

# **IDEATION PHASE**

## **Literature survey**

Modeling the quality of water resources is vitally important for water scheduling and management. In the past, scientists regularly sampled the water in water quality monitoring stations and assessed the components in the water sample in a lab. However, this process takes a long time, and thus, the detected results are not timely. With the emergence of artificial intelligence (AI) techniques since the last decade, researchers have begun to adopt multivariate linear regression (MLR), artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), and Fuzzy time series (FTS) model to predict water quality by exploring the linear and non-linear relationships residing in water quality datasets. In addition, the wavelet denoising method and intelligent algorithms are also proposed to combine with machine learning techniques to enhance the prediction accuracy. In the following, we will review these related work in four categories of machine learning methods.

### **1. MLR**

MLR is a kind of statistical analysis method which is used to estimate the target value based on given values collected from a set of independent variables. It is adopted to predict the water quality because of its speed and simplicity. In [3], the MLR model is used to predict biochemical oxygen demand (BOD) and chemical oxygen demand base on four independent variables, temperature, pH, total suspended solid, and total suspended. The system quickly receives relatively good result in BOD prediction with a correlation coefficient value of 0.5. MLR model has also been used to predict the water quality index in [10] and found to be reliable in formulating the relationship excluding the parameter chloride. However, the MLR model can only be used to formulate linear relationship. It is likely to have a large prediction error if the MLR model is used to predict non-linear relationship. .

### **2. ANN**

Various ANN models have been designed to predict water and wastewater discharge quality based on previous existing datasets. A two-layer ANN model has been applied to predict the DO concentration in the Mathura River [11], and the experimental result showed that the ANN model worked well. In [12], various neural network types are compared in predicting water temperatures in streams. A radial basis function neural network has also been proposed to describe the water quality parameters in [13]. The summary of the experiment result shows the model outperforms the linear regression

model in conductivity, turbidity, and total dissolved solids prediction. A time series prediction model, namely the autoregressive integrated moving average, was integrated with the ANN model to improve the prediction performance. The experimental results showed that the hybrid model provided better accuracy than ARIMA and ANN models [14]. Additionally, a comprehensive comparison between ANN and MLR models in biochemical oxygen demand and chemical oxygen demand prediction has been performed [3]. The experimental results show that a three-layer neural network model outperforms an MLR model. The other comparison between ANN and MLR models in water quality index prediction furtherly proves that the ANN model is a better option [10]. Although ANN models can effectively improve the prediction accuracy of water quality parameters, shortcomings still exist. Especially in some scenarios where the input parameters are ambiguous, neural networks struggle to formulate a non-linear relationship. In [15], wavelet transformation was applied to the ANN model to improve the prediction accuracy of a variety of ocean water quality parameters. An integration of a particle swarm optimization algorithm with ANN models has also been investigated to improve the forecasting performance [16]. In [17], 120 data samples, collected from 2002 to 2012, are used to verify whether the integration of fuzzy logic and ANN models can improve the water quality prediction performance. The experimental results confirm that the proposed method works.

### **3 .ANFIS**

Many studies have proven that ANFIS, which can integrate linear and non-linear relationships hidden in the dataset, is a better option in this scenario [5]. The experimental results in [6] show that an ANFIS model works much better than an ANN model in predicting dissolved oxygen, even though there are only 45 data samples available. An ANFIS model with eight input parameters is used to predict total phosphorus and total nitrogen, the experiment result based on 120 water samples shows the proposed model is reliable [18]. The ANFIS model has also been applied to estimate the biochemical oxygen demand in the Surma River [19]. The testing results from 36 water samples confirmed that the ANFIS model could accurately formulate the hidden relationship and correlation analysis can improve the prediction accuracy. Two different kinds of ANFIS model, fuzzy c-means and subtractive clustering-based was compared in [20], the experiment result shows the ANFIS model built by fuzzy c-means provides more accurate prediction result. In [21], the ensemble models of wavelet ANNs are found to be superior to the best single model for forecasting chlorophyll and salinity concentrations in coastal water. An ensemble of ANN and ANFIS is proposed in [22] to improve the prediction performance of the ANN and ANFIS model, the test result shows there is a significant improvement in the Ensemble ANN-ANFIS model. According to the developer of the ANFIS model, the size of the training dataset should be no less than

