**Environmental prediction analysis based on GM (1, 1) model and ARIAM model**

**1 Introduction**

In China, environmental pollution is an issue of widespread concern. Environmental pollution not only affects economic development, but also affects people's health.

The Chinese government has issued some policies to deal with environmental pollution. In order to better help the government to make policies, the study will use the GM model and the ARIMA model to predict and analyse the three aspects of China's future environmental pollution: exhaust gas, wastewater, and solid waste. The study used data from 2004 to 2019 released by the National Bureau of Statistics of China. Due to the limited data, our group first used the GM model for predictive analysis. In order to improve the accuracy of forecasting, based on the principles and ideas of time series modelling, our group implemented the ARIMA model for further forecasting analysis. The reform study will use Matlab to get the predictions, so as to obtain some analysing information for the government to refer to.

**1.1 Prediction Model and Environment Pollution**

**1.1.1 Environment Pollution**

Environmental pollution is a far-reaching issue, which will not only affect economic growth but also bring about a series of human health problems (Khan & Ghouri, 2011). Environmental pollution is also a much-developed problem in China. The Chinese government is also paying close attention to the current environmental pollution situation and has put forward some targeted policies. Environmental pollution includes three categories in this study: water pollution, air pollution, and solid waste pollution.

Water is essential to human life and industrial development. However, according to the European Public Union (2009), water pollution occurs worldwide. Water pollution can affect ecology, human health, and economic production activities (Carter, 1985). Water pollution refers to the pollution caused by harmful substances entering the water body. Pollutants that pollute water include radioactive materials, oil pollution, wastewater, sewage, and agricultural water pollution (Denchak, 2018). In this study, water pollution includes wastewater, ammonia nitrogen emissions, and petroleum emissions. Volatile phenols, lead substances, cadmium, chromium, hexavalent chromium.

The air we breathe is vital to human health. According to Holland et al, air pollution is indeed harmful to human health. But air pollution is indeed widespread (EPHA, 2009), and China is also facing the problem of air pollution. Mishra pointed out that air pollution is caused by urbanization, industrialization, and bad environmental regulations. Air pollution refers to the air containing one or more other harmful substances that are harmful to health (Health and Energy, 2007). According to the European Public Health Union (2009), the main pollutants found in the air include particulate matter, lead, ground-level ozone, heavy metals, sulfur dioxide, polycyclic aromatic hydrocarbons, benzene, carbon monoxide and nitrogen carbon dioxide. In this study, air pollution contains three kinds of pollutants: sulfur dioxide, nitrogen oxides, and smoke and dust. In the statistical process of the National Bureau of Statistics of China, since the ships operated by mobile sources across the country cannot be surveyed at the provincial administrative unit, the air pollution generated by them is not included in the total emissions.

Solid waste pollution is an important cause of environmental pollution. Solid waste refers to any material that is discarded through waste (Ejaz et al., 2010). Solid waste can cause the spread of infectious diseases, water pollution and loss of biodiversity. Therefore, proper disposal of solid waste is particularly important. The solid waste in this article refers to domestic waste.

**1.1.2 Prediction Model**

Predictive modeling, also known as predictive analysis, is a statistical method intended to forecast future events or results by evaluating trends that predict future outcomes. Based on historical data, the main purpose of predictive modeling is to predict what will happen in the future. This study uses GM model , and ARIMA model to do the prediction.

The GM model is also called the gray system. It was established in the 1980s by Professor Deng Julong of Huazhong University of Science and Technology. As a new cross-cutting discipline, the system has been developed by leaps and bounds in just two decades. It has become one of the important methods for forecasting, decision-making, evaluation, planning, control, system analysis and modeling in many fields such as society, economy, science and education, and technology. Grey forecasting is a method for forecasting systems with uncertain factors. Grey forecasting finds the law of system changes by identifying the degree of difference in the development trend between system factors, that is, performing correlation analysis, and processing the original data to find the law of system changes, generating a data sequence with strong regularity, and then establishing the corresponding differential equation Model to predict the future development trend of things. In particular, it has a very significant effect on the modeling and analysis of systems with short time series, few statistical data, and incomplete information. For example, one can use the GM model to predict SARS well.

The ARIMA model was proposed by Box and Jenkins in the 1970s and is a well-known time series forecasting method. The basic idea of the ARIMA model is to treat the data sequence formed by the prediction object over time as a random sequence. Use a certain mathematical model to predict future values. One application of ARIMA is to predict the future development of a country’s GDP.

**Key Words:** GM Model, ARIMA Model, Environmental Prediction, Pollution

**2 Methodology**

The purpose of this paper is to predict the quantities of some pollutants in wasted water, wasted gas and wasted solids. Our group get the data from the National Bureau of Statistics because they are quite accurate and reliable. Our group used GM model and ARIMA model to predict the value of some pollutants in wasted water, wasted gas and wasted solids in the following years. The process of these two models are as follows:

**GM model**

1. Given original data series x (0)

x (0) = (x (0)(1), x (0)(2) …, x(0)(n)) (1)

1. Then using AGO (1-AGO) to generates the sequence in order to weaken the randomness and volatility of the original sequence

x (1) = (x (1)(1), x (1)(2) …, x(1)(n)) (2)

x(1)(k) = (3)

1. The grey differential model

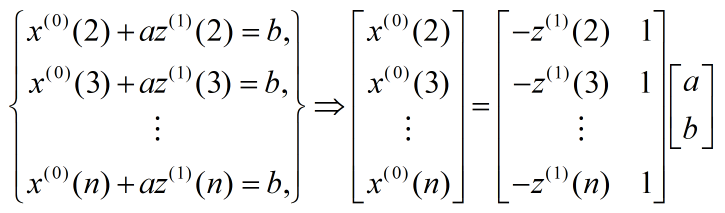
+ az(1)(k) = b, k = 2, 3, … n (4)

= 0.5x(1)(k) + 0.5x(1)(k-1), k = 2, 3, …, n (5)

a is the development coefficient

b is the grey input

1. Computing a and b



According to the least square method,

= (6)

1. The whitened differential equations:

+ ax(1)(t) = b (7)

The complete solution of the equation is

(t) = (8)

Suppose t = k + 1

(k+1) = (9)

1. The predicting value can be calculated as follows

(t) = (k+1) - (k) (10)

1. Class ratio test

k = 2, 3, …, n; r = 0, 1, 2, … (11)

If the series can be used to build a GM(1, 1) model, then the class ratio must satisfy that

(12)

Else we should add a value to the data.

k = 2, 3, …, n (13)

Until the data satisfies the class ratio test.

