1. Title:

Using Markov Chain to analyze the concentration of pollutants in northern Taiwan

2. Background and Motivation

2.1 Motivation

In recent years, the air quality of Taiwan and especially southern Taiwan has been tremendously worsen to a degree that it could be harmful and even detrimental to human health. Outdoor activities are not recommended under certain conditions. The aim is to find out the concentration of different contaminants with respect to time (i.e month, season) in order to find the relation between pollutants and also the relation between pollutants and month or season. In that sense, it is possible to figure out the cause of the deadly air pollution and it is likely that by the result, people or government can be able to know the genesis and can try to reduce or ban the polluting activities.

2.2 Background

2.2.1 Introduction to Air Quality in Taiwan

Breathing is essential and vital for human or animal to survive. However, currently in Taiwan, it is harmful to breathe and people even say that going out is somehow chronically suicidal. It is to some degree a national safety issue and people will even be deprive of freedom in order to keep themselves away from pollutants.

According to an article published by Air Clean Taiwan Organization, the air quality is especially low in fall and winter in Taiwan. The degree of air pollution differ from the source of pollution, weather conditions (wind speed, wind direction, humidity etc.) and terrain.

2.2.2 Factors and Influences to the Air Quality in Taiwan

In the aspect of weather conditions, If in conditions when the atmosphere is in weak convection (no wind etc.), the pollutants can't diffuse easily and accumulate in the air, which cause the air quality to get worsen. Due to a lower specific heat of the land, in winter, temperature of the land drops faster and creates a temperature inversion , consequently reduce convection. Moreover, the northeast monsoon and continental air mass in fall and winter will bring the smog in China directly to northern Taiwan. In summer, due to cyclones and the active vertical convection of atmosphere, the air quality is better than in other seasons.

In the aspect of the sources of pollutants, the extremely common usage of automobiles and motorcycles in metropolitan areas of northern Taiwan causes serious exhaust emission. According to the data announced by Directorate General of Budget, Accounting and Statistics, the population density of Taiwan is the second highest in the world, and the density of the vehicles is even the highest in the world. In Taiwan, there are 403 motorbikes per kilometer square in average, which is ten times higher than that of Japan. In another expression, that is every Taiwanese has number of 1.5 vehicles. Moreover, in winter, engines operate in a lower temperature and gas pressure, these condition reduce the combustion efficiency and cause incomplete combustion, which makes the emission of PM2.5 and precursors increase.

2.2.3 Air Quality Index (AQI)

To compare the conditions of air quality under the same scales, we referenced to the official air quality measurement standard used in Taiwan to label our raw data, which is provided as the pollutants' concentration. By labeling different concentration of PM10 and PM2.5, we were able to divide daily average data into several groups, by using the AQI as the standard levels.

Air Quality Index(AQI)							
AQI	O_3 (ppm) 8hr	O_3 (ppm) 1-hr (1)	$PM_{2.5}$ ($\mu g/m^3$) 24-hr	PM_{10} ($\mu g/m^3$) 24-hr	CO (ppm) 8-hr	SO_2 (ppb) 1-hr	NO_2 (ppb) 1-hr
Good 0～50	0.000 - 0.054	-	0.0 - 15.4	0 - 54	0 - 4.4	0 - 35	0 - 53
Moderate 51～100	0.055 - 0.070	-	15.5 - 35.4	55 - 125	4.5 - 9.4	36 - 75	54 - 100
Unhealthy for Sensitive Groups 101～150	0.071 - 0.085	0.125 - 0.164	35.5 - 54.4	126 - 254	9.5 - 12.4	76 - 185	101 - 360
Unhealthy 151～200	0.086 - 0.105	0.165 - 0.204	54.5 - 150.4	255 - 354	12.5 - 15.4	186 - 304 (3)	361 - 649
Very Unhealthy 201～300	0.106 - 0.200	0.205 - 0.404	150.5 - 250.4	355 - 424	15.5 - 30.4	305 - 604 (3)	650 - 1249
Hazardous 301～400	(2)	0.405 - 0.504	250.5 - 350.4	425 - 504	30.5 - 40.4	605 - 804 (3)	1250 - 1649
Hazardous 401～500	(2)	0.505 - 0.604	350.5 - 500.4	505 - 604	40.5 - 50.4	805 -1004 (3)	1650 - 2049

Figure 1. Air pollutants concentration level in the Air Quality Index standard

The details of each air quality levels are listed in the following figure:

Impact on Health	Good	Moderate	Unhealthy for Sensitive Groups	Unhealthy	Very Unhealthy	Hazardous
Status Color						
Impact on Human Health	Air quality is considered satisfactory, and air pollution poses little or no risk.	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.	Members of sensitive groups may experience health effects. The general public is not likely to be affected.	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.	Health alert: everyone may experience more serious health effects.	Health warnings of emergency conditions. The entire population is more likely to be affected.

