Team Number :	apmcm2213503
Problem Chosen :	E

2022 APMCM summary sheet

Analysis and Prediction of Global Nuclear Weapon

Abstract

Nuclear weapons are the most powerful weapons ever developed in the history of mankind, and it is because of the power of nuclear weapons that countries want to protect themselves through the deterrent effect of nuclear weapons. Although nuclear weapons are powerful, when they explode, they release enormous energy and cause serious damage to the ecological environment.

In response to question one, this paper makes a **basic data analysis** based on the relevant data on nuclear weapons of various countries given in the question. First, visualize the status of nuclear weapons owned by all countries, the number of nuclear weapons stored and the number of nuclear tests conducted. Use python to traverse relevant data and complete **descriptive statistical analysis**. Build a **critical weight scoring model**, select seven indicators to score the attitude of different countries towards nuclear weapons, and judge that **North Korea** is the most active country towards nuclear weapons.

For the second question, in order to predict the number and trend of nuclear weapons in the future, this paper constructs a prediction model based on **ARIMA time series analysis**. First, the data is preprocessed, and the **ADF stability test and white noise test** are performed. Then we evaluate and test the model to find the most appropriate **super parameter** to bring into the **prediction model**. Next, the data of possession status and storage of nuclear weapons of each country are brought in, and the prediction and numerical analysis are carried out. Finally, we get that the total number of nuclear weapons in **2123 will reach 2152** and the number of nuclear weapons in each country.

For the third question, this paper gives two definitions of destroying the earth. One is to blow the earth to pieces with nuclear weapons. Through the calculation of the **gravitational binding energy** of the earth and the **explosive capacity of nuclear weapons**, it is found that the current number of nuclear weapons is **far from enough**. The second definition is to destroy the majority of living things on the earth with **nuclear weapons and its radiation**. We convert the world map into a grayscale image, build the **DBSCAN clustering model**, and set the explosion radius and propagation radius of nuclear weapons to obtain the total number of nuclear bombs that destroyed the earth finally.

In response to question four， we submitted a **non-technical article** to the United Nations, explained our team's views on nuclear weapons and related findings with the previous model analysis results, and made several suggestions for all countries. We believe that the ultimate goal of developing nuclear weapons is to **eliminate nuclear weapons**.

Key words: Critical weight method; ADF stability test; ARIMA time series analysis; DBSCAN clustering model

Contents

1 Introduction **1**
1.1 Background 1
1.2 Problem Restatement 1

2 Analysis of Question **2**
2.1 Analysis of Question One 2
2.2 Analysis of Question Two 2
2.3 Analysis of Question Three 3
2.4 Analysis of Question Four 3

3 General Assumptions and Notations **4**

4 Model and Solution of Question One **4**
4.1 Data Collection 4
4.2 Problem Solving 5
4.2.1 The Solution of Problem a 5
4.2.2 The Solution of Problem b 5
4.2.2.1 Data Visualization 5
4.2.2.2 Problem Solving 6
4.2.3 The Solution of Problem c 7
4.2.4 The Solution of Problem d 8
4.2.4.1 Data Processing 8
4.2.4.2 Establishment and Solution of the Model 9
4.2.5 The Solution of Problem e 11

5 Model and Solution of Question Two **12**
5.1 Prediction Model of Nuclear Weapon Number Based on ARIMA Time Series Analysis 12
5.1.1 Prediction Model Based on ARIMA Time Series 12
5.1.2 Stationarity Test 12
5.1.3 White Noise Test 13
5.1.4 Model Estimation and Test 14
5.2 The Number of Nuclear Weapons and National Projections 15

6 Model and Solution of Question Three **17**
6.1 Measurement of Nuclear Weapon Energy 17
6.2 Definition of Earth Destruction 17
6.3 Model Establishment and Solution 17
6.3.1 Model Establishment and Solution Based on Definition ① 17
6.3.2 Model Establishment and Solution Based on Definition ② 19
6.3.2.1 Main Components of Nuclear Weapon Damage Radius 19

6.3.2.2 Nuclear Weapon Detonation Location Model Based on DBSCAN Clustering Algorithm 20

7 A Non-Technical Article **24**

8 Model Evaluation and Further Discussion **25**

8.1 Strengths 25

8.2 Weaknesses 25

8.3 Further Discussion 25

Reference **25**

Appendix **26**

1 Introduction

1.1 Background

Nuclear weapons, including neutron bomb, atomic bomb and hydrogen bomb, are related to nuclear reaction and have great destruction. Among all the weapons invented in human history, nuclear weapons are one of the most powerful weapons. It is because of the power of nuclear weapons that many countries hope to protect their homeland from invasion through the deterrent of nuclear weapons.

A single atomic bomb has the power to destroy a city. Considering the specific explosion parameters, the temperature of the explosion moment of the atomic bomb can reach tens of millions of degrees, and the explosive yield is similar to tens of thousands to hundreds of thousands of tons of TNT. But although nuclear weapons are very powerful, but the destruction of the ecological environment is extremely serious. The detonation of a nuclear weapon can make it difficult for the ecological environment of the explosion area to recover for many years.

1.2 Problem Restatement

Question one：Basic data analysis

According to the requirements of the title, data analysis should be conducted on the acquired data. Firstly, it needs to determine which countries the nuclear weapons have existed in the past. The countries that have reduced their stockpiles the most in the last 20 years and those that have increased their stockpiles the most; Taking five years as a span, the time period of the most nuclear weapon tests from 1945 to 2019 is calculated, and the most active countries in the research of nuclear weapons in the recent 10 years are selected. Finally, the countries that have changed from "not considering nuclear weapons" to "possessing nuclear weapons" are analyzed.

Question two：Forecast nuclear weapons

Mathematical models are built from the data to predict the number of nuclear weapons and predict which countries will have nuclear weapons within the next 100 years. A mathematical model is used to predict the number of nuclear weapons in the next 100 years, the total number of nuclear weapons in the world in 2123, and the total number of nuclear weapons possessed by each country.

Question three：Protect our planet

Build a mathematical model of where a nuclear weapon will detonate, and calculate how many nuclear bombs it will take to destroy the Earth. From the perspective of the individual nuclear weapon, the mathematical model is established to determine the maximum destructive power of the current nuclear bomb and analyze whether it can destroy the Earth. The above study is about the destructive power of nuclear weapons, while the protection of the ecological environment is a topic of constant human research. In order to protect the ecological environment, the limit of the total number of nuclear bombs in the world and the theoretical limit of nuclear weapons of the states possessing nuclear weapons should be analyzed.

Question four：Complete a non-technical essay

Combining the team's findings and writing to the UN in a non-technical article, it makes several recommendations for all countries.

2 Analysis of Question

2.1 Analysis of Question One

The primary purpose of Question 1 is to conduct a basic analysis of the data, which contains five small questions:

Question a: status=3 in the data (country-position-nuclear-weapons.csv) indicates that the country possessed nuclear weapons in that year. Therefore, by traversing the data, we can find that the country with status=3 is the country that once possessed nuclear weapons.

Question b: In order to analyze the problem more accurately, we will first visualize the data of nuclear weapons stockpile (nuclear-chirf-stock.csv), and then analyze which country's nuclear weapons stockpile has decreased or increased the most. Since there are different measures, two methods have been developed in order to analyze the problem more comprehensively. The first is the difference between the number of nuclear weapons in each country's last year and the number in each country's earliest year. The second approach is to look at the very different number of nuclear weapons in each country. Then the results of the two methods are compared. If they are consistent, the accuracy of the results can be better explained.

Question c: Firstly, the number of nuclear weapon tests in each year is obtained by processing the number-of-nuclear-weapons-weapons tests. Then, the number of nuclear weapon tests occurring in each "5-year" is calculated, and the five years with the highest number of nuclear weapon tests can be obtained after comparison.

Question d: By establishing CRITIC model and combining with nuclear weapon test data, we scored each country's nuclear weapon research activity, so as to get the most active country in nuclear weapon research.

Question e: We will analyze the selected countries that once possessed nuclear weapons, subtract the earliest year when status=3 and the latest year when status=0 to get the time interval for these countries to change from "not considering nuclear weapons" to "possessing nuclear weapons", and then compare them to get the result.

2.2 Analysis of Question Two

The main objective of question two is the projection of nuclear weapons, including the projection of the future nuclear-weapon States and the projection of the future number of nuclear weapons. Based on the data analysis, we choose to establish the nuclear weapon number prediction model based on ARIMA, and the establishment flow chart is as follows

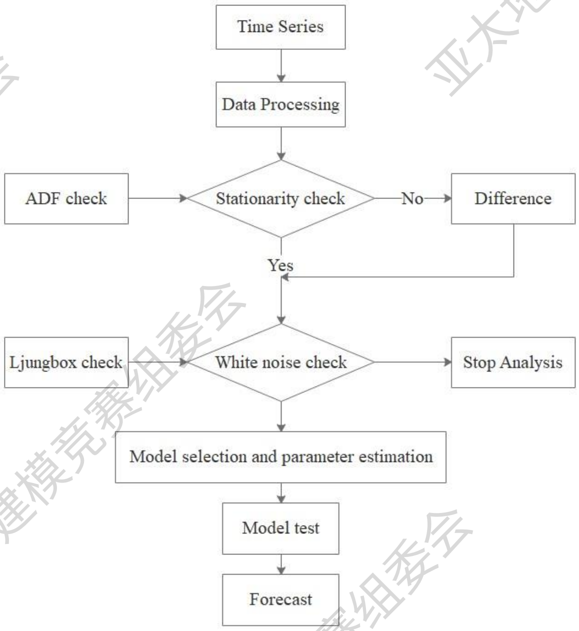

Figure 1 ARIMA model building flowchart

As can be seen from the figure, we need to preprocess the time series data first. Due to the prerequisite requirements of ARIMA model establishment, we need to conduct stationarity test and white noise test. After passing the stationarity test and white noise test, the model estimation and test are carried out, and finally the ARIMA model can be successfully established.