When getting in the traditional grey model, researchers always using the default weight 0.5 and 0.5. However, using this default weight does not mean the model is the most accurate. So we try to find a better model through changing the weight until we get the least error.

So the formula will be changed to:

= wx(1)(k) + (1-w)x(1)(k-1), k = 2, 3, …, n (14)

And we define the error as follows:

= , k = 2, 3, …, n (15)

When we get , we find the final weight.

**ARIMA model**

ARIMA (p, d, q) model is an extension of ARMA (p, q) model, which consists of three parts: autoregressive model AR (p), order of difference (d) and moving average model MA (q).

1. Plot the data and observe whether it is a stable time series. If the data is a non-stationary time series, the d-order difference operation is performed to transform the data into a stationary time series.

1.1 judge whether the time series are stable or not

Method 1: observe the time series directly. If the time series fluctuates at a certain point, it means that the time series is stable. Then d = 0.

Method 2: unit root test

There are six test methods (ADF, SFGLS, PP, KPSS, ERS, NP) and the judgment method of the test is as follows:

P value:

if it is less than the critical value, the original hypothesis will be rejected

Above the critical value, the original hypothesis is accepted

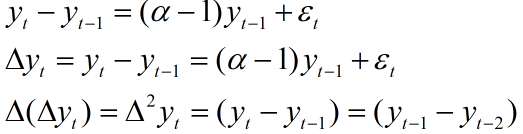
Critical value:

ADF, DFGLS, PP, NP left side: if the statistical value is greater than the critical value, accept the original hypothesis

KPSS, ERS: if the statistical value is less than the critical value, accept the original hypothesis

1.2 Time series stabilization. Time series need differential processing until the time series is stable. Then d = order of differentiation.

The common formula of difference operation is as follows:

 (16)

(17)

(18)

2. Calculate autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) to obtain p and q.

2.1 ACF

(19)

2.2 PACF (20)

(21)

(22)

(23)

Here is the previous autocorrelation coefficient ACF. Then the partial correlation coefficient of k is solved according to Yule Walker equation.

|  |  |  |
| --- | --- | --- |
| Model | ACF | PACF |
| ARMA(p，q） | After q-order, the attenuation tends to 0 | The attenuation tends to 0 after p-order |

(24)

We need to analyze ACF graph and PACF graph to get the best level p and order q.

3. Model test

3.1 test the randomness of residual sequence

4. Data prediction

The forecast function in MATLAB is used to estimate the value of the next period at the end of the data

**3 Analysis and interpretation of results**

1. Wasted solids: predicting results.
2. Domestic waste clearing and transportation volume (ton)

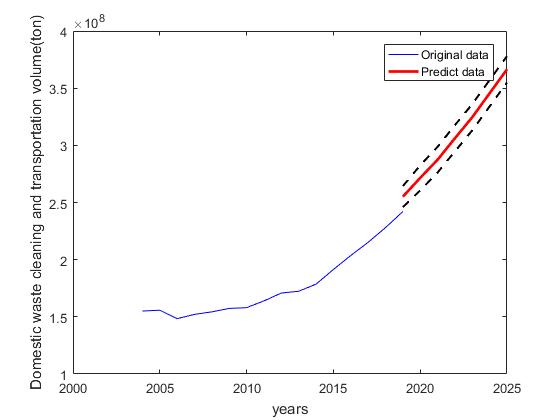


Figure 1: prediction of domestic waste clearing and transportation volume (ton)

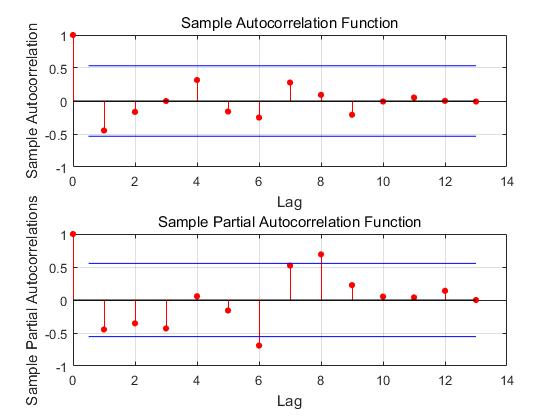


Figure 2: sample autocorrelation function and sample partial autocorrelation function for the second difference

When analyzing the original data, the time sequence was not stable until we did two differences to the original data. So we set d to 2. According to the sample autocorrelation function and sample partial autocorrelation function, the p and q firstly get close to 0 at point 2 and 4. So our group set p as 2 and q as 4.

Through ARIMA model, the red fitted curve showed that the domestic waste clearing and transportation volume will keep increasing steadily because the slope of the curve is relatively stable. We are 95% confident that the number may fluctuate among the dotted lines. According to Fig. 19 in appendix, there is no regularity in standardized residuals graph and data are around the line in QQ plot, so the model is reasonable and data is normally distributed. Based on our model, the number may reach to 30000 tons in 2021.

