Spatiotemporal analysis and forecasting model of hemorrhagic fever with renal syndrome in mainland China

Ling Sun[1] and Lu-Xi Zou[2]

[1]Department of Nephrology, Xuzhou Central Hospital, Medical College of Southeast University, Xuzhou, Jiangsu, China and [2]School of Management, Zhejiang University, Hangzhou, Zhejiang, China

Abstract

Hemorrhagic fever with renal syndrome (HFRS) caused by hantaviruses is a serious public health problem in China, accounting for 90% of HFRS cases reported globally. In this study, we applied geographical information system (GIS), spatial autocorrelation analyses and a seasonal autoregressive-integrated moving average (SARIMA) model to describe and predict HFRS epidemic with the objective of monitoring and forecasting HFRS in mainland China. Chinese HFRS data from 2004 to 2016 were obtained from National Infectious Diseases Reporting System (NIDRS) database and Chinese Centre for Disease Control and Prevention (CDC). GIS maps were produced to detect the spatial distribution of HFRS cases. The Moran's I was adopted in spatial global autocorrelation analysis to identify the integral spatiotemporal pattern of HFRS outbreaks, while the local Moran's I_i was performed to identify 'hotspot' regions of HFRS at province level. A fittest SARIMA model was developed to forecast HFRS incidence in the year 2016, which was selected by Akaike information criterion and Ljung–Box test. During 2004–2015, a total of 165 710 HFRS cases were reported with the average annual incidence at province level ranged from 0 to 13.05 per 100 000 persons. Global Moran's I analysis showed that the HFRS outbreaks presented spatially clustered distribution, with the degree of cluster gradually decreasing from 2004 to 2009, then turned out to be randomly distributed and reached lowest point in 2012. Local Moran's I_i identified that four provinces in northeast China contributed to a 'high–high' cluster as a traditional epidemic centre, and Shaanxi became another HFRS 'hotspot' region since 2011. The monthly incidence of HFRS decreased sharply from 2004 to 2009 in mainland China, then increased markedly from 2010 to 2012, and decreased again since 2013, with obvious seasonal fluctuations. The SARIMA $((0,1,3) \times (1,0,1)_{12})$ model was the most fittest forecasting model for the dataset of HFRS in mainland China. The spatiotemporal distribution of HFRS in mainland China varied in recent years; together with the SARIMA forecasting model, this study provided several potential decision supportive tools for the control and risk-management plan of HFRS in China.

Introduction

Hantaviruses cause two rodent-borne infectious diseases in human, hemorrhagic fever with renal syndrome (HFRS) in Europe and Asia and hantavirus cardiopulmonary syndrome in Americas. HFRS is endemic in all over China with exception of Taiwan province [1, 2], caused mainly by two types of hantaviruses, Hantaan virus and Seoul virus, and characterised by fever, bleeding and acute kidney injury [3]. In recent decades, China has the highest incidence of HFRS, accounted for nearly 90% of global HFRS cases [1].

Hantaviruses are unexpectedly stable in air and can survive for more than 10 days at room temperature [4], mainly carried by rodents, insectivores and bats, transmitted to human mainly via inhalation of virus-contaminated aerosols of excreta and secreta, and virus-contaminated food [5]. Generally speaking, HFRS emergence in human depends on reservoir host density, level of exposure to infectious viruses and frequency of contact between human and rodent populations, which could be influenced by temperature, rainfall, relative humidity, urbanisation, living and working conditions of local residents [6, 7]. All of the above factors result in the seasonal and regional variations of HFRS outbreaks. To better understand the changing trend of HFRS in China, it is necessary to identify its epidemiological distribution in the past and predict its spatiotemporal trend in the future.

Previous studies have been performed to analyse the spatiotemporal distribution of HFRS and to forecast HFRS epidemic in several provinces of China [8–11]. Few studies have investigated the spatiotemporal variation of HFRS outbreak all over China. In this study, we described the spatiotemporal characteristics of HFRS from January 2004 to December 2016. Furthermore, based on these historical data, we developed a seasonal autoregressive-integrated moving average

(SARIMA) model to forecast the seasonal trend of HFRS incidence in mainland China. This study could provide valuable information for the hygiene authorities to design and implement effective measures for the control and prevention of HFRS.

Methods

Data sources and collection

The data of confirmed HFRS cases from January 2004 to December 2016 were obtained from the National Infectious Diseases Reporting System (NIDRS) database and Chinese Centre for Disease Control and Prevention (CDC). HFRS cases were first diagnosed according to clinical symptoms, then blood samples were collected in the hospital and serological identification was performed in the laboratory of each provincial CDC to confirm the clinical diagnosis. All serologically confirmed cases were collected and reported to China CDC [3]. There were no surveillance data of HFRS cases obtained from Hong Kong, Macao and Taiwan. The data from HFRS surveillance system were aggregated as secondary data without personal information, thus informed consent was not required. This study was reviewed by the research institutional review board of the Xuzhou Central Hospital, Southeast University. The review board concluded that the utilisation of disease surveillance data did not require oversight by an ethics committee.

Geographical information system mapping

To conduct a geographical information system (GIS)-based analysis of the spatial distribution of HFRS, a province-level polygon map at 1:1 000 000 scale was obtained by the National Geomatics Centre of China, on which the province-level point layer that contained information regarding latitudes and longitudes of central points of each province was created. To lessen variations, the annual incidence of HFRS per 100 000 persons in each province was calculated. The annual incidence of HFRS in each province was mapped using a GIS technique in software ArcGIS (version 10.3, ESRI, Redlands, CA, USA). According to the annual average incidence, all provinces were grouped into six categories: no data areas; very low endemic areas with annual average incidence between 0 and 0.01/100 000 persons; low endemic areas with annual average incidence between 0.01 and 0.5/100 000 persons; medium endemic areas with incidence between 0.5 and 1.0/100 000; high endemic areas with incidence between 1.0 and 5.0/100 000; and very high endemic areas with incidence >5.0/100 000. The six types of categories were colour-coded on the maps.

Spatial autocorrelation analysis

Spatial autocorrelation is characterised by a correlation in a signal among nearby locations in space. Spatial autocorrelation is more complex than one-dimensional autocorrelation because spatial correlation is multi-dimensional and multi-directional. Moran's I is both the leading measure of and leading test on spatial autocorrelation [12]. Spatial autocorrelation measures and tests can be differentiated by the scope or scale of analysis. Traditionally, they are separated into 'global' and 'local' categories.

