Study on COVID-19 prevention and control based on the principle of sewage epidemiology

Jingwen Liang*, Xiner Huang, Yuanni Niu

School of Economics and Finance, Zhanjiang University of Science and Technology, Zhanjiang, Guangdong, 524084

*Corresponding author. Email: jingwenliang22413@163.com

Abstract. Through the analysis and study of the current virus level of wastewater at each site on the website, the percentage change in the past 15 days, and the percentage of wastewater samples that can detect the virus in the past 15 days, this paper establishes an evaluation model based on the distribution of sewage monitoring sampling points and a K-Means clustering model, and finally helps to add ten new sewage monitoring sampling points, and analyzes the severity of the risk of future large-scale epidemic transmission in four major regions of the United States, providing early warning and corresponding prevention and control measures for local governments.

In this paper, the analytic hierarchy process is used for modeling, three keywords "population, environment and epidemic prevention" are extracted, and they are used as the criterion layer, which is then extended from the criterion layer. A total of nine three-level indicators are found as the index layer, and the scoring standard is established to judge the rationality of the distribution of sewage detection sampling points. The results show that the scores of the four major regions of the United States are more than 200, which is reasonable. In order to select ten new sampling points, K-means clustering algorithm is used to divide the data points into 10 regions based on the Euclidean distance between coordinate points. Finally, Louisiana, Nevada, West Virginia, Hawaii, South Dakota, Illinois, Georgia, Massachusetts, New Mexico and Montana are selected. In order to predict the possible large-scale epidemic situation in the four major regions of the United States, three indicators are selected: the current virus level of wastewater at each site, the percentage change data in the past 15 days and the percentage of wastewater samples that can detect the virus. The data are preprocessed by regions. The severity of the epidemic situation is estimated according to the epidemic transmission risk threshold. Finally, five suggestions are given: ① call on people to fight the epidemic together, Strengthen publicity and education on epidemic prevention and control. ②Timely respond to the complaints of local governments and epidemic prevention institutions, and implement epidemic prevention policies. ③ An emergency plan should be established to strengthen the emergency handling capacity. ④ Establish and improve the supervision mechanism to reduce the infection rate of new crown.

Key words: analytic hierarchy process, K-means cluster analysis, sewage detection

1. Introduction

Covid-19, New Coronavirus (SARS-CoV-2), called "new crown pneumonia", is named after WHO. "2019 coronavirus" refers to the virus strain first found in Wuhan, Hubei Province since December 2019. The World Health Organization attaches great importance to this and calls on the world to fight the epidemic together. However, New Coronavirus's vitality is tenacious and easy to survive. New Coronavirus has undergone several variations, and 11 variants have been found, including the new strain of BA.. Subtype 2 strain, delta variant strain, b.1.640 strain, ay 4.2 strain, secondary variant ay 4. B.1.621 variant strain, c.1.2 variant strain, Mu variant strain, etc. Among them, Omicron is the most popular and has strong transmission power. It has the characteristics of strong concealment, which makes the virus more prone to multi-point distribution or concentrated outbreak. Sewage monitoring can expand the scope of epidemic monitoring and evaluate the overall infection situation of large-scale communities. Even the asymptomatic infection of COVID-19 can also detect the virus in the excreta. Through the investigation and examination of the waste water and domestic water in the waste water, the monitoring of the waste water can catch the COVID-19 contained in the
body. That is to say, the monitoring of wastewater plays an early warning role in COVID-19's investigation.

2. Evaluation of reasonable selection of sampling points for sewage detection

The selection of indicators is the cornerstone of the construction of the evaluation system. In order to make the indicator system scientific and standardized, the selection of indicators should follow the principles of systematization, typicality, dynamics, conciseness, scientificity, conciseness and comprehensiveness. A good indicator system can monitor business changes. When there are problems in the business, the problem can be traced through the indicator system, which can accurately locate the problems and feed back to the business to solve the corresponding problems.

