2021

High School Mathematical Contest in Modeling (HiMCM) & Middle Mathematical Contest in Modeling (MidMCM)
Summary Sheet
Team Control Number: **11606**
Problem Chosen: **B**

Summary

Predict the Water Level Based on the Time Series Model

Lake Mead which is the largest water reservoir in the United States provides and satisfies the demands of water for homes, businesses and industries. However, it reaches its lowest level in the summer of 2021. To relieve the degree of drought, we build a **ARMA time series model** to predict the water level and put forward several plans to recycle the wastewater.

In problem 1, we first identify and describe the factors that impact inflow, outflow, and loss in Lake Mead. The factors include **evaporation capacity, precipitation, industrial, agricultural and living consumption, policy** and so on. Then we discuss the relationship of these factors and their relative influences on the volume and water level of Lake Mead. For example, the excessive industrial consumption aggravates the degree of drought. Next, we verify the relationships between the elevation, area, and volume based on the **non-linear regression** and calculus method.

In problem 2, we discuss the overall patterns in the historical data for Lake Mead water levels. Then we define the criteria for drought periods based on the comprehensive evaluation model named **Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS)**. Next, the **Comprehensive Drought Index (CDI)** of each year or month is calculated and we identify the beginnings and ends of periods of drought. When compared with the earlier drought period, the most recent one is the period in 2016 from 1-Jul to 18-Feb.

To predict the water level of Lake Mead, we build two models based on the time series analysis named as **exponential smoothing prediction model** and **Auto Regressive Moving Average Model (ARMA)**. In the model 1, we consider the data from only the most recent drought period, and in the model 2, we use the water level data from 2005 – 2020. The results of the two predict models reveal the water level in the years 2025, 2030, and 2050 is 1087.38, 1098.24, 1105,49 in model 1, and 1085.28, 1095.36, 1100.38 in model 2, respectively.

In problem 3, to address the impact on future water usage demands, we identify and describe the factors included in the plans to recycle wastewater. Then we build a comprehensive evaluation model based on the **Analytic Hierarchy Process (AHP)** to measure the impact of implementing our plan. At last, we write a one-page non-technical news article reporting the key takeaways and recommendations from our investigation.

Keywords: Drought of Lake Mead, TOPSIS, AHP, ARMA Model, Time Series Analysis

Content

TACKLING THE DROUGHT

RESEARCH ON THE DROUT OF LAKE MEAD AND SOLUTIONS

NOVEMBER 14, 2021.

RESEARCH ON THE DROUGHT

Perhaps the people who live nearby are sensing the dwindling water storage capacity of Lake Mead, located in Las Vegas, and the drought that is plaguing it.

Lake Mead is the **largest freshwater lake** in the United States, supporting the water needs of more than 25 million people. However, due to climate change and increasing water demand, historical data show that lake Mead has become **increasingly dry in recent years** and its water storage has continued to decline. In order to study the drought situation of Lake Mead, we made a **line chart and Heat Map** for visual processing and analysis of historical data, defined a comprehensive drought index (CDI), and divided the trend and cycle of drought through the index. According to the indicator, **Lake Mead experienced two severe drought periods: one between 1954 and 1957 and between 1964 and 1966. After 2000, both the highest and lowest elevation curves went down at a high rate. Although there was a slight rebound in 2012, it cannot change the trend of running out of the water of the Lake.**

So how can we predict future droughts in Lake Mead? Using two different models, we conclude that Lake Mead will grow more and more slowly over the next two decades -- much drier -- and will not be able to meet the needs of a growing population. **In the decade between 2025-2030, it grew by 0.9%, and in the two decades between 2030-2050, it grew by only 0.4%.**

SOLUTIONS!

In order to protect the U.S. freshwater supply and the environment around Lake Mead, what strategies could be used to address or mitigate this drought? **Wastewater recycling** could play a significant role. We know that there are many types of wastewater, which can be classified by toxicity or by how it is produced. According to the evaluation index established by us, the **proportion of each wastewater in the optimal wastewater discharge** is calculated through the analytic hierarchy process to achieve the optimization of the wastewater recovery plan. At the same time, we also consider that different leaders have **different preferences** for evaluation indicators, so we set up the preferences of three templates to provide more choices.

1. Introduction

1.1 Background

Being the largest water reservoir in the United States, Lake Mead in Colorado River has shrunk approximately 36 percent of its full capacity since 1930s and reaches its lowest level in the summer of 2021. This shrunk result from drought that is fueled by climate change and increasing demand for water. The Bureau of Reclamation published on August 16, 2021 announced the first water shortage declaration on the Colorado River, resulting in less water deliveries. Researchers are developing ways to recycle water more efficiently and effectively. One method is to recycle wastewater which it undergoes several treatments and is finally released into local waterways. Water from a certain source is treated in order to meet the quality level for a particular purpose. This technique is referred to as fit-for-purpose specifications.

1.2 Problems Restatment

1.2.1 Question 1

- Many factors directly or indirectly influence the amount of inflow, outflow, and loss in Lake Mead. Use geographical knowledge to analyze what factors directly influence the inflow, outflow, and loss.
- List factors that influence them. And describe why do they influence the volume of Lake Mead. To develop a model that reflects the relationship of the lake's elevation, area, and volume. Because the shape of the lake is very irregular and the depths are varied. We need to know which kind of fitting method should be used and the best function type.

1.2.2 Question 2

- After the data pre-processing, make statistical charts for the water level data. Find out the pattern. And set the criterion of drought based on the analysis.
- Model1: Develop a mathematical model according to geographical knowledge and data from the most recent drought. Then, use this model to predict the future.
- Model2: Develop a mathematical model according to geographical knowledge and data from 2005-2020. Then, use this model to predict the future.

1.2.3 Question 3

- Give a list of factors that affect the wastewater recycling system. Then, establish a comprehensive evaluation model to order the priority.
- Make a specific implementation plan for the wastewater recycling system. And give a feasible method to evaluate its effect.

1.2.4 Question 4

- We need to write a non-technical news article. It shall include key conclusions from the investigation and modelling, making it easy to understand.

1.3 Model Summary

To relieve the degree of drought, we build a **ARMA time series model** to predict the water level and put forward several plans to recycle the wastewater.

- We verify the relationships between the elevation, area, and volume based on the **non-linear regression** and calculus method.
- We define the criteria for drought periods based on the comprehensive evaluation model named **Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS)**.
- To predict the water level of Lake Mead, we build two models based on the time series analysis named as **exponential smoothing prediction model** and **Auto Regressive Moving Average Model (ARMA)**.
- To address the impact on future water usage demands, we identify and describe the factors included in the plans to recycle wastewater. Then we build a comprehensive evaluation model based on the **c** to measure the impact of implementing our plan.

The models developed in this paper can be summarized as shown in Figure 1.