Global indicators of spatial association (GISA)

Global indicators of spatial association (GISA) is calculated to evaluate the spatial relationships between the regions in a whole dataset. As a global statistic, Moran's I indicates not only the existence of spatial autocorrelation (positive or negative) but also the degree of spatial autocorrelation [13, 14]. GISA describes the associations of all spatial units, using Moran's I as a principal parameter. Moran's I is defined in equation (1) [15]:

$$I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{1}^{n} (y_i - \bar{y})^2} \quad i \neq j, \tag{1}$$

where n is the number of spatial units indexed by i and j; y_i and y_j are the variables of interest at points i and j (with $i \neq j$); $\bar{y}$ is the mean of y; w_{ij} is an element of the weight matrix ($n \times n$), which is defined as follows: when location i is contiguous to location j, the weight w_{ij} is given the weight of 1, otherwise the w_{ij} is given the weight of 0. If $I \in (0,1]$, there is a positive autocorrelation; if $I = 0$, the spatial distribution is random. If $I \in [-1,0)$, there is a negative autocorrelation. In this study, n referred to 31, the number of province-level regions in mainland China; y_j referred to the incidence of HFRS in province i; and $\bar{y}$ was the average value of HFRS incidence in 31 province-level regions in mainland China.

Local indicators of spatial association

Local indicators of spatial association (LISA) statistics is created by Anselin (1995), whose motivation is to decompose global statistics such as Moran's I into their local components for the purpose of identifying influential observations and outliers. Anselin local Moran's I_i is defined as equation (2) [16]:

$$I_i = \frac{y_i - \bar{y}}{(1/n) \sum_{i=1}^{n} (y_i - \bar{y})^2} \sum_{j=1}^{n} w_{ij}(y_i - y_j) \quad i \neq j, \tag{2}$$

$$\text{for } j \text{ within } d \text{ of } i,$$

where j is within d (distance) of i, I_i refers to the value of local Moran's I at points i. Meanwhile, the parameters y_i, y_j, $\bar{y}$ and w_{ij} have the same meaning as they are in equation (1).

LISA describes the spatial associations among a studied spatial unit and its contiguous spatial units. The analysis of LISA provides a better method to identify hotspots all over the studied areas by identifying clustering pairs of neighbouring values [17]. In our study, a positive local Moran's I_i implied that the HFRS incidence at province i had a similarly value with its neighbours; in other words, these spatial units were in a spatial cluster. There were two types of spatial clusters, 'high–high' cluster and 'low–low' cluster. If the province i and its contiguous provinces all had high values of HFRS incidence, they were in 'high–high' cluster, otherwise they were in 'low–low' cluster. A negative local Moran's I_i value suggested that the HFRS incidence at province i had a very different value with its neighbours; in other words, these spatial units were in a spatial outlier. There were two types of spatial outliers, 'high–low' outlier and 'low–high' outlier. If the province i had a high value of HFRS incidence and its contiguous provinces all had low values, they were in 'high–low' outlier, otherwise they were in 'low–high' outlier. The calculation of spatial analysis was performed using the software GeoDa (version 1.10, Spatial Analysis Laboratory, Urbana, IL, USA).
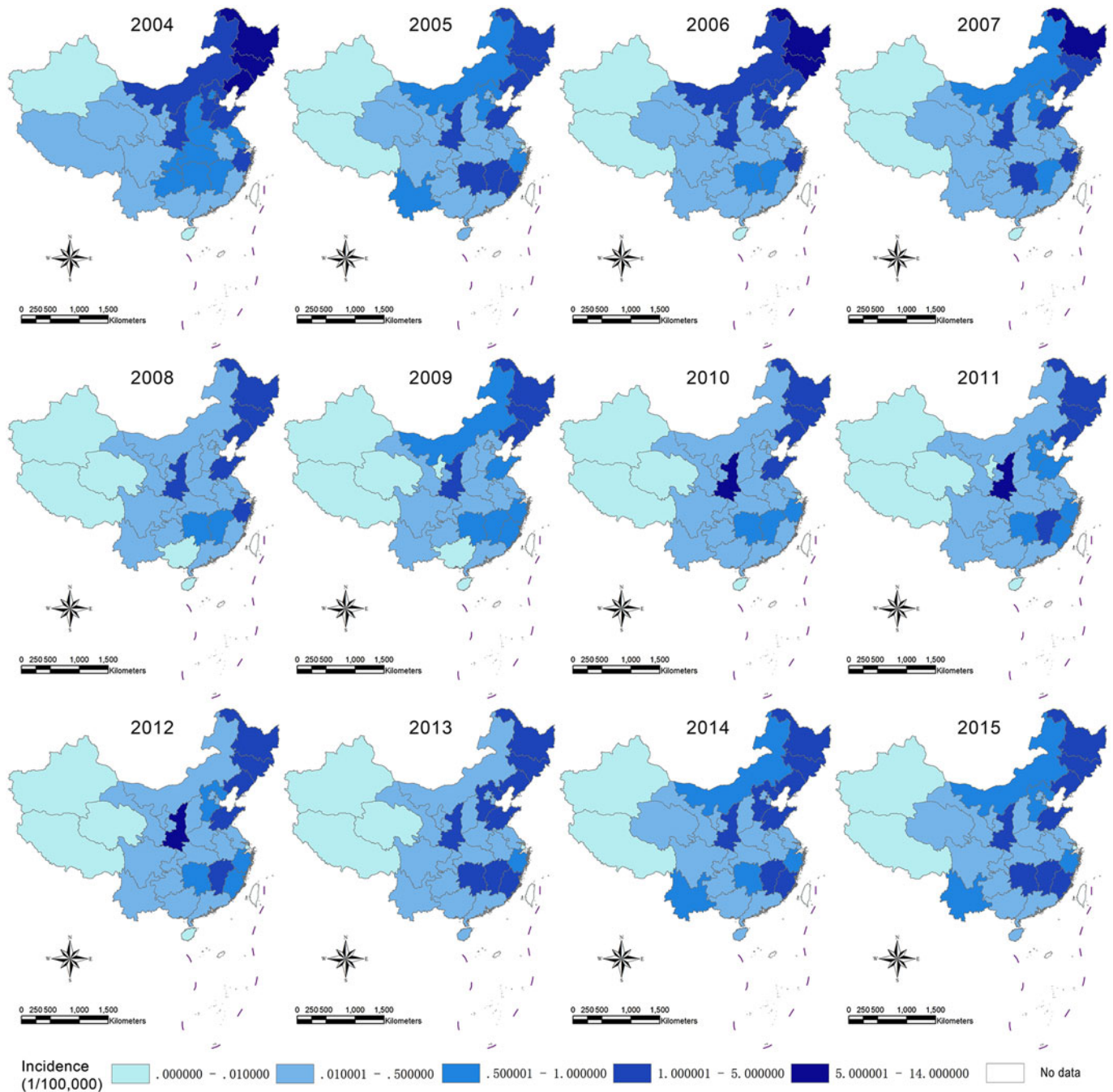
Fig. 1. Yearly distribution of HFRS incidence in mainland China, 2004–2015.

