Study on the correlation between water conservancy facilities operation efficiency and social demand data

Jiayou Xia[1] Liu Yanfang[2*]

1 Hangzhou Dipont School of Arts and Science，Hangzhou, Zhejiang, China，310000；

2 School of Information Engineering, Qingdao Hengxing University of Science and Technology, Qingdao, Shandong, 266041；

Abstract： Against the backdrop of water resource shortages and evolving water use structures, improving the operating efficiency of water conservancy facilities has become an important direction for optimizing resource allocation. This paper constructs a correlation analysis model driven by social demand data to explore the matching relationship between multiple demands such as population, industry, and ecology and the operating efficiency of facilities. The grey correlation and multiple regression methods are used to construct the correlation structure, and the statistical data of Hefei from 2021 to 2023 are used as an empirical basis to reveal the synchronization between industrial water use and operating energy consumption, as well as the lag between ecological demand and water supply results. Visual analysis further verifies the structural evolution trend of the supply and demand relationship, and based on this, dynamic control strategies such as zoning compensation, load forecasting, and recycled water joint regulation are proposed. The research results show that changes in social demand have become an important variable affecting the operating efficiency of water conservancy facilities, and data-aware flexible scheduling mechanisms will play a key role in future water conservancy systems.

Keywords： water conservancy facility efficiency; social demand structure; grey correlation analysis; matching evaluation; dynamic regulation strategy

DOI:10.69979/3041-0843.25.02.002

1 Introduction

With the continuous evolution of the social and economic structure and the acceleration of urbanization, the operation of water conservancy facilities faces more complex challenges in the coordination of supply and demand. On the one hand, population growth, agricultural structural adjustment, and changes in industrial water demand have led to strong temporal and spatial differences in social water demand; on the other hand, the traditional water conservancy facility operation and control model has significant lags and insufficient adaptability in responding to changing social needs. To improve the operating efficiency of water conservancy facilities, it is necessary to start from the macro-social needs, build a data-driven scientific control mechanism, and achieve accurate docking of operating capabilities with real needs.

Wang and Zhao [1] (2023) explored the role of social network relationships in promoting farmers' participation in the supply of small-scale water-saving irrigation facilities, and pointed out that social structural variables have a profound impact on the operation of water conservancy facilities. Silva [2] (2023) systematically reviewed the application of recycled water treatment and reuse technology in sustainable water resources management, emphasizing the importance of demand -side management . Raheli et al. [3] (2020) analyzed the cognitive drivers of Iranian farmers' water-saving behavior based on the health belief model, revealing the impact of social psychological variables on water resource use patterns. Scanlon et al. [4] (2023) analyzed the key role of groundwater resources in the future resilience of water systems from a global perspective, and proposed the need to strengthen the coordinated response capabilities of supply and demand. Molajou et al. [5] (2021) incorporated social system factors into the coupling of water-food-energy systems, demonstrating the applicability and explanatory power of social data in water resource allocation models.

2 Analysis of the correlation mechanism between water conservancy facility efficiency and social needs

2.1 Definition and evaluation index design of water conservancy facility operation efficiency

As shown in Figure 1, this study constructed an efficiency evaluation system based on "goal orientation-structural hierarchy-quantifiable indicators". The first-level indicators include four core dimensions: technical efficiency, management efficiency, resource efficiency, and benefit efficiency[6]. Among them, technical efficiency reflects the functional completeness and technical application level of facility operation; management efficiency measures the level of operation organization and decision-making response capability; resource efficiency reflects the utilization rate and energy consumption of unit resources; and benefit efficiency integrates the two types of output benefits, economic and ecological. On the basis of the first-level indicators, the second-level evaluation indicators are further set to realize the quantification and comprehensive evaluation of multi-dimensional data, providing a basis for subsequent model input and empirical analysis[7].

