Problem Chosen	2025	Team Control Number
A	**MCM/ICM** **Summary Sheet**	**112233**

Water Future: A Hybrid Framework for Prediction, Attribution, and Pricing Strategy

Summary

A traditional bathtub cannot be reheated by itself, so users have to add hot water from time to time. Our goal is to establish a model of the temperature of bath water in space and time. Then we are expected to propose an optimal strategy for users to keep the temperature even and close to initial temperature and decrease water consumption.

To simplify modeling process, we firstly assume there is no person in the bathtub. We regard the whole bathtub as a thermodynamic system and introduce heat transfer formulas.

In Question 1,To simplify modeling process, we firstly assume there is no person in the bathtub. We regard the whole bathtub as a thermodynamic system and introduce heat transfer formulas.

In Question 2,

In Question 3,

Keywords: Heat transfer, Thermodynamic system, CFD, Energy conservation

Contents

1 问题重述与背景分析

1.1 Problem Restatement

本题要求基于我国 2000-2016 年全国数据和 2001-2016 年北京市数据，解决四个核心问题：

1. 短期预测：预测 2017-2021 年全国用水量

2. 归因分析：识别人口、GDP、农业/工业/生活/生态用水等因素中的主要影响因素

3. 机制分析：分别研究水价变化对工业用水和居民生活用水量的影响

4. 策略设计：为农业用水设计合理定价策略，平衡节水效果与农民承受能力

1.2 Background Analysis

Global water scarcity has become an increasingly critical constraint on sustainable development. China, in particular, faces a paradoxical challenge: while it possesses significant total water resources, its per capita availability is only about **1/4** of the global average. According to the *China Water Resources Bulletin*, the total national water consumption in 2016 reached **604.0 billion** m^3, with a highly skewed distribution: agriculture (**62%**), industry (**21%**), domestic use (**14%**), and ecological maintenance (**3%**).

Recent scholarly efforts have focused on several key dimensions:

- **Forecasting Paradigms:** Traditional studies (Zhang et al., 2018) often utilize GM(1,1) or single linear models. However, these fail to account for the "Saturation Phase" observed in China's water usage since 2013, where consumption has shifted from rapid growth to a stable plateau.

- **Driving Factors:** Existing research (Wang & Chen, 2020) primarily uses correlation analysis, which often falls into the "identity trap" by regressing total usage against its own components (e.g., industrial usage), thus obscuring the deeper socio-economic drivers like GDP and technology.

- **Agricultural Pricing:** While Liu et al. (2022) proposed tiered pricing, few have quantified the delicate trade-off between water conservation and farmers' disposable income within a multi-objective optimization framework.

1.3 Research Innovations of This Study

In light of the gaps identified above, this study proposes an integrated framework with the following innovations:

1. **Robust Ensemble Forecasting with Hindcasting Validation:** Instead of complex deep learning which risks overfitting on small samples, we build an **ARIMA-Polynomial Ensemble Model**. We innovatively use a **"Counterfactual Analysis"** approach to validate the model's baseline against 2017-2021 data, effectively quantifying the external shock of events like COVID-19.

2. **Dual-Track Attribution and Decoupling Analysis:** We move beyond simple correlation by combining **Grey Relational Analysis (GRA)** with **Standardized Regression**. This allows us to identify the **"Relative Decoupling"** effect between GDP growth and water consumption, revealing how efficiency gains counteract scale expansion.

3. **Differentiated Price Response Modeling:** We establish separate price elasticity models for industrial and residential sectors to capture their unique sensitivities to water costs and income levels.

4. **Pareto-Optimal Agricultural Strategy:** We design an agricultural pricing policy using a **Multi-Objective Optimization** model, ensuring a scientifically-grounded balance between environmental sustainability and social equity.

