Bayesian Analysis on the Spatial Difference of Input Risk of Overseas Cases of COVID-19 in China

Bo Yang\textsuperscript{1}, Yunyuan Yang\textsuperscript{2}, Wei Zheng\textsuperscript{1}, Yanmei Li\textsuperscript{1}, Xinping Yang\textsuperscript{1, *}

\textsuperscript{1}School of Mathematics and Computer Science, Chuxiong Normal University, Chuxiong, China
\textsuperscript{2}School of Environment and Chemistry, Chuxiong Normal University, Chuxiong, China

Email address: Yangxp@cxtc.edu.cn (Xinping Yang)
*Corresponding author

doi: 10.11648/j.ijsd.20230901.15

Received: February 24, 2023; Accepted: March 20, 2023; Published: March 31, 2023

Abstract: To analyze the spatial difference of COVID-19 import risk is helpful for scientific prevention and control. On the basis of clustering 25 provinces and cities with epidemic input in study time, a multinomial distribution model was established under the Bayesian framework. All parameters Bayesian estimation was obtained by MCMC method. 25 provinces and cities with overseas input were divided into 9 categories from March 3 to April 23, 2020. 468 overseas input risk values are regarded as parameters, and the maximum MC-error estimated by Bayesian is only 0.677% of the standard deviation. During the study period, 25 provinces and cities have input risk. The highest risk areas of overseas import are 12 provinces and cities in the first category represented by Beijing, Shanghai and Guangdong Province, including 10 provinces and cities along the coast / border. The lowest risk areas are the eighth category (Henan Province) and the ninth category (Anhui Province); the fourth category (Heilongjiang Province and Shanxi Province) risk is higher than the first category in 7 days and it has the largest input vary fluctuation. Taking 2020-3-22, 4-7 and 4-18 as time nodes, the overseas input risk is divided into four stages. In the first stages, the highest risk of overseas import is the first category (59.613%); in the second and third stages are the first category (decline from 60.505% to 37.056%), the fourth category (increase from 16.071% to 33.852%); in the fourth stage, the first category (42.622%), the third category (Shaanxi Province and Jilin Province, 17.556%) and the fourth category (10.056%).

Keywords: COVID-19, Overseas Input Risk, Multiple Distribution, MCMC Method, Bayesian Estimation

1. Introduction

As of 2020-3-28, there were 11 days in Wuhan, the worst affected area of the epidemic in China, with 0 new cases. On the same day, the Ministry of Foreign Affairs and the State Administration of Immigration issued a notice suspending the entry of foreigners holding valid Chinese visas and residence permits from 0:00 on March 28. 2020-4-8, Wuhan lifted the blockade, but the number of imported cases abroad was 1103. By 2020-4-23, 1618 cases had been recorded for overseas imports. China's epidemic prevention work has shifted to the importation of overseas cases (including overseas importation of associated cases) and asymptomatic infection prevention and control. Therefore, the spatial difference analysis of the risk of overseas imported cases can serve the decision-making of epidemic prevention and control in various provinces (cities).