SARIMA model construction

Several statistical models have been used in the forecasting of infectious diseases [8–11, 18–20]. SARIMA model is a traditional method to study the time-series dataset, and is powerful in applying reference data to study the control, prevention and forecast of seasonal infectious diseases [20, 21]. In China, the outbreaks of HFRS on record have strong seasonality trends; therefore, we aimed to construct a SARIMA (p, d, q) × (P, D, Q)$_S$ model to predict the HFRS incidence accurately in future. The notation (p, d, q) × (P, D, Q)$_S$ describes the composition of temporal patterns considered for forecasting: these include autocorrelation over a maximum of p months or over P periods, each of length $S = 12$

months in our dataset; differencing over d adjacent months or D periods; and moving averages sustained over q months or Q periods. If the series is not stationary, it can be converted into a stationary series through differencing [22]. The SARIMA model is defined as equation (3), which comprises non-seasonality and seasonality components as equations (4) and (5) [18]:

$$\Phi(B^s)\phi(B)(x_t - \mu) = \Theta(B^s)\theta(B)\varepsilon_t. \tag{3}$$

The non-seasonality components are:

$$\emptyset(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right),$$

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right). \quad (4)$$

The seasonality components are:

$$\Phi(B^s) = \left(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}\right),$$

$$\Theta(B^s) = \left(1 - \Theta_1 B^\omega - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}\right). \quad (5)$$

In these equations, B represents the backward shift operator, s is the rotation period and the rest of the cases. ε_t stands for estimated residual error at t, and x_t represents the observed values at t ($t = 1, 2, \ldots, k$). In this study, the dataset of annual HFRS incidence was split into a training period and a validation period, the latter was used to test the predictive ability of models and select out the fittest one, using the R statistical package (version 3.4.3) and Akaike's information criterion (AIC) [18]. The Ljung–Box test was used to examine the distribution of the residuals from the selected SARIMA model to validate the goodness of fit of the model [23, 24].

Results

Spatial distribution of HFRS incidence in mainland China, 2004–2015

From 2004 to 2015, there were a total of 165710 HFRS cases in mainland China, all cases had been confirmed by the laboratories of Chinese CDC, then reported to the NIDRS database. The annual average incidence at province level ranged from 0 to 13.05 per 100 000. During 2004–2015, the HFRS incidence varied among these provinces (Fig. 1). Tibet was non-endemic since 2005. Xinjiang remained non-endemic, with the exception of 2009, in which year its HFRS incidence was 0.01/100 000. Tibet and Xinjiang altogether cover 29.94% of the total land and 1.96% of the total population in China. In 2004, Heilongjiang, Jilin, Liaoning presented the highest incidence (covering 8.17% of the total land and 7.99% of the total population), and Inner Mongolia, Hebei, Shandong, Shaanxi, Zhejiang and Jiangxi took the second place (covering 20.72% of the total land and 24.57% of the total population); the HFRS incidence of these provinces all had a declined trend since 2005, except for Shaanxi. From 2010 to 2012, Shaanxi surpassed Heilongjiang and became the province with highest HFRS incidence. Since 2013, the annual HFRS incidence in all provinces had not exceeded 5.0/100 000 persons, and HFRS incidence in Heilongjiang rose to the first place again, followed by Shaanxi, Liaoning and Jilin. The original time-series data were presented in File S1.

GISA analysis for HFRS incidence

The spatial autocorrelation was analysed based on annual average province-level incidence of HFRS in mainland China. The statistic was defined to be significant for Moran's I at significance level of $P < 0.01$. The statistical significance of Moran's I represented the spatial cluster of HFRS outbreaks. GISA analysis (Table 1) showed that the HFRS outbreaks presented spatially clustered distribution with Z-score >2.81 and P-value <0.01, and the clustering degree of HFRS cases was gradually decreased from 2004 to 2009. Except 2014, HFRS outbreaks turned out to be randomly distributed since 2010, and reached to its lowest point with a high discrete

Table 1. The Global Moran's I analysis for HFRS incidence

Year	Moran's I	Z-score	P	Pattern
2004	0.388	3.896	<0.001	Cluster
2005	0.528	5.113	<0.001	Cluster
2006	0.468	4.949	<0.001	Cluster
2007	0.296	3.601	<0.001	Cluster
2008	0.225	2.390	0.017	Cluster
2009	0.275	2.810	0.005	Cluster
2010	0.056	0.886	0.376	NS
2011	0.014	0.455	0.649	NS
2012	−0.034	−0.037	0.970	NS
2013	0.172	1.797	0.072	NS
2014	0.297	2.975	0.003	Cluster
2015	0.162	1.712	0.087	NS

NS, not significant.

pattern in 2012 (Moran's $I = -0.034$ with Z-score $= -0.037$ and P-value $= 0.970$).

LISA analysis for HFRS incidence

The LISA statistic was performed to identify the spatial clusters of HFRS epidemic. High values of LISA indicated that the features inside the fixed neighbourhood were homogeneous; otherwise, the features inside the fixed neighbourhood were heterogeneous. The yearly LISA cluster maps of HFRS incidence (Fig. 2) demonstrated that Jilin, Liaoning and Inner Mongolia constituted a 'high–high' cluster in 2004; Heilongjiang, Jilin, Liaoning and Inner Mongolia constituted a 'high–high' cluster in 2005. Since 2006, Inner Mongolia became a 'low–high' zone with the exception of 2014. Shaanxi used to be not different from its surrounding provinces, then it became an HFRS 'hotspot' and an obvious 'high–low' zone since 2011.

SARIMA model building for HFRS forecasting

The monthly HFRS incidence in mainland China were calculated and plotted to show seasonal fluctuations (Fig. 3a). The results showed that the monthly incidence of HFRS decreased sharply from 2004 to 2009, then increased markedly from 2010 to 2012, and have decreased again since 2013. The HFRS outbreaks were found to vary seasonally, most cases occurred in the winter (November to January) and early summer (May to July), and usually peaked in June and November.