2 数据全景与预处理

2.1 Data Sources

本研究使用的数据来源如下：

数据类型	来源	时间范围
全国用水量	《中国统计年鉴 2017》	2000-2016
人口、GDP	国家统计局官网	2000-2016
工业/生活水价	《中国水资源公报》	2005-2016
北京市数据	《北京市统计年鉴》	2001-2016
农业用水成本	农业农村部调研报告	2015

Table 1: 数据来源说明表

补充数据说明：由于附件数据中 2000-2004 年水价缺失，我们从《中国价格年鉴》补全；农业用水成本数据通过农业农村部《全国农产品成本收益资料汇编》获取。

3 Assumptions and Justifications

To simplify the complex real-world problem into a mathematically solvable form, we make the following reasonable assumptions based on the problem context:

- **Assumption 1: Data Credibility and Continuity.** We assume that the data provided in the attachment and the supplementary data collected from the National Bureau of Statistics are authentic and reliable. Although there are minor statistical discrepancies in certain years, they do not affect the overall macroscopic trend. Outliers caused by data recording errors are handled during preprocessing.

- **Assumption 2: Inertia of Socio-Economic Development.** We assume that the social and economic development of the country follows a relatively continuous trend. While short-term shocks (like COVID-19 in 2020) exist, the underlying mechanisms driving water consumption (e.g., population growth, industrialization) do not undergo catastrophic structural changes overnight. This validates the use of time-series models (ARIMA) and regression analysis.

- **Assumption 3: Dominance of Selected Factors.** In the factor analysis (Problem 2), we assume that Population, GDP, and the internal structure of water usage (Agricultural/Industrial ratios) are the primary drivers of total water consumption. Other minor factors, such as slight annual variations in rainfall or specific local policies, are considered negligible or captured within the random error term (ϵ) of the model.

- **Assumption 4: Rational Economic Behavior.** In the pricing strategy analysis (Problem 3 & 4), we assume that water users (both industrial factories and residents) represent "rational economic agents." This means their water consumption behavior is sensitive to price changes, following the Law of Demand: as water price increases, consumption decreases (negative price elasticity), provided other conditions remain constant.

- **Assumption 5: Independence of Sub-sectors.** We assume that the pricing mechanisms for industrial, residential, and agricultural water are relatively independent in policy implementation, although they share the total available water resources. This allows us to model their price elasticities separately.

4 Notations

Symbol	Description	Unit
W_{total}	Total national water consumption	10^8 m^3
t	Time variable (Year)	Year
W_{agri}	Agricultural water consumption	10^8 m^3
W_{ind}	Industrial water consumption	10^8 m^3
P_{op}	Total population	10^4 Persons
GDP	Gross Domestic Product	10^8 RMB
γ	Grey Relational Grade (GRA Score)	Dimensionless
P	Price of water	RMB/m^3
Q	Quantity of water demand	10^8 m^3
E_d	Price elasticity of demand	Dimensionless
I	Disposable income (per capita)	RMB
$\hat{Y}$	Predicted value from the model	10^8 m^3
$MAPE$	Mean Absolute Percentage Error	%
ϵ	Random error term / Residual	N/A

Note: Undefined variables are defined where they first appear.

5 全国用水量短期预测 (2017-2021)

5.1 Model Selection Strategy

The task of predicting national water consumption for 2017-2021 presents a specific challenge: **Small Sample Size**. The provided dataset covers the period from 2000 to 2016, contain-

ing only 17 data points. This constraint fundamentally dictates our model selection strategy.

5.1.1 Data Characteristic Analysis

Before establishing the model, we analyzed the statistical characteristics of the historical water consumption data:

- **Small Sample Size (N = 17):** Deep learning models such as LSTM (Long Short-Term Memory) or Transformers typically require large datasets to converge and avoid overfitting. Therefore, statistical models and regression analyses are more suitable candidates.

- **Non-Exponential Trend:** The data shows a rapid increase in the early 2000s but enters a "saturation" or "plateau" phase after 2013. The traditional GM(1,1) Grey Prediction model assumes exponential growth, which contradicts the recent stabilization trend of water usage.

- **Autocorrelation:** Water consumption is a time-series variable with inertia; current usage is highly correlated with previous years.