1. Harmless treatment capacity of domestic waste (ton/day)

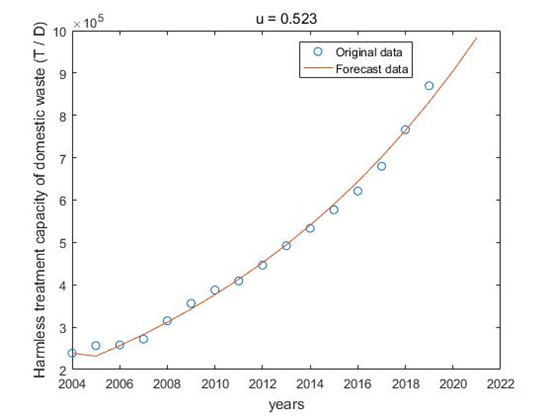


Figure 3: prediction of harmless treatment capacity of domestic waste (tons/day)

When calculating , we used a different computing process. After the test, when weight is equal to 0.523, the model can get the minimum predicting error E. For more details, please check appendix Fig.22.

Through the model, the red fitting curve showed an upward tendency of harmless treatment capacity of domestic waste, and it will increase year by year because the slope is gradually increasing. In 2021, the processing capacity may reach to 9.9\*105 tons per day.

1. Harmless treatment capacity of domestic waste sanitary landfill (tons/day)

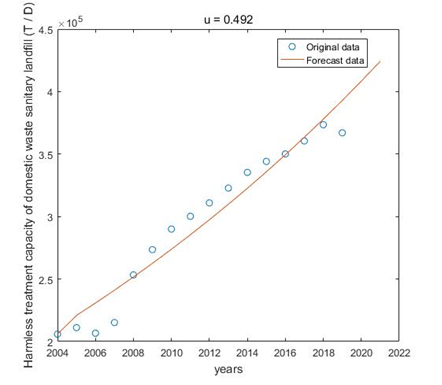


Figure 4: prediction of harmless treatment capacity of domestic waste sanitary landfill (ton/day)

In Fig.4, when the best weight is 0.492, the model can get the minimum predicting error E. For more details, please check appendix Fig.23.

Through GM(1, 1) model, the red fitting curve showed that in 2021, harmless treatment capacity of domestic waste sanitary landfill will reach to 4.25\*105 tons per day.

1. Harmless treatment capacity of domestic waste incineration

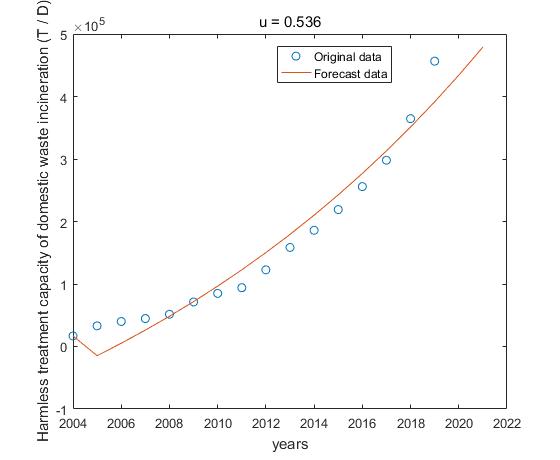


Figure 5: prediction of harmless treatment capacity of domestic waste incineration (tons/day)

In Fig.5, when the best weight is 0.536, the model can get the minimum predicting error E. For more details, please check appendix Fig.24.

Through GM(1, 1) model, the red fitting curve showed that in 2021, harmless treatment capacity of domestic waste incineration will reach to 4.9\*105 tons per day.

1. Harmless treatment capacity of domestic waste (tons)

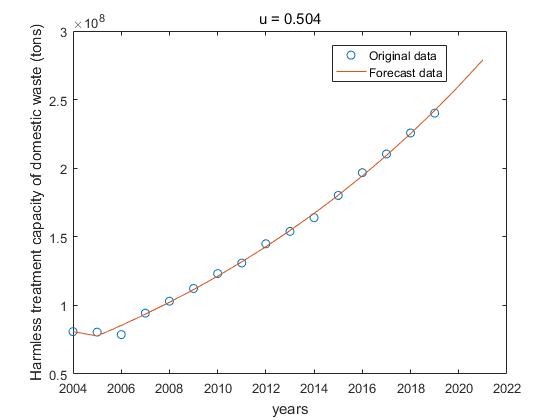


Figure 6: prediction of harmless treatment capacity of domestic waste (tons)

In Fig.6, when the best weight is 0.504, the model can get the minimum predicting error E. For more details, please check appendix Fig.25.

Through GM(1, 1) model, the red fitting curve showed that in 2021, harmless treatment capacity of domestic waste will reach to 2.75\*108 tons.

1. Harmless treatment capacity of domestic waste sanitary landfill (ton)

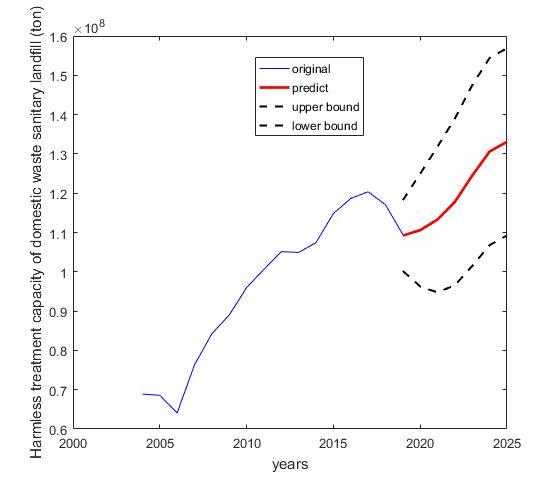


Figure 7: prediction of harmless treatment capacity of domestic waste sanitary landfill (ton)

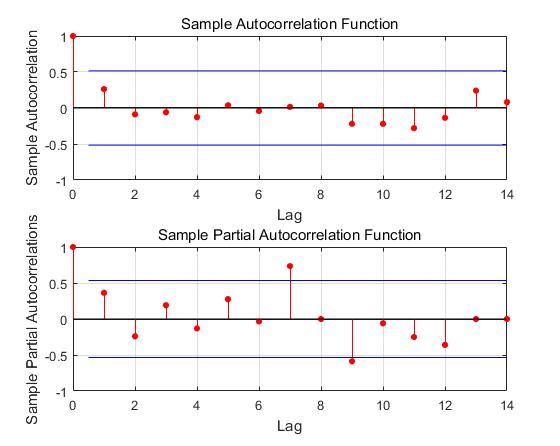


Figure 8: sample autocorrelation function and sample partial autocorrelation function for the first difference

When analyzing the original data, the time sequence was not stable until we did one differences to the original data. So we set d to 1. According to the sample autocorrelation function and sample partial autocorrelation function, the p and q firstly get close to 0 at point 2 and 6. So our group set p as 2 and q as 6.

