

GL-PaD: Great Lakes' Predictor-and-Decision maker

Summary

This paper focuses on optimizing water level in the Great Lakes. It aims to coordinate the interests of stakeholders and develop a better water level management plan by combining environmental factors and historical data. The superiority of the algorithm is verified through simulation.

For requirement 1, we propose a model to evaluate the interests of each stakeholder, using historical water level data to categorize them into primary and secondary groups. Additionally, the lakes are classified based on water level fluctuation magnitude, leading to tailored optimization strategies for each. The established water levels for the five lakes at 183.35, 176.33, 175.10, 174.28, and 74.82 meters serve as benchmarks for future optimization control strategies.

For requirement 2, we presents a joint differential predictor and MPC decision-maker model, modified for iterative optimization via machine learning. Trained on 2002-2022 data, the model achieved convergence and outperformed the traditional MPC decision model in water level control. It consistently maintained optimal water levels across the Great Lakes, within a defined error margin.

For requirement 3, we follows the model from requirement 2, retaining its structure in the prediction phase. Inputs like 2017's actual rainfall were used to simulate water levels under the dam decision model. Comparing these simulations with actual 2017 water levels shows our algorithm effectively lowered lake levels in the watershed, aligning them more closely with optimal levels. These results confirm the model's effectiveness in extreme weather and its lower sensitivity to specific environmental variables.

For requirement 4, we assesses the model's sensitivity to environmental factors using the Sobol algorithm via the SAlib Python library. The algorithm separates the impact of environmental variables into two measures: a main effect index for singular impacts and a total effect index for combined effects. Utilizing the Monte Carlo method, numerous independent, normally distributed samples were generated in a hypercube. Findings show that tributary inflow and ice stagnation rate significantly influence the model, more so than other variables. Moreover, simultaneous changes in multiple variables have a greater impact on the model than individual variable changes.

For requirement 5, we analyze stakeholders around Lake Ontario and the St. Lawrence River, evaluating their importance based on economic, ecological, and sensitivity aspects. The analysis not only adapt optimal water levels for each season, but also incorporates the influence of seasonal changes in river dynamics and spring ice melt on the St. Lawrence River. This leads to a comprehensive, multi-objective optimization model, which is validated through training and analysis, affirming the effectiveness of our control strategy.

Keywords: Water level control; Predictor; Desicion maker; Machine learning

Contents

1	Introduction	3
1.1	Background	3
1.2	Restate of the Problem	3
1.3	Literature Review	4
1.4	Our Work	4
2	Assumptions	5
3	Notations	5
4	Requirement 1	6
4.1	Model Preparation: Historical Water Level Data Exploration	6
4.2	Modeling Optimal Water Level Determination for the Great Lakes	7
4.3	Model Solving and Result	8
5	Requirement 2	9
5.1	PaD: Predictor-and-Decision Maker	9
5.2	Model Revision	11
5.3	Model Solving: Based on Machine Learning	12
5.4	Result	13
6	Requirement 3	15
6.1	Integration of Environment Variables for Predictive Simulation	15
6.2	Analysis of Result	16
7	Requirement 4	17
7.1	Sobol-based Sensitivity Analysis	17
7.2	Model Solving Based on SAlib	18
7.3	Result	19
8	Requirement 5	20
8.1	Multi-Objective Optimization of Lake Ontario Water Level	20
8.2	Model Solving: Based on Genetic Algorithms	22
8.3	Result	23
9	Conclusion	23

1 Introduction

1.1 Background

The Great Lakes, straddling the United States and Canada, represent the world's largest collective body of freshwater, facilitating a range of activities, including fishing, recreation, and shipping. Effective management of these water levels is essential to mitigate flooding risks and sustain economic pursuits. Water level regulation is influenced by a confluence of factors such as temperature, wind, and precipitation, and is partly managed through engineered structures like the Soo Locks and Moses-Saunders Dam. Nevertheless, uncontrollable natural phenomena such as rainfall and evaporation significantly impact the ecosystem and stakeholder interests. The flow dynamics of the Great Lakes have intricate interdependencies and uncertainties that are challenging to resolve[1].



Figure 1: The Great Lakes and Dams

1.2 Restate of the Problem

To assist in managing the control mechanisms (two dams) that directly affect water levels in the Great Lakes flow network, we made a network model for the Great Lakes and considered the following problems:

- Determine the optimal water level for the Great Lakes at any given time of the year, taking into account the desires of the various stakeholders.
- Building algorithms based on inflow and outflow data from the five lakes to maintain optimal water levels in the five lakes. The algorithms need to be able to adapt to weather changes, seasonal fluctuations, and other factors.
- Analyze the sensitivity of the control algorithm to outflows from the two control dams, and conduct an in-depth analysis of the 2017 data to assess whether the new controls better meet the expectations of different stakeholders.
- Analyze the impact of environmental changes on the algorithm and provide insight into the algorithm's adaptability and sensitivity to environmental changes such as precipitation, winter snowpack, and ice jams.
- Concentrate on analyzing the stakeholders and factors affecting Lake Ontario.

1.3 Literature Review

Hydrologic modeling's inception in the Great Lakes region dates back to the early 20th century, initially serving to assess water level changes due to diversions. Early works, exemplified by Freeman in 1926, using outflow rating curves for each lake to predict changes in water surface elevation due to diversions[2]. In 1977, Frank H. Quinn's innovative water quantity model encompassed Lakes Michigan, Huron, St. Clair, and Erie, integrating their channels through the continuity equation of lake flows. He delved into the Runge-Kutta and Newton-Raphson algorithms, as well as second-order finite difference methods, for precise water level predictions[3]. By 1994, Deborah H. Lee and her team advanced this field by modeling Lake Ontario's water dynamics using stochastically generated Great Lakes basin supplies. Their work highlighted existing regulatory frameworks' shortcomings under severe water supply conditions, proposing essential model adjustments for enhanced simulation accuracy in extreme scenarios[4]. Nevertheless, considerations for the model's real-time optimization and adaptability to extreme weather conditions remained unaddressed.

1.4 Our Work

As Figure 2, This paper makes the following contributions:

- Devised an algorithm to balance stakeholder interests for optimal water level determination.
- Engineered a dynamic dam control strategy, integrating predictive models with decision-making frameworks.
- Based on the predicted 2024 water level, we put the algorithm into simulation and analyzed the sensitivity and robustness of the model

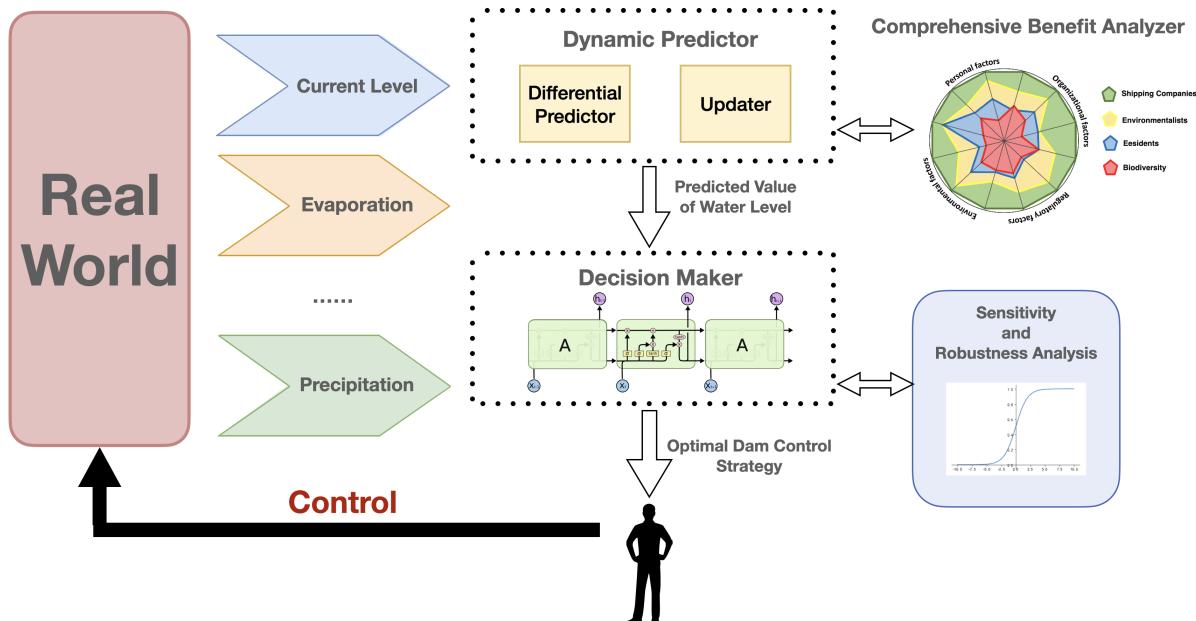


Figure 2: The Flow Chart of Our Work

2 Assumptions

1. Variations in lake level do not influence the lake's surface area.
2. For each lake, the precipitation level is uniform across all locations on the lake's surface.
3. The time required for dam adjustments is negligible..
4. Human activities (e.g., residential water use) have no effect on lake level.

