Team Control Number

11823

Problem Chosen

B

2021
HiMCM
Summary Sheet

With the increasingly intensified global warming and the burgeoning population, the water resources around the world have been significantly affected. Lake Mead, which provides water supply for 25 million people, reached its historically low elevation during the summer of 2021. Therefore, we explored the impact of drought on Lake Mead, constructed a model to predict the change of its water level, and used wastewater recycling as a solution to address water shortages.

Firstly, we made a qualitative analysis on factors that impact the inflow, outflow, and loss of water, thus getting a brief understanding of the water flow cycle. We weighed out factors that might influence Lake Mead more than others and verified the **relationship between elevation, surface area and volume** by the volume formula of frustum of the cone.

Afterwards, we focused on the analysis of drought and the forecast of the water level. We first employed several past data of hypothetical volume at controlled elevation to create a fitting polynomial trend line, which illustrates **the relationship between elevation and volume**. We made a data cleansing to eliminate outliers of volume and calculated the average capacity over years. The percentage error formula was used to define the drought index by the current water capacity and average capacity. By using our drought standards, we concluded that **there were three droughts** since 1935, and the 2004 onward drought is the most severe one by comparing its overall trend, capacity trough, and duration with other two droughts.

After noticing the current drought period has not ended as it hits a historical low this year, we used **Time Series Analysis** to explore its future trend. Exploring the data at the same period over years with given elevations, we evaluated the change of monthly trend by calculating monthly difference. Finding out all elevation data passed the stability test and white noise sequence evaluation, **ARMA** was used to predict the future trend of elevation. In order to verify the accuracy of model, we employed data from 1935-2010 to predict 2011-2021s elevation. Two predicted trend line are made, one used most recent drought period data and the other used data from 2005-2020.

Based on our analysis, the elevation of Lake Mead is going to keep dropping, and cities need to take immediate action. A plan to recycle wastewater has been proposed. It takes manifold factors into account. To solve the water shortage and ensure the quality of recycling water, different **optimization models** are established according to the demand of different cities. Two essential indexes are percentage filling up the water gap and circulating water quality index. As a city strives to optimize one indicator, the other one also need to ensure. Lastly, a sensitivity test is conducted, the 3D graph reveals that the wastewater recycling program is a bit sensitive to **wastewater recycling rate**, a factor that needs to be dealt with cautiously when implementing the plan.

Keywords: Polynomial fitting, Time Series Analysis, Stability Test, ARMA, Optimization Model

DROUGHTS

By Lili Pike, Apr 21, 2021, 8:00am EDT, on VOX

OF LAKE MEAD

I We studied the past data of Lake Mead and defined the criteria for dry periods. We first managed to find the volume of water from its elevation by identifying the relationship between the two. At start, we only had water elevation information over the years. However, to clearly understand each period's drought data, the water volume is more intuitive and accurate than the elevation. For this reason, we also used another set of data from 1935 and 2010, which contains and compares the information of elevation, area, and volume in the same period. We used this information to plot a graph and generated a polynomial best-fit line. From the polynomial expression, as long as we have either the elevation or volume information, we can know the value of the other one by converting it.

Then, we standardized the drought index and noticed the three main droughts. After the conversion, we used the volume and made a graph with colors marking the severeness of drought. As shown on the graph, over the three drought periods, the severeness of drought is increasing since the trough is getting lower and last for longer time.

II We forecasted the future trend of the lake level rise and found that the drought in the future is very serious and plans need to be implied. A time-series model was built using the monthly elevation data, from what we observed, the drought is most likely to become more severe in the short term future, but might be restored in the long term. Before using the model, the data's stability was examined using two methods, this is used to determine the suitability of the time-series model's application.

III We formulated two different types of wastewater recycling formulas according to different city types which can effectively alleviate the water shortage crisis. The formulas has the types of wastewater split into three different types because it is very difficult to gain accurate results for wastewater as a whole, so wastewater has been divided into three different types: domestic water, agricultural water, and industrial water. These are the three main types of wastewater and it would be able to cover almost all of the wastewater that can be produced by a city. The first recycling formula is to solve the water shortage within the city and the second is to consider the water quality of the recycled water. Solving the water shortage would need a optimization method to calculate the water gap that exists in the city using the factors of the lake water volume, lake minimum water volume, proportion of water flowing into the lake and the total water demand to form a formula. The formula to indicate the points of the water quality of the water that is recycled uses the different types of water. With the two programs, there is an optimization method which would allow the city's wastewater to be managed.

IV Even though there is the wastewater recycling program mentioned above, we still need to put an end to water waste to protect the environment. People around the world, not only the areas and cities around Lake Mead, needs to start to considering the problems that wasting water would bring. Not only the waste of water but also polluting the water. Even though these water could be recycled to an extent, the process of recycling the water would use energy and harm the environment in another way. The fastest and not environmental friendly was to, not end the problem because it is very hard to end the problem, but to mitigate the issue is from the start, the wasting of water. There should be more advocating of the issue in order to more people to take notice of the problem of the waste of water.

Steps Towards Solution

BRIEF SUMMARY ⬇

ANALYSIS
OF PAST DATA

FORECAST
FUTURE TREND

FORMULATE
FORMULA

END
BY PROTECTION

Contents

1 Introduction

1.1 Background

Lake Mead is the largest water reservoir by volume in the United States that lies on the main stem of the Colorado River. It is located approximately 30 miles east of Las Vegas by the Nevada-Arizona border. It has enough capacity to hold the entire annual flow of the Colorado River for two years. When it is at a full capacity, Lake Mead extends from Hoover Dam to Pearce Ferry over 65 miles, with the greatest width of 9.3 miles and shoreline of 550 miles in length[12]. Lake Mead has four connected underground basins that cover canyons along the historical Colorado River channels.

The Colorado River provides approximately 97 percent of the inflow to Lake Mead[12], the remaining inflow comes from Las Vegas Wash, the Virgin and the Muddy Rivers. The flow from Las Vegas has doubled over the past 30 years caused by the rapid growth of the population in Las Vegas.

The construction of the Hoover Dam was started in 1931, done in 1936 and started to impound water by 1935. The Hoover Dam successfully controlled the water and formed Lake Mead. Currently, it provides municipal needs of water for the cities of Las Vegas, Henderson, North Las Vegas, and Boulder City, NV as well as industrial and irrigation needs of water for the downstream users. Lake Mead is a crucial factor to the development of the Southwest. Altogether, around 25,000,000 people rely on Lake Mead[12].

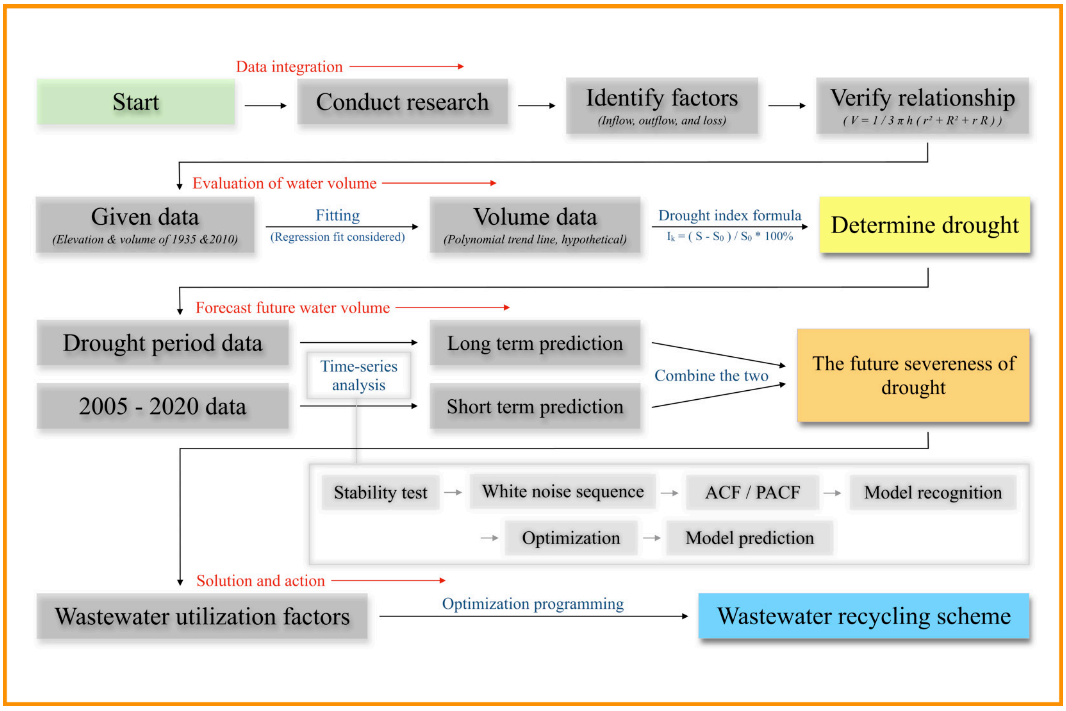
1.2 Problem Restatement

Droughts has been increasing severely all around the globe over the past few decades, researchers have been continuously developing and exploring ways to recycle water more efficiently and effectively. In the summer of 2021, Lake Mead reported to its lowest level of water on record since 1930. Climate change propelled drought around these areas and with the increasing demand of water by these 25 million people that relies on Lake Mead, the reservoir has shrunk around 36 percent from its full capacity. The surface elevations have dropped by about 131 feet (39.9288 meters) since 1999 due to the droughts and increasing water demands. The elevation change affected the water quality of the lake. The Bureau of Reclamation announced the first ever shortage declaration on August 16, 2021. Although many researchers have been developing ways to recycle water, there are other factors which would affect the inflow and outflow of Lake Mead.

In order to address to the problem of drought of Lake Mead, we will identify the factors that would impact the inflow, outflow and loss of Lake Mead. Finding the relationship between the factors and their influence to the water levels of Lake Mead with its irregular shape and varying depths throughout. Then using these factors and relationships to develop models of the water levels in Lake Mead for future years. Then based on the models that are made, find the factors that could be included in a plan for recycling the water usage and developing a plan for recycling the water.

The purpose of this is to discover and investigate in the area of recycling the water and addressing to the problem of drought that Lake Mead is facing. In the end, there would be two models that are used to predict the water levels of Lake Mead, a plan that considers the factors of inflow, outflow and loss of Lake Mead and a non-technical news article reporting the recommendations and takeaways from this investigation.

1.3 Our work



2 Assumptions and Justification

2.1 General Assumptions

- **Assumption 1:** Data for Lake Mead capacity and the polynomial trend line are true and reliable.
- **Justification 1:** The National Park Services provided people two set of datas which showed hypothetical capacity of Lake Mead at controlled elevations. We assumed these data are true and by plotting them, we created a polynomial trend line that aimed to fit the line with coordinates as much as possible. We assume this fitted trend line as reliable.
- **Assumption 2:** The impact of oil, plant, and animal on Lake Meads inflow, outflow, and loss is almost negligible.
- **Assumption 3:** The riverbed is smooth and the distance between two elevations is a straight line.
- **Justification 3:** Since the formula to calculate the frustum of the cone will be used to verify the relationship between elevation, surface area and volume, the distance between two elevations should be a straight line.