2.1. Sewage monitoring index structure

When setting the index of Sewage Monitoring for the purpose of testing COVID-19, the sewage monitoring data are taken as the target level. This paper studies the three key words "population, environment and epidemic prevention" and takes it as the standard level. In order to measure the correlation between the population, environment, epidemic prevention and the establishment of sampling points in the area, the nine level three level indicators are used as indicators level to establish the standard for evaluation. Judge the rationality of the distribution of sewage detection sampling points.

2.2. Analytic hierarchy process

AHP is a research method that combines qualitative and quantitative methods to calculate decision weight to solve multi-objective complex problems. It is difficult to use the weight of each decision-making method to compare the relative importance of each decision-making method with that of each decision-making method, and it is difficult to use the weight of each decision-making method to measure the relative importance of each solution.

Establish a hierarchical structure, which is divided into target layer, criterion layer, index layer and scheme layer. Take the sewage detection distribution as the target layer, the primary influencing factors as the criterion layer, the secondary influencing factors as the index layer, and the evaluation as the scheme layer.

Construct a judgment matrix, which represents the relative importance of elements in this layer, usually 1, 2, 5 and the reciprocal. For the criterion layer judgment matrix and index layer judgment matrix, due to the lack of basic data, judge the relative importance of each element according to experience, and then assign values to each element of the judgment matrix.

Eigenvector calculation: calculate the n-th power root of the product of elements in each row of the judgment matrix.

\[ M_i = \sqrt[1^n]{\prod_{j=1}^{n} a_{ij}} \]  

(1)

Weight calculation: normalize MI.

\[ \omega_i = \frac{M_i}{\sum_{i=1}^{n} M_i} \]  

(2)

Calculate the maximum eigenvalue of the judgment matrix: this value will be used in combination with the CI value for consistency inspection

\[ \lambda = \sum_{i=1}^{n} \frac{(A\omega)_i}{nw_i} \]  

(3)
Consistency test analysis, when constructing the judgment matrix, there may be logical errors, such as a is more important than B, B is more important than C, but C is more important than a. Therefore, it is necessary to use the consistency test to check whether there is a problem. The consistency test uses the CR value for analysis. If the CR value is less than 0.1, it indicates that it has passed the consistency test, otherwise it indicates that it has not passed the consistency test. If the data fails to pass the consistency test, it is necessary to check whether there are logical problems and re-enter the judgment matrix for analysis.

\[
CI = \frac{(\lambda - n)}{(n-1)}
\]

(4)

\[
CR = \frac{CI}{RI}
\]

(5)

Analysis conclusion: if the weight has been calculated and the judgment matrix meets the consistency test, can draw a conclusion and continue to make further analysis

2.3. Weight analysis

| Criterion layer   | Northeast China | Central and Western | Southern Region | Western Region |
|-------------------|-----------------|---------------------|-----------------|----------------|
| population        | 0.5             | 0.5                 | 0.3             | 0.3            |
| environment       | 0.3             | 0.2                 | 0.3             | 0.2            |
| Epidemic prevention| 0.2             | 0.3                 | 0.4             | 0.5            |

The northeast of the United States, including 11 states such as New York, Pennsylvania and Vermont, is the richest region in the US. Therefore, the population density is large, which has also become one of the reasons why the new crown mortality rate in this region accounts for the highest proportion in the US. In terms of environment, the climate in this area is temperate continental climate, with precipitation higher than 400mm and average temperature In July, the highest temperature was 28 degrees Celsius. The average minimum temperature in January was -10 degrees. New Coronavirus generally survived below 56 degrees. Under these conditions, the survival time was longer and it was difficult to inactivate. In addition, because of the problem of climate and environment, many people have migrated from the northeast "cold zone" to the southern "sunshine belt", increasing the risk of infection of New Coronavirus in the process of population migration, thus increasing the difficulty of epidemic prevention and control in the area, and also causing the mortality of new crown infection to rise.

