Spatiotemporal analysis of hand, foot and mouth disease data using time-lag geographically-weighted regression

Zhi-Min Hong,[1,2] Hu-Hu Wang,[1,3] Yan-Juan Wang,[1] Wen-Rui Wang[3]

[1]School of Sciences, Inner Mongolia University of Technology, Hohhot; [2]Inner Mongolia Key Laboratory of Statistical Analysis Theory for Life Data and Neural Network Modeling, Inner Mongolia, Hohhot; [3]Institute for infectious disease and endemic disease control, Inner Mongolia Autonomous Region Center for Disease Control and Prevention, Hohhot, China

Abstract

Hand, Foot, and Mouth Disease (HFMD) is a common and widespread infectious disease. Previous studies have presented evidence that climate factors, including the monthly averages of temperature, relative humidity, air pressure, wind speed and Cumulative Risk (CR) all have a strong influence on the transmission of HFMD. In this paper, the monthly time-lag geographically-weighted regression model was constructed to investigate the spatiotemporal variations of effect of climate factors on HFMD occurrence in Inner Mongolia Autonomous Region, China. From the spatial and temporal perspectives, the spatial and temporal variations of effect of climate factors on HFMD incidence are described respectively. The results indicate that the effect of climate factors on HFMD incidence shows very different spatial patterns and time trends. The findings may provide not only an in-depth understanding of spatiotemporal variation patterns of the

effect of climate factors on HFMD occurrence, but also provide helpful evidence for making measures of HFMD prevention and control and implementing appropriate public health interventions at the county level in different seasons.

Introduction

Hand, Foot and Mouth Disease (HFMD) is a common, acute infectious disease resulting in millions of cases and even death among children under five years old. It is caused by different species of enteroviruses, most commonly Coxsackievirus A16 and Enterovirus 71 (Yang *et al*., 2011; Zeng *et al*., 2015) that may change with the variation of climate. In addition, human activity and contact frequency may greatly determine the eventual disease occurrence (Cheng *et al*., 2018). China is one of the most affected countries where HFMD has been made statutorily notifiable in May 2008 (Xing *et al*., 2014). The intensity and frequency of the HFMD outbreaks are also different from place to place due to the extensive area where the disease occurs in China, as there may be very different climatic and environmental conditions.

Various research approaches have been used to investigate the spatial and temporal patterns of HFMD outbreaks and the risk factors, such as climate factors and socio-economic environment. In the state of Sarawak in Malaysia, spatial and temporal distributions of HFMD were studied based on Geographical Information Systems (GIS), (Sham *et al*., 2014). In China, time series analysis methods were adopted to achieve a short-term epidemic prediction for HFMD occurrence (Lai *et al*., 2016; Wang *et al*., 2018), and space–time scanning was applied for the detection of spatiotemporal clusters and the identification of high-risk locations (Hong *et al*., 2017; Qian *et al*., 2016) revealing that HFMD occurrence shows autocorrelation characteristics. Using linear regression model, Ma *et al*. (2010) demonstrated that climate factors play important role in predicting HFMD activity, which could assist in explaining the winter peak detected in recent years. The geographical detector method was implemented to test the association between HFMD with a single climate factor and their interaction (Wang *et al*., 2013), while a distributed lag non-linear model was used to quantify the effect of the Average Temperature (AT) and relative Humidity (AH) on HFMD incidence in Huainan and Hefei, China, respectively (Xu *et al*., 2016; Yang *et al*., 2017). The boosted regression tree model, combining the advances of the traditional regression models and the machine-learning methods, were employed by Zhang *et al*., 2016 to determine the optimal lag for the climate factors at which the variance of HFMD occurrence was most explained, and to evaluate the effects of these climate factors on HFMD occurrence at the optimal lag. A Bayesian spa-

tiotemporally varying coefficients model was used to detect the effect of climate factors on HFMD incidence in Sichuan, China (Song *et al*., 2019). The results showed that the epidemic patterns of HFMD are closely related to climate factors. Meanwhile, the impact of climate factors on HFMD is significantly different across space and time. Therefore, the study on spatiotemporal variation patterns of the effect of climate factors on HFMD activity is of great importance to understand the transmission mechanism of HFMD and take proper protection measures against HFMD at different regions and times.

The Geographically Weighted Regression (GWR) model, originally proposed by Brunsdon *et al*. (1996), is a popular and powerful tool in exploring spatial nonstationarity of a regression relationship. Subsequently, GWR models with their variants have been applied to various fields for geo-referenced data analysis. In addition of the recent applications in, for instance, hedonic house price modelling (Lu *et al*., 2014), geology (Duan and Li, 2016), ecology (Song *et al*., 2016) and public health (Wabiri *et al*., 2016), GWR models with their various variants have also been employed to detect the association between HFMD incidence and climate factors, social and economic factors, air pollutants, *etc*. (Hu *et al*., 2012; Dong *et al*., 2016; Wang *et al*., 2017; Hong *et al*., 2018; Yu *et al*., 2019). In the GWR model, temporal effects are assumed to be constant over space because the weighting matrix is calculated only based on spatial information. In fact, however, there is possibility that the effect of explanatory variables on the response variable is not only spatially non-stationary, but also temporally so, meaning that the underlying process is spatiotemporally non-stationary. Time is also an essential dimension relating to various physical processes that often provides valuable information. To fill this gap, Huang *et al*. (2010) proposed a so-called Geographically and Temporally Weighted Regression model (GTWR) by incorporating the time factor into the GWR model. Fotheringham *et al*. (2015) also extended GWR to a new version of GTWR (here called TL-GWR) by separating the selection of the optimal spatial bandwidth from that of the temporal bandwidth. Only previous time (lag time) neighbours are taken into account and their impact on the current state are estimated in the TL-GWR model. This model is useful because a set of spatial bandwidths are selected adaptively at different time points. In contrast to GWR models, the TL-GWR model considers the interaction of space and time and make it possible to simultaneously explore the spatial and temporal non-stationarity of a regression relationship based on the constructed spatiotemporal weight matrix. As mentioned above, the transmission and dispersal of HFMD are closely associated with climate factors and the effect of these parameters on HFMD incidence may change over space and time. To the best our knowledge, the TL-GWR model has not yet been applied to the analysis of HFMD data. Therefore, this model is of great potential in analysing the spatiotemporal variation patterns of the effect of climate factors on HFMD incidence by providing useful evidence for finding disease epidemic patterns that could indicate suitable control measures.

Materials and Methods

HFMD data based on the 102 counties from January to December in 2016 were provided by the Inner Mongolia Autonomous Region's Center for Disease Control and Prevention (CDC). Of the total 13,928 reported HFMD cases, we retained 13,416 cases of the children under 9 years old in 2016 for analysis.

These data from each county were provided by the Inner Mongolia Autonomous Region Statistics Bureau. The climate data, consisting of the monthly Average Temperature (AT), relative humidity (AH), Air Pressure (AP), wind speed (AW) and monthly cumulative precipitation (CR) for each county, came from the Ecological and Agricultural Meteorology Center of Inner Mongolia Autonomous Region. In addition, the geographic information of each county, including name, code, and longitude and latitude, came from the National Catalogue Service for Geographic Information.