The existing quantitative study of COVID-19 mainly focuses on the description statistics of epidemic trends [1, 2] and trend estimation. For examples, estimates of the size of the COVID-19 outbreak were made using SEIR or modified SEIR models [3-5], self-regression moving average models (ARIMAs) [6], random transmission models [7], etc; COVID-19 regeneration estimation was carried out using the Marocf Monte Carlo Method (MCMC) [8-9] and so on. Bayesian method has some advantages in data analysis in the field of medicine, the uncertainty parameter value can be quantified under the premise of known data, and the quantitative accuracy of parameters can be improved by a priori information and data information [10]. For example, Han Ke et al. [11] used the Poisson distribution model under
Bayesian framework to estimate the number of COVID-19
reurrences in first-tier cities. On the study of spatial
differences in the COVID-19 outbreak, the Johns Hopkins
Center for Systems Science and Engineering produced a map
of the global outbreak [12], with a maximum of more than 2
billion daily visits. Guan Weijie et al. [13] and Qi Cuifang et
[14] carried on the study of nationwide clinic characteristic of
epidemic situation and COVID-19 inter-provincial
communication and influencing factor analysis respectively,
and both of them used GIS mapping to reflect the distribution
of confirmed cases in each province and city. Yi Dali [15] and
others carried out cluster analysis with the epidemic data in 34
provinces and cities from 2020-1-19 to 2-16, a total of 6
categories, among them, it was the high risk areas of Hubei
Province and Henan Province that needed to be strictly
controlled. According to Tencent location data and Baidu
migration data, Liu Zhang [16] et al. have completed the
spatial distribution estimate of people who moved out of
Wuhan during the COVID-19 outbreak. In the scale of Wuhan City
1140 traffic analysis area, Feng Mingxiang [17] and
others carried out COVID-19 space-time diffusion estimate
combined with mobile phone user space interaction data. GIS,
multi-source data and big data platform are effective methods
for the study of COVID-19 outbreak simulation and spatial
distribution differences. However, the analysis and processing
of zero expansion data, missing data and short-term data by
these methods often result in a great deviation from the actual
situation. At present, foreign imported case data in China have
the characteristics of zero expansion, geographical absence,
data size differences, the result may deviate from the actual
situation of the current domestic and foreign input if using the
traditional method for analysis and processing. Based on the
number of overseas imported cases from 2020-3-3 to 4-23 and
the clustering results, this paper constructs a multinomial
distribution model of the probability of input risk in each
province and city under the framework of Bayesian, solves the
model by MCMC method, and analyzes the spatial differences
in the input risk of cases of COVID-19 outside China. The
research results are expected to serve for epidemic prevention
and control abroad.

2. Data and Methods

Collect daily new case data from 25 provinces (cities) in
China involved in overseas imported cases from 2020-3-3 to
4-23 as a sample (data from the Bulletin of the National Health
And Health Commission and the Provincial and Municipal
Health and Construction Commission). The zero expansion
characteristics of this data are obvious, the daily data of
provinces and cities are different, and the regional differences
are obvious. For example, among 25 provinces and cities,
Heilongjiang imported 86 cases in 2020-4-7, and it was the
largest number of imported cases day by day, however there
were 0 imported cases in 18 provinces in the same day. From
the frequency (proportion) point of view, the frequency of
foreign import cases was 0, but this could not be explained
that there were no risk of overseas import in 4-7 in these 18
provinces and cities. If we choose the methods used in the
literatures to deal with these data, it inevitably results in most
time points not in line with the actual situation.

By using two clustering methods to cluster data from 25
provinces and cities with overseas outbreak input in China, the
results can be obtained relatively close. Based on the cluster
results, the model of the probability of input risk abroad under
Bayesian is established, the appropriate Dirichlet distribution
is used as its priori distribution, the model parameters (input
risk) are solved by MCMC method, and the GIS mapping is
used to reveal the spatial difference of the input risk of
COVID-19 cases abroad. The software used are R language
and OpenBUGS, and the map mapping software are Adobe
Illustrator and Photoshop.

3. Model of Import Risk for Overseas
Cases

3.1. Clustering of Imported Cases from Abroad

Programs are written in R language to cluster sample data
(2020-3-3 to 4-23) from 25 provinces and municipalities. The
name of the province and city is considered an indicator, the
similarity coefficient between the two indicators $X_i, X_j$ ($i \neq
j$) is defined as

$$c_{ij} = \frac{x_i^T x_j}{\|x_i\|^2 \|x_j\|^2},$$

among $\| \cdot \|_2^{1/2}$ represents
the square root of two norm of the vector [18], and the
distance between the two variables is $d_{ij} = 1 - c_{ij}$.