The number of nuclear weapons forecast model based on ARIMA is established to predict the countries with nuclear weapons in the next 100 years and the change trend of the number of nuclear weapons, and to predict the total number of nuclear weapons in the world and the number of nuclear weapons in each country in 2123.

2.3 Analysis of Question Three

There are different definitions of destroying the earth. In order to analyze the problem more comprehensively, we set up two definitions of destroying the earth. The first definition is to tear the earth apart, and on the basis of the first definition a proper analysis is carried out. The second definition is to cause serious damage to the ecological environment on the earth surface. Based on this definition, we will establish the detonation location model of nuclear weapons through DBSCAN clustering algorithm based on the population density as the standard, and then analyze and solve the problem, judge how many nuclear weapons are needed to destroy the earth, and calculate the limit of nuclear weapons in the world.

2.4 Analysis of Question Four

Based on the results of the above solution and our findings, we will write a non-technical article to the United Nations with our recommendations for all countries.

3 General Assumptions and Notations

To simplify our modeling, we make the following assumptions:

1) It is assumed that the collected data ignores the influence of information disclosure degree.
2) Suppose that the standard for the earth to be destroyed is not the earth to shatter, but the ecological environment on the earth surface to be seriously damaged.
3) The analysis of the number of nuclear weapons in this paper is limited to the countries in the collected data, and other countries are not analyzed.

Furthermore, we have a list of the main notations we define during our modeling, as shown in Table 1.

Table 1 notations

Symbol	Description
x_{ij}	The number of nuclear bombs that the i country has in the j year
δ_i	The number of nuclear bombs that the i country has possessed in the last 20 years is extremely low
$score$	CRITIC review score
U	Gravitational binding energy
R	The damage range of a nuclear weapon explosion
Eps	The minimum distance between two points
minPoints	The minimum number of points that form a dense area

4 Model and Solution of Question One

4.1 Data Collection

Question one requires a quantitative analysis of the situation related to nuclear weapons, so we study the question by collecting sufficient data. The data include four aspects: (1) data on the proliferation of nuclear weapons. This part contains two data, one is the annual proliferation of nuclear weapons in the world (nuclear-weapons.csv), and the other is the proliferation of nuclear weapons in each country (country-position-nuclear-weapons.csv). In the latter, "proliferation" is a description of the status of countries' possession of nuclear weapons. In the data, the number of states of countries' nuclear weapons has four different values, each of which has its own significance. status=0 means not considering nuclear weapons, status=1 means considering nuclear weapons, status=2 means pursuing nuclear weapons, and status=3 means possessing nuclear weapons. (2) Storage data of nuclear weapons (nuclear-warhead- stock.csv). To some extent, the stockpile of nuclear weapons can fully reflect the distribution of nuclear weapons. (3) number of nuclear weapon tests (number-of-nuclear-weapons-tests.csv). To some extent, the number of nuclear weapons tests can reflect a country's activity in nuclear weapons research and the degree of demand for nuclear weapons.

4.2 Problem Solving

4.2.1 The Solution of Problem a

In order to determine which country has ever possessed nuclear weapons, we look for data (country position-nuclear-weapons.csv) on the number of states of possession of nuclear weapons by different countries from 1938 to 2022, and traverse all the years with python. A state of 3 indicates that the country had nuclear weapons in that year. As is shown below, a total of ten countries have possessed nuclear weapons.

Table 2 Countries that once possessed nuclear weapons

Country	China	France	India	Israel	North Korea
	Pakistan	Russia	South Africa	United Kingdom	United States

South Africa, in particular, only had nuclear weapons between 1979 and 1991. Nine other countries still have nuclear weapons by 2022.

4.2.2 The Solution of Problem b

4.2.2.1 Data Visualization

Firstly, we searched the data of nuclear weapons storage (nuclear-kit-stock.csv), and visualized the data to show the changing trend from 2003 to 2022 with the help of python drawing, as shown in the figure below.

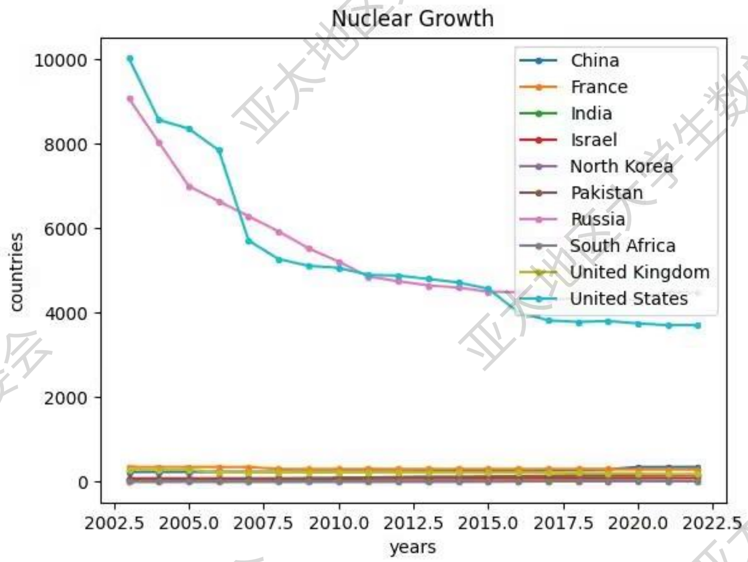

Figure 2 Nuclear growth

The image shows that the number of nuclear weapons in the United States and Russia far exceeded that of other countries in the past 20 years, and showed a rapidly decreasing trend. The number of nuclear weapons in other countries during this period was basically no more than 500.

Since the number of nuclear weapons of the United States and Russia is far more than that of other countries, in order to make the trend of the change of the number of nuclear weapons of other countries clearer, we have made the chart of the change of the number of nuclear weapons of other countries except the United States and Russia, as shown in the figure below.

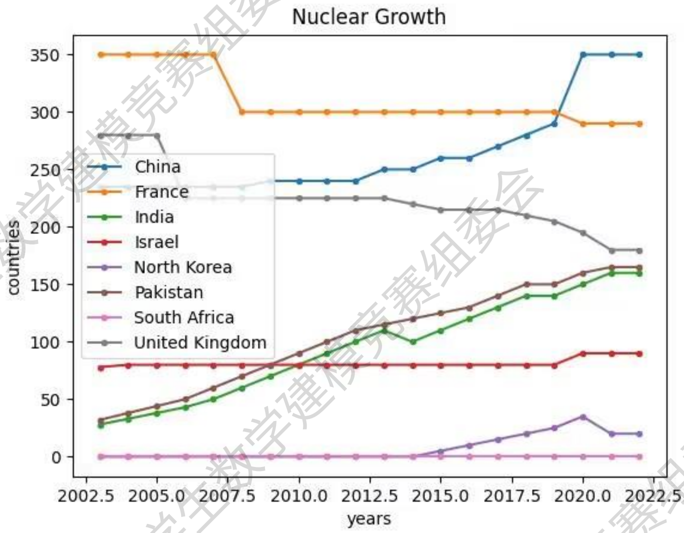

Figure 3 Nuclear growth(others)

As can be seen from the chart, the general trend of changes in the number of nuclear weapons varies from country to country. There are countries that are on an overall downward trend, and there are countries that are on an overall upward trend.

4.2.2.2 Problem Solving

To find out which country's nuclear weapons stockpile has decreased or increased the most over the past 20 years, we objectively divide it into two parts.

Method one： Will ever have nuclear weapons by the first letter sorting, respectively $x_1, x_2 \cdots x_{10}$, might as well t for years, is the first i country in the first j the number of years the bomb for x_{ij}, 20 years of this variable for $change$, among them $j = t - 2003$, $change = x_{i20} - x_{i1}$, Calculate the final result by plugging in the following formula. If the change is greater than 0, then

$$in = \max(x_{i20} - x_{i1}), i = 1, 2 \cdots 10 \tag{1}$$

If the change is less than 0, then

$$de = \min(x_{i20} - x_{i1}), i = 1, 2 \cdots 10 \tag{2}$$

in Represents the maximum increase, de represents the maximum decrease.

Method two： Note that the difference between the most and least number of nuclear bombs possessed by i countries in the past 20 years is δ_i, which is calculated as

$$\delta_i = \max\{x_{ij}\} - \min\{x_{ij}\}, i = 1, 2 \cdots 10, j = 1, 2 \cdots 20 \tag{3}$$

And compare the results with Method one, as shown in the following table.

Table 3 Variation of nuclear weapons

Country	Variation（Method one）	Range（Method two）
United States	-6319	6319
China	115	115
France	-60	60
India	132	132
Israel	12	12
North Korea	20	35
Pakistan	133	133
Russia	-4599	4776
South Africa	0	0
United Kingdom	-100	100

It is easy to see that according to Method one, the stockpile of nuclear weapons in the United States has been reduced the most, by 6319, or 60.43%. Pakistan's nuclear weapons stockpile increased by the most, by 516%, with 133 weapons. According to Method two, America has the largest range in the past 20 years. Combined with the visual chart, it is not difficult to see that America has the largest reduction during this period. The second is Pakistan, which has the largest increase in the past 20 years.