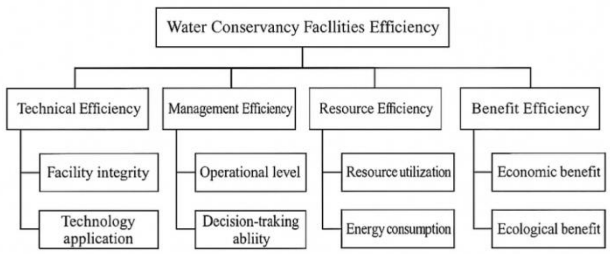

Figure 1. Evaluation system for water conservancy facility operation efficiency

2.2 Classification structure and quantification method of social demand data

Social demand data is essential for evaluating water conservancy facility efficiency, as its accuracy and structure affect model reliability. Originating from population growth, industrial activity, ecological needs, and unconventional water use, it is dynamic and diverse[8]. For effective analysis, demand is grouped into four types: urban-rural, industrial-agricultural, ecological, and unconventional use. Indicators like population, industrial water use, and recycled water utilization are drawn from yearbooks, remote sensing, and surveys, then standardized to ensure consistency and comparability[9]. As shown in Figure 2:

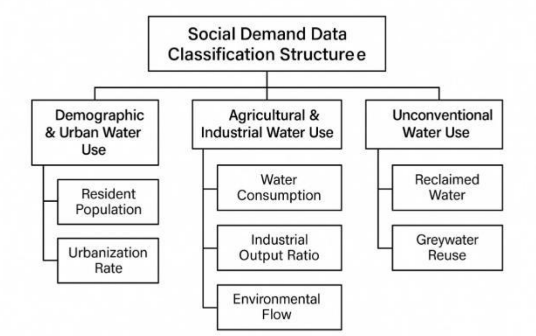

Figure 2. Classification structure of social demand data

3 Construction of Correlation Analysis Model

To reveal the dynamic coupling between water conservancy facility efficiency and social demand, the model follows a four-step process: data, indicator, structure, and analysis. Social demand and operational indicators are first normalized using Min-Max standardization, with outliers and missing values removed to ensure robustness[10]. Based on the established indicator system, input and output matrices are constructed—inputs include population and ecological flow; outputs cover energy consumption and response time. The model integrates grey correlation analysis to identify key relationships and regression modeling to quantify impact strength. A coupling evaluation function is then built to support dynamic, region-specific decision-making.

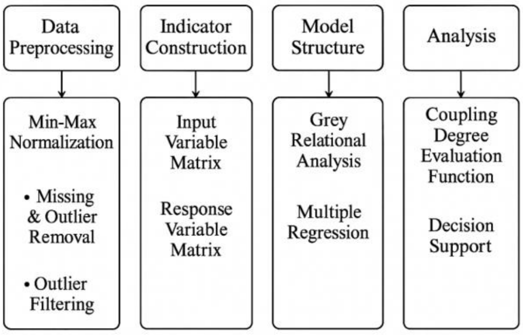

Figure 3. Model building architecture diagram

3.1 Collection and preprocessing of data required for modeling

The data relied on for model construction are mainly divided into two categories: one is data representing social demand, such as permanent population (P), industrial water consumption (W_i), ecological flow lower limit (Q_e), etc.; the other is data describing the operating efficiency of water conservancy facilities, such as unit dispatching energy consumption (E_u), operation response time (T_r), comprehensive water supply satisfaction rate (R_s), etc. The data sources mainly include the National Bureau of Statistics "Statistical Yearbook", the Ministry of Water Resources "National Water Resources Bulletin", and the provincial water system monthly dispatch report. After data collection, the time scale needs to be unified to the annual average level, and missing items and extreme outliers need to be eliminated to ensure data quality.

The variable processing adopts the normalization method to transform the data of different dimensions into the interval [0,1]. The formula is as follows:

$$x_{ij}^{*}=\frac{x_{ij}-\min\left(x_j\right)}{\max\left(x_j\right)-\min\left(x_j\right)} \quad (1)$$

Among them, x_{ij} represents the value of the original i sample under j the index, x_{ij} which is the normalized data. The triple standard deviation method is further introduced to eliminate the anomalies for the violently volatile indicators, and finally the structured data matrix required for modeling is generated $X \in ð^{n\times m}$, providing a data basis for subsequent modeling.

3.2 Establishment of the relationship between the indicator system and the variable mapping

In the modeling stage, it is necessary to systematically map the original multidimensional social demand data and the operating efficiency indicators. This study aims to construct a correlation function $f: X \rightarrow Y$, in which the input variable set $X=\left\{x_1,x_2,\ldots,x_m\right\}$ represents the standardized indicator set of social demand, and the output variable set corresponds to the $Y=\left\{y_1,y_2,\ldots,y_k\right\}$ performance indicators of each dimension of the operating efficiency of water conservancy facilities . The redundancy and linear correlation between indicators are reduced through expert scoring and principal component analysis (PCA), and a mapping system without multicollinear interference is constructed.