Usually the results of different clustering methods are
different, and the determination of the number of
classifications is not yet fully resolved. In the actual study,
according to the purpose of the study, we generally choose a
variety of methods of clustering, through comparative analysis,
to determine the final clustering method and classification
number to get better results [18].

Through the comparative analysis of various clustering
results, it is found that the clustering results of the class
average method (Average) and the similar method (Mcquitty)
are close, and the average method of the class is not
concentrated or expanded, which is the clustering method
recommended by many literatures [19]. By combining the
advantages of the two clusters, the following clustering results are
obtained:

- Category I (G1): Zhejiang, Beijing, Shanghai, Guangdong,
  Sichuan, Shandong, Yunnan, Guangxi, Fujian, Tianjin,
  Liaoning, Jiangsu, a total of 12 provinces and cities. The
corresponding sample data from 2020-3-3 began to form a
data matrix of the lower trapezoidal structure, the largest
number of cases, up to 917 cases.

- Category II (G2): Jiangxi, Chongqing, Guizhou, 3 provinces
  and cities. Input cases were concentrated in 3-21 to 3-28, and
  the number was small, only 6 cases.

- Category III (G3): Shaanxi, Jilin, 2 provinces. Input cases
  were concentrated in 3-21 to 4-23, with continuous
  importation of overseas cases, a total of 49 cases.

- Category IV(G4): Heilongjiang, Shanxi, 2 provinces. Input
cases were concentrated in 3-18 to 4-23, with continuous importation of overseas cases, a total of 445 cases.

Category V (G_5): Hebei, Hunan, 2 provinces. Input cases were concentrated in 3-21 to 4-15 cases, a total of 11 cases.

Category VI (G_6): Inner Mongolia. Input cases were concentrated in 3-24 to 4-15 cases, a total of 118 cases.

Category VII (G_7): Gansu. Input cases were concentrated in 3-5 to 4-5 cases, a total of 47 cases.

Category VIII (G_8): Henan. Input cases were concentrated in 3-11 to 3-25 cases, a total of 3 cases.

Category IX (G_9): Anhui. Only one input case in 4-8.

By the definition of multinomial distribution, we get

\[ p(\theta_1 | data_t) = \Gamma(\alpha_t + 1) \prod_{j=1}^9 \frac{p_{tj}}{\Gamma(\alpha_t + 1)}, \tag{1} \]

where \( \Gamma(\gamma) = \int_0^\infty x^{\gamma-1}e^{-x}dx \). According to proposals such as NguyenX (2016) [20], the Dirichlet distribution is used as a priori distribution, and the number of provinces and cities contained in each category in the clustering results is used to select prior parameters, namely,

\[ \alpha = (12,3,2,2,1,1,1,1)^T, \]

The priori density is \( p(\theta_i) \propto \prod_{j=1}^9 \theta_{ij}^{\alpha_j - 1} \)

and we obtain the post-test exact distribution of parameter \( \theta_i \):

\[ \theta_i \sim \text{Dirichlet}(\alpha_1 + g_{i1}, \alpha_2 + g_{i2}, \alpha_3 + g_{i3}, \ldots, \alpha_9 + g_{i9}) \]

and the corresponding post-test density function is

\[ p(\theta_i | data_t, \alpha) = \Gamma(\alpha_t + 25) \prod_{j=1}^9 \frac{p_{tj} \prod_{i=1}^{25} \sum_{j \in G_i} p_{tj}}{\Gamma(\alpha_t + 1)} \tag{2} \]

By (2) we can get the full condition distribution of nine parameters, for example, in the case of \( \theta_1 \), the corresponding full-condition distribution is

\[ p(\theta_1 | \theta_{(-1)}, data_t, \alpha) \propto p_1^{12+g_{11}} = \text{Beta}(13 + g_{11}, 1). \tag{3} \]

Formula (3) is a standard distribution, the other eight full-condition distributions are also beta distribution. We can use Gibbs sampling to obtain all the parameters (a total of 52 days, 9 categories per day, 52 x 9 = 468 parameters) of the post-test sample, and complete Bayesian inference.

