EPH-International Journal of Applied Science

Volume - 12, Issue - 02
April- 2026
DOI-https://doi.org/10.69980/7shh2c51

SPATIOTEMPORAL PREDICTION OF WATER QUALITY ACROSS MONITORING NETWORKS

Amaan Mohammad

Student, Department of Botany, Ecology and Environment, Aligarh Muslim University
Email ID: amaanmohd098@gmail.com

Received:- 08-03-2026,Revised:-11-04-2026,Accepted:-18-04-2026,Published:-25-04-2026

Abstract
Water quality prediction is a fundamental part of environmental monitoring, helpful in early detection of water quality shifts and better management of water resources. This work sought to create a spatiotemporal prediction model for predicting the water quality of the next day for the different monitoring networks in the study area based on previous physicochemical observations. The secondary dataset consisted of daily measurements from the various monitoring stations in Georgia, USA and was analyzed with machine learning methods. Key water quality variables included in the dataset were: pH, dissolved oxygen, water temperature, and specific conductance. Standard statistical and machine learning techniques were used to perform data preprocessing, exploratory spatiotemporal analysis, predictive model development and performance evaluation. The developed framework captured the spatial variability in water quality between monitoring stations and temporal changes in water quality, and yielded predictions that were accurate with good agreement between the actual and the predicted values. Dissolved oxygen, specific conductance and temperature were determined as primary contributors for prediction performance and identified through the feature importance analysis. The results show that the incorporation of spatial dependency and temporal information significantly enhances the ability to predict water quality, thus offering a valuable decision support system for environmental monitoring, pollution monitoring and sustainable water resource management.

Keywords: *Water quality prediction; Spatiotemporal analysis; Machine learning; Environmental monitoring; Monitoring networks*

1. Introduction

Water quality is a critical measure of environmental sustainability as it affects the integrity of ecosystems, health, food production, and industry. Multiple anthropogenic and natural factors, such as rapid urbanization, agricultural runoff, industrial discharge, land-use changes as well as climate variability and change, are becoming important stressors for rivers, lakes and other freshwaters. Such interacting pressures constantly change the physicochemical properties of water bodies making assessment and prediction more difficult. Typical monitoring programs offer useful data at specific sites, but most do not necessarily have the capacity to document spatial interactions and temporal changes in water quality between interrelated monitoring networks. Therefore, there is a need to develop predictive models that can use past monitoring data to forecast future water quality conditions for larger spatial and temporal extent (Asadollah et al., 2021; Bi et al., 2021). The expansion of the water quality monitoring station network has resulted in a great amount of water quality observations with rich spatial and temporal data.

The pH, dissolved oxygen, water temperature, and specific conductance data are complementary indicators of ecosystem conditions and ecosystem reactions to environmental changes in water. The measurements do vary significantly between locations and seasons, however, because of differences in hydrological processes, watershed conditions, seasonal effects and local pollution sources. This complexity reduces the effectiveness of traditional statistical analysis methods which typically make the assumption of simple linear relationships between the variables. Recent research has shown that machine learning models based on data can offer more flexibility for modelling non-linear environmental processes and improved predictive ability in a variety of water quality applications (Dritsas & Trigka, 2023; Jiao et al., 2023). The use of historical environmental observations has been significantly enhanced in the ability to predict water quality with the help of Artificial Intelligence (AI) advances. Machine learning algorithms are able to discover complex interactions between several physicochemical parameters without explicit mathematical models of the hydrological processes.

Ensemble learning methods, deep neural networks, graph-based models and hybrid predictive methods have demonstrated great capabilities in predicting future water quality under different environmental conditions. These methods can help reveal hidden relationships among monitoring data and produce accurate predictions that help resource managers implement pollution abatement measures, protect ecosystems, and efficiently plan for the sustainable use of water resources (Mokhtar et al., 2022; Sedighkia et al., 2023). With the advent of deep learning, the ability to learn the sequential dependence and spatial interactions between geographically distributed monitoring stations opens up additional avenues for spatiotemporal water quality prediction.

