Research on lake water level dynamic adjustment model based on ARIMA and pPSA

Xinyi Wu [1,*], Huaizheng Lu [2], Kexin Huang [1], Ruoyu Wang [1]

[1] School of Ocean Information Engineering, Jimei University, Xiamen, China, 361021

[2] College of Computer Engineering, Jimei University, Xiamen, China, 361021

* Corresponding Author Email: 202121306003@jmu.edu.cn

Abstract. The aim of this study is to construct a time-water level dynamic adjustment model to effectively manage lake levels and maximize the needs of relevant stakeholders. First, the time-series information was stabilized by constructing difference equations and using lagged differences, thereby eliminating data fluctuations. Second, by fuzzy modeling natural factors (e.g., evaporation and precipitation), we constructed an autoregressive integral moving average model (ARIMA) for fuzzy modeling of water levels. Finally, we use perturbed particle swarm algorithm (pPSA) for dynamic adjustment of water level to avoid local optimal solutions and improve model adaptability. Experimental results show that our model has better performance in stabilizing water level, accurate prediction and sensitivity analysis.

Keywords: Difference equations, pPSA, water level adjustment, sensitivity analysis Introductio.

1. Introduction

Dynamic adjustment of lake levels plays a critical role in water resource management and ecological preservation. Previous studies have often employed time series analysis methods to forecast changes in lake water levels, with the ARIMA (Autoregressive Integrated Moving Average) model and pPSA (Periodic Pattern of Seasonal Adjustment) model being frequently utilized. The ARIMA model is adept at modeling and predicting time series data by capturing both short- and long-term dependencies, while the pPSA model excels in identifying and describing trends, seasonality, and periodicity within time series.

This paper aims to investigate dynamic adjustment modeling of lake levels using both the ARIMA and pPSA models, offering insights for decision-makers involved in water resource management and lake ecological preservation. The ARIMA model is employed to model and forecast time series data, leveraging autoregression, differencing, and moving average techniques. Similarly, the pPSA model, based on time series analysis, is utilized for analyzing and predicting lake water level data.

Furthermore, this paper provides a detailed description of the dynamic adjustment model of lake water levels based on the ARIMA and pPSA models. It explores the synergies between these models to construct a comprehensive framework for predicting Lake water level changes, thereby furnishing decision-makers with valuable guidance for effective water resource management and ecological conservation efforts. Lastly, the paper underscores the significance and potential of utilizing the ARIMA and pPSA models in dynamic adjustment modeling, offering recommendations for further enhancements and optimizations in time series analysis applications.

2. Related Works

2.1. ARIMA Model

The ARIMA model has produced many excellent results in time series forecasting this year, and is combined with different models to achieve high-precision forecasting on specific topics. Tarmanini et al. [1] evaluated the performance loss of ARIMA in a specific environment by using two forecasting methods, ARIMA and artificial neural network (ANN), to address the knowledge gap in machine learning (ML)-based load forecasting techniques. Chodakowska et al. [2] used the ARIMA model to study the seasonal prediction of solar radiation under different climate conditions, further analyzed the modeling mechanism of the ARIMA model as well as the performance and prediction capabilities of the model, evaluated the model through data obtained in the field, and also evaluated the basic prediction model. Proposed for analysis. Luo et al. [3] proposed an ARIMA-WOA-LSTM model, which uses ARIMA to extract the linear part of air pollution data and output the nonlinear part, uses the WOA-LSTM model to predict the nonlinear part, and further adjusts the hyperparameters of the combined model. Explore possibilities for reducing model error.