Through ARIMA model, the red fitted curve showed that the harmless treatment capacity of domestic waste sanitary landfill will keep increasing but it may fluctuate. We are 95% confident that the number may fluctuate among the dotted lines. According to Fig. 20 in appendix, there is no regularity in standardized residuals graph and data are around the line in QQ plot, so the model is reasonable and data is normally distributed. Based on our model, the number may reach to 1.1\*108 tons in 2021.

1. Harmless treatment capacity of domestic waste incineration (tons)

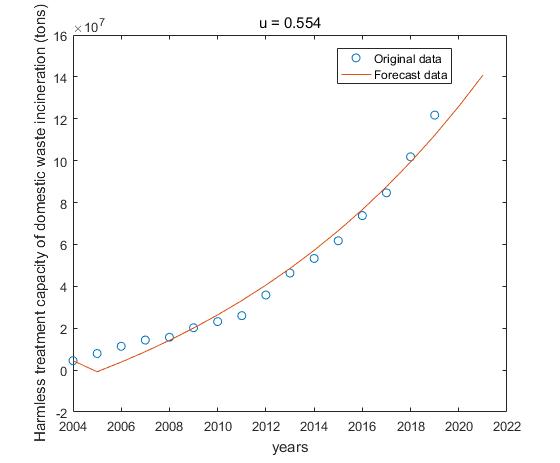


Figure 9: prediction of harmless treatment capacity of domestic waste incineration (tons)

In Fig.9, when the best weight is 0.554, the model can get the minimum predicting error E. For more details, please check appendix Fig.26.

Through GM(1, 1), the red fitting curve showed that in 2021, harmless treatment capacity of domestic waste will reach to 1.4\*108 tons.

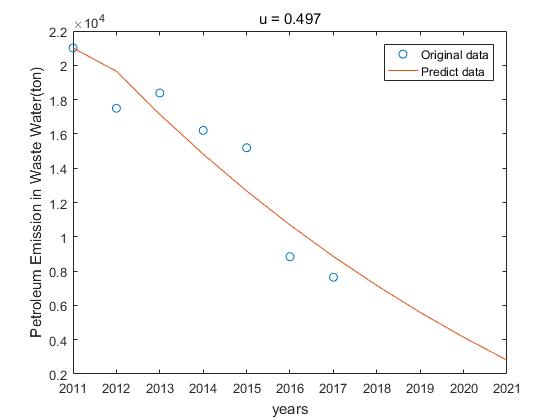
1. Wasted water: predicting results.
2. prediction of petroleum emission in wasted water (tons)

Figure 10: prediction of petroleum emission in wasted water (tons)

In Fig.10, when the best weight is 0.497, the model can get the minimum predicting error E. For more details, please check appendix Fig.27.

Through GM(1, 1) model, the red fitting curve showed that in 2021, petroleum emission in wasted water will drop to 3000 tons.

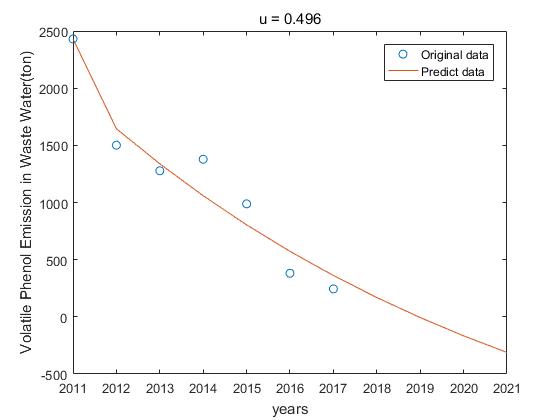
1. Volatile phenol emission in wasted water (ton)

Figure 11: prediction of volatile phenol emission in wasted water (tons)

In Fig.11, when the best weight is 0.496, the model can get the minimum predicting error E. For more details, please check appendix Fig.28.

Through GM(1, 1) model, the red fitting curve showed that in 2021, the volatile phenol emission in wasted water will drop to a negative number. So it implies that the volatile phenol emission will going to be very low or disappear.

1. Hexavalent chromium emission in wasted water (ton)

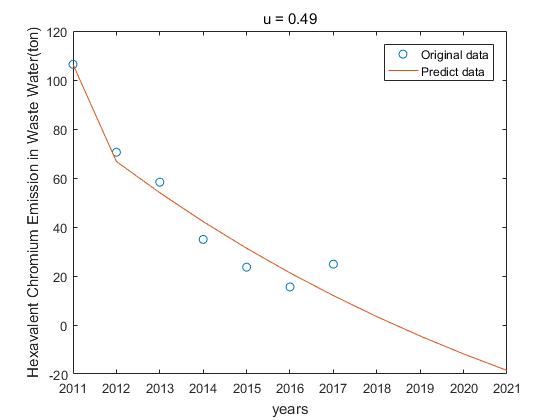


Figure 12: prediction of hexavalent chromium emission in wasted water (tons)

In Fig.12, when the best weight is 0.490, the model can get the minimum predicting error E. For more details, please check appendix Fig.29.

Through GM(1, 1) model, the red fitting curve showed that in 2021, the hexavalent chromium emission in wasted water will drop to a negative number. So it implies that the hexavalent chromium emission will going to be very low or disappear.