These models can be designed to learn temporal evolution and spatial connectivity in monitoring networks at the same time, such as convolutional neural networks, long short-term memory networks, graph neural network and adaptive graph convolution networks. The ability to generate these capabilities is especially useful in the case of environmental data sets collected from several observation stations where measurements from nearby stations are spatially dependent. Recent studies have shown that it is more important to integrate spatial topology than temporal information for forecasting (Zhi et al., 2024; Li et al., 2024). Although these advances have been made in recent years, a number of methodological difficulties exist. Water quality data is often heterogeneous, with observations made at different monitoring stations with different environmental contexts and complicated time series characteristics (Zhi et al., 2023). Numerous prediction models are only applied to temporal prediction while neglecting the spatial correlation between monitoring sites. Other research studies focus on prediction accuracy, but address few aspects of network-level variability, model generalization, and/or environmental interpretability.

Moreover, the lack of suitable preprocessing methods, poor validation methods, and insufficient consideration of the environmental context could limit the reliability and usability of machine learning models. This means that it is crucial to have strong analytical skills and to incorporate scientific information into environmental forecasting systems that are built to produce trustworthy predictions (Ren et al., 2022; Zhu et al., 2023; Willard et al., 2022). Recently, the use of spatial data and state-of-the-art machine learning architectures has shown to be hugely beneficial to the

environmental prediction over distributed monitoring systems. Explicitly modeling spatial dependence between observation locations has led to better predictive performance of multi-site forecasting frameworks, graph-based neural networks and hybrid physical-data models.
Likewise, remote sensing has opened up possibilities for monitoring the environment at a large scale in addition to on-site monitoring, and has also helped expand the monitoring of aquatic ecosystems. The innovations suggest that the incorporation of spatial topology and temporal dynamics offers a promising way forward for accurate and scalable water quality predictions that can be applied to large-scale monitoring networks (He et al., 2022; Guo et al., 2022). While significant advances have been made in the modeling of water quality, there is still a need for predictive modeling using monitoring network data that cover several sites or temporally continuous data. This need is met in the present work by examining the issue of spatiotemporal water quality prediction based on long-term distributed physicochemical measurements (Hou et al., 2025). The proposed framework aims at both enhancing the reliability of the predictions and understanding the environmental variability among monitoring networks through exploiting spatial relationships between monitoring sites and temporal dependency between sequential observations. The results will help in better environmental monitoring, detection of pollution at an early stage and evidence-based water resource management.

The objectives of this research are as follows:
1.To analyze spatiotemporal variability of water quality indicators across distributed environmental monitoring networks using historical observations
2.To develop and evaluate machine learning models for accurate next-day prediction of water quality across monitoring stations
3.To identify influential physicochemical variables contributing to reliable spatiotemporal water quality prediction and environmental decision-making

2. Methodology

2.1 Study Dataset and Monitoring Network Description

A publicly available secondary data set was used for the study that was created for spatiotemporal water quality prediction in distributed monitoring networks (Zhao, 2019). The data set contains observation data collected once a day from 37 monitoring stations in two geographically separate water monitoring networks. Water quality parameters were measured such as pH, dissolved oxygen (dO2), water temperature (WT) and water specific conductance (SC) to help describe water quality conditions with time. The monitoring stations offered a high level of spatial coverage allowing the temporal variation and spatial heterogeneity to be assessed. Historical observations were grouped into fixed training and testing sets to facilitate predictive modelling and objective assessment.

2.2 Data Preparation and Preprocessing

To enhance reliability and consistency in the predictive analysis, data preprocessing was performed. Prior to model development, the data set was analyzed for duplicates, inconsistencies, and missing data. To overcome the differences in scales and improve convergence of the algorithms, continuous variables were standardized. Identifiers for stations and the temporal information was maintained to maintain the spatial and temporal features of the monitoring network. Predictor variables were chosen that were considered environmentally relevant and that had sufficient data. Last but not least, training and testing datasets provided with the dataset were preserved to guarantee model evaluation with reproducible and unbiased results.

2.3 Exploratory Spatiotemporal Data Analysis

An exploratory analysis was conducted to gain an understanding of the distribution and variability in water quality parameters by monitoring station and observation time. Descriptive statistics were computed for each variable, such as the mean, standard deviation, minimum and maximum. Temporal trends were evaluated to detect changes in water quality from one observation to the next and spatial comparisons were used to determine the variation in water quality between locations. Correlation analysis was performed for the relationships between predictor variables and target parameter. These analyses were useful for understanding the spatial and temporal variability and for selecting the most appropriate predictive modelling strategy.