2.2. pPSA

Tao et al. [4] proposed a dual population PSO algorithm (BPPSO) with a random perturbation strategy based on the traditional particle swarm optimization algorithm (PSO). After reaching the set threshold, random perturbations are added to the positions of all particles in the two subgroups. Thereby increasing diversity and enhancing the ability to escape local optimality, and experimentally verifying it through three benchmark functions and four environment models, the feasibility of BPPSO in solving mobile robot path planning challenges was finally discovered. Dong et al. [5] proposed a hybrid algorithm (HKSOPSO-CP) based on kernel search optimization (KSO) and particle swarm optimization (PSO) with Cauchy perturbation, which was successfully applied to solve the EED problem in power systems, and considered valve point effect. At the same time, the algorithm can also use different kernel functions to solve some optimization problems, and obtain better results by paying attention to the calculation skills of kernel mapping. Jin et al. [6] proposed a novel improved discrete particle swarm optimization algorithm (IDPSO) to solve the multi-objective flexible job shop scheduling problem (MOFJSP) under dynamic disturbance. By enhancing the ability of the discrete particle swarm algorithm to jump out of local extremes, they designed the inertia weight was reduced nonlinearly and the individual optimal and global optimal particle update methods of the algorithm were redesigned. Finally, the effectiveness of the algorithm was verified on the Kacem data set. The results show that the algorithm has strong performance in finding the optimal.

3. Background

3.1. Great Lakes Water-system

The Great Lakes are five neighboring freshwater lakes in North America, including Lake Superior, Lake Michigan, Lake Huron, Lake Erie, and Lake Ontario. Located between the United States and Canada, these lakes are considered to be the largest group of freshwater lakes in the world and are typical transboundary watersheds. The Great Lakes not only play an important role geographically, but also have a profound impact on the economy and livelihoods of the surrounding region, and regulating water levels for the benefit of all stakeholders is a major issue. The water levels of the Great Lakes depend on the input and output of water flows, which can be controlled by the two main flow control mechanisms, but are still influenced by rainfall, evaporation, and factors that humans cannot intervene with. [7] The management of the water levels of the Great Lakes is a very complex and challenging issue, involving many conflicting interests. It is necessary to face the dynamic situation, take into account both natural and anthropogenic factors, balance various interests, and seek feasible solutions to ensure the ecological balance of the lakes and sustainable socio-economic development. [8]

3.2. Our Work

1) Dynamic flow modeling to reflect the hydrologic regime of the Great Lakes and their linked rivers from Lake Superior to the Atlantic Ocean.

By sorting out the water flow and many natural influence values of the Five Great Lakes and adjacent waters in recent years, a dynamic water flow network of the entire continent was constructed to select the optimal water flow path and facilitate managers to make dynamic adjustments.

2) A multi-objective planning model was constructed to consider the different needs of different stakeholders for water levels and find the optimal level for each lake

Based on the built water network, a multi-objective optimization model is constructed by introducing different stakeholders to find the best water level to ensure maximum revenue for different lake areas and river basins.

4. Experiment

4.1. Build Dynamic Flow Network Model

Initially, all feature information is subtracted from the average of its dimensions to perform data centering processing, and then the covariance matrix is obtained. By analyzing the Great Lakes data using Principal Component Analysis (PCA), the Great Lakes weight distribution in Table 1 can be obtained, laying the foundation for further processing of the subsequent model.

Table 1. PCA weight analysis of Great Lakes water level information.

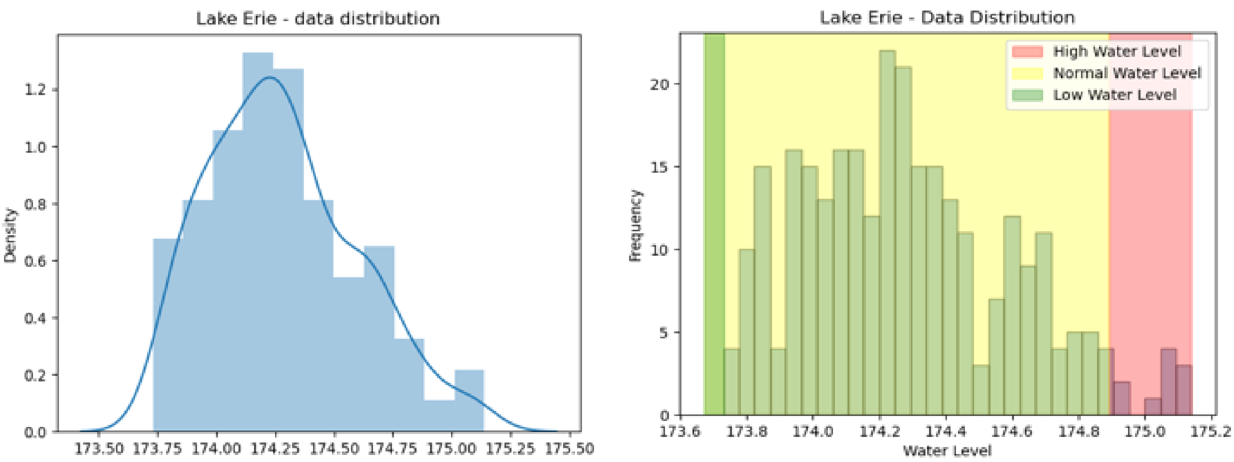
Lake name	PCA weight
Lake Michigan Huron	0.9566
Lake St. Clair	0.9205
Lake Superior	0.9040
Lake Erie	0.8761
Lake Ontario	0.5218