Figure 2. Details regarding to each standard level

2.3 Problem Definition:

This study aims to find out the probabilities of the concentration levels of different pollutants in northern Taiwan. The results are compared to the official publication by the Taiwan Environmental Protection Agency (EPA) and PM2.5 and PM10 concentration levels are discussed with the respect to time.

3. Methodology

3.1 Method to Consider Sequential Transition between Statuses

In this study, the target parameter is recorded with the respect to time, and thus the form is a sequential data. Under this requirement, a method that can examine the sequence is recommended. The Markov chain is a stochastic model describing a sequence of possible events in which in the long term the probability of each event depends only on the state attained in the previous event. It suits the scenario of our project that is a sequence of timeline and what we invision is that by using the Markov chain method, we will be able to find a transition probabilities matrix that converges. The statistics we obtained from the Taiwan Environmental Protection Agency, is the source of creating the initial transition matrix from our calculation, the resulting matrix will then be used to represent and predict the long term air quality levels.

3.2 Flow Chart Markov Chain Application on Long Term Air Quality

The overview of our processes is shown as the figure below. We began with preprocessing the raw data and rearranging the format to our convenience for later computation and analysis. Then by the daily average data, we can use python scripts to compute the frequency of each status-to-status transition, and by that we are able to obtain its annual and seasonal transition matrices. With the annual and seasonal Markov transition matrix, we can use python script to execute a for loop and find at which iteration that the matrix will began to converge. The results will be shown in part 4.3 and the detail process of preprocessing will be introduced in the next part.

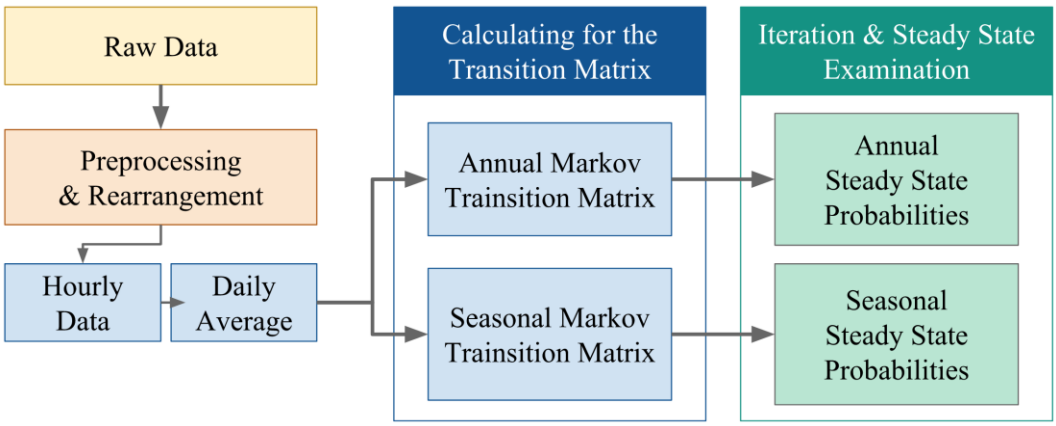


Figure 3. Flow chart of our data processing and application of Markov chain on obtaining the long term air quality

4. Data Collection and Analysis Result

4.1 Data Collection and Preprocessing:

Though the most severe location should be southern Taiwan. However, there doesn't exist that much monitoring stations in southern Taiwan. Thus, in this study we select Taipei as our candidate and this data is from Environmental Protection Administration, Executive Yuan, R.O.C. (Taiwan). There is air quality data and meteorological monitoring data in hourly basis for research and analysis (only include northern Taiwan 2015). There are 25 observation stations in Taipei that is taken into account including (Banqiao, Cailiao, Datong, Dayuan, Guanyin, Guting, Keelung, Linkou, Longtan, Pingzhen, Sanchong, Shilin, Songshan, Tamsui, Taoyuan, Tucheng, Wanhua, Wanli, Xindian, Xinzhuang, Xizhi, Yangming, Yonghe, Zhongli, Zhongshan).