3 Notations

Table 1: Description of Notations

Symbol	Description
Z_i	Instantaneous water level at the i-th lake
Z_{best}^i	Optimal water level of the i-th lake
Z_{max}^i	Maximum water level of the i-th lake
Z_{mean}^i	Average water level of the i-th lake
A_i	Area of the i-th lake
P_i	Precipitation of the i-th lake
R_i	Tributary inflow to the i-th lake
D_i	Subsystem diversion for the i-th lake
EV_i	Evaporation rate of the i-th lake
QR_i	Stagnant ice rate for the i-th lake
K_i, ym_i	Optimizable parameters describing the i-th lake
$QSM(t)$	Outflow from Lake Superior controlled by dam
$QSL(t)$	Outflow from Lake Ontario controlled by dam
$\mathbf{Q}, \mathbf{R}, \mathbf{F}$	The weight coefficient matrix of MPC algorithm
\mathbf{J}	The objective function of a optimization problem
\mathbf{F}, \mathbf{G}	The Predictor model and decision maker model
$loss_i$	Economic sensitivity measurement indicators
F_j	Flow of the St. Lawrence River in season i
V_i	Effect of parameter i on the output of the objective function
S_i	Contribution of i to the output variance
d^+, d^-	The positive and negative deviation variable

4 Requirement 1

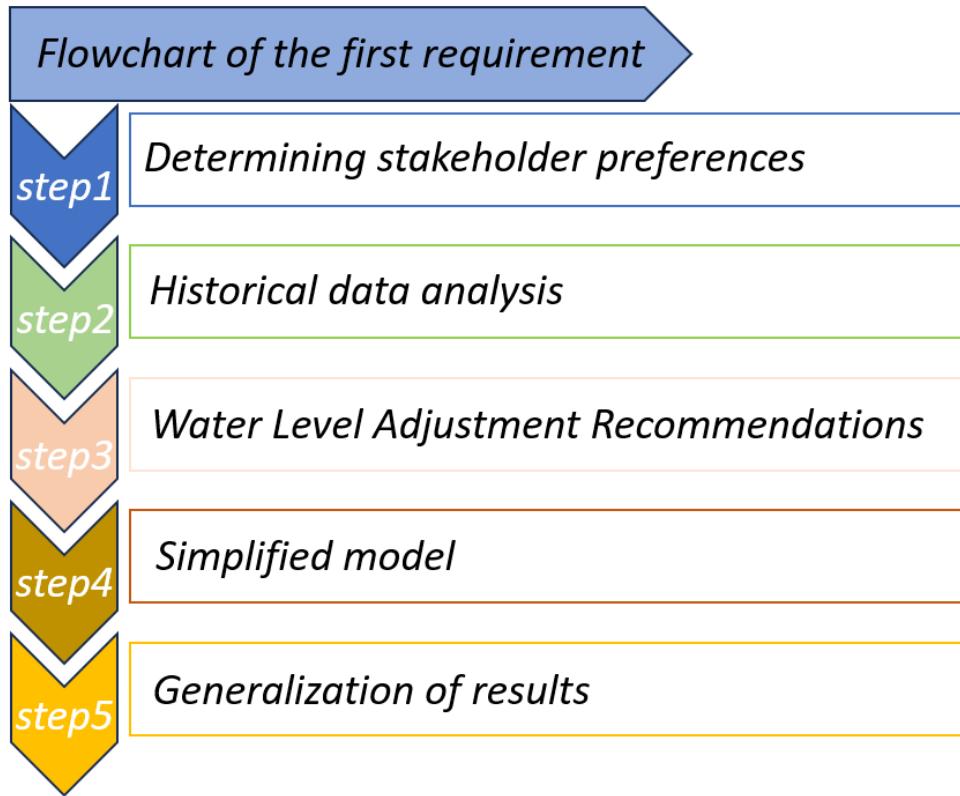


Figure 3: The Flow Chart of the Requirement 1

4.1 Model Preparation: Historical Water Level Data Exploration

Firstly, a meticulous evaluation of stakeholder preferences is imperative. Property owners and visitors near Lake Ontario prefer a moderate and stable water level. Conversely, environmentalists advocate for the preservation of natural seasonal water level fluctuations to protect the ecosystem. Additionally, shipping entities operating on the St. Lawrence River require sustained high water levels. These diverse considerations necessitate a comprehensive analysis, which requires first analyzing the historical data.

We conducted an analysis of historical lake level data, charting the average, maximum, and minimum monthly water levels for four of the lakes. Subsequently, we made a graphical representation of water level fluctuations for all five lakes.

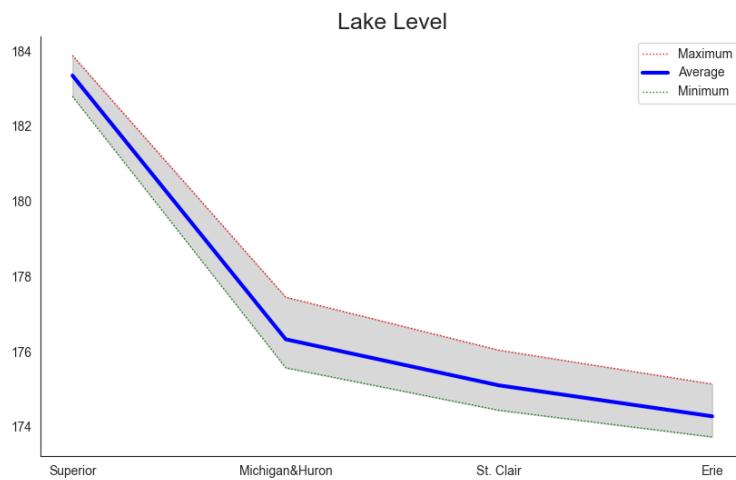


Figure 4: Historical Lake Level

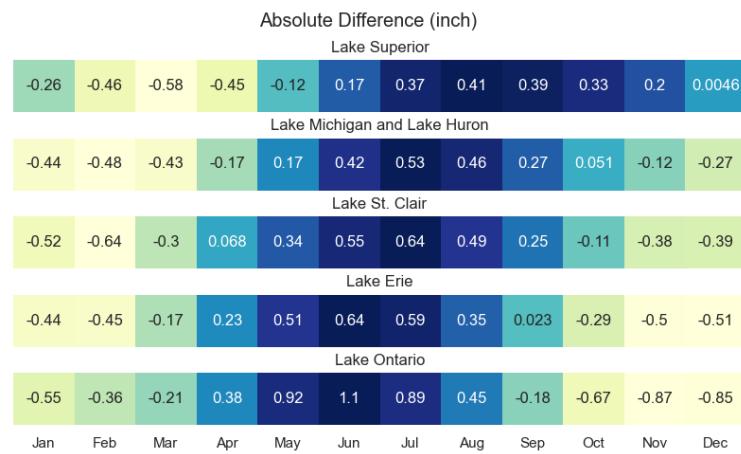


Figure 5: Lake Level Fluctuation

Figure 4 and Figure 5 show that, with the exception of Lake Ontario, the levels of the four major lakes have fluctuated less over the years and are in a relatively stable state.

4.2 Modeling Optimal Water Level Determination for the Great Lakes

Among these stakeholders, shipping occupies a more important position[5]. Therefore, in order to take care of the interests of the major stakeholders, we would like to raise the optimum water level and keep it as stable as possible while respecting the laws of nature and historical precedents. Historical data suggests that with the exception of Lake Ontario, lake levels are inherently more stable. Considering the floods that occurred in 2017 and 2019, it is more reasonable to use 75% of the maximum water level as the optimal

level for these four lakes. We therefore modeled the following:

$$\begin{aligned}
 Z_{\text{best}}^i &= 0.75 \times Z_{\text{max}}^i \\
 Z_{\text{max}}^i &= \max(Z_{\text{mean}}^i) \\
 Z_{\text{mean}}^i &= \frac{1}{N_{\text{Year}}} \sum_{N_{\text{Year}}} Z_{\text{month}}^i \\
 i &= \{1, 2, 3, 4\}
 \end{aligned} \tag{1}$$

As for Lake Ontario, we can use the following single-objective optimization model:

$$\begin{aligned}
 \min f(Z_{\text{best}}^5) &= \sum_{m=1}^{12} (Z_{\text{best}}^5 - Z_{\text{mean}}^5)^2 \\
 \text{s.t. } &Z' \leq Z_{\text{best}}^5 \leq Z'' \\
 &Z'' = Z_{\text{max}}^5 \times 0.75 \\
 &Z' = Z_{\text{min}}^5 \times 1.1
 \end{aligned} \tag{2}$$

4.3 Model Solving and Result

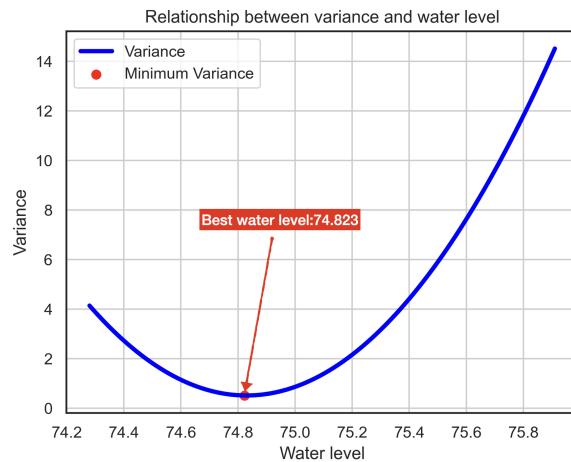


Figure 6: Relationship Between Variance and Water Level

To resolve Equation 2, the feasible domain was systematically traversed in increments of $0.01(L'' - L')$ to identify the minimum variance. The corresponding water level at this juncture is deemed the optimal water level Z_{best}^5 for Lake Ontario. The calculation culminates in the following result:

Table 2: Requirement 1 Result

Superior	Michigan&Huron	St. Clair	Erie	Ontario
183.35	176.33	175.10	174.28	74.82

5 Requirement 2

5.1 PaD: Predictor-and-Decsion Maker

As Figure 7, We have developed a comprehensive schematic network model for the Great Lakes, which is based on their geographical proximities. This model incorporates key environmental factors such as evaporation, precipitation, runoff and dicsersions.