Table 2. Scoring criteria for measuring the rationality of the distribution of sewage detection sampling points

| population | environment | Epidemic prevention |
|------------|-------------|---------------------|
| Weight index | Weight index | Weight index |
| Below 0.2: 60 points | Below 0.2: 60 points | Below 0.2: 60 points |
| 0.3-0.4:70 | 0.3-0.4:70 | 0.3-0.4:70 |
| 0.4-0.5:80 | 0.4-0.5:80 | 0.4-0.5:80 |
| Above 0.5: 90 points | Above 0.5: 90 points | Above 0.5: 90 points |

To sum up, although the distribution of Sewage Monitoring and sampling points in the four major regions of the United States is basically reasonable, there are two deficiencies. First, the distribution of sewage monitoring sampling points is uneven, and there are no sampling points in a small number of areas, which can not timely predict the future development trend of the epidemic in this area and formulate corresponding prevention and control measures. Second, the sampling points in some areas are too dense, resulting in cross coincidence, resulting in a waste of resources.
3. Establish ten sewage detection and sampling points

3.1. K-means cluster analysis

Cluster analysis is a center based clustering algorithm (K-means clustering). Through iteration, the samples are divided into k classes to make the similarity of data objects in the same cluster as large as possible, and the difference of data objects not in the same cluster as large as possible. That is, after clustering, the data of the same class shall be gathered in DEUTZ as much as possible, and different data shall be separated as much as possible.

If K objects are arbitrarily selected from data set D as the initial cluster center, the number of coarse results can be determined by hierarchical aggregation algorithm, and an initial cluster can be found, and then the cluster can be improved by iterative relocation.

Calculate the distance from P to K cluster centers of each cluster object. The commonly used distance measurement methods include Euclidean distance (Euclidean distance) and cosine similarity. Euclidean distance calculation:

\[ d_{ij} = \sqrt{\sum_{k=1}^{n}(x_{1k} - x_{2k})^2} \]  

Assigning the cluster object p to the cluster closest to it (the shortest distance); Again, calculate the mean value of each cluster as the center of the new cluster. Until the cluster center of K clusters does not change, it stops, that is, the maximum number of iterations or the objective function is optimal.

| Table 3. Cluster center value |
|-------------------------------|
| Cluster type | Center longitude | Central latitude |
| 1            | -78.01197143     | 38.54422714     |
| 2            | -107.9468        | 36.3859         |
| 3            | -155.58          | 19.8903         |
| 4            | -97.2204         | 42.60924        |
| 5            | -112.5585        | 45.4679146      |
| 6            | -92.63825        | 33.377225       |
| 7            | -118.803         | 39.7889666      |
| 8            | -86.2637333      | 40.4117         |
| 9            | -72.3939375      | 41.9671800      |
| 10           | -81.8661667      | 31.2160821      |

Firstly, the relevant data of 42 states with sewage detection sampling points are collected. Using their longitude and latitude data and using the method of cluster analysis (K-means), the K value is set to 10, that is, 44 states are clustered into ten clusters, and the cluster center points of these ten clusters are obtained, that is, the central longitude and latitude of ten clusters are obtained (Table3). Based on the regional distribution, the clustering centers mentioned above are the ten central points that represent the sewage test data of the United States, corresponding to the ten central regions of the United States. The data of the sewage detection data in the state closer to the cluster center are more representative, and more sewage sampling points are needed. More sewage test data are needed, reflecting the situation of COVID-19 in the whole country from the virus level in the sewage.