HFRS incidence data from January 2004 to December 2016 were used to construct a fittest SARIMA model. The dataset was split into a training period (January 2005 to December 2015), used as a platform for creating the SARIMA models, and a validation period (January 2016 to December 2016), used to test the models' predictive ability. Autocorrelation function (ACF) plots and partial autocorrelation function (PACF) plots were used to determine the key parameters (p, P, d, D, q, Q) of SARIMA models. If all plots of ACF are close to zero, then this dataset should be in a white noise series [25]. Figs 3b and c showed that the monthly HFRS incidence was not white noise.
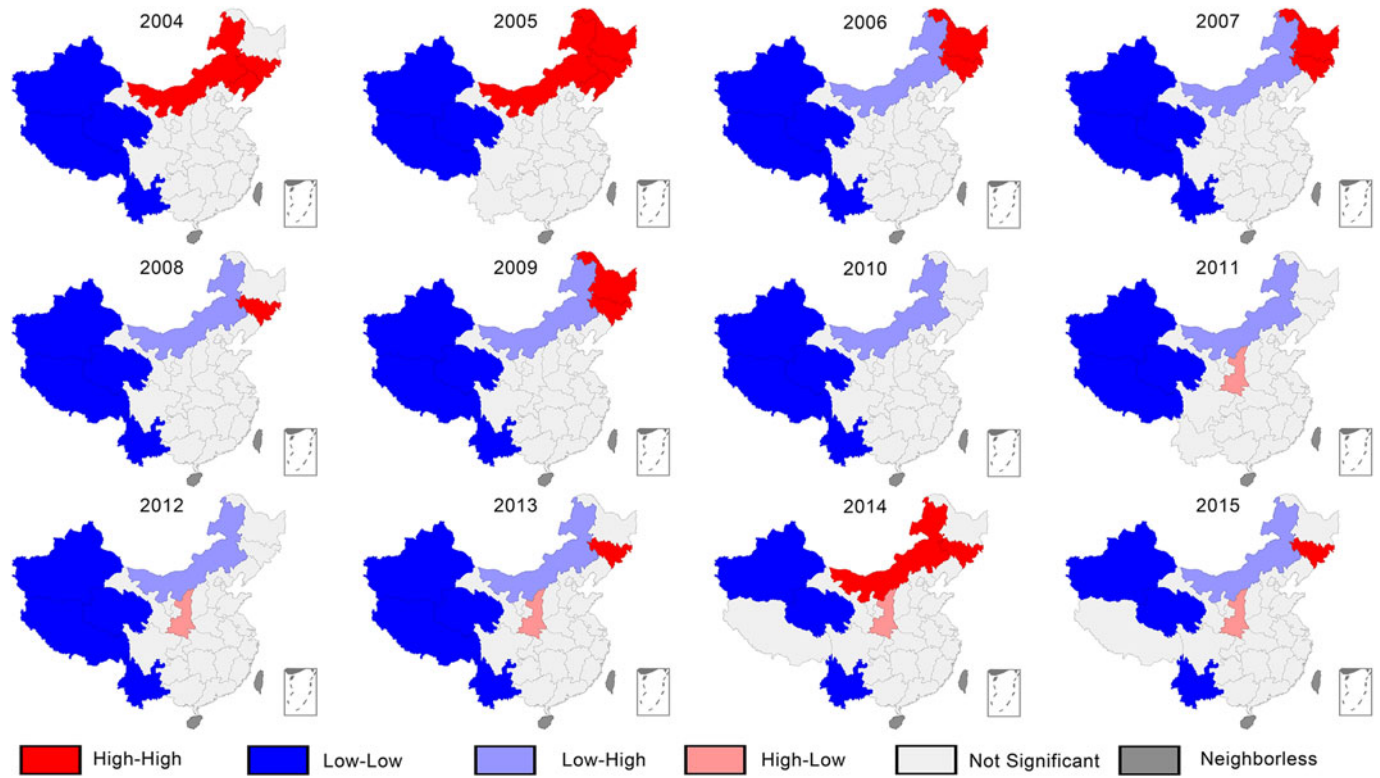
Fig. 2. Yearly LISA cluster maps of HFRS at province level in mainland China, 2004–2015. *Per 100 000 persons.
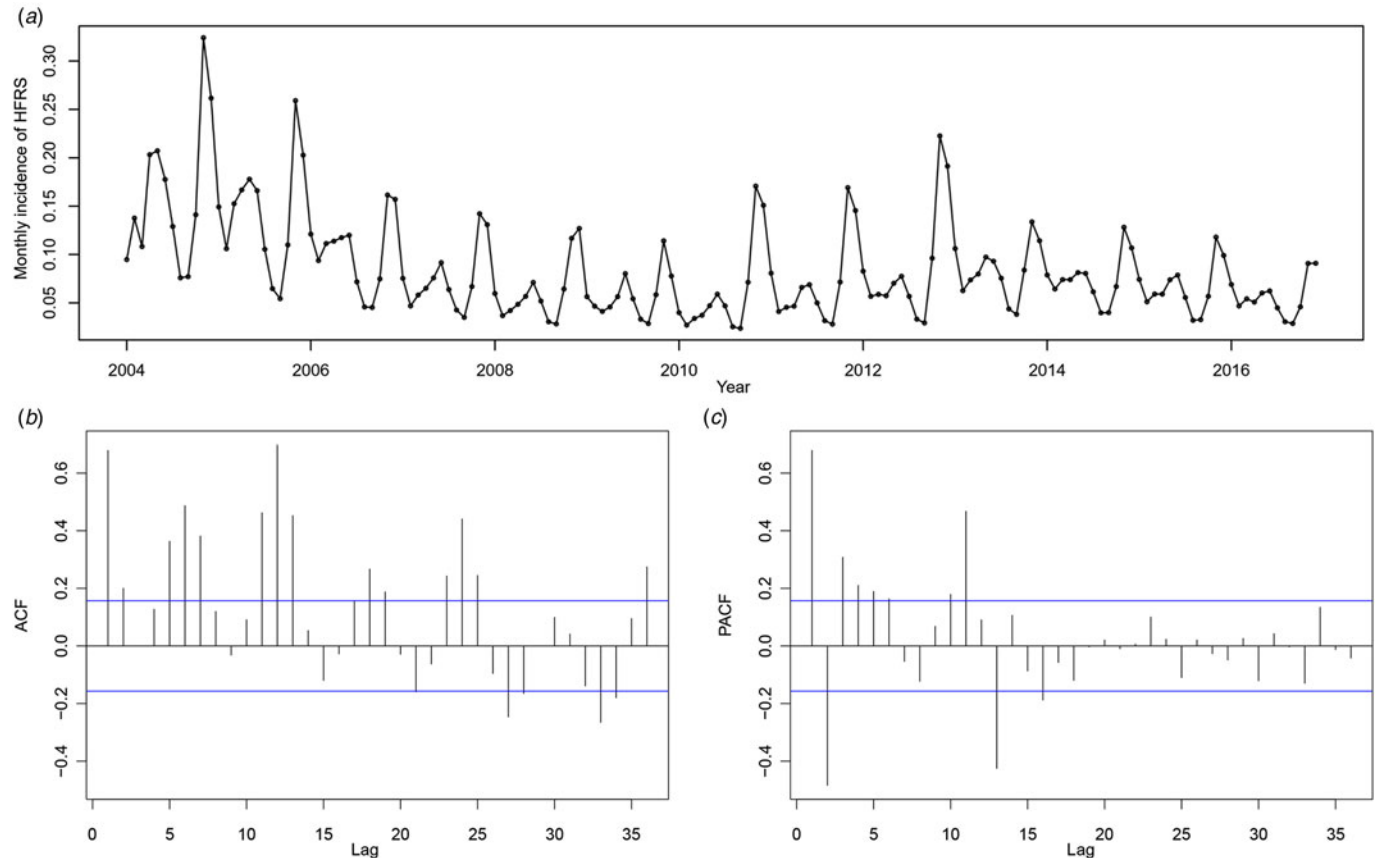

