

abstract

apmcmthesis \LaTeX template is designed by <https://www.latexstudio.net> for <http://www.apmcm.org>. The template is designed to let everyone focus on the content writing of the paper, without spending too much effort on the customization and adjustment of the format.

Note that users need to have some experience with \LaTeX , at least some of the features of common macro packages, such as references, math formulas, image usage, list environment, etc. Templates have added commonly used macros. Package, no additional user added.

This template is located on <https://github.com/latexstudio/APMCMthesis>. You can update files from the repository.

Keywords: Keywords1 Keywords2 Keywords3

Contents

1. Introduction	1
2. Question restatement	1
3. Terms, definitions and symbols	2
4. Models assumptions	2
5. Solution for question 1	2
5.1 (a)	2
5.1.1 <i>Analysis of problem 1(a)</i>	2
5.1.2 <i>Solution of problem</i>	2
5.2 (b)	4
5.2.1 <i>Analysis of problem 1(b)</i>	4
5.2.2 <i>Model 1: ARIMA</i>	4
5.2.3 <i>Model 2: LSTM</i>	6
5.3 (c)	7
5.3.1 <i>Result from model ARIMA</i>	7
5.3.2 <i>Result from model LSTM</i>	8
5.4 (d)	8
5.4.1 <i>Analysis of model ARIMA</i>	8
5.4.2 <i>Analysis of model LSTM</i>	9
6. Solution for question 2	9
6.1 (a)	9
6.1.1 <i>Analysis of problem 2(a)</i>	9
6.1.2 <i>Relationship about location</i>	9
6.1.3 <i>Relationship about time</i>	10
6.2 (b)	11
6.2.1 <i>Relationship analysis</i>	11
6.3 (c)	13
6.3.1 <i>Data of more factor</i>	13
6.3.2 <i>Analysis of relationship</i>	13
6.4 (d)	13
7. Solution for question 3	14
8. Future Work	14
8.1 Another model	14
8.1.1 <i>The limitations of queuing theory</i>	14
8.1.2	14
8.1.3	14
8.1.4	14
9. References	14
10. Appendix	15

I. Introduction

In order to indicate the origin of problems, the following background is worth mentioning.

Since this year, we have seen a large number of amazing temperature reports. There is no doubt that the earth is burning. Following the terrible high temperature in these areas from the end of June to the beginning of July, Italy set a European temperature record again, reaching an astonishing 48.8 ° C, and many countries declared a state of emergency.

Global warming is a phenomenon related to nature. It is precisely because of the continuous accumulation of the greenhouse effect that the energy absorbed and emitted by the Earth's atmospheric system is unbalanced. The continuous accumulation of the energy of the Earth's atmospheric system leads to temperature rise and global warming.

Before the industrial revolution, the carbon dioxide (CO₂) in the atmosphere was about 280 parts per million. In March 2004, the atmospheric CO₂ concentration reached 377.7ppm, the largest average increase in 10 years so far. According to the statistics of scientists from the National Oceanic and Atmospheric Administration (NOAA) and the Scripps Institution of Oceanography (SIO), the monthly average concentration of CO₂ reached a peak of 421ppm in May 2022. An OECD report predicted that the CO₂ concentration would reach 685ppm by 2050.

II. Question restatement

1) Do you agree with the global temperature statement? Data given in attachment and other data sets collected by the team analyze changes in global temperature. Are you suggesting that the rise in global temperature in March 2022 has caused more growth than any decade ago. Based on historical data, please develop two or more mathematical models to show the past and predict the future level of global temperature. Use each model at 1(b) to estimate global temperature in 2050 and 2100. Which model of 1(b) do you think is the right one, and why.

2) What are the causes that affect temperature change? Use the results of question 1 and annex Data. Data from other data sets collected by csv and your team, developed a mathematical model to verify the relationship between global temperature, time and location (whatever), and explain the relationship between them or prove that there is no relationship between them. Collect native data and check actions that have an impact on natural disasters (such as volcanic eruptions, forest fires and COVID-19). Does it affect global temperature? What do you think is the main reason that is affecting global temperature change? Do you think there are steps that can stop or slow global warming?

3) Prepare a non-technical article (mostly on 1 page). Write a non-technical article (mostly on 1 page) to the APMCM Organization Committee to explain your team's findings and suggestions for the future. The PDF solution with a total number of not more than 25 pages contains: a full table of one page; and complete solution catalogue; A non-technical article page.