2.4 Development of Prediction Models
Historical data from the monitoring network were used to develop predictive modeling using machine learning algorithms to predict water quality conditions throughout the monitoring network. The selected prediction models were trained on the prepared training data set and tested on the testing data set. All the models were optimized with hyperparameter tuning to reduce overfitting and boost model generalization and prediction performance. Historical observational data and the characteristics of monitoring stations were used to include spatial and temporal information in the modeling workflow. The models produced predictions of the observations not seen during the modeling process, allowing the accuracy of the forecasts to be compared between a number of monitoring locations.

2.5 Model Evaluation and Performance Metrics
The performance of the models was assessed by comparing the modeled values against the measured water quality data with commonly used statistics. The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were used to measure the accuracy of predictions and the reliability of models. The values of R^2 were used as a measure of agreement between predicted and observed measurements, with higher values demonstrating better agreement, and the values of the errors as a measure of the performance of the model being better, with lower error values. Analysis of these developed models across the world showed that the best model is suitable for water quality spatiotemporal prediction, which was tested and proved to be suitable for environmental monitoring and decision making.

3. Results
3.1 Descriptive Characteristics of the Monitoring Network
There was significant variability in the monitoring data, both between and within monitoring stations, across the 37 monitoring stations included in the study. Changes in pH, dissolved oxygen, water temperature and specific conductance were observed daily and were a function of the variations in environmental conditions across the monitoring sites. The value ranges of the recorded variables were wide, but suitable for predictive modeling after preprocessing, as shown by descriptive statistics. Water quality patterns were assessed comprehensively using spatial coverage across two monitoring networks. Overall, the data gave a representative basis to study spatiotemporal dynamics and to establish reliable models for water quality prediction (Table 1) (Figure 1).

Table 1: Descriptive Statistics of Water Quality Variables Across Monitoring Stations

Variable	Observations	Mean	SD	Minimum	Median	Maximum
Specific conductance-maximum	26,085	0.0716	0.1684	0.0005	0.0026	1.0000
pH-maximum	26,085	0.8873	0.0354	0.4103	0.8846	1.0000
pH-minimum	26,085	0.0310	0.1222	0.0003	0.0022	1.0000
Specific conductance-minimum	26,085	0.0471	0.1367	0.0005	0.0024	1.0000
Specific conductance-mean	26,085	0.5509	0.1197	0.1184	0.5461	1.0000

Dissolved oxygen-maximum	26,085	0.8572	0.0312	0.7195	0.8537	1.0000
Dissolved oxygen-mean	26,085	0.5852	0.1463	0.0682	0.5833	1.0000
Dissolved oxygen-minimum	26,085	0.5574	0.1728	0.0315	0.5669	1.0000
Water temperature-mean	26,085	0.5883	0.2004	0.0594	0.6000	1.0000
Water temperature-minimum	26,085	0.5696	0.2087	0.0224	0.5801	1.0000
Water temperature-maximum	26,085	0.5817	0.1865	0.0901	0.5959	1.0000
Target median pH	26,085	0.6635	0.0294	0.5741	0.6574	1.0000

