# Modelling And Research on Water Level Control of Great Lakes Based on Neural Network PID Algorithm

Lingfei Zhang [*, #], Jiayu Zhou [#], Gaofeng Zhang [#]

Hanhong College, Southwest University, Chongqing, China, 400715

* Corresponding Author Email: brinca@email.swu.edu.cn

[#]These authors contributed equally

**Abstract.** The Great Lakes are situated in the border region between the United States and Canada, which has a significant impact on the climate and the lives of those living in the surrounding areas. The objective of this paper is to establish a network of the Great Lakes through Pearson's correlation coefficient analysis and to construct a two-tank water level model based on a PID control system in order to effectively manage the dynamics of the Great Lakes. Firstly, the strength and direction of the linear relationship between two variables is quantified through Pearson's correlation coefficient analysis, which involves the collection of observational data and the calculation of mean values. This analysis serves as a fundamental basis for predictive modelling and hypothesis testing. Secondly, based on the flow balance principle, mathematical expressions are constructed to simulate the water flow, and a PID control system is constructed to achieve optimal water level maintenance. By analysing the Pearson's correlation coefficient, the interrelationships among the variables in the Great Lakes network can be understood, thereby providing guidance for scientific research and decision-making. The results demonstrate that the constructed two-tank water level model combined with the PID control system and SHAP algotithm can effectively manage the water level of the Great Lakes and achieve optimal water level regulation.

**Keywords:** Pearson correlation coefficient, network, PID control system, SHAP algotithm.

## 1. Introduction

The Great Lakes, the world's largest freshwater lake system, straddle the United States and Canada. They are key locations for fishing, recreational enjoyment, power generation, and shipping for a wide range of stakeholders. However, they also face complex water level management challenges. These challenges are driven by a combination of climate change, environmental impacts, and management measures such as the Soo Locks and Moses Sanders Dam.

The extent of human control is limited, and natural processes such as rainfall, evaporation and glacial blockages are completely out of the control of humans. Small changes in water levels can have serious consequences.

Due to the intertwined interests of numerous stakeholders and the interaction with natural factors, the problem has been described as a complex and intractable "wicked problem".

In this paper, we construct a dynamic model using the MATLAB Simulink environment in order to efficiently and dynamically manage the water level of the Great Lakes. This is achieved by applying both classical and neural network-based PID algorithms to fine-tune the water regulation system. Our results are not only robust but also demonstrate adaptability to environmental changes. Furthermore, through comprehensive sensitivity and error analyses, we confirm the reliability and accuracy of the model.

## 2. Establishment of Great Lakes'Network

### 2.1. Pearson correlation coefficient

To establish network of the great lakes, we employed Pearson correlation coefficient analysis.The Pearson correlation coefficient, denoted as r, is a statistical method for measuring the strength and

direction of the linear relationship between two quantifiable variables.The Pearson's coefficient calculation schematic show in Figure 1.

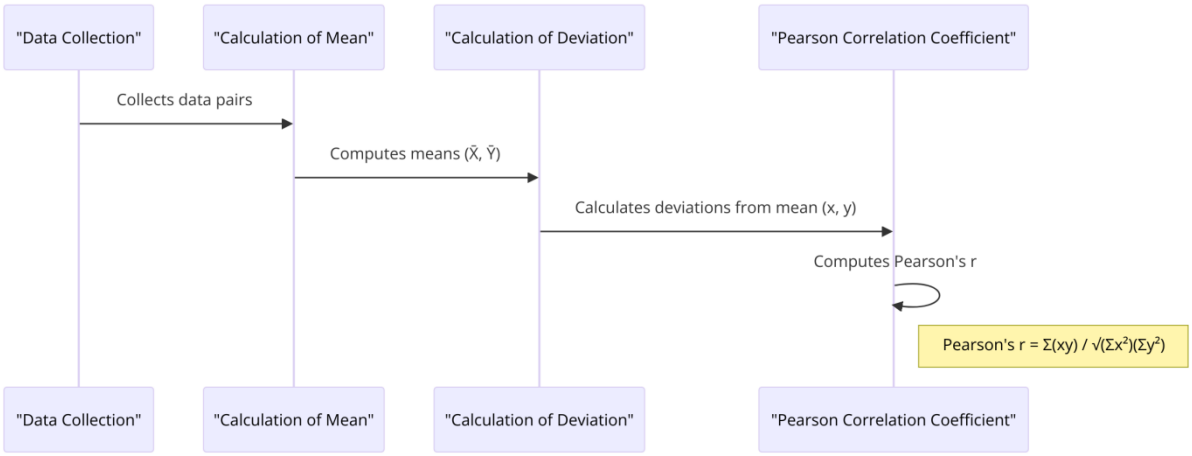



**Figure 1.** Pearson's coefficient calculation schematic

The essence of the formula is in calculating the numerator, which is the sum of the products of corresponding deviations,