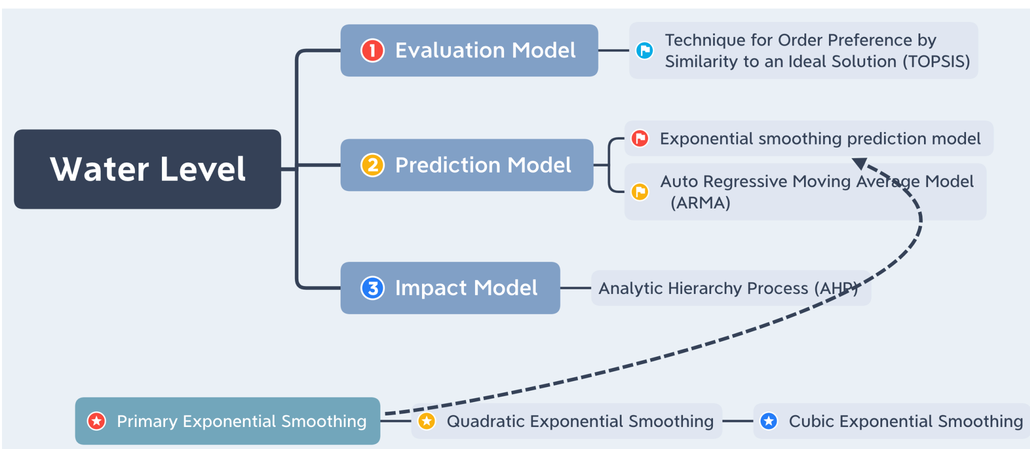

Figure 1. The models developed in this paper.

2. Assumptions and Justifications

- **Assumption 1**: We assume the data provided in the attachments is correct and reliable. The data of water level can reflect the degree of the drought in Lake Mead.

Justification: Because the data of water level in Lake Mead is collected and summarized from the authoritative websites, the data is reliable and convincing.

- **Assumption 2**: We assume the most recent drought period's pattern continues in the Model 1, when we consider data from only the most recent drought period.

  Justification: Because in the predict Model 1, the data comes from the most recent drought period, the most recent drought period's pattern will continue.

- **Assumption 3**: We assume the early drought period's pattern continues in the Model 2, when we consider data from early drought period during 2005-2020.

  Justification: Because in the predict Model 2, the data is collected during 2005-2020 from the early drought period, the drought period's pattern will continue.

- **Assumption 4**: We assume the influence of earthquake, volcano eruption, tornado and other extreme conditions on the Lake Mead can be neglected.

  Justification: Because during the past decades, the extreme conditions of weather barely happened, the influence of these extreme conditions has little impact on the water level in Lake Mead and can be neglected to some extent.

3. Variables Description

Variables	Description
C_i	Score of each year's water elevation
CDI_i	Comprehensive Drought Index
R^2	R^2 refers to coefficient of determination
$X = (x_{ij})_{n\times m}$	The evaluation matrix
$S_t^{(1)}$	The smoothed statistic in primary exponential smoothing
$S_t^{(2)}$	The smoothed statistic of quadratic exponential smoothing
$S_t^{(3)}$	The smoothed statistic of cubic exponential smoothing

4. Factors Impact Model

4.1 Selecting the Factors

In Figure 2 below, the mind map demonstrates the over all factors that we considered. However, we have chose some of them(as highlighted) to be the ultimate factors after analysis. As shown in the Table 1, for inflow, outflow and loss of the Lake

Mead, we have considered 3 aspects for each and have determined the factors that affect each aspects.

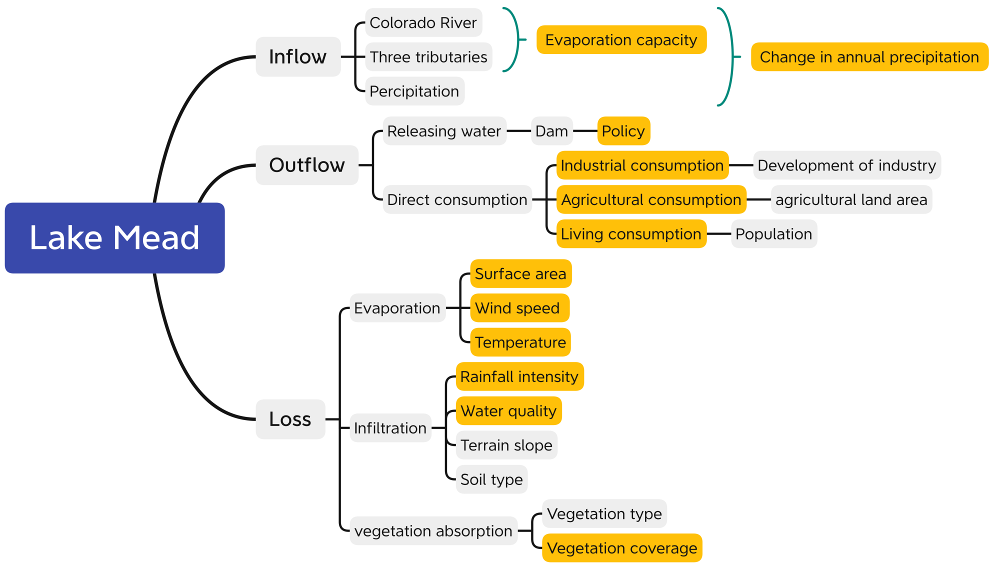

Figure 2. Factor Determining Mind Map.

- **Inflow:** For inflow, we consider three aspects: inflow from the Colorado River, inflow from the three tributaries, and precipitation. Since stream flow is mainly influenced by precipitation and stream evaporation rate, we finally determined the factors as evaporation capacity and precipitation.
- **Outflow:** For outflow we have three aspects: Releasing water, direct consumption and outflow rivers. The releasing water and direct consumption are human aspects, while the outflow of the rivers is natural aspects. Among them, the releasing water depends largely on the discharge of water from the dam, which determined by the government's policy or strategy. Direct consumption can be affects by the factors of industrial consumption, agricultural consumption, and living consumption.
- **Loss:** The loss of lake water is intricate, however, it mainly consists of evaporation, infiltration and vegetation absorption. Evaporation from lakes depends on three main factors: temperature, wind and surface area. Infiltration refers to that under the action of molecular force, capillary force and gravity, water enters into the soil pores and is absorbed by the soil to supplement it [1]. Mostly, water in the soil gathers to form a subsurface runoff, which is continuously replenished by lake water. The infiltration is determined by the factors: precipitation intensity and water quality. Last, the vegetation around will also absorb the water from the lake. And the vegetation absorption is affected by the factor vegetation coverage.

Table 1. Established selections for factors affects inflow, outflow and loss.