Preliminary hand, foot and mouth disease data processing

Consideration that the number of reported HFMD cases in each county is closely associated with the size of the exposed population in that county, it was felt suitable to use the monthly Cumulative Incidence (CI) in each county as the response variable to model the relationship to the climate factors. The monthly CI is defined by the ratio of the number of monthly reported HFMD cases $O(i)$ under 9 years old to the size of the total population $P(i)$ of this age group at risk in that county. However, the crude CI may suffer from the problem of zero CI values, mainly caused by the geographic coded error or missing reports. Here, we adopted a hierarchical Bayesian model with both structured and unstructured random effects (Haining 2003; Huang *et al*., 2014) to adjust the crude CI of HFMD. Subsequently, the adjusted monthly CI was used as the response variable for analysis. The adjusting process is described as follows illustrating the spatial distribution and temporal trend of the values of the adjusted CI:

Let $r(i)$ be the risk probability that a child in the ith county infects HFMD. The hierarchical Bayesian model is then formulated as:

$$O(i) \sim B(r(i), P(i)), \text{logit}(r(i)) = \mu + \upsilon(i) + e(i) \qquad \text{(Eq. 1)}$$

where μ represents the mean effect of disease risk; $e(i)$ independently and identically distributed $N(0,\sigma_e^2)$ and captures the spatial unstructured nonstationarity; and $\upsilon(i)$ a spatial structured random variable describing the spatial dependence of the relative risk and defined by the intrinsic Gaussian spatial autoregression process with the conditional distribution given $\upsilon(j)$ being:

$$\upsilon(i)|\upsilon(j) \sim N\left(\sum_{j=1}^{n} w_{ij}^* \upsilon(j), \frac{\sigma_\upsilon^2}{\sum_{j=1}^{n} w_{ij}}\right), j \neq i \qquad \text{(Eq. 2)}$$

where j denotes the neighbour of site i with the spatial contiguity weights w_{ij} for a given i defined as $w_{ij} = 1$ if there is a common edge between j and i; 0 if otherwise; and $w_{ii} = 0$, $w_{ij}^* = w_{ij}/\sum_{j=1}^{n} w_{ij}$. The gamma distributions are taken as prior distributions of $1/\sigma_e^2$ and $1/\sigma_\upsilon^2$, namely, $1/\sigma_e^2 \sim \gamma(0.001, 0.001)$ and $1/\sigma_e^2 \sim \gamma(0.5, 0.0005)$ with a non-information prior assumed for μ. By means of the WinBUGS 1.4 package, the Markov chain Monte Carlo (MCMC) simulation was used to solve the hierarchical Bayesian model, where the length of burn-in sequence was set at 6,500. The outputs of the hierarchical Bayesian model were the adjusted CI values for the 102 counties at a given month. Figure 1 shows the spatial distributions of the adjusted CI (1/100000) values of the counties averaged by the cold season (from May to September) and the warm season (from November to April), respectively, and the temporal trend of those values averaged over the 102 spatial units. It

was observed that, on average, the adjusted CI values exhibited both significant spatial non-stationarity and notable temporal non-stationarity.

TL-GWR model and bandwidth optimization

Fotheringham *et al.* (2015) proposed a geographically and temporally weighted regression, called the TL-GWR model, in which the spatial bandwidths are adaptively selected at different time periods. That is, the TL-GWR model is of the form:

$$y_i = \beta_0(u_i,v_i,t) + \sum_{j=1}^p \beta_j(u_i,v_i,t)x_{ij} + \varepsilon_i, \quad i=1,2,\cdots,n_t \quad \text{(Eq. 3)}$$

where $\{y_i; x_{i1}, x_{i2}, \cdots, x_{ip}\}_{i=1}^{n_t}$ are the observations of the response variable Y and the explanatory variables $X_1, X_2, \cdots, X_p$ at a set of spatial locations $[(u_1,v_1), (u_2,v_2), \cdots, (u_{n_t}, v_{n_t})]$, where n_t is the number of spatial locations of data observations at time period t; and q the number of time lags in addition to those from the same time period. $\{\beta_j(u_i,v_i,t)\}_{j=1}^p$ represent the unknown coefficients of local coordinates and the time for describing spatiotemporally varying patterns of each explanatory variable. $\{\varepsilon_i\}_{i=1}^{n_t}$ denote errors assumed to be independent and identically distributed random variables with zero mean and the same variance $\sigma^2 > 0$.

The calibration method of GWR model proposed by Brunsdon *et al.* (1996) can still be used to estimate the coefficients $\{\beta_j(u_i,v_i,t)\}_{j=1}^p$, but in this case, in addition to those from the time period t, while the data collected at prior time periods are named t-1, t-2, $\cdots$, t-q, with q being the number of time lags is incorporated in the GWR model, in which the weight matrix W is constructed in a different way in order to capture simultaneously both spatial and temporal non-stationarities.

Using the locally weighted Ordinary Least Squares (OLS) procedure, the estimator of the coefficient vector $\beta(u_i,v_i,t) = (\beta_0(u_i,v_i,t), \beta_1(u_i,v_i,t), \cdots, \beta_q(u_i,v_i,t))^T$ at (u_i,v_i,t) can be computed by,

$$\hat{\beta}(u_i,v_i,t) = [X^T W_i(b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}, b_T)X]^{-1} X^T W_i(b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}, b_T)Y \quad \text{(Eq. 4)}$$

where X is the design matrix of the explanatory variables and Y is the observation vector of the response variable

$$W_i(b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}, b_T) = \text{Diag}\left(w_i^t(b_{s_t}, b_T), w_i^{t-1}(b_{s_{(t-1)}}, b_T), \cdots, w_i^{t-q}(b_{s_{(t-q)}}, b_T)\right) \quad \text{(Eq. 5)}$$

is the weight matrix with the segregated spatial bandwidths over time, where

$$w_i^{t_0}(b_{s_{t_0}}, b_T) = \text{Diag}\left(w_{i1_{S,T}}^{t_0}, w_{i2_{S,T}}^{t_0}, \cdots, w_{in_{t_0 S,T}}^{t_0}\right), t_0 = t, t-1, \cdots, t-q \quad \text{(Eq. 6)}$$

The elements in $w_i^{t_0}(b_{s_{t_0}}, b_T)$ can be generated by the Gaussian spatiotemporal kernel function, that is
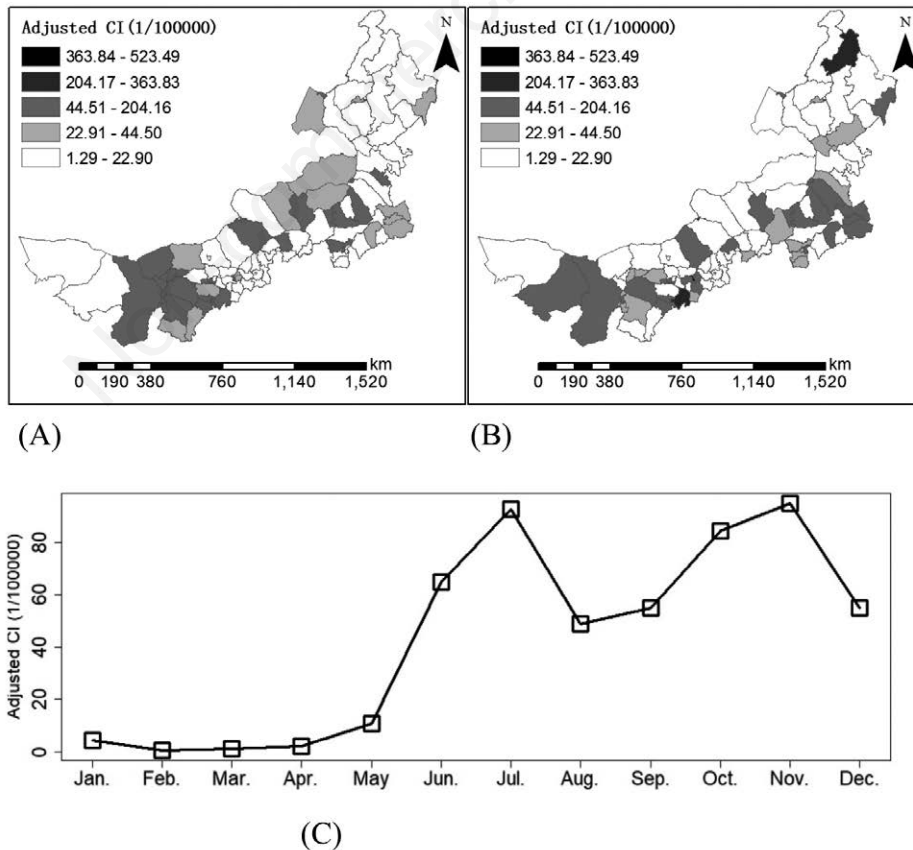