2.2 Variables and Definitions

The variable table is on the next page.

Table 2.1: Variables Table

Variables	Description
F	Volume of lake water
t	Time in year
V_{inflow}	Inflow to Lake Mead
$V_{outflow}$	Outflow of Lake Mead
V_{loss}	Loss of Lake Mead
$V_{tributary\,inflow}$	Tributary inflow
$V_{damn\,water\,release}$	Dam water release inflow
$V_{precipitation}$	Precipitation inflow
$V_{groundwater}$	Groundwater inflow
$V_{water\,released}$	Water release outflow
$V_{water\,consumed}$	Consumed water outflow
$V_{evaporation}$	Evaporation loss
$V_{household\,use}$	Household use of water
$V_{agricultural\,use}$	Agricultural use of water
$V_{public\,use}$	Public use of water
V	Frustum of the cone
h	Height
E_1	Elevation of larger frustum cone
E_0	Elevation of smaller frustum cone
A_1	Area of larger frustum cone
A_0	Area of smaller frustum cone
R	Radius of larger frustum cone
r	Radius of smaller frustum cone
I_k	Distance in percentage of current capacity away from average capacity over years
S	Experimental values in percentage errorformula as current reservoir capacity
S_0	Actual value in percentage error formula as the average storage over years
$Score$	Score of wastewater recycling scheme
V_{waste}	Volume of wastewater generated by the city
V_{consum}	Volume of waster used by residents
$V_{provide}$	Volume of water that Lake Mead can supply
$rate$	Recovery rate
$Cost$	Money spent

3 Identify factors and relationship

3.1 Inflow, Outflow and Loss

General Function

There are a number of factors that could influence the inflow, outflow and loss of water in lakes. Letting F be the volume, t = time in year, the inflow, outflow and loss forms a dynamic cycle which usually balances the volume of water in lakes. The could be calculated by:

$$F(t+\Delta t) = F(t) + V_{inflow} - V_{outflow} - V_{loss} \tag{3.1}$$

Nevertheless, factors that impact the inflow, outflow and loss varies from lake to lake depending on its geographic location, water resources nearby, human environment, to name but a few. For Lake Mead, specific factors that contributed to its

$$\begin{aligned} V_{Inflow} &= V_{tributary\,inflow} + V_{dam\,water\,release} \\ &+ V_{precipitation} + V_{groundwater} \end{aligned} \tag{3.2}$$

$$V_{Outflow} = V_{water\,released} + V_{water\,consumed} \tag{3.3}$$

$$V_{Loss} = V_{evaporation} \tag{3.4}$$

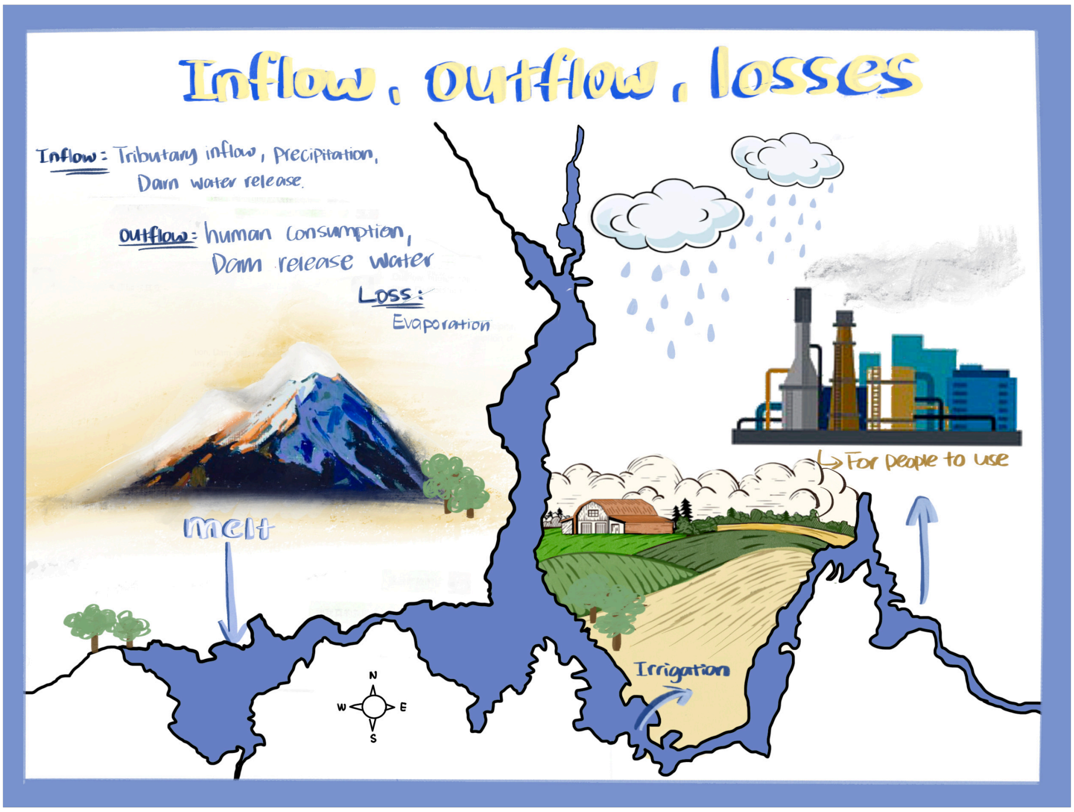


Figure 3.1: Inflow and Outflow

Inflow ($V_{tributary\,inflow}$):

Lake Meads inflow is mostly depending on one of its tributary lake, the Colorado River, which contributed to 96% of its inflow each year, while only 4% are from other 3 tributaries or rainfall and snowfall. However, such difference in weigh of impact make Lake Meads water level closely related to the Colorado River, which makes it more susceptible to the conditions of changes happen to the Colorado River. During drought periods, if Colorado river suffers, the amount of inflow decreases, Lake Meads water level will drop significantly as the amount of inflow is less than the amount of outflow plus loss.

Inflow ($V_{Precipitation}$):

Precipitation (rainfall and snowfall) and snow melt are direct inflow factors for Lake Mead, and precipitation each year could be calculated by ave. precipitation in Arizona (annually) * ave. surface area of Lake Mead (annually). Though precipitation seems to not weigh much, several severe snow melt from Rocky Mountains has impacted the water level of Lake Mead in the short-term enormously. For example, the heavy snowmelt caused a 3.3 million acre-feet release from Glen Canyon to Lake Mead, which resulted in a small fluctuation in water level between 2011-2013.

Inflow ($V_{Dam\,water\,release}$):

Glen Canyon Dams releasing of water which surpassed the demand and wasnt conform to the contract has made the water level of Lake Mead stays high during the wet years. However, the contribution of dams to the inflow of water is also depends on many other factors such as temperature, precipitation, etc, indicating that it is almost neglectable in comparison to the Colorado rivers inflow.

Inflow ($V_{groundwater}$):

Groundwater is an important factor in the gain or loss of Lake Mead. In rainy seasons, the soil and rocks around the lake are soaking wet, and the water passes through underground fissures or channels to supply the lake with additional water, contributing to the lakes content. However, in times of drought, the vicinity of the lake will be terribly dry, and instead of providing water, it takes the lakes water away. In this scenario, the lakes water seeps through the soil, rocks, and pores, causing the elevation to drop.

Outflow ($V_{waterconsumed}$)

$$V_{waterconsumed} = V_{household\,use} + V_{agricultural\,use} + V_{industrial\,use} + V_{public\,use}$$

Lake Mead provide water supplies and generate electricity across its tribal lands for about 25 million people, and serves as the main source of drinking water for Las Vegas which obtains 90% of it from Lake Mead annually. Human consumption of Lake Meads water includes the following categories: household water consumption, agriculture use, industrial use and public use. Human water consumption each year could be calculated by the addition of the following uses:

Household use = ave.use per household per day (gallons)*number of households* 365,
Agriculture use = farmland area (acre) * ave. use of water per acre annually (gallons),
Industrial use = adding up the use of water annually for each factory,
Public use = construction use + afforestation use + business services use + facilities use,

where ave. stands for average. Nonetheless, 40% of water should be subtracted from the sum of water used for all above categories as they will be recycled to enter back Mead Lake. By employing this calculation, people can obtain the most accurate data of human water consumption each year, however, we are not able to provide it here because of lacking of data.

Outflow ($V_{water\,released}$):

Theoretically, the dam allows water to impound in and release out every year. Water release in Lake Mead is because of several factors. For instance, during the wet years or water abundant times, water should be released out to reduce the pressure endured by the dam. Besides, water in Lake Mead is also used for balancing water in other areas, making water released a factor of outflow.

Evaporation

Evaporation is directly proportional to temperature and is inversely proportional to water level. The higher the temperature, the more the lake evaporates, the lower the water level, and vice versa. Impact of evaporation spurred by warmer temperatures to Lake Mead is relatively larger than other rivers because of its geographic location and weather - locating in an arid region with deserts surrounding it. According to Climate Hot Map, approximately 3% of the stored water in Lake Mead is being evaporated each year, making evaporation an important factor for loss. Along with the intensified global warming, Lake Mead might even face a more severe decline in water levels in the future.

3.2 Relationship between elevation, area, and volume

Since calculating the irregular shape is too complicated, we can approximate the shape of Lake Mead as a frustum of a cone, with a circular surface area. Hence, the volume of slice of any irregular shape can be obtained from the volume formula of the frustum of a cone.

The formula of the frustum of the cone is

$$V = \frac{1}{3}\pi h(r^2 + R^2 + rR) \tag{3.5}$$

Several substitution should be done to some of the variables in the formula, making the formula clearer for bringing in data of elevation, surface area and volume.

1.Substitute the height (h) in the formula as the elevation of the larger circle (E_1) minus the height/elevation of the smaller circle (E_0).

2.Assuming the area of the larger circle of the frustum of the cone as A_1, radius as R. and the smaller one A_0. Assuming the area of the smaller circle of the frustum of the cone as A_0, radius as r.

$$\begin{aligned} V &= \frac{1}{3}\pi h\left(r^2 + R^2 + rR\right) \\ &= \frac{1}{3}h\left(\pi r^2 + \pi R^2 + \sqrt{\pi^2 r^2 R^2}\right) \\ &= \frac{1}{3}\left(H_1 - H_0\right)\left(A_0 + A_1 + \sqrt{A_0 \times A_1}\right) \end{aligned} \tag{3.6}$$

where $V_1 - V_0$ is the difference between the volume for the given two dates, E_1 and H_0, E_1 and A_0 are the elevations and surface areas respectively for the corresponding dates.