$$\sum(x_i - \bar{x})(y_i - \bar{y}) \tag{1}$$

This part measures the degree of co-variation of the two variables from their means. The denominator is the product of the square roots of the sums of squared deviations,

$$\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} \tag{2}$$

ensuring the ratio is dimensionless, allowing for the correlation coefficient to be compared across different studies and datasets.The specific formula is as follows:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \tag{3}$$

This formula not only reflects the strength of the correlation between variables but also indicates the direction of correlation through its sign. Though the calculation process may appear complex, it can be efficiently completed with the help of modern statistical software and programming languages.[2][3] The heat map of lakes shows in Figure 2.

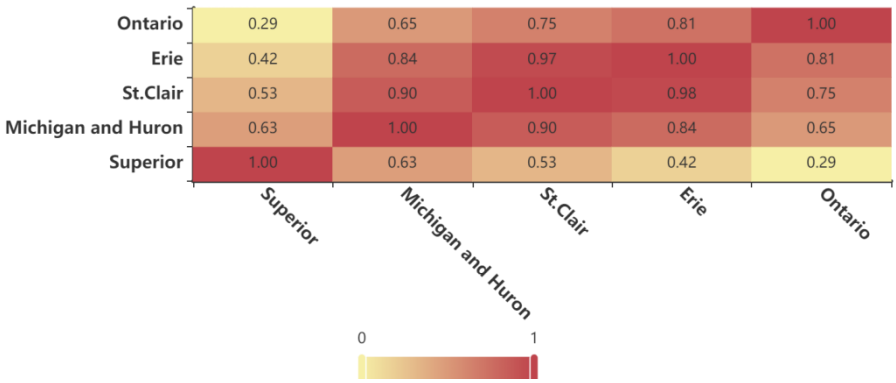



**Figure 2.** Heat map

## 2.2. Double-tank water level model based on a PID control system

### 2.2.1 Abstraction of Specific Models

In abstracting the Great Lakes system, we represent each lake as an irregular cylindrical shape to simplify the analysis of water level changes. This approach allows the calculation of lake volume changes by multiplying the lake's surface area by the change in water level, under the assumption that such changes are minimal relative to the overall size of the lakes assuming 1. Similarly, rivers connecting these lakes are abstracted as rectangular prisms, with their average cross-sectional area determined by dividing the average flow rate by the average velocity. This method of abstraction transforms the complex dynamics of the Great Lakes and their river systems into a manageable model composed of geometric shapes, facilitating easier calculations and understanding of water flows and level changes. By conceptualizing the lakes and rivers in this way, we create a simplified yet effective framework for analyzing the water dynamics within the Great Lakes system, offering a clear and accessible means to visualize and quantify interactions in environmental studies and water management.

### 2.2.2 Constructing Mathematical Expressions

To accurately describe system states and facilitate data processing, we utilize the water levels of the Great Lakes and river flow rates to construct mathematical expressions. Based on the principle of flow balance, equations are formulated to simulate water flow in rivers.Building upon this foundation, we can calculate the flow rates of the rivers corresponding to the Great Lakes system using the Manning-Strickler.[1]

$$\begin{cases} \frac{dH_1}{dt} = \frac{1}{A_1}(X - Q_1) \\ \frac{dH_2}{dt} = \frac{1}{A_2}(Q_1 + V_1 - V_2) \\ \frac{dH_1}{dt} = 0.626\frac{dH_2}{dt} + \Phi_1 \\ \frac{dH_4}{dt} = \frac{1}{A_4}(V_3 - Q_2 - Y) \\ \frac{dH_4}{dt} = 0.81\frac{dH_3}{dt} + \Phi_2 \\ \frac{dH_3}{dt} = \frac{1}{A_3}(V_2 - V_3) \end{cases} \tag{4}$$

$$\begin{cases} V_1 = \frac{1}{n} \cdot S_1 \cdot R_1^{\frac{2}{3}} \cdot \left(\frac{H_1 - H_2}{L_1}\right)^{\frac{1}{2}} \\ V_2 = \frac{1}{n} \cdot S_2 \cdot R_2^{\frac{2}{3}} \cdot \left(\frac{H_2 - H_3}{L_2}\right)^{\frac{1}{2}} \\ V_3 = \frac{1}{n} \cdot S_3 \cdot R_3^{\frac{2}{3}} \cdot \left(\frac{H_3 - H_4}{L_3}\right)^{\frac{1}{2}} \end{cases} \tag{5}$$

With these two equation sets, we essentially replicate the constructed system through mathematical expressions.

