Optimization of Water Level from Great Lakes Based on Vector Autoregressive Model and Goal Programming Model

Ziheng Zhou [1, *], Binzhe Li [2]

[1] School of Statistics, Southwestern University of Finance and Economics, Chengdu, China, 611130

[2] School of Economics, Southwestern University of Finance and Economics, Chengdu, China, 611130

* Corresponding Author Email: adamzz5@163.com

Abstract. The Great Lakes, the largest group of freshwater lakes in the world, have profound impacts on residents, ecosystems, water resource utilization, shipping, and tourism industries. Addressing water level variability, this study integrates network science, goal programming algorithm, and Model Predictive Control to establish a comprehensive and adaptive model for optimizing dam adjustment mechanisms and maximizing stakeholder benefits. Initially, a Vector Autoregression Model is developed for the Great Lakes and connecting river flows toward the Atlantic Ocean to conditionally project future paths of specified variables. This model yields a network representation of the Great Lakes system. Subsequently, a Goal Programming Model is constructed to determine optimal water levels throughout the year based on extensive literature review and priority rankings. Leveraging insights from the 2014 plan, a detailed analysis is conducted on Lake Ontario water levels, focusing solely on stakeholders and influential factors. This research contributes a robust methodology for managing water levels in the Great Lakes region, providing valuable insights for sustainable water resource management.

Keywords: Water Level, Network, VAR, Goal Programming.

1. Introduction

Human activities have led to serious degradation of lakes in terms of their ecological functions, making the study of the ecological water level of lakes exceedingly important. The variation of water levels is influenced by various factors.

However, the dynamic nature of the water flow, coupled with the conflicting interests of stakeholders, presents a highly challenging and "wicked" problem.

On this basis, this paper proposes a comprehensive and adaptive model and management plan to address these complexities, with a particular focus on understanding the sensitivity of control algorithms to environmental changes.

Based on the above understanding, we need to solve the following problems:

(1) Conflicts among stakeholders: Different stakeholders' water resource demands can conflict, such as balancing flood control and navigation needs.

(2) Dynamic changes and uncertainties: Water levels are affected by natural and human factors like climate change, precipitation, and evaporation, increasing the complexity of water resource management due to uncertainties.

(3) The complex water resource management demands: The diversified uses of the water resources of great lakes make the management work complex and challenging.

(4) Limitations of control mechanisms: While control mechanisms like the compensatory project of the Suez Canal and the Moses-Saunders Power Dam exist, natural phenomena such as rainfall and evaporation are beyond human control.

Historically, the management and optimization of water levels in the Great Lakes have been key areas in environmental studies and water resource management. Previous studies have primarily focused on using various statistical and forecasting models to understand and predict water level changes, such as Autoregressive Models (AR), Autoregressive Moving Average Models (ARMA), and Autoregressive Integrated Moving Average Models (ARIMA). These studies have laid the groundwork for understanding the seasonal and long-term variations in water levels and have guided effective management strategies for water resources. In the current study, time series modelling technique was adopted by AS Azad [1] for the RWL prediction in RHR using Box–Jenkins autoregressive seasonal autoregressive integrated moving average (SARIMA) and artificial neural network (ANN) hybrid models. WY Xing [2] proposes a com-bined model based on variational mode decomposition (VMD), a genetic algorithm-the ELMAN neural network-VMD-the autoregressiveintegrated moving average (ARIMA) model (GA-ELMAN-VMD-ARIMA). a water level prediction method combining The Autoregressive Integrated Moving Average (ARIMA) Model, Exponential Smoothing (ES) model and Long Short-term Memory (LSTM) model through nonlinear programming genetic algorithm is proposed by C Wang [3]. By combining the advantages of local search of nonlinear programming and global search of genetic algorithm, this method uses nonlinear programming genetic algorithm to allocate the weights of ARIMA model, ES model and LSTM model, and obtains the final water level prediction result by weighting.

The novelty of this research lies in its use of the Vector Autoregressive model to analyze and predict changes in the water levels of the Great Lakes and in integrating the Goal Programming model to provide a multi-objective optimization framework for water level management. This combined approach allows researchers to consider hydrological cycles, the impacts of climate change, and socio-economic demands simultaneously, enabling the development of more effective and sustainable water resource management strategies. Moreover, this study enhances the predictive capacity of the models and the precision of decision-making by applying the latest statistical techniques and algorithms, offering new scientific evidence and decision-making tools for water resource management in the Great Lakes region.