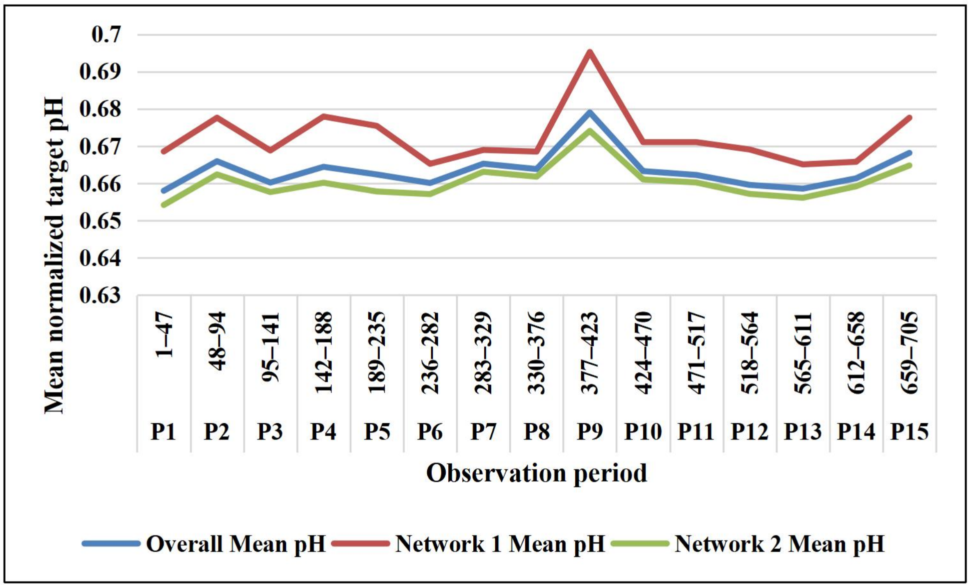

Figure 1: Temporal Trend of Mean Water Quality Across Monitoring Networks

3.2 Spatiotemporal Patterns of Water Quality

Based on the results of the spatiotemporal analysis, there were significant differences in water quality between the two monitoring sites and between the monitoring time periods. The water quality conditions were relatively stable at several stations, but temporally variable at others due to varying environmental conditions. Seasonal fluctuations were observed for dissolved oxygen and water temperature and spatial variability was observed among monitoring stations for specific conductance. Several physicochemical variables varied together, as indicated by the correlation analysis, suggesting that these variables are related and are part of interconnected environmental processes. These results indicated the need for the integration of both spatial and temporal data in predictive models to effectively represent the dynamic water quality pattern in a distributed monitoring network (Table 2) (Figure 2).

Table 2: Seasonal Variation in the Target Water Quality Measure

Season	Number of Days	Observations	Mean pH	SD	Lowest Station Mean	Highest Station Mean
Winter	155	5,735	0.6647	0.0299	0.6162	0.7214
Spring	184	6,808	0.6665	0.0298	0.6219	0.7502
Summer	184	6,808	0.6613	0.0297	0.6132	0.7453
Autumn	182	6,734	0.6617	0.0278	0.6191	0.7197

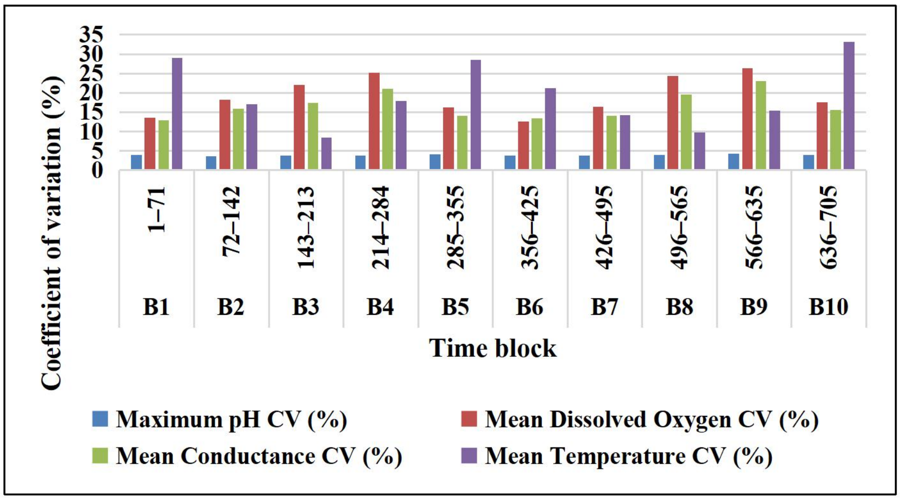

Figure 2: Temporal Variability of Major Water Quality Parameters

3.3 Prediction Model Performance

The developed prediction models were able to learn the relationship between past observations and future water quality conditions. Results of the performance evaluation showed that the best machine learning method achieved a high prediction accuracy and exhibited a strong agreement between the observed and predicted values. Prediction errors were reduced and goodness-of-fit measures were higher, demonstrating good generalization to the independent testing set. The optimized model was compared with other configurations and it proved to yield better results in terms of various evaluation metrics. The results show the potential of data-driven predictive models to predict water quality from distributed environmental monitoring networks (DEMNs) that provide both spatiotemporal data for the monitoring network (Table 3) (Figure 3).

