Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Modeling and Analysis of Lake Water Levels Using System Dynamics and Principal Component Analysis

Jinsheng Liu [1, #], Tong Jing [1, #, *], Xuejiayi Xu [1, #], Guohao Song [1, #], Xiaotong Li [2, #]

[1] College of Computer Engineering, Qingdao City University, Qingdao, China, 266106

[2] School of Mechanical and Electrical Engineering, Qingdao City University, Qingdao, China, 266106

* Corresponding author: 13655338595@163.com

[#]These authors contributed equally.

Abstract. Understanding the intricate dynamics of the Great Lakes water levels, shaped by environmental factors like rainfall and evaporation, is crucial for devising sustainable water management strategies. This study develops a model to manage the Great Lakes' water levels by analyzing the interconnectedness of the lake network. Utilizing principal component analysis, key variables like rainfall and evaporation were evaluated for their impact on water levels. System dynamics models integrated with flow conservation laws allowed for a detailed hydrodynamic assessment. The study introduces a "narrow seam method" for dry river scenarios, enhancing model accuracy. Annual water level trends from 2000 are depicted, offering insights into future water management practices in response to environmental changes.

Keywords: Principal component analysis; Hydrodynamic assessment; Great Lakes; System Dynamics; water level.

1. Introduction

Water levels in the Great Lakes have a profound impact on regional ecology, economy, and community well-being [1]. However, understanding and predicting their fluctuations pose a significant challenge due to the complex interplay of environmental factors such as rainfall and evaporation [2, 3]. Traditional models often fall short of capturing the nuanced dynamics of this vast lake network [4].

Our methodology begins with a detailed examination of the lake network's interconnectedness, followed by the application of PCA to identify key variables influencing water levels. System dynamics models are then employed, incorporating flow conservation laws to provide a robust hydrodynamic assessment. A unique aspect of our study is the introduction of the "narrow seam method" for modeling dry river scenarios, which enhances the accuracy of our models in predicting water levels under various conditions. The results of this study, particularly the annual water level trends from 2000 onwards, offer valuable insights for policymakers and stakeholders involved in water management. By providing a clearer understanding of the factors driving water level changes in the Great Lakes, our research lays the groundwork for developing adaptive management strategies that can mitigate the impacts of environmental variability and ensure the long-term sustainability of this critical freshwater resource.

2. Data Processing and Framework

To better understand the dynamics of lake networks, this study has developed a lake network model that elucidates the interrelationships among lakes and the mechanisms through which water level changes propagate. This model is capable of predicting how alterations in the water level of one lake may impact other lakes, thereby offering a more holistic view of the dynamics at play across the entire lake system. These models were crafted to deliver comprehensive lake water level management solutions that cater to the requirements of all stakeholders involved.

The data for this study was obtained from https://www.comap.com/contests/mcm-icm. Upon loading the 'Lake Superior' data table, we confirmed that the first row correctly labels the columns with the year and each month. The data collation process comprised setting these column headers, converting years to integers and monthly water levels to floating-point numbers, checking and addressing missing values, and finally reformatting the data into a data format appropriate for time series analysis. Through these measures, the 'Lake Superior' dataset was effectively prepared for accurate and comprehensive analysis.

3. Decoding Hydrological Variability with PCA

Principal component analysis (PCA) streamlines the complexity of multivariate problems by transforming correlated variables into a smaller number of uncorrelated factors, called principal components [5, 6]. This technique not only simplifies data with high dimensionality but also preserves its essential variability, facilitating the identification of the most influential variables [7]. This study examines the hydrological variability of Lake Superior as a case example.

The preliminary steps in PCA involve data standardization, calculation of the covariance matrix, and selection of principal components. Data standardization here opts for the z-score method [8]. Principal components are the eigenvectors of the covariance matrix, ordered by the magnitude of their eigenvalues [9]. A larger eigenvalue indicates that the principal component accounts for a greater amount of variability in the data. Typically, the first few principal components are selected to cumulatively explain most of the variability present in the data. Afterward, through data conversion, the data is transformed into the principal component space. In this study, variables that may affect hydrological changes, such as rainfall, evaporation, water level of surrounding bodies, land use type, soil type, etc., were selected. Here, the R language is used to implement the code, with specific results presented in Table 1.