2. The basic fundamental

2.1. Vector Autoregressive Model

The vector autoregression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models. Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model.

The research methodology for constructing the water level vector autoregressive (VAR) model is illustrated in Figure 1, comprising three main steps: stationarity testing, Granger causality testing, and variance decomposition. In particular, The ADF (Augmented Dickey-Fuller) test is used to check for stationarity in the time series variables. It is important to ensure that the variables are stationary before proceeding with VAR modeling.[4]

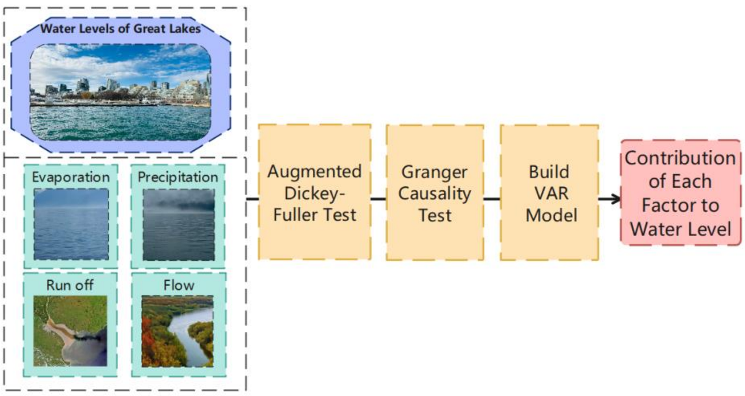

Figure 1: Research Idea of’ Influencing Factors Based on VAR Model

Nnconstrained regression mode:

$$Y_t = c + \rho_1 WY_{t-1} + \rho_2 WY_{t-2} + \cdots + \rho_p WY_{t-p} + \varepsilon_t \quad (1)$$

Constrained regression mode:

$$Y_t = c + \sum_{i=1}^{p} \rho_{\mathrm{i}} WY_{t-\mathrm{i}} + \sum_{i}^{p} X_{t-\mathrm{i}} \beta_{\mathrm{i}} + \varepsilon_t \quad (2)$$

ρ and β are the parameters to be estimated, c is the constant term; p is the number of lags; and ε is the white noise.

Particularly, introduce the relationship matrix W to reduce the number of parameters to be estimated. W is a 5×5 matrix, where the value is 1 if two lakes are connected by a river, and 0 otherwise.

2.2. Goal Programming Model

The lakes support various activities like fishing, recreation, power generation, drinking water supply, navigation, wildlife habitat, construction, and irrigation. This diversity of uses creates a wide range of stakeholders involved in managing the lakes' water. Considering these stakeholders' interests, the planning challenge becomes a multi-objective decision task. For each decision objective, introduce positive and negative deviation variables d^+ and d^--, representing the portions by which the decision value exceeds or falls short of the target value. By definition, it holds that:

$$d^+ >= 0, d^- >= 0 \; and \; d^+ d^- = 0 \quad (3)$$

The prioritization of different objectives involves two distinctions. One is absolute, which can be represented by a priority factor P_t. The satisfaction of lower-priority objectives is contingent upon the fulfillment of higher-priority ones. The larger the value of t, the lower the priority level. The general form of the goal programming model is as follows:

$$min\{P_l \sum_{k=1}^{K} (W_{lk}^- d_k^- + W_{lk}^+ d_k^+), \; l = 1,2, \ldots, L\} \quad (4)$$

$$s.t. \begin{cases} c_{kj} x_j + d_k^- - d_k^+ = g_k \; (k = 1,2, \ldots, K) \\ \sum_{j=1}^{n} a_{ij} x_j \leq (=, \geq) b_i \; (i = 1,2, \ldots, m) \\ x_j \geq 0 \; (j = 1,2, \ldots, m) \\ d_k^-, d_k^+ \geq 0 \; (k = 1,2, \ldots, K) \end{cases} \quad (5)$$

In the model, g_k represents the expected target value of the k-th objective constraint, where W_{lk}^- and W_{lk}^+ are the weight coefficients of each objective corresponding to priority factor P_l.

3. Results

3.1. The establishment of Vector Autoregressive Model

Due to the high requirement for data stationarity in VAR model construction, prior to building the VAR model, it is necessary to conduct Augmented Dickey-Fuller (ADF) tests on multiple influencing factors including net flow, precipitation, evaporation, and water levels of Great Lakes to verify the presence of spurious regression in the monthly time series of each factor (Prob < 0.05). The ADF test values and non-stationary probabilities for each factor are presented in Table 1 They become stationary after first-order differencing.