3.2. Euclidean distance

Euclidean distance is a commonly used definition of distance, which refers to the real distance between two points in m-dimensional space, or the natural length of the vector (that is, the distance from the point to the origin). The Euclidean distance in two-dimensional and three-dimensional space is the actual distance between two points.
3.3. Analysis results

Table 4. States closest to the cluster center

| Clustering | State (USA)   | Difference between cluster centers |
|------------|--------------|-----------------------------------|
| 1          | Louisiana    | 2.734916336                       |
| 2          | Nevada       | 2.57671829                        |
| 3          | West Virginia| 0.933713344                       |
| 4          | Hawaii       | 0                                 |
| 5          | South Dakota | 2.029473295                       |
| 6          | Illinois     | 1.135510656                       |
| 7          | Georgia      | 0.191318079                       |
| 8          | Massachusetts| 0.840433734                       |
| 9          | New Mexico   | 3.490265637                       |
| 10         | Montana      | 2.601911564                       |

By using the Euclidean distance formula, the distance difference between each state and the cluster center of ten clusters can be obtained. The point with the least distance is the most reasonable location. The smaller the difference from the cluster center point, it shows that the sewage detection data is more representative for the region. As shown in the above figure (Table 4), the difference between Hawaii, Georgia, Massachusetts and West Virginia and the cluster center is small, which indicates that their sewage detection data are more representative for their cluster. According to the above table (Table 4), increasing the number of sewage sampling points in ten states can make the sewage monitoring data more perfect and have a clearer understanding of the situation of COVID-19 in different parts of the United States.

4. Selection of risk indicators to evaluate the spread of large-scale epidemic in four regions

4.1. Solution of index weight

With the evaluation index defined above, the index is standardized to eliminate the influence of different attributes of the sample with different orders of magnitude. Based on the attribute type of the original index, the standard 0-1 transformation and the given optimal interval method are used for dimensionless and normalization. The current virus level of wastewater at each site, the percentage change data in the past 15 days and the percentage of wastewater samples that can detect virus in the past 15 days are set as x1, x2 and x3 respectively. Therefore, there are:

\[
d_{ij} = \sqrt{\sum_{k=1}^{3} (x_i - x_j)^2}
\]  

Where \(d_{ij}\) is the standardized value of each evaluation index and the maximum and minimum values of max (x) and min (x) of the evaluation index.

Set a transmission risk threshold. Exceeding this threshold indicates the risk of large-scale epidemic transmission. If Q is below 0.1, it is serious, if Q is 0.1-0.14, it is very serious, if Q is 0.15-0.19, it is very serious, and if Q is 0.2 and above, it is super serious. As shown in Table 5

Table 5. Scoring criteria for measuring the severity of transmission risk in four regions

| Q is below 0.1 | Q is 0.1-0.14 | Q is 0.15-0.19 | Q is 0.2 and above |
|----------------|--------------|---------------|-------------------|
| Serious        | Very Serious | Very Serious  | Super serious     |
Since $x_3$ only shows the percentage of wastewater samples with positive sars-cov-2 rna in each location in the past 15 days, and does not show the overall level of sars-cov-2 in wastewater, the average of $x_1$ and $x_2$ is finally selected for weight $Q$ calculation:

$$Q = x_1 \cdot x_2$$

(9)

Where $q$ is the transmission risk threshold of the four regions, and the data are shown in Table 6.

**Table 6. Transmission risk value of four regions**

| index | Northeast | Southern | Central and Western | Western |
|-------|-----------|----------|--------------------|---------|
| X1    | 0.47      | 0.31     | 0.21               | 0.39    |
| X2    | 0.53      | 0.48     | 0.45               | 0.43    |
| Q     | 0.25      | 0.15     | 0.09               | 0.18    |

In conclusion, the threshold of epidemic situation in Northeast China is 25.0; The risk threshold of epidemic transmission in southern China is 0.15, which is very serious; The risk threshold of epidemic spread in the central and western regions is 0.09, which is serious; The risk threshold of epidemic transmission in the western region is 0.18, which is very serious.