III. Terms, definitions and symbols

IV. Models assumptions

V. Solution for question 1

In whole question 1, we need to analyze global temperature changes using dataset in the attachment and other datasets we collect.

5.1 (a)

5.1.1 *Analysis of problem 1(a)*

In this question, we need to judge if the increase of global temperature in March 2022 resulted in a larger increase than observed over any previous 10-year period. So we can plot the global temperature in 2022. Firstly, consider if global temperature in March 2022 is an abnormal value. Then calculate the average temperature between 2011 to 2021 and its rising rate. We let them be visual and analyze their trend and differences. Finally we conclude a judgement of the statement.

5.1.2 *Solution of problem*

We download global temperature dataset from <https://berkeleyearth.org/data/>. And we choose land only monthly average temperature to plot the figure. The picture about annual average temperature between 1760 to 2022 is the following.

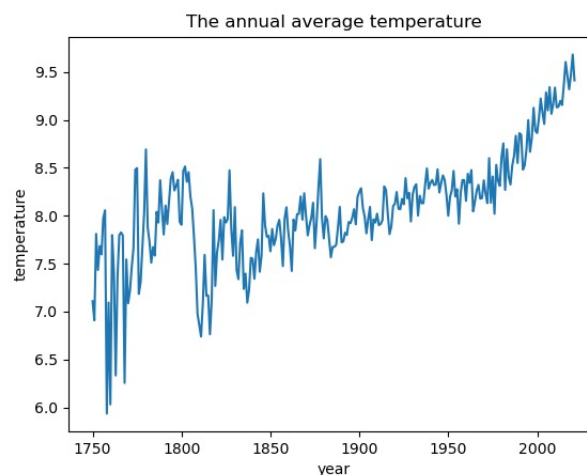


Figure 1 picture 1

Then we get the average temperature in March between 2011 to 2021 is the following.

[6.0780, 5.9890, 6.2030, 6.3420, 6.7220, 7.4750, 7.1340, 6.6680, 7.0950, 7.0450, 6.6840].

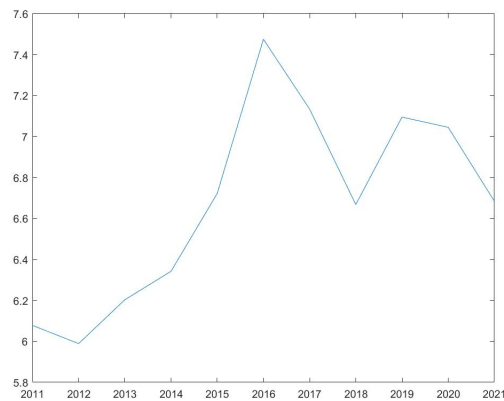


Figure 2 picture 2

The total increasing rate is 1.100%, and increase 0.606°C in value. Increasing rate respectively is given by

$[-0.0941, -0.0146, 0.0357, 0.0224, 0.0599, 0.1120, -0.0456, -0.0653, 0.0640, -0.0070, -0.0512]$.

Past 10 years average temperature is 6.6759°C .

And the monthly average temperature between January 2022 to October 2022 except March we can calculate is

$[3.989, 4.358, 6.812, 9.665, 12.313, 14.742, 15.588, 15.07, 13.151, 10.811]$.

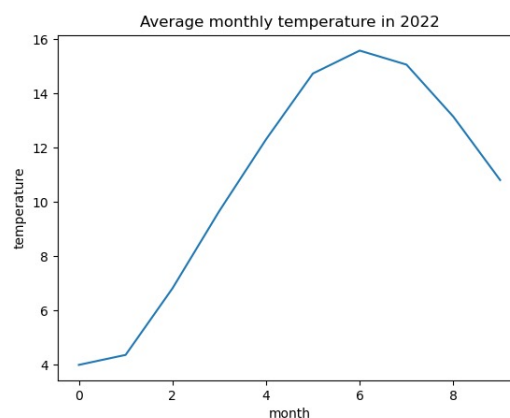


Figure 3 picture 3

Especially, the increasing rate between 2022 to 2021 is 0.0192. This is much smaller than previous 10 years.

5.2 (b)

5.2.1 Analysis of problem 1(b)

Question (b) require us to build 2 or more mathematical models to describe the past and predict the future global temperature level. We can pick a observation as example to analyze the data annually. That is we build models using yearly average temperature. We decide to use Autoregressive integrated moving average (ARIMA) model and long short term memory algorithm(LSTM) to describe the global temperature.