Table 1: ADF Test Results of Data

	ADF Statistic	*p-value*		*ADF Statistic*	*p-value*
Superior_MWL	-2.262	0.184	Ontario_evap	-1.986	0.293
Superior_evap	-3.028	0.032	Ontario_fall	-14.631	0.000
Superior_fall	-2.337	0.160	Ottawa River flow	-4.849	0.000
St. Mary's River flow	-1.372	0.596	St. Lawrence River flow	-1.025	0.744
MH_MWL	-1.360	0.601	Superior_MWL	-3.560	0.007
MH_evap	-2.154	0.224	Superior_fall	-11.814	0.000
MH_fall	-4.258	0.001	St. Mary's River flow	-4.350	0.000
St. Clair River	-1.113	0.710	MH_MWL	-4.175	0.001
Clair_MWL	-1.496	0.536	MH_evap	-15.068	0.000
Clair_evap	-3.493	0.008	St. Clair River	-6.177	0.000
Clair_fall	-8.867	0.000	Clair_MWL	-4.598	0.000
Detroit River	-1.160	0.690	Detroit River	-6.061	0.000
Erie_MWL	-1.764	0.399	Erie_MWL	-5.114	0.000
Erie_evap	-3.611	0.006	Niagara River flow	-6.604	0.000
Erie_fall	-13.309	0.000	Ontario_evap	-14.584	0.000
Niagara River flow	-0.733	0.838	St. Lawrence River flow	-8.057	0.000
Ontario_MWL	-4.024	0.001			

Subsequently, the Granger causality test was utilized to assess the associations among the selected variables, eliminating factors that did not exhibit causal relationships. It was noted that only a small subset of variables did not pass thesignificance test. In cases of small sample sizes or non-normal data, the results of Granger causality tests may be less stable, leading to a degree of fragility. Therefore, these variables can serve as input factors for constructing the VAR model.

3.2. The establishment of Goal Programming Model

(1) We build the final goal programming model as follows:

$$\begin{aligned} minf = {} & P_1(d_1^- + d_2^- + d_3^- + d_4^-) + P_2(d_5^+ + d_6^- + d_7^- + d_8^+ + d_9^- + d_{10}^+ + d_{11}^+ + d_{12}^-) \\ & + P_3 d_{13}^+ + P_4(d_{14}^- + d_{15}^+) + P_5 d_{16}^+ \end{aligned} \tag{6}$$

Hard constraints:

Starting from practical considerations, we stipulate that the deviation ratio of the optimal monthly water level and its variation rate from the historical monthly actual values should not exceed 1%; similarly, and the deviation ratio of the optimal monthly flow rate and its variation rate from the historical monthly actual values should also not exceed 1%.

$$s.t.\begin{cases} 0.99 s_{it} \le x_{it} \le 1.01 s_{it} \\ 0.99 \triangle s_{it} \le \triangle x_{it} \le 1.01 \triangle s_{it} \\ 0.99 d_{it} \le z_{jt} \le 1.01 d_{it} \\ 0.99 \triangle d_{it} \le \triangle z_{jt} \le 1.01 \triangle d_{it} \end{cases} \tag{7}$$

Soft constraints:

Scholars have different views on how stakeholders are classified. The "Mitchell classification," a method proposed by Mitchell, is widely used for stakeholder identification due to its effectiveness. This study applies the "Mitchell classification" to the water rights market in the Shule River Basin, using legitimacy, power, and urgency as key indicators. Stakeholders are categorized as direct, indirect, potential, or non-stakeholders based on a comprehensive assessment, resulting in a stakeholder identification matrix for the basin water rights market, as shown in Table 2.

Table 2: Stakeholders in Water Levels

Category	*Legitimacy*	*Power*	*Urgency*	*Classification*
Agricultural Users	High	High	High	Direct
Industrial Users	High	High	High	Direct
Hydropower Company	High	High	High	Direct
Government Agencies	High	High	Medium	Indirect
Water Users Association	Medium	High	High	Indirect
Environmental Protection Agency	Medium	Medium	High	Indirect
Irrigation Water Supply Units	Medium	Medium	Medium	Indirect
Domestic Water Management Department	Medium	Medium	Low	Potential
Intermediary Institutions	Low	Medium	Low	Potential
Media	Low	Medium	Low	Potential
Scholars	Low	Medium	Low	Potential
Others	Low	Low	Low	Non-stakeholders

Based on the research above, we can categorize soft constraints into five priority levels.