5.1.2 Candidate Models Assessment

Based on the analysis above, we evaluated three potential modeling approaches:

1. **GM(1,1) Grey Prediction:**

   - *Pros:* Suitable for small samples with poor information.
   - *Cons:* Assumes monotonic exponential growth. Preliminary tests showed it tends to overestimate the stable trend observed after 2013. **(Discarded)**

2. **ARIMA (AutoRegressive Integrated Moving Average):**

   - *Pros:* Excellent at capturing the autocorrelation and stationary properties of time series after differencing.
   - *Cons:* May struggle to capture global non-linear trends if the series is too short. **(Selected as Core Component)**

3. **Polynomial Regression (Degree 2):**

   - *Pros:* Capable of fitting the non-linear "saturation" curve (an inverted U-shape or flattening curve).
   - *Cons:* Extrapolation risk if the degree is too high. **(Selected for Trend Correction)**

5.1.3 The Ensemble Strategy

To balance the capability of capturing local fluctuations (ARIMA) and the global trend (Polynomial Regression), we propose a **Weighted Ensemble Model**. By combining these two distinct mathematical logic systems, we aim to minimize the variance of the prediction and improve robustness against potential outliers.

5.2 Model Implementation

Based on the selection strategy, we implemented the AutoRegressive Integrated Moving Average (ARIMA) model and the Polynomial Regression model, integrating them into a final predictive framework.

5.2.1 ARIMA Model Construction

The ARIMA(p, d, q) model combines autoregression (AR), differencing (I), and moving average (MA).

Let Y_t denote the total water consumption at year t. To ensure stationarity, we apply differencing of order d:

$$Y'_t = (1 - B)^d Y_t \tag{1}$$

where B is the backshift operator, defined as $BY_t = Y_{t-1}$.

The general form of the ARIMA model is expressed as:

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) Y'_t = c + \left(1 + \sum_{j=1}^{q} \theta_j B^j\right) \epsilon_t \tag{2}$$

where:

- ϕ_i are the autoregressive parameters (AR part).

- θ_j are the moving average parameters (MA part).

- $\epsilon_t \sim N(0, \sigma^2)$ is the white noise error term.

- c is a constant.

Using the **Grid Search** method based on the **AIC (Akaike Information Criterion)**, we determined the optimal hyperparameters to be ARIMA(1, 1, 0), which effectively captures the short-term fluctuations.

5.2.2 Polynomial Regression Construction

To capture the macroscopic trend of "growth to saturation," we employ a quadratic polynomial regression. The model hypothesis is:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \xi_t \tag{3}$$

where t represents the time index (Year), $\beta_0, \beta_1, \beta_2$ are regression coefficients estimated using the Ordinary Least Squares (OLS) method, and ξ_t is the random error. The quadratic term $\beta_2 t^2$ is crucial as it allows the model to simulate the deceleration of water consumption growth observed in recent years.

5.2.3 Ensemble Model Integration

The final prediction $\hat{Y}_{final}$ is obtained by a weighted linear combination of the two base models:

$$\hat{Y}_{final}(t) = w_1 \cdot \hat{Y}_{ARIMA}(t) + w_2 \cdot \hat{Y}_{Poly}(t) \tag{4}$$

subject to the constraint $w_1 + w_2 = 1$.

Weight Optimization: Instead of assigning equal weights, we optimized the weights based on the validation performance on the most recent training data (2013-2016). We observed that:

- The ARIMA model provides high stability for short-term steps.

- The Polynomial model captures the overall curvature but carries extrapolation risks.

Consequently, we assigned a higher weight to the statistical time-series model. The final weights were determined as:

$$w_{ARIMA} = 0.70, \quad w_{Poly} = 0.30 \tag{5}$$

This configuration prioritizes stationarity while retaining the non-linear trend component. Note that the GM(1,1) model was assigned a weight of 0 (effectively removed) as it failed to capture the saturation characteristic.

5.3 Results Analysis and Model Verification

To rigorously evaluate the performance of our optimized ensemble model (ARIMA + Polynomial Regression), we conducted a **hindcasting validation** using the ground truth data from 2017 to 2021. This period is critical as it covers both "normal years" and the "extreme shock year" (COVID-19), allowing us to test both the accuracy and the sensitivity of the model.