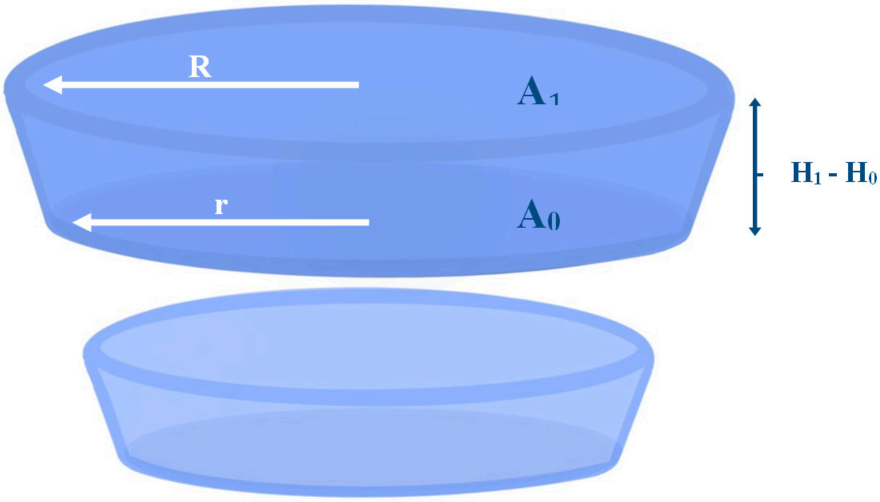


Figure 3.2: Frustum of the cone

Actual Value of V (V_1-V_0)	Calculation of V
29,686,054 - 28,229,730 = 1,456,324 (acre-feet)	1/3(1229.0-1219.6) (159,866+152,828+ √159,866x152,828) = 1469537.697 (acre-feet)
18012331 (acre-feet)	18797958.56 (acre-feet)
7641004 (acre-feet)	7789209.543 (acre-feet)

Figure 3.3: Actual value of V

In order to verify the applicability of the derived formula for the relationship between elevation, surface area and volume, percentage error could be used. If the subtraction of the two volumes on the left is close to the answer derived from the formula on the right, that is, the percentage error is less than 5%, the formula is proved to be applicable and successful. However, if the percentage error is bigger than 5%, the formula is not acceptable.

$(1469537.697 - 1,456,324)/1,456,324 \times 100 = 0.90\%$

$(18797958.56 - 18012331)/18012331 \times 100 = 4.36\%$

$(7789209.543 - 7641004)/7641004 \times 100 = 1.94\%$

It could be concluded that by using the formula derived from the frustum of a cone, the volume we get is usually a little bigger than the actual volume change. Luckily, as 0.90%, 4.36%, 1.94% are all less than 5%, the formula is proved to be acceptable. It could be discovered that the higher the elevation difference is, the larger the frustum, the more inaccuracy there is. This is because we assume the difference between elevation a straight line, but the surface of the riverbed could never be a straight line. As the frustum of the cone gets bigger, we are assuming more parts of the riverbed as a smooth straight line, which caused more inaccuracy.

4 Drought Definition Model

4.1 Overall patterns for Lake Mead water levels

Lake Mead was formed by the Hoover Dam on the Colorado River, making the water level rise almost in a straight line during its early years. Since then, Lake Mead has remained filled, with fluctuating water levels due to a variety of inflow and outflow factors. After that, prior to the filling of Lake Powell, a reservoir that has a similar size to Lake Mead, the huge amount of water in the Colorado River flows into it uncontrollably, making Lake Mead more vulnerable to drought. Obviously, due to these factors, Lake Mead has endured several large drops in elevation: elevation fell from 1200 feet to 1092 feet between 1952-1955, from 1204 to 1090 feet between 1962-1965. Paul Lutus speculated that, between 1966-1983, people who oversaw the filling operation of the Lake Powell took water from the Lake Mead at its rapid flow times to balance the water level and make Lake Powell act as a buffer, which smoothed out the annual peaks and troughs water levels in Lake Mead. Afterward, wet years from the 1970s to the 1990s make both lakes to their full capacity, making Lake Mead reach a record of 1,225 feet in July 1983. However, factors that make Lake Mead stays at a high capacity are also because of Glen Canyon Dams releasing of excess water to Lake Mead, which that quantity is significantly higher than what it was contracted for. Nevertheless, since 2000, persistent drought experienced by the Colorado River has influenced Lake Mead significantly. With the increasing human consumption, water release, and temperature, elevation in Lake Mead has steadily declined. During the ongoing water level declining years, there was heavy snowmelt in the Rocky Mountains in 2012, in which approximately 3.3 million acre-feet of snow enters Lake Mead, causing a fluctuation in water level. Despite this heavy snowmelt buffer, Lake Mead's

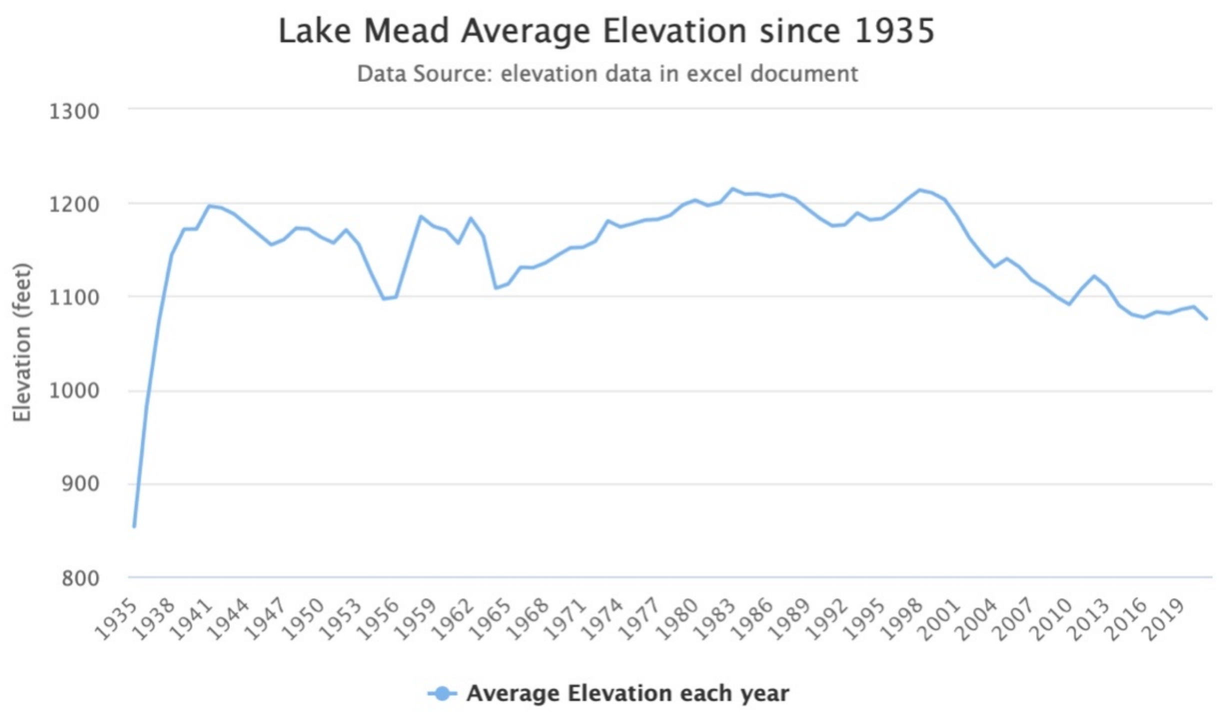


Figure 4.1: Lake Mead Average Elevation since 1935

water level continues falling, suggesting an urgent need for action.

4.2 Drought Definition Criteria

The drought index of each year can be obtained through the percentage of water storage in the reservoir at the time compared with the average storage over years. We discovered that the formula of percentage error could be used to calculate the drought index as they have similar characteristics.

Firstly, we set the actual value in percentage error formula as the average storage over years (S_0), which is a fixed variable that we take as a standard to measure how and to what extent other volume values differ from it. Then, assuming experimental values in percentage error formula as current reservoir capacity (S), which is from 1935-2021.

$$I_k = \frac{(S - S_0)}{S_0} \times 100\% \tag{4.1}$$

We should first make a thorough consideration to get the most accurate S_0 and other S values by using hypothetical data provided on the National Park Services website. On the website, we got hypothetical capacity data of controlled elevation from 1935 and 2021 respectively. Nevertheless, we can neither only use the capacity data for 1935, nor can we only use the capacity data for 2010 because we found out that capacity for all controlled elevations in 1935 is greater than that of 2010 as a whole. After research, we speculated that this phenomenon is because the outflow and loss for Lake Mead gradually yield more than inflow, therefore making some of the riverbeds dry thoroughly overtimes. In order to balance out the data from the water abundant times and drought times, we averaged out the capacity data of the two set at the same controlled elevation. By plotting these data on a graph, in which we took elevation as x-coordinate and volume as y-coordinate and plotted 5 pairs of data on the graph, we created a polynomial line aiming to make it fit these points as much as possible, as shown in the following graph.

$$V = 195.87x^2 - 33364x + 1.444 \times 10^8 \tag{4.2}$$

Usually, if $0.8 < R^2 < 1$, a trend line is proved to be reliable. The closer the value of R to 1, the better the regression line fits to plotted values. In this case, as the regression fit for H_v is 0.9975, it proves the trend line to be valid. By employing the trend line formula which shows the relationship between elevation and volume, we created a table that calculated all average data for volume by elevation from 1935 to 2021. As long as one set of data is given, the other can be found. This allowed us to calculate value for S_0, which stands for average water storage value over years.