	Aspect	Factor
Inflow	Colorado River Three tributaries Precipitation	Evaporation capacity Precipitation
Outflow	Releasing water Direct consumption Outflow rivers	Industrial consumption Agricultural consumption Living consumption Policy
Loss	Evaporation Infiltration Vegetation absorption	Surface area Wind speed Temperature Precipitation intensity Water quality Vegetation coverage

4.2 Relationship of Factors

The outflow of Lake Mead is highly related to human activities, including industrial consumption, agricultural consumption, living consumption, and policies. These factors are influenced by population and economic developments, and are relatively independent of natural factors.

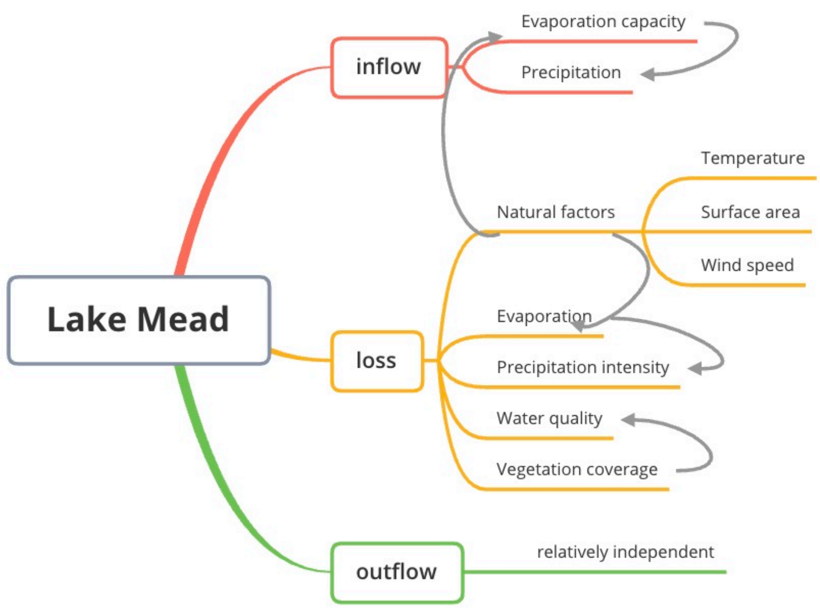

Figure 3. The factors related to the inflow, outflow and loss we considered.

Some aspects that contribute to the inflow and outflow of Lake Mead are similarly influenced by the same factor. The evaporation capacity that contributes to inflow and evaporation loss are similarly determined by temperature, surface area, and wind speed. The higher the temperature, surface area, and wind speed, the higher the evaporation rates, and thus, the higher the inflow and loss.

Factors within one category can also affect each other. Firstly, evaporation positively influence precipitation. Secondly, the better the quantity and quality of vegetation coverage, the better the water quality.

4.3 Relative Influences

We discuss the relationship between the factors concerning the inflow, outflow and loss of Lake Mead with water volume and relative influences, then we summarize them into the Table 2 as following.

Table 2. The relationship with water volume and relative influences analyses.

	Factor	**Relationship with water volume**
Inflow	Evaporation capacity Precipitation	• When the evaporation capacity of the Colorado River and the three tributaries rises, the precipitation that contributes to the inflow of Lake Mead rises, resulting in an increase in water volume.
Outflow	Industrial consumption Agricultural consumption Living consumption Policy	• When the direct consumption rises, outflow increases and water volume decreases. • When government policies change, leading to less discharge of water(or more), the water volume will increase generally(or decrease).

Loss	Surface area Wind speed Temperature Water quality Precipitation intensity Vegetation coverage	• The larger the surface area, wind speed and temperature, the higher the evaporation. Hence, the lower the water volume. • The lesser the vegetation coverage, the higher the water volume. • The higher the precipitation intensity, the higher infiltration will be. Thus, the lower the water volume. As for water quality, the higher the turbidity degree or the lower the salinity degree, the more infiltration there will be.

4.4 Calculus Approach Model

Next, we verify the relationships between the elevation, area, and volume based on the **non-linear regression** and calculus method.

Since Lake Mead has irregular shape, the cross-sectional area of each height of the lake and the shape of the cross-sectional area of the lake cannot be determined. Thus we can approximate Lake Mead as an irregular non-rotating body and calculate its volume using the element method. The central idea of the infinitesimal method. is to calculate the volume of a non-rotating body by calculating the definite integral of the cross section area perpendicular to an axis. Let A(y) be the cross section area of the crossing point y perpendicular to the Y-axis, and A(y) is a continuous function of y. Here, the altitude of the lake is equal to Y, and A(y) is a function of the cross-sectional area of the lake. Thus, the calculation formula of lake volume can be obtained:

$$V=\int_{a}^{b} A(y)\,dy$$

If we get additional information, such as lake cross section configuration and lake cross section area function, then we can approximate the lake volume as a rotating body and calculate the volume of the rotating body using the infinitesimal method.

Here, we provide three examples to calculate Lake Mead's volume:

- **Cone:** $$V=\pi\int_{0}^{h}\left(\frac{Ry}{h}\right)^{2}dy$$
- **Semi ellipsoid**: $$V=\pi\int_{0}^{h}\left(r+\left(\frac{R-r}{h}\right)y\right)^{2}dy$$
- **Truncated cone:** $$V=\frac{\pi\int_{-a}^{a}\frac{b^2}{a^2}\sqrt{a^2-x^2}\,dx}{2}$$

4.5 Non-linear Regression

4.5.1 Data Analysis

In order to explore the relationship between the elevation, area, and volume of Lake Mead, we first plot the curves between them and list the curves in Figure 4.

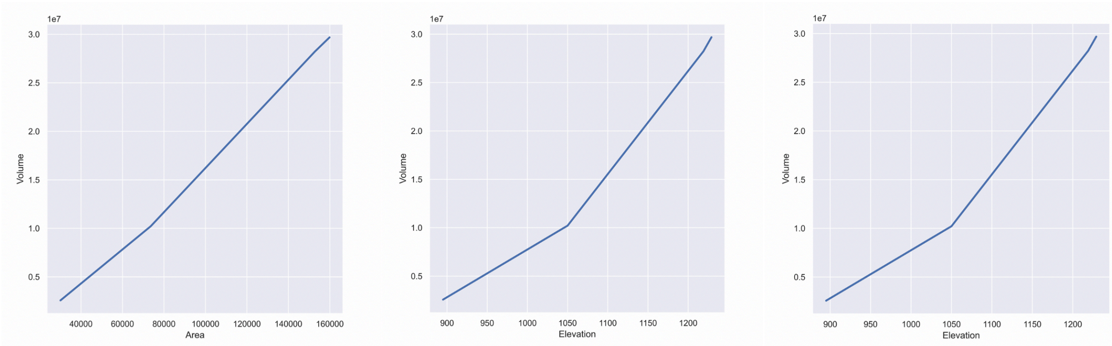

Figure 4. The curves between the elevation, area, and volume of Lake Mead.