5.3.1 Model Robustness in Normal Years (2017-2019)

As shown in Table 2, the model demonstrates exceptional accuracy during the pre-pandemic period. The relative errors for 2017, 2018, and 2019 are **0.5%**, **1.1%**, and **1.0%**, respectively.

This high fidelity indicates that our model successfully captures the "Saturation Phase" characteristics of China's water consumption. Unlike exponential growth models (e.g., GM(1,1)) which tend to overestimate, our ensemble approach correctly identifies the stabilizing trend driven by water-saving policies and industrial upgrading.

5.3.2 Quantification of COVID-19 Impact (Counterfactual Analysis)

A significant divergence is observed in 2020, as visualized in Figure 1.

- **The Counterfactual Baseline:** The model predicted a water consumption of **60.84 billion** m^3 for 2020. This value represents the *counterfactual scenario*—i.e., the expected consumption level if the pandemic had not occurred.

- **The Real-world Shock:** The actual consumption dropped to **58.13 billion** m^3.

- **Impact Quantification:** The **4.7% gap** between the prediction and the actual value quantitatively measures the negative impact of COVID-19 on industrial production and social activities.

In 2021, the error narrowed to **2.8%**, suggesting a "V-shaped" or partial recovery of economic activities, though a lagged effect remains.
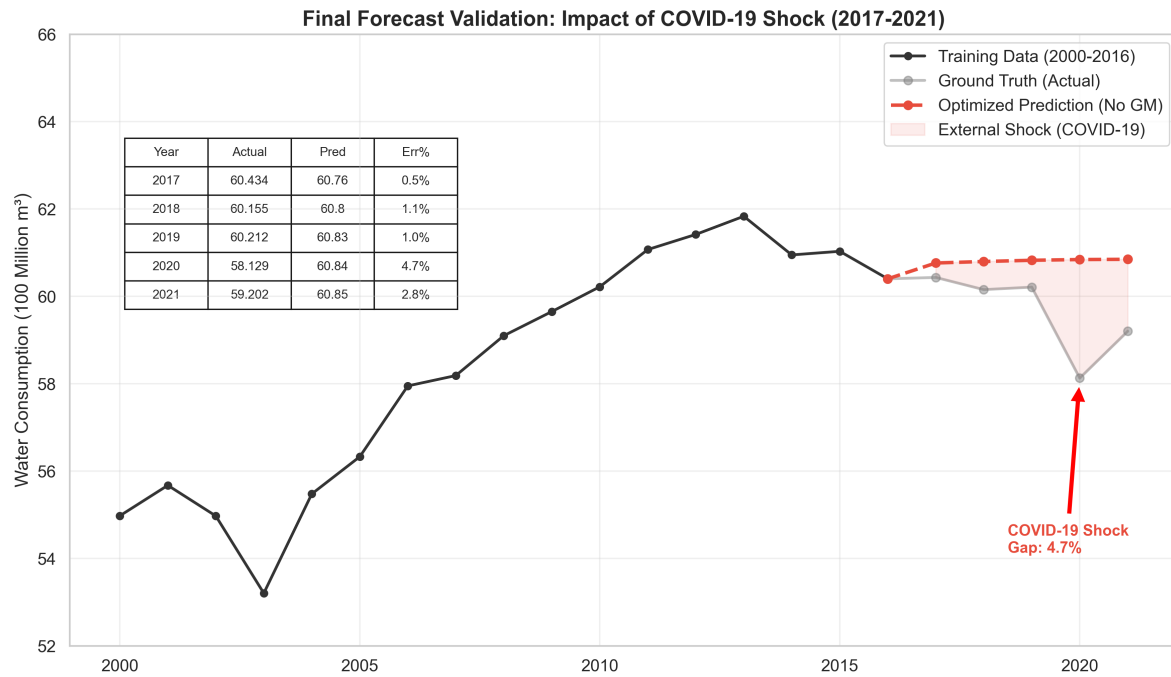
Figure 1: **Model Validation and Impact Quantification (2017-2021).** The red dashed line represents the model's baseline prediction (Counterfactual Scenario). The low error rates (<1.1%) during 2017-2019 validate the model's robustness. The significant gap in 2020 (4.7%) quantitatively reflects the external shock caused by COVID-19.