In order to obtain a more accurate S_0 value to help define the drought, we conducted data cleansing for all annual average volume from 1935 to 2021 to avoid outliers which might resulting in a systematic higher or lower drought index for storage volumes. We first arranged average volumes from 1935 to 2021 in an ascending order, then calculated out

$$Q_1 = \frac{(n+1)}{4} = \frac{[(2021 - 1935 + 1) + 1]}{4} = 22 \tag{4.3}$$

which is 16886103.95 acre-foot, and

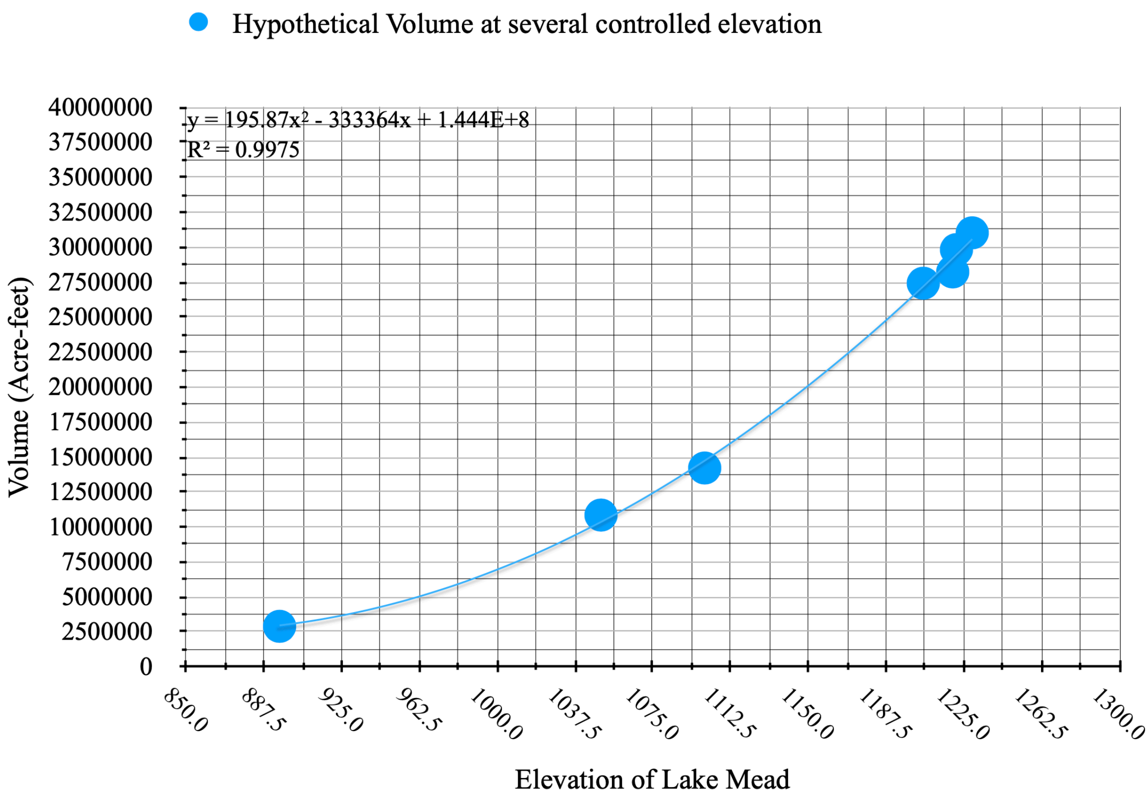


Figure 4.2: Lake Mead Hypothetical Volume since 1935

$$Q_3 = \frac{(3n+1)}{4} = \frac{[3(2021-1935+1)+1]}{4} = 65.5 \tag{4.4}$$

which is the mean of 65th and 66th number, 24439920.29 acre-feet, where n is stands for number of terms.

Then, we used $Q_1 - 1.5(IQR)$ and $Q_3 + 1.5(IQR)$ to set up the lower and upper boundary for outliers, where IQR is

$$Q_3 - Q_1 = 24439920.29 - 16886103.95 = 7553816.34 \,\text{acre-feet} \tag{4.5}$$

Lower boundary:

$$16886103.95 - 1.5 \times 7553816.34 = 5555379.45 \,\text{acre-feet} \tag{4.6}$$

Upper boundary:

$$24439920.29 + 1.5 \times 7553816.34 = 35770644.79 \,\text{acre-feet} \tag{4.7}$$

After that, we compare all 87 annual average volume data from 1931-2021 with the lower and upper boundary volume, attempting to clean out the outliers. We found out that the volume for 1935, that is, 2557448.71 acre-feet, is smaller than the lower boundary of 5555379.45 acre-feet, hence, mathematically, this data should be eliminated.

Nevertheless, after reading articles from National Park Services, we also decided to eliminate the volume data for 1936. Despite the average volume for 1936, 5911854.07 acre-feet, is not an outlier as its smaller than the lower boundary, we still consider this data as unreliable from the logic perspective. According to National Park Services, the construction of the Hoover dam was completed in 1936, which impounded water to form Lake Mead. This indicated that the volume for the year 1936 is inaccurate because it at least took Lake Mead several months to a year to store water. Hence, logically, we should eliminate the volume data for both 1935 and 1936.

After data cleansing, we averaged out S_0, getting the measuring standard for all volume values. Then, we employ the formula, calculating I_k, that is, the distance in percentage of current water storage away from average storage over years. Lastly, we ran these data of I_k in python to define the drought index for each years storage, the standard is shown in the below table:

Drought Index	No drought	Mild drought	Moderate drought	Severe drought	Extreme drought
I_k	≥ -10	-10 ~ -30	-30 ~ -50	-51 ~ -80	< -80
Elevation (h) in feet	≥1140.71	1140.71 ~ 1100.76	1100.76 ~ 1053.06	1053.06 ~ 943.19	< 943.19
Volume (V) in acre-feet	≥18998328.13	18998328.13 ~ 14776477.43	14776477.43 ~ 10554626.74	10554626.74 - 4221850.69	< 4221850.69

Figure 4.3: Drought Index

4.3 Drought Period Comparison

In figure 4.4 below, we used the polynomial equation provided to graph out all volumes since 1935 and used different colored areas to represent no drought to extreme drought, making the degree of drought for each year easy to judge.

According to our drought index standard (ignoring the storage period), since Hoover Dam was built and filled in 1935, Lake Mead has experienced three droughts, from 1954 to 1956, 1964 to 1968, and 2004 onward. By contrast, the current drought is the most severe and longest one. It has the lowest trough.

Firstly, looking at the overall trend, the two droughts in the 1900s have recovered, and the water level and capacity of the river rose rapidly after falling. However, since 2000, the capacity has decreased significantly. After the beginning of the drought period in 2004, even if the capacity trend fluctuated slightly from 2012 to 2013, the overall trend of Lake Mead's capacity has always been declining without any signs of reversal. Since 2014, the capacity of Lake Mead has been relatively stable, but it has always been in moderate drought.

Besides, judging from the capacity trough of these three drought periods, the trough of the second dry period (1964-1968) was slightly higher than that of the first dry period (954-1956), which seemed to be an improvement, moving onwards to the normal level. However, during the current drought period (2004), Lake Mead's capacity has broken the historically low record in 2021. Thus, this could prove that the current dry spell is the worst.

Finally, in terms of drought duration, each drought is longer than the last, it lasts for 3 years in 1954, 5 years in 1964, and lasts for 18 years in the drought that began in 2004. As a result, the current drought period is the most severe one in comparison to several others. As of now, there are signs of continued deterioration, and we will be forecasting drought trends after 2021.

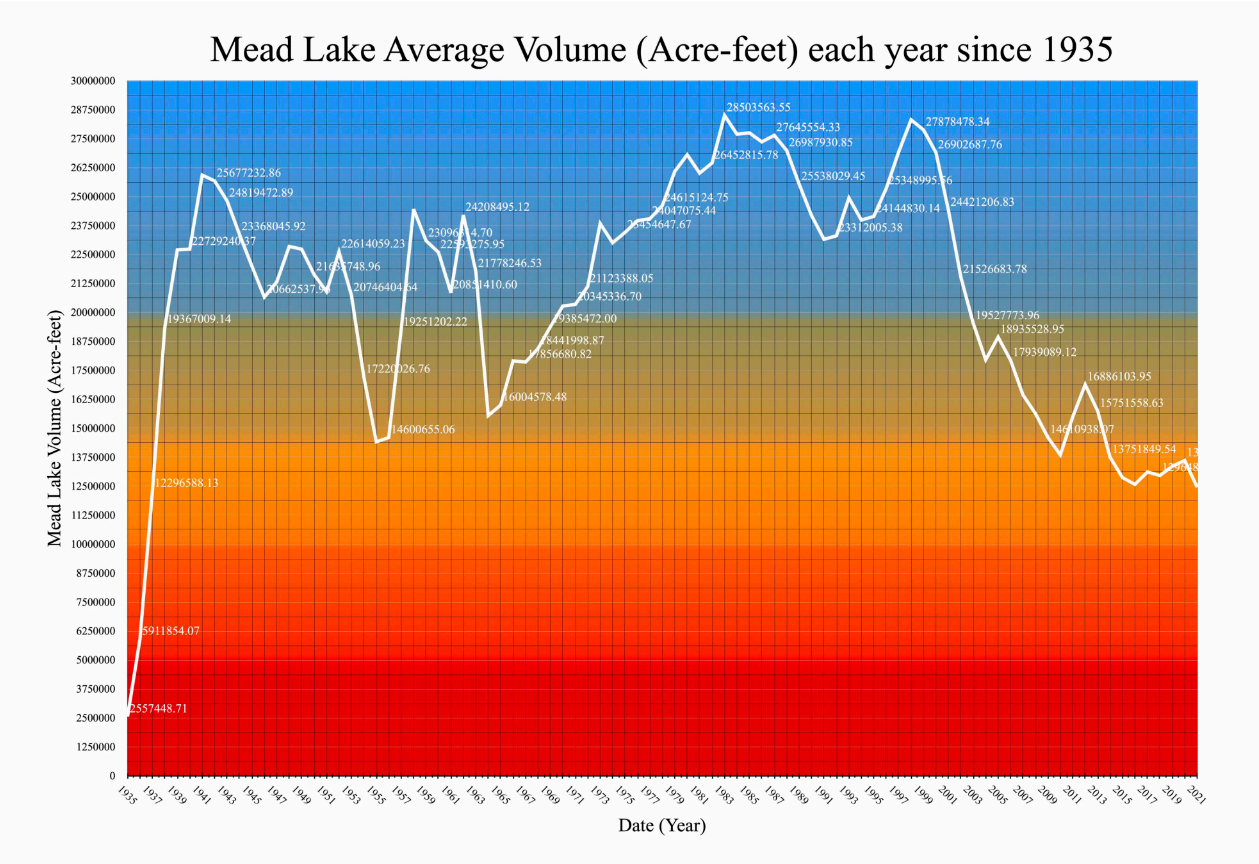


Figure 4.4: Lake Mead Average Volume (Acre-feet) each year since 1935

5 ARMA Prediction Model

As the only non-hypothetical data we got was elevation with its dates and we desired to look at the future trend of drought, we discovered that Time Series Analysis fitting our needs because of its characteristics.

Time Series Analysis, refers to a type of sequence that arrange values of the same statistical indicator in time order. The processing information in this model is highly dependent on time and it convert each additional unit time to a step. The main purpose for applying Time Series Analysis is to make predictions for future trend based on existing data. All above properties of Time Series are consistent with our data and demand. Therefore, we decided to use Times Series to predict the future trend of drought.

However, one important prerequisite for using this analysis is that data should pass a range of stability tests. Time Series Analysis can only predict trend for relatively stable information. Thus, we started stability tests. Step 1. Taking month as unit, analyzing the water level of each month in chronological order.