Figure 1. Spatial distribution of the adjusted CI values averaged by time (A) in the cold season; (B) in the warm season; (C) temporal trend of those values averaged over the counties in Inner Mongolia in 2016.

$$w_{ik_{S,T}}^{t_0} = \exp\left(-\frac{1}{2}\frac{d_{s_{ik}}^2}{b_{s_{t_0}}^2}\right) * \exp\left(-\frac{1}{2}\frac{d_{t_{ik}}^2}{b_T^2}\right), k = 1,2,\cdots,n_{t_0} \qquad \text{(Eq. 7)}$$

where $d_{s_{ik}} = \sqrt{(u_i - u_k)^2 + (v_i - v_k)^2}$ is the Euclidean distance between (u_i,v_i) and (u_k,v_k); and $d_{t_{ik}} = |t_i - t_k|$ the distance in time between the regression point i and the data point k. The data point k can be located at any set $S_t, S_{t-1}, \cdots, S_{t-q}$.

A set of separated spatial bandwidths $b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}$ and the temporal bandwidth b_T can be determined by minimizing the Cross Validation (CV) function. That is, select $b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}$ and b_T that minimizes

$$CV\left(b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}, b_T\right) = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_{-i}(b_{s_t}, b_{s_{(t-1)}}, \cdots b_{s_{(t-q)}}, b_T))^2/n} \qquad \text{(Eq. 8)}$$

Taking the time period t as an example, the procedure of selecting the optimal spatial/temporal bandwidth according to Fotheringham *et al*. (2015) is described in detail as follows. Given a temporal bandwidth, for example, $b_T = 1$, the spatial bandwidth b_{s_t} is first i) calculated by using data points only from time t.

In this case, $d_{t_{ik}} = 0$, $w_{ik_{S,T}}^t = \exp\left(-\frac{1}{2}\frac{d_{s_{ik}}^2}{b_{s_t}^2}\right)$, the weight matrix

$$W_i(b_{s_t}, b_T) = w_i^t(b_{s_t}, b_T) = \text{Diag}\left(w_{i1_{S,T}}^t, w_{i2_{S,T}}^t, \cdots, w_{in_{t_{S,T}}}^t\right)$$

and $\hat{y}_{-i}(b_{s_t}, b_T)$ can be obtained using the calibration method of the GWR model by setting $w_{ii_{S,T}}^t = 0, (i=1,2,\cdots,n_t)$. The optimal spatial bandwidth $b_{s_t}^*$ is confirmed by minimizing the CV score,

$$CV(b_{s_t}, b_T) = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_{-i}(b_{s_t}, b_T))^2/n}.$$

Furthermore, the first diagonal elements of the weight matrix in Eq. 5 have been kept fixed by setting $b_{s_t}^*$ in Eq. 6 as $t_0 = t$.

Then ii) the data points at time period $t - 1$ are incorporated into the model. In this case, the observations of the response variable and explanatory variables are from t and $t - 1$ time periods. As for time period $t - 1$, $d_{t_{ik}} = 1$ and

$$w_{ik_{S,T}}^{t-1} = \exp\left(-\frac{1}{2}\frac{d_{s_{ik}}^2}{b_{s_{(t-1)}}^2}\right)\exp\left(-\frac{1}{2}\frac{1}{1}\right).$$

The weight matrix is $W_i\left(b_{s_t}^*, b_{s_{(t-1)}}, b_T\right) = \text{Diag}\left(w_i^t(b_{s_t}^*, b_T), w_i^{t-1}\left(b_{s_{(t-1)}}, b_T\right)\right).$

The optimal spatial bandwidth $b_{s_{(t-1)}}^*$ is obtained by minimizing

$$CV\left(b_{s_t}^*, b_{s_{(t-1)}}, b_T\right) = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_{-i}(b_{s_t}^*, b_{s_{(t-1)}}, b_T))^2/n}.$$

Then, the second set of $n_{\{t-1\}}$ diagonal elements can be derived and kept fixed. Then iii) the above described process is repeated by incorporating one-by-one data points from the past time $t-2$, $t-3, \cdots, t-q$ to derive in turn optimal spatial bandwidths, denoted by $b_{S\{(t-2)\}}^*$ $b_{s_{(t-3)}}^*$, $b_{s_{(t-q)}}^*$ for these time periods, and finally iv) the process is repeated described from i) to iii) for the other q–1 possible temporal bandwidths according to the number of time lags in the model, that is for b_T equal to $2, 3, \cdots$, or q temporal unit in the past.

Thus, a set of CV scores are obtained, say, $CV_{b_T=1}, CV_{b_T=2}$, $CV_{b_T=q}$. Now the optimal temporal bandwidth can be selected as follows

$$b_T^* = \arg\min_{b_T > 0}\left(CV_{b_T=1}, CV_{b_T=2}, \cdots, CV_{b_T=q}\right)$$

Statistical outcome

Model specification

We used $CI \times 10^5$, as a common practice in HFMD analysis, namely the adjusted CI (1/100000), taking its logarithm as the response variable in the TL-GWR. The model is described as:

$$\log(CI) = \beta_0(u,v,t) + \beta_1(u,v,t)AT + \beta_2(u,v,t)AH + \beta_3(u,v,t)AP + \beta_4(u,v,t)CR + \beta_5(u,v,t)AW + \varepsilon \quad \text{(Eq. 9)}$$

where time $t=1,2,\cdots,12$, namely from January to December in 2016, (u,v) is the spatial coordinate denoting the spatial location of each county (the total 102 counties in Inner Mongolia) and AT, AH, AP, CR and AW standing for the monthly average temperature, relative humidity, air pressure, cumulative precipitation and wind speed, respectively.

Selection of spatial and temporal bandwidths

In this paper, because of the HFMD data availability from January to December in 2016, local coefficients for regression points located at months April, March, February, and January were estimated using only 3, 2, 1, and 0 time lag periods respectively. The number of time lags was taken as 4 months because data points located more than 4 months away from the regression month in which the regression point is located would have little or negligible influence on the estimation of local coefficients at the regression point.

A set of optimal spatial bandwidths and the optimal temporal bandwidth for each regression month are listed in Table 1. A temporal bandwidth of value 0 would mean that only a spatial bandwidth is used, for example, January. As shown in Table 1, in general, a decreasing trend of the optimal spatial bandwidths is exhibited when the data points incorporated into model were located farther away in time from the regression month. However, the decreasing trend of the optimal spatial bandwidths was interrupted for regression months March, April and November.