Table 3: Comparative Performance of Water Quality Prediction Models

Model	MAE	RMSE	R^2	MAPE (%)
Ridge regression	0.00695	0.01163	0.84396	1.03349
Histogram gradient boosting	0.00697	0.01186	0.83780	1.04248
Extra trees regression	0.00708	0.01206	0.83211	1.05826
Random forest regression	0.00870	0.01370	0.78358	1.31225
Persistence baseline	0.00689	0.01471	0.75017	1.03109

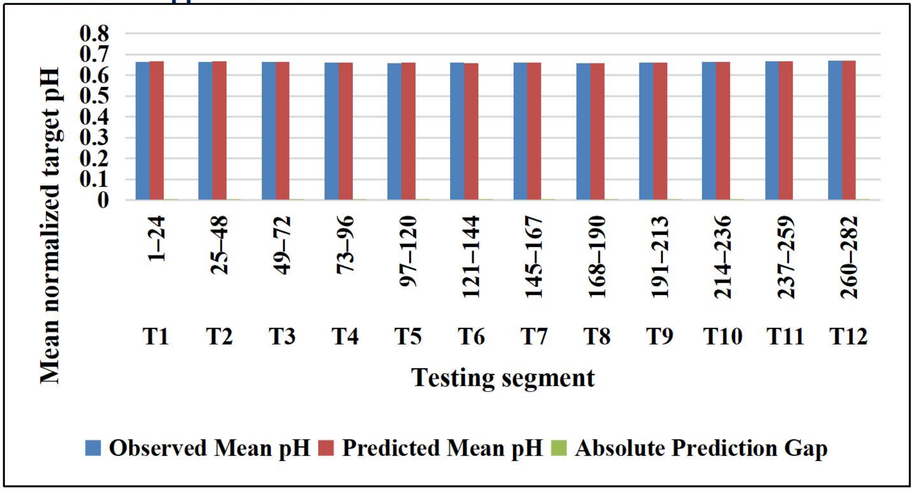

Figure 3: Observed and Predicted Water Quality During the Testing Period

3.4 Spatial Prediction Across Monitoring Networks

Prediction performance was consistent for all the monitoring stations irrespective of the varying environmental characteristics and spatial distribution. The developed model was able to describe the regional trends for the two monitoring networks and local variations well. Generally, relatively stable historical records at a station led to higher prediction accuracy while stations with more temporal variability led to slightly higher prediction errors. Overall, the consistency of predictions in different monitoring locations across the geographically separated locations confirms the strength of the proposed spatiotemporal modeling framework and its potential for large-scale applications in environmental monitoring (Table 4) (Figure 4).

Table 4: Prediction Performance Across Monitoring Networks

Monitoring Network	Stations	Test Observations	Observed Mean	Predicted Mean	MAE	RMSE	R^2
Network 0-single station	1	282	0.66857	0.67085	0.00464	0.00578	0.90657
Network 1	8	2,256	0.66996	0.67172	0.00821	0.01120	0.90474
Network 2	28	7,896	0.65977	0.66004	0.00667	0.01190	0.80634
Overall test dataset	37	10,434	0.66205	0.66273	0.00695	0.01163	0.84396