As shown in Figure 1, it's clear that there is linear or quadratic relationship between the elevation, area, and volume of Lake Mead in the function of V=F(E, A). The volume of Lake Mead increases with the increase of elevation level and area.

4.5.2 Determine the Relationship

Next, in order to obtain the quantitative function of the elevation, area, and volume, we verify their relationships based on the non-linear regression method. The original data of the elevation, area, and volume of Lake Mead is shown in Table 3.

Table 3. Area and Volume of Lake Mead by Elevation Level.

Elevation (feet)	Area of Lake (acres)	Volume of Lake (acre-feet)
1229.0	159,866	29,686,054
1219.6	152,828	28,229,730
1050.0	73,615	10,217,399
895.0	30,084	2,576,395

Multi-Variables regression:

By observing the line chart, we found that the independent variable volume is correlated with the area and altitude, so we selected the function type $a+bx+cy$ and $a+bx^2+cx+dy^2+ey$ to do the regression by MATLAB software. The result:

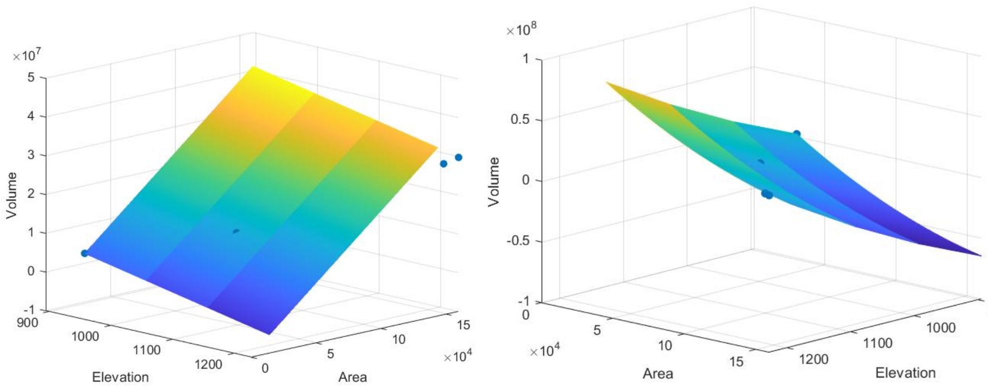

Figure 5. The surface of the three variables based on the linear and nonlinear methods.

Table 4. The results of the linear and nonlinear regression methods.

Regression Type:	Linear	Non-Linear
Function Type:	a+bx+cy	a+bx^2+cx+dy^2+ey
Best-Fit Result:	a=2.3991*1.0e+07 b=-0.0034*1.0e+07 c=0.0000*1.0e+07	a =0.0000*1.0e+05 b= 0.0020*1.0e+05 c=-1.5080*1.0e+05 d= 0.0000*1.0e+05 e=-0.0092 *1.0e+05
R of residuals:	-1.4881*1.0e+05 1.6953*1.0e+05 -0.3437*1.0e+05 0.1365*1.0e+05	0.8941*1.0e-07 0.5960*1.0e-07 0.2980*1.0e-07 0.0373*1.0e-07
R^2	0.0000	1.0000

Where R^2 refers to coefficient of determination. We can see in the linear regression, the coefficient of determination R^2 equals to 0.000, which means the explanatory variables can represent 0% of the response variables. It means the effect of linear regression is bad. The value of R^2 in the non-linear regression approximately equals to 1. It means the explanatory variables can represent approximately 100% of the response variables. It is a good regression result. However, because of the small scale of training set, and the extremely high R^2, the result may be overfitting and lack representativeness. We may have a better result if there are more available data.

5. Overall Patterns Analysis

5.1 Historical Data Analysis

5.1.1 The impoundment period

As shown in Figure 6, in the impoundment period, the highest and lowest elevation level both increased with a constant speed. The monthly elevation increased with a constant rate as well.

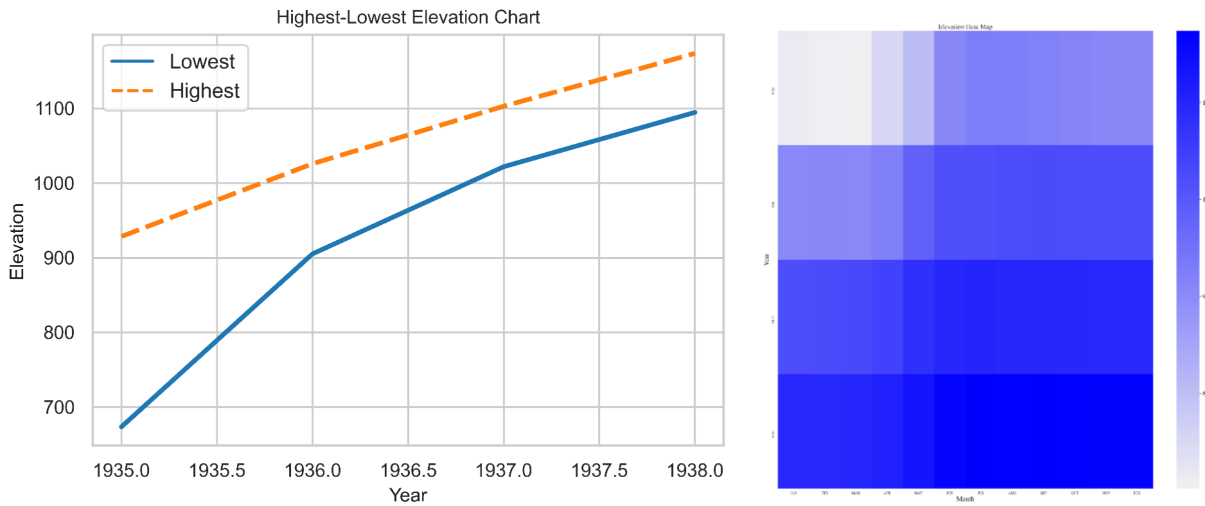

Figure 6. The line plot of the (a) highest and lowest water level during the impoundment period; (b) heatmap of the water level during the impoundment period.

5.1.2 Recognizing the Overall Patterns

After the impoundment period, the highest and lowest elevation curve comparably fluctuates more drastically. Geographically, it is caused by the artificial change of the characteristic hydrological regime and the unchanged climate regime.