### 3.2. Bayesian Model

#### 3.2.1. Construction of Model

Any confirmed case of new import outside the province \( i \) of \( t \)-day is indicated by \( b_{ti} \), the \( \alpha_{te} \) indicates the number of new cases imported from outside on the \( t \)-day, and the number of cases entered from the province of category \( j \) on \( t \)-day is indicated by \( g_{tj} \), then \( \alpha_t = \sum_{i=1}^{25} b_{ti} = \sum_{j=1}^{9} g_{tj} \). The risk of foreign input for category \( j \) in \( t \)-day is defined as:

\[ P(b_{ti} \in G_j | \alpha_t \in G_j) = p_{ti} \sum_{j=1}^{9} p_{tj} = 1. \]

This probability expresses the risk of foreign input for category \( j \) in \( t \)-day. The higher the value, the greater the risk of the overseas COVID-19 input of category \( j \) on \( t \)-day, the greater the pressure of prevention and control from overseas input. The number \( \alpha_{te} \), new overseas imported cases in the nine categories on day \( t \)-day is considered as \( \alpha_{te} \) independent tests, and each confirmed case imported from abroad is considered to be a test. The test result can only belong to one of the nine categories, the risk probability of the case belonging to category \( j \) is \( p_{tj} \) and it is an evaluable parameter. Vectors consisting of the parameters to be evaluated and data of new overseas imported cases in the nine categories on \( t \)-day are recorded as \( \theta_t \) and \( data_t \), namely:

\[ \theta_t = (p_{t1}, p_{t2}, p_{t3}, p_{t4}, p_{t5}, p_{t6}, p_{t7}, p_{t8}, p_{t9})^T, data_t = (g_{t1}, g_{t2}, g_{t3}, g_{t4}, g_{t5}, g_{t6}, g_{t7}, g_{t8}, g_{t9})^T. \]

Figure 1. Cluster tree of daily number of imported cases of 25 provinces or cities from March 3 to April 8, 2020.

Note: Use Chinese Pinyin as the variable name of 25 provinces or cities, in order to avoid the same name in Pinyin, use shangxi to represent Shanxi.

### 3.2.2. Solution of the Model

The solving of the model is completed in the OpenBUGS software environment. After seeding random numbers, the software automatically generates the initial value of 468 parameters. In order to reduce the self-correlation among the post-test samples and to ensure the convergence of the MC chain, the sampling step of the sampling interval is 100. The orderly loosening algorithm [21] is used to eliminate random walking in the MCMC. After \( 10^5 \) iterations, posterior samples (468 MC chains) of 468 parameters to be evaluated was obtained. After each chain throws away the first 4999 samples of the burnin period, the parameters \( \theta \) are inferred by the MC method using the remaining 29619 samples. The corresponding parameter estimates are shown in Table 1 (9 categories, 52 values per category). Table 1 includes mean estimates for each parameter (Mean), median estimates (Median), 95% confidence interval CI: (2.5% q1, 97.5% q2), standard deviation (SD), and MC error.
(MCerror). See table 2 for the MC errors of category I to IX overseas input with greatest risk and the time point of occurrence. Each has a maximum MC error of less than $5.56 \times 10^{-4}$. The maximum of $(\text{max.MCerror}/\text{sd}) \times 100\%$ is $0.677\%$ (far smaller than $1\%$), which means that the model (2) has high precision [22].