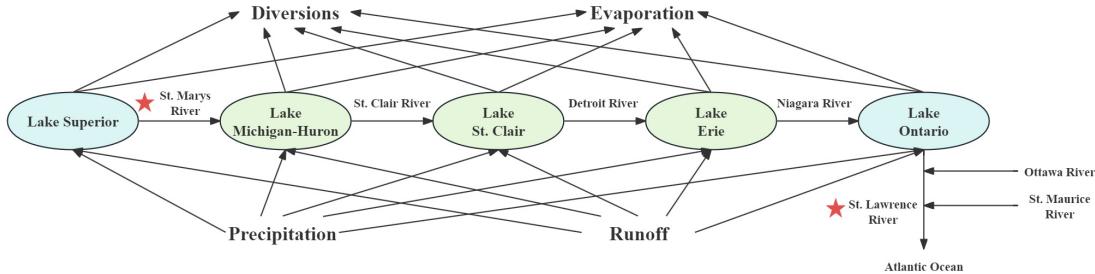


Figure 7: Comprehensive Schematic Network Model for the Great Lakes

We formulated five differential equations based on the flow continuity equation[3] (change in lake volume = volume of water entering the lake - volume of water flowing out of the lake) to delineate the interplay between water level, evaporation, precipitation, runoff and dicsersions, as derived from the Figure 7 :

$$\begin{aligned}
 \frac{dZ_1}{dt} &= \frac{1}{A_1} [P_1 + R_1 - D_1 - EV_1 - QSM(t)] \\
 \frac{dZ_2}{dt} &= \frac{1}{A_2} [P_2 + R_2 + QSM + QR_2 - EV_2 - D_2 - K_2 (Z_2 - ym_2)^2 (Z_2 - Z_3)^{0.5}] \\
 \frac{dZ_3}{dt} &= \frac{1}{A_3} [P_3 + R_3 + K_2 (Z_2 - ym_2)^2 (Z_2 - Z_3)^{0.5} - QR_2 - EV_3 \\
 &\quad - K_3 (Z_3 - ym_3)^2 (Z_3 - Z_4)^{0.5} + QR_3] \\
 \frac{dZ_4}{dt} &= \frac{1}{A_4} [P_4 + R_4 + K_3 (Z_3 - ym_3)^2 (Z_3 - Z_4)^{0.5} - QR_3 - EV_4 - D_4 \\
 &\quad - K_4 (Z_4 - ym_4)^2 + QR_4] \\
 \frac{dZ_5}{dt} &= \frac{1}{A_5} [P_5 + R_5 + K_4 (Z_4 - ym_4)^2 - QR_4 - EV_5 - D_5 - QSL(t)]
 \end{aligned} \tag{3}$$

The term K_i, ym_i in the stage-fall-discharge equations represents the nature of the river itself. They can be calculated by measuring the flow of each river.

In order to address this problem, we firstly approach the solution by treating key factors — precipitation, evaporation rate, runoff, and water usage — that influence lake levels as constants over a specific period. By utilizing historical data on river flow and lake levels, we calculate the parameters K_i and ym_i for the stage-fall-discharge equations. This process allows us to represent this multi-input, multi-output system as a nonlinear

set of differential equations, which are as follows:

$$\frac{d\mathbf{Z}}{dt} = f(\mathbf{Z}, \mathbf{u}) \quad (4)$$

Within this framework, the discharge volumes from two dams, labeled as \mathbf{u} , serve as the control variables, while the water levels in the five lakes, denoted by \mathbf{z} , act as the state variables. We adopt the model predictive control approach, setting the optimization goal to minimize the sum of squared deviations between the actual lake levels and their optimal states, under reasonable constraints. This method facilitates the determination of the discharge volumes, \mathbf{u} .

Algorithm 1 Model Prediction Control Algorithm

Require: Actual water level \mathbf{Z} ,

Ensure: discharge volumes of the two dams \mathbf{u}

Step 1: Data Acquisition

$$\mathbf{Z}_{k+1} = f(\mathbf{Z}_k, \mathbf{u}_k)$$

Step 2: Optimization Based on $\mathbf{u}_k, \dots, \mathbf{u}_{k+N}$

$$\min \mathbf{J} = \sum_{k=1}^{N-1} \mathbf{Z}_k^T \mathbf{Q} \mathbf{Z}_k + \mathbf{u}_k^T \mathbf{R} \mathbf{u}_k + \mathbf{Z}_N^T \mathbf{F} \mathbf{Z}_N$$

Step 3: Exert the First Control Variable \mathbf{u}_k

REPEAT Steps 1-3 to get all the control variables \mathbf{u}_k

In practical applications, various factors besides river flow influence water levels. These factors not only fluctuate over time but also demonstrate specific probabilistic distribution traits in addition to their functional relationships. Consequently, simplified models often lack precision, particularly under significant disturbances, leading to inaccuracies in the MPC approach.

To address these challenges, we propose a dynamic dam control strategy, employing differential equations as the foundational predictor \mathbf{F} [6]. This model is formally defined as:

$$\mathbf{F} : \mathbf{x}_t \mapsto \hat{\mathbf{Z}}_{t+\Delta t} \quad (5)$$

where \mathbf{x}_t represents the vector of input variables at time t , which includes the current water level, precipitation rate, evaporation rate, and the existing dam control strategy. Providing these inputs to the predictor \mathbf{F} computes the forecasted water level $\hat{\mathbf{Z}}_{t+\Delta t}$ for the next time.

Furthermore, the decision-making model, represented by \mathbf{G} , determines the optimal dam control operations for the current time based on the predicted future water level. The function \mathbf{G} is articulated as follows:

$$\mathbf{G} : \hat{\mathbf{Z}}_{t+\Delta t} \mapsto (QSM(t), QSL(t)) \quad (6)$$

In this context, the input for the decision maker G is the projected water level $\hat{Z}_{t+\Delta t}$, and the outputs are the optimal dam control strategies $QSM(t)$ and $QSL(t)$.

Algorithm 2 Dam Control Strategy Algorithm

Require: Sensor measurements M , Predictor model F , Decision maker model G .

Ensure: Iterative optimization and application of dam control strategies.

Step 1: Data Acquisition and Input

$$\mathbf{x}_t = M(\text{WaterLevel}, \text{Rainfall}, \text{Evaporation}, \text{DamStrategy}, \dots)$$

Step 2: Apply Predictor Model

$$\hat{Z}_{t+\Delta t} = F(\mathbf{x}_t)$$

Step 3: Generate Decision

$$(QSM(t), QSL(t)) = G(\hat{Z}_{t+\Delta t})$$

Step 4: Optimization Loop

$$\mathbf{x}_{t+\Delta t} = M(t + \Delta t)$$

$$G = \text{Optimize}(G, \mathbf{x}_{t+\Delta t}, \hat{Z}_{t+\Delta t})$$

Step 5: Convergence and Model Refinement

REPEAT Steps 1-4 UNTIL convergence of G

5.2 Model Revision

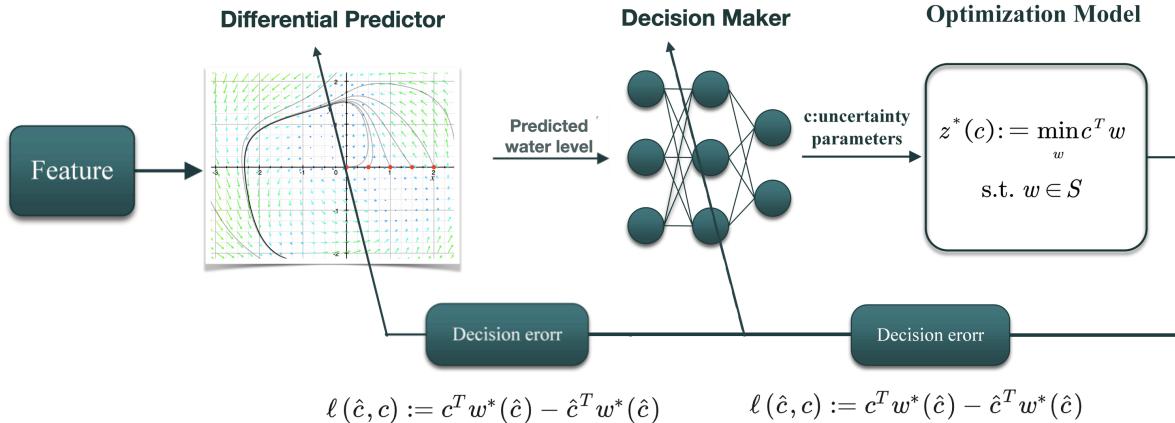


Figure 8: Revised Model

In Equations 3, the variables K_i and ym_i originate from empirical formulas[7], indicating that our predictor would initially be static and incapable of optimization over time. Consequently, we have designated K_i and ym_i as parameters to be dynamically learned and optimized with each prediction. This approach ensures simultaneous optimization of both the predictor and the decision-maker.

Algorithm 3 Dynamic Dam Control Strategy Algorithm

Require: Sensor measurements M, Predictor model F, Decision maker model G.

Ensure: Iterative optimization and application of dam control strategies.

Step 1: Data Acquisition and Input

$$\mathbf{x}_t = \mathbf{M}(\text{WaterLevel}, \text{Rainfall}, \text{Evaporation}, \text{DamStrategy}, \dots)$$

Step 2: Apply Predictor Model

$$\hat{Z}_{t+\Delta t} = \mathbf{F}(\mathbf{x}_t)$$

Step 3: Generate Decision

$$(QSM(t), QSL(t)) = \mathbf{G}(\hat{Z}_{t+\Delta t})$$

Step 4: Optimization Loop

$$\mathbf{x}_{t+\Delta t} = \mathbf{M}(t + \Delta t)$$

$$\mathbf{F}, \mathbf{G} = \text{Optimize}(\mathbf{F}, \mathbf{G}, \mathbf{x}_{t+\Delta t}, \hat{Z}_{t+\Delta t})$$

Step 5: Convergence and Model Refinement

REPEAT Steps 1-4 UNTIL convergence of \mathbf{F} and \mathbf{G}

5.3 Model Solving: Based on Machine Learning

For the MPC strategy, we set the predictive horizon to 12 and the control horizon to 2. In pursuit of optimization, the weight matrix is adjusted to emphasize control weight with a strong inclination toward water level management. For constraints, we take into account both the upper limit of dam discharge and the speed of gate adjustment.

