterial harmful algal blooms through AI integration of physical, chemical, and biological data. *ACS ES&T Water.* 4, 844-858. https://pubs.acs.org/doi/10.1021/acsestwater.3c00369

Massey, I.Y., Yang, F., 2020. A mini review on microcystins and bacterial degradation. *Toxins.* 12, 268. https://doi.org/10.3390/toxins12040268

Massey, I.Y., Yang, F., Ding, Z., Yang, S., Guo, J., Tezi, C., Al-Osman, M., Kamegni, R.B., Zeng, W., 2018. Exposure routes and health effects of microcystins on animals and humans: a mini-review. *Toxicon.* 151, 156-162. https://doi.org/10.1016/j.toxicon.2018.07.010

Mchau, G.J., Machunda, R., Kimanya, M., Makule, E., Gong, Y.Y., Mpolya, E., Meneely, J.P., Elliott, C.T., Greer, B., 2021. First report of the co-occurrence of cylindrospermopsin, nodularin and microcystins in the freshwaters of Lake Victoria, Tanzania. *E&H.* 13, 185–194. https://doi.org/10.1007/s12403-020-00372-7

Méresse, S., Larrigaldie, V., Oummadi, A., de Concini, V., Morisset-Lopez, S., Reverchon, F., *et al.* Mortaud, S., 2022. β-N-Methyl-Amino-L-Alanine cyanotoxin promotes modification of undifferentiated cells population and disrupts the inflammatory status in primary cultures of neural stem cells. *J. Toxicol.* 482, 153358. https://doi.org/10.1016/j.tox.2022.153358

Metcalf, J.S., Tischbein, M., Cox, P.A., Stommel, E.W., 2021. Cyanotoxins and the nervous system. *Toxins.* 13, 660. https://doi.org/10.3390/toxins13090660

Millie, D.F., Weckman, G.R., Fahnenstiel, G.L., Carrick, H.J., Ardjmand, E., Young, W.A., Sayers, M.J., Shuchman, R.A., 2014. Using artificial intelligence for cyanoHAB niche modeling: discovery and visualization of microcystis–environmental associations within western Lake Erie. *CJFAS.* 71, 1642-1654. https://doi.org/10.1139/cjfas-2013-0654

Mohamed, Z.A., 2016. Breakthrough of *Oscillatoria limnetica* and microcystin toxins into drinking water treatment plants-examples from the Nile River, Egypt. *Water SA.* 42, 161-165. https://hdl.handle.net/10520/EJC184341

Montgomery D.C., Peck E.A., Vining G.G., 2021. *Introduction to Linear Regression Analysis*. John Wiley & Sons, New Jersey.

Muhetaer, G., Asaeda, T., Jayasanka, S.M., Baniya, M.B., Abeynayaka, H.D., Rashid, M.H., Yan, H., 2020. Effects of light intensity and exposure period on the growth and stress responses of two cyanobacteria species: *Pseudanabaena galeata* and *Microcystis aeruginosa. Water.* 12, 407. https://doi.org/10.3390/w12020407

Nielsen, M.C., Jiang, S.C., 2020. Can cyanotoxins penetrate human skin during water recreation to cause negative health effects?. *Harmful Algae.* 98, 101872. https://doi.org/10.1016/j.hal.2020.101872

Nunes Carvalho, T.M., Lima Neto, I.E., Souza Filho, F.D.A., 2022. Uncovering the influence of hydrological and climate variables in chlorophyll-A concentration in tropical reservoirs with machine learning. *ESPR.* 29, 74967-74982. https://doi.org/10.1007/s11356-022-21168-z

Oliver, T.H., Gillings, S., Girardello, M., Rapacciuolo, G., Brereton, T.M., Siriwardena, G.M., Roy, D.B., Pywell, R., Fuller, R.J., 2012. Population density but not stability can be predicted from species distribution models. *J. Appl. Ecol.* 49, 581-590. https://doi.org/10.1111/j.1365-2664.2012.02138.x

Park, J., Patel, K., Lee, W.H., 2024. Recent advances in algal bloom detection and prediction technology using machine learning. *Sci. Total Environ.* 938, 173546. https://doi.org/10.1016/j.scitotenv.2024.173546

Pham, T.L., Tran, T.H.Y., Shimizu, K., Li, Q., Utsumi, M., 2021. Toxic cyanobacteria and microcystin dynamics in a tropical reservoir: assessing the influence of environmental variables. *ESPR.* 28, 63544–63557. https://doi.org/10.1007/s11356-020-10826-9

Qin, B., Zhu, G., Gao, G., Zhang, Y., Li, W., Paerl, H.W., Carmichael, W.W., 2010. A drinking water crisis in Lake Taihu, China: linkage to climatic variability and lake management. *J. Environ. Manage.* 45, 105-112. https://doi.org/10.1007/s00267-009-9393-6

Rastogi, R.P., Madamwar, D., Incharoensakdi, A., 2015. Bloom dynamics of cyanobacteria and their toxins: environmental health impacts and mitigation strategies. *Front. Microbiol.* 6, 1254. https://doi.org/10.3389/fmicb.2015.01254