The weight coefficient for establishing the indicator mapping w_j is determined e_j by the entropy weight method . First, the information entropy of each variable is calculated :

$$e_j=-k\sum_{i=1}^{n}p_{ij}\ln\left(p_{ij}\right),\quad p_{ij}=\frac{x_{ij}^{*}}{\sum_{i=1}^{n}x_{ij}^{*}},\quad k=\frac{1}{\ln(n)} \quad (2)$$

Then the objective weight of each indicator is derived as $\sum_{j=1}\left(1-e_j\right)$ Finally, the weighted aggregation function of the input variables is constructed to model the input feature set $X = w_1x_1+w_2x_2+\cdots+w_mx_m$.

3.3 Establishing a preliminary demand impact structure model

After the variable system is constructed, it is necessary to clarify the impact path of social demand variables on the operational response of water conservancy facilities. The preliminary structural model adopts a linear regression framework of multiple dependent variables to construct the following functional relationship:

$$Y = \beta_0 + \sum_{j=1}^{m} \beta_j X_j + \varepsilon \quad (3)$$

Among them, $Y \in ð^{n\times k}$ represents the operating efficiency output matrix, X_j represents the j social demand indicator variable, β_j is the estimated coefficient, ε and is the residual term of compliance $N\left(0,\sigma^2\right)$. The model is fitted by the least squares method (OLS), and the objective function is to minimize the sum of squared errors:

$$\min_{\beta} \| Y - X\beta \|^2 \quad (4)$$

In order to avoid the model from falling into overfitting, the normality and heteroscedasticity tests were performed on the fitting residuals, and the partial least squares (PLS) model was introduced as a supplementary comparison. The fitting results provide parameter support for subsequent coupling calculation and dynamic regulation simulation, and can also be used to analyze the response amplitude and directional differences caused by changes in social demand structure at the operating level.

3.3 Establishing a preliminary demand impact structure model

The impact of social demand on the operating efficiency of water conservancy facilities has systematic and structural characteristics. Based on the selection of the above variable set, a preliminary demand impact structural model is constructed. The model mainly adopts a multivariate linear structure, and the basic form is as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ij} + \varepsilon_i \quad (5)$$

Among them, Y_i represents the operating efficiency result corresponding to the i-th sample, X_{ij} represents the i - th sample j social demand index values, β_j are the coefficients to be estimated , ε_i and are the error terms, which satisfy independent and identical distribution $\varepsilon_i \sim N\left(0,\sigma^2\right)$.

In order to test the explanatory power of each variable, partial least squares regression (PLSR) is introduced to reduce the dimension of the variables and model them , which is especially suitable for the problem of multicollinearity of high-dimensional covariates. The core of PLSR is to extract the latent variable TTT and establish the following structure:

$$Y = TQ \ + E, \quad X = TP \ + F \quad (6)$$

by maximizing T the covariance with , which improves the stability of modeling. The preliminary structure of the model is formed at this stage, laying the foundation for correlation evaluation. Y 3.4 Constructing the correlation evaluation model and determining the parameters

On the basis of establishing the structural model, in order to further quantify the coupling intensity and directionality between social demand and water conservancy facility efficiency, a grey correlation analysis (GRA) model is constructed to evaluate the correlation. Assume that the social demand sample data sequence is $X_j = \left\{x_j(1), x_j(2), \ldots, x_j(n)\right\}$ and the facility operation efficiency target sequence is $Y = \{y(1), y(2), \ldots, y(n)\}$. The grey correlation coefficient calculation formula is as follows:

$$\xi_j(k) = \frac{\min_j \min_k \left|x_j(k) - y(k)\right| + \rho \cdot \max_j \max_k \left|x_j(k) - y(k)\right|}{\left|x_j(k) - y(k)\right| + \rho \cdot \max_j \max_k \left|x_j(k) - y(k)\right|} \quad (7)$$

Where $\rho \in [0,1]$ is the resolution coefficient, usually 0.5. The grey relational degree of each variable is:

$$\gamma_j = \frac{1}{n}\sum_{k=1}^{n}\xi_j(k) \tag{8}$$

According to the calculation results, the input variables with strong correlation with the target variable are selected, and the final optimization model is constructed accordingly . The following is a table of some sample parameters and their correlation results :