### 2.2.3 Building a PID control system

In subsequent steps, we'll refine our model using a PID control system [4]to achieve optimal water level maintenance. Initially, we correlate Lake Surperior & Lake Michigan with Lake Huron as Entity 1, and Lake Erie and Lake Ontario as Entity 2, with Soo Locks and Moses- Saunders Power DAM acting as controllers—Soo Locks controls Entity 1 and indirectly affects Entity 2, while Dam 2 controls Entity 2 and indirectly affects Entity 1. Next, we linearize and simplify our nonlinear equation set to a linear system, apply Laplace transformations, calculate the transfer function, and simulate in MATLAB's Simulink, aiming to obtain results.

$$\begin{cases} Q = f\big(u(t)\big) \\ u(t) = P \cdot e(t) + I \cdot \int_0^t e\,(\tau)d\tau + D \cdot \frac{de(t)}{dt} \\ e(t) = H_{\text{predict}} - H_{\text{real}} \end{cases} \tag{6}$$

Next, we will import the obtained transfer function into Simulink of MATLAB for simulation,[5][6] while we compare the control function with the results based on adaptive neural network fitting. Again, based on this, further debugging of the parameters was carried out. Through the Simulink simulation process, we derived two sets of PID parameters-$K_p$, $T_i$and $T_d$for controlling the dam [7]. Comparison between Classic PID and Neural Network-based PID show in Figure 3.

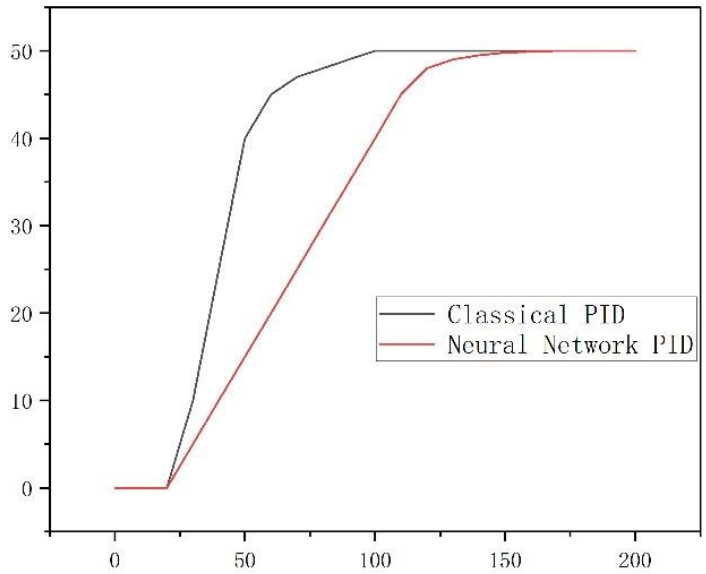



**Figure 3.** Comparison between Classic PID and Neural Network-based PID

Plug the obtained parameters into the standard form to obtain our final algorithm, the Parameter Values show in Table 1.

**Table 1.** Parameter Values

| Parameter Set | P | I | D | N |
|---|---|---|---|---|
| **Set 1** | **$1.9043\times 10^4$** | **14.23** | **$5.377\times 10^6$** | **0.06567** |
| **Set 2** | **$4.215\times 10^5$** | **595.6** | **0** | **100** |

$$\begin{cases} u(t) = (1.904 \times 10^4) \cdot e(t) + 14.23 \cdot \int_0^t e\,(t)dt + (5.377 \times 10^6) \cdot \frac{de(t)}{dt} \\ u(t) = (4.215 \times 10^5) \cdot e(t) + 595.6 \cdot \int_0^t e\,(t)dt \end{cases} \tag{7}$$

### 2.3. SHAP algotithm

In order to evaluate how well our model responds to changes in the external environment, we optimised the algorithm in Task II. Specifically, a function f (x) is introduced to represent the effect of the external environment on our control algorithm. Then we introduced the SHAP algorithm.SHAP (Shapley Additive Explanations) is a groundbreaking approach in the field of machine learning for explaining the output of any model. It is grounded in the game theory concept of Shapley values, which offers a method to fairly distribute the "payout" among players based on their contribution to the game, regardless of the order in which they joined. In the context of machine learning, "players" are the features of a model, and the "payout" is the prediction output by the model.