In this project, we only selected PM10 and PM2.5 as the main concern subjects. However, there are some missing values in our data. Theoretically, there should be 25*60*365=219,000 data in PM10 and PM25 dataset respectively, but there are only 218640 data in the dataset actually. Besides, there are also some invalid values due to improper inspection. To simplified, we set all the missing value and invalid value to zero, and because the air quality ranking of PM10 and PM2.5 are in daily basis, we calculate the mean concentration of them by adding up the 24 data in the same day and divide by 24 minus the number of zeros. This is the method how we convert the hourly data into daily average data.

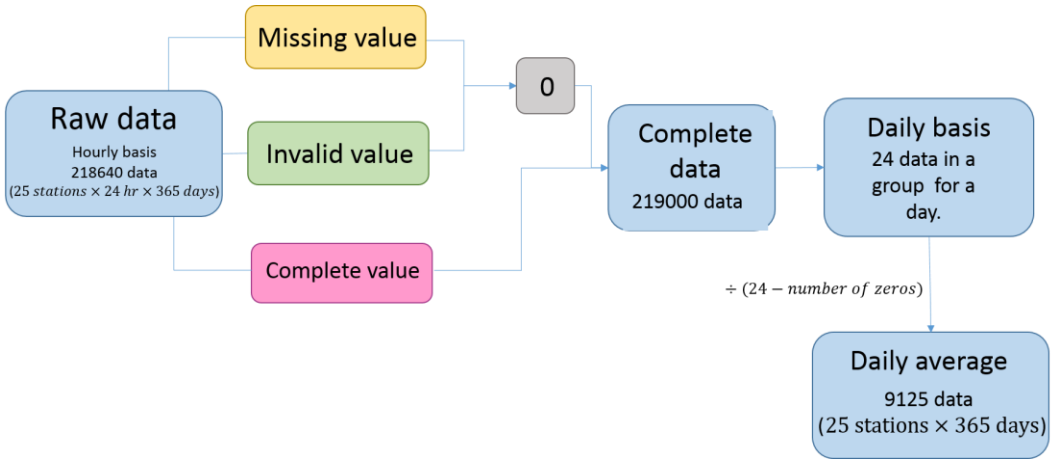


Figure 4. Flow chart of our data processing method

4.2 Model Formulation

In this study, the Markov chains method is used. The probabilities to form the transition matrix is obtained by calculation on the raw data obtained from the Taiwan Environmental Agency. After formulating the transition matrix, the long-term steady-state probabilities can also be calculated, and will be used to compare with the official

report and we will compare the insights provided by using Markov chain in the later discussion section.

From the raw dataset acquired, we found that only four color labeling status were reached, thus reducing the Markov transition matrix to a four-by-four matrix.

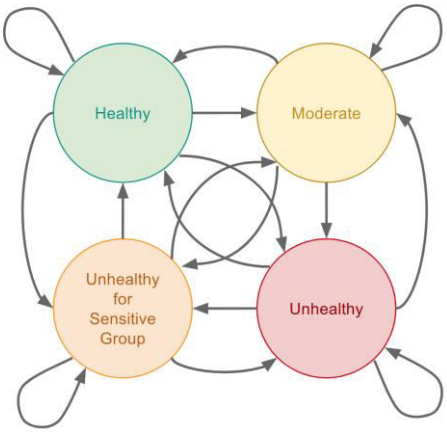


Figure 5. State diagram of the air quality level Markov chain

4.3 Result and Visualization

4.3.1 Annual PM10&2.5 probability

In this result, we can see that for PM10, the annual condition probable air condition is appropriate for 99.9% and for PM2.5 it is about 91.9%, This is in the aspect of good air condition. The other concerned is the bad condItion for PM10&PM2.5. We can see that for PM2.5, it has two high concentration class that PM10 doesn't reach or for very little portion which is "unhealthy for sensitive group" and "unhealthy".

This is due to the standard criteria differences for different pollutants. Though PM2.5 is included in PM10, we can draw some insight on this sense that under the same production condition, the origin might more prone to create PM2.5 rather than pure PM10 ranging from 2.5~>10μm.