TABLE I. Grey correlation evaluation table of social demand variables and operating efficiency

Index No.	Indicator name	Series mean	Series standard deviation	Grey correlation coefficient
X1	Permanent population (10,000 people)	185.2	23.4	0.712
X2	Industrial water consumption (billion cubic meters)	12.6	4.1	0.801
X3	Recycled water utilization rate (%)	18.4	5.6	0.678
X4	Urbanization rate (%)	59.7	2.3	0.745
X5	Ecological flow compliance rate (%)	91.2	3.8	0.769

This model not only improves the scientific nature of variable selection, but also provides a quantitative basis for the parameter setting of subsequent coupling functions.

3.5 Model training, verification and evaluation

After the correlation analysis model is built and the parameters are set, the model needs to be trained and verified to test its fitting performance and generalization ability. First, the structured sample data is divided into a training set and a verification set in a ratio of 7:3. The training set is used to fit the model parameters, and the verification set is used to evaluate the predictive ability and robustness of the model. The training phase estimates the regression coefficients based on the least squares method (OLS) or partial least squares method (PLSR), and the mathematical expression is as follows:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}\ \mathbf{X}\right)^{-1}\mathbf{X}\ \mathbf{Y} \tag{9}$$

Among them, $\mathbf{X}$ is the training input variable matrix, $\mathbf{Y}$ is the response variable matrix, $\boldsymbol{\beta}$ and is the estimated regression coefficient vector.

4 Empirical Research: Analysis of Demand-Efficiency Data Based on a Certain Region

4.1 Research area background and data collection instructions

Hefei City, a key urban center in the Yangtze River Delta with rapid economic and population growth, was selected to verify the correlation model between social demand and water conservancy efficiency. Its combined urban water demand and ecological pressures make it representative. The study uses data from 2021 to 2023, sourced from Hefei's economic communiqués and Anhui's water bulletins, covering indicators such as population, water consumption, recycled water use, ecological compliance, energy use, and supply satisfaction. In order to ensure the comparability of data and suitability for model analysis, this paper normalized the original data. The minimum-maximum normalization method was used to convert the values of each indicator to the [0,1] interval. The processed data are shown in Table 2 below: Preprocessed data

years	Permanent population (10,000 people)	Urbanization rate (%)	Industrial water consumption (billion m^3)	Recycled water utilization rate (%)	Agricultural irrigation water consumption (billion m^3)	Ecological flow guarantee compliance rate (%)	Energy consumption per unit of water supply (kWh/m^3)	Comprehensive water supply satisfaction rate (%)
2021	963.4	82.1	5.12	16.5	3.85	90.3	0.59	96.4

2022	975.6	82.7	5.46	17.4	3.88	91.8	0.61	96.9
2023	987.3	83.2	5.64	18.4	3.91	92.6	0.62	97.3

The above data provide a solid foundation for subsequent model analysis, and help to further explore the changes in the matching degree between the social demand structure and the operating efficiency of water conservancy facilities, as well as its lag or advance.

4.2 Result analysis :

From 2021 to 2023, Hefei's data reveal three key insights into the relationship between social demand and water conservancy facility performance. First, the matching degree between demand and efficiency steadily increased, with the coefficient rising from 0.74 to 0.81, reflecting better coordination driven by industrial growth and higher recycled water use. Second, different demands affect facilities differently—industrial water use quickly raises energy consumption, while ecological demand impacts supply satisfaction with a one-year lag. Third, increased recycled water use helps lower energy consumption, forming a positive feedback mechanism that enhances overall operational efficiency..