Mathematically, the Shapley value for a feature value is defined as the average marginal contribution of a feature value over all possible coalitions. Let's denote v(S) as the prediction function for a subset of features S where S ⊆ F and F is the set of all features. The Shapley value ψi(v) for a feature i can be formally expressed as:

$$\Psi_i(v) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [v(S \cup \{i\}) - v(S)] \tag{8}$$

Here, $|S|$ is the number of features in subset $S$, $|F|$ is the total number of features, and $\frac{|S|!(|F|-|S|-1)!}{|F|!}$ calculates the weight of each possible subset of features, considering all permutations. The term $[v(S \cup \{i\}) - v(S)]$ computes the marginal contribution of adding feature $i$ to the subset $S$.

SHAP values provide insights into how each feature contributes to the prediction for a specific instance, making it a powerful tool for interpretability. It not only offers explanations for individual predictions but also helps in understanding the model behavior globally,by aggregating SHAP values across multiple instances.[7]The SHAP value show in Figure 4.

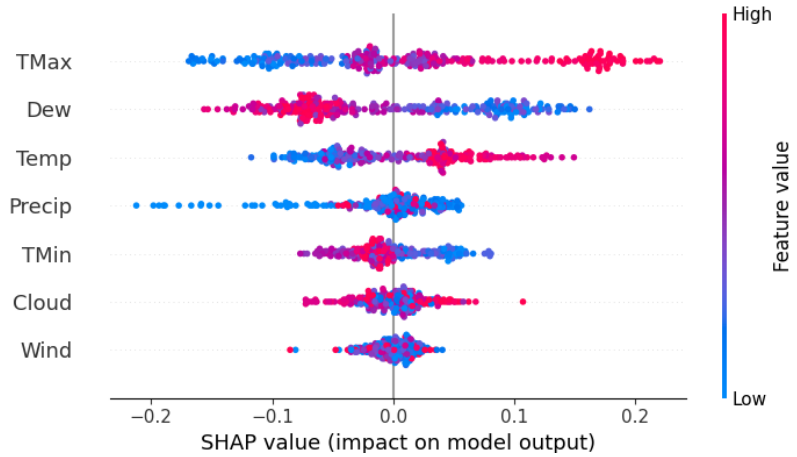



**Figure 4.** SHAP value

After obtaining the SHAP value for each feature through the SHAP model, the corresponding output value $OP\ can$ be calculated, $OP\ being$ the influence parameter:

$$O_P = \sum_{i=1}^{7} \Psi_i(v) + O_T \tag{9}$$

$OP$ By linking O functionally to the original model and recalculating the corresponding parameters, our control model is updated to take into account the effects of climate change, the new Parameter Values show in Table 2.[8][9]

**Table 2.** Parameter Values

| Parameter Set | P | I | D | N |
|---|---|---|---|---|
| Set 1 | $1.263 \times 10^6$ | 2393 | $1.858 \times 10^7$ | 0.02571 |
| Set 2 | $3.344 \times 10^4$ | 43.09 | $5.002 \times 10^6$ | 0.03576 |

$$\begin{cases} u(t) = (1.263 \times 10^6) \cdot e(t) + 2393 \cdot \int_0^t e(t)dt + (1.858 \times 10^7) \cdot \frac{de(t)}{dt} \\ u(t) = (3.344 \times 10^4) \cdot e(t) + 43.09 \cdot \int_0^t e(t)dt + (5.002 \times 10^7) \cdot \frac{de(t)}{dt} \end{cases} \tag{10}$$

