Journal of Innovation Economics ISSN: 2960-1835
DOI: 10.5281/zenodo.16421788 Volume 1, Issue 2, 2025

Statistical Analysis of Influencing Factors on Water Quality Compliance Rate in the Yellow River Basin

Yuanyuan Zhai[1]

[1] Henan University, Advanced Institute of Finance
Mingli Road, Zhengzhou 45000, China
18638113950@163.com

Abstract: **This study is based on water quality monitoring data from nine provinces and regions in the Yellow River Basin from 2018 to 2024. Descriptive statistics, hypothesis testing and multiple linear regression methods were employed to systematically evaluate the effectiveness of water quality improvement and its influencing factors. The result shows that the proportion of Class I-III water quality sections in the Yellow River Basin increased from 74.3% in 2018 to 93.9% in 2024, with an average annual growth of 3.2 percentage points. Multiple regression analysis indicates that environmental protection investment intensity and sewage treatment rate are the primary drivers of water quality improvement, while rising temperatures have a significant negative impact. This study provides a scientific basis for water environment governance in the Yellow River Basin.**

Keywords: Water Quality Compliance Rate; Correlation Analysis; Multiple Linear Regression; Robustness Test

1. Introduction

The Yellow River Basin stands as a vital ecological shield and a cornerstone of China's economic growth. Its role in safeguarding national ecological security and driving sustainable development cannot be overstated. Recent years have witnessed the remarkable progress under the deepening implementation of the "Yellow River Strategy." Reflected in steadily improving water quality, water environment governance has achieved breakthroughs. The numbers speak volumes: Ministry of Ecology and Environment data reveals that by 2024 (MEE, 2024), Class I to III water quality sections accounted for 93.3% of the basin, marking a striking increase of 25.6 percentage points from 2018 levels (MEE, 2018). These figures underscore the tangible success of concerted conservation efforts. However, challenges such as complex pollution sources, regional development imbalances and climate change continue to exert pressure on further water quality improvements. Against this backdrop, a systematic assessment of water quality and its influencing factors is essential for scientifically evaluating the effectiveness of environmental governance. Such an assessment can also provide theoretical support for optimizing water management strategies. To this end, this study constructs a statistical analysis model to quantitatively assess the actual progress in water quality improvement in the Yellow River Basin, identifying key driving factors and offering decision-making references for more targeted and effective water environment protection policies.

2. Data and Methods

2.1. Data Sources

The data used in this study were sourced from the following authoritative channels:

(1) “The China Ecological Environment Status Bulletin” (2018–2024) published by the Ministry of Ecology and Environment, providing basic information and classification results of water quality monitoring sections in the Yellow River Basin.

(2) Provincial statistical yearbooks of the Yellow River Basin, including indicators such as environmental protection expenditures, regional GDP and sewage treatment rates.

(3) Temperature monitoring data released by the China Meteorological Administration to analyze the potential impact of climate change on water quality. All data were standardized to ensure temporal consistency and regional comparability.

2.2. Variable Descriptions

To comprehensively evaluate the factors influencing water quality improvement in the Yellow River Basin, the following variables were selected:

Dependent Variable: Water quality compliance rate (proportion of Class I-III water quality sections, %), reflecting the overall water environment quality.

Independent Variables: (1) Environmental protection investment intensity (percentage of environmental protection expenditure to GDP, %), measuring local government funding for environmental protection. (2) Sewage treatment rate (%), reflecting urban sewage treatment capacity and pollution control levels. (3) Abnormal temperature (℃), serving as a proxy for climate factors to analyze its impact on water self-purification and pollutant dispersion.

2.3. Analytical Methods

2.3.1. Descriptive Statistical Analysis

Descriptive statistics (mean, standard deviation, maximum, minimum and range) were calculated for water quality compliance rates and related influencing factors to reveal their basic characteristics and temporal trends.

2.3.2. Hypothesis Testing

A paired sample t-test was conducted to compare water quality compliance rates between 2018 and 2024, verifying whether the improvement before and after the implementation of the "Yellow River Strategy" was statistically significant.