1. Total chromium emission in wasted water (ton)

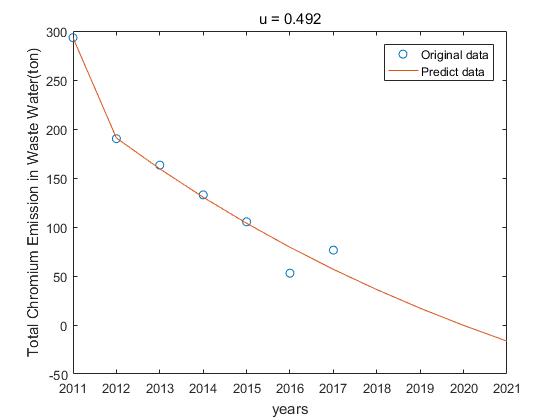


Figure 13: prediction of total chromium emission in wasted water (tons)

In Fig.13, when the best weight is 0.492, the model can get the minimum predicting error E. For more details, please check appendix Fig30.

Through GM(1, 1) model, the red fitting curve showed that in 2021, the total chromium emission in wasted water will drop to a negative number. So it implies that the total chromium emission will going to be very low or disappear.

1. Plumbum emission in wasted water (ton)

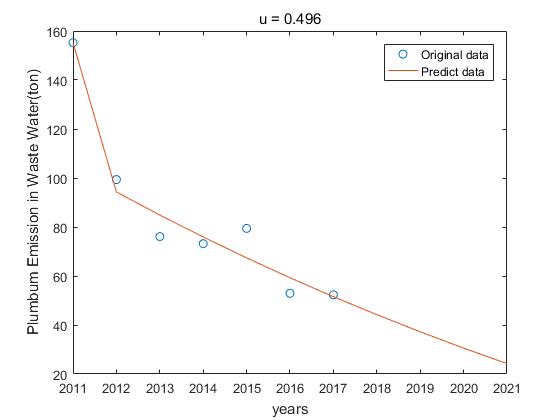


Figure 14: prediction of plumbum emission in wasted water (ton)

In Fig.14, when the best weight is 0.496, the model can get the minimum predicting error E. For more details, please check appendix Fig.31.

Through GM(1, 1) model, the red fitting curve showed that in 2021, plumbum emission in wasted water will drop to 25 tons.

1. Cadmium emission in wasted water (ton)

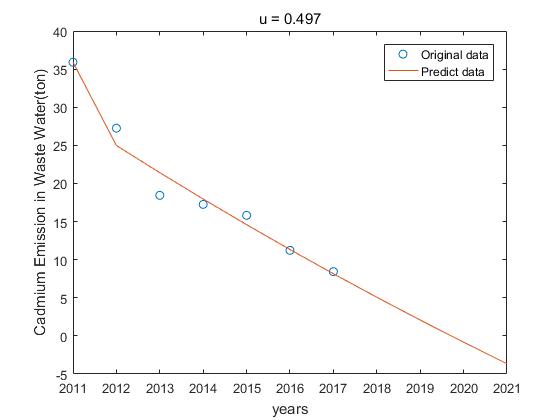


Figure 15: prediction of cadmium emission in wasted water (ton)

In Fig.15, when the best weight is 0.497, the model can get the minimum predicting error E. For more details, please check appendix Fig.32.

Through GM(1, 1) model, the red fitting curve showed that in 2021, the cadmium emission in wasted water will drop to a negative number. So it implies that the cadmium emission will going to be very low or disappear.

1. Wasted gas: predicting results.
2. Sulfur dioxide (tons)

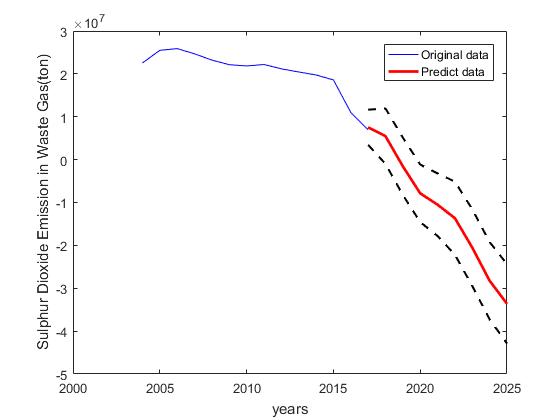


Figure 16: prediction of sulfur dioxide (tons)

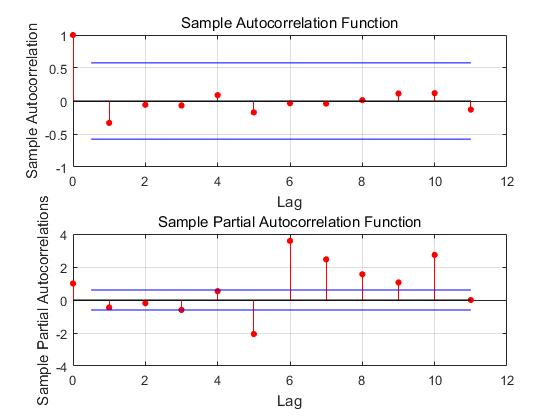


Figure 17: sample autocorrelation function and sample partial autocorrelation function for the first difference

When analyzing the original data, the time sequence was not stable until we did one differences to the original data. So we set d to 2. According to the sample autocorrelation function and sample partial autocorrelation function, the p and q firstly get close to 0 at point 2 and 1. So we set p as 2 and q as 1.

Through ARIMA model, the red fitted curve showed that the sulfur dioxide will keep decreasing but it may fluctuate. We are 95% confident that the number may fluctuate among the dotted lines. According to Fig. 21 in appendix, there is no regularity in standardized residuals graph and data are around the line in QQ plot, so the model is reasonable and data is normally distributed. Based on our model, the number will drop to a negative number, in 2021. So it implies that the sulfur dioxide will going to be very low or disappear.