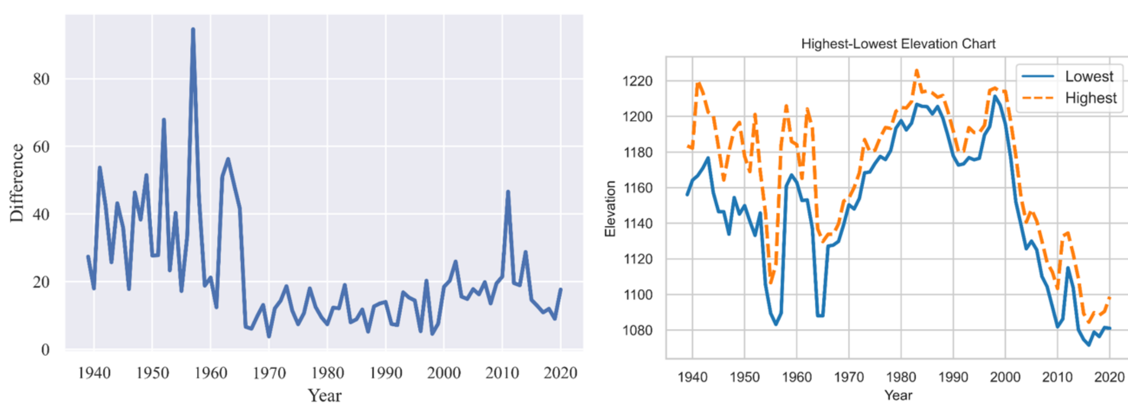

Figure 7. The line plot of the (a) difference of water level in the adjacent year; (b) highest and lowest water level during different years.

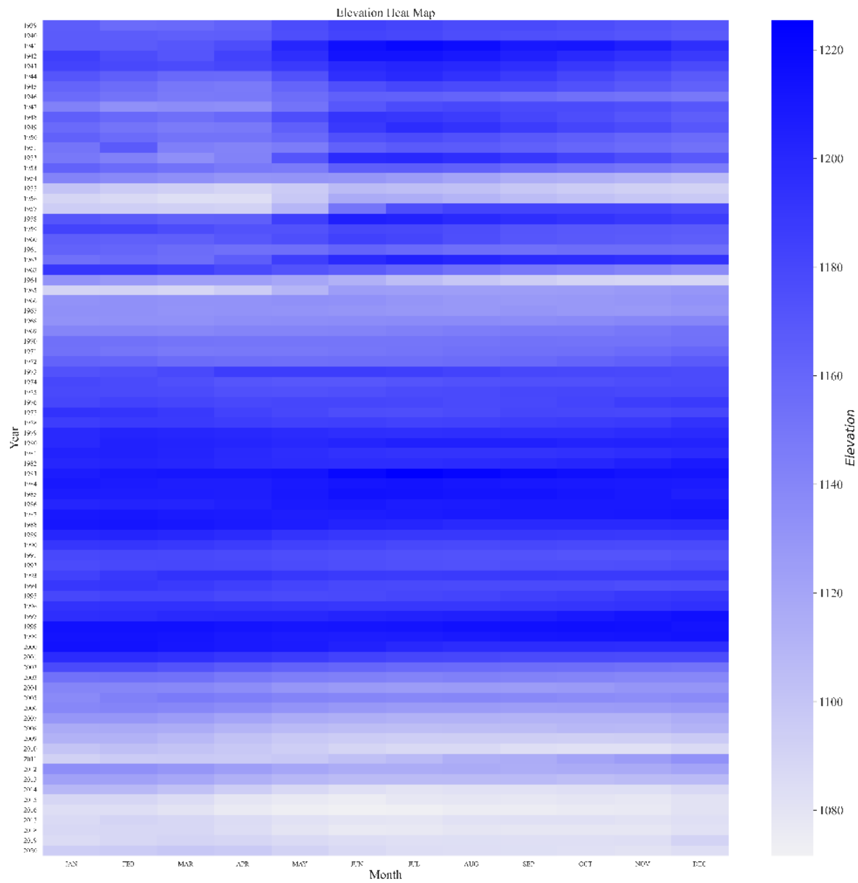

Figure 8. The heatmap of the water level among different years and months.

As shown in Figure 7 and 8, it experienced two severe drought periods: one between 1954 and 1957 and between 1964 and 1966. After 2000, both the highest and lowest elevation curves go down with a high rate. Although there was a slight rebound in 2012, it cannot change the trend of running out of the water of the Lake Mead.

5.2 TOPSIS Evaluation Model

5.2.1 Building the TOPSIS Model

In 1981, a paper named "Methods for Multiple Attribute Decision Making" proposed a new common intragroup comprehensive evaluation method: Technique for Order Preference by Similarity to an Ideal Solution, known as TOPSIS. Its authors, Ching-Lai Hwang and Kwangsun Yoon utilized the this objective evaluation method [2].

The basic process is based on the normalization of the original data matrix, the cosine method is used to find the best and the worst scheme in the finite scheme, and then calculate the distance between each evaluation object and the best alternative and the worst alternative respectively, to obtain the relative proximity of each evaluation object and the optimal scheme, which is used as the basis for evaluation. This method has no strict limitation on data distribution and sample size, and data calculation is simple.

Step 1. Create the Evaluation Matrix

In this method, we first build up an evaluation matrix $X = \left(x_{ij}\right)_{n\times m}$:

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

Where n refers to n alternatives, m refers to m criteria, all values have been regularized.

Step 2. Data Normalization

Normalize the matrix $\left(x_{ij}\right)_{n\times m}$ to form a new matrix $Z = \left(z_{ij}\right)_{n\times m}$, using the the normalization method:

$$z_{ij}=\frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{ij}^2}}$$

Step 3. Determine the worst alternative(Z^-) and the best alternative (Z^+):

$$Z^+=(\max\{z_{11},z_{21},...,z_{n1}\},\max\{z_{12},z_{22},...z_{n2}\},...,\max\{z_{1m},z_{2m},...,z_{nm}\})=(Z_1^+,Z_2^+,...,Z_m^+)$$

$$Z^-=(\min\{z_{11},z_{21},...,z_{n1}\},\min\{z_{12},z_{22},...z_{n2}\},...,\min\{z_{1m},z_{2m},...,z_{nm}\})=(Z_1^-,Z_2^-,...,Z_m^-)$$

Step 4. Calculate the L^2-distance between the target alternative i and the best condition.

$$D_i^+=\sqrt{\sum_{j=1}^{m}(Z_j^+-z_{ij})^2}$$

And the L^2-distance between the target alternative i and the worst condition Z^-:

$$D_i^-=\sqrt{\sum_{j=1}^{m}(Z_j^--z_{ij})^2}$$

Step 5. Calculate the scoring according to the distance to the best and worst conditions:

$$C_i=\frac{D_i^-}{D_i^+ + D_i^-}$$

$$0\leqslant C_i\leqslant 1,\ \textit{larger } C_i \textit{ refers to higher score}$$

Step 6. According to the order of C_i to give the evaluation result.

5.2.2 Comprehensive Drought Index (CDI)

The **Comprehensive Drought Index (CDI)** of each year or month can be calculated as the following formula based on the sum of the weighted factors.

$$CDI_i = \frac{1}{C_i}$$

Where the weight of each factor C_i is the scoring according to the distance to the best and worst conditions of drought, and the F_i is the factor we select for evaluating the degree of drought including the lowest water level, highest water level, average water level, difference of the water level per year or month.