By comparing the two methods, it can be seen that the results are consistent: the United States is the country with the largest reduction in nuclear weapons inventory, while Pakistan is the country with the largest increase in nuclear weapons inventory, which can better illustrate the accuracy of the results.

4.2.3 The Solution of Problem c

By adding the test times of each country in the same year in the data table of nuclear weapon test times (number-of-nuclear-weapons-tests) through python, we get the number of nuclear weapon test times of each year in the world. Then, taking each year as the basis, we extend back for five years, respectively. Each "five years" is regarded as a group, and the number of nuclear weapon test occurrence of each group is calculated. To better visualize the change in the data, the graph below shows the change in the number of nuclear weapons tests.

Figure 4 Changes in the number of nuclear weapon tests

Due to limited space, only some groups are shown in the abscissa annotations in the figure. The figure can intuitively observe the changes in the number of nuclear weapon tests, and it can be known that over time, the number of nuclear weapon tests in the world is gradually declining.

A comparison of the total number of tests in each group gives the five years with the highest number of nuclear weapons tests, as shown in the table below.

Table 4 The five years with the highest number of nuclear weapons tests

Year	Number of nuclear weapons tests
1962	178
1963	50
1964	60
1965	58
1966	76
Total	422

It can be seen from the table that the five years with the largest number of nuclear weapon tests were 1962, 1963, 1964, 1965 and 1966, with a total of 422 times.

4.2.4 The Solution of Problem d

4.2.4.1 Data Processing

The valid data is obtained by secondary processing of the data found on the previous web address. Firstly, the following indicators are selected as the criteria to judge whether a country is active in nuclear weapons, namely "nuclear weapons status", "nuclear weapons stockpile" and "nuclear weapons tests". "the difference over ten years", "deployed strategic", "centrally deployed nonstrategic", "centrally deployed non deployed", "retired warhead". There are 8 indicators in total, which are respectively denoted as $s_1, s_2 \cdots s_n, n = 8$.

Where s_1 is the state of nuclear weapons, respectively 0,1,2,3 according to the meaning of the question to represent different degrees of state. Since s_2 and s_3 were given data in different years, we added the average value of the data into the table. In order to describe the activity degree of a country more objectively, From the initial data itself, we construct a new list of indicators called s_4, representing the difference before and after the decade. A table with 10 rows and 8 columns can be obtained by adding the following 4 indicators.

Since the order of magnitude of each index is different, we need to compare them in the same range, that is to say, all indexes need to be normalized. For convenience, we will directly standardize them here. There are *m* objects to be evaluated and *n* evaluation indicators, which can constitute the data matrix $X = \left(x_{ij}\right)_{m \times n}$. Let the elements in the data matrix be $x_{ij}^{'}$ after the positive and standardized processing of indicators. All the selected elements in this paper are positive indicators, so there is no need for positive processing.

4.2.4.2 Establishment and Solution of the Model

Here, the CRITIC method is adopted to evaluate and score these countries, and the country with the highest score is selected as the most active country. This method considers the comparison intensity of evaluation indicators and the conflict between indicators to comprehensively measure the weight of indicators, while taking into account the correlation between indicators. In essence, it means that the larger the number is, the greater the weight is not assigned. Instead, it makes full use of the attributes of data to evaluate projects, which is an objective and scientific weighting method.

The first point to be considered here is comparison intensity, which refers to the value gap between the same index and other evaluation schemes, that is, to consider the variability between different indicators, which is expressed by standard deviation. In other words, the higher the standard deviation, the greater the fluctuation, the higher the weight should be. The specific calculation is as follows:

$$\begin{cases} \overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ S_j = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2}{n-1}} \end{cases} \tag{4}$$

Next, we consider the conflicts among indicators, which can be expressed by the correlation coefficient R_j. The calculation method is as follows:

$$R_j = \sum_{i=1}^{p}\left(1 - r_{ij}\right) \tag{5}$$

When there is a strong positive correlation between two indicators, the smaller the conflict is, the lower the weight will be. For the CRITIC method, when the standard deviation is constant, when the degree of positive correlation between the two indicators is greater, that is, when the correlation coefficient is close to 1, the conflict will reach the minimum, which indicates that the information reflected by the two indicators in the evaluation scheme has great similarity.

Finally, calculate the amount of information C_j, which takes the product of the conflict and variability of the previously calculated indicators and actually represents the objective proportion of the j evaluation indicator subchild in the whole evaluation system. The larger the value, the more weight should be assigned, so it is positively correlated with the final weight.

$$C_j = S_j \sum_{i=1}^{p}\left(1-r_{ij}\right) = S_j \times R_j \tag{6}$$

Finally, the objective weight is calculated as follows:

$$W_j = \frac{C_j}{\sum_{j=1}^{p} C_j} \tag{7}$$

Add the weights and indices here to get the final score

$$score = \sum_{j=1} w_j * x_{ij} \tag{8}$$

Here we use MATLAB calculation to write corresponding programs to solve this problem, the final score is as follows

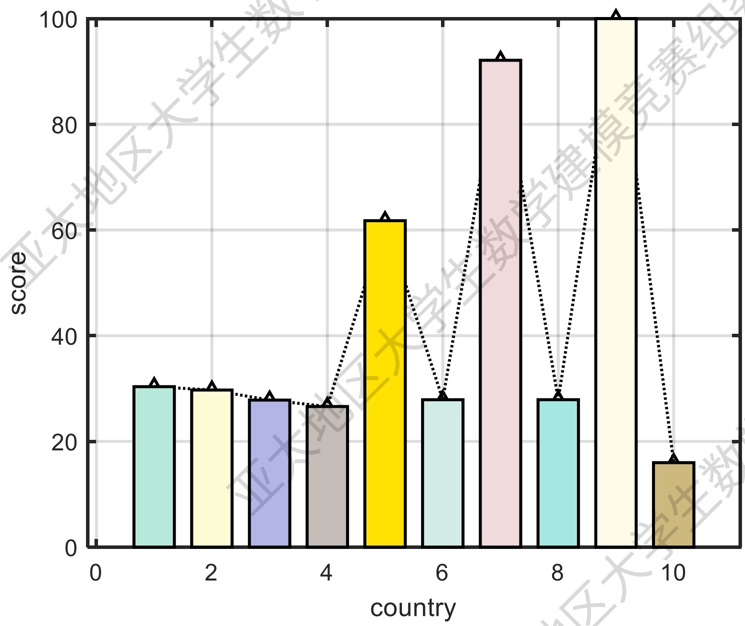

Figure 5 CRITIC score bar chart

The figure above intuitively shows the rating of countries by CRITIC, among which three countries have significant scores and others have low scores. The table below shows the specific scores of these countries.

Table 5 CRITIC rating result

Country	Evaluation score	Country	Evaluation score
North Korea	100.0000	Indian	27.8934
China	92.1177	Israel	27.8865
France	61.7613	Russia	27.8156
Britain	30.3602	United States	26.6016
Pakistan	29.7192	South Africa	15.9810

According to the table, North Korea received the highest score of 100.0. Although North Korea has a small stockpile of nuclear weapons, it has increased its nuclear weapons significantly in recent years, has a large stockpile of nuclear weapons, has tested more nuclear weapons, and scored the highest score of 100.0. So overall, North Korea has been the most active in nuclear weapons research over the past decade.

4.2.5 The Solution of Problem e

We searched the data of different countries' possession states of nuclear weapons from 1938 to 2022 (country-position-nuclear-weapons.csv), screened out 10 countries from 3.2.1 for analysis, and iterated all the years through python. When the status quantity is 0, it means that the country does not consider researching nuclear weapons in that year (status=0); when the status quantity is 3, it means that the country has possessed nuclear weapons in that year (status=3). In these countries, we subtract the earliest year when status=3 appears from the latest year when status=0 to obtain the time interval for these countries to change from "no consideration of nuclear weapons" to "possessing nuclear weapons", and the results are shown in the table below

Table 6 The time interval of transformation

Country	Year	Experience time（year）
China	1954~1964	10
France	1953~1960	7
India	1963~1987	24
Israel	1954~1967	13
North Korea	1979~2006	27
Pakistan	1971~1987	16
Russia	1942~1949	7
South Africa	1973~1979	6
United Kingdom	1940~1952	12
United States	1941~1945	4

As can be seen from the table, the United States experienced the fastest transition from "not considering nuclear weapons" to "possessing nuclear weapons" in only four years, from 1941 to 1945.

5 Model and Solution of Question Two

5.1 Prediction Model of Nuclear Weapon Number Based on ARIMA Time Series Analysis

5.1.1 Prediction Model Based on ARIMA Time Series

A time series is a sequence of numbers in chronological order that shows the values of different periods of time. ARIMA model is an analysis of random time series, its core is the probability analysis of the sequence itself, that is, the random attribute. ARIMA(p,d,q) model is used to establish a stationary sequence, where d is the order of difference, AR is autoregressive, and p is the corresponding autoregressive term. MA is the moving average and q is the number of moving average terms.