Table Ⅲ: Matching of Key Indicators of Social Demand and Water Conservancy Facility Operation Efficiency in Hefei City from 2021 to 2023

years	Industrial water consumption (billion m³)	Ecological compliance rate (%)	Comprehensive water supply satisfaction rate (%)	Specific energy consumption (kWh/m³)	Gray matching coefficient γ	Response hysteresis cycle
2021	5.12	90.3	96.4	0.59	0.74	—
2022	5.46	91.8	96.9	0.61	0.78	—
2023	5.64	92.6	97.3	0.62	0.81	Ecology lags behind by 1 year

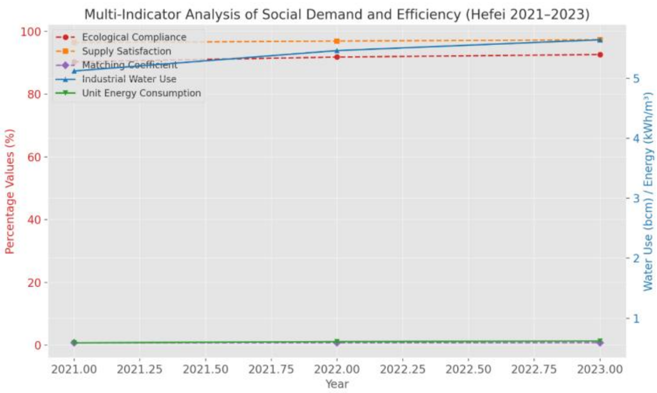

Figure 4. Dual-axis multi-indicator visualization

The figure 4 illustrates a steady upward trend in ecological compliance, supply satisfaction, and the matching coefficient from 2021 to 2023, reflecting improved alignment between social demand and water conservancy operations. Industrial water use and unit energy consumption also increased, indicating higher system load but improved efficiency due to recycled water utilization and enhanced coordination mechanisms.

4.3 Propose dynamic control strategy based on modeling results

Based on the analysis of Hefei's 2021–2023 data, a strong coupling between social demand and water conservancy facility performance is evident, with certain response lags. To enhance operational adaptability, it is recommended to establish a real-time demand forecasting mechanism that integrates data such as water usage, weather, and industrial plans, enabling short-term scheduling optimization. Given the delayed response in ecological demands, a lag compensation mechanism should prioritize ecological allocation through advance planning and predictive analysis. Additionally, to improve the integration of recycled water, a feedback linkage system should be built for real-time data sharing and dynamic scheduling. Overall, transitioning to a data-driven, predictive, and coordinated scheduling model will improve the flexibility and efficiency of water resource management.

5 in conclusion

This study explores the correlation between water conservancy facility operation efficiency and social demand by building an indicator system and applying grey correlation and regression models. Using Hefei's 2021–2023 data, the analysis reveals a clear matching relationship and time lag between demand structure changes and facility responses, particularly in ecological water allocation. Visual verification supports the model's reliability. Based on the findings, a dynamic, data-driven scheduling strategy is proposed to enhance operational flexibility. The results offer a theoretical and practical basis for optimizing regional water resource management, with future research suggested on multi-region, multi-period models to improve applicability.

References

[1]Wang L, Zhao Y. Will social network relationship significantly enhance farmers' participation in the supply of small water-saving irrigation and water conservancy facilities in China?[J]. Agriculture, 2023, 13(1): 216.
[2]Silva J A. Wastewater treatment and reuse for sustainable water resources management: a systematic literature review[J]. Sustainability, 2023, 15(14): 10940.
[3]Raheli H, Zarifian S, Yazdanpanah M. The power of the health belief model (HBM) to predict water demand management: A case study of farmers' water conservation in Iran[J]. Journal of environmental management, 2020, 263: 110388.
[4]Scanlon BR, Fakhreddine S, Rateb A, et al. Global water resources and the role of groundwater in a resilient water future[J]. Nature Reviews Earth & Environment, 2023, 4(2): 87-101.
[5]Molajou A, Pouladi P, Afshar A. Incorporating social system into water-food-energy nexus[J]. Water Resources Management, 2021, 35: 4561-4580.
[6]Katal A, Dahiya S, Choudhury T. Energy efficiency in cloud computing data centers: a survey on software technologies[J]. Cluster Computing, 2023, 26(3): 1845-1875.
[7]Katusiime J, Schütt B. Integrated water resources management approaches to improve water resources governance[J]. Water, 2020, 12(12): 3424.
[8]Gupta AD, Pandey P, Feijóo A, et al. Smart water technology for efficient water resource management: A review[J]. Energies, 2020, 13(23): 6268.
[9]Peng T, Deng H, Lin Y, et al. Assessment on water resources carrying capacity in karst areas by using an innovative DPESBRM concept model and cloud model[J]. Science of The Total Environment, 2021, 767: 144353.
[10]Grigg N S. Water Resources Management: Principles, Methods, and Tools[M]. John Wiley & Sons, 2023.