Fig. 3. Temporal distribution of HFRS in mainland China from January 2004 to December 2016. (a) The values of monthly HFRS incidence; ACF plots (b) and PACF plots (c) for monthly HFRS incidence.
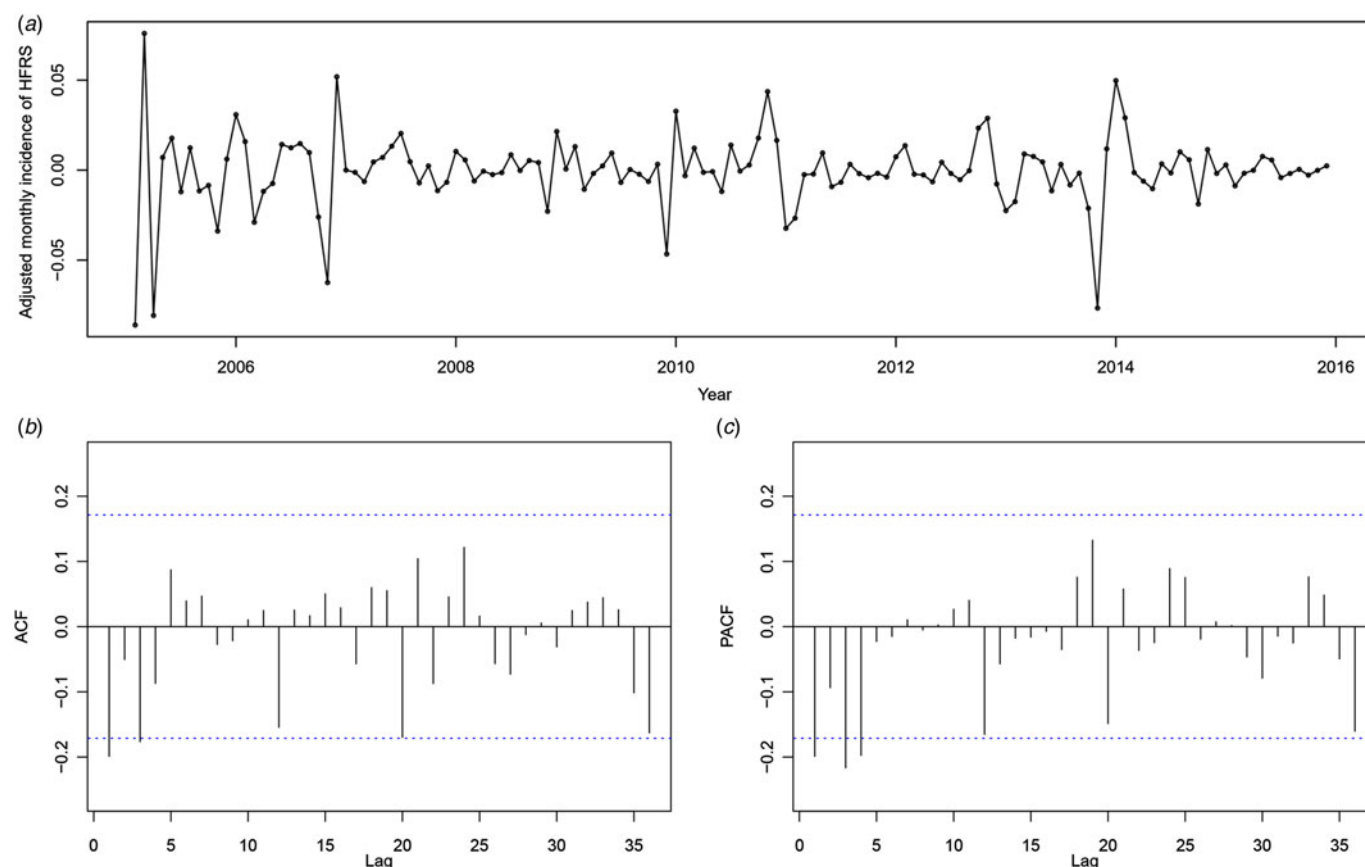
Fig. 4. Temporal distribution of HFRS in mainland China adjusted by first differencing and seasonal differencing, 2004–2016. (a) The values of adjusted monthly HFRS incidence; ACF plots (b) and PACF plots (c) for adjusted monthly HFRS incidence.

Table 2. Coefficients and standard errors of the parameters in SARIMA $((0,1,3) \times (1,0,1)_{12})$ model

	Coefficient	Standard error
MA1	−0.3289	0.0851
MA2	−0.1115	0.0980
MA3	−0.3418	0.0926
Seasonal AR1	0.9310	0.0309
Seasonal MA1	−0.1694	0.1196

Then we performed one-order trend differencing, seasonal differencing and augmented Dickey–Fuller test, which were necessary to stabilise the variance of HFRS incidence. Figure 4a described the temporal distribution of HFRS incidence in mainland China from 2004 to 2015 after one-order trend differencing and seasonal differencing. Figs 4b and c showed the ACF plots and PACF plots for seasonality-adjusted monthly HFRS incidence after differencing.

Concerning each parameter between zero and five (P, p, Q, q = 0, 1, 2, 3, 4, 5), various models were constructed and tested. The SARIMA $((0,1,3) \times (1,0,1)_{12})$ was selected to be the best-fit model, which had the lowest value of AIC (AIC = −702.35) of all constructed models. The coefficients and standard errors of the parameters in SARIMA $((0,1,3) \times (1,0,1)_{12})$ model were listed in Table 2.

The Ljung–Box test was performed to examine the distribution of the residuals, which were the differences between the observed values from dataset and predicted values from the SARIMA $((0,1,3) \times (1,0,1)_{12})$ model (Fig. 5a). All spikes in the ACF plots of residuals were in significance limits (Fig. 5b). Figure 5c indicated that the residuals had no autocorrelations and distributed independently ($P > 0.05$). Figure 5 validated that these residuals were white noise, which was desirable. Therefore, the fittest SARIMA model appeared, which passed all the required checks and was ready for prediction. Using this SARIMA $((0,1,3) \times (1,0,1)_{12})$ model, we predicted the values of monthly HFRS incidence from 2016 to 2020, then compared the predicted values with the observed values in 2016. As shown in Figure 6, the line of observed values almost coincided with the line of predicted values. The SARIMA prediction model always has a relatively wide 95% confidence interval, which is statistically acceptable [26, 27].