The ARIMA model will capture a series of different standard time structures in the time series, and the ARIMA model can separate noise and signal, and deduce them, so as to obtain accurate prediction results, especially suitable for the fitting of non-stationary data. The mathematical expression of ARIMA model is as follows:

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)(1-L)^d X_t=\alpha_0+\left(1+\sum_{i=1}^{q}\beta_i L^i\right)\varepsilon_t \tag{9}$$

$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)$ is $\mathrm{AR}(p)$, $(1-L)^d X_t$, $(1-L)^d X_t$ is d Degree Differential, $\left(1+\sum_{i=1}^{q}\beta_i L^i\right)\varepsilon_t$ is $\mathrm{MA}(q)$ .The ARIMA model is $X_t \sim \mathrm{ARIMA}(p,d,q)$.

5.1.2 Stationarity Test

Before constructing the ARIMA model, the stationarity test of the sequence should be carried out, and the following three conditions should be met to determine that $\{x_t\}$ is covariance stationarity:

1) $E(x_t)=E(x_{t-s})=u$ (The mean is a constant)
2) $Var(x_t)=Var(x_{t-s})=\sigma^2$ (The variance is constant)
3) $\mathrm{cov}(x_t,x_{t-s})=\gamma_s$ (The covariance is only somewhat related to the interval s)

If any $t_1,t_2,\cdots,t_k$ (k is an arbitrary value) and h are satisfied, the multidimensional random variables $(x_{t_1},x_{t_2},\cdots,x_{t_k})$ and $(x_{t_1+h},x_{t_2+h},\cdots,x_{t_k+h})$ have the same joint distribution, then $\{x_t\}$ can be considered to be in a strictly stationary condition.

Logarithmic sequence of the number of nuclear weapons of each country is taken for ADF test, and the results are shown in the table below.

Table 7 ADF test results

Variable name	t statistic	1%	5%	10%	p value	Inspection result
Num	-2.2912	-3.8724	-3.3958	-3.2132	0.1379	unstable
Δ*Num*	-12.3285	-2.6123	-1.8921	-1.6037	0.0000	steady

As can be seen from the table, the test statistics after the difference of the sequence show significance at the level of 1%. Therefore, the sequence is stable, and the ARIMA model can be used to predict the number of nuclear weapons in each country.

5.1.3 White Noise Test

In this part, we carry out white noise test on the sequence to further explore the rationality of the model. If it is white noise, it means that the sequence is extremely random, and the historical situation has nothing to do with the future development, so there is no need for more research. If the sequence is not white noise, the model can be used for quantitative prediction. In order to prevent the data from being full of randomness, this paper must detect the white noise of the residual sequence. Its essence is to judge whether the data has autocorrelation, and use autocorrelation graph to further test whether the sequence data has validity.

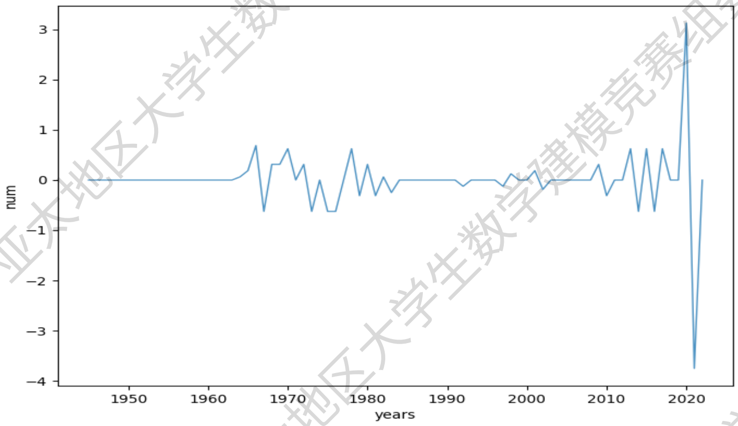

Figure 6 White noise

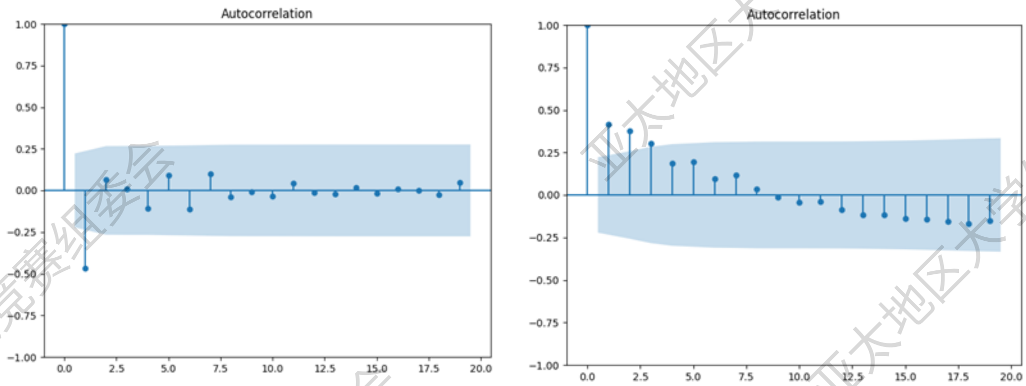

Figure 7 Autocorrelation(left) and Partial Autocorrelation(right)

Autocorrelates and partial autocorrelates are bounded by light blue bars in the graph, and the length of the sequence may be denoted as L. Since the data follows the normal distribution, it can be considered that the data is autocorrelated. Meanwhile, it is found that the characteristics of the data and white noise are inconsistent in many places. Therefore, the data successfully passes the white noise test, so the ARIMA model can be established.

5.1.4 Model Estimation and Test

We substituted the data set into the model for repeated training, so as to obtain the results meeting the error requirements, as shown in the figure below

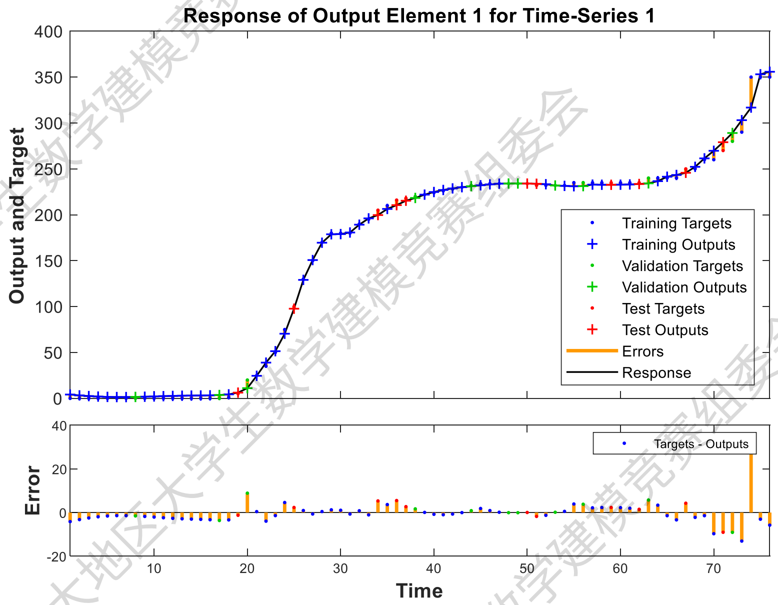

Figure 8 Error analysis diagram

As can be seen from the figure, the errors are generally within a reasonable range and are always distributed around 0, which indicates that the model has high accuracy and can obtain effective analysis results. After repeated experiments, the optimal model is determined. In the ARIMA(2,1,2) model, AIC and BIC were the smallest, and the best model ARIMA(2,1,2) with better fitting effect was obtained after experimental comparison. The specific model is as follows:

$$\Delta Num_t = 1.3124\Delta Num_{t-1} - 0.6524\Delta Num_{t-2} - 1.0058\mu_{t-1} + 0.4632\mu_{t-2} \tag{10}$$

Where, Num_t represents the sequence of the number of nuclear weapons after logarithm. μ_t indicates the random error term.

Because of

$$\begin{gathered}\Delta Num = dNum = Num - Num(-1) \\ \Delta Num = (1-L)Num_t\end{gathered} \tag{11}$$

namely

$$(1-L)Num_t = 1.3124(1-L)Num_{t-1} - 0.6524(1-L)Num_{t-2} - 1.0058\mu_{t-1} + 0.4632\mu_{t-2}$$

$$Num_t = 2.3121Num_{t-1} - 1.8947Num_{t-2} + 0.6524Num_{t-3} - 1.0079\mu_{t-1} + 0.4632\mu_{t-2}$$

In order to ensure the correctness and applicability of the model, the correlation test of residual error is carried out in this paper. By running python, it is known that the values of AC and PAC both tend to be close to 0, indicating that the residual sequence of ARIMA(2,1,2) mode has no autocorrelation, which means that the residual is white noise.