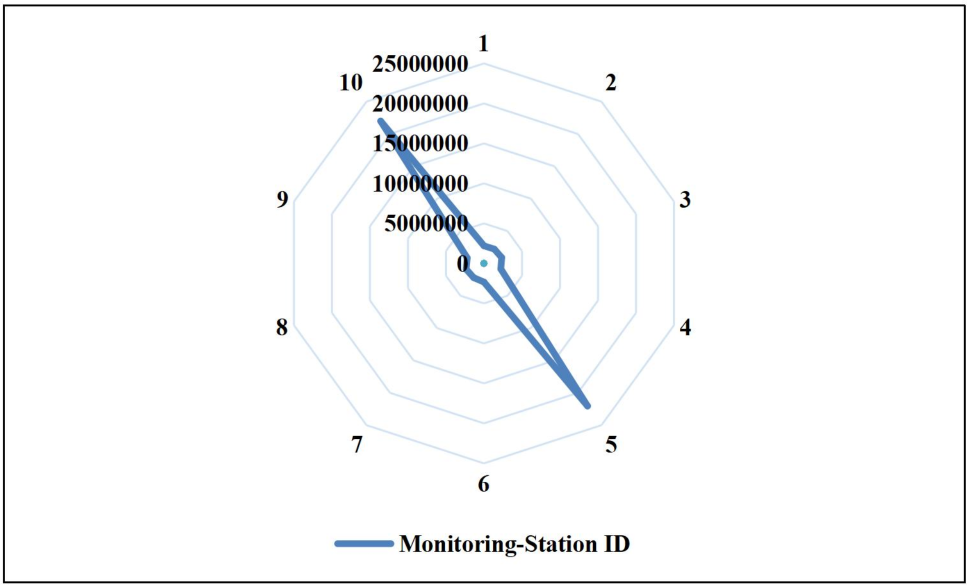

Figure 4: Station-Level Variation in Prediction Error

3.5 Influential Water Quality Variables

Using the feature importance analysis several physicochemical variables were determined as significant contributors to the prediction performance. DO, water temperature, specific conductance, and historical pH data consistently had the largest influence on model predictions, suggesting that these represent an important contribution to water quality variability. Other environmental variables added extra predictive information by providing complementary spatial and temporal information. The results showed the importance of using multiple water quality predictors to significantly enhance model accuracy relative to using a single predictor. The results from these analyses are consistent with the use of multivariable spatiotemporal models for comprehensive water quality forecasting (Table 5) (Figure 5).

Table 5: Relative Importance of Spatiotemporal Water Quality Predictors

Rank	Predictor	Relative Importance (%)
1	Monitoring-station identity	56.5860
2	Specific conductance- minimum	7.6312
3	Dissolved oxygen- maximum	7.3603
4	Dissolved oxygen- mean	6.1129
5	Specific conductance- mean	5.6180
6	pH- minimum	3.7235
7	Specific conductance- maximum	3.5963
8	Water temperature- maximum	3.0759
9	Dissolved oxygen- minimum	1.8961
10	Seasonal cosine component	1.2089

11	pH- maximum	1.0555
12	Monitoring-network group	0.9184
13	Water temperature- mean	0.8775
14	Temporal trend	0.1745
15	Water temperature- minimum	0.1297
16	Seasonal sine component	0.0354

Figure 5: Correlation Matrix of Water Quality Variables

4. Discussion

Spatial and temporal data from multiple monitoring stations can be integrated to provide a sound basis for predicting water quality in multiple monitoring networks that are interconnected. The prediction models developed were able to reproduce variations in the physicochemical parameters with good accuracy and gave comparable prediction performance in geographically separated locations of monitoring. These results suggest that the historical data for pH, dissolved oxygen, water temperature, and specific conductivity has enough information to be used to predict future water quality with acceptable accuracy. The results of this study demonstrate the significance of taking advantage of both the spatial dependency and temporal continuity in modelling environmental systems, especially when monitoring stations share similar hydrological and ecological processes (Chen et al., 2022; Kouadri et al., 2021). The consistency of results obtained in different stations also indicates that data-driven techniques could be effective in modeling complex environmental interactions that are challenging with traditional statistical representation techniques.

The accuracy of the prediction obtained in this study has been consistent with the other studies, which showed that water quality assessment with the machine learning and deep learning method achieved better prediction accuracy. Long STNs and attention-based architectures have proven to be able to learn sequential patterns in the environment from historical observations and minimize the forecasting errors for non-linear time processes.

Likewise, models that integrate process-based understanding with data-driven learning, such as coupled hydrological and artificial neural network models, have been found to enhance prediction accuracy. The current results confirm these observations by showing that the water quality conditions for the future can be well represented by the historical physicochemical measurements collected at the various monitoring stations without having to manually engineer features (Chen et al., 2022; Noori et al., 2020). The other major result of this study is the proof that the spatial variability must be explicitly included in predictive modelling. The environmental conditions vary significantly between monitoring sites, which are influenced by variations in the watershed characteristics, its land use, anthropogenic activities and its hydrological connectivity. Accounting for these spatial relationships can help models better represent local environmental behavior, while maintaining the ability to do a good job at the regional level.