Priority 1:

(1) Ensure that the actual value aligns with the optimal value in terms of the changing trends on a monthly and seasonal basis.

$$s.t.\begin{cases}\frac{\triangle s_{it}}{\triangle x_{it}}+d_1^- - d_1^+ = 0\ (t{:}\,month)\\ \frac{\triangle s_{it}}{\triangle x_{it}}+d_2^- - d_2^+ = 0\ (t{:}\,season)\\ \frac{\triangle d_{jt}}{\triangle z_{jt}}+d_3^- - d_3^+ = 0\ (t{:}\,month)\\ \frac{\triangle d_{jt}}{\triangle z_{jt}}+d_4^- - d_4^+ = 0\ (t{:}\,season)\end{cases} \tag{8}$$

Priority 2: The IJC's criteria for regulating outflows explicitly recognizes the needs of three major interest groups: riparian, hydropower and commercial navigation. To protect the surrounding ecological environment and ensure the sustainable utilization of water resources, the following constraints are imposed to ensure that the water levels meet specific management requirements and environmental standards:

(2) According to AB Griffioen [5], from May to September, the absolute difference between the water level of Lake Ontario and the average water level does not exceed 2 inches (0.6069meters), aimed at controlling the magnitude of water level fluctuations during this period to minimize potential impacts on the ecological environment and human activities around the lake.

(3) According to Scott-Shaw B C Direnzo [6], from March 3rd to December 12th, the water level of Lake Ontario should be maintained between 74.15 meters and 75.37 meters. This regulation is based on water resource management policies by relevant government agencies, established according to long-term hydrological data, environmental protection regulations, shipping safety standards, and public interests.

(4) According to Cho [7], the flow of the Niagara River should be greater than or equal to the minimum hydroelectric flow and less than or equal to the design flow, using the average flow of 6000 cubic meters per second at the Sir Adam Beck Hydroelectric Generation Station as the standard.

(5) According to Brammeier [8], the flow of the St. Lawrence River should be greater than or equal to the minimum hydroelectric level and less than or equal to the design flow, with 8200 cubic meters per second as the standard at the Beauharnois generating station.

(6) According to Backer [9], the mean water level should be between the low water level and the high-water level which, is managed under the auspices of the IJC and its International St. Lawrence River Board of Control.

$$s.t.\begin{cases} x_{5t} - s_{5t} + d_5^- - d_5^+ = 0.6069\ (from\ May\ to\ September) \\ x_{5t} - s_{5t} + d_6^- - d_6^+ = -0.6069\ (from\ May\ to\ September) \\ x_{5t} - 74.15 + d_7^- - d_7^+ = 0\ (from\ March\ to\ December) \\ x_{5t} - 75.37 + d_8^- - d_8^+ = 0\ (from\ March\ to\ December) \\ z_{4t} - 6000 + d_9^- - d_9^+ = 0 \\ z_{5t} - 8200 + d_{10}^- - d_{10}^+ = 0 \\ x_{it} - h_i + d_{11}^- - d_{11}^+ = 0 \\ x_{it} - l_i + d_{12}^- - d_{12}^+ = 0 \end{cases} \tag{9}$$

Priority 3:

(7) According to Craiu [10], to prevent the occurrence of natural disasters, the optimal water level of Lake St. Clair should be lower than the historical monthly average water level.

$$t.x_{3t} - s_{3t} + d_{13}^- - d_{13}^+ = 0 \tag{10}$$

Priority 4: The region's glacial history and the tremendous influence of the lakes themselves create unique conditions that support a wealth of biological diversity, including more than 130 rare species and ecosystems. The constraints are as follows:

(8) According to Van [11], the average depth of the St. Mary River is 6 meters, and the average width of the rapids section is 810 meters. To ensure the spawning of local species, the flow velocity needs to exceed $0.4m^3/s$.

(9) According to Qiuhua L [12], to prevent the invasion of alien species, the flow velocity needs to be less than 0.5 (the spawning season for alien species is from May to July) $Q = WVD$ ($Q = Flow\ rate, W = Width\ of\ the\ river; V = Velocity\ of\ the\ water; D = Depth\ of\ the\ water$)

$$s.t.\begin{cases} z_{1t} - 1944 + d_{14}^- - d_{14}^+ = 0 \\ z_{1t} - 2430 + d_{15}^- - d_{15}^+ = 0\ (from\ May\ to\ July) \end{cases} \tag{11}$$

Priority 5:

(10) According to Yaseen[13], for normal navigation of ships, the water level of Lake St. Clair must not exceed the high-water level.

$$s.t.x_{3t} - h_3 + d_{16}^- - d_{16}^+ = 0 \tag{12}$$

3.3. Analysis of experimental results

According to the Bayesian Information Criterion (BIC), we have determined the optimal lag order to be 1. VAR model parameter estimation results and model accuracy is shown as Table 3.

