

Supplementary Information

1. Ripley's K function

Ripley's K function is mainly used to determine whether the elements show the clustering or discretization with statistical significance within a certain range of distances (Dixon 2014). We calculate this function in ArcGIS 10.4 and set the compute confidence envelope as 999 permutations. If the K observed value of a specific distance is larger than the K expected value, the clustering degree of the distribution is higher than that of the random distribution of the distance. If the K observed value is less than K expected value, the dispersion degree of the distribution is higher. If K observation value is larger than the upper limit of confidence interval (HiConf Env), the spatial clustering of the distance has statistical significance. If the K observation value is less than the lower limit of confidence interval (LwConfEnv), the spatial dispersion of the distance is statistically significant. In our results, all the observed values were larger than HiConf Env, so the spatial aggregation of patients with different time lags are statistically significant. We can see that the number of cases which imported cases spread to the local area is the largest, and the degree of aggregation is the highest. When the time lag is longer, although the distribution of cases presents a certain aggregation, its distribution is relatively more uniform. And when the time lag is the longest, the aggregation degree increases to a certain extent.

Table 1 Ripley's K functional indicators with different time lags, which the number of cases is represented by Number, ExpectedK is the expected K value, ObservedK is the observed K value, DiffK is the difference between the expected value and the observed value.

| Time lag | Number | ExpectedK | ObservedK | DiffK | LwConfEnv | HiConfEnv |
|----------|--------|-----------|-----------|----------|-----------|-----------|
| 2 | 420 | 6671.343 | 8865.127 | 2193.784 | 5741.112 | 6213.321 |
| 3 | 222 | 6438.904 | 7629.168 | 1190.265 | 5472.448 | 6094.686 |
| 4 | 230 | 5069.29 | 6012.815 | 943.5246 | 4437.184 | 5015.969 |
| 5 | 212 | 5346.744 | 6047.196 | 700.4523 | 4615.442 | 5199.175 |
| 6-8 | 514 | 6600.059 | 7614.314 | 1014.254 | 5791.398 | 6216.832 |

2. Sensitivity analysis of susceptibility

We set a susceptibility factor when population density is greater than 3 (1000 person / km²) to show that dengue fever easily spread in these areas. But it is not clear how adaptable the factor is, so we conducted a sensitivity analysis of the index and observed the impact on the model when the population density is greater than 2,5,10,20, respectively. The results show that when the definition of susceptibility index changes, the model quality decreases with the increase of population density threshold and is significantly smaller than the original model.

Table 2 Comparison of model results when susceptibility changes

| | Sf(origin) | pop>2 | pop>5 | pop>10 | pop>20 |
|---------------------------|------------|------------|----------|----------|-----------|
| Log-likelihood ratio test | -3093.5141 | -3090.0865 | -3091.96 | -3092.15 | -3093.131 |
| AIC | 6214.852 | 6216.173 | 6219.924 | 6220.303 | 6224.262 |
| pseudo-R2 (McFadden's) | 0.0717 | 0.0715 | 0.0709 | 0.0709 | 0.0706 |

3. Collinearity analysis among the variables

The model explores the effects of socio-economic factors on the diffusion of dengue fever, it includes population density from Census and POI data, and the concentration of POIs might act as an indicator of daytime human density. In order to explore the validity of POI categories and population density, we made a collinearity analysis of model variables (Table 2). Firstly, we do a simple linear regression for all variables, then test their collinearity, and select variance expansion factor (VIF) to evaluate. When the VIF is greater than 5 to 10, it shows that the variables have a weak dependence, and when the index is greater than 30, the dependence is from medium to strong. (Mason and Perreault 1991). Obviously, the VIF values of all variables are less than 5, which indicates that there is no serious collinearity among the variables we choose.

Table 3. The collinearity test between variables

| Variable | VIF |
|--------------------------------------|------|
| Convenience services | 4.63 |
| Residential and commercial buildings | 4.61 |
| Administrative agencies | 4.36 |
| Factories and companies | 3.34 |
| Healthcare facilities | 3.02 |
| Pop | 1.76 |
| Distance(km) | 1.46 |
| Transportation logistics facilities | 1.35 |
| Flow | 1.28 |
| Sf | 1.2 |
| Mean VIF | 2.7 |

Reference

- Dixon, Philip M. 2014. "Ripley's K Function." In *Wiley StatsRef: Statistics Reference Online*. Encyclopedia of Environmetrics.
- Mason, Charlotte H., and William D. Perreault. 1991. "Collinearity, Power, and Interpretation of Multiple Regression Analysis." *Journal of Marketing Research* 28 (3):268-280. doi: 10.2307/3172863.