### Table 1. Summary of Bayesian estimation and some statistics of 468 parameters.

<table>
<thead>
<tr>
<th>node</th>
<th>Mean</th>
<th>Sd</th>
<th>MCerror</th>
<th>2.50%ql</th>
<th>Median</th>
<th>97.50%qu</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.6066</td>
<td>0.08335</td>
<td>0.000494</td>
<td>0.4399</td>
<td>0.6088</td>
<td>0.7621</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.5179</td>
<td>0.09425</td>
<td>0.000518</td>
<td>0.3310</td>
<td>0.5190</td>
<td>0.6995</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$p_{23,1}$</td>
<td>0.4812</td>
<td>0.09475</td>
<td>0.000556</td>
<td>0.2978</td>
<td>0.4807</td>
<td>0.6674</td>
</tr>
<tr>
<td>node Mean</td>
<td>Sd</td>
<td>MCerror</td>
<td>2.50%ql</td>
<td>Median</td>
<td>97.50%qu</td>
<td></td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.06057</td>
<td>0.04082</td>
<td>0.000230</td>
<td>0.007569</td>
<td>0.05202</td>
<td>0.1612</td>
</tr>
<tr>
<td>$p_{25}$</td>
<td>0.07441</td>
<td>0.04936</td>
<td>0.000286</td>
<td>0.009637</td>
<td>0.06430</td>
<td>0.1960</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$p_{25,2}$</td>
<td>0.07440</td>
<td>0.04953</td>
<td>0.000281</td>
<td>0.009567</td>
<td>0.06392</td>
<td>0.1978</td>
</tr>
</tbody>
</table>

Note: Categories 6-8 are omitted from the table.

![Figure 2. Convergence diagnosis diagram. The upper is the history strace plot graph of posterior samples of $p_{11}$, the lower left is the curve graph of kernel density estimation of $p_{11}$, and the lower right is the autocorrelation graph of $p_{11}$.

3.2.3. Diagnosis of Model

For the MC chain of 468 parameters in model (2), the History-strace-plot plot, the density estimation graph, and the Autocorrelation graph are drawn respectively (in the case of $p_{11}$, see Figure 1, the remaining 467 plots are omitted, the same below). The History-strace-plots of 468 parameters show that after discarding the first 4999 burnin values, the MC chain of 468 parameters converge and each limit distribution
is their own posterior distribution. The autocorrelated graph of each MC chain shows that after the lag period ≥2, the autocorrelation coefficient is close to 0, and each MC chain can be regarded as a MC chain of the independent and identically distribution. Statistical inferences can be made using the corresponding MC chain as a posterior sample (see Figure 1). the category I \( p_{t,1} \) presents symmetrical distribution characteristics (it is also proved by 52 box diagrams of the first category in Figure 2), and the probability of the other nine categories presents a more severe right-biased distribution, with mean estimates being more affected by extreme values. Their robust estimates (median value) and corresponding 95% confidence interval are taken as the risk probabilities of the nine categories for discussion.