In known conditions, changes in water level exceeding 2 feet will have a greater impact on overall revenue. We can use this distance to define the height of the water level, as shown in Table 2:

Table 2. Water level interval definition.

Water level classification	PCA weight
High water level	[Average Water Level + 0.6096, Highest Water Level)
Normal water level	(Average Water Level - 0.6096, Average Water Level + 0.6096)
Low water level	(Lowest Water Level, Average Water Level - 0.6096]

Taking Lake Erie as an example, after obtaining the water level information and interval division standards, we can visualize the information about the lake. Picture (a) in Figure 1 shows the water level information of Lake Erie, and picture (b) divides the three water level intervals using different colors. The data distribution shows a normal distribution trend, further proving the rationality of the data.



(a) Water level distribution (b) Water level division.

Figure 1 Water level information visualization.

We use the binary method to simulate the interest correlation of each relevant beneficiary to different lakes at different times as a benefit function:

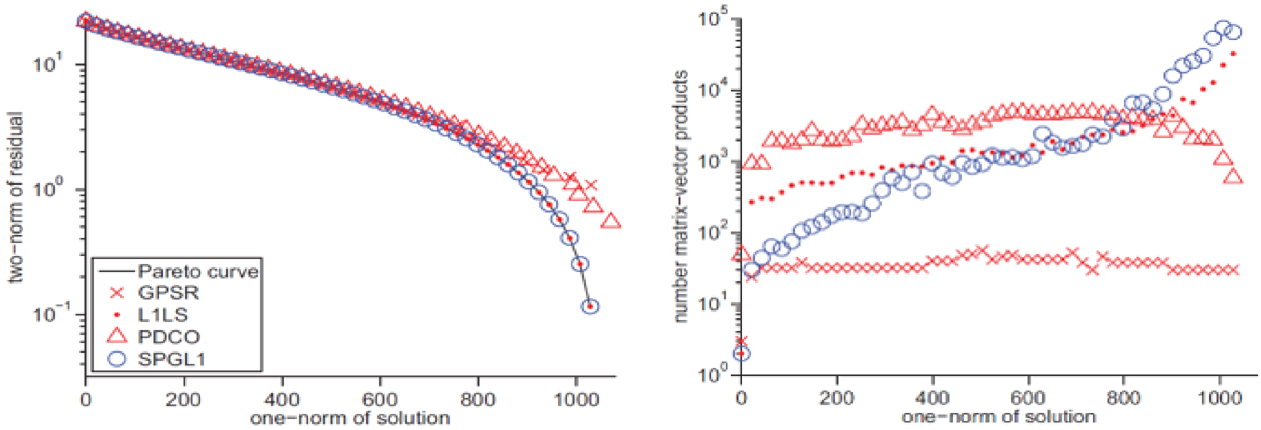
$$f(x)=\begin{cases} -1, & positively_{related} \\ 0, & irrelevant \\ 1, & negative_{related} \end{cases} \tag{1}$$

Then use the ideal point method to construct a multi-objective programming model for solution. The objective function can be expressed as:

$$\min/\max F_i(x)\times f_i(x) \tag{2}$$

Where $F_i(x)$ represents the objective function of the *i* stakeholder, and $f_i(x)$ binary represents the correlation degree of the *i* stakeholder, $i\in\{0,1,\dots,n\}$ n represents the total number of stakeholders.