5.3 Criteria for Drought Periods

We define the criteria for the drought periods as when the **Comprehensive Drought Index (CDI)** of each year or month is less than the threshold of drought λ, this condition can be regarded as the period of drought.

$$CDI_i = \frac{1}{C_i} > \lambda$$

Where λ is the threshold of drought, and the CDI_i is the Comprehensive Drought Index (CDI) of each year or month.

5.4 Identify the Lasting Periods of Drought

In this part, we define the threshold of drought λ as 1.60, and then we calculate the Comprehensive Drought Index (CDI) of each year or month as shown in Table 5.

Table 5. The Top twenty Comprehensive Drought Index (CDI) for drought period.

Rank	Year	CDI	Begin	End	Rank	Year	CDI	Begin	End
1	2016	inf	1-Jul	18-Feb	11	1956	5.55	26-Apr	9-Jul
2	2015	39.42	26-Jun	22-Jan	12	2008	4.33	31-Jul	11-Mar
3	2018	29.08	11-Jul	12-Mar	13	2011	4.23	2-Jan	31-Dec
4	2017	22.10	2-Aug	1-Mar	14	2013	4.17	12-Nov	5-Feb
5	2019	15.14	1-Jan	31-Dec	15	1964	4.16	30-Dec	1-Jan
6	2020	11.20	26-Nov	30-Mar	16	1965	3.58	3-Apr	31-Dec
7	2014	9.68	13-Aug	2-Feb	17	2007	3.48	4-Oct	14-Feb
8	2010	9.29	27-Nov	26-Feb	18	2012	3.15	21-Nov	22-Jan
9	1955	6.64	22-Apr	7-Jul	19	1954	2.91	31-Dec	1-Jan
10	2009	6.30	6-Nov	25-Jan	20	1967	2.62	22-Nov	13-Feb

5.5 Comparison with Early Drought

Compared to earlier years, current years has relatively higher CDI, representing a higher drought degree. Moreover, the time during which the drought occurs shows a pattern: the starting time changes from June to January, and the ending time changes from January to December.

6. Water Level Prediction Model

6.1 Model 1: Exponential Smoothing of Time Series

6.1.1 Primary Exponential Smoothing

Assume time series to be $y_1, y_2, \cdots, y_t, \cdots$, α is the smoothing parameter, $0 < \alpha < 1$, the single exponential smoothing formula is given by:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha\left(y_t - S_{t-1}^{(1)}\right)$$

Which is modified from the moving Average formula. The formula of deriving the rolling average is:

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}$$

We use $M_{t-1}^{(1)}$ as the best estimate of y_{t-N}, there is

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - M_{t-1}^{(1)}}{N} = \frac{y_t}{N} + \left(1 - \frac{1}{N}\right) M_{t-1}^{(1)}$$

Let $\alpha = \frac{1}{N}$..., use S_t to replace $M_t^{(1)}$, we get:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)}$$

In order to understand the essence of exponential smoothing, we expand and get

$$S_t^{(1)} = \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha) S_{t-2}^{(1)}\right] = \cdots = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j \, y_{t-j}$$

Which shows that $S_t^{(1)}$is the weighted mean of all historical data, obviously, we get

$$\sum_{j=0}^{\infty} \alpha(1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1$$

Because weighing parameter follows the exponential law, and also applies to smooth data, it is known as exponential smoothing. Predictions based on this smooth value is Primary Exponential Smoothing. The prediction model is:

$$\hat{y}_{t+1} = S_t^{(1)}$$

Which can be also written as:

$$\hat{y}_{t=1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

which is using the t time period to predict t+1.

6.1.2 Quadratic Exponential Smoothing

The Primary exponential smoothing method overcomes the disadvantages of the moving average method. However, when the change in the time series shows a linear trend, the primary exponential smoothing method still shows obvious hysteresis error.

Therefore, it must also be amended. The correction method is the same as what to do with the trend movement average method, which contains another quadratic exponential smoothing, and uses the law of the hysteresis error to establish a linear trend model. This is called the quadratic exponential smoothing method. The calculation formula is:

$$S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)}$$

In the formula, $S_t^{(1)}$ is the smoothed statistic in primary exponential smoothing, $S_t^{(2)}$ is the smoothed statistic of quadratic exponential smoothing:

$$\hat{y}_{t+T} = \alpha_t + b_t T, T = 1,2,\cdots$$

$$\begin{cases} \alpha_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right) \end{cases}$$

When time series $\{y_t\}$ starts to show a linear trend at one time period, just like Moving Average Method, can build the linear trend model to predict the water level.

6.1.3 Cubic Exponential Smoothing

When the changes in the time series shows a quadratici curve trend, the cubic exponential smoothing is required. The cubic exponential smoothing is based on the secondary exponential smoothing, with the formula of

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \end{cases}$$

Where $S_t^{(3)}$ is the smoothed statistic of cubic exponential smoothing.

The cubic exponential smoothing prediction model is

$$\hat{y}_{t+T} = \alpha_t + b_t T + C_t T^2, T = 1,2,\cdots$$

Consequently, the Cubic Exponential Smoothing can be represented as:

$$\begin{cases} \alpha_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}\left[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}\right] \\ c_t = \dfrac{\alpha}{2(1-\alpha)^2}\left(S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right) \end{cases}$$

Where $S_t^{(1)}$is the smoothed statistic in primary exponential smoothing, $S_t^{(2)}$ is the smoothed statistic of quadratic exponential smoothing, and $S_t^{(3)}$ is the smoothed statistic of cubic exponential smoothing.

6.1.4 Results of Prediction Model

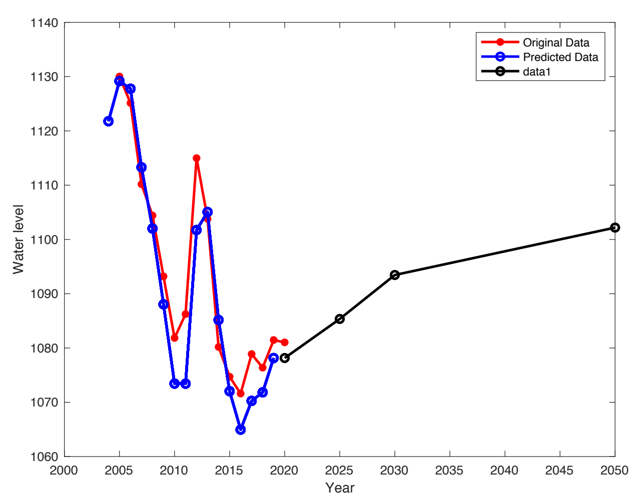

Figure 9. The predicted results of exponential smoothing of time series.