## 3. Evaluation of Sensitivity

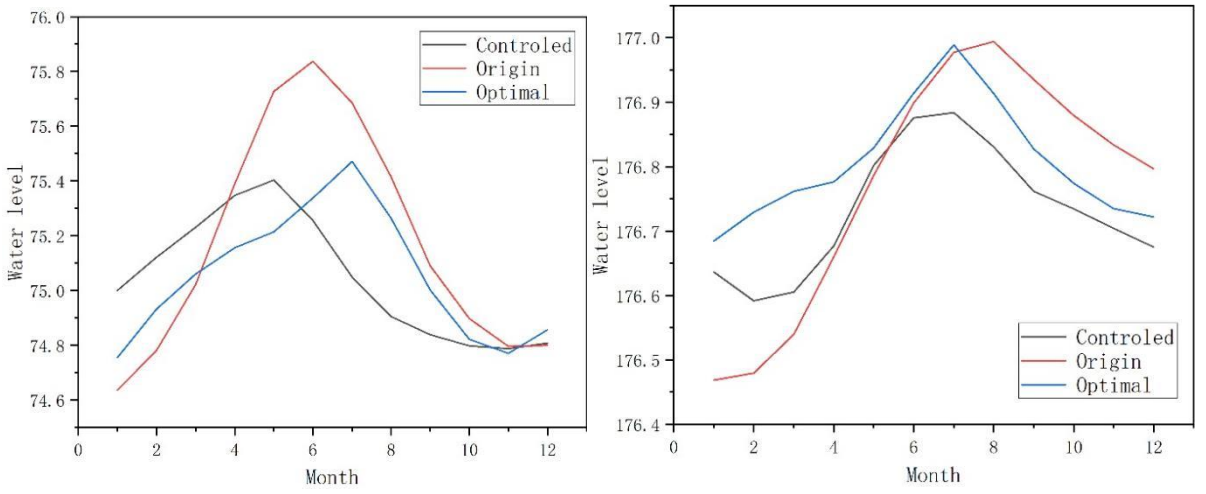
After obtaining the adjusted control model, we simulate the water level of Lake Ontario in the rainy season and the dry season in the face of environmental changes to observe the robustness of

the system, through Figure 5 and according to our optimal water level, we can know that the system can maintain the water level near the optimal water level in the face of the environmental changes in different seasons, thus We know that our system is robust.

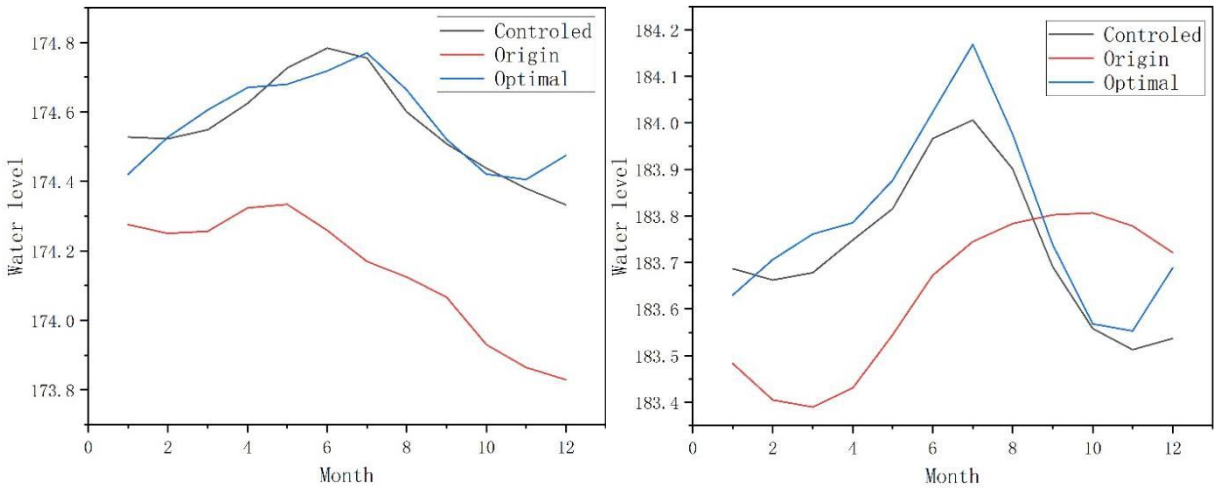
The Euclidean distance [10] represents the shortest distance between two points in a Euclidean space. This concept is an extension of the "ordinary" distance measured in a physical world, applicable to any number of dimensions. The formula for calculating the Euclidean distance between two points, P = (p1, p2, . . ., pn) and Q = (q1, q2, . . ., qn), in an n-dimensional space is given by:

$$d(A,B) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \tag{11}$$

This formula generalizes the Pythagorean theorem to n dimensions, where d (P, Q) denotes the Euclidean distance between points P and Q. It calculates the straight-line distance by squaring the differences between corresponding coordinates of the two points, summing these squares, and then taking the square root of this sum.[10]




(a) Lake Michigan and Lake Huron (b) Lake Ontario




(c) Lake Superior (d) Lake Erie

**Figure 5.** Comparison of controlled data with optimal water levels and original data in a line Graph

Subsequently, by calculating the Euclidean distance, we obtain Euclidean distance show in Figure 6.