5.2.2 Model 1: ARIMA

ARIMA model, differential integrated moving average autoregression model, also known as integrated moving average autoregression model, is one of the time series prediction and analysis methods. In ARIMA (p, d, q), AR is 'autoregressive', and p is the number of autoregressive terms; MA is the 'moving average', q is the number of moving average terms, and d is the number of differences (orders) made to make it a stationary sequence. This model is built to describe the relationship between current value and historical value using the historical time data of the variable itself to predict itself. So it fit to the question in 1(b).

Firstly, we need to determine the order p which means the lagged order i.e how many previous variable related to the current value. This leads an autoregressive model(AR):

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t$$

where y_t is current value, μ is constant.

Then we consider more about the error along the time series. We get the moving average model(MA):

$$y_t = \mu + \sum_{i=1}^q \theta_i \epsilon_{t_i} + \epsilon_t$$

This can reduce the random fluctuation when estimating.

Finally, we combine previous 2 models then obtain the model ARMA

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t_i} + \epsilon_t$$

But this model has some limitations. The results that ARMA obtains might not be stationary, then some test will not be credible. So we can add a difference order d to let them be stationary.

We load the dataset 'land year std data'

Before we obtain the expression of ARIMA, we need to determine the value of p, q and d. In general, we choose low value of d i.e low order for the model. So it is important to find a reasonable value of p, q. We can check the Akaike information criterion(AIC) and Schwarz criterion(SC) given

by respectively:

$$AIC = \ln\left(\frac{SSE}{T}\right) + \frac{2K}{T}, \quad K = p + q + 2.$$

$$SC = \ln\left(\frac{SSE}{T}\right) + \frac{K \ln(T)}{T}.$$

If AIC and SC are small enough, then the choices of p and q are preferred. So we can use the force algorithm to find the best combination of p, q.

Moreover we need to check which model we should use to estimate the data. We can check the autocorrelation function(ACF):

$$ACF(k) = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}.$$

k is the lagged index.

We calculate the statistic ACF and plot a scatter so that we can really determine which model is the better one.

After we get the annual average temperature, we first perform a stationarity check on the data. The most intuitive way is to use the data to make a sequence diagram and then observe the trend of the image. At the same time, we calculate the first-order difference and the second-order difference of the data and make a graph to observe the trend.

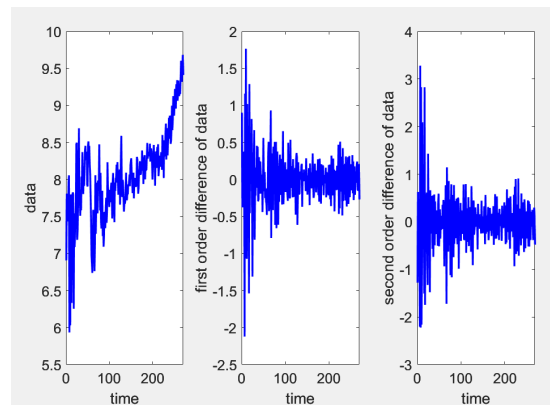


Figure 4 picture 4

It is not obvious from the image to confirm the stationary relationship of the data, so we use the mathematical test method——ADF test and KPSS test.

In the ADF test, if the return value is 0, the data is not stationary, and if the return value is 1, the data is stationary. Then in KPSS test, if the return value is 0, the data is stationary, and if the return value is 1, the data is not stationary.

From the test result we can know the raw data is not stationary, so we need to calculate the first order difference of data, then we do the same test to the first order difference of data. We can get

Therefore, the data after the first order difference can be used for time series modeling analysis. And then what we need to do is to confirm the parameters (p, q, d). The first step is to examine the autocorrelation and partial autocorrelation plots of a stationary time series.

We can't get the values of p and q from the figure, so we adopt the AIC and BIC criteria to find the most suitable p and q values within a certain range. When we have confirmed the appropriate

```
d_adf =
logical
0

d_kpss =
logical
1
```

Figure 5 picture 5

```
d1_adf =
logical
1

d1_kpss =
logical
0
```

Figure 6 picture 6

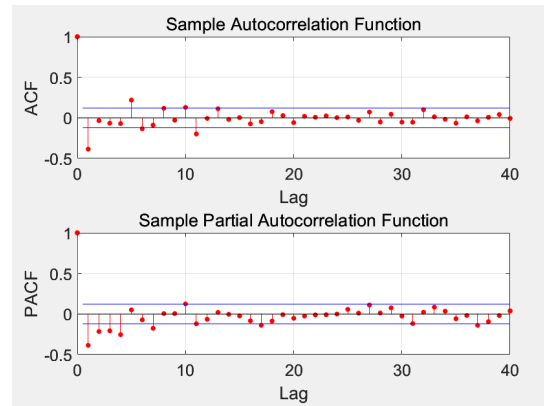


Figure 7 picture 7

(p, q, d) value, we can build a model to predict the subsequent data. In the process of forecasting, we also set the upper and lower forecast limits under the 95% confidence interval.