Discussion

HFRS is a kind of highly fatal infectious disease with murine being the major source of infection, and has caused severe influences worldwide. HFRS has been recognised as a serious public health problem in mainland China since it remains one of the top 10 communicable diseases for decades [28]. The incidence of HFRS is highly variable at province level. In this study, the application of GIS and spatial autocorrelation analysis provided
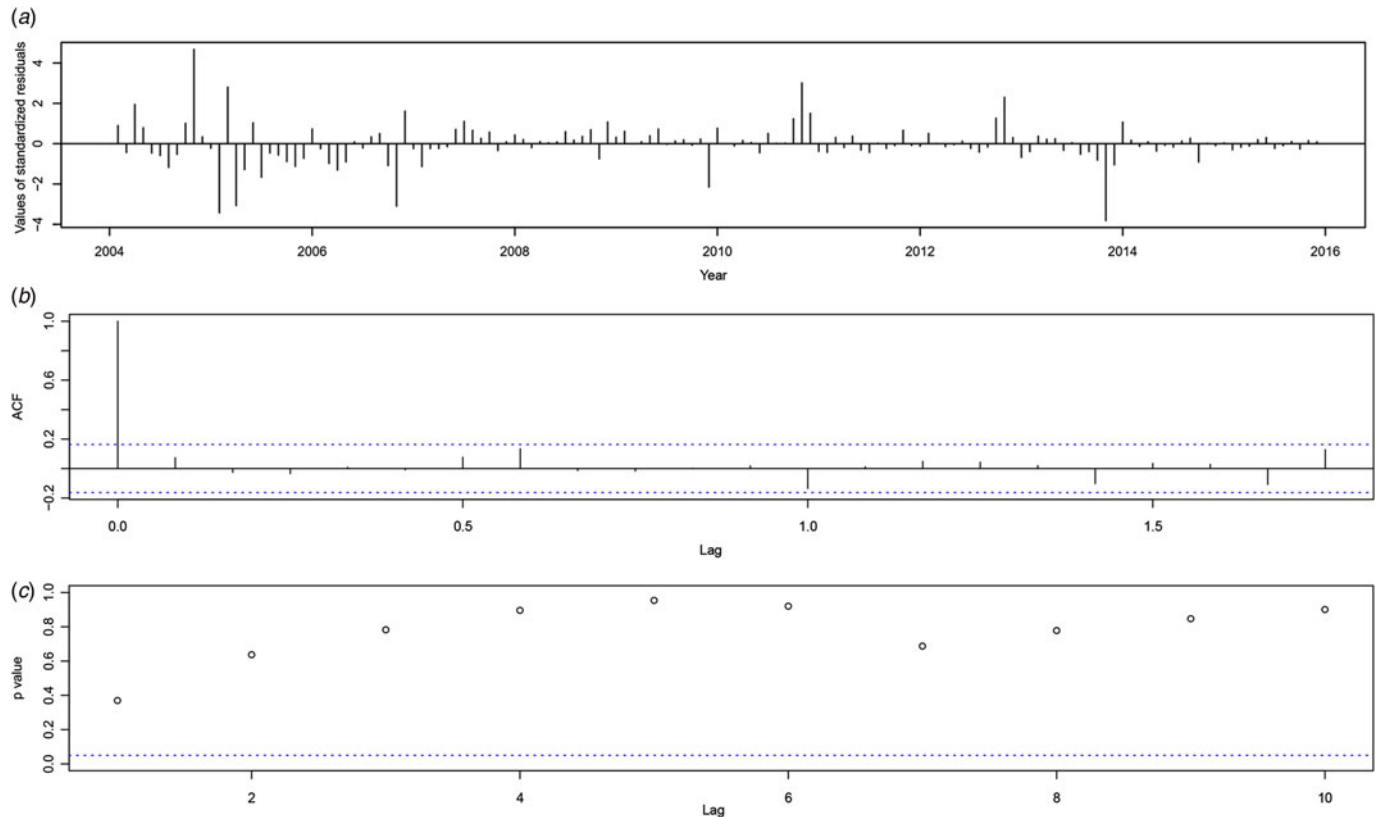
Fig. 5. Standardised residuals from the SARIMA $((0,1,3) \times (1,0,1)_{12})$ model applied to HFRS incidence, 2004–2015. (a) Values of standardised residuals of monthly HFRS incidence; (b) ACF plots for standardised residuals in (a); (c) P values for the standardised residuals in (a) by Ljung–Box test.
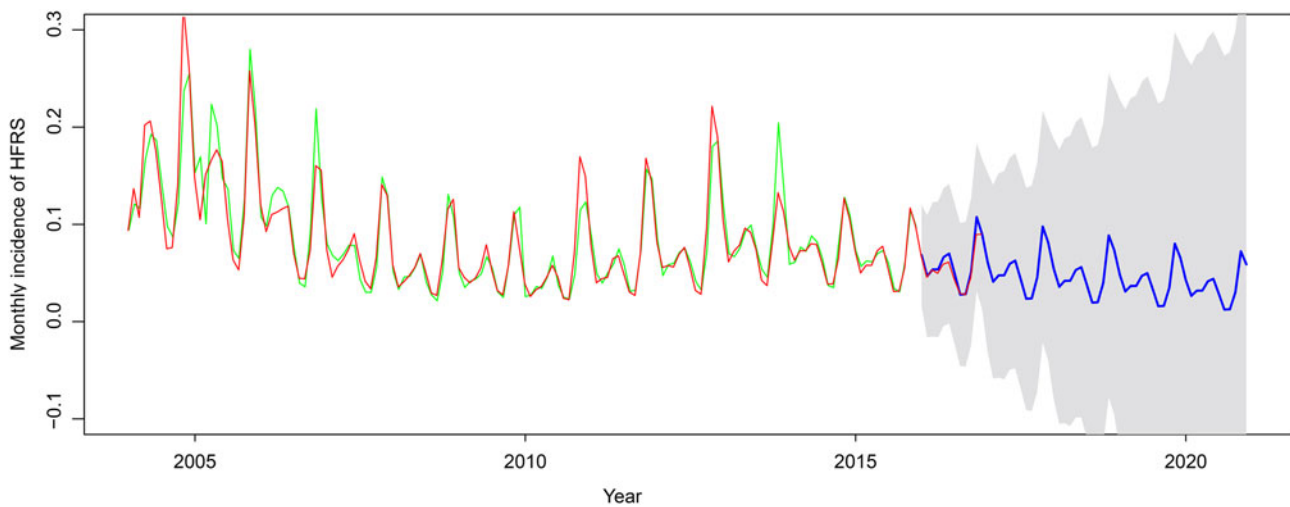

Fig. 6. Forecasted counts of HFRS incidence in 2016 according to the SARIMA $((0,1,3) \times (1,0,1)_{12})$ model. The solid red line represented the observed values; the solid green line followed by blue line indicated the forecasted curve; the grey area showed the upper and lower 95% confidence limits for the forecasted counts.

ways to quantify HFRS outbreaks and to further identify geographical risk factors for the disease.