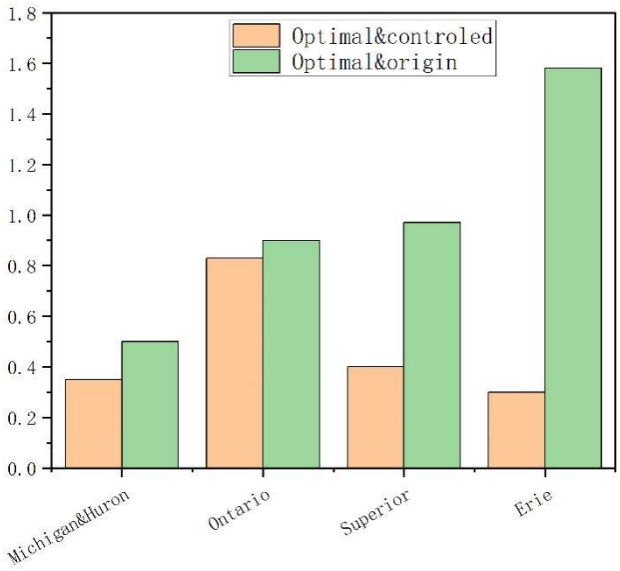



**Figure 6.** Euclidean distance

## 4. Conclusions

This paper's findings are far-reaching, offering potential advances in water level management in the Great Lakes, with practical applications ranging from water level management in the Great Lakes to influencing policy and decision-making processes. However, the study also acknowledges limitations, including the need for larger and more diverse datasets to validate the applicability of the model in different domains. Future research should focus on addressing these limitations, exploring alternative modeling techniques, and extending the applicability of models to a wider range of problems. This research lays the groundwork for new directions in water level management in the Great Lakes, highlighting the potential for significant advances in understanding and applying adaptive neural networks, and dual-capacity tank PID control.

## References

[1] U.S. Geological Survey. Manning's Roughness Coefficients for South Carolina Streams[R]. Washington, DC: U.S. Geological Survey, 2022.

[2] Benesty, J., Chen, J., Huang, Y. On the Importance of the Pearson Correlation Coefficient in Noise Reduction[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2020, 28(4): 757-765.

[3] Gabryel, M., Scherer, R., Damaševičius, R. Pearson Correlation-Based Feature Selection for Document Classification Using Balanced Training[J]. Sensors, 2020, 20(23): 6793.

[4] Adrian Gambier. MPC and PID Control Based on Multi-Objective Optimization[C]// Proceedings of the 2008 American Control Conference. IEEE, 2008: 4727-4732.

[5] Chen, M., Liu, W. Robust Self-Learning PID Control of an Aircraft Anti-Skid Braking System[J]. Mathematics, 2022, 10(8): 1290. DOI: 10.3390/math10081290.

[6] Hossein Bonakdari, Isa Ebtehaj, Pijush Samui, Bahram Gharabaghi. Lake Water-Level Fluctuations Forecasting Using Minimax Probability Machine Regression, Relevance Vector Machine, Gaussian Process Regression, and Extreme Learning Machine[J]. Water Resources Management, 2019, 33: 3965-3984.

[7] Bahrudin Hrnjica, Ognjen Bonacci. Lake level prediction using feed forward and recurrent neural networks[J]. Water Resources Management, 2019, 33(7): 2471-2484.

[8] Ahmet Sancak Sanlı, Erdal Kesgin, R. Ilayda Tan, Hayrullah Agaccioglu, Ibrahim Demir, Ilke Sinan. Effect of lake-water budget management preferences on optimum operating conditions and neighboring basins interacting: case of lake Beysehir (Turkey)[J]. Sustainable Water Resources Management, 2022, 8: 1-18.

[9] Pengfei Xue, Aditya Wagh, Gangfeng Ma, Yilin Wang, Yongchao Yang, Tao Liu, Chenfu Huang. Integrating deep learning and hydrodynamic modeling to improve the Great Lakes forecast[J]. Remote Sensing, 2022, 14(11): 2640.

[10] E. Maria, E. Budiman, M. Taruk, et al. Measure distance locating nearest public facilities using haversine and euclidean methods[C]// Journal of Physics: Conference Series, 2020, 1450: 012080. IOP Publishing.