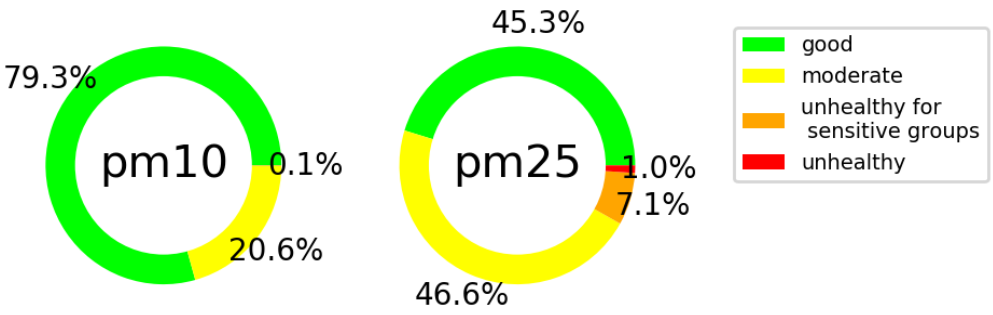


Figure 6. Results of the annual PM10 and PM2.5 air quality index level probability

4.3.2 Seasonal PM10 probability

In this figure, we can examine the effect of season more clearly. Roughly,the concentration of PM10 is acceptable for most people in northern Taiwan. We can observe that there are only three different status of PM10, and even the "unhealthy for sensitive groups" level, labeled in orange, seldom happens in fall and winter (only 0.1% for each). That is, in spring and summer, the concentration of PM10 only fall into the "good " and "moderate" level. For summer, the air quality is the best in all season with quite a big difference and it matches with the seasonal and the weather effect.

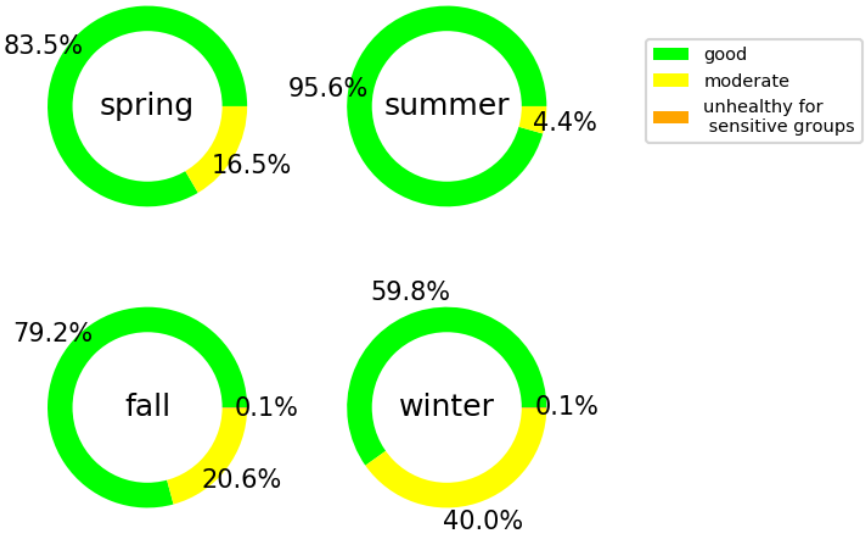


Figure 7. Results of the seasonal PM10 air quality index level probability

4.3.3 Seasonal PM2.5 probability

We can observe four different status for the PM2.5. The "unhealthy" level, labeled in red and means that the concentration of PM2.5 exceed the acceptable level for everyone, happens only in fall and winter. If we calculate how many days with high concentration of PM2.5 are there in the four seasons respectively, we can see that there are about 18 days in winter that are in "unhealthy for sensitive groups" and "unhealthy" level. Compare that with summer, which is only less than one days in "unhealthy for sensitive groups", it is significant that the condition of PM2.5 in winter is much severe than that in summer.