DOI: 10.5281/zenodo.16421788 Volume 1, Issue 2, 2025

2.3.3. Multiple Linear Regression Analysis

A multiple linear regression model was constructed to quantitatively analyze the impact of environmental protection investment intensity, sewage treatment rate and temperature anomalies on water quality compliance rates. The model is expressed as:

$$\text{Compliance} = \beta_0 + \beta_1\text{Investment} + \beta_2\text{Treatment} + \beta_3\text{Temperature} + \varepsilon$$

Compliance: Water quality compliance rate
Investment: Environmental investment
Treatment: Sewage treatment rate
Temperature: Abnormal temperature
β_0 : Intercept
β_1, β_2, β_3: Estimated coefficients.
ε: Random error term.

This model identifies the contribution of each factor to water quality changes, providing a basis for formulating scientific water environment governance policies.

3. Statistical Analysis

3.1. Correlation Analysis

Pearson correlation analysis was conducted to preliminarily examine statistical relationships between variables.

Table 1: Pearson Correlation Coefficient Matrix

Variable Quantity	Compliance	Protection	Treatment	Temperature
Compliance	1			
Protection	0.62	1		
Treatment	0.58	0.43	1	
Temperature	-0.41	-0.12	-0.19	1

Tips: $p<0.001$, $p<0.01$, $p<0.05$

The analysis revealed several significant relationships between the variables and the water quality compliance rate. First, a strong positive correlation was observed between environmental investment and water quality improvement, with a correlation coefficient of $r = 0.62$ ($p < 0.001$) (Zhang & Li, 2021). Second, the sewage treatment rate also showed a statistically significant and moderately strong positive association with the compliance rate ($r = 0.58$, $p < 0.001$). In contrast, temperature anomalies exhibited a moderate negative correlation with the compliance rate ($r = -0.41$, $p = 0.002$) (Wang & Smith, 2022), indicating that rising temperatures may hinder water quality improvement efforts. Importantly, the correlation coefficients among the independent variables were all below 0.5, suggesting that multicollinearity is not a serious concern in the model.

Analytical significance:

This result provides a solid foundation for the subsequent regression analysis. The observed correlations confirm statistically meaningful relationships between the key independent variables and the dependent variable. These relationships, in turn, offer both theoretical support and empirical justification for variable selection in model construction. Correlation tests also help identify and address multicollinearity among predictors. This step is essential to ensure the stability and interpretability of the regression model. Moreover, the findings affirm the suitability of using a multiple linear regression approach. Together with the correlation results, they establish a strong methodological and data-driven basis for building robust and scientifically sound models in the next stages of the study.

3.2. Paired Sample t-Test (2018 vs. 2024)

Verify whether there has been a significant change in the rate of water quality compliance in the Yellow River basin between 2018 and 2024 (MEE, 2018, 2024).

Data preparation: Select water quality compliance rate data of 9 provinces in the basin in 2018 and 2024, and Calculate the difference for each province.

Table 2: Descriptive Statistics

Statistics	2018 (%)	2024 (%)	d-value (%)
Mean	68.2	85.7	17.5
SD	10.3	6.5	8.2
SE	3.4	2.2	2.7

Because it is one-tailed test, so $\alpha=0.05$, $t=1.86$. As shown above, the t-value is 6.4, which is substantially higher than the critical value. Therefore, the water quality compliance rate in the Yellow River Basin has undergone a highly significant improvement from 2018 to 2024.

3.3. Multiple Linear Regression Analysis

Verify whether there has been a significant change in the rate of water quality compliance in the Yellow River basin between 2018 and 2024 (MEE, 2018, 2024).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Y: Water quality compliance rate (proportion of sections with class I-III water quality, %)
X_1: Environmental protection input intensity (environmental protection expenditure as a percentage of GDP, %)
X_2: Sewage treatment rate (%)
X_3: Abnormal temperature (℃)
X_4: The proportion of industrial added value (the proportion of secondary industry added value in GDP, %) is taken as the control variable.
β_0 : Intercept
β_1, β_2, β_3: Estimated coefficients.
ε: Random error term.

To test for homoscedasticity, the Breusch-Pagan test was employed to examine the constancy of model residuals' variance.