1. NOx emission (ton)

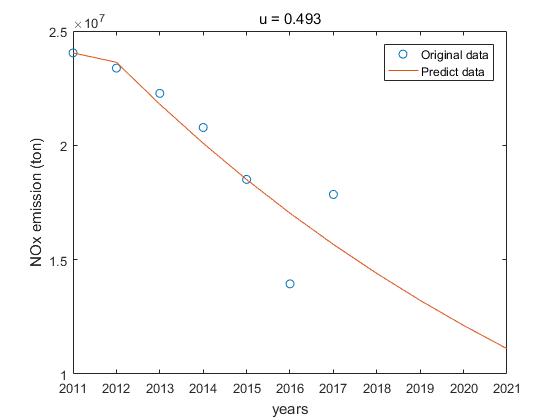


Figure 18: prediction of NOx emission (ton)

In Fig.18, when the best weight is 0.493, the model can get the minimum predicting error E. For more details, please check appendix Fig.33.

Through GM(1, 1) model, the red fitting curve showed that in 2021, NOx emission will drop to 1.1\*107 tons.

**4 Conclusion**

In this paper, we used GM (1, 1) model and ARIMA model to predict the pollution in wasted solids, wasted water and wasted air. The results and error analysis showed that the models are proper and the prediction are reliable. It is believed that these models can also be used to predict industrial data or some other area.

**5 Further Work**

In the paper, we used grey model and ARIMA model to predict the pollution. However, our data was insufficient since we can only get most of the data from 2004 to 2019, which may lead to some errors. If we can get more data, for example, the data before 2020 or 2004, our prediction can be more accurate. Also we can build models for each type of data and that can make the prediction closer to the reality.

**6 Recommendations**

We can see that most of the pollution are decreasing since China legislated laws to prevent pollution. So the government can keep rigid supervision on the companies. Also they can focus on the petroleum emission and plumbum emission in wasted water and NOx emission in wasted gas because they cannot reach to a low level according to our prediction.

References

Carter, F. W. (1985). Pollution Problems in Post-War Czechoslovakia, *Transactions of the Institute of British Geographers*, 10(1), pp. 17-44.

Denchak.M (2018) *Water Pollution: Everything You Need to Know*. Retrieved from <https://www.nrdc.org/stories/water-pollution-everything-you-need-know#whatis>

European Public Health Alliance, (2009). *Air, Water Pollution and Health Effects*.

Retrieved from <http://www.epha.org/r/54>

Environment Research Network (PERN) Cyberseminar, December 1-15. Retrieved from http://www.mnforsustain.org/climate\_health\_effects\_of\_air\_pollution\_mishra\_pern.htm

Ejaz, N., Akhtar, N., Hashmi, H., & Naeem, U. A. (2010). Environmental impacts of improper solid waste management in developing countries: A case study of Rawalpindi city. *The sustainable world*, *142*, 379-387.

Health and Energy, (2007). *Air Pollution Health Effects*, Retrieved from

<http://healthandenergy.com/air_pollution_health_effects.htm>

Holland, W. W., Bennett, A. E., Cameron, I. R. , Florey, C. V., Leeder, S. R., Shilling, R. S. F.,Swan, A. V. & Waller, R. E. (1979). Health Effects of Particulate Pollution: Reappraising the Evidence. Am Journal Epidemiol, 110(5), pp. 525- 659.

Khan, M. A., & Ghouri, A. M. (2011). Environmental pollution: its effects on life and its remedies. *Researcher World: Journal of Arts, Science & Commerce*, *2*(2), 276-285.

Mishra, V. (2003). Health Effects of Air Pollution, Background paper for Population-

Vinz,S. (2015) *Developing your theoretical framework*. Available: https://www.scribbr.com/dissertation/ (Accessed: 12/01/2021)

**Appendix:**

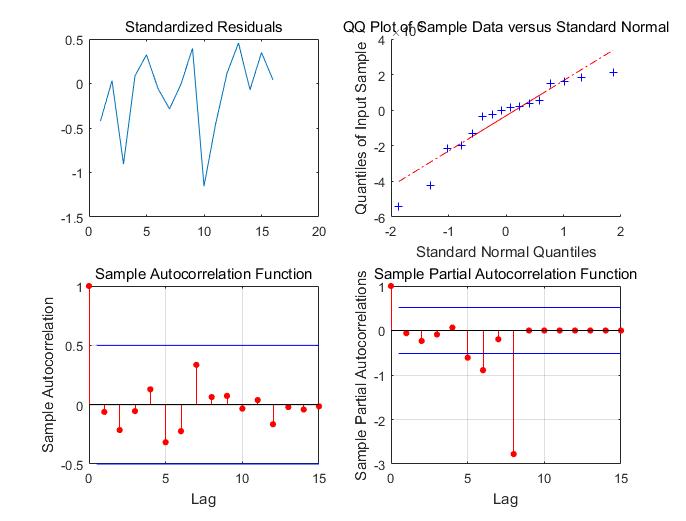


Figure 19: Error analysis of domestic waste cleaning and transportation volume

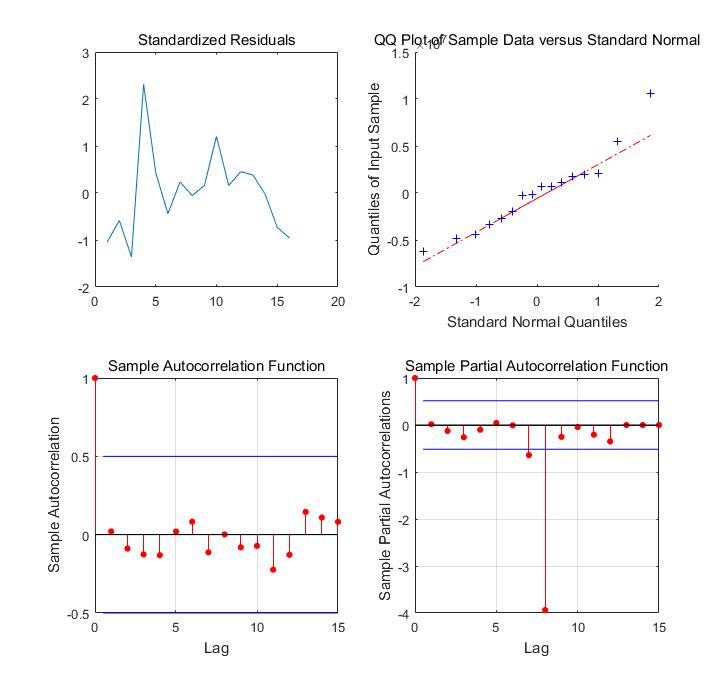


Figure 20: Error analysis of harmless treatment capacity of domestic waste sanitary landfill