Finally we get that the best combination of p, q is:

$$p = 3, \quad q = 2$$

5.2.3 Model 2: LSTM

The recurrent neural network has memory, parameter sharing and Turing completeness, so it has certain accuracy and advantages to select the recurrent neural network for learning and prediction.

LSTM, a short - and long-term memory network, is used in this paper. LSTM is formed by a chain of repeated modules, each of which is called a cell (cell).

Each cell has a specific gate structure to achieve selective information transmission. Through the transmission of LSTM gate structure (forgetting gate, input gate, cell state update, output gate) information, each cell state is updated.

The function of Amnesia Gate is to selectively update the cell state. The first step of LSTM is to determine the cell status update.

It uses the Sigmoid function to h_{t-1} and x_t processing, outputting numbers between 0-1. 1 means completely reserved, while 0 means completely forgotten. Where h_{t-1} is the weight matrix, b_f is the offset term.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

The function of the input gate is to input data and determine the information stored in the cell state. This part is divided into two steps. First, the Sigmoid layer determines the weight of the input information in the update, that is, which information needs to be updated. Next, build a candidate vector \tilde{C}_t through tanh calculation at this time

$$i_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

Next update the previous cell status C_{t-1} , add this matrix and the last cell state after the forgetting gate processing to form C_t . Multiply the previous state matrix by f_t , representing selective memory of the last cell state. Then we add $i_t \cdot \tilde{C}_t$ to the value we get, the new cell state C_t is obtained.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

Finally, the output gate will output based on the current cell state, passing the current input data through the Sigmoid layer. Then pass the current cell state through the tanh layer, normalize the matrix between -1 and 1, and perform the basic product operation with the output of the sigmoid layer, So far, the data of the output part has been determined.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

Through the module chain network composed of LSTM cell units, Be able to predict the future through continuous learning of input data.

5.3 (c)

According to the model in question 1(b), we can easily predict global temperatures in 2050 and 2100 respectively using different models. Moreover we need to find when the average temperature of observation points in our prediction models will reach 20.00°C .

5.3.1 Result from model ARIMA

First, let's predict the annual average temperature for the next 100 years.

From the predicted results, when it comes to 2050, the annual average temperature will reach 9.6147 °C . when it comes to 2100, the annual average temperature will reach 9.9985 °C .

Since 20 °C has not been reached, we continue to extend the forecast range. We will predict until 1400 years later. We found that when the time comes to 3404 years , the annual average temperature in the world will reach 20.0075 °C and exceed 20 °C for the first time.

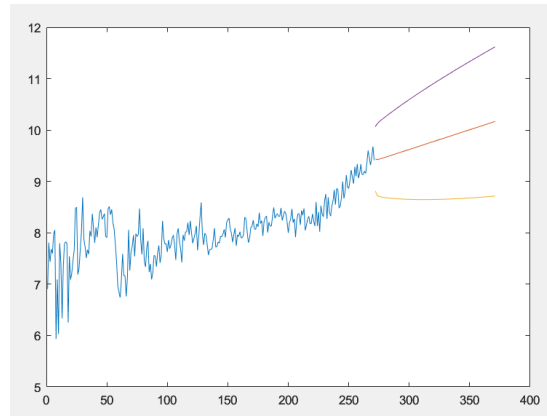


Figure 8 picture 8

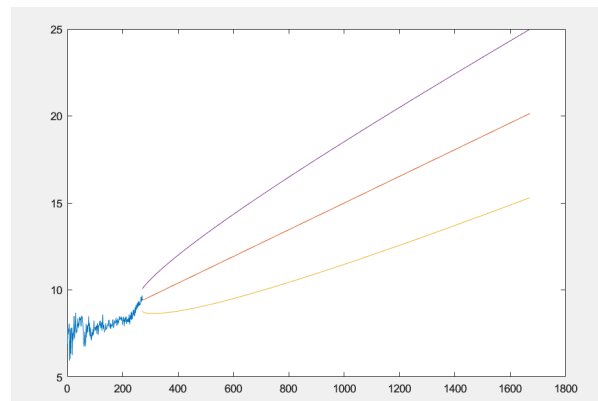


Figure 9 picture 9

5.3.2 Result from model LSTM

5.4 (d)

We need to calculate important statistics to determine which model is better. We compare SE, MSE and RMSE and draw a figure as the following.