Tests for non-stationarity of parameter estimation

For each regression month (from January to December), a specific TL-GWR model was constructed by incorporating data from lag time period into GWR model. The spatial non-stationarity of parameter estimates was assessed by comparing twice the standard errors (2sd) of the global ordinary OLS estimates with the interquartile (IQ) of parameter estimates for the TL-GWR model. The larger values of IQ mean that there was significantly spatial nonstationarity at the corresponding time period (Fotheringham *et al*., 2002). The testing results are summarized in Table 2, in which the statistically significant factors are marked with "*". As shown in Table 2, the statistical significance of effects of climate factors on HFMD incidence mainly concern occurrences in the cold season.

Evaluation of the TL-GWR model

It is worth comparing the performance of the TL-GWR model, in which data points from previous time periods were incorporated

Evaluation of the TL-GWR model

It is worth comparing the performance of the TL-GWR model, in which data points from previous time periods were incorporated into the GWR model, against the GWR model, in which the local parameters were estimated at each time point only using data points from this time period. The CV scores for each month of both TL-GWR and GWR are described in Figure 2, where it can be seen that there is a similar trend of the CV score of the both model (GWR and TL-GWR). However, the CV score of the TL-GWR model was lower than that of the GWR model. The result means that the performance of the TL-GWR is superior to that of the GWR for all months of the study except for January where it had the same CV score.

Table 1. Optimal temporal and spatial bandwidths January-December in 2016.

	Dec.	Nov.	Oct.	Sep.	Aug.	Jul.	Jun.	May	Apr.	Mar.	Feb.	Jan.
b^*_T	3	2	3	4	2	3	2	4	3	2	1	-
b^*_{S12}	311	-	-	-	-	-	-	-	-	-	-	-
b^*_{S11}	35	359	-	-	-	-	-	-	-	-	-	-
b^*_{S10}	17	33	850	-	-	-	-	-	-	-	-	-
b^*_{S9}	0.01	38	44	559	-	-	-	-	-	-	-	-
b^*_{S8}	-	50	20	17	803	-	-	-	-	-	-	-
b^*_{S7}	-	-	22	22	27	886	-	-	-	-	-	-
b^*_{S6}	-	-	20	18	25	47	635	-	-	-	-	-
b^*_{S5}	-	-	-	-	26	20	22	645	-	-	-	-
b^*_{S4}	-	-	-	-	6	0.001	1.17	9	890	-	-	-
b^*_{S3}	-	-	-	-	-	-	0.00	0.00	54	303	-	-
b^*_{S2}	-	-	-	-	-	-	-	-	7	13.8	665	-
b^*_{S1}	-	-	-	-	-	-	-	-	25	24	19	432

*The unit of bandwidth b^*_{si} is kilometers (km). The unit of bandwidth b^*_T is month.

Table 2. Non-stationarity test of parameters for the TL-GWR model January-December in 2016.

	Jan.		Feb.		Mar.		Apr.	
Variable	IQ	2sd	IQ	2sd	IQ	2sd	IQ	2sd
Inter	0.442*	0.258	0.058	0.124	0.395*	0.242	0.038	0.237
AT	0.65*	0.615	0.259	0.336	1.062*	0.567	0.124	0.534
AH	0.283	0.629	0.125	0.339	0.539	0.603	0.083	0.534
AP	0.29	0.355	0.044	0.168	0.399*	0.255	0.052	0.272
CR	0.29*	0.279	0.107	0.158	0.596*	0.296	0.063	0.252
AW	0.121	0.262	0.048	0.126	0.091	0.260	0.051	0.290
	May.		Jun.		Jul.		Aug.	
Variable	IQ	2sd	IQ	2sd	IQ	2sd	IQ	2sd
Inter	0.049	0.284	0.079	0.310	0.041	0.272	0.046	0.232
AT	0.283	0.502	0.137	0.511	0.089	0.510	0.118	0.275
AH	0.446	0.551	0.245	0.588	0.058	0.531	0.089	0.351
AP	0.120	0.372	0.115	0.427	0.045	0.428	0.059	0.266
CR	0.099	0.386	0.424*	0.415	0.170	0.359	0.051	0.314
AW	0.154	0.348	0.162	0.358	0.052	0.319	0.086	0.274
	Sept.		Oct.		Nov.		Dec.	
Variable	IQ	2sd	IQ	2sd	IQ	2sd	IQ	2sd
Inter	0.081	0.215	0.043	0.235	0.344*	0.271	0.411*	0.282
AT	0.219	0.373	0.111	0.294	0.815*	0.568	1.938*	0.620
AH	0.285	0.468	0.06	0.329	0.228	0.570	0.218	0.589
AP	0.161	0.415	0.062	0.279	0.530*	0.332	0.710*	0.352
CR	0.170	0.315	0.051	0.335	0.365*	0.354	0.620*	0.305
AW	0.071	0.250	0.077	0.262	0.341*	0.290	0.289*	0.285

Results and discussion

Spatial patterns of the estimated coefficients

Figures 3 to 8 shows the spatial variation patterns of the effect of each explanatory variable including the intercept term on the response variable for the two seasons (cold and warm seasons).

Figure 3 reflects the basic spatial variation patterns of HFMD incidence in different seasons. It can be shown that, HFMD incidence varies obviously in the two seasons with the largest values takes place over space in the warm season.

Figure 4 shows the choropleth maps of the estimated coefficient of the AT in cold and warm seasons. From Figure 4A, it can be observed that, in the cold season, AT has a negative influence on HFMD incidence in a number of counties (almost all) of the north-eastern and eastern parts of the region with the strongest negative influence in Genhe City, Hailaer District, Ewenki Autonomous Banner, Arhorqin Banner and Wengniute Banner. This influence

turns to be positive in the south-western area, which is relatively consistent with the region of the higher HFMD incidence in Figure 1A. However, in the warm season, AT influences HFMD incidence in all counties positively (Figure 4B), which is consistent with the results of existing literatures (Zhang *et al.*, 2016; Xu *et al.*, 2016; Song *et al.*, 2019).

The seasonal maps of the coefficient estimator of AH are depicted in Figure 5. The negative influence of AH on HFMD incidence in almost all counties, except for the middle part of the region and some other counties, appears in the cold season (Figure 5A). In the warm season, the influence turns to be positive in almost all counties (in Figure 5B), which is relatively consistent with the study results of Onozuka and Hashizume (2011) and Yang *et al.* (2017; 2018).

Figure 6 reflects the spatial variation characteristics of the estimated coefficient of monthly average AP in different seasons. As can be observed in Figure 6, from cold season to warm season, except for the north-eastern area, the significant variation where the positive influence turns into negative, appears in the other counties. In addition, except for the slightly positive effect in a few of counties, Figure 6B shows that the AP has basically a negative influence on HFMD incidence in the warm season. In these two seasons, the negative influence of the AP on HFMD incidence is upheld in the north-eastern part of whole region, which might mean that the AP is a protective factor for the epidemic of HFMD in these counties.