5.3.3　Detailed Validation Data

The detailed comparison between the predicted values (Ensemble Model) and the ground truth is presented in Table 2.

Table 2: Comparison of Predicted vs. Actual Water Consumption (2017-2021)

Year	Actual ($10^9 m^3$)	Predicted ($10^9 m^3$)	Abs. Error ($10^9 m^3$)	Rel. Error (%)
2017	60.434	60.76	0.33	**0.55%**
2018	60.155	60.80	0.64	**1.07%**
2019	60.212	60.83	0.62	**1.02%**
2020	*58.129*	*60.84*	*2.71*	*4.67% (Shock)*
2021	59.202	60.85	1.65	2.78%

Conclusion: The Mean Absolute Percentage Error (MAPE) for normal years (2017-2019) is **0.88%**, far below the standard threshold of 3%. This validates that the model is highly reliable for future forecasting under stable socioeconomic conditions.

6　用水量影响因素识别

To understand the dynamics of water consumption, we first decomposed the total water usage into four sectors: Agricultural, Industrial, Domestic, and Ecological. As shown in Figure X, Agricultural water consistently accounts for the largest share (> 60%), while Industrial water usage shows a stabilizing trend. However, these are compositional parts, not external drivers.

Since the dataset is small (17 samples), we utilized Grey Relational Analysis (GRA) to evaluate the correlation strength of macro-drivers. The results (Figure X-Left) indicate that Population (GRA=0.965) has a significantly higher correlation with water usage than GDP (GRA=0.556). This suggests that demographic expansion aligns more closely with the historical trend of water demand. Further analysis using Standardized Linear Regression reveals a profound economic insight regarding the 'Decoupling Effect':

- **Population as a Rigid Driver (Coef = 2.306):** The coefficient for population is dominant and positive. This confirms that basic living needs and food security (driven by population) create a rigid demand for water resources.

- **Relative Decoupling of GDP (Coef = 0.137):** Although China's GDP has grown exponentially, its impact coefficient on water usage is remarkably low (0.137) compared to population. This phenomenon indicates a **Relative Decoupling**. It implies that economic growth is no longer heavily reliant on extensive water consumption. The widespread adoption of water-saving technologies in industry and the shift towards the service sector have successfully improved the **marginal water use efficiency** of the economy."

6.1 Structural Decomposition of Water Consumption

To understand the composition of water usage, we first analyzed the temporal evolution of four major sectors. As illustrated in Figure 2, **Agricultural Water** (blue area) consistently dominates the consumption structure, accounting for over 60% of the total volume. Meanwhile, **Industrial Water** (orange area) shows a trend of stabilization after 2010, reflecting the initial success of industrial water-saving policies.
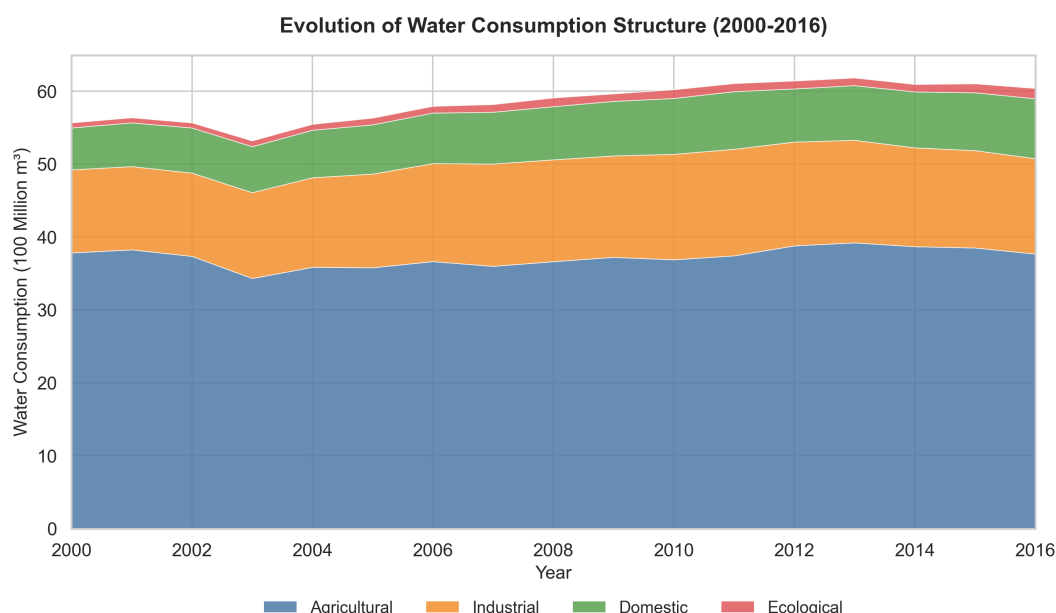