Recknagel, F., Orr, P.T., Bartkow, M., Swanepoel, A., Cao, H., 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. *Harmful Algae.* 69, 18–27. https://doi.org/10.1016/j.hal.2017.09.003

Rigosi, A., Hanson, P., Hamilton, D.P., Hipsey, M., Rusak, J.A., Bois, J., Sparber, K., Chorus, I., Watkinson, A.J., Qin, B., Kim, B., Brookes, J.D., 2015. Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. *Ecol. Appl.* 25, 186-199. https://doi.org/10.1890/13-1677.1

Robert, C., Tremblay, H., DeBlois, C., 2005. Cyanobactéries et cyanotoxines au Québec: suivi à six stations de production d'eau potable (2001-2003). Service des eaux municipales, Direction générale des politiques, Développement durable, environnement et parcs Québec.

Schindler, D.W., 2006. Recent advances in the understanding and management of eutrophication. *L&O.* 51, 356-363. https://doi.org/10.4319/lo.2006.51.1_part_2.0356

Shanmugam, B.K., Vardhan, H., Raj, M. G., Kaza, M., Sah, R., Hanumanthappa, H., 2022. ANN modeling and residual analysis on screening efficiency of coal in vibrating screen. *Int. J. Coal Prep. Util.* 42, 2880-2894. https://doi.org/10.1080/19392699.2021.1910505

Shi, L., Du, X., Liu, H., Chen, X., Ma, Y., Wang, R., Tian, Z., Zhang, S., Guo, H., Zhang, H., 2021. Update on the adverse effects of microcystins on the liver. *Environ. Res*. 195, 110890. https://doi.org/10.1016/j.envres.2021.110890

Shi, X., Li, Y., Yao, B., Wang, S., Ni, S., 2025. Impact of high temporal resolution data on water quality modeling: insights from erhai case study. *Processes*. 13, 1726. https://doi.org/10.3390/pr13061726

Sinang, S.C., Reichwaldt, E.S., Ghadouani, A., 2013. Spatial and temporal variability in the relationship between cyanobacterial biomass and microcystins. *Environ. Monit. Assess.* 185, 6379-6395. https://doi.org/10.1007/s10661-012-3031-0

Sivonen, K., Jones, G., 1999. Cyanobacterial toxins. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management, 1, 43-112.

Spoof, L., Catherine, A., 2017. Appendix 3: Tables of Microcystins and Nodularins, in: Meriluoto, J., Spoof, L., Codd, G.A. (Eds.), *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis*. John Wiley & Sons, Ltd., Chichester, West Sussex, pp. 526-537. https://doi.org/10.1002/9781119068761.app3

Srisuksomwong, P., Pekkoh, J., 2020. Artificial neural network model to prediction of eutrophication and microcystis aeruginosa bloom. *Emerg. Sci. J.* 4, 129-135. https://doi.org/10.28991/esj-2020-01217

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *JMLR.* 15, 1929-1958. http://jmlr.org/papers/v15/srivastava14a.html

Svirčev, Z., Baltić, V., Gantar, M., Juković, M., Stojanović, D., Baltić, M., 2010. Molecular aspects of microcystin-induced hepatotoxicity and hepatocarcinogenesis. *J. Environ. Sci. Health Part C.* 28, 39-59. https://doi.org/10.1080/10590500903585382

Turner, A.D., Dhanji-Rapkova, M., O'Neill, A., Coates, L., Lewis, A., Lewis, K., 2018. Analysis of microcystins in cyanobacterial blooms from freshwater bodies in England. *Toxins.* 10, 39. https://doi.org/10.3390/toxins10010039

Vieira, J.M.D.S., Azevedo, M.T.D.P., de Oliveira Azevedo, S.M.F., Honda, R.Y., Corrêa, B., 2005. Toxic cyanobacteria and microcystin concentrations in a public water supply reservoir in the Brazilian Amazonia region. *Toxicon*. 45, 901-909. https://doi.org/10.1016/j.toxicon.2005.02.008

Wang X., Ji Y., Li X. 2011. Research on the prediction of water treatment plant coagulant dosage based on feed-forward artificial neutral network, in: 2*011 International Conference on Consumer Electronics, Communications and Networks (CECNet)*. IEEE, Xianning, pp. 16-18. https://doi.org/10.1109/CECNET.2011.5769402

Wiegand, M.C., do Nascimento, A.T.P., Costa, A.C., Neto, I.E.L., 2021. Trophic state changes of semi-arid reservoirs as a function of the hydro-climatic variability. *J. Arid Environ.* 184, 104321. https://doi.org/10.1016/j.jaridenv.2020.104321

[WHO] World Health Organization. 2022. Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum. World Health Organization, Geneva. Available at: https://www.who.int/publications/i/item/9789240045064. [Date acceseed: 24 September 2024]

Ying X. 2019. An overview of overfitting and its solutions. *J. Phys.: Conf. Ser.* 1168, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022