By constructing the objective function and adding the conditional constraints of the water level, we constructed a multi-objective model for solving the optimal value. Next, we find the optimal solution by looking for the Pareto front. By using quadratic extrapolation for the last point, and requiring the extrapolated minimum to coincide exactly with the horizontal axis, this works well when the Pareto curve ends up approaching quadratic. The solution process is shown in Figure 2.



(a) Solution quality (b) Solution performance

Figure 2. Solver performance on multi-objective problems.

The plot on the left shows the norm of the calculated residual solution for fifty parameter values; the plot on the right gives the corresponding number of matrix-vector products required.

By solving the Pareto surface, we evaluate the impact of different solutions on the interests of all parties and their balance between environmental protection and economic development, and obtain the optimal water level, as shown in Table 3.

Table 3. PC Optimal water levels for the Great Lakes taking into account stakeholder visions and other factors.

Lake name	Optimal water level(meter)
Lake Michigan Huron	176.3289
Lake St. Clair	175.1030
Lake Superior	183.3464
Lake Erie	174.2807
Lake Ontario	74.8298

4.2. Build Time-water_level Dynamic Adjustment model

We choose to construct a difference equation and use lagged differences to eliminate data fluctuations and thereby stabilize the time series information. Here the second-order difference is defined as:

$$\triangle^2 Y_t = \triangle (Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t+2} \tag{3}$$

The time series is represented by \$Y_{t}\$. Instead of subtracting neighboring observations, the difference operation is performed by subtracting observations separated by a certain number (i.e., the number of lags). On this theoretical basis, taking Lake Erie as an example, we perform multiple differences on the water level data, and the effect is shown in Figure 3.

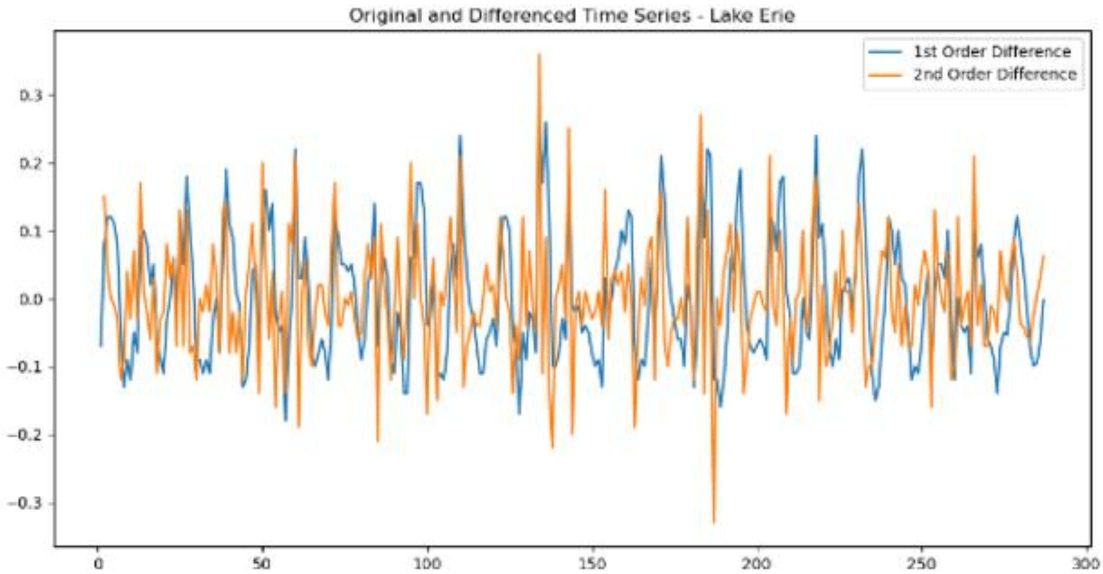


Figure 3. Changes in Lake Erie's water leve
l time-series information after two differencing sessions.

From the information in the Figure 4, it is not difficult to see that the time series volatility of only one difference is significantly smaller than that of multiple differences, which is because the water level time series data does not have obvious seasonal trends, or it can be said that the magnitude of numerical changes with the seasons is relatively small, so when dealing with the time series data of the network, we chose to eliminate the fluctuation of the time series by only performing a single difference process.