From the choropleth maps of the coefficient estimator of CR in Figure 7A, it can be seen that, in the cold season, except for the slightly positive influence in the north-eastern part of the area, the negative effect of the CR on HFMD incidence takes place in most of counties of Inner Mongolia. The results shown in Figure 7B indicate that the CR influences HFMD incidence positively in the whole region with the intensity increasing from east to west, with the strongest positive impact of the CR on HFMD incidence in the south-western part of the region.
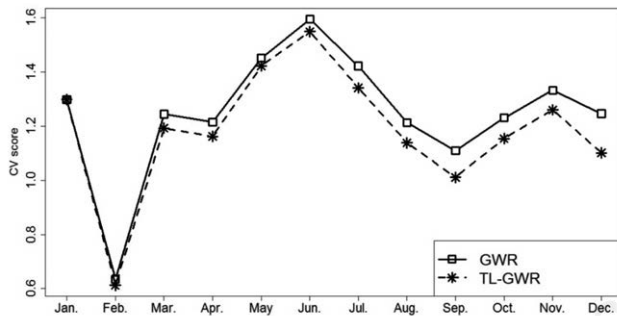

Figure 2. Cross Validation (CV) score from GWR and TL-GWR in 2016 (January to December).
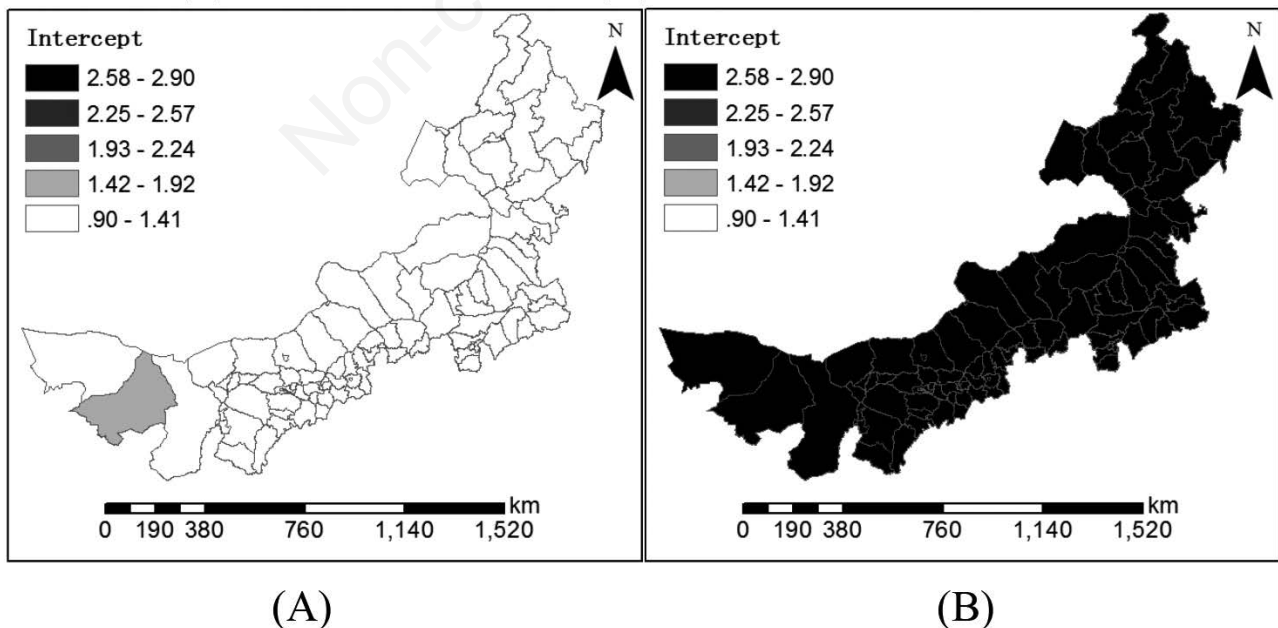

(A)   (B)

Figure 3. Spatial variation of the estimated intercept averaged by the seasons: (A) in the cold season; (B) in the warm season.

Figure 8 depicts the choropleth maps of the estimated coefficient of AW. As seen in Figure 8A, in the cold season, except for the strongest negative effect in the middle counties, the impact in the other counties is positive with the strongest positive impact of the AW on the CI appearing in the north-eastern area and the western-most counties. However, because there are relatively sparse sample points in the extreme West, the explanation should be taken as cautious. From Figure 8B, in the warm season, the areas influence by the negative effect of the AW on HFMD incidence gradually extend from the middle sideways. The slight positive influence mainly appears in the north-eastern area, while the strongest positive influence is observed in the westernmost counties.
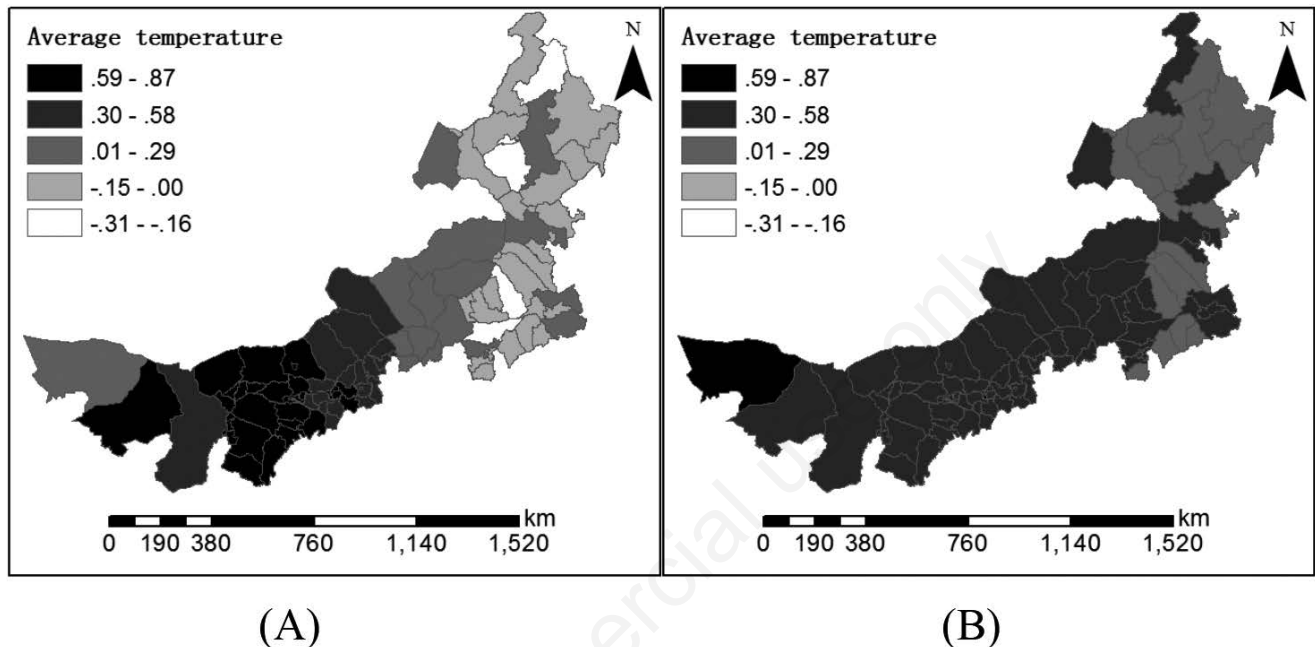

(A)　　　　　　　　　　　　　　　　(B)