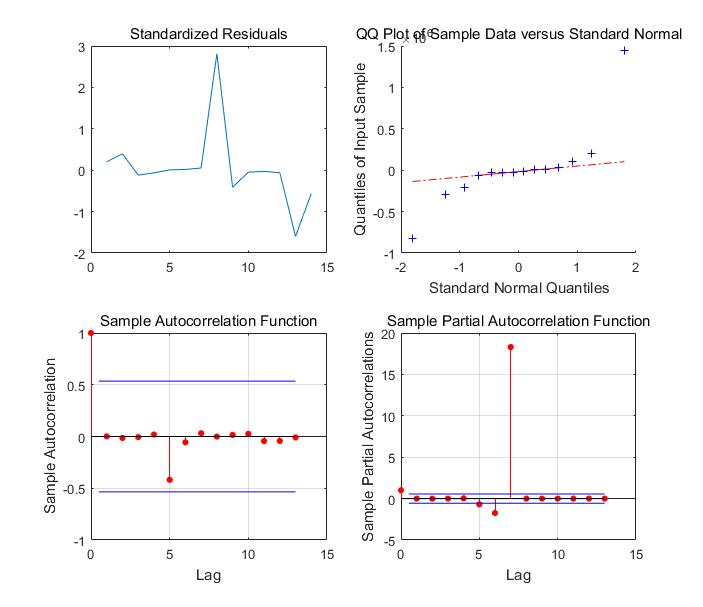


Figure 21: Error analysis of ammonia nitrogen emission

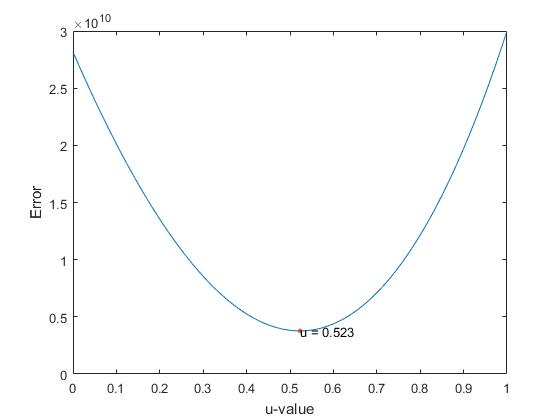


Figure 22: Error analysis of harmless treatment capacity of domestic waste (tons/day)

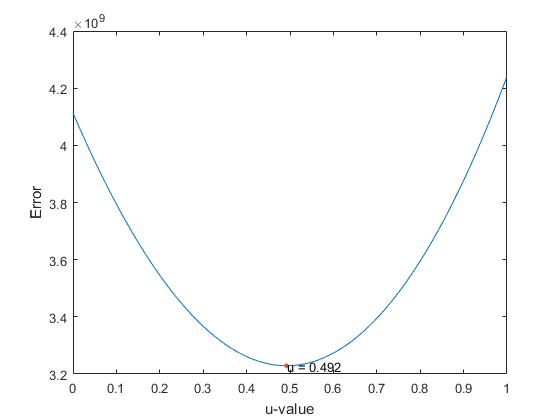


Figure 23: Error analysis of harmless treatment capacity of domestic waste sanitary landfill (ton/day)

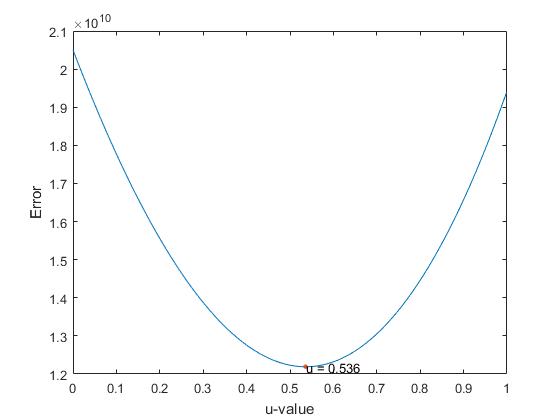


Figure 24: Error analysis of harmless treatment capacity of domestic waste incineration (tons/day)

l (ton/day)

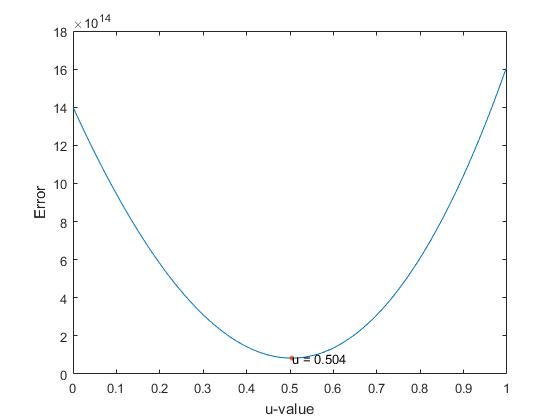


Figure 25: Error analysis of harmless treatment capacity of domestic waste (tons)

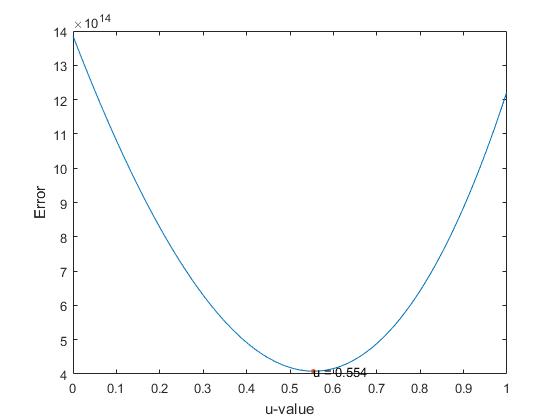


Figure 26: Error analysis of harmless treatment capacity of domestic waste incineration (tons)