Table 3: Regression Coefficient Estimation Results

Non-stand. Coefficients	SE	Stand. Coefficients	T-value	P-value	95% CI
52.41	2.33		22.49	<0.001	[47.82,57.00]
0.39	0.11	0.36	3.45	0.001	[0.17,0.61]
0.34	0.1	0.31	3.27	0.002	[0.14,0.54]
-0.16	0.07	-0.16	-2.24	0.028	[-0.30,-0.02]

Tips: $p<0.001$, $p<0.01$, $p<0.05$

DOI: 10.5281/zenodo.16421788 Volume 1, Issue 2, 2025

In the result determination, the degrees of freedom corresponding to the number of predictor variables is 3 based on relevant statistical analysis. Under the chi-square distribution, when the significance level α=0.05 and the degrees of freedom is 3, the critical value $\chi^2_{0.05,3} = 7.815$ and the test statistic is 2.15. Since 2.15 is far less than 7.815, the significance level corresponding to the critical value is not reached. Combined with the test background, the current data situation supports the conclusion represented by the null hypothesis. The specific interpretation can be further accurately analyzed in combination with the test scenario.

The overall significance test of the model revealed an F-statistic of 31.7 with degrees of freedom of 3 and 80, yielding a p-value <0.001. This indicates the model demonstrates high statistical significance, demonstrating that independent variables collectively exert significant explanatory power over the dependent variable. In terms of explanatory power, the R^2 value reached 0.69, explaining 69% of the variance in the dependent variable. The adjusted R^2 value of 0.66 maintained strong explanatory capacity even when accounting for the number of variables. The standard error of the prediction stands at 4.98, suggesting an average prediction error of approximately 5 percentage points.

The research findings reveal a significant positive relationship between environmental protection investment and water quality compliance rates. For every 1% increase in environmental investment, the compliance rate improves by 0.39% (p = 0.001). This highlights the critical role of financial input in improving water quality. In addition, sewage treatment rates also show a strong positive effect. A 1 percentage point increase leads to a 0.34% improvement in compliance rates (p = 0.002), underscoring the importance of infrastructure development and enhancing pollution control in achieving better water quality outcomes. However, rising temperatures have a negative impact on water quality. Each 1 °C increase is associated with a 0.16% decline in compliance rates (p = 0.028). These results suggest that climate change may pose increasing challenges to water environment management.

Interpretation of effect size analysis results:

In the effect size analysis of influencing factors, the standardized regression coefficient for environmental protection investment was 0.36. This indicates that, among all variables included in the analysis, environmental protection investment had the strongest impact on the dependent variable and served as the core driving factor for water quality improvement. In contrast, temperature variables exhibited a negative effect with statistically significant significance. However, considering its standardized coefficient (β = -0.16), the actual influence of temperature on water quality changes was relatively minor, classifying it as a secondary influencing factor.

To further validate the robustness of the research findings, this study conducted multidimensional robustness tests. Firstly, through sub-sample analysis divided by the upper, middle and lower reaches of the river basin, it was found that environmental protection investments in upstream regions demonstrated more significant effects. The standardized coefficient reached 0.45, significantly higher than the estimated values in the overall sample, further confirming the importance and effectiveness of increased environmental protection investments in upstream areas. In downstream regions, however, the impact of sewage treatment rates proved more pronounced, with a standardized coefficient of 0.40, indicating that enhancing sewage treatment capacity in these areas had a more direct effect on water quality improvement.

Furthermore, this study conducted time-segmented regression analysis, dividing the period into two phases with 2020 as the cutoff. The results indicate that during 2018-2020, the negative impact of temperature on water quality compliance rates was not statistically significant (p = 0.152), showing that climate change had no notable effect during this phase. However, from 2021 to 2024, the negative impact of temperature became significantly stronger, with a standardized coefficient of-0.21 and passing statistical significance tests. This demonstrates that over time, the influence of temperature on water quality gradually emerged and intensified, indicating that future research and policy-making should pay greater attention to the potential impacts of climate change.

3.4. Fault Tolerance Test

3.4.1. Heteroskedasticity Test

The purpose is to verify whether the variance of the residual is constant (homogeneity). If heteroskedasticity exists, the standard error estimation of the regression coefficient may be inaccurate, which will affect the statistical inference.