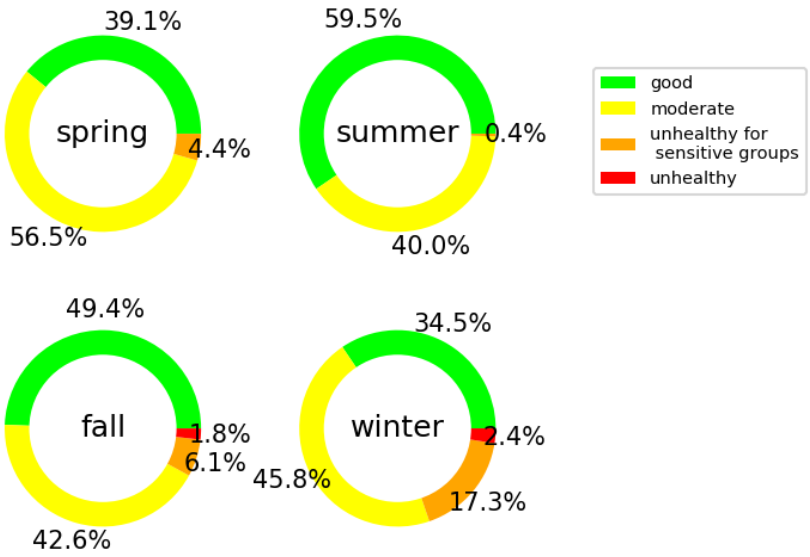


Figure 8. Results of the seasonal PM2.5 air quality index level probability

5. Conclusion

With the annual Markov transition probability matrix, there are 8.1% of the days that are not recommended for outdoor activities. That is, about a month that we should stay indoor, especially for the sensitive groups. Moreover, our results for the four seasons showed seasonality and matched the with the report published by the Taiwan EPA-- the air quality in winter is much more worse than that in summer.

With other weather and pollutant data , we can find out the steady state of them and further discuss whether they influence each other. Plus, if we can access to those data of other area in Taiwan (southern Taiwan, eastern Taiwan ,etc.), we can compare the difference of air pollution conditions between different areas to see that what factors may cause the difference.

6. Feedback from Professor Lee

6.1 Comment 1- NOx & SOx is highly related to PM2.5 & PM10

In order to answer some feedback given form professor. We tried to find more data from official website. In this case, we can see that NOx and SOx truely has a high relation with PM2.5.

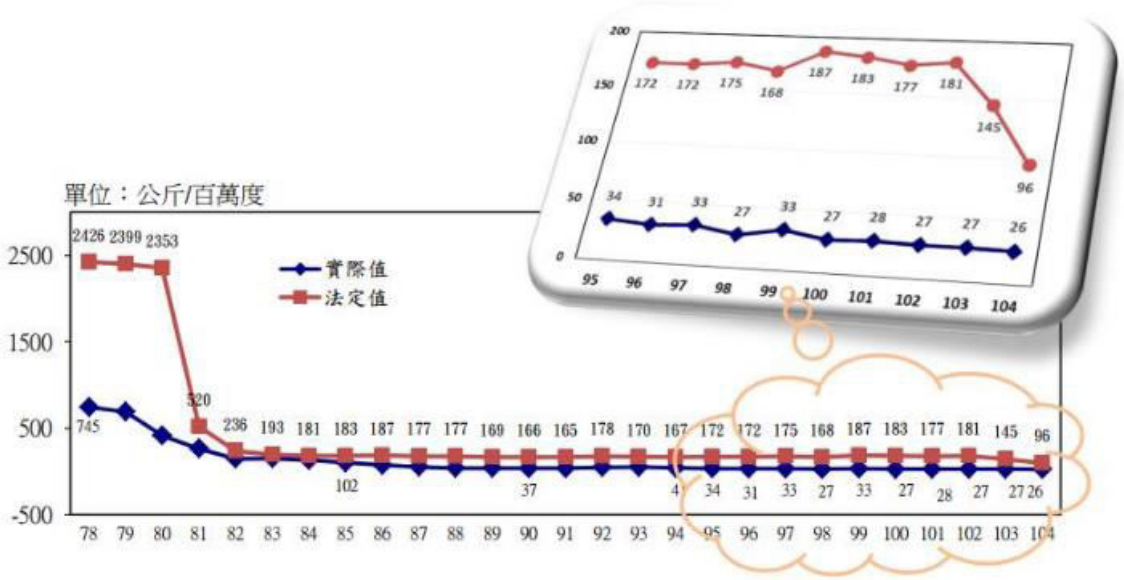
6.2 Comment 2 - The dominance of power plant on air pollution