5.4.1 Analysis of model ARIMA

Using method of statistics, we can get the following analysis of ARIMA(3,2,1).

	Value	StandardError	TStatistic	PValue
Constant	0.0083569	0.0052983	1.5773	0.11473
AR1	-0.45471	0.15899	-2.86	0.0042368
AR2	0.27647	0.079342	3.4845	0.00049308
AR3	0.089483	0.0525	1.7044	0.0883
MA1	-0.14389	0.15508	-0.92784	0.35349
MA2	-0.63333	0.14147	-4.4768	7.5784e-06
Variance	0.10201	0.0054319	18.78	1.1057e-78

We can find that, most of the p-values are very small. So we can conclude that most of the coefficients are reliable except MA1.

5.4.2 Analysis of model LSTM

VI. Solution for question 2

In question 2, we are required to build a global temperature model not only including the value of temperature but also more other reasons such as position, time, volcanic eruptions, forest fires and the COVID-19 etc.

6.1 (a)

6.1.1 Analysis of problem 2(a)

According to the data given in attachment 1, we analyze the correlation between global temperature, time and location. We can use multiple regression to describe the relationship between global temperature, time and location.

6.1.2 Relationship about location

We first analyze the relationship between latitude and temperature. And we conclude that only latitude effects temperature by geometry. Since we only consider the relationship between latitude and temperature, so we can arbitrary choose data in one year. We use different observation points to denote different latitude. Then we can get average temperature in 2010, for example, while every temperature represents different latitude. And we draw a heat picture to describe the relationship.

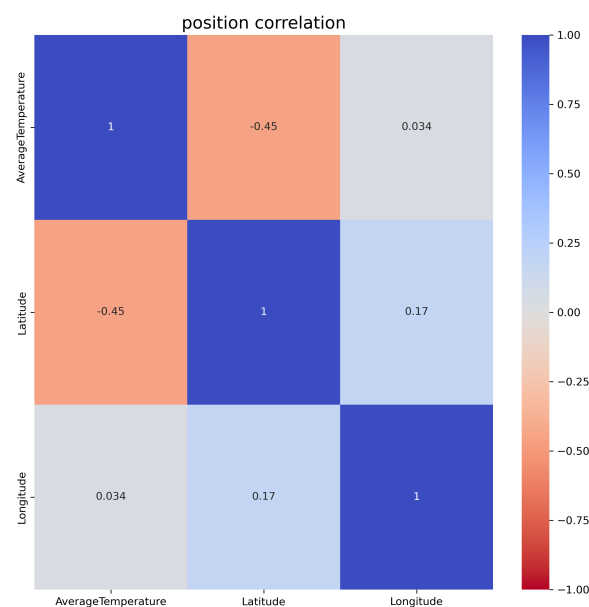


Figure 10 position correlation

Moreover, we make a statistic table for regression model.

OLS Regression Results						
=====						
Dep. Variable:	AverageTemperature	R-squared:	0.218			
Model:	OLS	Adj. R-squared:	0.202			
Method:	Least Squares	F-statistic:	13.49			
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	6.75e-06			
Time:	17:31:22	Log-Likelihood:	-329.87			
No. Observations:	100	AIC:	665.7			
Df Residuals:	97	BIC:	673.6			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	22.4571	0.964	23.297	0.000	20.544	24.370
Latitude	-0.1565	0.030	-5.181	0.000	-0.216	-0.097
Longitude	0.0131	0.010	1.271	0.207	-0.007	0.033
=====						
Omnibus:	4.789	Durbin-Watson:	1.989			
Prob(Omnibus):	0.091	Jarque-Bera (JB):	4.820			
Skew:	-0.508	Prob(JB):	0.0898			
Kurtosis:	2.645	Cond. No.	118.			
=====						

Figure 11 regression table

We can see that p-value is small enough, so we can conclude that this model is good. And we conclude the relationship between temperature and location:

$$Avgtemperature = -0.1565 \cdot latitude + 0.0131 \cdot longitude + 22.4571.$$

6.1.3 Relationship about time

Then, we analyze the relationship between global average temperature and time. We choose one city Shanghai to analyze since we need to reduce the effect of location. So we pick several years monthly average temperature of Shanghai. And we divide every year as a period to use the seasonal statsmodel package.