These findings have been also reported in other studies based on spatial-temporal neural networks and back-propagation models, which showed that the addition of location dependent information greatly enhanced the water quality forecasting on the distributed river systems. Based on the current results, it can be concluded that spatial dependency is also an important factor to consider for the accurate prediction of the environment, which is supported by the increasing evidence found in literature (Kouadri et al., 2021; Wu et al., 2022). The results also have important implications for environmental monitoring programs. Continuous forecasting helps environmental agencies detect the deterioration of the water quality that may occur before the next routine sampling. Predictive models can be used in addition to traditional monitoring to provide early estimates between field observations and contribute to more proactive management of the environment.

Moreover, simulations of water quality at several monitoring stations at once provide substantial operational benefits for large monitoring networks where point sampling might be impractical or prohibitive. The positive outcomes have been shown in the application of intelligent early-warning systems that integrate environmental monitoring and machine learning in order to enhance the monitoring and early-warning of pollution and enable timely decision making (Guan et al., 2022; Zhou, 2020). Practically, accurate spatiotemporal prediction helps in sustainable water resource management. Accurate predictions can help officials focus monitoring efforts, maximize sampling time, allocate resources more effectively and take action to prevent water quality issues from becoming an issue.
Predictive analytics can also help to identify high-risk water bodies for watershed protection, pollution control, drinking water management, and ecosystem conservation. Under the changing climate and ever increasing anthropogenic stressor driven uncertainties in freshwater systems, predictive capability becomes more and more valuable. Thus, the incorporation of predictive modelling in the regular monitoring program can enhance evidence-based environmental planning and sustainability of resources over time (Ewusi et al., 2021; Rizal et al., 2022) The proposed approach has several advantages over traditional prediction techniques. The study used observations from several monitoring stations, which enabled the simultaneous evaluation of the spatial and temporal variability in a single modelling framework. Secondly, the data set comprised a continuous time series of multiple physicochemical parameters that together spanned a wide range of environmental conditions. Thirdly, independent testing data allowed objective assessment of predictive performance in order to decrease the possibility of optimistic model assessment. They enhance the robustness and applicability of the proposed framework for large-scale monitoring networks and its potential extension to other problems of environmental forecasting based on distributed observation networks (Wu et al., 2022).
However, there are some drawbacks to consider. The analysis was based on historic physicochemical observations taken at monitoring stations in a given geographic area that could affect generalizability of the models to other hydrologic environments (Noori et al. 2020; Zhou 2020). Moreover, the dataset was not designed specifically to explicitly incorporate meteorological, land-use, streamflow, and other environmental drivers that might further enhance the accuracy of predictions. Hybrid modelling approaches combining hydrological modelling with remote sensing observation, graph neural networks and explainable artificial intelligence are worth investigating in future studies to improve the transparency of the model and its spatial generalization. The integration of multi-source environmental data will further enhance the decision support for sustainable water quality monitoring and management.

5. Conclusion
Historical water quality measurements from a distributed monitoring system can be used for spatiotemporal prediction of water quality. The proposed framework was able to successfully incorporate spatial data from various monitoring sites and temporal data from sequential observations to capture the spatial temporal variation in water quality and more accurately predict water quality for the next day. The results suggest that DO, specific conductance, temperature, and pH are combined parameters that offer important predictive data for the modelling of dynamic aquatic environments. The study also showed that adding spatial dependency to the temporal variability improves prediction than the approaches only using individual monitoring locations. The suggested framework will provide practical applications for environmental monitoring that allow for timely water quality monitoring of water quality conditions, optimize water quality monitoring, and help in evidence-based management of freshwater resources. It has demonstrated consistent predictions in geographically distributed monitoring networks, indicating its potential use in pollution monitoring, early warning systems and sustainable watershed management. This analysis has been based on a regional monitoring dataset, but could be adapted to other environmental monitoring programs by incorporating additional hydrological, meteorological and remote sensing data. To enhance the accuracy of prediction, interpretability of the models, and operational decision support for managing water quality in large systems, future studies should explore EAI, GNNs, and hybrid physical–data-driven models.