Table 1. Results of principal component analysis

Influencing factor	The first principal component	The second principal component
rainfall	0.904	0.700
evaporation capacity	0.700	-0.700
standard deviation	1.047	0.451
Contribution rate of variance	0.548	0.451
The cumulative contribution rate of the variance	0.548	1

This means that, by analyzing the monthly water level data in the "Lake Superior" table, we find that the main variability of the data can be represented by a few principal components. The first principal component dominates, explaining most of the variability, which may indicate the existence of a strong, overall common factor affecting the water levels in all months, such as overall annual precipitation or evaporation rate. The second principal component provides additional information that may reflect secondary trends or cyclical changes in addition to the dominant factors. Such analyses are very useful for understanding the main drivers of groundwater table variation, especially when preparing and evaluating water management strategies. By reducing the dimensionality of the dataset, PCA helped us identify the most important factors affecting water level.

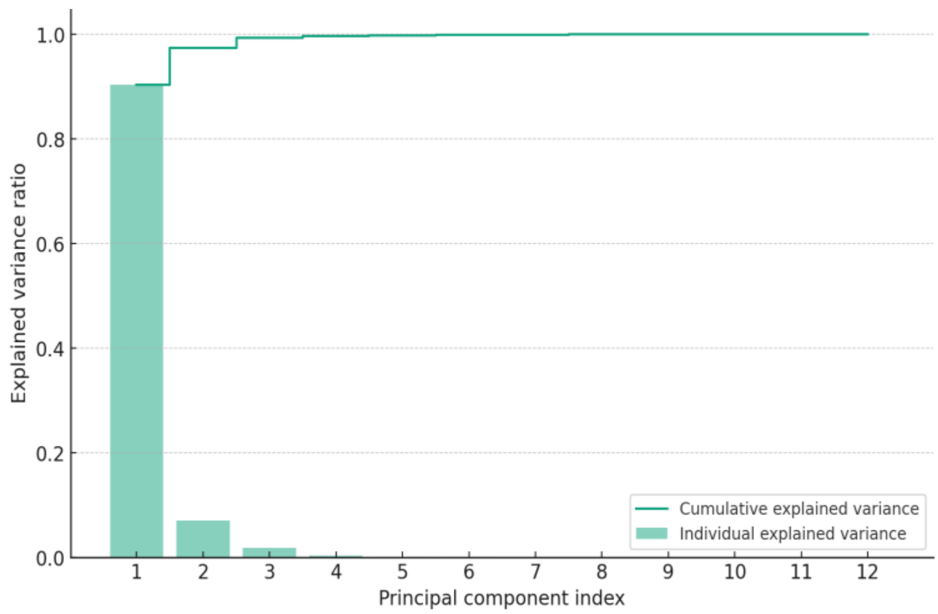

Figure 1. PCA score plot

The result of the principal component analysis shown in Figure 1, according to the first and second principal components, the weight of rainfall and evaporation influencing factors is expressed as:

$$F_1 = 0.904X_1 + 0.700X_2 \tag{1}$$

$$F_2 = 0.700X_1 - 0.700X_2 \tag{2}$$

According to the formula and Table 1, the contribution of the variance of the first principal component is the highest, reaching 54.8%. It contains information about rainfall and evaporation. The weights of these two sets of information are 0.904 and 0.700, respectively, both of which are positively correlated with the first principal component. The second principal component variance contribution rate of 45.1%, contains information on rainfall and evaporation, its weight is 0.700 and-0.700 respectively, the rainfall sequence and the second principal component of the negative correlation, evaporation, and rainfall, the second principal component sequence information and moisture loss and similar, water gain sequence by 0.700 on the weight of the second principal component sequence. Theoretically, the second principal component contains information in line with the relationship between rainfall and evaporation, but the variance contribution of the information contained by the first principal component is large. The timing diagram of output rainfall, evaporation, water surplus, and the two principal components were further analyzed.