the number of training parameters [23]. In the aforementioned papers, though the ANFIS models have received higher prediction accuracy, the sizes of the training datasets are data have a large value range and there exist some extreme data value points, an out-of-range error is likely to occur, which happens when the testing dataset cannot find any insight from the training model. A few out-of-range errors can cause a very large prediction error, even though the model can accurately predict most of the data samples. In [24], a dataset collected from 122 wells in Mashhad plain (Iran) is used to investigate the performance of ANFIS, ANN, and geostatistical models in groundwater quality prediction. The experimental result shows that the ANFIS model has poor performance in the testing stage because the limited training dataset cannot build a robust or reliable model. Recently, a few researchers have tried to integrate a machine learning model with a wavelet de-noising technique to improve prediction accuracy. Wavelet support vector regression and wavelet artificial neural networks have been proposed to model monthly pan evaporation [25]. The experimental results confirm that wavelet artificial neural networks outperform other models. An integrated wavelet de-noising ANFIS model was proposed to predict electrical conductivity (EC) and total dissolved solids (TDS) in [26]. Although the size of the dataset was smaller than the requirement, the model still achieved good prediction performance. In [27], eight different kinds of membership functions, with different wavelet de-noising schemes, were investigated to improve the performance of an ANFIS model. Based on the above two research studies, an optimized wavelet-ANFIS model is proposed in [28] and the experimental results show that a bell-shaped membership function with random sampling has the best prediction performance. In [29], a wavelet-ANFIS model is proposed to predict the groundwater level. Compared to ARIMA and ANFIS, the proposed model provides a more precise prediction result. A comparative study of different wavelet-based ANN models to predict sewage sludge quantity is given in [30], the experiment result also proves wavelet can improve the accuracy to the ANN models. On the other hand, many researchers have also tried to integrate intelligence algorithms with the ANFIS model to improve the performance of the proposed model. An application of genetic algorithm (GA), ant colony optimization for continuous domains, and differential evolution is introduced in [31] to improve the performance of the ANFIS model in predicting parameter electrical conductivity, sodium absorption ratio, and total hardness. The experiment result confirms that the proposed model can improve the performance of the ANFIS model for predicting EC and pH and the root mean square error (RMSE) value of the proposed model in the testing stage is 73.03 and 49.55, respectively. In [32], the genetic algorithm and particle swarm optimization (PSO) algorithm are integrated with the ANFIS model to optimize the threshold bank profile prediction. This method is also used in precipitation modeling. The experimental result indicates that the integrated ANFIS models with hybrid GA/PSO achieve better accuracy than the simple ANFIS model [33].

## 4 FTS

A water quality data is a kind of time series dataset which is likely to have complicated linear and nonlinear relationships. The Fuzzy time series (FTS) model was first proposed by Song and Chissom in 1993 to address an enrollment prediction problem [34]. Chen improved this model by replacing complicated max-min composition operations with simplified arithmetic operations [35]. In [8], a Heuristic Gaussian cloud transformation was integrated with an FTS model to forecast water quality. The experimental results showed that the proposed model significantly improved the prediction accuracy. However, there were only 520 water quality samples available to build the cloud, and thus, the model was not reliable or robust. Time series analysis is also proposed to address dissolve oxygen prediction, and the experimental results show that the proposed analysis method can find out valuable knowledge from water quality historical timeseries data [36]. In this dissertation, MLR, ANN, ANFIS, and FTS models are integrated with statistical analysis, wavelet denoising, and intelligence algorithm to explore the prediction of water quality.

## REFERENCES

- [1] L. Metcalf and H. P. Eddy, Wastewater Engineering: Treatment and Resource Recovery, 5th Edition.
- [2] McGraw Hill, Chapter 3, 2014. Journal of Water, vol. 8, no. 2, pp. 1-15, Jan. 2016
- [3] A. H. Zare, "Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters," Journal of Environmental Health Science & Engineering, vol. 12, no. 1, pp. 1-8, Jan. 2014.
- [4] L. Li, P. Jiang, H. Guang, L. Dong, G. Wu, and H. Wu " Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang River, China," Environmental Science and Pollution Research, vol. 26, no. 19, pp. 19879-19896, Mar. 2019.
- [5] J.-S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference Systems," IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665-685, May. 1993.
- [6] A. Najah, A. El-Shafie, O. A. Karim, and A. H. El-Shafie, "Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring," Environmental Science and Pollution Research, vol. 21, no. 3, pp. 1658 1670, Aug. 2013.
- [7] J. Wan, M. Huang, Y. Ma, W. Guo, Y. Wang, H. Zhang, W. Li, and X. Sun, "Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based

fuzzy inference system," *Applied Soft Computing*, vol. 11, no. 3, pp. 3238-3246, Apr. 2011.

[8] W. Deng, G. Wang, and X. Zhang, "A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 39-49, Dec. 2015.

[9] P. Singh, "Rainfall and financial forecasting using fuzzy time series and neural networks based model," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 3, pp. 491- 506, May. 2018.

[10] A. K. Kadam, V. M. Wagh, A. A. Muley, B. N. Umrikar, and R. N. Sankhua, "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*, vol. 5, no. 3, pp. 951-96, Mar. 2019.

[11] A. Sarkar and P. Pandey, "River Water Quality Modelling Using Artificial Neural Network Technique," *Aquatic Procedia*, vol. 4, pp. 1070-1077, 2015.

[12] A. P. Piotrowski, M. J. Napiorkowski, J. J. Napiorkowski, and M. Osuch, "Comparing various artificial neural network types for water temperature prediction in rivers," *Journal of Hydrology*, vol. 529, no. 1, pp. 302-315, Oct. 2015.