Table 3: The Evaluation Results for Model Accuracy

	MSE	*RMSE*	*MAE*
E1	0.024463743	0.156408897	0.103084438
E2	0.013909942	0.117940417	0.075942305
E3	0.020051806	0.141604401	0.090776206
E4	0.024526989	0.156610949	0.107094609
E5	0.023075898	0.151907532	0.105924581

From the table, it can be observed that the model's MSE (Mean Squared Error), RMSE (Root Mean Square Error), and MAE (Mean Absolute Error) are relatively small, indicating that the models' predictive results exhibit minimal deviation from the actual values, hence reflecting a high level of accuracy. The optimal water levels of the Great Lakes at any time of the year are shown in Table 4.

Table 4: The Optimal Water Levels of the Great Lakes throughout the Year

Month	*Lake Superior*	*Lake Michigan and Lake Huron*	*Lake St. Clair*	*Lake Erie*	*Lake Ontario*
Jan	183.3155861	176.1572047	174.9631	174.168	74.67472
Feb	183.2289762	176.2178408	174.8659	174.1243	74.68309
Mar	183.1211431	176.176869	175.062	174.2786	74.74877
Apr	183.1654665	176.2693245	175.152	174.3859	74.90138
May	183.3179922	176.3950378	175.1793	174.4408	75.11171
Jun	183.3813976	176.468336	175.2248	174.4863	75.16894
Jul	183.4960094	176.4816575	175.2948	174.4309	75.11329
Aug	183.5086713	176.4590498	175.2366	174.3596	74.99834
Sep	183.5130997	176.3748768	175.1833	174.2532	74.74964
Oct	183.4694093	176.347734	175.0855	174.1813	74.63704
Nov	183.4446449	176.2658941	175.0162	174.0899	74.54463
Dec	183.3297911	176.2733069	174.9353	174.1276	74.52844

The historical optimal water levels of Great Lakes, as shown in Figure 2, lead to the conclusion that the majority of historical optimal water levels fall within a small range of variation around the historical average levels.

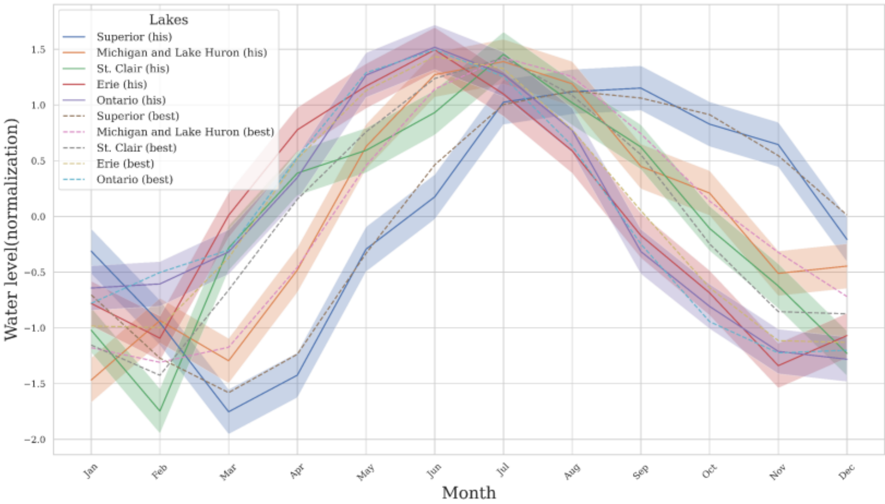

Figure 2: The Optimal Water Levels of the Great Lakes throughout the Year

Specifically, as shown in Figure 3 the historical optimal water levels of Lake Ontario are situated around the historical average water level, without surpassing the thresholds of high and low water levels.