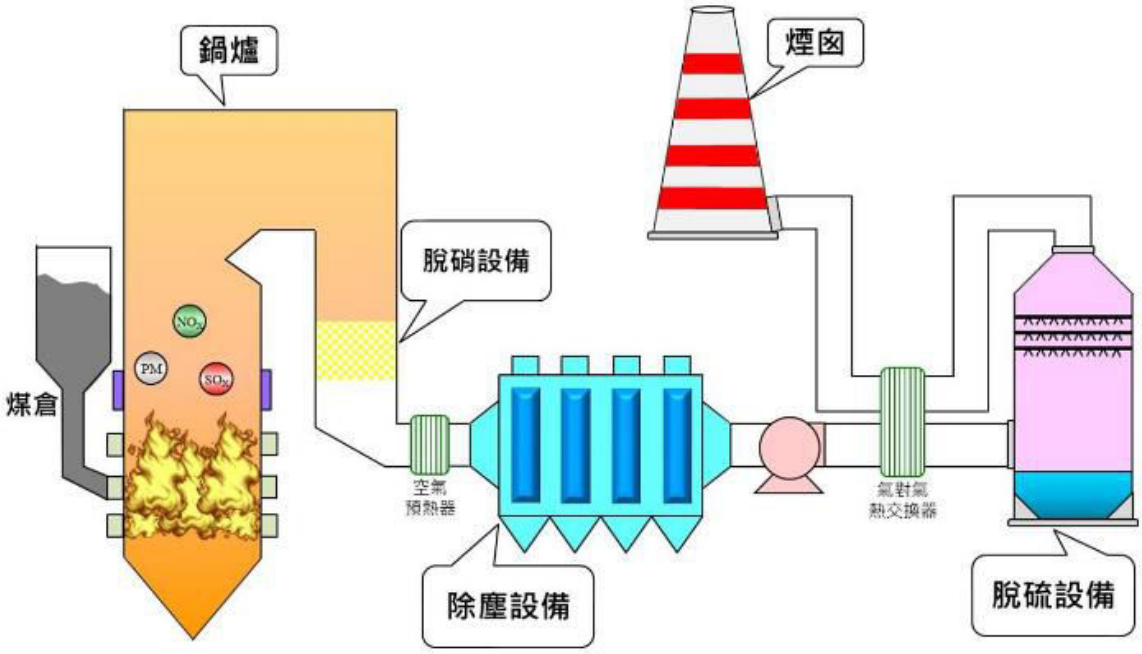
Actually, for the cause of PM2.5, the proportion of industrial and transportation factor weight almost the same. Moreover, industrial includes all means of high technology factories to traditional industries and power plants as well. It indicates that the portion of transportation is dramatically significant.

99年行業別						
	PM2.5		Sox		Nox	
	百萬噸/年	比例	百萬噸/年	比例	百萬噸/年	比例
工業	16865	23%	105261	88%	171600	40%
車輛	16756	23%	343	0%	217109	51%
非公路運輸	601	1%	8559	7%	21033	5%
商業	6440	9%	3651	3%	3684	1%
營建/道路揚鑣	27662	37%	0	0%	0	0%
露天燃燒	4601	6%	574	0%	6577	2%
其他	930	1%	574	0%	9657	2%
	73855		118962		429660	

6.2.1 The progress of local power plants

Though the politicians likes the attribute air pollution to the exhaust of power plant, Taiwan Power Company has put a lot effort on minimizing the power plant exhaustion on pollutants including the facilities to eliminate SOx, NOx and PMs. The result of elimination is even more successful that the overall exhaustion is lowered by 95% and the PM is even to 98% which is huge.





7. References

http://www.airclean.tw/index.php/2016-05-27-14-06-07/2016-06-09-03-52-44/120-2016-08-30-13-55-21
http://www.airclean.tw/index.php/2016-05-27-14-06-07/2016-06-09-03-52-44/121-2016-08-30-13-56-28
http://www.airclean.tw/index.php/2016-05-27-14-06-07/2016-06-09-03-52-44/121-2016-08-30-13-56-28
https://kknews.cc/zh-tw/science/g28l6xm.html
https://kknews.cc/zh-tw/science/g28l6xm.html
https://www.cw.com.tw/article/article.action?id=5054397
https://kknews.cc/zh-tw/science/g28l6xm.html
http://big5.cri.cn/gate/big5/gb.cri.cn/40151/2013/02/04/6351s4011820.htm