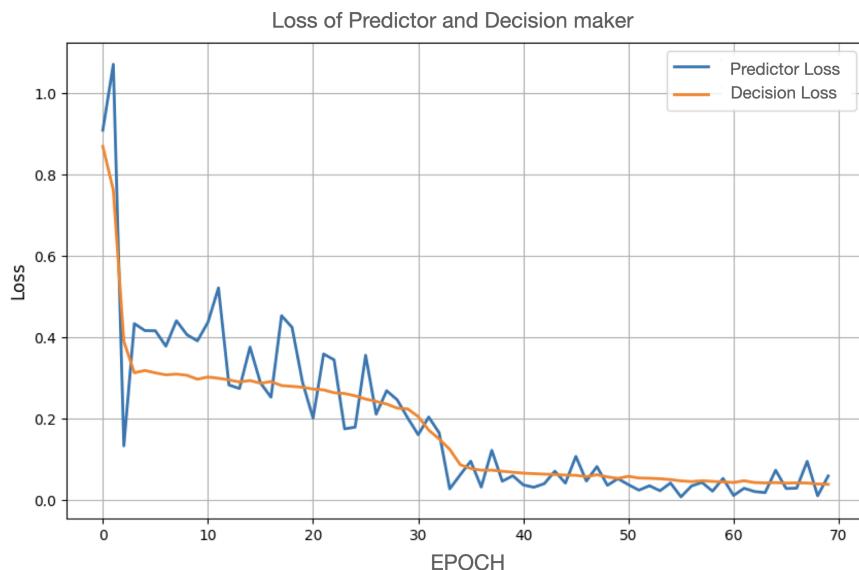


Figure 9: Loss of Predictor and Decision maker

As Figure 9, We conducted training for 70 epochs simultaneously on both the predictor and the decision maker[8], utilizing Great Lakes water level data. Furthermore, the monthly precipitation, surface runoff, and evaporation data used in the GL-PaD algorithm were obtained from two sources. Source I is the Reports on Great Lakes including water levels, forecasts[9], basin conditions, and outflows, retrieved from the US Corps of Engineers Detroit District. Source II is the Great Lakes Regional Water Use Database[10], which provides information on diversions, withdrawals, and consumption related to water usage by residents and industries in the Great Lakes area. The data from these sources were aggregated to form our model. However, it's important to note that some data lack earlier statistical results, limiting our model to using data only from 2002 onwards.

The Adam[11] optimizer was employed with a learning rate of 3e-4. It was observed that concurrent model convergence presented challenges, manifesting in a fluctuating yet overall decreasing loss trend. However, as the number of epochs increased, the loss for both models gradually diminished, ultimately achieving convergence.

5.4 Result

We conducted simulations of the dam control algorithms, utilizing forecasts of the Great Lakes' surface levels for 2024. This enabled us to determine the dam-controlled releases from Lake Superior and Lake Ontario at various points in time, shown in Figure 10:

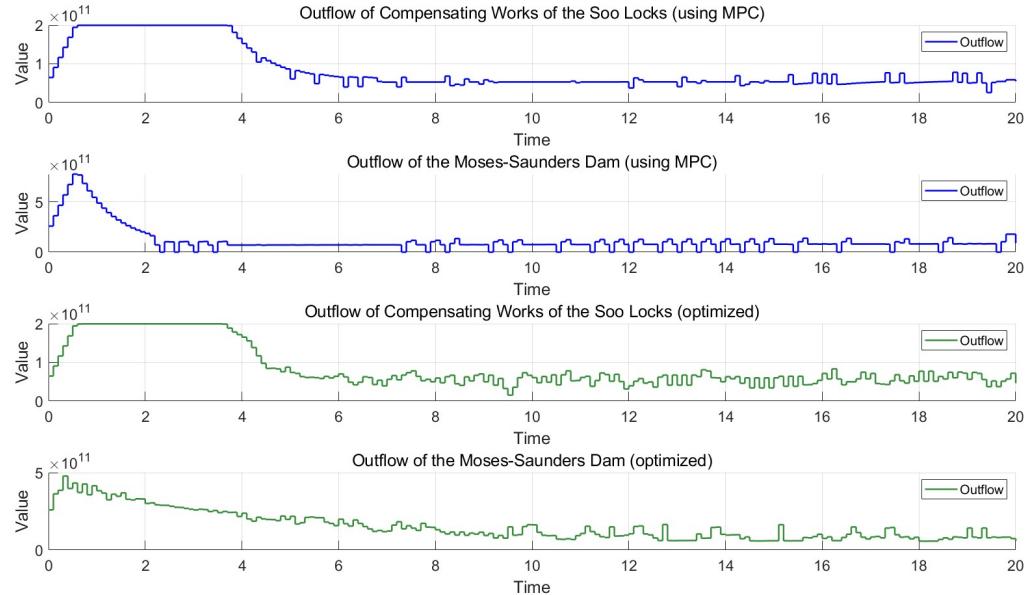


Figure 10: Outflow of the Each Dam From Each Control Strategy

Employing these strategies, we simulated the fluctuations in the Great Lakes' water levels. As Figure 11 and Figure 12 , Initial observations revealed significant deviations from the optimal levels across all the Great Lakes. However, subsequent adjustments in dam control strategies gradually aligned their water levels with the optimal benchmarks. Within an acceptable margin of error, the water levels were effectively regulated to remain close to these optimal levels.