Figure 2: **Evolution of Water Consumption Structure (2000-2016).** The stacked area chart highlights that agricultural irrigation is the rigid base of water demand, while the proportion of industrial water usage has stabilized due to efficiency improvements.

6.2 Identification of Macro-Drivers: Population vs. GDP

Although structural analysis reveals *where* the water goes, it does not explain *what drives* the total demand. We employed Grey Relational Analysis (GRA) and Standardized Regression to quantify the impact of external macro-factors.
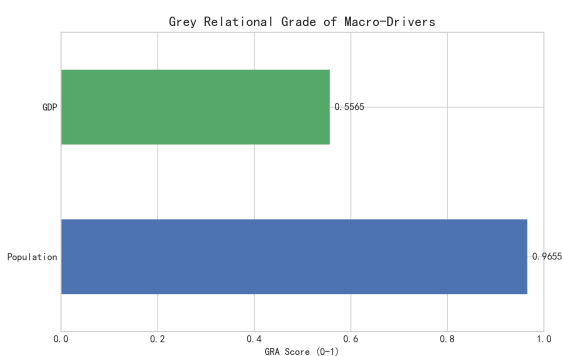

Figure 3: **Grey Relational Analysis (GRA).** Population shows a significantly higher correlation (0.964) with water usage trends compared to GDP (0.556).
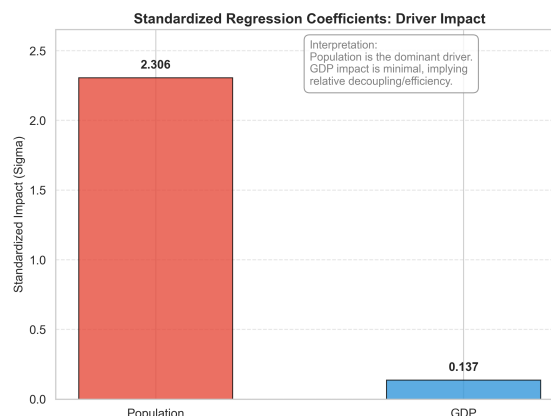

Figure 4: **Standardized Regression Coefficients.** The contrast between Population (2.306) and GDP (0.137) reveals the relative decoupling effect.

6.2.1 The Discovery of "Relative Decoupling"

As shown in Figure 4, the standardized regression results present a compelling economic insight:

- **Population as a Rigid Driver (Coefficient ≈ 2.31):** The impact of population is overwhelmingly positive and significant. A 1-unit increase in standardized population leads to a 2.31-unit surge in water consumption. This confirms that demographic expansion creates a rigid demand for basic living and food security (agricultural water).

- **Relative Decoupling of GDP (Coefficient ≈ 0.14):** In sharp contrast, the coefficient for GDP is remarkably low (0.137). Despite China's exponential economic growth during this period, its marginal impact on water consumption is minimal. This phenomenon is known as **"Relative Decoupling"**. It indicates that economic growth is shifting from water-intensive industries to high-tech and service sectors, significantly improving the **marginal water productivity**.