### Table 2. Summary of maximum MC error and corresponding information of each MC chain of model (2).

<table>
<thead>
<tr>
<th>( p_{t,1} ) of ( G_i )</th>
<th>Max.MC error</th>
<th>Sd</th>
<th>Time point of occurrence</th>
<th>The frequency of the sample calculation</th>
<th>Prediction of the model (2)</th>
<th>(Max.MC error/sd) 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{12} ) of ( G_2 )</td>
<td>0.000556</td>
<td>0.09475</td>
<td>2020-4-23</td>
<td>0.5000</td>
<td>0.4807</td>
<td>0.5877%</td>
</tr>
<tr>
<td>( p_{13} ) of ( G_3 )</td>
<td>0.000371</td>
<td>0.05730</td>
<td>2020-3-12</td>
<td>0.0000</td>
<td>0.09741</td>
<td>0.647%</td>
</tr>
<tr>
<td>( p_{14} ) of ( G_4 )</td>
<td>0.000408</td>
<td>0.07016</td>
<td>2020-4-20</td>
<td>0.0000</td>
<td>0.5004</td>
<td>0.581%</td>
</tr>
<tr>
<td>( p_{15} ) of ( G_5 )</td>
<td>0.000443</td>
<td>0.07363</td>
<td>2020-4-17</td>
<td>0.73333</td>
<td>0.3549</td>
<td>0.602%</td>
</tr>
<tr>
<td>( p_{16} ) of ( G_6 )</td>
<td>0.000201</td>
<td>0.04888</td>
<td>2020-3-9</td>
<td>0.0000</td>
<td>0.06333</td>
<td>0.616%</td>
</tr>
<tr>
<td>( p_{17} ) of ( G_7 )</td>
<td>0.0000318</td>
<td>0.05884</td>
<td>2020-4-10</td>
<td>0.05263</td>
<td>0.4106</td>
<td>0.540%</td>
</tr>
<tr>
<td>( p_{18} ) of ( G_8 )</td>
<td>0.0000432</td>
<td>0.06811</td>
<td>2020-3-6</td>
<td>0.6875</td>
<td>0.3565</td>
<td>0.6345%</td>
</tr>
<tr>
<td>( p_{19} ) of ( G_9 )</td>
<td>0.0000276</td>
<td>0.04386</td>
<td>2020-3-11</td>
<td>0.0000</td>
<td>0.05557</td>
<td>0.629%</td>
</tr>
<tr>
<td>( p_{20} ) of ( G_9 )</td>
<td>0.0000232</td>
<td>0.03429</td>
<td>2020-3-12</td>
<td>0.0000</td>
<td>0.02538</td>
<td>0.677%</td>
</tr>
</tbody>
</table>

Note: \( p_{t,i} \) denotes the probability of \( G_i \) (\( i = 1,2,\ldots,9 \)).

### 4. Interpretation of Result

The 52-day 2020-3-3 to 4-23 days are considered to be 52 time points, with 0 new cases of overseas input in nine categories with many time points. The frequency (proportion) of overseas input is calculated with sample data, and the input frequency of the corresponding time point is 0 or 1. However, this does not mean that the risk of overseas input in these points is 0 or that the risk of overseas input is 100%. According to probability theory, the estimation error is large at these time points corresponding to these extreme values [23]. The calculation frequency of each of the five time points (2020-3-9, 3-11, 3-12, and 4-20) in Table 2 is 0, which is also explained by the maximum MC error (the second column in Table 2). Using the model (2) calculated by the nine categories of overseas input risk (converted to percentages). The nine categories show the risk change characteristics of four stages. In order to clearly express the change characteristics of each stage, the four categories (G1, G2, G4, G6) with large overseas input risks and the remaining five categories are respectively plotted as point and line graphs. The statistical values of each category are calculated (Table 3). The results show that: (1) In the first stage (2020-3-3 (No. 1) to 3-21 (No. 21)), except for the rapid rise of category G1 and the rapid decline of category G7, other types of input risks decline slowly. (2) In the second stage (2020-3-22 (No. 22) to 4-7 (No. 39)), except for the rapid decline of category G1 and the rapid rise of category G4, other types of input risks rose slowly and gradually declined after seven days. (3) In the third stage (2020-4-8 (No. 40) to 4-18 (No. 49)), the G4 category of overseas input risks declined rapidly, while other types of input risks showed an obvious upward trend. (4) In the fourth stage (after 4-18 (No. 50)), the input risks of each category fluctuate steadily (Figure 3). The overseas input risk probability of nine categories is between 0 and 1, which indicates that the overseas input risk probability obtained by model (2) can truly reflect the spatial distribution of overseas input cases, which is more practical than the frequency to describe.
Figure 4. Risk probability of imported cases from 9 categories of provinces (cities) from March 3 to April 8, 2020 (%).

Table 3. Relevant statistics of overseas input risk in 9 categories of provinces (cities) from March 3 to April 8, 2020 (%).