Step 2. Performing first-order difference to detect the change in value between months and understand the relationship.

$$\Delta y_x = y_{x+1} - y_x \; (x = 1, 2, 3, ...) \tag{5.1}$$

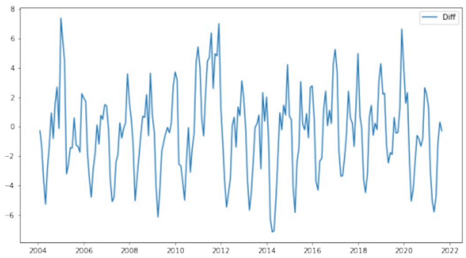


Figure 5.1: first-order difference(predict1)

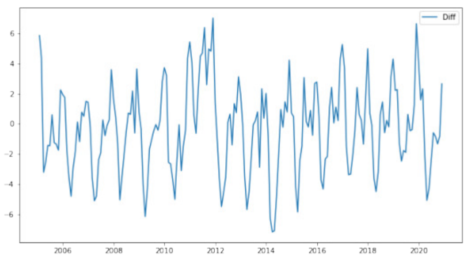


Figure 5.2: first-order difference(predict2)

Step 3. Running ADF test on python to see the stability of data. Looking at 4 values in the results: p-value, critical value 1%, 5% and 10%, we were able to examine the stability of data. The closer these values get to zero, the more stable the examined data is.

	value
Test Statistic Value	-4.326257
p-value	0.0004
Lags Used	14
Number of Observations Used	197
Critical Value(1%)	-3.463987
Critical Value(5%)	-2.876326
Critical Value(10%)	-2.574652

Figure 5.3: ADF test(predict1)

	value
Test Statistic Value	-3.7091
p-value	0.003987
Lags Used	13
Number of Observations Used	177
Critical Value(1%)	-3.467845
Critical Value(5%)	-2.878012
Critical Value(10%)	-2.575551

Figure 5.4: ADF test(predict2)

As the tables shown above, all mentioned values for both set of data in ADF tests approaches to 0, which represented that elevation data being examined (from 1937-2010) is stable.

Step 4. IC values

AIC value examine:

$$AIC = 2k - 2ln(L) \tag{5.2}$$

BIC value examine:

$$BIC = kln(n) - 2ln(L) \tag{5.3}$$

QH value examine:

$$QH = -2ln(L) + ln(ln(n)) * k \tag{5.4}$$

Step 5. ARMA Afterward, we used ARMA to examine the monthly value difference between existing elevation data, for observing the degree of fitting. We put parts of the given elevation data in ARMA, which a line was then resulted.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t + \alpha_1 \epsilon_{t-} \tag{5.5}$$

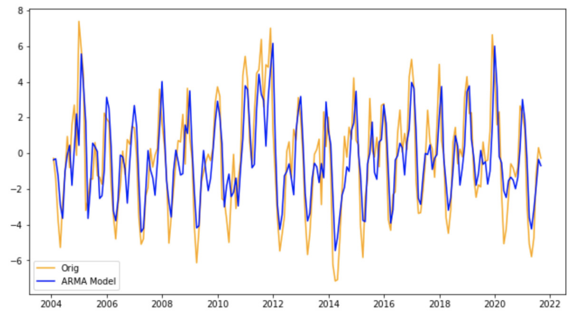


Figure 5.5: ARMA(predict1)

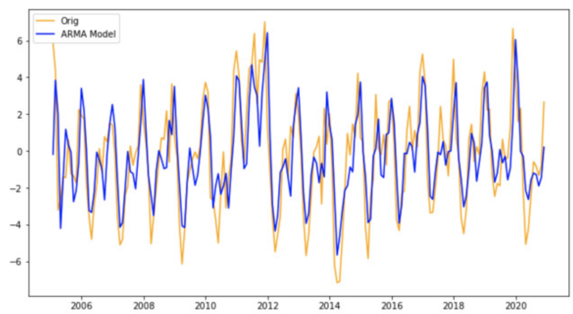


Figure 5.6: ARMA(predict2)

In the above figure, the orange line is the difference analysis between each two month and the blue line is the prediction trend line by ARMA model. It could be proved that ARMA model is reliable as these two lines are almost coincident except peaks and troughs.

After that, we did a in-sample prediction. Since the predicted data is calculated according to the monthly value difference, it needs to be restored by first-order difference. As we should measure from a point of time, we firstly translated the series one moment later. Next, we devised two model evaluation index, which calculated a fitting score and evaluated the in-sample fit of the model respectively. In the process of evaluating in-sample fit, we used Root Mean Square Error (RMSE) to evaluate the effectiveness of fitting. When using this criterion to discriminate, it is necessary to eliminate the influence of "unpredicted" data. We filtered out all no predicted records.

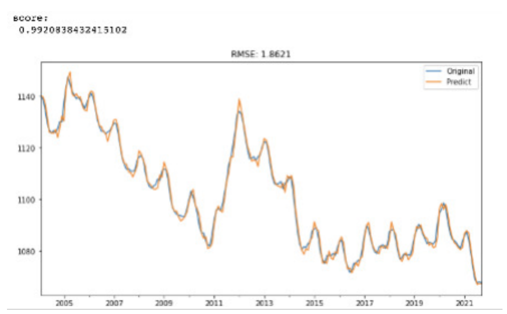


Figure 5.7: ARMA(predict1)

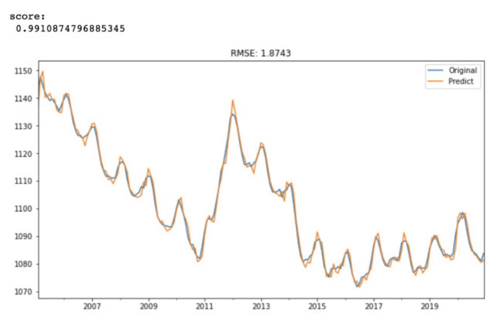


Figure 5.8: ARMA(predict2)

Then, we intercepted part of the form. Through this step, we can more intuitively see the difference between the original data and in-sample data, as well as the size of the gap.

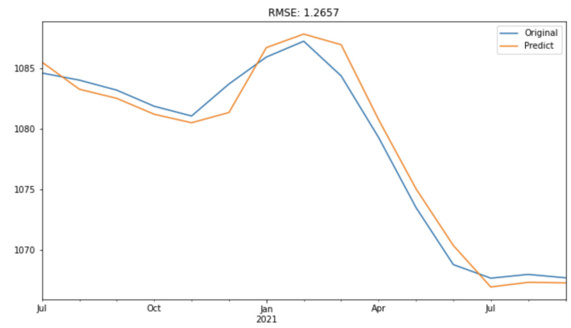


Figure 5.9: ARMA(predict1)

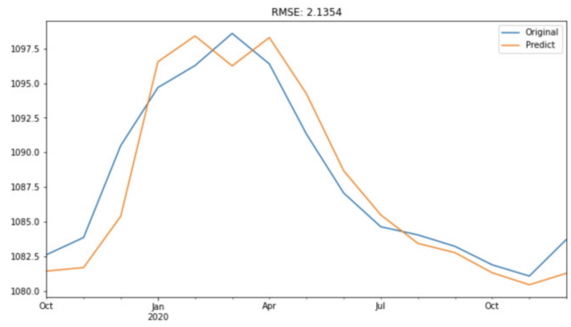


Figure 5.10: ARMA(predict2)

Finally, after a series of information tests (stability) and model tests (coincidence), we used the model to predict water level values for the next 29 years, and future drought trends could be predicted by comparing the Drought Index.