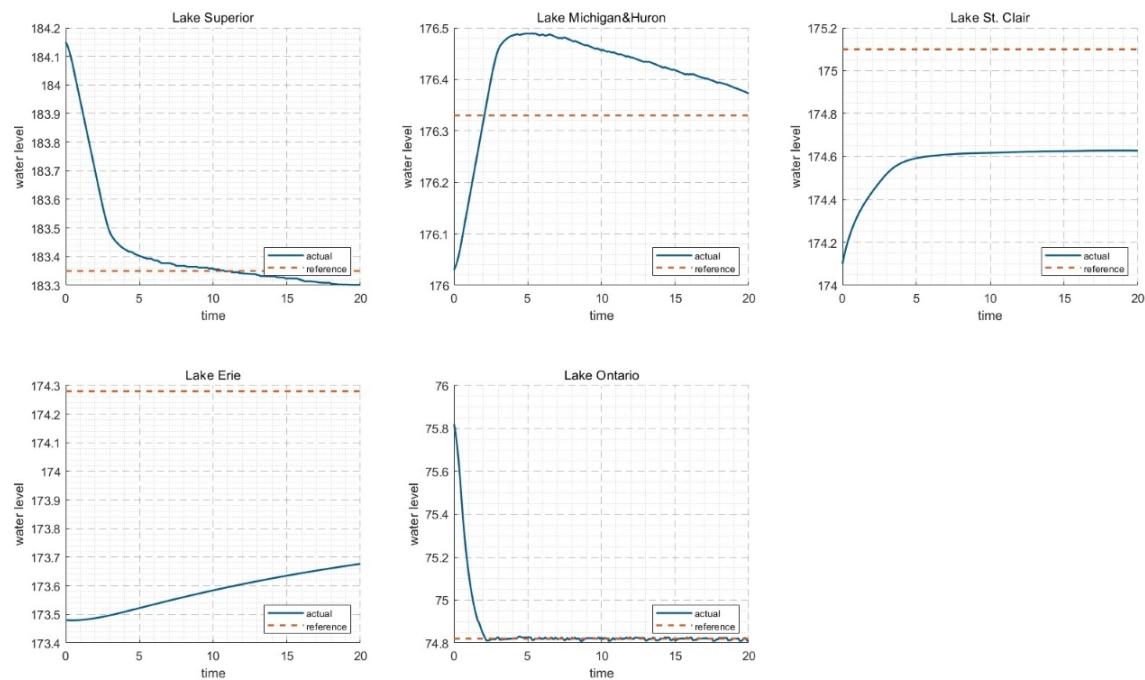


Figure 11: Water Level Under MPC Dam Control Strategy

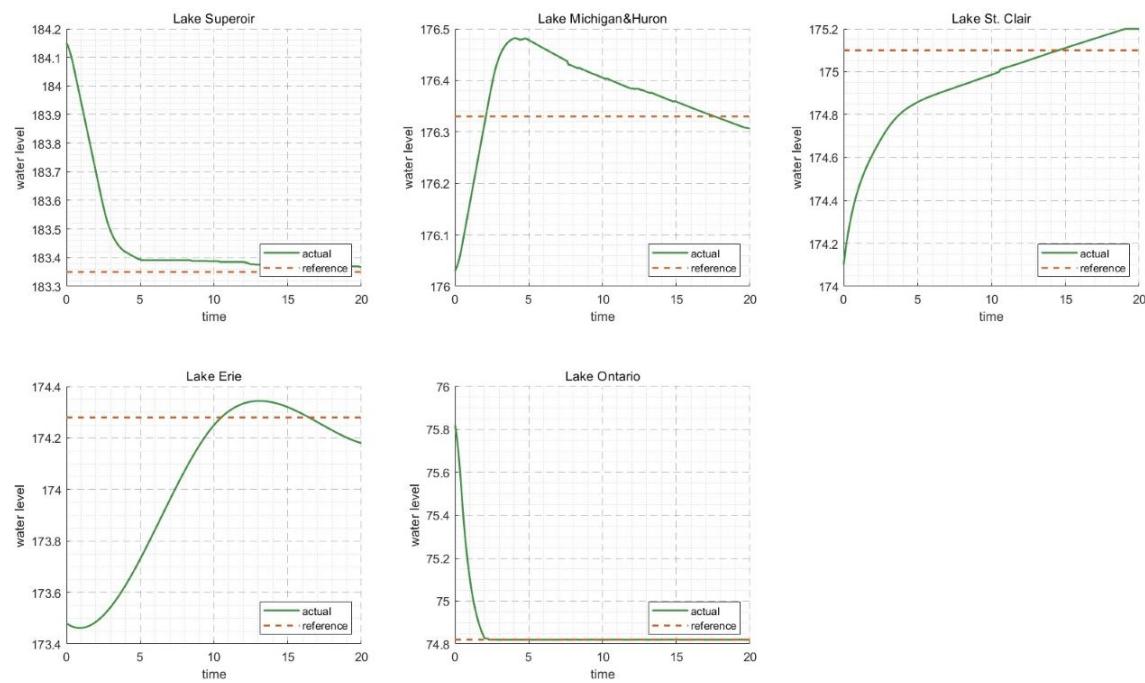


Figure 12: Water Level Under GL-PaD Control Strategy

Finally, we calculate the objective function values under the two control strategies, in order to better evaluate the control strategy for each lake level control situation, the difference between the actual water level and the expected water level in the five lakes is calculated and the sum of squares is presented in Table 3:

Table 3: Value of the objective function from two strategy

Lakes	MPC Strategy	Our Strategy
Lake Superior	8.4839	8.4971
Lake Michigan-Huron	3.1711	1.9951
Lake St. Clair	57.6949	15.4822
Lake Erie	99.6226	32.2187
Lake Ontario	5.5413	5.3031

It can be found that the revised control strategy can better control the water level of each lake, and then make the overall objective function value to achieve better.

6 Requirement 3

6.1 Integration of Environment Variables for Predictive Simulation

In order to understand the sensitivity of the two control algorithms for controlling the outflow from the dam, the level of each lake for each month of 2017 can be predicted by substituting the precipitation and evaporation rate of each lake in each month of 2017 into the above model.

Algorithm 4 2017 Water Level Simulation Algorithm

Require: Predictor model F, Decision maker model G, Real data for 2017 X.

Ensure: 2017 Simulated Water Levels $[\hat{Z}_2, \dots, \hat{Z}_{12}]$

Step 1: Calculate the Strategy for the First Month

$$\mathbf{x}_1 = [Z_1, QSM(1), QSL(1), P_1, R_1, D_1, \dots]$$

$$\hat{Z}_2 = \mathbf{F}(\mathbf{x}_1)$$

$$Q\hat{S}M(1), Q\hat{S}L(1) = \mathbf{G}(\hat{Z}_2)$$

Step 2: Calculate Projected Water Level for March-December

$$\hat{\mathbf{x}}_i = [\hat{Z}_i, Q\hat{S}M(i), Q\hat{S}L(i), P_i, R_i, D_i, \dots]$$

$$\hat{Z}_{i+1} = \mathbf{F}(\hat{\mathbf{x}}_i)$$

$$Q\hat{S}M(i), Q\hat{S}L(i) = \mathbf{G}(\hat{Z}_{i+1})$$

Step 3: Repeat Step 2 until the December WaterLevel is Predicted

6.2 Analysis of Result

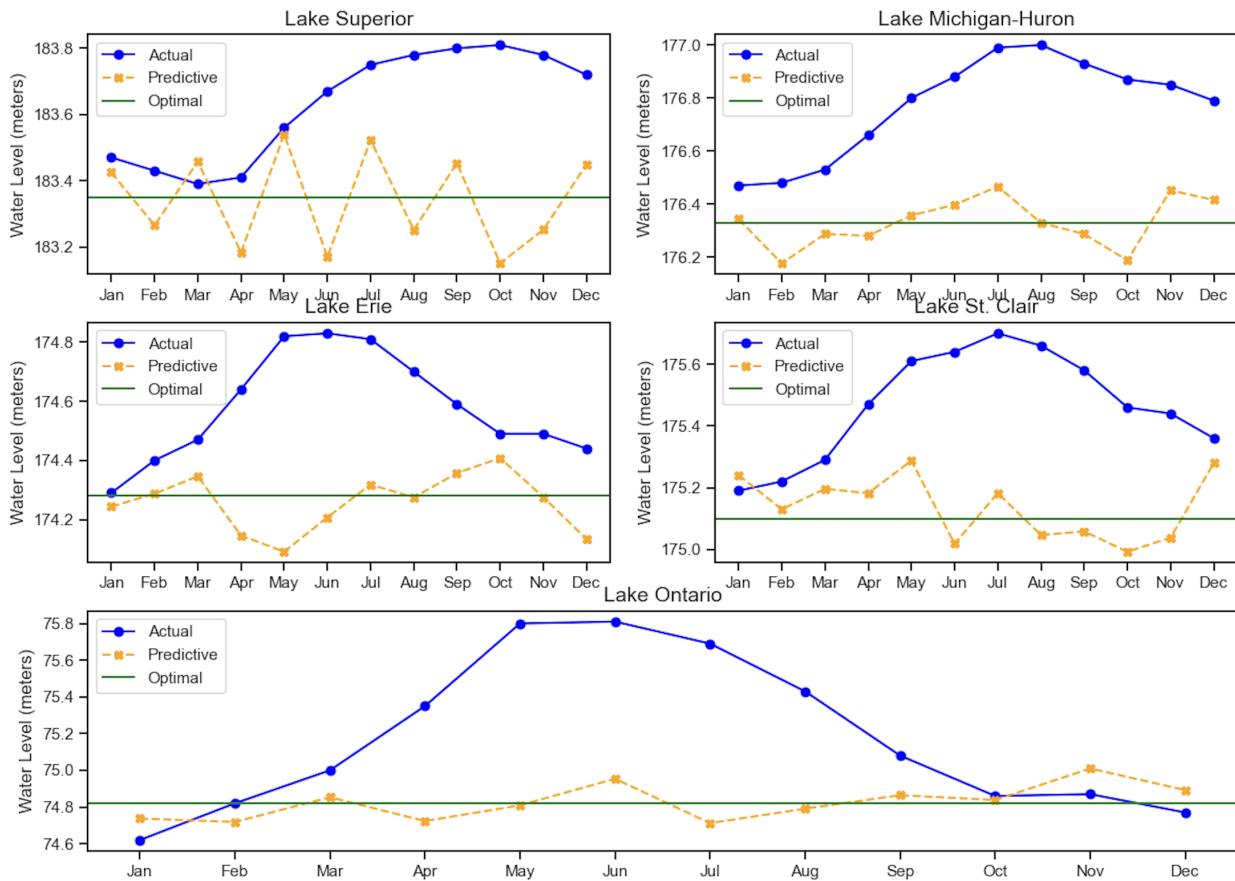


Figure 13: 2017 Monthly Water Levels: Actual vs. Predictive for the Great Lakes

The 2017 basin flooding outbreak, as depicted in the Figure 13, resulted in the Great Lakes' water levels exceeding the optimal level identified in the first question for an entire year, posing a significant threat to stakeholders' interests. Our dam control algorithm successfully lowered these levels, aligning them more closely with the optimal level than the actual recorded data. However, this required adopting a more aggressive control strategy, leading to increased fluctuations in water levels, particularly in Lake Superior. This approach might also compromise some stakeholders' interests. Despite these challenges, our model effectively managed the water levels during the 2017 flood, safeguarding the benefits of the majority.

7 Requirement 4

7.1 Sobol-based Sensitivity Analysis

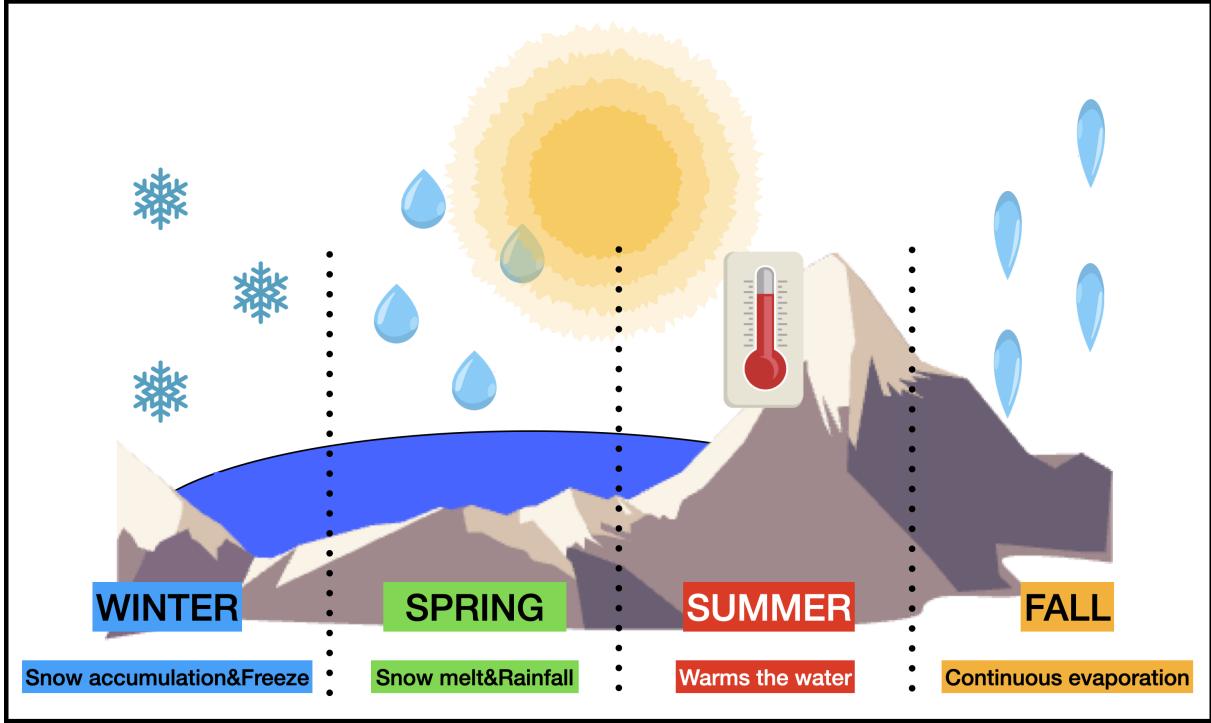


Figure 14: Environmental Impacts of Different Seasons

Clearly, the water level at any given moment is influenced by the combined effect of various environmental parameters. This influence manifests as follows: each environmental parameter impacts the predictor's accuracy, resulting in an increased discrepancy between the predicted and actual values in subsequent moments. Consequently, decisions made by the decision maker, based on these inaccurate predictions, are likely to be erroneous.