As shown in Figure 2, sequences similar to the water profit and loss sequence are obtained through principal component analysis, which can be used to obtain effective information in multifactor analysis and simplify the complexity of multivariate sequence analysis. The variance contribution rate of the first principal component is large, so the relationship between the first principal component and the second principal component can be used to verify the effect of the principal component analysis statistics.

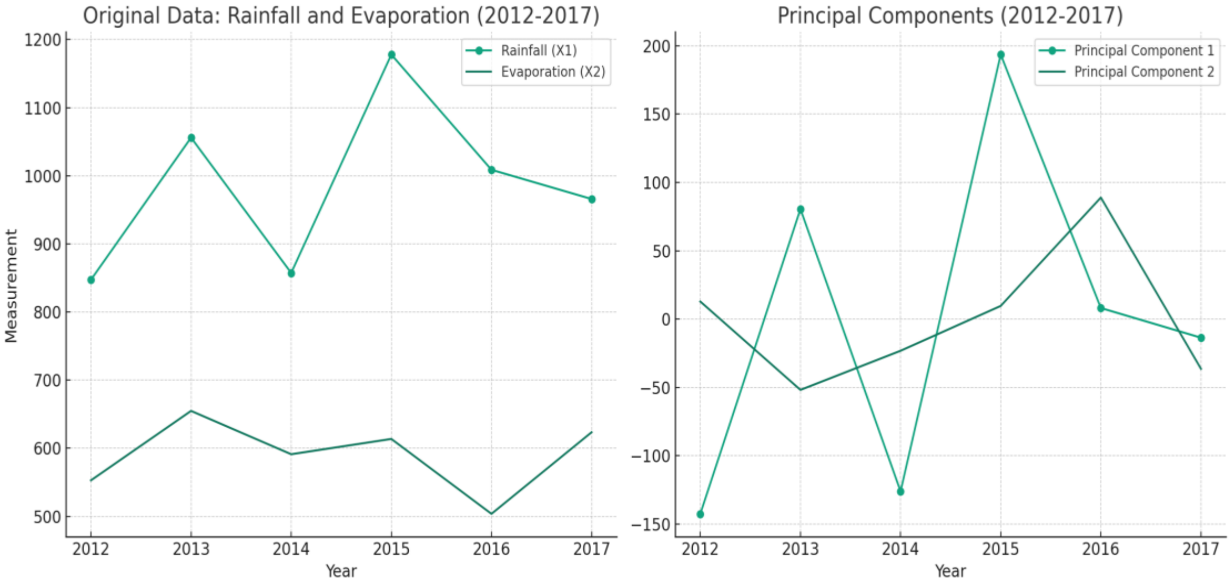

Figure 2. Principal component analysis and time comparison of rainfall, evaporation, and water profit and loss

4. Assessing Hydrological Variability in the Great Lakes

In the Great Lakes system map provided, we can see the geographical layout of the Great Lakes, including information about the connecting rivers, dams, and the height of the water level. Combining the previous discussions, we can conduct SLT (water level-flow-time) analysis to solve this problem, namely, how to manage and model the water level of the Great Lakes [10].

Firstly, this study used the water quantity balance equation to describe the change in water quantity over time:

$$[\Delta S = (Q_{in} - Q_{out} + P - E) \times \Delta t] \tag{3}$$

$\Delta S,\ Q_{in,}\ Q_{out}$Among them, it is the change of water quantity, the incoming flow rate, the outflow flow rate, P is the precipitation, the evaporation quantity, and the time interval were $E\Delta t$.