6.2 Model 2: ARMA Model with the Periods Correction

6.2.1 Building the ARMA Model

The Auto Regressive Model is abbreviated as the AR model. The Moving Average Model is abbreviated as the MA model. The Auto Regressive Moving Average Model is abbreviated as the ARMA model [4]. X_t is a zero mean stationary sequence.

(1) Auto Regressive Model $AR(n)$

Assume time series X_t has a linear relationship with $X_{t-1}, X_{t-2}, \cdots, X_{t-n}$ only. Under the condition of known $X_{t-1}, X_{t-2}, \cdots, X_{t-n}$, $X_t$is irrelevant with $X_{t-j}(j = n+1, n+2, \cdots)$, α_t is a white noise sequence independent of $X_{t-1}, X_{t-2}, \cdots, X_{t-n}$, $\alpha_t \sim N(0, \sigma_\alpha^2)$

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_n X_{t-n} + \alpha_t$$

Above formula can also be expressed as

$$\alpha_t = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_n X_{t-n}$$

It could be seen that the response of $AR(n)$ system X_t has n order dynamics. $AR(n)$ model eliminates $X_{t-1}, X_{t-2}, \cdots, X_{t-n}$ that are dependent of X_t and convert X_t into independent sequence α_t. Thus, the process of fitting $AR(n)$ model is the independence of relevant sequence.

(2) Moving Average Model $MA(m)$

$AR(n)$'s response X_t at time *t* is only relevant to former responses $X_{t-1}, X_{t-2}, \cdots, X_{t-n}$, but is irrelevant from the previous disturbance that enter the system. If a system's response X_t at time *t* is irrelevant with responses $X_{t-1}, X_{t-2}, \cdots$, at time $t-1, t-2, \cdots$, but is related to $t-1, t-2, \cdots, t-m$ the disturbance that enter the system $\alpha_{t-1}, \alpha_{t-2}, \cdots, \alpha_{t-m}$. Then, this system is called the $MA(m)$ system.

$$X_t = \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \cdots - \theta_m \alpha_{t-m}$$

(3) Auto Regressive Moving Average Model

In a system, if X_t in time *t* is not only related to its previous value, but also related to the disturbance that enter the system before, then, this system is an Auto Regressive Moving Average System. $ARMA(n, m)$ model is

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_n X_{t-n} = \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_m \alpha_{t-m}$$

Because AR、MA、$ARMA(n, m)$ are extreme cases of $ARMA(n, n-1)$, we use $ARMA(n, n-1)$ as general form to build our model.

6.2.2 Results of Prediction Model

Table 6. The predicted results of water level based on the ARMA model.

Year	2025	2030	2050
Predicted Value	1085.28	1095.36	1100.38

7. Comprehensive Evaluation Model

7.1 Selection of the Factors

In order to propose the plan of recycling the wastewater and measure the impact of the decision made by the local leaders, we select the factors as shown in Figure 10.

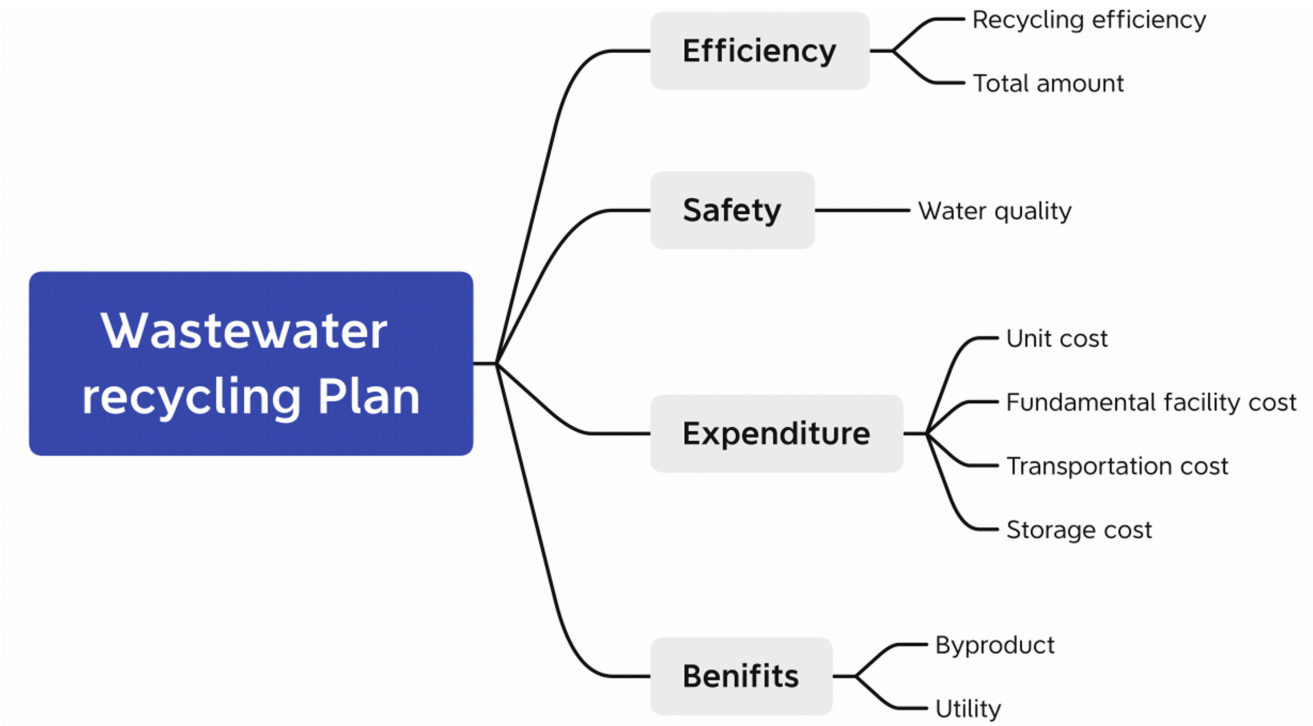

Figure 10. The factors we considered in the plan of recycling the wastewater.

7.2 Analytic Hierarchy Process (AHP)

7.2.1 Construct the Hierarchy Structure

After determining the influencing factors in 7.1, we need to determine the influence of each indicator on the wastewater recovery plan by determining the weight of each indicator. The Analytic Hierarchy Process (AHP), as a systematic hierarchical analysis method combining qualitative and quantitative analysis [3], is very suitable for determining the weight to decide the score for recycle distribution of each type of wastewater. First, we construct the hierarchy structure for our model as shown below.