To model the effect of environmental covariates on the model, we consider the dam control model in Requirement 2 as a function:

$$QSM(t), QSL(t) = f(\mathbf{X}) \quad (7)$$

Here, \mathbf{X} represents the model's d uncertain inputs $\{X_1, X_2, \dots, X_d\}$. To maintain generality, we assume that \mathbf{X} is independently and uniformly distributed within a hypercube. This assumption allows for the decomposition of the model's output variance into individual effects of each input and their collective effects[12]:

$$\text{Var}(Y) = \sum_{i=1}^d V_i + \sum_{i < j}^d V_{ij} + \dots + V_{12\dots d} \quad (8)$$

$$V_i = \text{Var}_{X_i} (E_{\mathbf{X}_{\sim i}} (Y | X_i)), \quad (9)$$

$$V_{ij} = \text{Var}_{X_{ij}} (E_{\mathbf{X}_{\sim ij}} (Y | X_i, X_j)) - V_i - V_j \quad (10)$$

Where X_i represents each parameter of the input model, including factors like rainfall and evaporation. As environmental conditions vary, so does X_i . The influence of X_i on the output is quantitatively assessable by determining V_i , which reflects the impact of changes in individual environmental variables on the model.

$$S_i = \frac{V_i}{\text{Var}(Y)} \quad (11)$$

$$\sum_{i=1}^d S_i + \sum_{i < j}^d S_{ij} + \dots + S_{12\dots d} = 1 \quad (12)$$

S_i is the “main effect index” of X_i on the model, indicating the contribution of X_i to the output variance.

Further, we can analyze the variance of the model output (V_{ij}) when multiple environmental variables (X_i, X_j) all change, i.e., analyze the effect on the model when multiple union variables change:

$$S_{ij} = \frac{V_{ij}}{\text{Var}(Y)} \quad (13)$$

7.2 Model Solving Based on SALib



Figure 15: SALib[13]

In this research, we used SALib, a Python library, for sensitivity analysis to assess the impact of input parameters on our model. We configured these parameters to mirror diverse environmental conditions. The Sobol Sensitivity Analysis method in SALib helped calculate the Sobol indices, capturing both main and total effects. These indices quantitatively evaluated how each parameter influenced the model’s output variance. This analysis provided essential insights into the parameters significantly affecting the model’s effectiveness.

7.3 Result

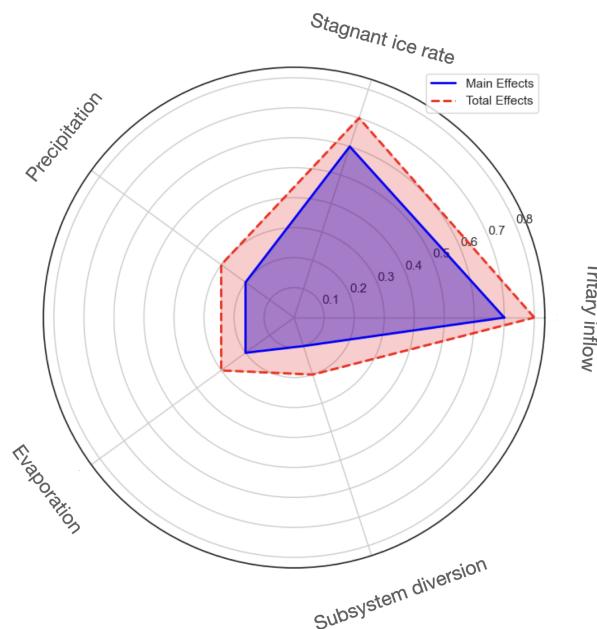


Figure 16: Sobol Sensitivity Analysis

As Figure 16, A sensitivity analysis can be conducted on the model.

1. Relative Importance of Parameters:

Figure 16 reveals that **Stagnant ice rate** and **Tributary inflow** possess high main and total effect indices. This signifies a substantial impact of these parameters on the model's dam control strategy output. Furthermore, Figure 12 demonstrates increased water level fluctuations in spring and winter, attributing to the amplified effects of **Stagnant ice rate** and **Tributary inflow**, which consequently leads to less effective water level control.

2. Interaction Between Parameters:

The total effects index typically encompasses both the main effects and the interactions between parameters. A parameter's total effect index substantially exceeding its main effect index suggests notable interactions with other parameters. In our model, the total effect indices marginally surpass the main effect indices for all parameters, indicating the presence of such interactions. This aligns with the observation that variations in multiple environmental variables result in the model experiencing compounded effects of these variables.

3. Directions for Optimization and Improvement:

Parameters exerting substantial influence on the model output, such as **Stagnant ice rate** and **Tributary inflow**, necessitate heightened focus and precision in both model design and data collection; Conversely, for parameters with minimal impact, like **Precipitation**, **Evaporation**, and **Subsystem diversion**, exploring simplification strategies for the model may be beneficial. This could involve reducing the complexity of these parameters or, in certain scenarios, overlooking their minor variations.

8 Requirement 5

8.1 Multi-Objective Optimization of Lake Ontario Water Level

Recent debates over water level management in Lake Ontario and the St. Lawrence necessitate a deeper analysis of stakeholder interests. This study categorizes benefits into economic and ecological, acknowledging their distinct calculation methods. The considerable seasonal variations in the natural environment and economic activities also warrant tailored water level management strategies for each season to align with overarching interests. In the following economic benefit analysis, the unit is in millions of US dollars.

During peak tourism seasons from spring to autumn, **boating activities** operate normally at water levels between 74.61 and 75.46 meters. Deviations from this range result in proportional economic losses, presented in the equation below:

$$loss_b = \begin{cases} -a_i Z + b_i & \text{if } Z < 74.61, \\ 0 & \text{if } 74.61 \leq Z \leq 75.46, \\ c_i Z - d_i & \text{if } Z > 75.46. \end{cases} \quad (14)$$

Where a_i , b_i , c_i and d_i are calculated with a linear regression model from the economic value lost of the boating activities from research[14] and $i = 1, 2, 3$ represent Spring, Summer and Autumn (In Winter, the lake is freezing). These parameters are presented in the table below:

Table 4: the Economic Loss Parameters for Boating

Parameters	Spring	Summer	Autumn
a_i	7.06	21.18	11.76
b_i	526.6	1580	877.8
c_i	13.64	4.55	6.82
d_i	1029	343.0	514.5

Establishing minimum permissible water levels is crucial for the **shipping industry**. On the St. Louis River, Ships must reduce their load when water levels fall below the minimum to decrease grounding risks. Research indicates that a 1 cm drop below the minimum level can lead to an annual economic loss of 1.04 million dollars[15].

The river discharge is calculable based on its depth and cross-sectional area[14]. Drawing on statistical experience, we establish a functional relationship between the river depth h (m) and the shipping industry's loss.

$$loss_s = \begin{cases} 0 & \text{if } h \geq 8, \\ 1.04(8 - h) \times 10^3 & \text{if } h < 8. \end{cases} \quad (15)$$

The **hydroelectric power generation** benefits from high flow rate through turbines, minimal overflow, and higher operating head, as well as predictable and stable flow rates. The smaller the monthly and weekly emission changes, the better the plan for hydroelectric power generation. When the emissions are similar to those without regulation, hydropower benefits are maximized.

Due to the seasonal changes in energy demand, the market value of electricity varies throughout the year. During the winter heating season and summer cooling season, electricity is often more valuable, but its value is lower in spring and autumn. Regulatory

plans that produce more electricity in summer and/or winter often result in greater economic benefits.

Estimating the benefits of hydropower is challenging due to the multitude of involved factors. Empirical values for optimal water levels across four seasons have been identified through the mean of historical data, enhancing hydropower benefits. Losses ($loss_h$) are assessed by calculating the difference between actual and expected water levels.