5. Conclusion

Through the research and analysis in this paper, it is considered that the environmental monitoring of sewage has a certain early warning effect on the prevention and control of COVID-19. Analyze and understand the current virus level of wastewater at each site, the percentage change data in the past 15 days and the percentage of wastewater samples that can detect viruses in the past 15 days. In order to predict the future risk transmission of the epidemic in the four major regions of the United States, set a transmission risk threshold, set the interval corresponding to different degrees of severity, and finally formulate corresponding protective measures and give suggestions according to the severity of each region.

This paper focuses on the evaluation and prediction of four regions in the United States, namely the northeast, the South and the middle Western region, western region. First, establish the transmission risk threshold. Exceeding this threshold indicates the risk of large-scale epidemic transmission. If it is less than 0.1, it is serious; if it is 0.1-0.14, it is very serious; if it is 0.15-0.19, it is serious; very serious, 0.2 and above is super serious. Northeast China is a heavy industrial area with an extremely developed economy, With large population density and concentrated distribution, the risk threshold of epidemic transmission in Northeast China is 0.25, which belongs to super serious area; The southern region has a warm and pleasant climate, profound and complex cultural heritage, many people of different races and a relatively large population. The risk threshold of epidemic transmission in the southern region is 0.15, which belongs to a very serious region; Most of the central and western regions have a temperate continental climate with four distinct seasons, sparse precipitation, sparsely populated land and developed agricultural economy. People mostly live in rural towns with scattered population and low density, so it is not easy to outbreak the epidemic. The risk threshold of epidemic transmission in the central and western regions is 0.09, which belongs to serious areas; As a link and bridge connecting Asia, the western region has continuously attracted a large number of outstanding talents to start businesses here and realize their dreams. The developed tertiary industry and high-tech industry have led to the rapid rise of the economy in the western region. Many people come here one after another, with strong population mobility and high population density. The climate of the western region is Mediterranean climate and temperate marine climate, with abundant precipitation and beautiful environment. More young people will choose to settle here. The risk threshold of epidemic transmission in the western region is 0.18, which is very serious.
Reference

[1] Cheng Rong, Shi Lei, Zheng Xiang Review on the hot spots of sewage epidemiology in 2021 [J] Science and technology Herald, 2022,40 (01): 150-160

[2] Gawolan zunong, Nubia amarjiang, Xiao Huidi, Hu Yifei Research Progress on epidemiological monitoring and early warning of New Coronavirus sewage based on community health level [J]. Chinese Journal of preventive medicine, 2021,55 (08): 1016-1021

[3] Liu ranbin, Hao Xiaodi, Mark van Loosdrecht, Jiang Han Advances in covid-19 epidemiology for new crown pneumonia early warning [J]. China water supply and drainage, 2021,37 (14): 37-45

[4] Pan Yuwei, Mao Kang, tuerk Franziska, Yang Zhugen Application progress of electrochemical biosensors in sewage analysis and sewage epidemiology [J] Electrochemistry, 2019,25 (03): 363-373

[5] Yang Shuo Study on the application of sewage epidemiological method to investigate urban tobacco consumption [D] Dalian Maritime University, 2016

[6] Zhang Tong, Xu Haoguang Sewage monitoring New Coronavirus anti epidemic [J]. Science Bulletin, 2021,66 (34): 4354-4357

[7] Li Guotao Strictly implement various prevention and control measures, and firmly build Hengshui defense line for epidemic prevention and control [n] Hengshui daily, 2022-04-08 (A01)

[8] Strive to achieve the maximum effect of epidemic prevention and control at the least cost [n] Public health news, March 29, 2022 (002)

[9] Our commentator Strictly and tightly grasp the normalized epidemic prevention and control [n] Yuxi daily, March 23, 2022 (001)

[10] Du Qingfeng, Zhou Zhiheng, Wu Xiaoliang, Wang Jiaji, Han Jianjun "General practice +" medical community mode New Coronavirus disease prevention community grid prevention and control experts recommend [J]. Chinese general practice: 1-9