We can get the time series decomposition.

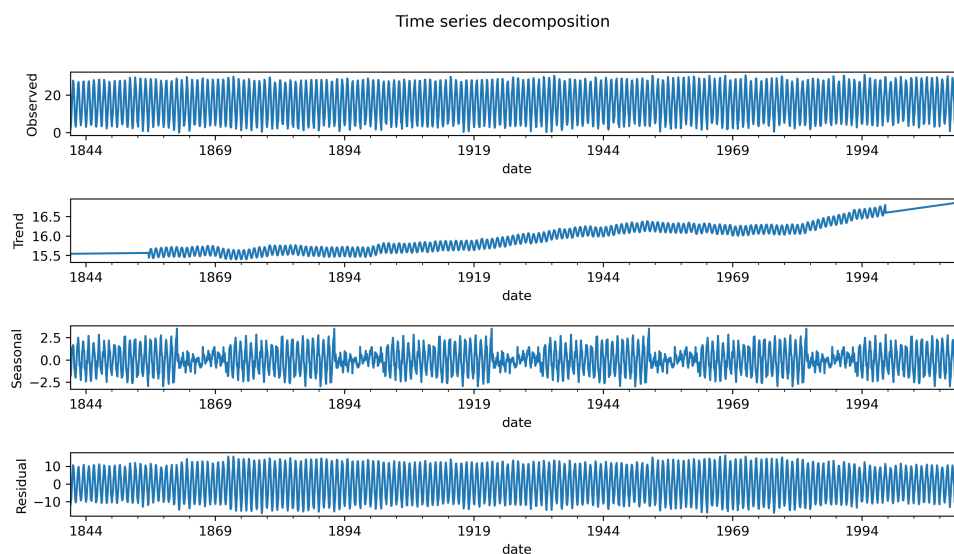


Figure 12 time series decomposition

We find that the estimation and resident are almost stable, and the observed temperature data has

very small increasing trend. We need to confirm that time has no relationship to global temperature.

Then we use augmented Dickey-Fuller test to judge if time factor is a noise for average temperature. We get that the p-value is $3.421929753273345e-10$. So we claim that time has no relationship to global temperature.

6.2 (b)

In this section, we should get data of the natural disasters, we find part of data(such as volcanic eruptions, forest fires) in <https://www.emdat.be/> and part of data(COVID-19) in <https://covid19.who.int/data>.

6.2.1 Relationship analysis

In this section, we choose 2010 volcanic eruptions in Eyafjallajökull, Iceland to analyze the average temperature change of Iceland in 2010. We plot the average temperature of Iceland from January 2009 to December 2011.

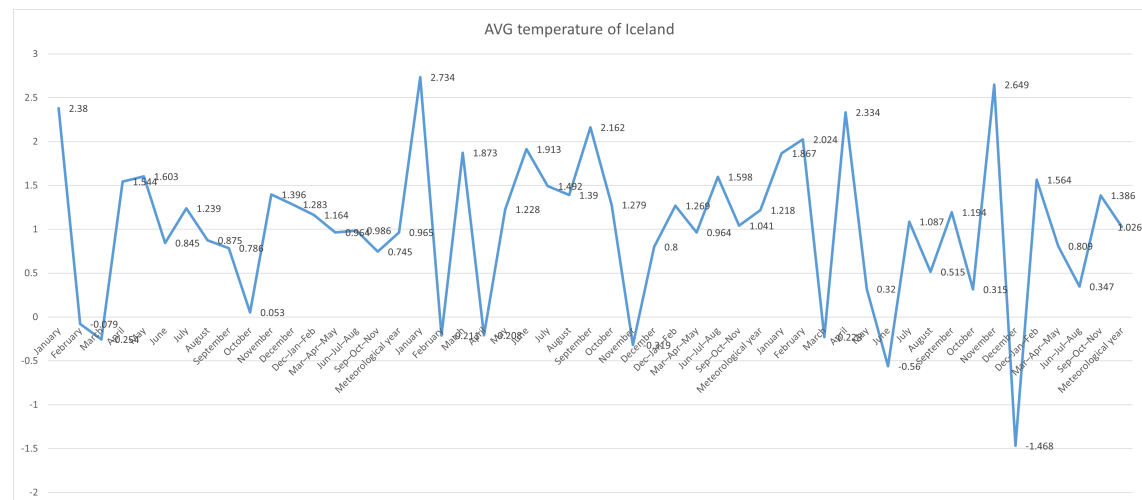


Figure 13 temperature of Iceland

Eyafjallajökull erupted on 20th March, 2010 first. Then it erupted on 14th April again. We want to discover if the eruption affected the local temperature.