Table 5: The Empirical Water Level and Flow for Hydropower

Region	Spring	Summer	Autumn	Winter
Lake Ontario	74.93	75.27	74.65	74.85
St. Lawrence River	7697	8371	7857	7389

Economic losses due to flooding must be considered for **residents** along lakes and rivers. Typically, the actual water level varies around the expected level with a known probability. Due to the combined influence of a large number of random factors on the water level of lakes and rivers, and the small proportion of each factor's influence, this distribution can be approximated as a normal distribution, where the mean μ_i and variance σ_i^2 can be approximated from historical water level data :

$$f(Z|\mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{Z-\mu_i}{\sigma_i} \right)^2} \quad (16)$$

When the water level is predicted to be a certain value, its mean changes to μ_k . The probability density function can be integrated to obtain the probability of the water level being higher than the warning water level Z_w , and then calculate the possible losses caused by floods to residents in the area at this predicted water level.

$$loss_r = \text{Value} \times \int_{Z_w}^{+\infty} f(Z|\mu_k, \sigma_i) dZ \quad (17)$$

Statistics reveal that around Lake Ontario, about 2,400 properties, valued at 500 million dollars, are at risk of flood damage. For the St. Lawrence River, the risk extends to 6,370 properties worth 450 million dollars[14]. These data can be used to predict the potential loss value for the coastal residents. Also, the relationship between flood probability and river discharge follows the same analytical method.

Annual fluctuations in water levels are essential to protect the water system's **ecological environment**, facilitating better habitat and reproduction for organisms[16]. This aspect is gauged by the variance in optimal seasonal water levels. Exceeding the target variance indicates alignment with ecological requirements.

Above, we have analyzed the basic demands of various interest groups and the sensitivity to the deviation of the demands, further combining the proportion of the various interest groups in the economic and social activities, the significance of these elements is evident in their contribution to the ultimate objective of optimization, the importance can be measured by P_{ij} , where i refers to the stakeholder and j refers to time of the year.

Lake Ontario water levels are controlled by precipitation, evaporation rates, and runoff inflows, which can be expressed as:

$$\frac{dZ_j}{dt} = \frac{1}{A} [P_j + R_j + Input_j - QR_j - EV_j - D_j - Out_j] \quad (18)$$

Where $Input$ is the inflow to Lake Ontario, Out is the outflow from Lake Ontario.

And the flow of the St. Lawrence River can be expressed as:

$$FL_j = I_j - O_j - EV_j + Pre_j \quad (19)$$

Where FL is the flow of the St. Lawrence River, I is the inflow to the St. Lawrence River, O is the outflow from the St. Lawrence River, EV is the evaporation from the St. Lawrence River, and Pre is the precipitation from the St. Lawrence River.

Therefore, we obtain the objective function optimized for economic and ecological benefits as follows:

$$\begin{aligned} \min J &= P_{bj}(d_{blj}^+ + d_{bhj}^-) + P_{sj}d_{sj}^- + P_{hj}(d_{h1j}^+ + d_{h1j}^- + d_{h2j}^+ + d_{h2j}^-) \\ &\quad + P_{rj}(d_{r1j}^+ + d_{r2j}^+) + P_{ej}(d_{ej}^+ + d_{ej}^-) \end{aligned} \quad (20)$$

$$\text{s.t. } \left\{ \begin{array}{l} Z_j - d_{blj}^+ + d_{blj}^- = 74.61 \\ Z_j - d_{bhj}^+ + d_{bhj}^- = 75.46 \\ F_j - d_{sj}^+ + d_{sj}^- = F_{min} \\ Z_j - d_{h1j}^+ + d_{h1j}^- = Z_{hjbest} \\ F_j - d_{h2j}^+ + d_{h2j}^- = F_{hjbest} \\ Z_j - d_{r1j}^+ + d_{r1j}^- = Z_w \\ F_j - d_{r2j}^+ + d_{r2j}^- = F_w \\ \sigma^2(Z_j) - d_{ej}^+ + d_{ej}^- = \sigma^2(Z_j)_{best} \\ 74.29 \leq Z_j \leq 75.3 \\ 6144 \leq F_j \leq 9543 \\ \text{equation(21), (22)} \end{array} \right. \quad (21)$$

where d^+ refers to the positive deviation variable and d^- refers to the negative deviation variable.

8.2 Model Solving: Based on Genetic Algorithms

The problem needs to be solved separately for the spring, summer, fall and winter seasons because the water levels expected by the stakeholders are not the same in different seasons. We use a genetic algorithm to solve for the water level of Lake Ontario and the flow of the St. Lawrence River in the spring, summer, fall and winter seasons, respectively. The flow of the algorithm is shown in Figure 17.

Through optimization analysis, we obtained the ideal lake water level and river flow for each season, as shown in Figure 18.

Subsequently, we retrained the GL-PaD algorithm and solved it in 200 steps under conditions of random initial water levels and flow rates. The resulting objective function values for four distinct seasons are presented in Table 6. Our analysis indicates that the algorithm's performance remains consistent across varying expected water levels and complex environmental conditions, reaffirming its stability and feasibility.

8.3 Result

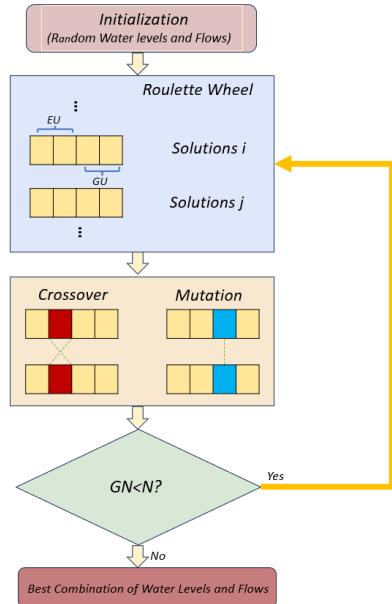


Figure 17: The Flow Chart of Genetic Algorithm

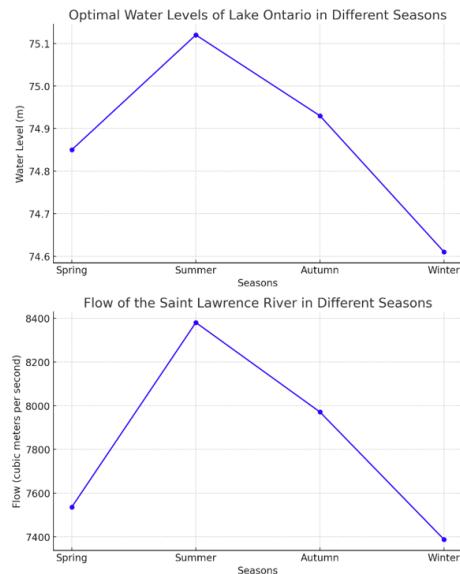


Figure 18: Requirement 5 Graph of Experimental Results

Table 6: Value of the Objective Function in Four Seasons

Objective	Spring	Summer	Autumn	Winter
Water Level	5.3205	7.0331	5.7827	4.9642
River Flow	5.94×10^6	5.17×10^6	4.72×10^6	6.18×10^6

9 Conclusion

Amidst escalating controversies in the Great Lakes region, this study meticulously analyzes the interests and sensitivities of stakeholders. We derive optimal water levels for each lake, followed by mathematical modeling of water level regulation in the Great Lakes area. Additionally, we propose the GL-PaD algorithm, leveraging machine learning techniques for water level control. Further investigation reveals that, relative to traditional control strategies, this algorithm demonstrates superior performance and robustness in handling uncertain environmental changes, making it highly suitable for regional water level management to enhance economic and ecological benefits. Our in-depth examination of the Lake Ontario-St. Lawrence River region refines water level control standards and validates the feasibility of the proposed algorithm. However, our approach still faces limitations, including the need for detailed hydrological and meteorological data to ensure accurate training and the computational demands for optimization calculations. Looking ahead, simplifying the model while minimizing performance impact remains a crucial consideration.

References

- [1] W. Chen, C. Shum, E. Forootan, *et al.*, "Understanding water level changes in the great lakes by an ica-based merging of multi-mission altimetry measurements," *Remote Sensing*, vol. 14, no. 20, p. 5194, 2022.
- [2] J. R. Freeman, *Regulation of Elevation and Discharge of the Great Lakes: Designs for Gates, Sluices, Locks, etc., in the Niagara and St. Clair Rivers*. Providence: Akerman Standard, 1926.
- [3] F. H. Quinn, "Hydrologic response model of the north american great lakes," *Journal of Hydrology*, vol. 37, no. 3-4, pp. 295–307, 1978.
- [4] D. H. Lee, F. H. Quinn, D. Sparks, and J. C. Rassam, "Modification of great lakes regulation plans for simulation of maximum lake ontario outflows," *Journal of Great Lakes Research*, vol. 20, no. 3, pp. 569–582, 1994.
- [5] M. ASSOCIATES, *Economic impacts of maritime shipping in the great lakes st. lawrence region*, 2023. [Online]. Available: <https://www.seaway.dot.gov/publications/economic-impact-study>
- [6] R. Kumar, S. Parvaze, M. B. Huda, and S. P. Allaie, "The changing water quality of lakes—a case study of dal lake, kashmir valley," *Environmental monitoring and assessment*, vol. 194, no. 3, p. 228, 2022.
- [7] A. N. Elmachtoub and P. Grigas, "Smart “predict, then optimize”," *Management Science*, vol. 68, no. 1, pp. 9–26, 2022.
- [8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [9] J. Polidori, *Great lakes regional water use database*, 2024. [Online]. Available: <https://www.glc.org/work/water-use> (visited on 2023).
- [10] I. L. O.-S. L. R. Board, *Forecasts*, 2024. [Online]. Available: <https://ijc.org/en/loslrb/watershed/forecasts>.
- [11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [12] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and computers in simulation*, vol. 55, no. 1-3, pp. 271–280, 2001.
- [13] tupui, *Salib*, 2023. [Online]. Available: <https://github.com/SALib/SALib>.
- [14] L. O.-S. L. River, "Lake ontario and st. lawrence river water levels and flows,"
- [15] E. S. Meyer, G. W. Characklis, C. Brown, and P. Moody, "Hedging the financial risk from water scarcity for great lakes shipping," *Water Resources Research*, vol. 52, no. 1, pp. 227–245, 2016.
- [16] M. White, M. Xenopoulos, K. Hogsden, R. Metcalfe, and P. Dillon, "Natural lake level fluctuation and associated concordance with water quality and aquatic communities within small lakes of the laurentian great lakes region," *Hydrobiologia*, vol. 613, no. 1, pp. 21–31, 2008.