Figure 4. Spatial variation of the estimated coefficient of the AT averaged by the seasons: (A) in the cold season; (B) in the warm season.
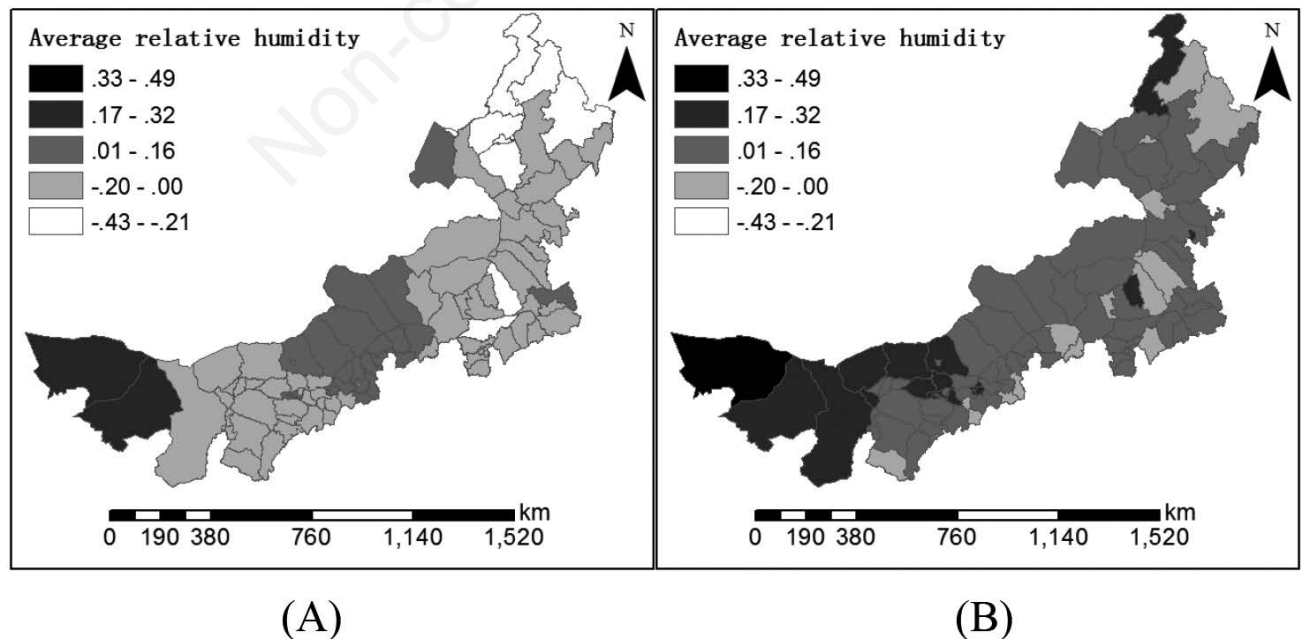

(A)　　　　　　　　　　　　　　　　(B)

Figure 5. Spatial variation of the estimated coefficient of the AH averaged by the seasons: (A) in the cold season; (B) in the warm season.

Temporal patterns of the estimated coefficients

Figure 9A depicts the temporal trend of the logarithm HFMD incidence adjusted by climate factors and is relatively consistent with the trend of HFMD incidence shown in Figure 1C. From the temporal trend chart of the AT in Figure 9B, it can be seen that it positively influences the CI from January to December. The largest

intensity appears in June and December, approximately corresponding to the highest incidence in Figure 1C. Combined with the findings in Figure 4, we may draw the conclusion that AT is a risk factor of HFMD transmission and dispersal during the warm season. In addition, in the cold season, the AT may be a risk factor for the epidemic of HFMD in the south-western part. Figure 9C illus-
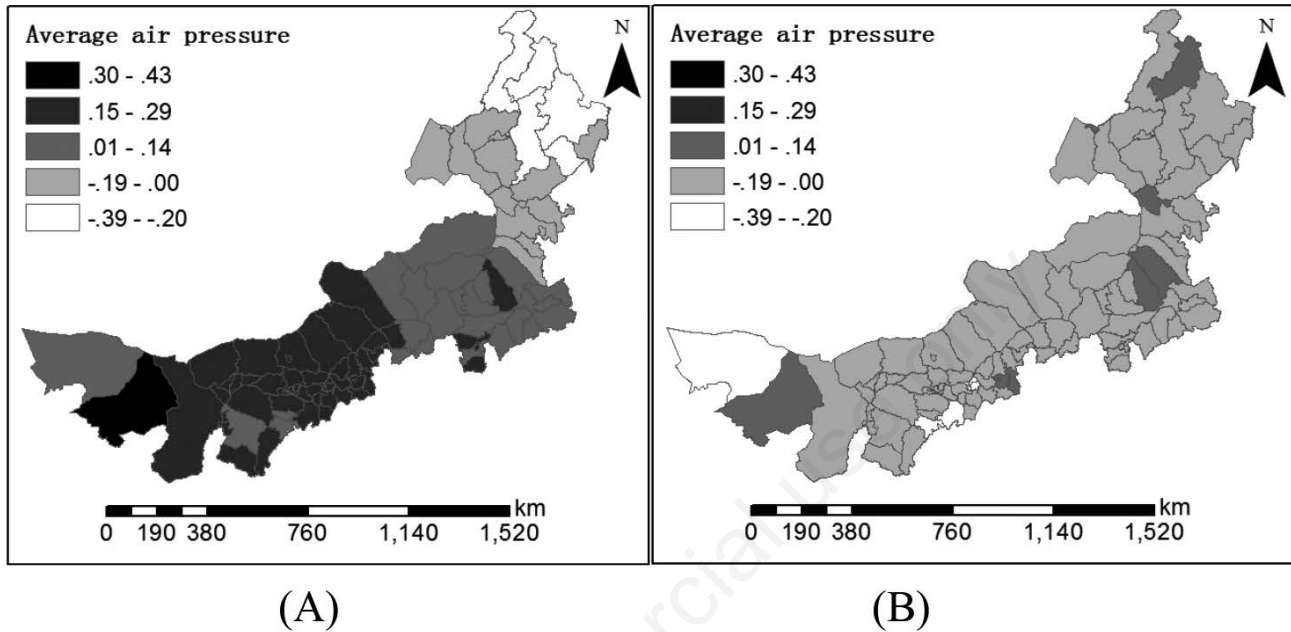

(A)

(B)