Let's focus on the figure. We find that the avg temperature in 2010 is different to 2009 and 2011, which has abnormal value in February and March. The temperature in this two months was on the opposite. In March, 2010, the temperature increase to 1.873°C , while -0.079°C and -0.228 in 2009 and 2011 respectively. Moreover, in April, 2010, the temperature was still a abnormal value which was -0.208°C . After that, there was a high avg temperature during May to October up to 1.577°C . It is 0.677°C greater than which in 2009, and 1.0985°C greater than which in 2011. And the avg temperature 1.577°C is almost 0.5°C greater than the annual temperature of Iceland. It is a very abnormal value for Iceland. And we claim that there is a positive relationship between volcanic eruptions and global temperature.

Next, we analyze the relationship between forest and global temperature. We choose the 2019 Australian forest fires. The forest fires in 2019 is a major natural disaster. And we plot the average temperature of Australia from January 2018 to December 2020.

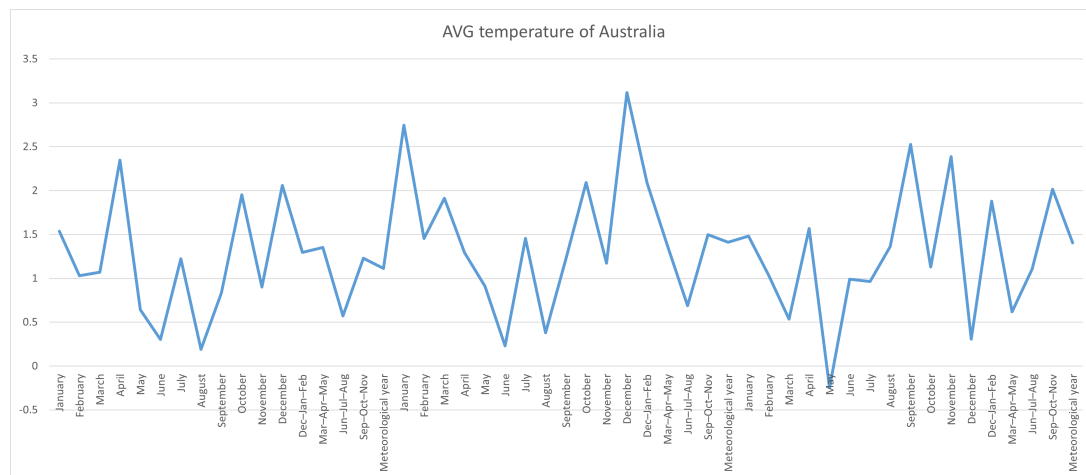


Figure 14 temperature of Australia

The Australian forest fires was last for 1 year. So we focus on July, 2019 to July, 2020.

The avg temperature during the forest fire is 1.268°C . And there is 1.093°C during July, 2018 to July, 2019. It is 0.175°C greater. We claim that there still a positive relationship between forest fires and global temperature.

Last but not least, we consider the relationship between COVID-19 to global temperature.

We can draw the picture about injected number and temperature from 2020 Jan.