Hydrodynamic models, incorporating the dynamic interplay between lake water levels and river flows, were created for each body of water. These models are interconnected using flow conservation and continuity equations to provide a unified representation of the entire system's groundwater flow, which is fundamentally described by the mass conservation equation.

$$[\frac{\partial z}{\partial t} = \frac{1}{B}\frac{\partial Q}{\partial x} - q] \tag{4}$$

In the equation, the variables from left to right represent the rate of change of the hydraulic head over time, the aquifer's storage coefficient, the rate of flow change per unit width in the direction of flow, and the rate of infiltration or pumping per unit area.

The dynamic equation describes the driving force of the groundwater flow, which is related to the properties of the flow medium and the head gradient. The mathematical expression is:

$$[\frac{\partial Q}{\partial t} = \frac{a}{\partial A}(\frac{\partial^2 z}{\partial A}) + \frac{a \partial z}{K \partial x} - R[\frac{\partial Q}{\partial x}] = 0] \tag{5}$$

The elements $\partial QaAzR$ are the flow rate per unit width, the cross-sectional area of the aquifer, the flow break area, the permeability coefficient, and the resistance coefficient, usually related to the roughness of the material. A single-channel dynamics equation is solved by Preissmann's four-point eccentric format 37. The M point in Figure 3 is located in the middle of the spatial step △x, favors the unknown moment n + 1 in the time step, the known moment n is △t, n + 1 is (1- θ) At, where θ is the weight coefficient (O θ 1). The nodes along the spatial direction are j, j + 1

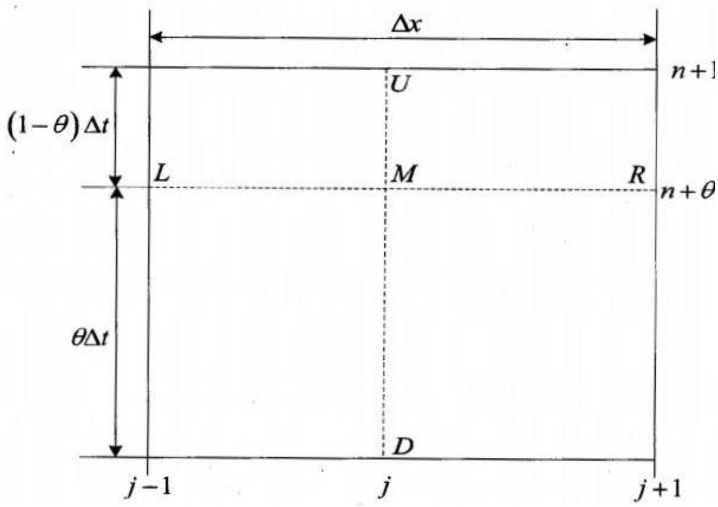

Figure 3. Four points of eccentricity diagram

The Han Point contains two forms: unregulated Han Point and regulated Han Point. If the storage capacity of the Han junction area can be omitted in the calculation process of the reach, it is no regulating branch point, otherwise it is the regulation Han point. According to the principle of mass conservation, the increase and decrease of the water storage volume in a Han point should be balanced with the flow rate in and out of this branch point.

The equation for the adjustment point is:

$$\sum_{i=1}^{n} Q_i = 0 \tag{6}$$

The equation for the regulated branch point is:

$$\sum_{i=1}^{n} Q_i = \frac{dW_k}{dt} \tag{7}$$

Where: k-point number, n-total number of single channel into (outflow) Han point, W. Water storage capacity of Han point, Q-Flow rate of the I channel into Han point (generally assume that inflow is positive and outflow is negative).

In some rivers in the north, the flow of rivers is generally small, especially in the dry season will often appear rivers dry up phenomenon. Therefore, the cross section needs to be treated somewhat, so that the model can run when the river is cut off. In this paper, the "narrow seam method" is used when the river dries up.

The "narrow slit method" was first used in the numerical simulation of water waves on the bank, and rarely used the dynamic boundary problems for the drying up of rivers. In this paper, the "narrow slit method" for the numerical simulation of water waves is used for the hydrodynamic simulation of the flow interruption of rivers in arid areas.