Then it is necessary to carry out fuzzy simulation of natural factors. We perform fuzzy simulations of natural factors such as evaporation and precipitation by constructing an Autoregressive Integrated Moving Average Model. The autoregressive portion of the time series is handled using an autoregressive (AR) model that incorporates the effect of past values on the outcome of the predicted values, which is calculated as:

$$AR: Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \xi_t \tag{4}$$

The moving average portion of the time series is also handled using a moving average model (MA) that incorporates the effect of past forecast errors on the current forecast value, which is calculated as follows:

$$MA: Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \tag{5}$$

Considering the two model parts comprehensively, the calculation formula of the autoregressive integrated moving average Model (ARIMA) is as follows:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \xi_t \tag{6}$$

Among them, Y_t is the time series of the current window, and θ_1 to θ_p be the parameters of the AR model, which are used to describe the relationship between the current value and the value of p time points in the past? φ_1 To φ_q are the parameters of the MA model, used to describe the relationship between the current value and the error of the past q time points, ξ_t is the error term at t time point, and c is the constant term. We use this ARIMA model to perform fuzzy simulations of natural factors such as evaporation and precipitation.

4.3. Build Dynamic adjustment model of water level

We set the final goal of dynamic adjustment of the water level to the optimal water level mentioned above, and update the water level iteratively through the opening and closing of the dam. Since the particle swarm algorithm often has a high convergence speed when synchronizing the local search of particles and updating the fuzzy global optimal value of particle trajectories, this will lead to a rapid loss of diversity during the optimization process and lead to model underfitting, and this simulation environment There are a large number of fuzzy variables under the model, which will increase the speed of model fitting and lead to more serious under-fitting. Even if the idea of the ant colony algorithm is used to quickly update the particle position to obtain a feasible solution space, it still cannot get rid of the shortcomings of local optimality. We use a perturbed particle swarm algorithm (pPSA) to avoid local optima. This particle updating strategy is different from other fuzzy particle swarms. pPSA with a perturbed particle updating strategy is 'possibly at gbest' instead of a crisp location which is different from other fuzzy PSOs. The model can be expressed as:

$$p(k) = \begin{cases} \sigma_{max}, & \text{cycles} < \alpha \times \text{max\_cycles} \\ \sigma_{min}, & \text{otherwise} \end{cases} \tag{7}$$

Whereσ_{max},σ_{min}, and α are manually set parameters. Then we get the average daily flow by calculating the mean value of the monthly flow according to the model assumptions, and control the dynamic adjustment of the inflow and outflow of the lake by opening and closing the gates, so that its water level can be kept at the optimal level as much as possible. The adjustment process is shown in Figure 4.

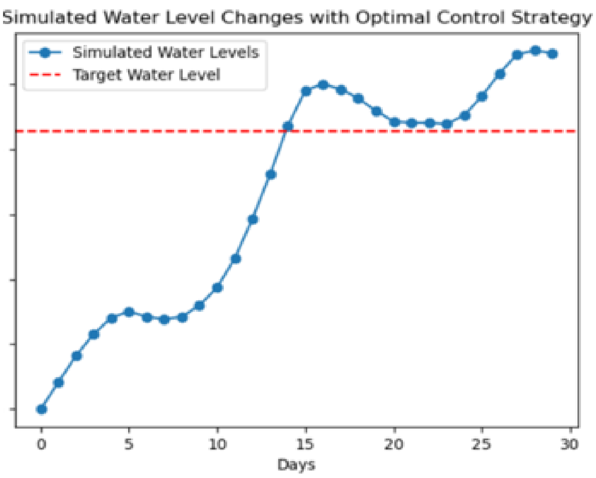
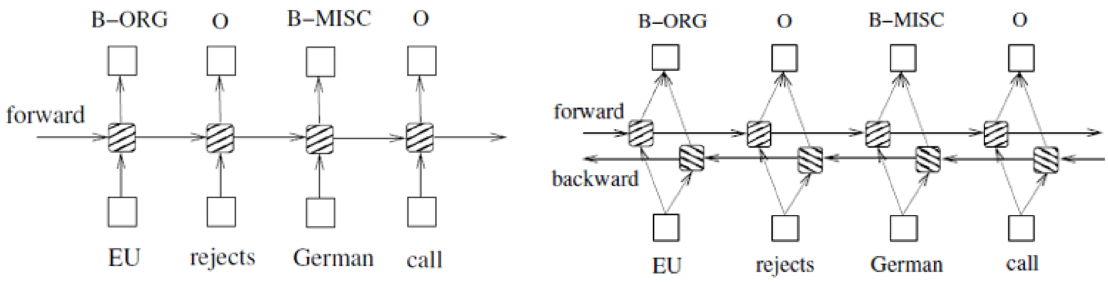


Figure 4. Use ARIMA to simulate environmental factors and dynamically adjust the water level by opening and closing gates based on the pPSA model.