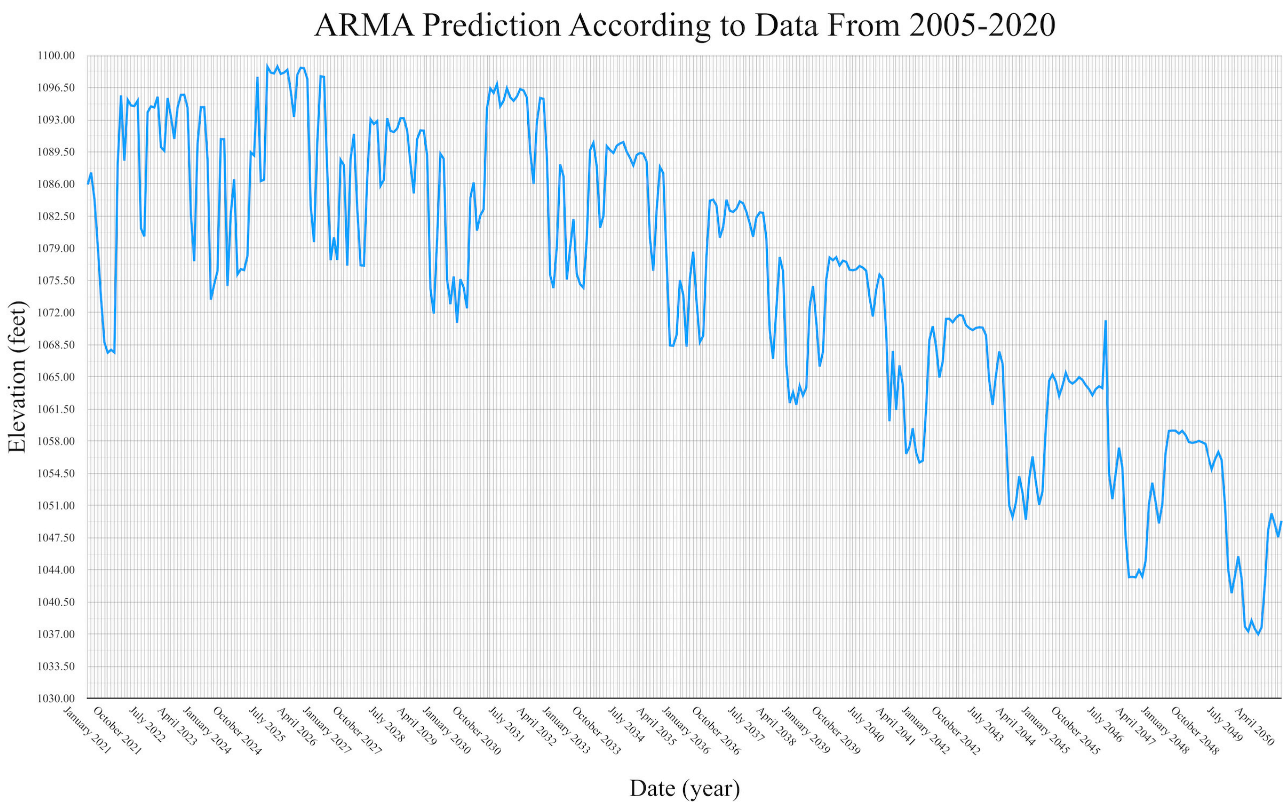


Figure 5.11: Predicting results from 2005-2020

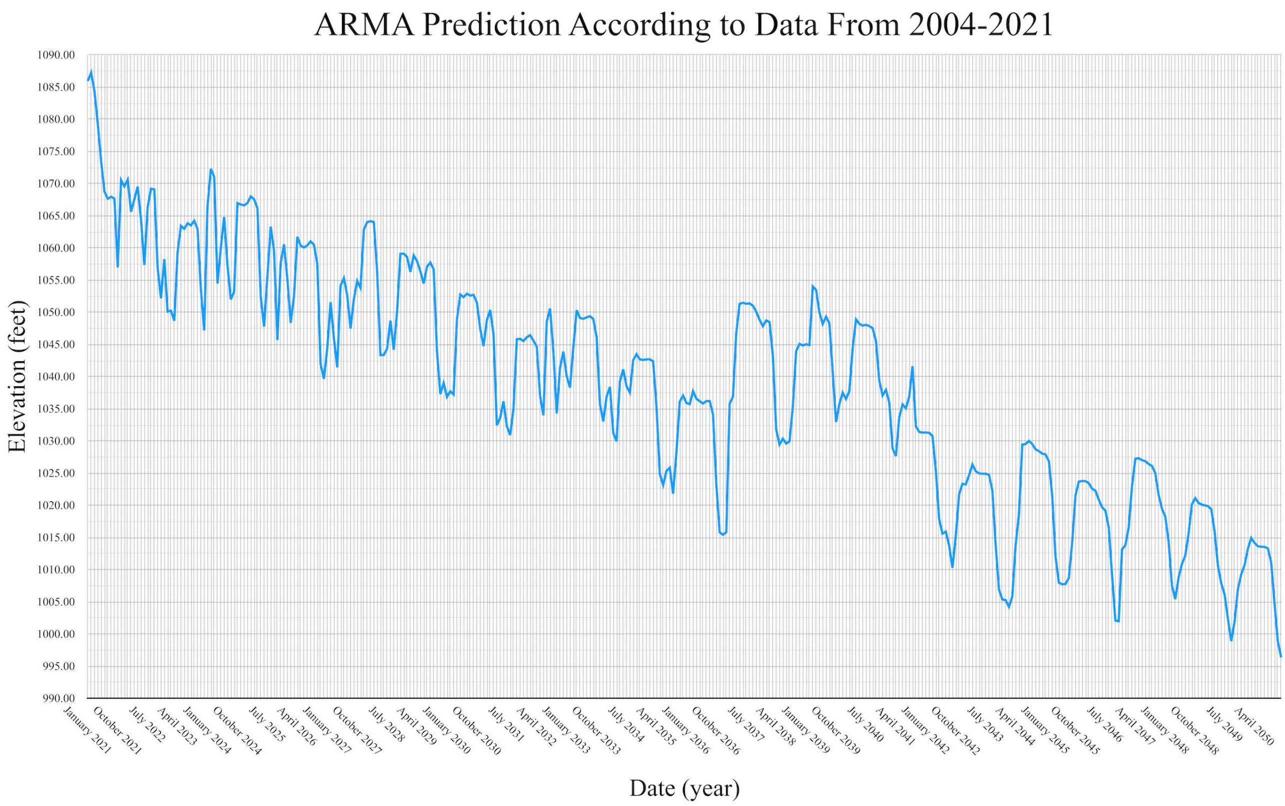


Figure 5.12: Predicting results from 2004-2021

Based on the prediction results of the time series model, we found that the prediction results with the source of water level data for the recent drought period would be a bit drier compared to using the water level data for 2005-2020. This is in line with expectations, since the latter uses less data and uses the same kind of data. Overall, however, the results obtained from the two prediction models are close, and both predict many prolonged dry periods in the future for Lake Mead.

6 Wastewater Reuse Planning

6.1 Factors in water recycle and Priorities

As the water level of Lake Mead keeps dropping, water availability becomes increasingly problematic in Nevada, Arizona and California, some of whose major cities depend on the lake for soaring water usage demands. For these states, wastewater recycling currently seems to be an effective, or possibly the only, solution to water scarcity in the foreseeable future. However, we should be optimistic that there are feasible wastewater-recycling plans to mitigate the effects of water scarcity. This plan should be multifaceted and take the following factors into consideration.

- Statewide Awareness Campaign

  A campaign that is meant to raise public awareness about water conservation and recycling should be vigorously promoted. Every citizen should be encouraged–via public service announcements, education, social media, and radio and television–to cut their water consumption by at least 20%, for example by levying higher water tariffs on households whose water usage exceeds a certain limit or by encouraging them to take shorter showers and replace lush lawns with less-thirsty native plants. If the campaign is met with opposition, the public should be informed that todays state-of-the-art recycling technologies are perfectly able to convert wastewater (even toilet water) into potable water. Perhaps a social-recognition program that recognizes households which save the most water also helps coax the public into a water-saving mindset.

- Sources of Wastewater

  Essential is the task of installing citywide facilities that can handle wastewater from nearly all types of sources. Pipes should be mounted to collect water from home appliances including clothes washers, dishwashers, kitchen sinks, toilets and bathtubs. Moreover, businesses such as restaurants, office buildings, hotels, casinos, hospitals, laundries, car washes, and other light industries should be equipped with wastewater-collecting facilities too. Another important source of wastewater is farmlands, parks and golf courses–areas which consume large amount of water. It is important that water collected from these sources should be treated to near-drinking standards at wastewater treatment plants and returned to Lake Mead for natural cleansing.

- Water Reuse Ratio

  One paramount criterion for the wastewater recycling program to succeed is whether the percentage of recycled water is high enough. Theoretically, it is impossible for this percentage to reach 100%, for exmaple due to evaporation. However, there are already cities that has done a wonderful job in this regard. For example, South Nevada has achieved a de facto water reuse in terms of indoor water sector through a community-wide water recycling process. All water used indoors–whether at home, the office, or a hotel–is captured, treated to near-drinking water standards, and returned to Lake Mead, according to SNWA Public Information Officer Bronson Mack. If the general population is willing to cooperate and the appropriate facilities are in place for water capturing purposes, then we expect the percentage of recycled water to be higher.

- Costs of Construction

  The costs of constructing a water-collecting infrastructure and building a network of wastewater treatment facility can be prohibitive. Usually, they are in the ballpark of hundreds of millions of dollars. Therefore, it is not surprising that local leaders might oppose the plan if they decide to go on a low budget in a fiscal year. Despite this, efforts should be made to change their mind to increase their political willingness in investing in this project.

- The Purposes of Recycled Water

  Recycled water may sound unpleasant, for invokes images of toilet to tap, but recycled water can in fact be treated in technologically advanced facilities to make it suitable for varied purposes including drinking. This process involves capturing wastewater, treating it in a state-of-the-art facility, and supplying it back to Lake Mead. Some reclaimed water is delivered to golf courses and parks for irrigation. Recycled water which eventually arrives at Lake Mead is purified by natural processes, and then becomes clean and safe. This fresh water is then delivered to the towns, cities and states, catering to the needs of people living there and serving a wide range of purposes: drinking, showering, cooking, toilet flushing, clothes washing, garden watering and car washing in the home environment; machine washing, textile coloring, clothes manufacturing, cooling, electricity generation, etc. in the industrial sector; livestock drinking, land irrigation, etc. in the agricultural sector.

- Application to Boulder City - considering its circumstances

  One of the close cities around Lake Mead is: Boulder city. It is located 30 minutes away from Las Vegas airport, right next to the Hoover Dam at the southern tip of Nevada. It is considered as the clean, green city. But it is a very dry area, that only has an average of 5.4 inches rainfall each year. Since it lacks water and is very close to Leake Mead and the Hoover Dam, water management is more important for this area. They do have wastewater treatment plants in Boulder city that is funded federally since they need to have good wastewater treatment. There is a chance where if the wastewater is not controlled well enough then it would flow into Lake Mead which would cause even a bigger problem. Considering these factors, the local leaders would want to solve the problem of the treatment of wastewater and not considering how much money there is needed to make the plan work. The country would also take notice of the issue and would assist them to complete the plan.

6.2 Planning and Optimization Model

We plan to develop a wastewater recycling program that varies according to different cities.

First we classify different cities and divide all cities into two categories: water-starved and capital-starved. These cities were classified according to the following criteria.

1. If the per capita water consumption exceeds 120L per day, the city is a water-starved city.
2. If the GDP per capita is less than \$1.50, the city is a capital-starved city.

We then develop a recycling index that is used to score a particular wastewater recycling scheme.

$$Score = \frac{V_{waste} * Rate_{recycle}}{V_{consum} - V_{provide}} * 100\% \tag{6.1}$$

Where V_{waste} is the volume of wastewater generated by the city, $Rate_{recycle}$ is the recycling rate of wastewater, V_{consum} is the volume of water used by residents, and $V_{provide}$ is the volume of water that Lake Mead can supply. In addition to this, we established the relationship between the wastewater recovery rate - money spent.

Wastewater Recovery Rate	Money Spent
0 - 50%	$c_1 = \$0.38/m^3$
50% - 70%	$c_2 = \$0.45/m^3$
70% - 80%	$c_3 = \$0.53/m^3$

Table 6.1: Optimization model

$$Cost = V_{waste} * \sum_{i=1}^{3} rate_i \times c_i \tag{6.2}$$

where $rate_i$ is the rate corresponding to the recovery rate of the i^{th} fraction.

For water-starved cities, our scheme aims to maximize Score, but ensures that Cost is less than a critical value. Similarly, for a capital-starved city, we need to ensure that Cost is minimized after Score is greater than a certain minimum.

We chose a city called Boulder for our case study. In this city, the per capita water consumption is 150L per day, the per capita GDP is about $2, and the annual volume of wastewater generated is $914,380m^3$. Therefore, this city is a water-starved city. Using the model, we obtain an optimal solution: when the wastewater recovery rate is 73.6%, we have a Score of 86 and a Cost of $273473.

According to our planning and optimization model, water shortage loopholes can be found effectively. But only recycling wastewater is not enough, we must save water from now on. Advocating the harms and negative events that would be brought by the lack of useable water is the way to end water shortage problems fundamentally.