Using GIS-based spatial statistics, we investigated the spatial distribution of HFRS cases and identified provinces with high endemic HFRS and clustering patterns. From 2004 to 2015, the high endemic areas of HFRS have changed from northeastern China (traditional epidemic areas) towards the middle and southeast parts of China. As a whole, HFRS incidence showed an obviously decreasing trend in most provinces, particularly in Heilongjiang, Liaoning and Jilin. However, HFRS epidemic presented a trend of rebound in several provinces, such as Shaanxi, Shandong and Jiangxi. Meanwhile, there was a decreasing trend in the Global Moran's I from 2004 to 2012, which indicated that the distribution of HFRS outbreaks changed from cluster to random, and then the Global Moran's I turned out to increase from 2012 to 2014, with reappearance of cluster aggregation of

HFRS in 2014. This phenomenon could possibly be explained by several factors. First, hantaviruses were mainly carried and transmitted by rodents; the population density and types of rodent species significantly impact on the occurrence of HFRS [6, 7, 29]. Previous reports demonstrated that the mice-positive rate dropped from 2005 to 2011, but suddenly rose in 2012 in mainland China [7, 11], which might be influenced by the changes of climatic factors [30]. Second, effective vaccination programmes have been conducted in traditional epidemic areas, such as Heilongjiang, Liaoning and Jilin, and reduced the proportion of infections resulting in HFRS in these traditional epidemic provinces [31–33]. Third, the ratio of rural population over total population is decreasing year-by-year in mainland China. The human migration (from rural areas to cities), urbanisation, improvement of housing and workplace conditions could all reduce the risk of human exposing to rodent excreta [3]. These may also affect the incidence of HFRS as a decreased trend in the whole period.

LISA detected spatial clusters of HFRS with high–high pattern. The clusters with the high–high pattern were recognised as 'hotspots'. Figure 2 showed that the northeast three provinces, Inner Mongolia and Shaanxi were high-risk areas of HFRS epidemic, closely correlated with meteorological factors including temperature, relative humidity, precipitation, vegetation type, land use and multivariate El Niño Southern Oscillation (ENSO) [8, 34, 35]. We assumed that the above meteorological factors could affect rodent density and activity as well as infectivity of hantaviruses. Moreover, socio-economic status may also contribute to high incidence. Compared with provinces with similar high population density in eastern and southern China, these areas have poorer living conditions and sanitation, which increased the frequency of contact between the human and rodent populations. Efficient allocation of health resources for HFRS control and intervention requires accurate information on its geographical distribution. The LISA cluster maps suggested that the targeted policies for prevention and control of HFRS should be made, particularly in the regions of 'high–high' cluster and 'high–low' zone.

There is no universal model applicable to any environment due to the inherent complexity of a time-series dataset in the real world [36]. In our study, the model was designed to improve the forecasting accuracy from the data driven by incorporating the intrinsic characteristics of the historical time-series data on HFRS incidence in mainland China. Due to the seasonal variations in HFRS epidemic, a seasonal ARIMA model can adequately simulate the HFRS. We applied SARIMA $((p, d, q) \times (P, D, Q)_S)$ models to analyse the surveillance data of HFRS in mainland China. After adjusting for these trends in HFRS incidence, the best fit to the dataset was a SARIMA $((0,1,3) \times (10,1)_{12})$ model, with the lowest value of AIC, which indicated that the number of monthly HFRS incidence can be estimated from the incidence occurring 12 months before, including differencing over 1 adjacent month. Again, the moving average parameters indicated a drop-in magnitude of average HFRS incidence in a given month compared with 3 and 12 months before. HFRS control in China follows the principle of 'three-early and one-in-place', namely, early discovery, early rest, early treatment and in-place isolation treatment, which renders great progress in the prevention of HFRS [37]. In accordance with the previous studies [14, 38], our time-series analytic results showed a decreasing trend of HFRS incidence since 2004, indicated that the epidemic trend of HFRS in China was under control and the control strategies had attained certain achievements. With the utilisation of a SARIMA model, the short-term predicting results expected that HFRS incidence would continue to decline over the next year, which implied that the national monitoring programme would continue to operate effectively in HFRS control in the near future.

Our research has some advantages in terms of study and prevent HFRS epidemic. Our study focused on the spatial and temporal epidemiology of HFRS in mainland China; the dataset was large and accurate. Several studies have been done on the spatiotemporal epidemiology of HFRS in specific provinces of China, but regarding to nationwide data, few research could be found. With the help of a SARIMA model, it is reasonable for the government to allocate health resources to control HFRS epidemic efficiently. If the forecasted values continue to rise, the government should allocate more resources into health interventions in advance. It can also be useful to evaluate the effectiveness of currently used intervention strategies by the change of forecasting trend of HFRS incidence. We confirm that our study is able to assist public health officials in HFRS controlling, epidemics prediction and medical service sources disposition.

Limitations of this study should also be acknowledged. In addition to the inherent features of time series, time-series data on other influencing factors were not mentioned, such as economic factors and human activities; it was difficult to further uncover the probable causes and shifts of HFRS outbreaks. Future researches are warranted to focus on the risk factors of HFRS to modify the ARIMA model, such as rodent population density, human activities, socio-economic and environmental factors, particularly in the 'hotspot' provinces.

Conclusions

This study explored the spatiotemporal features of HFRS from 2004 to 2015 in mainland China, using GIS, GISA and LISA analyses. A SARIMA model was constructed to monitor and predict the trends of HFRS outbreaks. Our results provided latest data and decision support tools for the hygiene authorities to design and implement effective measures for the control and prevention of HFRS in China.

Author contributions. LS and LZ conceived and designed the experiments; performed the experiments; analysed the data; and contributed reagents/materials/analysis tools. LS wrote the paper. LZ revised the manuscript.

Conflict of interest. The authors declare that there is no conflict of interest.

References

1. **Li SJ et al.** (2014) Spatiotemporal heterogeneity analysis of hemorrhagic fever with renal syndrome in China using geographically weighted regression models. *International Journal of Environmental Research & Public Health* 11, 12129–12147.

2. **Zhang YZ et al.** (2010) Hantavirus infections in humans and animals, China. *Emerging Infectious Diseases* **16**, 1195–1203.