Then we established a sensitivity analysis model of the control algorithm and adopted the Sobol method based on complex environments:

$$S_{T_i} = \frac{E_{X_i}(Var(Y|X_i))}{Var(Y)} \tag{8}$$

On this basis, the bidirectional LSTM model is used to perform information-based prediction on time series data. Different from traditional LSTM models, in order to more effectively utilize past features (through forward state) and future features (through backward state) within a specific time range, backpropagation time series is used to train the LSTM network, and then a bidirectional LSTM is established. The model structure is shown in Figure 5.



(a) LSTM network (b) Bidirectional LSTM network

Figure 5. LSTM model structure and bidirectional LSTM model structure based on it to obtain more contextual semantics by training reverse time series.

Taking Ontario Lake as an example, the training visualization using this model is shown in Figure 6. Use the model to make predictions for 2017 data.

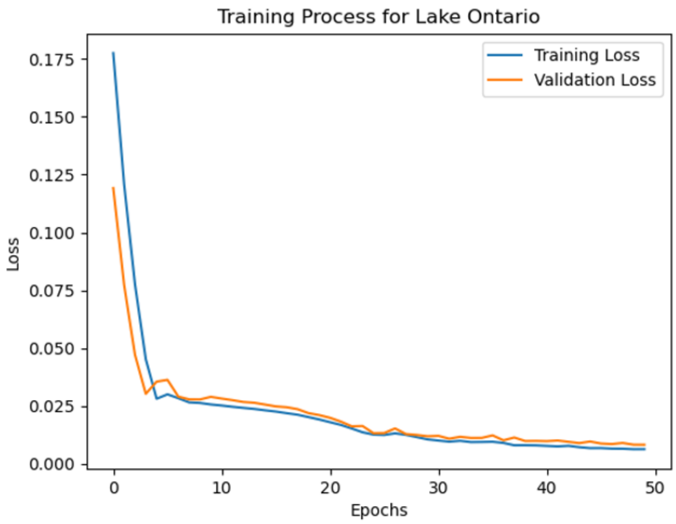


Figure 6. Visualization of Bidirectional LSTM Model training process.

Taking Ontario Lake as an example, the Bidirectional LSTM model is used to predict the 2017 data and the prediction results are compared with the original bad value results. The results are shown in Figure 7.

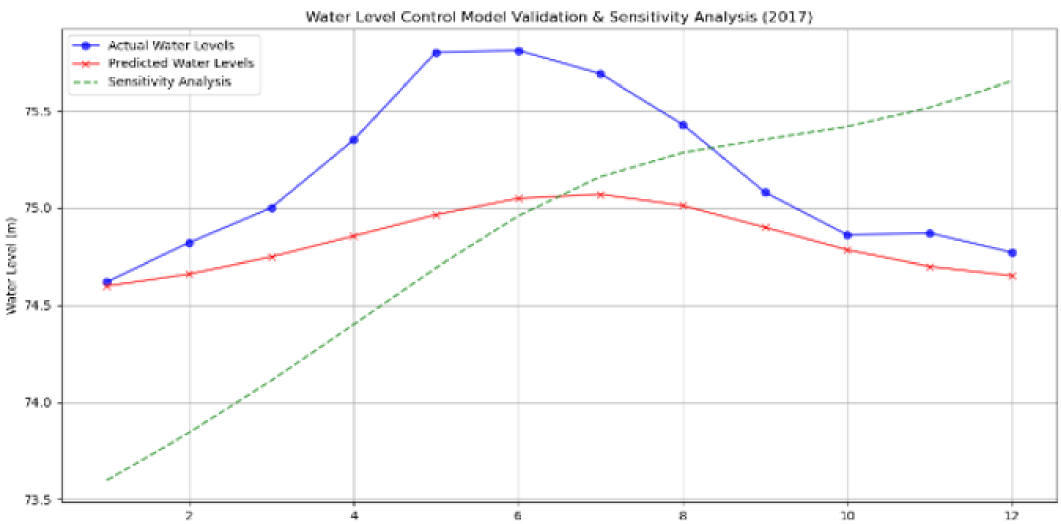


Figure 7. Comparison of recorded values in Lake Ontario in 2017 with predicted values after adding control algorithms.