Figure 6. Spatial variation of the estimated coefficient of the AP averaged by the seasons: (A) in the cold season; (B) in the warm season.
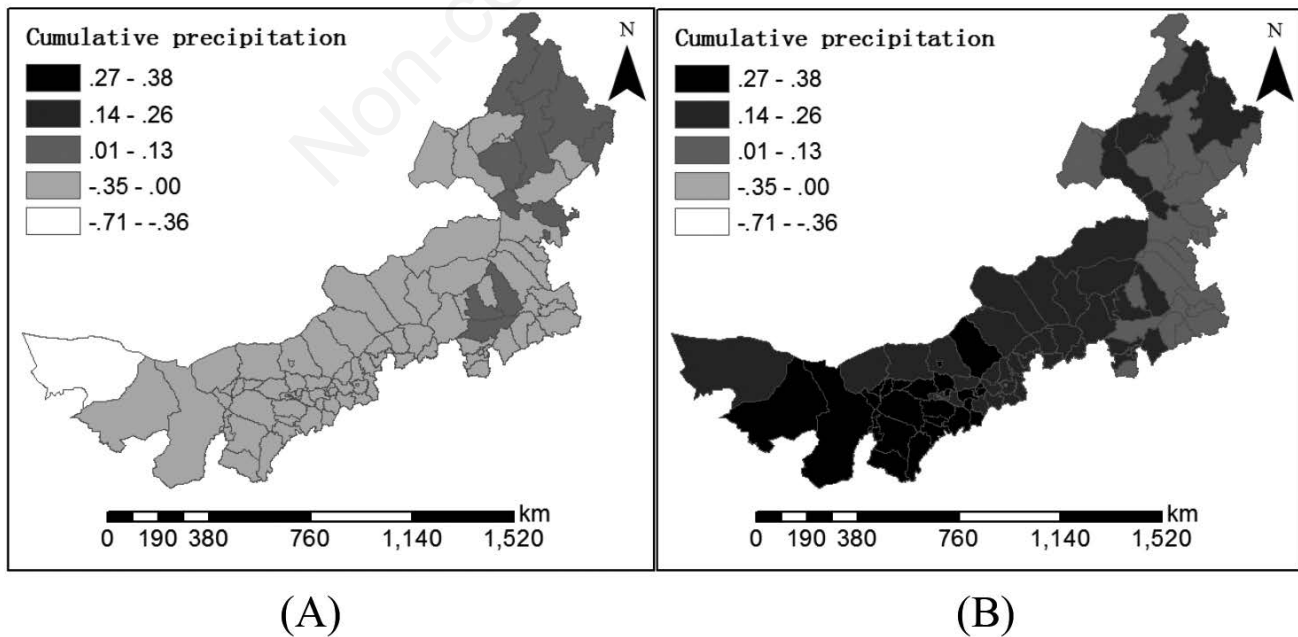

(A)

(B)

Figure 7. Spatial variation of the estimated coefficient of the CR averaged by the seasons: (A) in the cold season; (B) in the warm season.
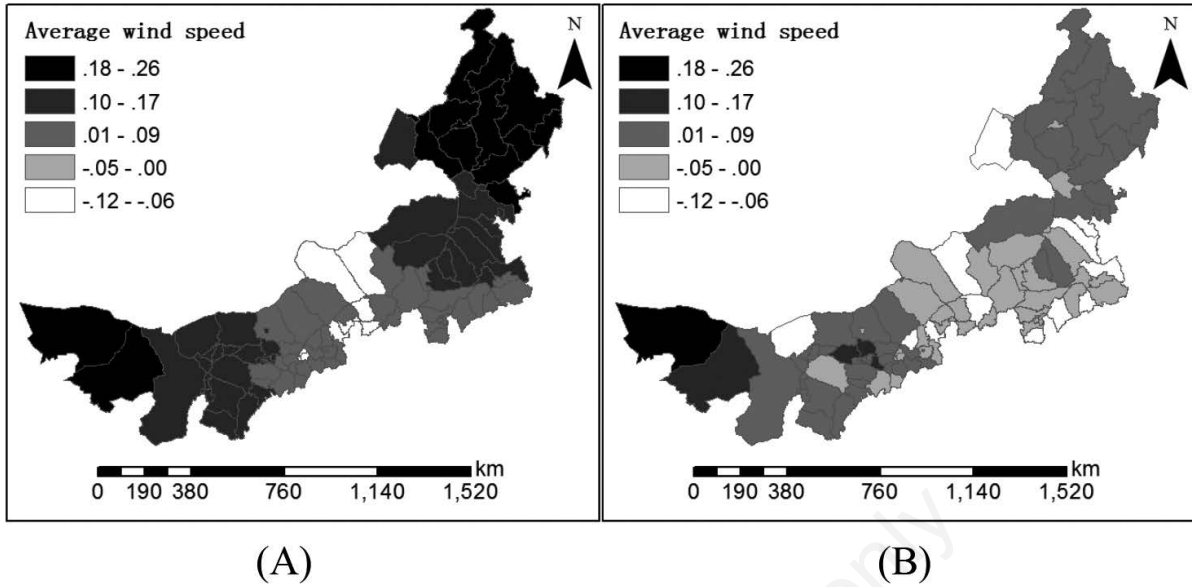
(A)

(B)

Figure 8. Spatial variation of the estimated coefficient of the AW averaged by the seasons: (A) in the cold season; (B) in the warm season.
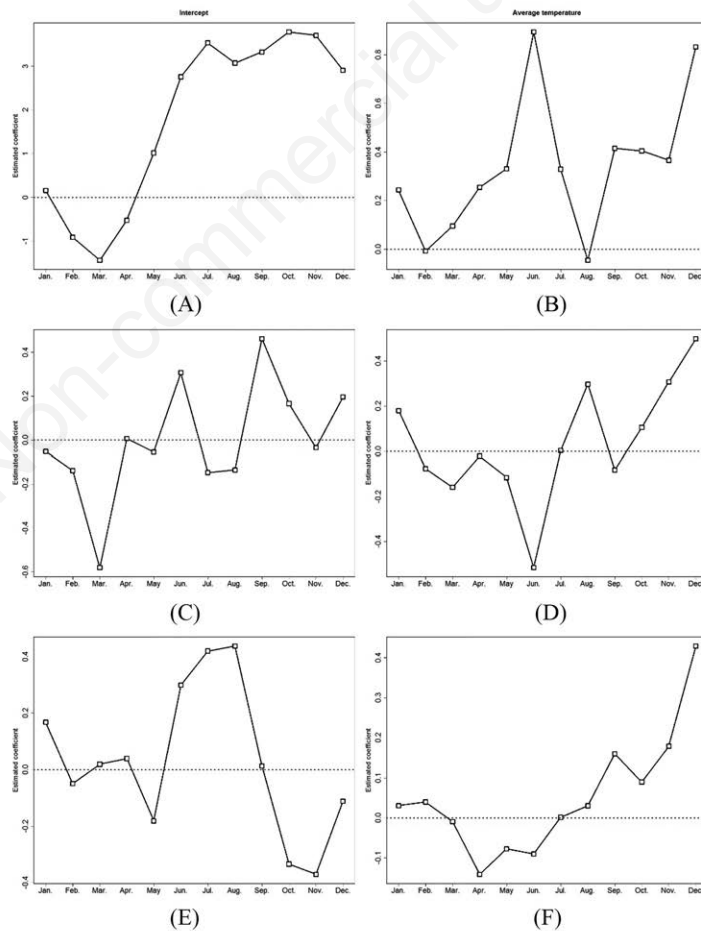

(A)

(B)

(C)

(D)

(E)

(F)

Figure 9. Temporal variation of the estimated intercept and coefficients of climate factors averaged over the counties: (A) Intercept; (B) Average temperature; (C) Average relative humidity; (D) Average air pressure; (E) Cumulative precipitation; and (F) Average wind speed.

trates that the AH has the strongest positive effect in September and the strongest negative influence in March. Figure 9D shows that the influence of the AP is consistently negative from February to June with the largest intensity appearing in June and it has a positive effect on the CI in the other months with the largest intensity appearing in December. Figure 9E illustrates that the effect of CR shows a temporal "∩-formed" trend in the warm season with the strongest positive effect taking place in July and August, approximately corresponding to the highest incidence of HFMD in Figure 1C. Combined with the findings shown in Figure 7, it might mean that the CR is a risk factor for HFMD transmission in the summer. Figure 9F shows that the effect of the AW is consistently negative from April to July with the largest intensity appearing in April. Except for the negligible influence in March and July, the effect in the other months is positive with the largest intensity taking place in December.