5.2 The Number of Nuclear Weapons and National Projections

After the time series successfully passed the stationarity test and white noise test, this paper established the nuclear weapon number prediction model based on ARIMA. By building a good model to predict the countries possessing nuclear weapons in the next 100 years, the specific results are shown in the figure below:

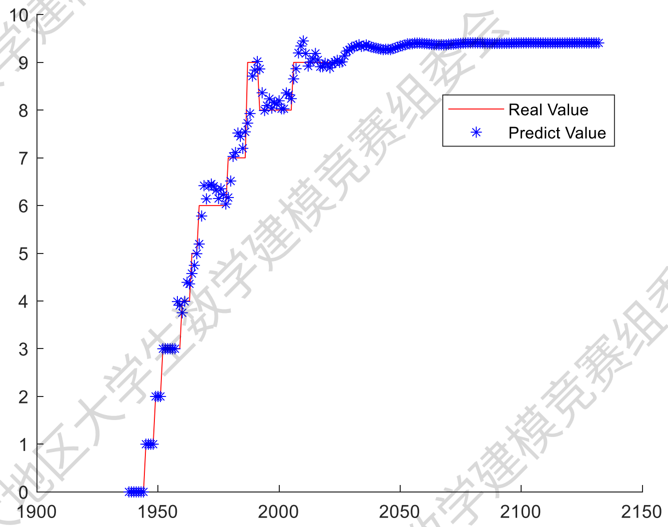

Figure 9 The predictions of nuclear-weapon states

As can be seen from the figure, with the passage of years, the number of states possessing nuclear weapons gradually rose to 9.62 and then leveled off, no longer increasing significantly, which is also related to the current environment of nuclear non-proliferation.It is easy to know that under the Nuclear Ban treaty, nine of the future nuclear weapon states should be the same as the current nine nuclear weapon states: China, France, India, Israel, North Korea, Pakistan, Russia, South Africa, United Kingdom, United States. Because of 9.62>9.5, hence the possibility of a tenth nuclear-weapon state in the future. Based on data analysis, it is presumed to be one of three countries: Brazil, Iran and South Africa. Brazil and Iran are pursuing nuclear weapons, so the possibility is high. There is also a strong possibility that South Africa has a certain research base because it once possessed nuclear weapons.

We built a good model to predict the number of nuclear weapons in the next 100 years, the total number of nuclear weapons in 2121. In order to more intuitively show the change of the number of nuclear weapons in the next 100 years, python is used to draw a graph to show the change trend of the number of nuclear weapons in the next 100 years, as shown in the figure below.

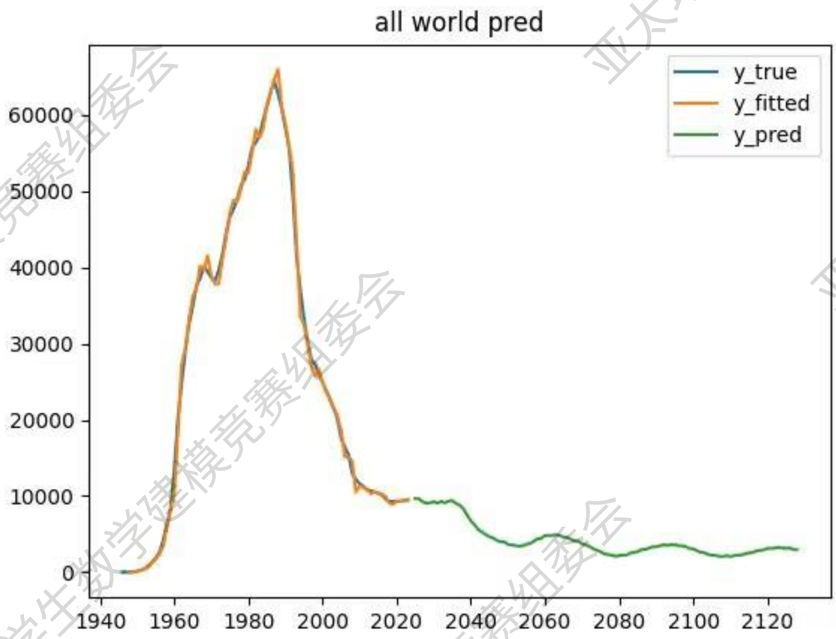

Figure 10 Trends in the number of nuclear weapons

The model predicts a total of 2,152 nuclear weapons in 2123. As can be seen from the figure, the real value is close to the predicted value, which indicates that the prediction effect of the model is good. As you can see from the graph, the number of nuclear weapons over the next 100 years will go up and down, but overall it will go down. The ban on nuclear weapons testing promoted in the Comprehensive Nuclear-Test-Ban Treaty adopted by the United Nations General Assembly has limited the growth in numbers to some extent, while the total number of nuclear weapons will continue to decline as nuclear weapons are decommissioned and dismantled, which is consistent with our model.

Since these countries have the same forecasting methods, we will use China as an example to analyze the number of nuclear weapons in China in the next 100 years.

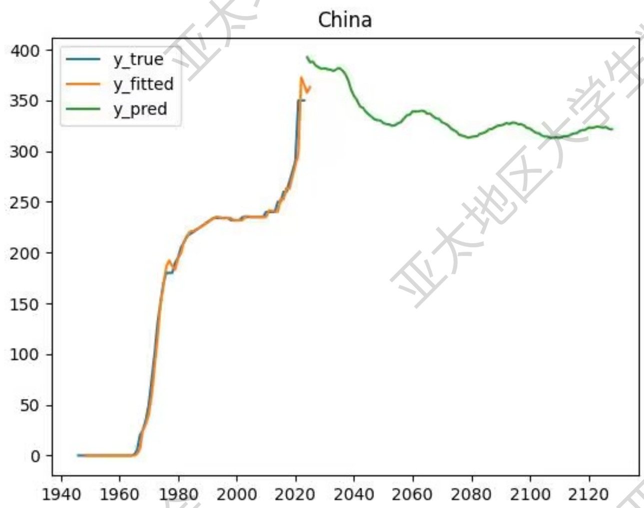

Figure 11 China's nuclear weapons numbers forecast

As can be seen from the figure, the number of nuclear weapons in China will generally show a downward trend in the next 100 years and gradually stabilize to 323.

6 Model and Solution of Question Three

6.1 Measurement of Nuclear Weapon Energy

（1）A brief introduction to nuclear Weapons

Nuclear bombs, also known as nuclear weapons, essentially use nuclear radiation, thermal radiation, shock waves and pulses to create enormous damage and radiation to the surrounding environment.

Nuclear weapons can be divided into tactical and strategic nuclear weapons. The use of tactical nuclear weapons is to detonate nuclear weapons in the strategic vanguard area of the target enemy, while the strategic nuclear weapons are to detonate intercontinental missiles in the important places far away from the enemy. The modes of occurrence are generally divided into sea, land and air. In contrast, launching in the sea is the most effective, directly reflected in the sea can reduce the likelihood of enemy detection.

（2）Nuclear weapon power measure

Equivalent is an important unit that reflects the power of nuclear weapons. It means the energy of nuclear weapons explosion is equivalent to the energy of TNT explosion. The most powerful nuclear weapon ever developed was the Soviet Czar Bomb, which had a yield of 50 megatons of TNT. So far, the total yield of nuclear weapons possessed by all countries adds up to about 15 billion tons. That's still a far cry from the 250 trillion tons produced when the planet hit Earth.

6.2 Definition of Earth Destruction

We currently define the destruction of the planet in two ways:

① Rocks were gravitationally bound together to form the Earth, so destroying the Earth would mean breaking it up into pieces of rock that would not be gravitationally bound back together.

② From the point of view of biological survival, the destruction of the earth means that the ecological environment and living conditions on the surface of the earth have been seriously destroyed, even can be said to be razed to the ground, resulting in the destruction of the earth is impossible to have biological survival.

6.3 Model Establishment and Solution

6.3.1 Model Establishment and Solution Based on Definition ①

Matter in nature is held together by its attraction to each other. If a body contains more than one part, the parts form a whole because of the gravitational interaction between them. If you want to separate the parts, you need to do external work. And the work needed to overcome gravity increases as the force of gravity increases. Binding energy means the total energy of the different parts of an object bound together.

In the nucleus as a whole, protons and neutrons are bound together by the strong nuclear force, and the binding energy of these microscopic particles together is the nucleus binding energy. Similar to that is the gravitational binding energy, which holds the Earth together from matter in a stable state. So to blow up the Earth, the energy from the explosion has to be greater than the binding energy of gravity. The gravitational binding energy of the earth is calculated by the following formula.

$$U=\frac{\frac{3}{5}GM^2}{r} \tag{12}$$

G stands for gravity constant, $6.67\times10^{-11}\mathrm{N}\cdot\mathrm{m}^2/\mathrm{kg}^2$; M is the mass of the earth, is $5.97\times10^{24}\mathrm{kg}$; r is the radius of the Earth, which is 6371 km. Through calculation, the gravitational binding energy of the earth can be known $2.24\times10^{32}\mathrm{J}$.

The total number of nuclear weapons currently in existence is about 9,440, of which the atomic bomb is the main. According to the literature, the yield of the atomic bomb is relatively small, only a few tons, while the yield of the hydrogen bomb is measured in megatons. If all the nuclear bombs in the world exploded at the same time, it would generate the energy of $3.14\times10^{19}\mathrm{J}$, and it would take at least 7 trillion times more energy to destroy a planet on the basis of definition ①.

According to the descriptive statistical analysis of the data above, it can be known that there are 9,440 nuclear weapons in the world at present. The most powerful nuclear weapon known to have ever been developed was the Soviet Czar Bomb, which had an explosive yield of about 5 kilotons. Since the emphasis in the title is on the minimum number of nuclear weapons needed, we can assume that all existing nuclear weapons are the strength of Czar bombs. Based on this assumption, the formula for calculating the sum of the current explosive energy of all nuclear weapons is as follows

$$W=Num\cdot m_{\mathrm{TNT}}\cdot W_{\mathrm{TNT}} \tag{13}$$

In the formula, Num represents the total number of existing nuclear weapons; m_{TNT} represents the yield of the assumed nuclear weapons; W_{TNT} represents the energy released when one ton of TNT is exploded.