Due to defects in the control algorithm that year, the highest water level in the past years occurred, which was significantly higher than the optimal value. After correction by our control algorithm, the result is smoother and closer to the optimal water level. At the same time, it can be seen that the sensitivity curve is relatively smooth, indicating that the control algorithm has a relatively gentle impact on the dam outflow, and the control algorithm can indeed control water level changes and thus affect the interests of relevant beneficiaries [9].

Then, we established a Monte Carlo simulation [10] combined with a hydrological analysis model to study the sensitivity model of the control algorithm to environmental variables. We queried precipitation, snowfall, and evapotranspiration for cities near the Great Lakes (e.g., Ontario, Minnesota, etc.), further quantified the effects of environmental conditions by inputting them into the SWAT hydrologic model, further incorporated environmental variables using Monte Carlo simulations, and finally calculated sensitivities to environmental variables, with the results shown in Figure 8.

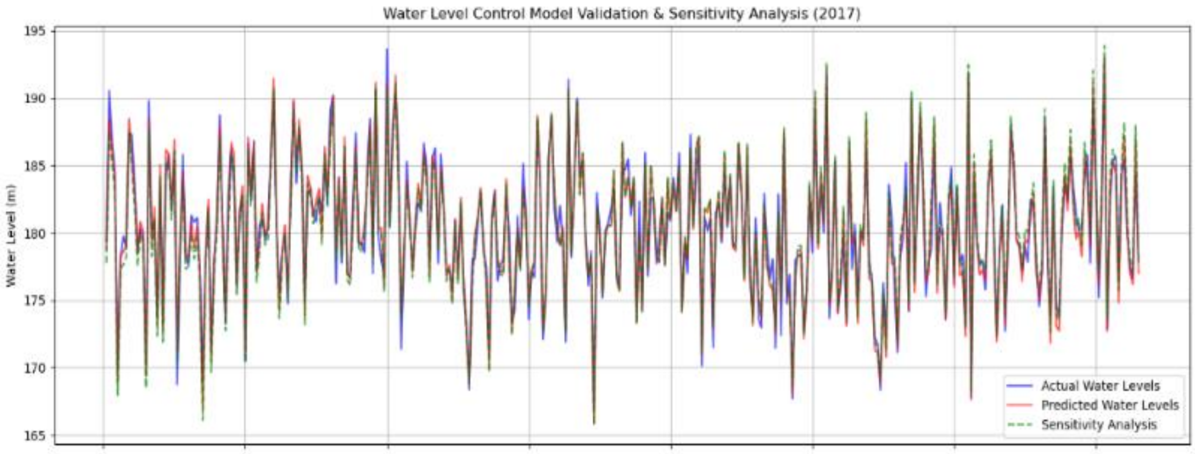


Figure 8 Control algorithm sensitivity to environmental variables.

At the same time, a quantitative and objective representation of sensitivity is needed. We quantify this performance in terms of economy to more intuitively represent the changes in the economy as seasons and water levels change, thereby more accurately proving the correctness of the sensitivity analysis. The impact of water level changes on the economy is shown in Figure 9.