Conclusions

Previous investigations found significant relationships between climate factors and HFMD incidence (Cheng *et al*., 2018; Dong *et al*., 2016; Hong *et al*., 2018; Ma *et al*., 2010; Onozuka and Hashizume, 2011; Song *et al*., 2019; Wang *et al*., 2017; Xu *et al*., 2016; Yang *et al*., 2017; Yang *et al*., 2018). The relationship of climate factors and the transmission of HFMD is spatiotemporally non-stationary. The GWR model incorporating the time factor (TL-GWR) in this paper was used for analysing the spatial and temporal non-stationarities of the impact of climate factors on the occurrence of childhood HFMD in Inner Mongolia, China. The detailed spatiotemporal variation patterns of the effect of climate factors on HFMD incidence were uncovered based on the outputs of the TL-GWR model.

The results presented here are not only useful for a full understanding of the relationship between climate factors and HFMD occurrence, but also helpful in making measures for protecting and controlling HFMD at the county level in different seasons. This study also demonstrates the strengths and potentials of the TL-GWR model in detecting spatiotemporal non-stationarity of a regression relationship. However, there are still much room for the improvement of the TL-GWR model. First, one of the important issues for the TL-GWR model is to develop the optimal spatial and temporal bandwidth selection procedure to deal with the computational complexity, especially for big datasets. Second, the local tests would be developed to judge the statistical inferences of spatiotemporal characteristics shown by the estimated coefficients.

References

Brunsdon C, Fotheringham AS, Charlton ME, 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr Anal 28:281-98.

Cheng Q, Bai L, Zhang Y, Zhang H, Wang S, et al., 2018. Ambient temperature, humidity and hand, foot, and mouth disease: a systematic review and meta-analysis. Sci Total Environ 625:828-36.

Dong WH, Li XE, Yang P, Liao H, Wang XL, Wang QY, 2016. The effects of weather factors on hand, foot and mouth disease in Beijing. Sci Rep 6:19247.

Duan SB and Li ZL, 2016. Spatial downscaling of MODIS land surface temperatures using geographically weighted regres-

sion: case study in northern China. IEEE T Geosci Remote 54:6458-69.

Fotheringham AS, Brunsdon C, Charlton, ME, 2002. Geographically Weighted regression: the analysis of spatially varying relationships. Chichester, U.K.: Wiley.

Fotheringham AS, Crespo R, Yao J, 2015. Geographical and temporal weighted regression (GTWR). Geogr Anal 47:431-52.

Haining R, 2003. Spatial data analysis: theory and practice. Cambridge, U.K.: Cambridge University Press. doi:10.1017/CBO9780511754944

Hong ZM, Hao H, Wang XL, Wang WR, Wei LD, 2017. [Analysis of spatio-temporal epidemiology of severe hand, foot and mouth disease in Inner Mongolia from 2009 to 2016.] Chinese J Dis Control Prev 21:1048-51. doi:10.16462 /j.cnki.zhjbkz. 2017.10.019 [Article in Chinese].

Hong ZM, Hao H, Li CY, Du W, Wei LD, Wang HH, 2018. Exploration of potential risks of hand, foot, and mouth disease in Inner Mongolia Autonomous Region, China using geographically weighted regression model. Sci Rep 8:17707.

Huang B, Wu B, Barry M, 2010. Geographically and temporally weighted regression for spatio-temporal modeling of house prices. Int J Geogr Inf Sci 24:383-401.

Huang JX, Wang JF, Bo YC, Xu CD, Hu MG, Huang DC, 2014. Identification of health risks of hand, foot and mouth disease in China using the geographical detector technique. Int J Environ Res Public Health 11:3407-23.

Hu MG, Li ZJ, Wang JF, Jia L, Liao YL, et al., 2012. Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. Plos One 7:e38978.

Lai CC, Jiang DS, Wu HM, Chen HH, 2016. A dynamic model for the outbreaks of hand, foot, and mouth disease in Taiwan. Epidemiol Infect 144:1500-11.

Lu B, Charlton M, Harris P, Fotheringham AS, 2014. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. Int J Geogr Inf Sci 28:660-81.

Ma E, Lam T, Wong C, Chuang SK, 2010. Is hand, foot and mouth disease associated with meteorological parameters? Epidemiol Infect 138:1779-88.

Onozuka D, Hashizume M, 2011. The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. Sci Total Environ 410-411: 119-25 .

Qian HK, Huo D, Wang XL, Jia L, Li XT, et al., 2016. Detecting spatial-temporal cluster of hand foot and mouth disease in Beijing, China, 2009-2014. BMC Infect Dis 16:206.

Sham NM, Krishnarajah I, Ibrahim NA, Lye MS, 2014. Temporal and spatial mapping of hand, foot and mouth disease in Sarawak Malaysia. Geospat Health 8:503-7.

Song XD, Brus DJ, Liu F, Li DC, Zhao YG, et al., 2016. Mapping soil organic carbon content by geographically weighted regression: a case study in the Heihe River Basin, China. Geoderma 261:11-22.

Song C, Shi X, Bo YC, Wang JF, Wang Y, Huang DC, 2019. Exploring spatiotemporal nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian spatiotemporally varying coefficients (STVC) model in Sichuan, China. Sci Total Environ 648:550-60.

Wabiri N, Shisana O, Zuma K, Freeman J, 2016. Assessing the spatial nonstationarity in relationship between local patterns of HIV infections and the covariates in South Africa: a geograph-

ically weighted regression analysis. Spat Spatiotemporal Epidemiol 16:88-99.

Wang JF, Xu CD, Tong SL, Chen HY, Yang WZ, 2013. Spatial dynamic patterns of hand-foot-mouth disease in the People's Republic of China. Geospat Health 7:381-90.

Wang JJ, Cao ZD, Zeng Daniel D, Wang QY, 2017. Assessing local risk factors of Beijing hand-foot-mouth disease in China. Online J Public Health Informatics 9:e009.

Xing W, Liao Q, Viboud C, Zhang J, Sun J, et al., 2014. Hand, foot, and mouth disease in China, 2008–12: an epidemiological study. Lancet Infect Dis 14:308–18.

Xu J, Zhao D, Su H, Xie M, Cheng J, et al., 2016. Impact of temperature variability on childhood hand, foot and mouth disease in Huainan, China. Public Health 134:86-94.

Yang F, Zhang T, Hu Y, Wang X, Du J, et al., 2011. Survey of enterovirus infections from hand, foot and mouth disease outbreak in china, 2009. Virol J 8:508.

Yang H, Wu J, Cheng J, Wang X, Wen L, et al., 2017. Is high rel-
ative humidity associated with childhood hand, foot, and mouth disease in rural and urban areas? Public Health 142:201-7.

Yang YW, You EQ, Wu JJ, Zhang WY, Jin J, et al., 2018. Effects of relative humidity on childhood hand, foot, and mouth disease reinfection in Hefei, China. Sci Total Environ 630:820-6.

Yu GQ, Li YH, Cai JS, Yu DM, Tang JX, et al., 2019. Short-term effects of meteorological factors and air pollution on childhood hand-foot-mouth disease in Guilin, China. Sci Total Environ 646:460-70.

Zhang W, Du Z, Zhang D, Yu S, Hao Y, 2016. Boosted regression tree model-based assessment of the impacts of meteorological drivers of hand, foot and mouth disease in Guangdong, China. Sci Total Environ 553:366-71.

Zeng H, Lu J, Zheng H, Yi L, Guo X, et al., 2015. The epidemiological study of coxsackievirus A6 revealing hand, foot and mouth disease epidemic patterns in Guangdong, China. Sci Rep 5:10550.