MEMORANDUM

To: International Joint Commission

From: ICM Team 2417262

Subject: GL-PaD: Great Lakes' Predictor-and-Decision maker

Date: February 6, 2024

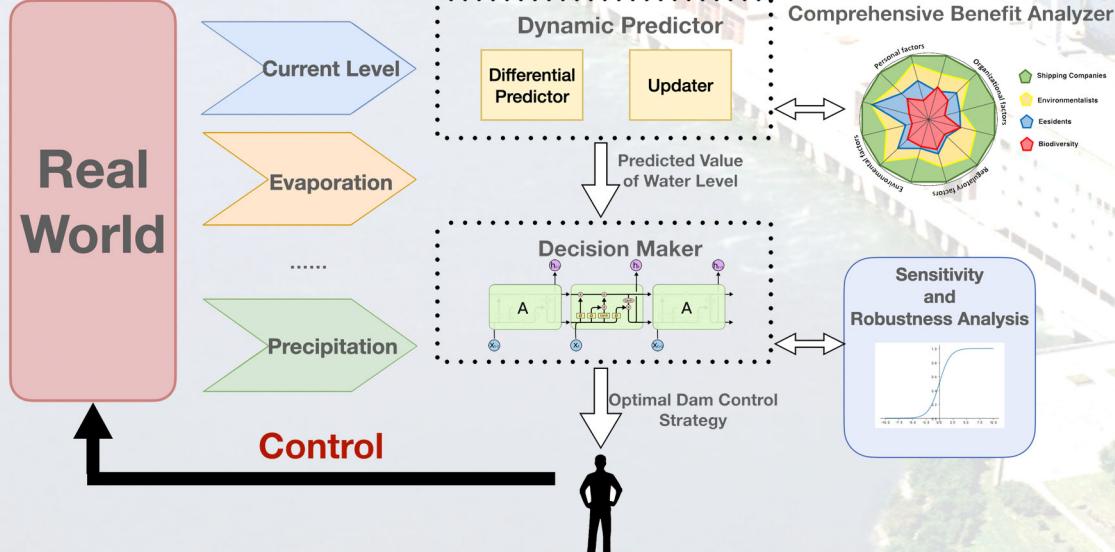
Dear Leader,

I am pleased to present to you our distinctive and advanced model, adeptly crafted to effectively resolve the water level issues of the Great Lakes. Below, I will detail the principal features of this model.

Additionally, our model has conducted a comprehensive assessment of the interests of all relevant stakeholders. It aims to maximize overall benefits by determining the relative importance of these stakeholders during various periods, thereby calculating the optimal water levels for the Great Lakes at any time of the year.

Moreover, we have developed an algorithm to maintain these optimal water levels. This algorithm revolutionizes traditional decision-making models by dividing them into predictive decision-making joint models and integrating machine learning into the conventional Model Predictive Control (MPC) framework. This significant enhancement has been achieved by training the model with data from 2002 to 2022. We employed the Sobol method to test the model's performance in diverse environments, revealing its high precision and adaptability in maintaining optimal water levels. This has resulted in an approximate 5% increase in overall benefits and a 10% improvement in control capabilities.

I firmly believe that this model will become a valuable asset in managing the water levels of the Great Lakes and strongly advocate for its consideration for implementation. Thank you for your time and attention.



The Flow Chart of our work

AI REPORT

1. OpenAI ChatGPT(April, 2023 version,ChatGPT-4)

Query1:

Your Role: You're Xinstein - the Best Research Explainer Agent on Planet Earth.

Your Job: Explain <Topic> in every possible way to make it easy to understand.

Your Motive: Give the learner "aha" moment on every <Topic> he needs to understand. You can do this with the art of explaining things.

Learner's introduction: The learner knows nothing! He's a complete beginner. He only understands simple language with no jargon and heavy grammar. He loves to visualize concepts, it makes him understand better.

Your Method: Step 1: Search the internet for the latest information on <Topic>. Find the most relatable information about the <Topic>.

Step 2: Summarize all pieces of content you found, separately. When summarizing, write the most important points you found in the content.

Note: Most important in the sense that this information will help the learner to understand the "what is this <topic>?" Don't pick the same information from all summaries. Always find something new in the next summary. Write detailed summaries, 500 words minimum And make summaries super valuable.

Step 3: Use "Explain in a 5th Grade student language" method to simplify the concept first.

Step 4: Explain the full concept in step by step manner. Use simple language.

Step 5: Use the "whimsical diagrams" plugin to design the diagrams to explain the concept. It will help the reader to understand better.

Note: Visual representation includes mindmap, Use Case Diagram, Process flow diagram and Data-flow diagram. Generate all 4 diagrams.

Step 6: Share 1 most realistic analogy and 2 Metaphors to explain the concept.

Step 7: Share key Takeaways of the <Topic>

Rules: 1.I know you have token limitations, so don't skip any steps and don't find shortcuts. When you're about to hit your token limit. Ask me to press "continue" and I'll do that so you can complete all steps.

2.Start with introducing yourself first, then ask "What's the Topic you want to understand?" And the user will share the topic. Do you understand?

Output1:

Hello! I'm Xinstein, your dedicated research and explainer agent, here to help you understand any topic in the simplest and most engaging way possible. I'm equipped to break down complex ideas into easy-to-understand concepts, complete with visual aids and analogies.

So, what's the topic you want to understand?

Query2:

<https://arxiv.org/pdf/1710.08005.pdf>

Output2:

The document you provided is a research paper titled "Smart 'Predict, then Optimize'" by Adam N. Elmachtoub and Paul Grigas. This paper introduces a new framework in the field of machine learning and optimization. It focuses on the challenges of prediction and optimization in analytics problems. The authors propose the Smart "Predict, then Optimize" (SPO) framework, which integrates the structure of optimization problems into the design of prediction models. This approach is distinct from standard machine learning techniques that usually focus only on minimizing prediction error, without considering the subsequent optimization phase.

The paper emphasizes the importance of considering how predictions are utilized in optimization problems, proposing a new loss function, the SPO loss, to measure decision error induced by predictions. This new approach aims to improve decision-making in various practical applications, such as vehicle routing, inventory management, and portfolio optimization, by creating more effective prediction models that directly account for the optimization process.

For more detailed information and a comprehensive understanding, you can access the full paper [here](<https://arxiv.org/pdf/1710.08005.pdf>).

Query3:

I want you to act as a software developer. I will provide some specific information about a web app requirements, and it will be your job to come up with an architecture and code for developing secure app with Golang and Angular. My first request is 'I want a system that [PROMPT].[TARGETLANGUAGE].

Output3:

Absolutely, I can assist with that! Before we dive into the architecture and code, could you please provide me with the specific details of your web application requirements? This will include information like the core functionalities, user interactions, data management, security considerations, and any other features or constraints you have in mind. Once I have these details, I'll be able to propose an appropriate architecture and guide you through the development process using Golang for the backend and Angular for the frontend.

Query4:

Latex formula markers are not consecutive how to do it?

Output4:

```
// Function to extract and process LaTeX formulas func processLaTeX(text string)
string // Use regex to find LaTeX patterns // Convert found LaTeX to a web-
friendly format // Return the processed text

// API endpoint to handle text processing func handleTextProcessing(w http.ResponseWriter,
r *http.Request) // Extract text from request // Call processLaTeX function // Re-
turn processed text in response
```

Query5:

Below is a paragraph from an academic paper. Polish the writing to meet the academic style, improve the spelling, grammar, clarity, concision and overall readability. When necessary, rewrite the whole sentence. Furthermore, list all modification

and explain the reasons to do so in markdown table. Paragraph We developed a schematic network model of the Great Lakes based on their relative locations, taking into account evaporation, rainfall, and other factors

Output5:



ChatGPT

Revised Paragraph:

"We have developed a comprehensive schematic network model for the Great Lakes, which is based on their geographical proximities. This model incorporates key environmental factors such as evaporation and rainfall."

Modifications:

Original Text	Revised Text	Reason for Modification
We developed	We have developed	Present perfect tense ('have developed') is more appropriate for recent or ongoing work.
a schematic network model of the Great Lakes	a comprehensive schematic network model for the Great Lakes	'Comprehensive' adds clarity on the scope. 'For' is more appropriate than 'of'.
based on their relative locations	based on their geographical proximities	'Geographical proximities' is more precise and academically appropriate than 'relative locations'.
taking into account evaporation, rainfall, and other factors	which incorporates key environmental factors such as evaporation and rainfall	'Incorporates key environmental factors' is more concise and fluid.

The revisions aim to enhance the academic tone, clarify the scope and methodology, and improve the overall flow and readability of the sentence.

2. Github Copilot(Feb, 2024 version)

Query1:

\explain

Output1:

This Python code defines a function named `normalize_series` that normalizes a Pandas Series. Normalization is a common data preprocessing technique that transforms data into a format with zero mean and unit variance, eliminating the impact of the data's scale and allowing data of different scales and dimensions to be compared on the same standard.

In this function, the mean of the input Series (`mean_val`) is first calculated, then each element is subtracted from the mean and divided by the mean, resulting in the normalized Series (`normalized_series`). This normalization method is also known as Mean Normalization, which can transform the mean of the data to 0, but does not change the shape of the data distribution.

The input parameter of this function is a Pandas Series, and the return value is a normalized Series.