The Breusch-Pagan test was employed to assess homoscedasticity by first fitting the original regression model and obtaining the squared residuals. Subsequently, an auxiliary regression model was constructed, with the squared residuals as the dependent variable and the original independent variables, including environmental investment, sewage treatment rate, and temperature anomalies. These serve as the explanatory variables.

$$\lambda_i^2 = \alpha_0 + \alpha_1 \text{Investment} + \alpha_2 \text{Treatment} + \alpha_3 \text{Temperature} + \mu_i$$

Investment: Environmental investment
Treatment: Sewage treatment rate
Temperature: Abnormal temperature
α_0 : Intercept
α_1, α_2, α_3: Estimated coefficients.
μ_i: Random error term.

To assess whether the regression model residuals satisfy the assumption of homoscedasticity, this study employed the Breusch-Pagan test. In this procedure, an auxiliary regression was first conducted by regressing the squared residuals from the original model onto all explanatory variables—namely, environmental protection investment, sewage treatment rate, and temperature anomalies. The resulting Lagrange Multiplier (LM) statistic was then calculated to evaluate the null hypothesis of constant error variance.

With a sample size of n = 84, the LM statistic follows a chi-square distribution with three degrees of freedom. The computed LM value was 2.15, which is notably lower than the critical value of 7.815 at the 0.05 significance level. The associated p-value was 0.342, indicating insufficient evidence to reject the null hypothesis. Therefore, it can be concluded that the residuals exhibit homoscedasticity. Under the current model specification, no significant heteroskedasticity is present. Therefore, no corrective measures such as weighted least squares (WLS) are required. Subsequent analysis and interpretation are thus based on the estimates obtained from ordinary least squares (OLS) regression.

3.4.2 Robustness Test

The stability of the results is verified by different subsamples or model Settings to ensure that the conclusions are not affected by data partitioning or time changes.

(1) Regional Regression

In the regional heterogeneity analysis, distinct patterns emerge across the upper and lower reaches of the basin. In the upper reaches, environmental investment exhibits the strongest effect (β = 0.45), which may be attributed to the ecological fragility of the region, making it more responsive to financial inputs. In contrast, in the lower reaches, the sewage treatment rate demonstrates the most significant impact (β = 0.40), reflecting the higher demand for pollution control from industrial and agricultural activities. Furthermore, the negative effect of temperature anomalies is more pronounced in the lower reaches (β = -0.20), potentially due to reduced water mobility, which exacerbates the impact of climatic factors on water quality.

Table 4: Test the Differences of Influencing Factors in Different River Basin Sections

Region	N	Environmental Protection Input (β)	Treatment (β)	Temperature (β)
Upstream	28	0.45	0.28	-0.12
Midstream	28	0.38	0.35	-0.18
Downstream	28	0.3	0.4	-0.2

(2) Time-period Regression

The analysis reveals a strengthening negative effect of temperature anomalies after 2021, with the coefficient declining from –0.11 to –0.21, suggesting an intensified impact possibly linked to the increasing frequency of extreme climate events. Meanwhile, the policy effect of environmental investment shows a slight decline (β= 0.42 to β= 0.37), which may indicate diminishing marginal returns as initial investments yield progressively smaller improvements in water quality.

The validation results demonstrate that the model does not exhibit heteroskedasticity, with regression outcomes demonstrating high stability and reliability. Analysis further reveals significant regional disparities: Water quality improvements in upstream areas primarily depend on increased environmental protection investments, while downstream regions require greater emphasis on enhancing sewage treatment capacity. Moreover, climate change impacts have become increasingly pronounced. Specifically, since 2021, the temperature increases nearly doubled the negative effects on water quality. This indicates that climate factors must be incorporated into long-term management strategies and policy formulation for future water quality governance.