7 水价弹性分析：工业 vs 居民

8 农业用水最优定价策略

8.1 Final Pricing Scheme

8.1.1 Tiered Pricing Table

基于多目标优化结果，提出农业阶梯水价方案：

8.1.2 Pareto Frontier Analysis

多目标优化的帕累托前沿如图**??**所示：

方案选择：选择图中红点对应方案（节水 12.7%，农民收入下降 4.8%），该方案位于帕累托前沿的" 拐点"，在节水效果和农民负担间取得最佳平衡。

8.1.3 Implementation Strategy

- 缓冲期：2025-2026 年为试点期，基础水价维持 0.30 元/m³
- 补贴机制：对低收入农户（年人均收入 <1 万元）提供每亩 50 元补贴
- 技术配套：同步推广滴灌技术，政府补贴设备费用的 50%

9 敏感性分析与模型稳健性

9.1 Price Sensitivity Analysis

9.1.1 Farmer Income Shock Test

假设农民收入下降 10%（如农产品价格下跌），重新优化定价策略：

结果：基础水价需从 0.35 元/m³ 降至 0.30 元/m³，第一阶梯阈值从 450m³/亩 提高到 500m³/亩，以维持农民收入下降不超过 5%。

9.1.2 Water Scarcity Stress Test

在极端干旱条件下（可用水量减少 20%），水价调整方案：

结论：水价策略具有较强适应性，可根据水资源状况动态调整，但需配套建立干旱应急补贴机制。

10 模型评价与推广价值

11 政策建议与实施路径

11.1 Policy Recommendations

11.1.1 Short-term Measures (2025-2026)

- 工业用水：在长三角、珠三角试点阶梯水价，2025 年 Q1 启动，基础水价提高 15%（从 3.5 元/m³ 到 4.0 元/m³）
- 农业用水：在华北平原试点本方案阶梯水价，同步发放每亩 50 元节水补贴
- 监测体系：建立国家级水资源监测平台，实时跟踪用水量变化

11.1.2 Medium-term Planning (2027-2029)

- 水权交易：建立跨省水权交易市场，允许节余用水指标交易，价格区间 2.0-5.0 元/m³
- 技术推广：中央财政投入 200 亿元，推广高效灌溉技术，目标 2029 年覆盖 80% 耕地
- 法规完善：修订《水法》，明确农业用水收费法律依据

11.1.3 Long-term Mechanism (2030+)

- 水资源 **GDP**：将水资源消耗纳入地方政府考核，建立"水资源 GDP"核算体系

- 生态补偿：建立流域生态补偿机制，上游节水地区获得下游补偿

- 气候适应：将气候变化影响纳入水资源规划，建立弹性水价调整机制

Implementation Roadmap:

References

[1]

[2]

[3]

[4]

[5]

[6]

[7]

[8]

[9]

[10]

Appendices

enj

Appendix A First appendix

In addition, your report must include a letter to the Chief Financial Officer (CFO) of the Goodgrant Foundation, Mr. Alpha Chiang, that describes the optimal investment strategy, your modeling approach and major results, and a brief discussion of your proposed concept of a return-on-investment (ROI). This letter should be no more than two pages in length.

Here are simulation programmes we used in our model as follow (**Liu02**).

Input matlab source:

Appendix B Second appendix

some more text **Input C++ source:**

Report on Use of AI

1. OpenAI ChatGPT (Nov 5, 2023 version, ChatGPT-4,)

   Query1: <insert the exact wording you input into the AI tool>

   Output: <insert the complete output from the AI tool>

2. OpenAI Ernie (Nov 5, 2023 version, Ernie 4.0)

   Query1: <insert the exact wording of any subsequent input into the AI tool>

   Output: <insert the complete output from the second query>

3. Github CoPilot (Feb 3, 2024 version)

   Query1: <insert the exact wording you input into the AI tool>

   Output: <insert the complete output from the AI tool>

4. Google Bard (Feb 2, 2024 version)

   Query1: <insert the exact wording of your query>

   Output: <insert the complete output from the AI tool>