Since the current total number of nuclear weapons is 9440, the yield of Czar bomb is 5000 tons, and the energy released by one ton of TNT explosive is about 4.18×10^{9} J, the total energy of nuclear weapons in the world can be obtained by substituting it into the formula: $W=1.97296\times10^{17}$ J. The earth's gravitational binding energy 2.24×10^{32} J, by comparison, we can see that the total energy of global nuclear weapons and the earth's gravitational binding energy is far from each other, so it is not possible to destroy the earth on the basis of the definition of ①. From another point of view, if the Earth's gravitational binding energy is converted into the energy of the Czar bombs, a total of 316 trillion Czar bombs would be needed, which is far from enough given the current level of human technology. Therefore, on the basis of definition 1, it is impossible to blow the earth apart, and the ARIMA model predicts that nuclear weapons will not blow the Earth up as the number of nuclear weapons is decreasing.

6.3.2 Model Establishment and Solution Based on Definition ②

6.3.2.1 Main Components of Nuclear Weapon Damage Radius

Based on the center of the explosion, the closest distance to the center is the radius of the fireball, that is, the central circle of the damage, followed by the detonation radius, radiation radius, and the first region of the shock wave. The power of a nuclear weapon decreases as its radius increases.

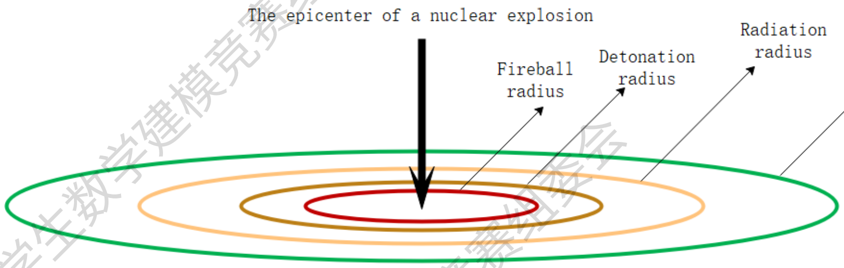

Figure 12 Nuclear explosion radius

1) The radius of the fireball is the central part of the explosion. Within this radius, the nuclear bomb is hot enough to evaporate everything. In other words, the nuclear weapon is energetic enough to burn everything within this radius.
2) The detonation radius is the second range circle outward from the central explosion point. The energy released by a nuclear weapon can severely destroy all buildings within the detonation radius. Since the radius of the fireball and the radius of the detonation are very small from the center of the explosion, the death rate of the creatures in it would be a horrible 100% if they were not hidden deep underground.
3) The radiation radius is the third circle of range outward from the center of the explosion. It's called the radius because the main thing that causes damage to living things in this area is radiation. Most living things will be seriously affected if they are exposed to severe radiation.
4) The first region of the shock wave is the fourth circle of range outward from the center of the explosion. An important component of the shock wave is atmospheric pressure, which is caused by the rapid expansion of the fireball and expands outward at a staggering speed of 200m/s, releasing the energy of a nuclear weapon explosion. After the shock wave, the expansion creates a vacuum, and the atmospheric pressure causes the atmosphere to backfill rapidly, which can cause serious damage.

The specific damage range of a nuclear weapon explosion is calculated by the following formula:

$$R = C \cdot \sqrt[3]{m_{\mathrm{TNT}}} \tag{14}$$

Where, C is the fixed value of 1.493885, and m_{TNT} is the explosive equivalent.

6.3.2.2 Nuclear Weapon Detonation Location Model Based on DBSCAN Clustering Algorithm

(1)Establishment of model

Based on the definition ② of destroying the Earth, we construct the DBSCAN clustering model to further judge the number of nuclear weapons destroying the earth. Since there are a large number of human beings living in areas with ecological environment and living conditions, we will perform DBSCAN clustering based on population density to obtain the total number of clustering categories, and the number of clustering is the number of nuclear weapons that destroy the earth. In order to visually display the situation of population density, we make the following figure to show.

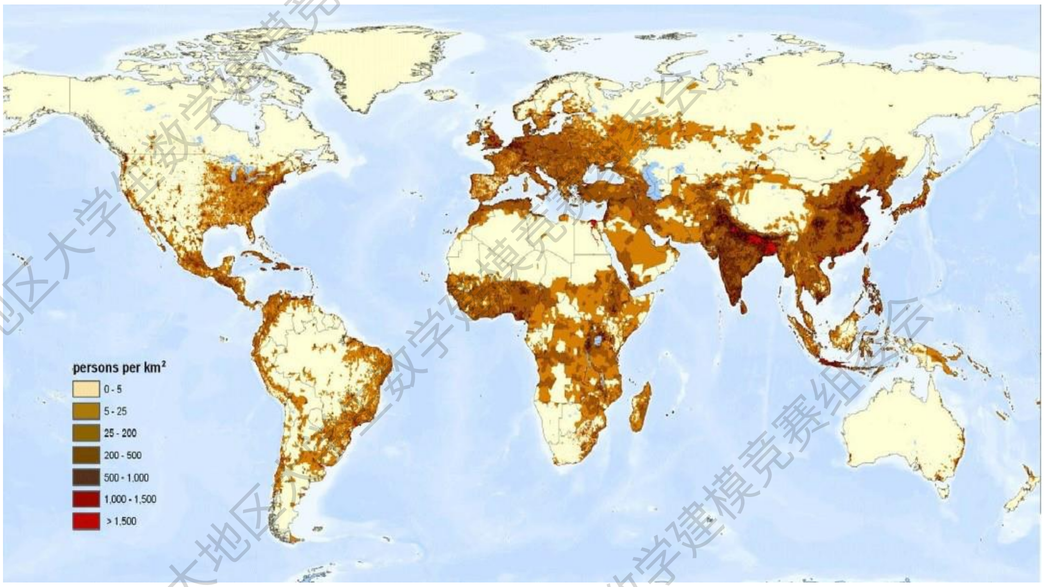

Figure 13 Global population density

The above figure is the distribution of global population density. We converted the above figure into a gray scale of 60000*10000 by using python as follows

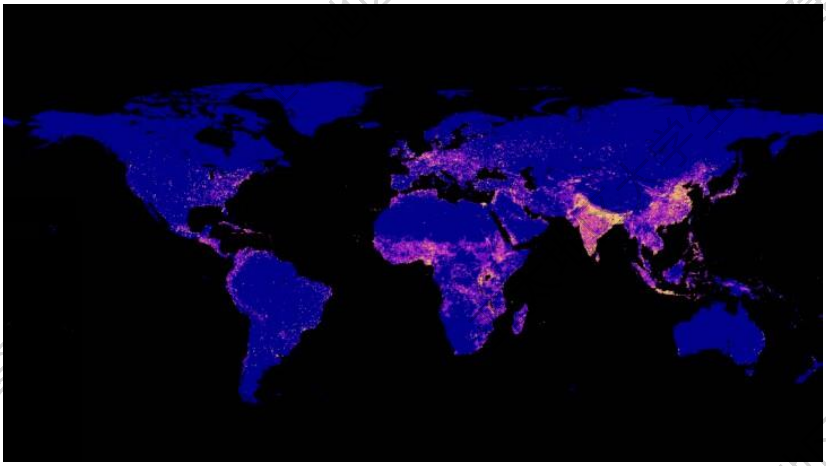

Figure 14 Grayscale map

The above figure is the gray level after the population density distribution map is converted, from which the general situation of population density distribution can be seen. Because we consider irreversible damage to the human and much of the ecological environment of the Earth as a way to destroy the planet, we need to calculate the range and number of nuclear explosions needed to cover the entire planet.

DBSCAB is a density-based spatial clustering method. It does not need to determine the number of clusters in advance, and the number of clusters generated is random. The algorithm adopts density-based clustering method, and the number of targets must be contained in the clustering space should not be lower than the preset threshold. Even if there are noise points in the spatial database, the algorithm can still find clusters of arbitrary shapes. Large density areas close to each other can be connected by this algorithm, which can improve the effectiveness of data processing.

DBSCAN clustering algorithm has three types of data points:

Core points: With the core point as the center of a circle and Eps as the radius, the number of points in Eps is not less than minPoints

Boundary points: With boundary points as the center of the circle and Eps as the radius, the number of points in Eps is less than minPoints, but they are in the Eps neighborhood of the first type of data points

Noise points: With noise points as the center of the circle and Eps as the radius, the number of points in Eps is less than minPoints, and they are not in the Eps neighborhood of the first type of data points

Eps represents the minimum distance between two points. When the distance between two points is $d \le Eps$, they are considered to be adjacent. Eps selection is based on data sets. minPoints indicates the minimum number of points to form a dense area.

The flowchart of DBSCAN clustering algorithm is as follows

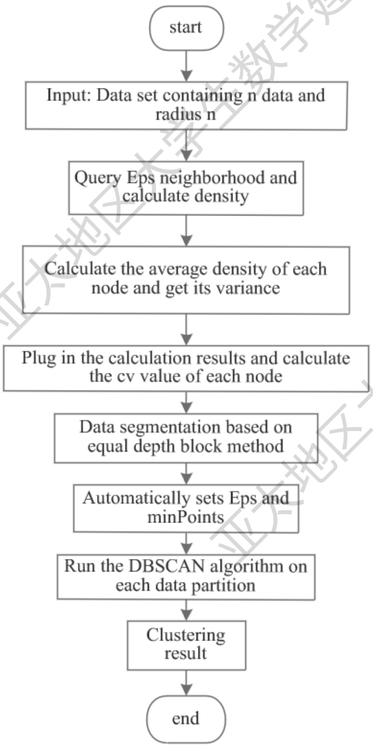

Figure 15 DBSCAN clustering algorithm flowchart

Based on DBSCAN clustering algorithm, this paper combined with the idea of density detection for analysis.