Table 5: Analysis of dynamic changes in the impacts of climate change (2018-2020 vs. 2021-2024)

Time period	Abnormal Temperature（β）	Environmental Protection Input（β）	Sewage Treatment Rate（β）
2018-2020	-0.11	0.42	0.31
2021-2024	-0.21	0.37	0.36

4. Conclusions and Suggestions

4.1 Main Conclusions

The study highlights significant improvements in water quality compliance rates across the Yellow River Basin from 2018 to 2024, with Class I-III water quality sections increasing from 74.3% to 93.9%. However, regional disparities reveal deeper challenges that require targeted solutions. Upstream regions like Qinghai and Sichuan achieved higher compliance rates due to effective ecological investments, while downstream areas such as Henan and Shandong lagged despite similar funding levels, indicating systemic imbalances in governance effectiveness. The analysis identifies environmental investment and sewage treatment capacity as primary drivers of improvement, with each 1% increase in GDP allocation and treatment rates boosting compliance by 0.39% and 0.34% respectively. However, the growing negative impact of climate change, particularly post-2021 (Wang & Smith, 2022) where temperature anomalies showed doubled adverse effects, underscores the urgent need for adaptive strategies (World Bank, 2023).

4.2 Recommendations

To address these challenges, a three-tier early warning system has been developed, alongside the promotion of temperature-resilient water purification technologies. Specifically, when temperature anomalies reach or exceed 1 °C, a yellow alert should be issued, accompanied by emergency water quality monitoring. If anomalies rise to 2 ℃ or higher, an orange alert is triggered, leading to temporary emission restrictions on major polluters. At 3 ℃ or above, a red alert is activated, calling for ecological water replenishment projects to mitigate the adverse effects of high temperatures on water resources. Additionally, to adapt to rising water temperatures caused by climate change, pilot programs are being implemented in regions such as Shandong and Henan to promote advanced purification technologies, such as biofilm-activated carbon hybrid systems, which are more resilient to high-temperature conditions. These measures aim to ensure water quality safety under changing climatic conditions.

Looking ahead, more measures need to be done. Policy interventions must move beyond uniform approaches and embrace precision governance tailored to regional realities. Upstream regions should maintain ecological investments at 2.1-2.3% of GDP but enhance accountability through performance-linked funding, while midstream industrial hubs like Shaanxi and Shanxi require elevated investments (2.5%-2.8% of GDP) focused on real-time emission monitoring and coal-sector upgrades. Downstream agricultural zones need targeted subsidies for precision farming technologies to curb non-point source pollution, complemented by decentralized sewage solutions like

DOI: 10.5281/zenodo.16421788 Volume 1, Issue 2, 2025

modular bioreactors in rural areas. Climate resilience demands proactive measures, including a tiered alert system triggering emergency monitoring at 1℃ anomalies and industrial restrictions at 3℃, alongside nature-based solutions like wetland restoration to mitigate heatwave impacts.

Table 6: Key Construction of Sewage Treatment Facilities in the Middle Reaches of Provinces

Province	Current Processing Rate	2025 Target	Major Project
Shaanxi	83%	90%	Xi'an fourth sewage treatment plant expansion
Shanxi	79%	88%	Fenhe River basin pipe network renovation

Cross-regional collaboration is equally critical. Establishing a unified data platform for real-time pollution tracking across provinces would enable coordinated action, while innovative financing mechanisms like water quality bonds could incentivize upstream conservation by allowing pollution credit trading. Looking ahead, research must delve into pollutant-specific responses to investments and integrate climate projections to model long-term water quality scenarios. By combining tiered investments, adaptive infrastructure, and basin-wide cooperation, China can transform the Yellow River Basin into a model of sustainable water governance that balances ecological protection with economic development, setting a precedent for managing complex river systems worldwide. The findings not only validate the success of current policies but also chart a course for addressing emerging challenges through evidence-based, forward-looking strategies.

References

[1] Ministry of Ecology and Environment. (2018–2024). Communiqué on the state of the ecological environment in China.
[2] Zhang, M., & Li, H. (2021). Evaluation on the effectiveness of water environment governance in the Yellow River Basin. Research of Environmental Sciences, 34(5), 1120–1128.
[3] Wang, L., & Smith, J. (2022). Climate change impacts on water quality. Nature Water, 1(1), 25–32.
[4] Yellow River Conservancy Commission. (2021). Water resources protection plan for the Yellow River Basin (2021–2030).
[5] Ministry of Ecology and Environment. (2024). China ecological environment bulletin.
[6] World Bank. (2023). Climate-smart water governance.
[7] Liu, X., et al. (2025). AI for river basins. Nature Sustainability, 8(3), 123–135.