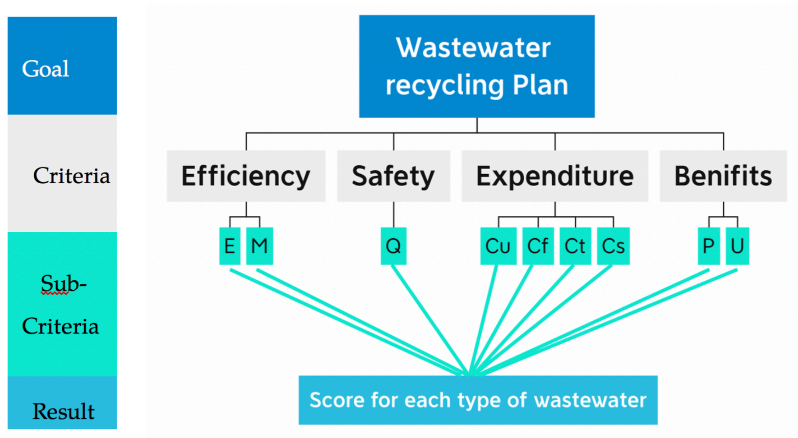

Figure 11. The hierarchy structure of the AHP model.

The factors we selected are listed in Table 7 as follows.

Table 7. The variable definition of the AHP model.

Symbol	Definition
E	The amount of wastewater that can be treat in unit time
M	Total amount of wastewater
Q	The water quality scale after recycled
C_u	Cost of recycling per unit amount of wastewater
C_f	Cost of fundamental facilities needed to recycle wastewater
C_t	Cost of water transportation
C_s	Cost of water storage
P	The profit made by the byproduct during the recycle process
U	Utility scale of the recycled water

7.2.2 Construct the Comparison Matrix

We used the 1-9 scored table to compare the importance of factors, so as to get the weight of each factor. When you compare two factors, you get a comparison matrix.

Table 8. The 1-9 scored table of the AHP model.

Scale	Definition
1	factor i is equally important as factor j
3	factor i is slightly more important than factor j
5	factor i is apparently more important than factor j
7	factor i is strongly more important than factor j
9	factor i is extremely more important than factor j

2,4,6,8	the intermediate values
Reciprocal	important scale between factor j and factor i

Each element Aij in the comparison matrix represents the ratio of importance of Ai to Aj. Namely, the following form:

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

However, the local government leader, which is the project implementer, may have personal preferences and priorities for these factors. These will affect the actual weight of each factor ultimately. In order to make the plan as close as possible to the actual needs, we pre-designed the weight distribution of three different priorities (namely, the comparison matrix).

- **Leader1:** This leader is more conservative in the implementation of the plan, and he puts costs and risks more important.
- **Leader2**: The leader is more ambitious and optimistic, arguing that the benefits of wastewater recycling are more important.
- **Leader3:** The leader is average, and he thinks costs are as important as benefits.

7.2.3 Checking the Data Consistency

In order to ensure the validity and rationality of our model, we check the consistency of the comparison matrix that we constructed by calculating the CI with the lambda max (greatest eigenvalue) ,which can be expressed as followed:

$$CI=\frac{\lambda_{\max}-n}{n-1}$$

The smaller the value of CI is, the better the consistency is, that is, the comparative matrix is reasonably constructed. Subsequently, we use the CI and RI to calculate the CR using the following formula:

$$CR=\frac{CI}{RI}$$

Where CR is the index to check the inconsistency of the matrix. CR<0.1indicates that the inconsistency degree of matrix is within the allowable range, and its eigenvectors can be used as weight vectors.

Table 9. The random consistency indexes.

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

The CR is 0.06 for A1, 0.08 for A2, and 0.05 for A3, which indicate that three comparison matrix are reasonable.

7.2.4 Weight of Each Index

Table 10. The weight of each index in the AHP model.

Indexes	Efficiency	Safety	Expenditure	Benefits
Weights	0.258	0.321	0.185	0.236

7.3 Impact on the Plan

According to the preferences and priorities made by the local leaders, we construct the comparison matrix and obtain the weight of each index in the AHP model. Therefore, the composition of the plan for recycling wastewater can be written as follows:

$$Plan = 0.258EF + 0.321S + 0.185EX + 0.236B$$

Which indicates that we should implement our plan considering the order of priorities as following:

Safety > Efficiency > Benefits > Expenditure

Moreover, 32.10% of the money or attention should be focused on the Safety, 25.80% for Efficiency, 23.60% for Benefits and 18.50% for Expenditure.

8. Conclusion

In conclusion, to relieve the degree of drought, we build a ARMA time series model to predict the water level and put forward several plans to recycle the wastewater. We first identify and describe the factors that impact inflow, outflow, and loss in Lake Mead. The factors include evaporation capacity, precipitation, industrial, agricultural and living consumption, policy and so on. Then we discuss the overall patterns in the historical data. Then we define the criteria for drought periods based on the comprehensive evaluation model named TOPSIS. Next, we identify the beginnings and ends of periods of drought. To predict the water level of Lake Mead, we build two models based on the time series analysis named as exponential smoothing prediction model and Auto Regressive Moving Average Model (ARMA). Then we build a comprehensive evaluation model based on the Analytic Hierarchy Process (AHP) to measure the impact of implementing our plan.

9. Advantages and Disadvantages

9.1 Advantages

- To verify the relationships between the elevation, area, and volume, we apply two methods including the non-linear regression and calculus method. The non-linear regression method can determine the quantitative relationship.
- In the process of data analysis of the historical data, we plot several figures to illustrate the tendency of the elevation. It's beneficial for us to analyze the overall patterns of the drought from the data visualization. To determine the drought and identify the beginnings and ends of drought, we develop the TOPSIS model.
- In order to predict the water level of Lake Mead, we construct two models based on the time series analysis named as exponential smoothing and Auto Regressive Moving Average Model (ARMA), which can predict the water level accurately.
- To address the impact on future water usage demands, we identify and describe the factors included in the plans to recycle wastewater. Then we build a comprehensive evaluation model based on the Analytic Hierarchy Process (AHP) to measure the impact of implementing our plan, which is proper and efficient.

9.2 Disadvantages

- In the process of predicting the water level of Lake Mead, we consider the data from only the most recent drought period, and the water level data from 2005 – 2020. If we can collect more data of the drought period, it will more accurate to predict the water level in the year of 2025, 2030, and 2050.
- We assume that the influence of earthquake, volcano eruption, tornado and other extreme conditions on the Lake Mead can be neglected. Therefore, if we consider the influence of extreme conditions the Lake Mead, the overall patterns of drought could be clearer and more comprehensive.

10. References

[1] Holdren, G. C., & Turner, K. (2010). Characteristics of Lake Mead, Arizona–Nevada. Lake and Reservoir Management, 26(4), 230–239.

[2] Hwang, C. L., & Yoon, K. (1981). Methods for Multiple Attribute Decision Making. Multiple Attribute Decision Making, 58–191.

[3] Saaty, T. L. (1988). What is the Analytic Hierarchy Process? Mathematical Models for Decision Support, 109–121. https://doi.org/10.1007/978-3-642-83555-1_5

[4] Makridakis, S., & Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. Journal of forecasting, 16(3), 147-163.