7 Sensitivity Analysis

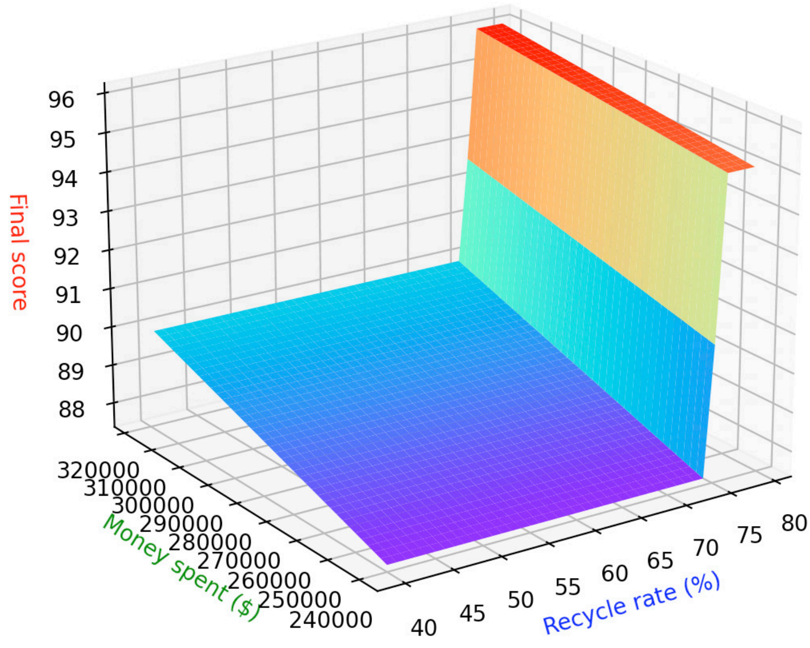


Figure 7.1: Sensitivity Analysis

The results of the sensitivity analysis are shown above, and from the figure, we can conclude that the final score of the wastewater reuse planning is less sensitive to the factor money spent than wastewater recycling rate. The image ushers in an inflection point at about 74% of the wastewater recycling rate, where the inflection point corresponds to our threshold value for filling the water shortage loophole, which is in line with the setup of our model (when the water shortage loophole is filled, the wastewater recycling program naturally approaches a full score).

In conclusion, if we want to construct a more efficient planning, we need to consider more about the wastewater recovery rate.

8 Strengths and Weaknesses

8.1 Strengths

- The elevation predicting model is reliable because past elevation (1937-2010) data were used to predict elevation for 2011-2021 to verify its conformity with real data, which reinforces the reliability of future predictions.
- As a range of tests such as ADF test and white noise sequences evaluation are used to test the stability of elevation data and were finally succeeded, Times Series Analysis is suitable to predict the future trend of drought.
- When the sample size is large (about 350 elevation data), the effect of Times Series Analysis is more accurate than the gray prediction, also it's easier to be realized than a neural network, and better to understand.
- Two indexes were considered comprehensively when developing the wastewater recycling plan(recycling index and costs), thus resulting in a planning scheme that is very comprehensive.

8.2 Weaknesses

- Not all factors for inflow, outflow and loss are considered, only the major ones are considered but not the minor ones. Even though these minor factors would affect the lake water volume, it would only lead to a slight difference in the results.
- Due to the lack of yearly data on factors that impact the inflow, outflow, and loss, only qualitative analysis can be used when discussing these factors, and the weight of each specific factor in them cannot be accurately measured.

9 Conclusion and Future work

After defining drought by calculating the distance in the percentage of current volume away from average volume over years, and defined drought index for each year by our own standards. We discovered that the 2004 onward drought is the most severe one. As the elevation hit a historical low this year, it drove us to use ARMA to predict the future trend. We discovered that the 2004 onward drought will be increasingly severe, though sometimes the trend line fluctuated upward or downward, the overall trend will continue to go downwards. This indicated that cities need to take immediate action to reduce the impact of a pending drought. Therefore, we listed and explained factors that need to be considered, planned to implement wastewater recycling, and established an optimal scheme that considers the optimal water recovery rate and cost.

Nevertheless, only by implementing a water recycling scheme, people can never address the water shortage problem fundamentally. Citizens awareness is always as important as advanced recycled technology or a perfectly designed plan. People should be conscious of todays water shortage situation and always remember to save water.

References

[1] "Acres of Open Space." Boulder County, 24 Mar. 2021, https://www.bouldercounty.org/open-space/management/acres/.

[2] Additional informationFundingThis material is based upon work supported by the National Science Foundation [Grant no. IIA-1301726]. "Socio-Economic Drought Assessment in Lake Mead, USA, Based on a Multivariate Standardized Water-Scarcity Index." Taylor amp; Francis, https://www.tandfonline.com/doi/abs/10.1080/02626667.2019.1593988.

[3] Additional Water to Be Released from Lake Powell to Lake Mead, Avoiding Shortages in Lower Basin in 2012, https://web.archive.org/web/20141214113152/http://home.doi.gov/news/pressreleases/Additional-Water-to-be-Released-from-Lake-Powell-to-Lake-Mead-Avoiding-Shortages-in-Lower-Basin-in-2012.cfm.

[4] authors, All, and G. Chris Holdren amp; Kent Turner. "Characteristics of Lake Mead, ArizonaNevada." Taylor amp; Francis, https://www.tandfonline.com/doi/full/10.1080/07438141.2010.540699.

[5] Barnett, Tim P., and David W. Pierce. "When Will Lake Mead Go Dry?" AGU Journals, John Wiley amp; Sons, Ltd, 29 Mar. 2008, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2007WR006704.

[6] "Boulder City, Nevada Population 2021." Boulder City, Nevada Population 2021 (Demographics, Maps, Graphs), https://worldpopulationreview.com/us-cities/boulder-city-nv-population.

[7] Castle, Stephanie L., et al. "Groundwater Depletion during Drought Threatens Future Water Security of the Colorado River Basin." AGU Journals, John Wiley & Sons, Ltd, 29 Aug. 2014, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014GL061055.

[8] Harbeck, Guy Earl, et al. Water-Loss Investigations: Lake Mead Studies. U.S. Govt. Print. Off., 1958.

[9] Lutus, Paul. "Lake Mead Water Levels", https://www.arachnoid.com/NaturalResources/.

[10] "Municipal Water Consumption." Encyclopædia Britannica, Encyclopædia Britannica, Inc., https://www.britannica.com/technology/water-supply-system/Municipal-water-consumption.

[11] "News Flash." Boulder City, NV, https://www.bcnv.org/CivicAlerts.aspx?AID=222.

[12] "Overview of Lake Mead." National Parks Service, U.S. Department of the Interior, https://www.nps.gov/lake/learn/nature/overview-of-lake-mead.htm.

[13] "Statsmodels.tsa.ARMA.model.ARMAű." Statsmodels, https://www.statsmodels.org/stable/ generated/statsmodels.tsa.ARMA.model.ARMA.htmlstatsmodels.tsa.ARMA.model.ARMA.

[14] "Total Gross Domestic Product for Boulder, CO (MSA)." FRED, 9 Dec. 2020, https://fred.stlouisfed.org/series/NGMP14500.

[15] "Wastewater Produce: Wastewater Management: Eco-Septic." Ecoseptic, 16 Apr. 2021, https://ecoseptic.com.au/how-much-wastewater-does-one-person-produce/.

[16] "Water Level Dropping at Lake Mead, NV/AZ, USA: Global Warming Effects." Water Level Dropping at Lake Mead, NV/AZ, USA | Global Warming Effects, https://www.climatehotmap.org/global-warming-locations/lake-mead-nv-az-usa.html.

[17] "Water Resource Recovery Program." City of Boulder, https://bouldercolorado.gov/services/water-resource-recovery-program.