As shown in Figure 4, this method is to assume that there is a very narrow gap, a narrow seam of water, and the water in front of the shore, this is equivalent to the beachfront water extends to the shore, so the beach in the narrow seam of the boundary section, make its formed with a certain depth of fixed boundary for the one-dimensional river hydrodynamic water quality model, this section from a single channel and river network is verified in detail. The single channel mainly uses three examples constant flow, non-constant flow, and narrow joint model. The river network is mainly analyzed with four examples: triangular bifurcation river network, dendrite symmetrical river network, ring river network, and trigeminal river network.

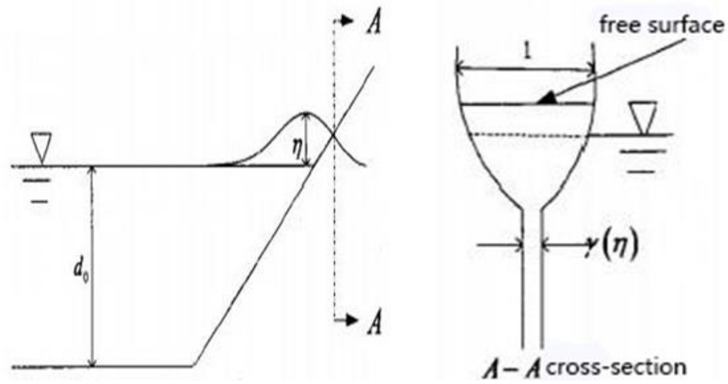

Figure 4. Schematic diagram of narrow seam

The annual average water levels for Lake Erie, along with those of Lake Superior and Lake Michigan-Huron, have been computed using the model analysis described above. This data will be utilized to create a trend map that charts the changes in water levels from 2000 to the present, aiding in the identification of long-term trends across these Great Lakes.