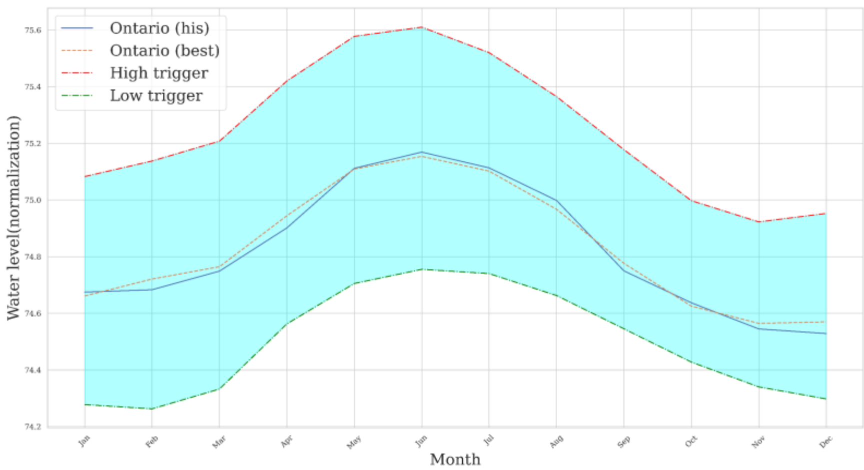

Figure 3: The Water Level Conditions of Lake Ontario throughout the Year

4. Conclusions

The Great Lakes, the largest group of freshwater lakes globally, wield considerable influence on residents, ecosystems, water resource management, shipping, and tourism industries. This paper aims to tackle water level variability by integrating network science, goal programming algorithms, and Model Predictive Control, thereby establishing a comprehensive and adaptable model. Initially, we developed a Vector Autoregression Model for the Great Lakes and connecting river flows from Lake Superior to the Atlantic Ocean, yielding the network mode. Subsequently, we constructed a Goal Programming Model to determine optimal water levels for the Great Lakes throughout the year, informed by extensive literature review and priority rankings, resulting in the depiction of optimal water levels. Leveraging insights from Plan 2014, we focused our analysis on Lake Ontario's water levels and the factors influencing them.

We conducted an analysis of the model's sensitivity to dam operations and environmental changes (evaporation, precipitation, runoff, ice concentration), and the result of test goes well.

References

[1] Azad A S, Sokkalingam R, Daud H, et al. Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study[J]. Sustainability, 2022, 14.

[2] Xing W Y, Bai Y L, Ding L, et al. Application of a hybrid model based on GA-ELMAN neural networks and VMD doubleprocessing in water level prediction [J]. Journal of hydroinformatics, 2022.

[3] Wang C, Cuan W, Jia L. Weighted Combined Water Level Prediction Based on Nonlinear Programming Genetic Algorithm [J]. 2022 7th International Conference on Computational Intelligence and Applications (ICCIA), 2022: 140-145.

[4] Ouma Y O, Moalafhi D B, Anderson G, et al. Dam Water Level Prediction Using Vector Auto Regression, Random Forest Regression and MLP-ANN Models Based on Land-Use and Climate Factors [J]. Sustainability, 2022, 14.

[5] Griffioen A B, Van Keeken O A, Hamer A L, et al. Passage efficiency and behaviour of sea lampreys (Petromyzon marinus, Linnaeus 1758) at a large marine-freshwater barrier [J]. River Research and Applications, 2022(5):38.

[6] Scott-Shaw B C, Everson C S. Water-use dynamics of an alien-invaded riparian forest within the summer rainfall zone of South Africa [J]. Hydrology and Earth System Sciences, 2019 (3).

[7] Cho S Y, Koo M H, Cho B W, et al. Factors Controlling the Spatial and Temporal Variability in Groundwater 222Rn and U Levels [J]. Water, 2019, 11(9):1796.

[8] Brammeier J. Our Great Lakes [J]. Proceedings of the Marine Safety & Security Council, 2022.

[9] Becker C D, Fickeisen D H, Montgomery J C. Assessment of impacts from water level fluctuations on fish in the Hanford Reach, Columbia River [J]. Environment Lences, 2019.

[10] Craiu I, Fidel N, Stan M F, et al. Measuring and Maintaining the Water Level in a Tank for Efficient Use in Irrigation [C]//2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). 2019.

[11] Van d C D, Van der Zee, S. E. A. T. M, Leijnse A. Anticipatory Drainage Base Management for Groundwater Level Optimization [J]. Water Resources Research, 2021(11): 57.

[12] Qiuhua L, Salmaso N. Impact of water level fluctuations on the development [J]. 2019.

[13] Yaseen Z M, Deo R C, Ebtehaj I, et al. hybrid data intelligent models and applications for water level prediction hybrid data intelligent models and applications for water level prediction [J]. 2019.