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>G9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>85.060</td>
<td>10.190</td>
<td>50.040</td>
<td>76.450</td>
<td>7.337</td>
<td>46.200</td>
<td>36.560</td>
<td>5.557</td>
<td>2.658</td>
</tr>
<tr>
<td>Min</td>
<td>14.020</td>
<td>1.817</td>
<td>1.138</td>
<td>2.172</td>
<td>1.142</td>
<td>0.634</td>
<td>0.476</td>
<td>0.473</td>
<td>0.467</td>
</tr>
<tr>
<td>Range</td>
<td>71.040</td>
<td>8.373</td>
<td>48.902</td>
<td>74.278</td>
<td>6.195</td>
<td>45.566</td>
<td>36.084</td>
<td>5.084</td>
<td>2.191</td>
</tr>
<tr>
<td>Sd</td>
<td>19.363</td>
<td>2.384</td>
<td>9.067</td>
<td>20.304</td>
<td>1.718</td>
<td>10.144</td>
<td>7.123</td>
<td>0.990</td>
<td>0.624</td>
</tr>
<tr>
<td>Mean</td>
<td>53.482</td>
<td>5.905</td>
<td>5.471</td>
<td>14.867</td>
<td>3.998</td>
<td>4.662</td>
<td>3.690</td>
<td>1.615</td>
<td>1.523</td>
</tr>
</tbody>
</table>

Note: The original map data is a public map of the Ministry of Natural Resources (GS (2019) 1829)

Figure 5. Risk map of overseas input from March 3 to April 23, 2020.
provinces and cities have input risks. The highest risk zone of provinces and cities represented by Beijing, Shanghai and In the long run, the highest risk area is still the first category of 12 combined with the input risk estimates (Figure 3) shows that: (1) the input frequency is 0, and there is still overseas input risk. (2) In the long run, the highest risk area is still the first category of 12 provinces and cities represented by Beijing, Shanghai and Guangdong, including 10 coastal/border provinces and cities. There are 7 days in the study period, and the risk value of the fourth category (Heilongjiang and Shanxi) is higher than that of the first category. (2) Before March 21, 2020 (the first stage), the first type of overseas input risk is the highest (59.613%). (3) From March 22 to April 18, 2020 (the second and third stages), the top two overseas input risks are both Category I and Category IV. However, the first category gradually decreased (from 60.50% to 37.00%), and the fourth category gradually increased (from 16.1% to 33.8%). (4) After April 19, 2020, the highest input risks are the first category, the third category (Shaanxi, Jilin), and the fourth category. The overseas input risk of other six categories are less than 5%. For the sixth, seventh, and eighth category, their overseas input risks of are stable around 2%. (5) The last two categories of overseas input risks are the eighth category (Henan) and ninth category (Anhui). (6) The fourth category has the largest fluctuation in overseas input risk.

5. Conclusion and Prospect

Using two clustering methods (Average) and similar (Mcquitty), the 25 provinces and cities with overseas inputs were clustered with indicators, resulting in nine distinct categories. The 12 provinces and cities represented by Beijing, Shanghai and Guangdong are the first category, accounting for 57.42% of the total number of imported cases from abroad. The fourth category Heilongjiang and Shanxi have 12 days of new overseas imported cases more than 11 people per day, and the total proportion of overseas imported cases are 27.86%. Shaanxi and Jilin are the third category, Inner Mongolia is the sixth category, and the total proportion of imported cases abroad are 3.07% and 2.95% respectively. there are more than 98.69% of foreign imported cases in 52 days in above-mentioned 4 categories, 17 provinces and cities. Based on the clustering results, a multinomial distribution model based on clustering results achieves the estimation of overseas input risk, which is superior to the traditional frequency characterization method. The relevant research needs to be continued by colleagues.

Funding

This research was funded by “Analysis of the Influence Mechanism of Modern Service Industry in Yunnan Province Base on Bayes Method”on the Project of Yunnan University Joint Fund. (2017FH001-068); Supported by the Scientific Research Foundation of Yunnan Provincial Department of Education (2023J1077).

References