Density:

Step1：Density formula of node i:

$$\rho_i = \frac{|Pts(i)|}{\pi \cdot Eps^2} \tag{15}$$

Where Eps is the radius, $Pts(i)$ is the point set with center i and radius Eps, and $|Pts(i)|$ is the number of elements.

k is other points in the neighborhood of node i, whose density is defined as

$$\rho_k = get\_\rho\left(Pts_k\right), Pts_k \in Pts(i) \tag{16}$$

In the circle neighborhood, k is the center of the circle and Eps is the radius

Step2：Mean density

$$\bar{\rho}_i = \frac{\sum_{k=1}^{|Pts(i)|} \rho_k}{|Pts(i)|} \tag{17}$$

Step3： Density variance reflects the degree of deviation between each point in the Eps neighborhood and the mean value

$$s^2 = \frac{1}{n-1}\left[\sum \rho_i^2 - n \cdot \left(\bar{\rho}_i\right)^2\right], \quad n = |Pts(i)| \tag{18}$$

Step4：Coefficient of density variation, labeled in the neighborhood. The value of node density decreases with increasing number of points.

$$cv_i = \frac{s}{\bar{\rho}_i}, s = \sqrt{s^2} \tag{19}$$

Automatic Eps parameter setting for each partition:

The Eps of each region is defined as the average distance between each point and the previous point within its neighborhood

① Determine the value of minPoints

② $|Pts'(i)| = \rho_i \cdot \pi \cdot Eps^2$ Calculate the number of nodes in the Eps neighborhood of each point

③ $Eps(i) = \frac{|Pts'(i)|}{MinPts} \cdot Eps$

④ $Eps(i)$ is averaged after summation， $E(Eps) = \frac{\sum_{i=1}^{n} Eps(i)}{n}$

(2) Solution of model

The nuclear weapon detonation location model based on DBSCAN clustering algorithm was established to solve the problem, and the results as shown in the following figure were obtained

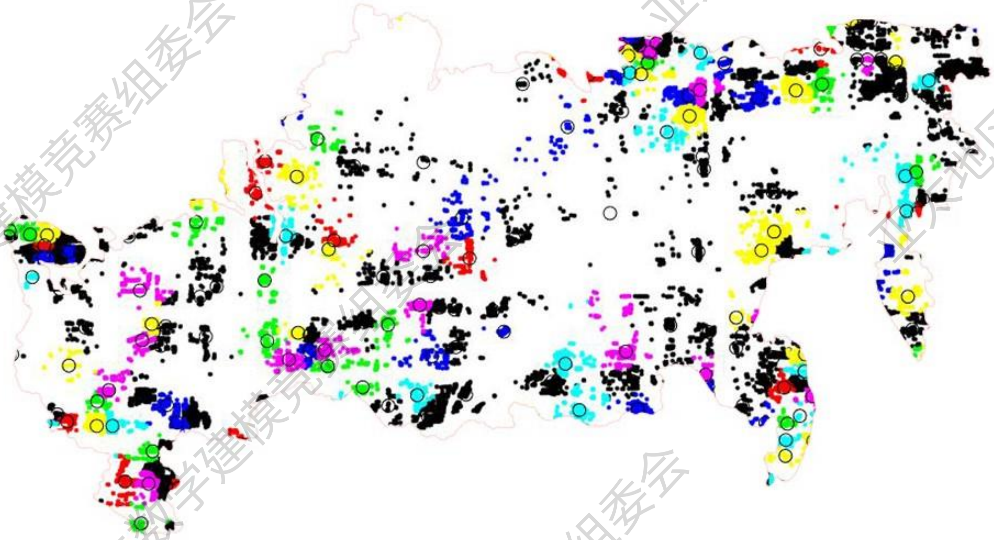

Figure 16 DBSCAN Cluster result

The graph above shows the population density of one of these countries. The same color represents the range of nuclear weapons and the range of nuclear radiation. According to the program, 5752 categories of clustering are obtained. Therefore, it is equivalent to 5752 Czar bombs or 14300 ordinary nuclear bombs are needed to destroy the earth on the basis of definition ②, which is not enough to completely radiate and contaminate the whole earth with the current scientific and technological level. With the development of science and technology, the power of nuclear weapons will continue to increase in the future, and the total number of nuclear weapons destroyed by the earth will decrease. Therefore, according to the results of the model, we believe that the total number of nuclear weapons in the world should be limited to 5,000 in theory, which is conducive to protecting the ecological environment. Russia and the United States have far more nuclear weapons than other countries, and the number of nuclear weapons in other countries is relatively small. Therefore, Russia and the United States are important factors in destroying the earth. Therefore, the first thing we should do is strictly limit Russia and the United States to 2000, and the rest of the countries cannot grow any more. Only on this basis can human beings further promote sustainable development and effectively protect the ecological environment.

7 A Non-Technical Article

Nuclear Weapons, Restriction or Development

Our findings:

In the 1960s, many countries conducted simultaneous high-volume nuclear tests to consolidate their dominance. Currently, there are about 10 countries with nuclear weapons in the world, and the number of nuclear bombs is negligible except for the P5, of which the United States and Russia have the most. To this day, many countries are still pursuing vigorous development of nuclear weapons.

From the prediction results of our model, even if all the harsh criteria are met, the energy generated by the explosion of all the existing nuclear bombs in the world will not be able to completely destroy the whole earth. However, if countries are allowed to develop nuclear weapons, once a large-scale war breaks out, it will cause enormous damage to the ecology of the entire planet. Therefore, it is imperative that we limit the number of nuclear weapons in the countries that possess them and change the direction of their development and competition.

Our suggestions:

From a military perspective, a nuclear threat policy is usually a policy that uses the power of nuclear weapons to threaten other countries. In the last century, countries have developed nuclear weapons to protect their territories. However, in the future, if countries still have conflicting interests, nuclear weapons will continue to exist due to their deterrent effect. Nuclear weapons will become the core military power of countries and promote the formation of mutual checks and balances among them. Therefore, in order to reduce the threat of nuclear weapons to human beings and the ecological environment, it is necessary to conclude agreements between countries to reduce the number of nuclear weapons in each country within a certain limit and to limit the use of nuclear weapons.

From the point of view of technological development, nuclear weapons are a great technological breakthrough in the history of mankind. In the 21st century, we should change the direction of its development and devote ourselves to the use of this highly sophisticated technology in agriculture, industry and medicine for the benefit of mankind. In order to achieve a sustainable planet and better use of resources in the future, we should also continue to study the theory and application of nuclear technology with the primary responsibility of protecting the ecological environment. We have always believed that the ultimate goal of developing nuclear weapons is to eliminate them.

8 Model Evaluation and Further Discussion

8.1 Strengths

1 Time series is a series of numbers obtained by arranging the total changes of a variable in time based on the sequence of time. All the information of the historical behavior that produced the sequence is included in the sequence, and the historical information can be used to predict the development trend through the time series model, so the prediction of the number of nuclear weapons in this paper is more accurate.

2 The ARIMA model, which requires only endogenous variables and does not require exogenous variables, is adopted in this paper, so it has a good effect on the long-term prediction of the number of nuclear weapons.

8.2 Weaknesses

1 As the degree of information disclosure has certain limitations on model building, the accuracy of prediction will be affected to some extent.

8.3 Further Discussion

To limit the number of nuclear weapons, we can analyze from the living standards of residents, especially ecological and environmental carrying capacity and other factors. These factors are important constraints on the number of nuclear weapons. We can do further analysis using the Vensim cause-and-effect flow chart to get a limit on the number of nuclear weapons.

Reference

[1] Hargreaves, Ben. Nuclear explosion.[J]. Professional Engineering, 2010.
[2] Thurber C , Trabant C , Haslinger F , et al. Nuclear explosion locations at the Balapan, Kazakhstan, nuclear test site: the effects of high-precision arrival times and three-dimensional structure[J]. Physics of the Earth & Planetary Interiors, 2001, 123(2):283-301.
[3] C, H, Thurber, et al. Accurate locations of nuclear explosions in Balapan, Kazakhstan, 1987 to 1989[J]. GEOPHYSICAL RESEARCH LETTERS, 1993, 20(5):399-402.
[4] Latter R , Karzas W J . Electromagnetic Radiation from a Nuclear Explosion in Space[J]. Phys Rev, 1962, 126(6):40.
[5] Karzas W J , Latter R . Detection of the Electromagnetic Radiation from Nuclear Explosions in Space. 1964.
[6] KeylisBorok, V.I, Malinovskaya, et al. DETECTION OF NUCLEAR EXPLOSIONS[J]. Trudy Instituta Fiziki Zemli Akademiya Nauk Sssr, 1962, vol: no. 20.
[7] Lars, Nielsen, Hans, et al. Origin of upper-mantle seismic scattering – evidence from Russian peaceful nuclear explosion data[J]. Geophysical Journal International, 2003.