3. **Zou LX, Chen MJ and Sun L** (2016) Haemorrhagic fever with renal syndrome: literature review and distribution analysis in China. *International Journal of Infectious Diseases* **43**, 95–100.

4. **Hardestam J et al.** (2007) Ex vivo stability of the rodent-borne Hantaan virus in comparison to that of arthropod-borne members of the Bunyaviridae family. *Applied & Environmental Microbiology* **73**, 2547–2551.

5. **Pedrosa PBS and Cardoso TAO** (2011) Viral infections in workers in hospital and research laboratory settings: a comparative review of infection modes and respective biosafety aspects. *International Journal of Infectious Diseases* **15**, E366–E376.

6. **Reusken C and Heyman P** (2013) Factors driving hantavirus emergence in Europe. *Current Opinion in Virology* **3**, 92–99.

7. **Xiao H et al.** (2013) Investigating the effects of food available and climatic variables on the animal host density of hemorrhagic fever with renal syndrome in Changsha, China. *PLoS ONE* **8**, e61536.

8. **Fang LQ et al.** (2010) Spatiotemporal trends and climatic factors of hemorrhagic fever with renal syndrome epidemic in Shandong Province, China. *PLoS Neglected Tropical Diseases* **4**, e789. doi: 10.1371/journal.pntd.0000789.

9. **Lin HL et al.** (2007) Analysis of the geographic distribution of HFRS in Liaoning Province between 2000 and 2005. *BMC Public Health* **7**, 207. doi: 10.1186/1471-2458-7-207.

10. **Wu W et al.** (2011) Clusters of spatial, temporal, and space-time distribution of hemorrhagic fever with renal syndrome in Liaoning Province, Northeastern China. *BMC Infectious Diseases* **11**, 229. doi: 10.1186/1471-2334-11-229.

11. **Ge L et al.** (2016) Spatio-temporal pattern and influencing factors of hemorrhagic fever with renal syndrome (HFRS) in Hubei Province (China) between 2005 and 2014. *PLoS ONE* **11**, e0167836. doi: 10.1371/journal.pone.0167836.

12. **Cliff AD and Ord JK** (1973) Spatial autocorrelation. *Trends in Ecology & Evolution* **14**, 196.

13. **Li RZ et al.** (2017) Epidemiological characteristics and spatial-temporal clusters of mumps in Shandong Province, China, 2005–2014. *Scientific Reports* **7**, 46328. doi: 10.1038/srep46328.

14. **Ge E et al.** (2016) Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009–2012. *Infectious Diseases of Poverty* **5**, 11. doi: 10.1186/s40249-016-0104-2.

15. **Moran PAP** (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society* **10**, 243–251.

16. **Anselin L** (1995) Local indicators of spatiala association – LISA. *Geographical Analysis* **27**, 93–115.

17. **Brown TT, Wood JD and Griffith DA** (2017) Using spatial autocorrelation analysis to guide mixed methods survey sample design decisions. *Journal of Mixed Methods Research* **11**, 394–414.

18. **Azeez A et al.** (2016) Seasonality and trend forecasting of tuberculosis prevalence data in Eastern Cape, South Africa, using a hybrid model. *International Journal of Environmental Research and Public Health* **13**, 757.

19. **Fernandez-Gonzalez M et al.** (2016) Prediction of biological sensors appearance with ARIMA models as a tool for integrated pest management protocols. *Annals of Agricultural and Environmental Medicine* **23**, 129–137.

20. **Ansari H et al.** (2015) Predicting CCHF incidence and its related factors using time-series analysis in the southeast of Iran: comparison of SARIMA and Markov switching models. *Epidemiology and Infection* **143**, 839–850.

21. **Allard R** (1998) Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization* **76**, 327–333.

22. **Wei WD et al.** (2016) Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng County, China. *PLoS ONE* **11**, e0156768. doi: 10.1371/journal.pone.0156768.

23. **Zhang H et al.** (2017) Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China. *Journal of the Air & Waste Management Association* **67**, 776–788.

24. **Ljung GM and Box GEP** (1978) Measure of lack of fit in time-series models. *Biometrika* **65**, 297–303.

25. **Anwar MY et al.** (2016) Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malaria Journal* **15**, 566. doi: 10.1186/s12936-016-1602-1.

26. **Wang KW et al.** (2017) Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiology and Infection* **145**, 1118–1129.

27. **Liu L et al.** (2016) Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model. *Epidemiology and Infection* **144**, 144–151.

28. **Bi P et al.** (1998) Seasonal rainfall variability, the incidence of hemorrhagic fever with renal syndrome, and prediction of the disease in low-lying areas of China. *American Journal of Epidemiology* **148**, 276–281.

29. **Khalil H et al.** (2014) The importance of bank vole density and rainy winters in predicting nephropathia epidemica incidence in Northern Sweden. *PLoS ONE* **9**, e111663. doi: 10.1371/journal.pone.0111663.

30. **Zhang S et al.** (2014) Epidemic characteristics of hemorrhagic fever with renal syndrome in China, 2006–2012. *BMC Infectious Diseases* **14**, 384. doi: 10.1186/1471-2334-14-384.

31. **Luo ZZ** (2002) Progress of epidemiology and vaccine research of epidemic hemorrhagic fever. *Chinese Journal of Disease Control & Prevention* **6**, 5–8.

32. **Liu W et al.** (2000) Safety and immunogenicity of inactivated bivalent EHF vaccine in humans. *Chinese Journal of Epidemiology* **21**, 445–447.

33. **Chen HX et al.** (2000) Preventive effects of three kinds of inactive vaccines against epidemic hemorrhagic fever (EHF) after 5 years of vaccination. *Chinese Journal of Epidemiology* **21**, 347–348.

34. **Liu XD et al.** (2011) Temporal trend and climate factors of hemorrhagic fever with renal syndrome epidemic in Shenyang City, China. *BMC Infectious Diseases* **11**, 331. doi: 10.1186/1471-2334-11-331.

35. **Zhang WY et al.** (2010) Climate variability and hemorrhagic fever with renal syndrome transmission in Northeastern China. *Environmental Health Perspectives* **118**, 915–920.

36. **Luo L et al.** (2017) Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Services Research* **17**, 469. doi: 10.1186/s12913-017-2407-9.

37. **Ke GB et al.** (2016) Epidemiological analysis of hemorrhagic fever with renal syndrome in China with the seasonal-trend decomposition method and the exponential smoothing model. *Scientific Reports* **6**, 39350. doi: 10.1038/srep39350.

38. **Zhang YH et al.** (2014) The epidemic characteristics and changing trend of hemorrhagic fever with renal syndrome in Hubei Province, China. *PLoS ONE* **9**, e92700. doi: 10.1371/journal.pone.0092700.