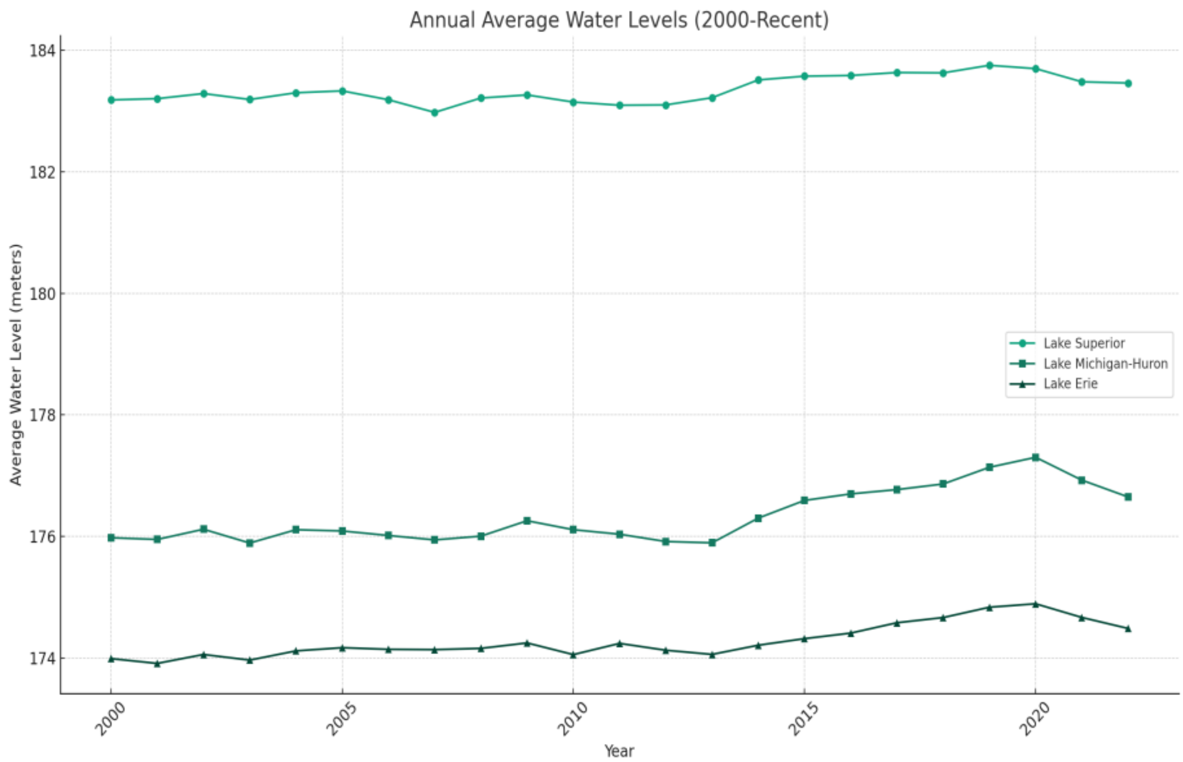

Figure 5. The average water level change

Figure 5 presents trends in annual mean water levels for Lake Superior, Lake Michigan-Huron, and Lake Erie from 2000 to the most recently available data years. The annual mean water level of each lake is represented in different markers and lines to facilitate the identification and comparison of variation between them. From this chart, we can observe different lake water levels over time, identifying possible trends, periodic fluctuations, or significant changes in a particular year.

5. Conclusion

Analyzing the Great Lakes drainage, seasonal decomposition of rainfall, evaporation, and water surplus data, along with the first principal component, shows a consistent upward trend indicating strong seasonal patterns likely linked to climate cycles. Conversely, moisture balance reveals a declining trend, suggesting that groundwater levels are significantly influenced by seasonal variations and short-term meteorological shifts. Although the first principal component reflects rainfall and evaporation, its direct impact on groundwater levels isn't clear-cut, highlighting the need for a tailored approach in applying PCA to the hydrologic system. These insights are crucial for devising effective water management strategies that accommodate both long-term trends and seasonal fluctuations due to climate and hydrological changes.

References

[1] ANGRADI T R, WILLIAMS K C, HOFFMAN J C, et al. Goals, beneficiaries, and indicators of waterfront revitalization in Great Lakes Areas of Concern and coastal communities [J]. Journal of Great Lakes Research, 2019, 45 (5): 851 - 863.

[2] BAXTER B K, BUTLER J K. Climate change and great Salt Lake [J]. Great Salt Lake biology: a terminal lake in a time of change, 2020: 23 - 52.

[3] MIANABADI A, DAVARY K, MIANABADI H, et al. Toward the development of a conceptual framework for the complex interaction between environmental changes and rural-urban migration [J]. Frontiers in Water, 2023, 5.

[4] MEMARI S. Combining Remote Sensing, Machine-Learning, and Mechanistic Modeling to Improve Coastal Hydrodynamics and Water Quality Modeling in the Laurentian Great Lakes [D]. Michigan State University, 2023.

[5] KHERIF F, LATYPOVA A. Principal component analysis [M]//Machine learning. Elsevier, 2020: 209 - 225.

[6] KARAMIZADEH S, ABDULLAH S M, MANAF A A, et al. An overview of principal component analysis [J]. Journal of Signal and Information Processing, 2020, 4.

[7] HASAN B M S, ABDULAZEEZ A M. A review of principal component analysis algorithm for dimensionality reduction [J]. Journal of Soft Computing and Data Mining, 2021, 2 (1): 20 - 30.

[8] MORAIS C L, PARASKEVAIDI M, CUI L, et al. Standardization of complex biologically derived spectrochemical datasets [J]. Nature Protocols, 2019, 14 (5): 1546 - 1577.

[9] SALEM N, HUSSEIN S. Data dimensional reduction and principal components analysis [J]. Procedia Computer Science, 2019, 163: 292 - 299.

[10] LONG Y, YANG T, GAO W, et al. Prevention and control of algae residue deposition in long-distance water conveyance project [J]. Environmental Pollution, 2024, 344: 123294.