Appendix

Appendix 1： Python program for problem a

```
import pandas as pd

def CountriesThatHaveNuclearBombs():
    data = pd.read_csv('../data/country-position-nuclear-weapons.csv').values.tolist()
    countries = {}
    for i in data:
        if i[3] == 3:
            countries[i[0]] = i[2]
    return countries

if __name__ == '__main__':
    res = CountriesThatHaveNuclearBombs()
    for item in res:
        print(item + ' has nuclear until ' + str(res[item]) + ' year')
```

Appendix 2： Python program for problem b

```
import pandas as pd
import matplotlib.pyplot as plt

def draw(countries, years, datas):
    plt.title('Nuclear Growth')
    plt.xlabel('years')
    plt.ylabel('num')
    for data in datas:
        plt.plot(years, data, marker='o', markersize=3)
    plt.legend(countries)
    plt.show()

if __name__ == '__main__':
    data = pd.read_csv('../data/nuclear-warhead-stockpiles.csv').values.tolist()
    datas=[]
    countries = []
    temp = []
    years=[]
    for i in range(2003, 2023):
        years.append(i)
    for i in data:
        countries.append(i[0])
        if i[2] > 2002:
            temp.append(i[3])
            if i[2] == 2022:
```

```
            datas.append(temp)
            temp = []
    countries_data = list(set(countries))
    countries_data.sort(key=countries.index)
    draw(countries_data,years,datas)
    del countries_data[-4]
    del countries_data[-1]
    del datas[-4]
    del datas[-1]
    draw(countries_data,years,datas)
```

Appendix 3：Python program for problem c

```
import pandas as pd
from matplotlib import pyplot as plt
import plotly.express as px
from matplotlib import pyplot as plt
import seaborn as sns
import statsmodels.api as sm

data = pd.read_csv('../data/number-of-nuclear-weapons-tests.csv').values.tolist()

res = []
temp = []
years = []
for i in range(1945, 2020):
    years.append(i)
# print(years)
for i in years:
    for item in data:
        if item[2] == i:
            temp.append(item[3])
    num = 0
    for j in temp:
        num += j
    res.append(num)
    num = 0
    temp = []
print(res)

res = pd.DataFrame(res, columns=['num'])
# res.index=pd.Index(sm.tsa.datetools.dates_from_range('1945', '2019'))
print(res)
fiveyears = []
fiveyearsnum = []
```

```
print(len(res.index))
for index in range(0, len(res.index) - 4):
    fiveyearsnum.append(res[index:index + 5].values.transpose().tolist()[0][0] +
res[index:index + 5].values.transpose().tolist()[0][
            1] + res[index:index + 5].values.transpose().tolist()[0][2] + \
            res[index:index + 5].values.transpose().tolist()[0][3] + res[index:index +
5].values.transpose().tolist()[0][
                4])
# +'to'+str(i+4)
[fiveyears.append(str(i)) for i in range(1945,2016)]
print(fiveyears)
print(fiveyearsnum)
plt.figure(figsize=(15, 10))
plt.plot(fiveyears,fiveyearsnum)
plt.xticks(rotation=90)
# Control chart notes
plt.annotate('The Max value is %d in %s' %
(max(fiveyearsnum),fiveyears[fiveyearsnum.index(max(fiveyearsnum))]),  # Control the
content of comments
             xy=(fiveyearsnum.index(max(fiveyearsnum)), max(fiveyearsnum)),  #
The point to comment, y1.index(max(y1)): The index of the maximum value in y1，
             # max(y1): The maximum value in y1
             xytext=(20, 400),  # The location of the comment display
             fontsize=30,
             arrowprops=dict(facecolor='black', width=0.5, headwidth=5)  # Control
arrow color, width, arrow width
             # arrowprops=dict(arrowstyle="<->") # Try this
             )
plt.show()
```

Appendix 4：MATLAB program for problem d

```
% Import data data1
function [Score,quan3]=CRITIC(data1) % comparability
the=std(data1);
%% contradiction
r=corr(data1);% Calculate the correlation coefficient between indicators
f=sum(1-r);
c=the.*f; %% Calculate information carrying capacity
% For each indicator
w=c/sum(c); % weight
quan3=w;
[n,m]=size(data1); % Calculate the score for each item
data= data1 ./ repmat(sum(data1.*data1) .^ 0.5, n, 1); % The matrix is normalized
% data=mapminmax(data1',0.002,1);% Standardize to a given interval interval
```

```
% data=data'; transpose
s=data*w';
Score=100*s/max(s);% Gain vector
end
```

Appendix 5: Python program for problem e

```python
import pandas as pd

data = pd.read_csv('../data/country-position-nuclear-weapons.csv').values.tolist()
countryname = 'none'
htime = -1
start = 0
end = 0
for item in range(0, len(data)):
    if htime == 3 and data[item + 1][0] != countryname:
        print(data[item][0], str(start) + 'to' + str(end), 'consume', end - start, 'years')
        htime = -1
        start = 0
        end = 0
        continue
    if data[item][0] != countryname:
        countryname = data[item][0]
        htime = -1
        start = 0
        end = 0

    if data[item][3] > htime:
        if data[item][3] == 1 or data[item][3] == 2:
            start = data[item - 1][2]
        if data[item][3] == 3:
            end = data[item][2]
        htime = data[item][3]
```

Appendix 6: ARIMA model python program code

```python
import datetime
import time
import tsod
import statsmodels.api as sm

import pandas as pd
import pmdarima as pm
import matplotlib.pyplot as plt
from pmdarima.arima import auto_arima
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.stattools import adfuller
```

```
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import explained_variance_score
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score

# get data
stockpiles = pd.read_csv('../data/nuclear-warhead-stockpiles.csv').values.tolist()
#    countriesindex=['China','France','India','Israel','North    Korea','Pakistan','Russia','South
Africa',
# 'United Kingdom','United States']
yearindex = [i for i in range(1945, 2023)]
# print(yearindex)
for j in yearindex:
    temp = []
    for i in stockpiles:
        if i[2] == j:
            temp.append(i[3])
    yearindex[yearindex.index(j)] = sum(temp)
# print(yearindex)
dta = pd.DataFrame(yearindex)
dta.index = pd.Index(pd.date_range('1945', '2023', freq='1Y'))
dta.plot(figsize=(12, 8))
plt.title('dta')
print('dta:', dta)

model = pm.auto_arima(dta, start_p=1, start_q=1,
                      max_p=8, max_q=8, m=1,
                      start_P=0, seasonal=False,
                      max_d=3, trace=True,
                      information_criterion='aic',
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=False)
forecast = model.predict(110)   # Forecast data for the next 10 years
print(forecast)

# Add the value of 2090 as the PredicValue first element for plot continuity
PredicValue = []
PredicValue.append(dta.values[-1])
for i in range(len(forecast)):
    PredicValue.append(forecast[i])
PredicValue = pd.Series(PredicValue)

PredicValue.index = pd.Index(sm.tsa.datetools.dates_from_range('2023', '2133'))
```

```
PredicValue=PredicValue.astype(float)

fig, ax = plt.subplots(figsize=(12, 8))
ax = dta.loc['1945':].plot(ax=ax, label='train')
PredicValue.plot(ax=ax, label='pred')
plt.legend(['train','pred'])
plt.show()
print(PredicValue)
```

Appendix 7: DBSCAN clustering algorithm MATLAB program code

```
clear all
A=xlsread('xian','B2:C1475');
B=xlsread('data','A2:B182807');
X=xlsread('data','C2:C182807');
k=612;
opts = statset('Display','final');
[cidx,ctrs]=dbscan(B,k, 'Distance','city', ...
                                    'Replicates',5, 'Options',opts);
X_Cent=ctrs(:,1);Y_Cent=ctrs(:,2);
X_Old=A(:,1);Y_Old=A(:,2);
%% Drawing code
for i=1:k
    plot(B(cidx==i,1),B(cidx==i,2),'k.')
    Point(i)=sum(X(cidx==i,1));
    hold on
end
hold on
for i=1:50
    plot(B(cidx==i,1),B(cidx==i,2),'r.')
    Point(i)=sum(X(cidx==i,1));
    hold on
end
hold on
for i=101:150
    plot(B(cidx==i,1),B(cidx==i,2),'g.')
    Point(i)=sum(X(cidx==i,1));
    hold on
end
hold on
for i=201:250
    plot(B(cidx==i,1),B(cidx==i,2),'b.')
    Point(i)=sum(X(cidx==i,1));
    hold on
end
```

```
hold on
for i=301:350
      plot(B(cidx==i,1),B(cidx==i,2),'c.')
      Point(i)=sum(X(cidx==i,1));
      hold on
end
hold on
for i=401:450
      plot(B(cidx==i,1),B(cidx==i,2),'m.')
      Point(i)=sum(X(cidx==i,1));
      hold on
end
hold on
for i=501:550
      plot(B(cidx==i,1),B(cidx==i,2),'y.')
      Point(i)=sum(X(cidx==i,1));
      hold on
end
hold on
for i=601:650
      plot(B(cidx==i,1),B(cidx==i,2),'k.')
      Point(i)=sum(X(cidx==i,1));
      hold on
end
hold on
for i=701:750
      plot(B(cidx==i,1),B(cidx==i,2),'r.')
      Point(i)=sum(X(cidx==i,1));
      hold on
end
hold on
plot(X_Cent,Y_Cent,'ko')
```