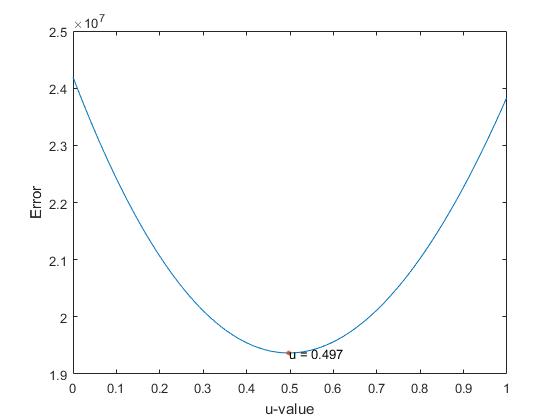


Figure 27: Error analysis of petroleum emission in wasted water (tons)

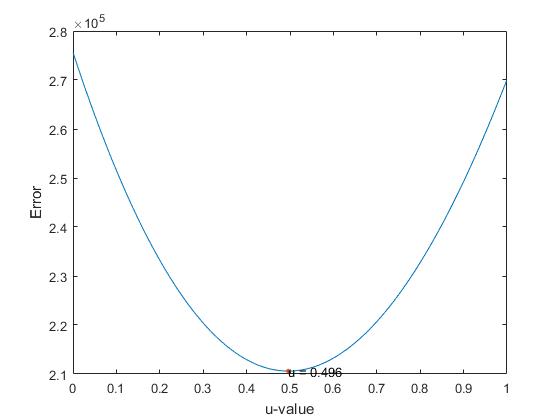


Figure 28: Error analysis of volatile phenol emission in wasted water (tons)

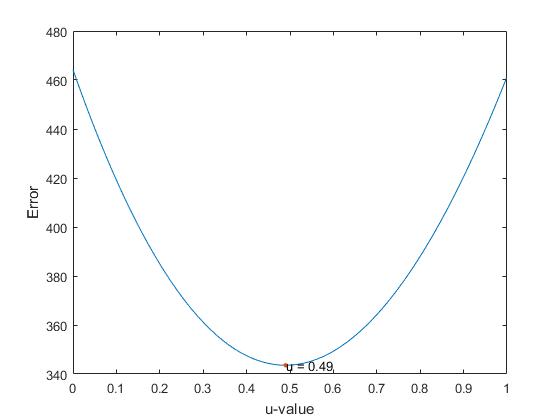


Figure 29: Error analysis of hexavalent chromium emission in wasted water (tons)

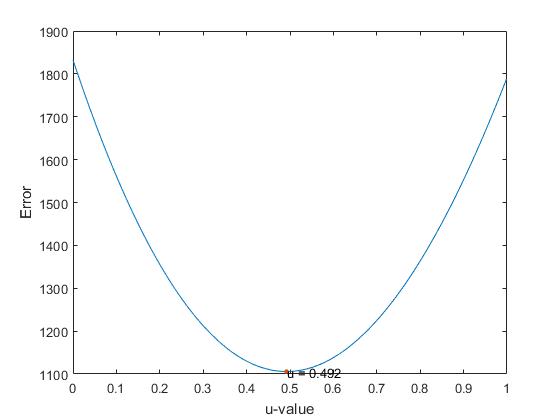


Figure 30: Error analysis of total chromium emission in wasted water (tons)

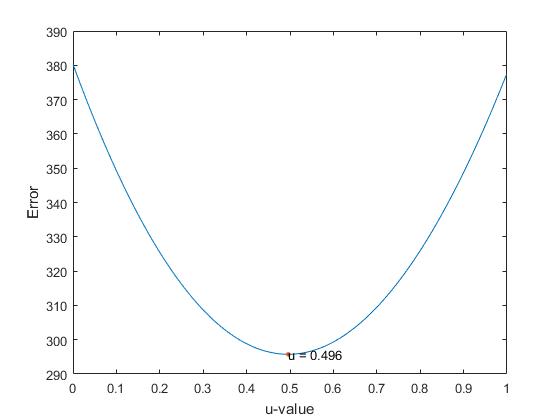


Figure 31: Error analysis of plumbum emission in wasted water (ton)

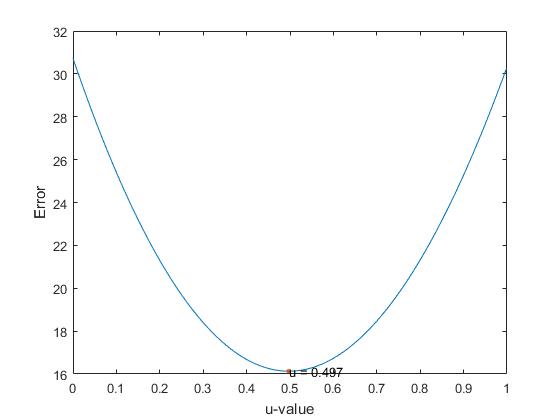


Figure 32: Error analysis of cadmium emission in wasted water (ton)

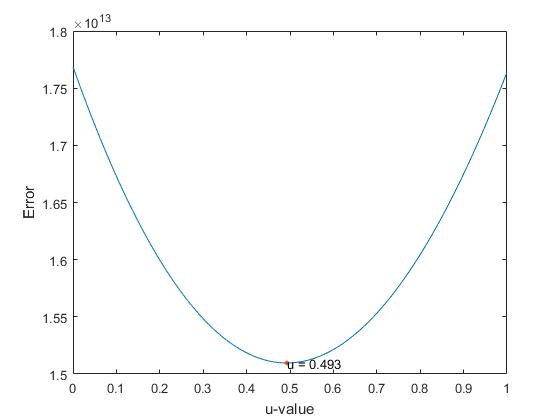


Figure 33: Error analysis of NOx emission (ton)