Appendices

Time-Series

```
import pandas as pd
import numpy as np
import seaborn as sns  #
import itertools
import datetime
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.tsa.stattools import adfuller  # ADF
from statsmodels.stats.diagnostic import acorr_ljungbox  #
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf  #
from statsmodels.tsa.arima_model import ARIMA  #
from statsmodels.tsa.arima_model import ARMA  #
from statsmodels.stats.stattools import durbin_watson  # DW
from statsmodels.graphics.api import qqplot  # qq


def data_handle():
    data = pd.read_csv('./data.csv', sep='\t')
    # print(data.columns)
    # print(data.describe()) #,-10000
    # print((data.isnull()).sum()) #
    # print((data.duplicated()).sum()) #
    #

    for data1 in data.columns:
        def change_mean(x):
            if x == 0:
                return mean
            else:
                return x

        mean = data[data1].mean()
        data[data1] = data[data1].fillna(mean)
        data[data1] = data[data1].replace(0, mean)
    print(data)
    #
    data.to_csv('./new_data.csv', index=0)


def Resampling():  #
    df = pd.read_csv('./new_data.csv', engine='python')
    #
    df['Year'] = pd.to_datetime(df['Year'], format='%Y')
    df = df.set_index('Year')
    data = df.loc['1936':'1977']  # 18-1-18-1
    test = df.loc['1977':'2021']
    data_train = data.resample('Y').mean()  # ,
    data_test = test.resample('Y').mean()
    return data_train, data_test


def stationarity(timeseries):  #
    # (),,1,dropna()
```

```
    diff1 = timeseries.diff(1).dropna()
    diff2 = diff1.diff(1)  #
    return diff1
    # diff1.plot(color = 'red',title='diff 1',figsize=(10,4))
    # plt.savefig(fname="scatter.png",figsize=[10,10])
    # diff2.plot(color = 'black',title='diff 2',figsize=(10,4))
    # plt.savefig(fname="scatter1.png",figsize=[10,10])
    #
    # rolmean = timeseries.rolling(window=4,center = False).mean()
    #
    # rolstd = timeseries.rolling(window=4,center = False).std()

    # rolmean.plot(color = 'green',title='Rolling Mean',figsize=(10,4))
    # plt.savefig(fname="scatter2.png",figsize=[10,10])
    # rolstd.plot(color = 'blue',title='Rolling Std',figsize=(10,4))
    # plt.savefig(fname="scatter3.png",figsize=[10,10])

    # for data1 in diff1.columns:
    #    x = np.array(diff1[data1])
    # adftest = adfuller(x, autolag='AIC')
    # print (adftest)

    # p_value = acorr_ljungbox(x, lags=1)
    # print (p_value)

    # plot_acf(diff1['FEB'],lags=40) #
    # plt.savefig(fname="scatter4.png",figsize=[10,10])
    # plot_pacf(diff1['JAN'],lags=10)
    # plt.savefig(fname="scatter5.png",figsize=[10,10])

    # AIC = sm.tsa.arma_order_select_ic(diff1['JAN'],\
    #    max_ar=6,max_ma=4,ic='aic')['aic_min_order']
    # BIC
    # BIC = sm.tsa.arma_order_select_ic(diff1['JAN'],max_ar=6,\
    #       max_ma=4,ic='bic')['bic_min_order']
    # HQIC
    # HQIC = sm.tsa.arma_order_select_ic(diff1['JAN'],max_ar=6,\
    #             max_ma=4,ic='hqic')['hqic_min_order']
    # print('the AIC is{},\nthe BIC is{}\n the HQIC is{}'.format(AIC,BIC,HQIC))

    # p_min = 0
    # q_min = 0
    # p_max = 5
    # q_max = 5
    # d_min = 0
    # d_max = 5
    ## Dataframe,BIC
    # results_aic = pd.DataFrame(index=['AR{}'.format(i) \
    #                            for i in range(p_min,p_max+1)],\
    #         columns=['MA{}'.format(i) for i in range(q_min,q_max+1)])
    ## itertools.product p,q
    # for p,d,q in itertools.product(range(p_min,p_max+1),\
    #                                range(d_min,d_max+1),range(q_min,q_max+1)):
    #    if p==0 and q==0:
    #        results_aic.loc['AR{}'.format(p), 'MA{}'.format(q)] = np.nan
    #        continue
    #    try:
    #        model = sm.tsa.ARIMA(diff1['JAN'], order=(p, d, q))
```

```
    #         results = model.fit()
    #         #pqmodelBIC
    #         results_aic.loc['AR{}'.format(p), 'MA{}'.format(q)] = results.aic
    #     except:
    #         continue
    # results_aic = results_aic[results_aic.columns].astype(float)
    ##print(results_bic)

    # fig, ax = plt.subplots(figsize=(10, 8))
    # ax = sns.heatmap(results_aic,
    #              #mask=results_aic.isnull(),
    #              ax=ax,
    #              annot=True, #
    #              fmt='.2f',
    #              )
    # ax.set_title('AIC')
    # plt.savefig(fname="scatter6.png",figsize=[10,10])


def ARMA_model(train_D, train, test, order):
    arima_model = ARMA(train, order=(1, 1))  # ARIMA
    result = arma_model.fit()  #
    # print(result.summary()) #
    ############ in-sample ############
    pred = result.predict()

    # pred.plot()
    # train.plot()
    # print('{}'.format(mean_squared_error(train,pred)))

    #
    resid = result.resid
    # QQ
    plt.figure(figsize=(12, 8))
    qqplot(resid, line='q', fit=True)
    plt.savefig(fname="scatter7.png", figsize=[10, 10])
    # D-W,
    print('D-W{}'.format(durbin_watson(resid.values)))

    pred_one = result.predict(start=len(train) - 5, end=len(train) + 30, \
                              dynamic=True)
    # print(pred_one)
    # print(len(test))
    # print(pred_one[6:-1])
    # pred_one.plot()
    # test.plot()
    print('{}'.format(mean_squared_error(test, pred_one[6:-1], sample_weight=None, \
                                         multioutput='uniform_average')))
#


def string_toDatetime(string):
    return datetime.datetime.strptime(string, "%Y-%m-%d %H:%M:%S")


def ARIMA_model(train_H, train, test):
    arima_model = ARIMA(train, order=(1, 1, 1))  # ARIMA
    result = arima_model.fit()
```

```
    # print(result.summary()) #
    #################
    pred = result.predict()

    ################
    ##2018-8-1 00:00  2018-9-1 00:00 ###
    # ,pred_restored
    idx = pd.date_range(string_toDatetime('2018-8-1 00:00:00'), periods=len(pred[4:20]),
    pred_list = []
    for i in range(len(pred[4:20])):
        pred_list.append(np.array(pred)[i + 4])
    pred_numpy = pd.Series(np.array(pred_list), index=idx)

    pred_restored = pd.Series(np.array(train_H)[5][0],
                              index=[train_H.index[5]]).append(pred_numpy).cumsum()
    x1 = np.array(pred_restored)
    x2 = np.array(train_H[5:22])
    y = []
    for i in range(len(pred_restored)):
        y.append(i + 1)
    y = np.array(y)
    fig1 = plt.figure(num=2, figsize=(10, 4), dpi=80)
    plt.plot(y, x1, color='blue')
    plt.plot(y, x2, color='red')
    plt.ylim(0, 0.8)
    plt.show()


x = Resampling()
order = (1, 1)
train = stationarity(x[0])
ARMA_model(x[0], train, x[1], order)
ARIMA_model(x[0], train, x[1], order)
```

Optimization

```
# coding=utf-8
from scipy.optimize import minimize
import numpy as np
import scipy
import matplotlib.pyplot as plt

V_waste = 741.3
i_1, i_2, i_3 = 1.5, 1.0, 0.75


# x[0] = k_1, x[1] = k_2, x[2] = k_3
def volume():
    v = lambda x: -(V_waste * (x[0] + x[1] + x[2])) / (3360.6 - (15697387 - 10314475) *
    return v


def score():
    v = lambda x: -100 * x[0] / (x[0] + x[1] + x[2]) + 75 * x[1] / (x[0] + x[1] + x[2])
                x[0] + x[1] + x[2])
    return v
```

```
def con1(args):
    V_recycle0, Score_recycle0 = args
    cons = ({'type': 'ineq', 'fun': lambda x: 100 * x[0] / (x[0] + x[1] + x[2]) + 75 * x
            {'type': 'ineq', 'fun': lambda x: x[0] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[1] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[2] - 0.0},
            {'type': 'ineq', 'fun': lambda x: 0.6 - x[0]},
            {'type': 'ineq', 'fun': lambda x: 0.7 - x[0]},
            {'type': 'ineq', 'fun': lambda x: 0.8 - x[0]},
            {'type': 'ineq', 'fun': lambda x: 0.8 - x[0] - x[1] - x[2]},
            )
    return cons


def con2(args):
    V_recycle0, Score_recycle0 = args
    cons = ({'type': 'ineq', 'fun': lambda x: (V_waste * (i_1 * x[0] + i_2 * x[1] + i_3
            {'type': 'ineq', 'fun': lambda x: x[0] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[1] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[2] - 0.0},
            {'type': 'ineq', 'fun': lambda x: 0.8 - x[0] - x[1] - x[2]},
            )
    return cons


if __name__ == "__main__":
    x0 = np.asarray((0.26, 0.27, 0.27))

    args = (V_waste * 0.6, 60)

    cons1 = con1(args)
    res1 = minimize(volume(), x0, method='COBYLA', constraints=cons1)
    print("<volume constrain>", "volume rate:", -res1.fun.real, ", k[i]:", res1.x)

    cons2 = con2(args)
    res2 = minimize(score(), x0, method='COBYLA', constraints=cons2)
    print("<score constrain>", "score:", -res2.fun.real, ", k[i]:", res2.x)

    k_1 = (res1.x[0]+res2.x[0])/2
    k_2 = (res1.x[1]+res2.x[1])/2
    k_3 = (res1.x[2]+res2.x[2])/2
    volume_rate = (V_waste * (k_1 + k_2 + k_3)) / (3360.6 - (15697387 - 10314475) * (1 /
    score = 100 * k_1 / (k_1 + k_2 + k_3) + 75 * k_2 / (k_1 + k_2 + k_3) + 50 * k_3 / (k
    print("Comprehensive results:", "k_1 =", k_1, ", k_2 =", k_2, ", k_3 =", k_3)
    print("Final volume rate =", volume_rate, ", score =", score)
```

Sensitivity Analyse

```
# coding=utf-8
from scipy.optimize import minimize
import numpy as np
import scipy
import matplotlib.pyplot as plt
from matplotlib import cm
from matplotlib.ticker import LinearLocator, FormatStrFormatter
import numpy as np
from mpl_toolkits.mplot3d import Axes3D
```

```
V_waste = 741.3
i_1, i_2, i_3 = 1.5, 1.0, 0.75


# x[0] = k_1, x[1] = k_2, x[2] = k_3
def volume():
    v = lambda x: -(V_waste * (x[0] + x[1] + x[2])) / (3360.6 - (15697387 - 10314475) *
    return v


def score():
    v = lambda x: -100 * x[0] / (x[0] + x[1] + x[2]) + 75 * x[1] / (x[0] + x[1] + x[2])
            x[0] + x[1] + x[2])
    return v


def con1(args):
    V_recycle0, Score_recycle0 = args
    cons = ({'type': 'ineq', 'fun': lambda x: 100 * x[0] / (x[0] + x[1] + x[2]) + 75 * x
            {'type': 'ineq', 'fun': lambda x: x[0] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[1] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[2] - 0.0},
            {'type': 'ineq', 'fun': lambda x: 0.6 - x[0]},
            {'type': 'ineq', 'fun': lambda x: 0.7 - x[0]},
            {'type': 'ineq', 'fun': lambda x: 0.8 - x[0]},
            {'type': 'ineq', 'fun': lambda x: 0.8 - x[0] - x[1] - x[2]},
            )
    return cons


def con2(args):
    V_recycle0, Score_recycle0 = args
    cons = ({'type': 'ineq', 'fun': lambda x: (V_waste * (i_1 * x[0] + i_2 * x[1] + i_3
            {'type': 'ineq', 'fun': lambda x: x[0] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[1] - 0.0},
            {'type': 'ineq', 'fun': lambda x: x[2] - 0.0},
            {'type': 'ineq', 'fun': lambda x: 0.8 - x[0] - x[1] - x[2]},
            )
    return cons


if __name__ == "__main__":

    V_recycle0s = np.arange(384, 504, 3)
    Score_recycle0s = np.arange(40, 80, 1)

    # Make data.
    X = V_recycle0s
    Y = Score_recycle0s
    X, Y = np.meshgrid(X, Y)

    Z1 = np.zeros((40, 40))
    Z2 = np.zeros((40, 40))
    for i in range(40):
        for j in range(40):
            x0 = np.asarray((0.26, 0.27, 0.27))

            args = (X[i], Y[j])
```

```
            cons1 = con1(args)
            res1 = minimize(volume(), x0, method='COBYLA', constraints=cons1)
            cons2 = con2(args)
            res2 = minimize(score(), x0, method='COBYLA', constraints=cons2)

            k_1 = (res1.x[0]+res2.x[0])/2
            k_2 = (res1.x[1]+res2.x[1])/2
            k_3 = (res1.x[2]+res2.x[2])/2
            volume_rate = (V_waste * (k_1 + k_2 + k_3)) / (3360.6 - (15697387 - 10314475
            score_ = 100 * k_1 / (k_1 + k_2 + k_3) + 75 * k_2 / (k_1 + k_2 + k_3) + 50 *

            Z1[i][j] = volume_rate
            Z2[i][j] = score_ + (100 - score_) / 40 * (0.2 * i)

    fig = plt.figure()
    ax = Axes3D(fig)
    X = np.arange(40, 80, 1)
    Y = np.arange(240000, 320000, 2000)
    X, Y = np.meshgrid(X, Y)
    ax.plot_surface(X, Y, Z2, rstride=1, cstride=1, cmap='rainbow')
    ax.set_xlabel('Recycle rate (%)', color='b')
    ax.set_ylabel('Money spent ($)', color='g')
    ax.set_zlabel('Final score', color='r')
    plt.draw()
    plt.show()
    # print(X.shape, Y.shape, Z1.shape, Z2.shape)
```