[13] A. Najah, A. El-Shafie, O. A. Karim, and A. H. El-Shafie, "Application of artificial neural networks for water quality prediction," *Neural Computing and Applications*, vol. 22, no. 1, pp. 187-201, Apr. 2012.

[14] L. Zhang, G. X. Zhang, and R. R. Li, "Water quality analysis and prediction using hybrid time series and neural network models," *Journal of Agricultural Science and Technology*, vol. 18, no. 4, pp. 975-983, Dec. 2015.

[15] M. J. Alizadeh and M. R. Kavianpour, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean," *Marine Pollution Bulletin*, vol. 98, no. 1, pp. 171-178, 2015.

[16] A. G. Mohammad, K. Reza, C. Kwok-Wing, S. Shahaboddin, and T. G. Pezhman,

[17] A. A. Nadiri, S. Shokri, F. T. Tsai, and A. A. Moghaddam, "Prediction of effluent quality parameters of a wastewater treatment plant using a supervised committee fuzzy logic model," *Journal of Cleaner Production*, vol. 180, no. 7, pp. 539-549, Jan. 2018.

[18] M. Khadr and M. Elshemy, "Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt," *Ain Shams Engineering Journal*, pp. 1-9, Sep. 2016.

[19] A. A. M. Ahmed and S. M. A. Shah, "Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River," *Journal of King Saud University - Engineering Sciences*, vol. 29, no. 3, pp. 237-243, Jul. 2017.

[20] S. Tiwari, R. Babbar, and G. Kaur, "Performance evaluation of two ANFIS models for predicting water quality Index of River Satluj (India)," *Advances in Civil Engineering*, vol. 2018, pp. 1-10, Mar. 2018.

- [21] S. Shahaboddin, J. N. Ehsan, E. A. Jason, A. M. Azizah, M. Amir, and C. Kwok-wing, -day ahead forecasting of chlorophyll-a Engineering Applications of Computational Fluid Mechanics, vol. 13, no. 1, pp. 91-101, Dec. 2018.
- [22] Y. Khan and S. Chai, "Ensemble of ANN and ANFIS for water quality prediction and analysis - a data driven approach," Journal of Telecommunication, Electronic and Computer Engineering , vol. 9, no. 2, pp. 117-122, 2017.
- [23] J.-S. R. Jang, "Frequently Asked Questions - ANFIS in the Fuzzy Logic Toolbox," [Online]. Available: <http://www.cs.nthu.edu.tw/~jang/anfisfaq.htm>.
- [24] A. Khashei-Siuki and M. Sarbazi, "Evaluation of ANFIS, ANN, and geostatistical models to spatial distribution of groundwater quality," Arabian Journal of Geosciences, vol. 8, no. 2, pp. 903-912, Nov. 2013.
- [25] N. Q. Sultan, S. Saeed, K. Salar, J. Salar, K. Ozgur, and S. monthly pan evaporation using wavelet support vector regression and wavelet artificial neural Engineering Applications of Computational Fluid Mechanics, vol. 13, no. 1, pp. 177-187, Jan. 2019.
- [26] A. A. Najah, A. El-Shafie, O. A. Karim and O. Jaafar, "Water quality prediction model utilizing integrated wavelet-ANFIS model with cross-validation," Neural Computing and Applications, vol. 21, no. 5, pp. 833-841, Nov. 2010.
- [27] R. Barzegar, J. Adamowski, and A. A. Moghaddam, "Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran," Stochastic Environmental Research and Risk Assessment, vol. 30, no. 7, pp. 1797-1819, Jan. 2016.
- [28] Z. Fu, J. Cheng, M. Yang, and J. Batista, "Prediction of industrial wastewater quality parameters based on wavelet-ANFIS model," in Proc. IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, 2018.
- [30] -based neural Environmental Monitoring and Assessment, vol. 191, no. 3, pp. 163-191, Feb. 2019. 78
- [31] A. Azad, H. Karami, S. Farzin, A. Saeedian, H. Kashi, and F. Sayyahi, "Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River)," KSCE Journal of Civil Engineering, vol. 22, no. 7, pp. 2206-2213, Sep. 2017.
- [32] A. Gholami, H. Bonakdari, I. Ebtehaj, M. Mohammadian, B. Gharabaghi, and S. R. swarm optimization with ANFIS to predict threshold bank profile shape based on digital laser Journal of the International Measurement Confederation, vol. 121, pp. 294- 303, Jun. 2018.
- [33] A. Azad, M. ManOochehri, H. Kashi, S. Farzin, H. Karami, V. Nourani, and J. Shirie, -fuzzy inference Journal of Hydrology, vol. 571 pp. 214-224, Apr. 2019. [34] Q. Song and B. S. Chissom, "Fuzzy time series and its models," Fuzzy Sets and Systems, vol. 54, no. 3, pp. 269-277, Mar. 1993.

- [35] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*, vol. 81, no. 3, pp. 311-319, Aug. 1996.
- [36] *International Journal of Nonlinear Science*, vol. 17, no.3, pp. 234-240, Dec. 2013.