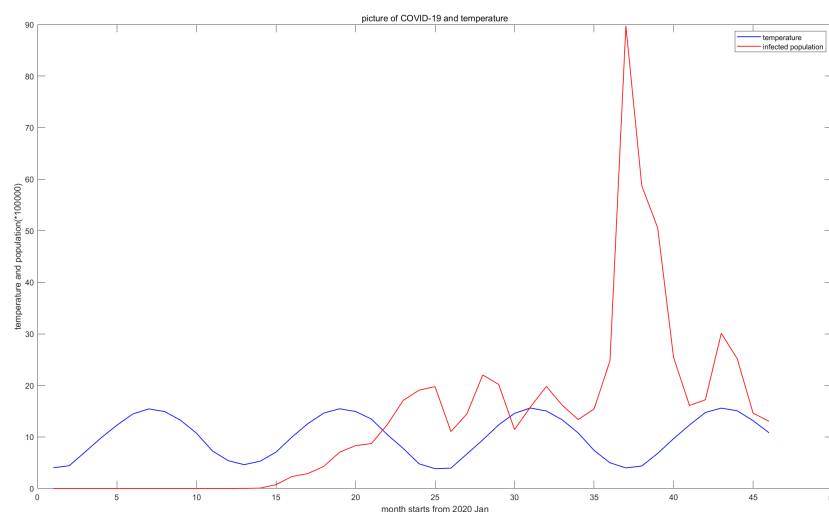


Figure 15 COVID-19 and temperature

We calculate the average temperature of first 15 months is $avg1 = 9.0693$, and $avg2 = 10.5949$ represents last 31 months when COVID-19 burst. We can see that after COVID-19 burst in 2020 Jan, the global average temperature arised about 1.5256°C . And this rise is reasonable because people need more activities to fight against COVID-19. So people around the world will take action such as large transport, factory produce etc. In other words, this will lead to entropy increase and be characterized by global average temperature rise. So we still claim that there is a positive relationship between global temperature and COVID-19.

6.2.2 Summary

6.3 (c)

6.3.1 Data of more factor

We find datasets of world energy, world population and Global forest resource. <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html> <https://population.un.org/wpp/Download/Standard/MostUsed/> <https://fra-data.fao.org/W0/fra2020/dataDownload/> <https://www.gddat.cn/newGlobalWeb/#/FeatureAnalysis> <https://www.fao.org/faostat/en/#data>

6.3.2 Analysis of relationship

Then we can use Grey Relation Analysis (GRA) to estimate the relationship between other factors

6.4 (d)

VII. Solution for question 3

VIII. Future Work

8.1 Another model

8.1.1 *The limitations of queuing theory*

8.1.2

8.1.3

8.1.4

IX. References

- [1] Author, Title, Place of Publication: Press, Year of publication.
- [2] author, paper name, magazine name, volume number: starting and ending page number, year of publication.
- [3] author, resource title, web site, visit time (year, month, day).
- [4] L^AT_EX 资源和技巧学习 <https://www.latexstudio.net>
- [5] L^AT_EX 问题交流网站 <https://wenda.latexstudio.net>
- [6] 模板库维护 <https://github.com/latexstudio/APMCMThesis>

X. Appendix

Listing 1: The matlab Source code of Algorithm

```

kk=2; [mdd, ndd]=size(dd);
while ~isempty(V)
    [tmpd, j]=min(W(i, V)); tmpj=V(j);
    for k=2:ndd
        [tmp1, jj]=min(dd(1, k)+W(dd(2, k), V));
        tmp2=V(jj); tt(k-1, :)= [tmp1, tmp2, jj];
    end
    tmp=[tmpd, tmpj, j; tt]; [tmp3, tmp4]=min(tmp(:, 1));
    if tmp3==tmpd, ss(1:2, kk)= [i; tmp(tmp4, 2)];
    else, tmp5=find(ss(:, tmp4)~=0); tmp6=length(tmp5);
    if dd(2, tmp4)==ss(tmp6, tmp4)
        ss(1:tmp6+1, kk)= [ss(tmp5, tmp4); tmp(tmp4, 2)];
    else, ss(1:3, kk)= [i; dd(2, tmp4); tmp(tmp4, 2)];
    end; end
    dd=[dd, [tmp3; tmp(tmp4, 2)]]; V(tmp(tmp4, 3))=[];
    [mdd, ndd]=size(dd); kk=kk+1;
end; S=ss; D=dd(1, :);

```

Listing 2: The lingo source code

```

kk=2;
[mdd, ndd]=size(dd);
while ~isempty(V)
    [tmpd, j]=min(W(i, V)); tmpj=V(j);
    for k=2:ndd
        [tmp1, jj]=min(dd(1, k)+W(dd(2, k), V));
        tmp2=V(jj); tt(k-1, :)= [tmp1, tmp2, jj];
    end
    tmp=[tmpd, tmpj, j; tt]; [tmp3, tmp4]=min(tmp(:, 1));
    if tmp3==tmpd, ss(1:2, kk)= [i; tmp(tmp4, 2)];
    else, tmp5=find(ss(:, tmp4)~=0); tmp6=length(tmp5);
    if dd(2, tmp4)==ss(tmp6, tmp4)
        ss(1:tmp6+1, kk)= [ss(tmp5, tmp4); tmp(tmp4, 2)];
    else, ss(1:3, kk)= [i; dd(2, tmp4); tmp(tmp4, 2)];
    end;
end
    dd=[dd, [tmp3; tmp(tmp4, 2)]]; V(tmp(tmp4, 3))=[];
    [mdd, ndd]=size(dd);
    kk=kk+1;
end;
S=ss;
D=dd(1, :);

```