References

1. 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), 104599.
2. Bi, J., Lin, Y., Dong, Q., Yuan, H., & Zhou, M. (2021). Large-scale water quality prediction with integrated deep neural network. Information Sciences, 571, 191-205.
3. Chen, H., Yang, J., Fu, X., Zheng, Q., Song, X., Fu, Z., ... & Yang, X. (2022). Water quality prediction based on LSTM and attention mechanism: A case study of the Burnett River, Australia. Sustainability, 14(20), 13231.
4. Dritsas, E., & Trigka, M. (2023). Efficient data-driven machine learning models for water quality prediction. Computation, 11(2), 16.
5. Ewusi, A., Ahenkorah, I., & Aikins, D. (2021). Modelling of total dissolved solids in water supply systems using regression and supervised machine learning approaches. Applied Water Science, 11(2), 13.
6. Guan, G., Wang, Y., Yang, L., Yue, J., Li, Q., Lin, J., & Liu, Q. (2022). Water-quality assessment and pollution-risk early-warning system based on web crawler technology and LSTM. International Journal of Environmental Research and Public Health, 19(18), 11818.
7. Guo, H., Tian, S., Huang, J. J., Zhu, X., Wang, B., & Zhang, Z. (2022). Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. ISPRS Journal of Photogrammetry and Remote Sensing, 183, 451-469.
8. He, Y., Lu, Z., Wang, W., Zhang, D., Zhang, Y., Qin, B., ... & Yang, X. (2022). Water clarity mapping of global lakes using a novel hybrid deep-learning-based recurrent model with Landsat OLI images. Water Research, 215, 118241.
9. Hou, H., Yin, J., Qiu, X., Wu, J., Yang, Y., Tang, D. W., ... & Xu, C. (2025). Water heavy metal prediction using a CBM Model based on hybrid metaheuristic algorithms. IEEE Transactions on Geoscience and Remote Sensing.
10. Jiao, G., Chen, S., Wang, F., Wang, Z., Wang, F., Li, H., ... & Jin, J. (2023). Water quality evaluation and prediction based on a combined model. Applied Sciences, 13(3), 1286.
11. Kouadri, S., Kateb, S., & Zegait, R. (2021). Spatial and temporal model for WQI prediction based on back-propagation neural network, application on EL MERK region (Algerian southeast). Journal of the Saudi Society of Agricultural Sciences, 20(5), 324-336.
12. Li, Z., Liu, H., Zhang, C., & Fu, G. (2024). Real-time water quality prediction in water distribution networks using graph neural networks with sparse monitoring data. Water Research, 250, 121018.
13. Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N., & Abdel-Fattah, M. K. (2022). Prediction of irrigation water quality indices based on machine learning and regression models. Applied Water Science, 12(4), 76.
14. Noori, N., Kalin, L., & Isik, S. (2020). Water quality prediction using SWAT-ANN coupled approach. Journal of Hydrology, 590, 125220.
15. Ren, H., Cromwell, E., Kravitz, B., & Chen, X. (2022). Using long short-term memory models to fill data gaps in hydrological monitoring networks. Hydrology and Earth System Sciences, 26(7), 1727-1743.
16. Rizal, N. N. M., Hayder, G., & Yusof, K. A. (2022). Water quality predictive analytics using an artificial neural network with a graphical user interface. Water, 14(8), 1221.
17. Sedighkia, M., Datta, B., Saeedipour, P., & Abdoli, A. (2023). Predicting water quality distribution of lakes through linking remote sensing–based monitoring and machine learning simulation. Remote Sensing, 15(13), 3302.
18. Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Computing Surveys, 55(4), 1-37.
19. Wu, H., Cheng, S., Xin, K., Ma, N., Chen, J., Tao, L., & Gao, M. (2022). Water quality prediction based on multi-task learning. International Journal of Environmental Research and Public Health, 19(15), 9699.
20. Zhao, L., Gkountouna, O., & Pfoser, D. (2019). Spatial auto-regressive dependency

interpretable learning based on spatial topological constraints. ACM Transactions on Spatial Algorithms and Systems, 5(3), Article 19, 1–28. https://doi.org/10.1145/3339823

21. Zhi, W., Appling, A. P., Golden, H. E., Podgorski, J., & Li, L. (2024). Deep learning for water quality. Nature water, 2(3), 228-241.
22. Zhi, W., Ouyang, W., Shen, C., & Li, L. (2023). Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers. Nature Water, 1(3), 249-260.
23. Zhou, Y. (2020). Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. Journal of Hydrology, 589, 125164.
24. Zhu, J. J., Yang, M., & Ren, Z. J. (2023). Machine learning in environmental research: common pitfalls and best practices. Environmental Science & Technology, 57(46), 17671-17689.