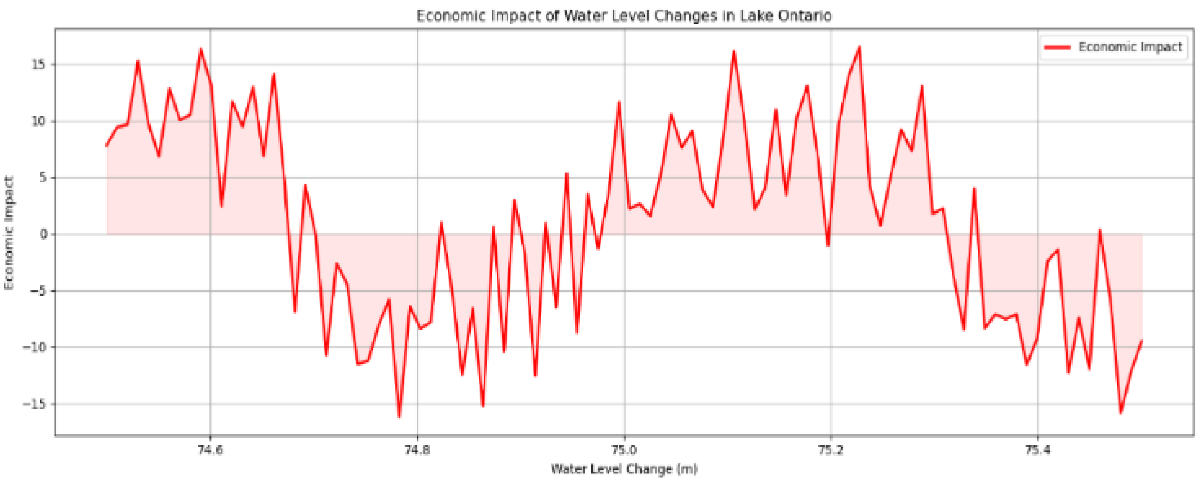


Figure 9. Impact of water level changes on economic conditions.

It is not difficult to see that after adding environmental changes, the sensitivity fluctuates greatly, and the economy is greatly affected by water levels.

5. Conclusions

In conclusion, our study demonstrates the effectiveness of integrating the ARIMA and pPSA models for dynamic adjustment modeling of lake water levels. By leveraging the strengths of both models, we have provided valuable insights into forecasting Lake water level changes. Our findings offer practical guidance for decision-makers involved in water resource management and lake ecological preservation. Further research could focus on refining and optimizing the proposed model to accommodate additional factors and stakeholders, thereby enhancing its applicability in real-world scenarios.

References

[1] Tarmanini C, Sarma N, Gezegin C, et al. short-term load forecasting based on ARIMA and ANN approaches [J]. Energy Reports, 2023, 9: 550-557.

[2] Chodakowska E, Nazarko J, Nazarko Ł, et al. Arima models in solar radiation forecasting in different geographic locations [J]. Energies, 2023, 16(13): 5029.

[3] Luo J, Gong Y. Air pollutant prediction based on ARIMA-WOA-LSTM model [J]. Atmospheric Pollution Research, 2023, 14(6): 101761.

[4] Tao B, Kim J H. Mobile robot path planning based on bi-population particle swarm optimization with random perturbation strategy [J]. Journal of King Saud University-Computer and Information Sciences, 2024: 101974.

[5] Dong R, Ma L, Chen H, et al. Hybrid kernel search and particle swarm optimization with Cauchy perturbation for economic emission load dispatch with valve point effect[J]. Frontiers in Energy Research, 2023, 10: 1061408.

[6] Jin H, Jin Q, Zhang Y. An improved discrete particle swarm optimization for multi-objective flexible job shop scheduling problem under dynamic perturbation[C]//International Conference on Automation Control, Algorithm, and Intelligent Bionics (ACAIB 2023). SPIE, 2023, 12759: 354-361.

[7] Bunnell D B, Ackiss A S, Alofs K M, et al. A science and management partnership to restore coregonine diversity to the Laurentian Great Lakes [J]. Environmental Reviews, 2023, 31(4): 716-738.

[8] McGregor D, Latulippe N, Whitlow R, et al. Towards meaningful research and engagement: Indigenous knowledge systems and Great Lakes governance [J]. Journal of Great Lakes Research, 2023, 49: S22-S31.

[9] Koelmans A A, Redondo-Hasselerharm P E, nor N H M, et al. On the probability of ecological risks from microplastics in the Laurentian Great lakes [J]. Environmental Pollution, 2023, 325: 121445.

[10] Li J, Zhang J, Yang X, et al. Monte Carlo simulations of deformation behaviour of unbound granular materials based on a real aggregate library [J]. International Journal of Pavement Engineering, 2023, 24(1): 2165650.