

RESEARCH ARTICLE

Open Access



# Prevalence of hemorrhagic fever with renal syndrome in Yiyuan County, China, 2005–2014

Tao Wang<sup>†</sup>, Jie Liu<sup>†</sup>, Yunping Zhou<sup>†</sup>, Feng Cui, Zhenshui Huang, Ling Wang<sup>\*</sup> and Shenyong Zhai<sup>\*</sup>

## Abstract

**Background:** Hemorrhagic fever with renal syndrome (HFRS) is highly endemic in mainland China, where human cases account for 90 % of the total global cases. Yiyuan County is one of the most serious affected areas in China. Therefore, there is an urgent need for monitoring and predicting HFRS incidence in Yiyuan to make the control of HFRS more effective.

**Methods:** The study was based on the reported cases of HFRS from the National Notifiable Disease Surveillance System. The demographic and spatial distributions of HFRS in Yiyuan were established. Then we fit autoregressive integrated moving average (ARIMA) models and predict the HFRS epidemic trend.

**Results:** There were 362 cases reported in Yiyuan during the 10-year study period. The human infections in the fall and winter reflected a seasonal characteristic pattern of Hantaan virus (HTNV) transmission. The best model was ARIMA (2, 1, 1) × (0, 1, 1)<sub>12</sub> (AIC value 516.86) with a high validity.

**Conclusion:** The ARIMA model fits the fluctuations in HFRS frequency and it can be used for future forecasting when applied to HFRS prevention and control.

**Keywords:** Hemorrhagic fever with renal syndrome, Autoregressive integrated moving average model, Prevalence

## Background

Hemorrhagic fever with renal syndrome (HFRS), a rodent-borne disease caused by hantaviruses (family Bunyaviridae), is characterized by fever, acute renal dysfunction, and hemorrhage manifestations [1–3]. HFRS was first recognized in northeastern China in 1931 and has been prevalent in many other parts of China since 1955. At present, it is highly endemic in mainland China accounting for 90 % of the total cases reported in the world [4–7]. In response to the spread of HFRS in China, the Chinese Center for Disease Control and Prevention (CDC) established the National Notifiable Disease Surveillance System in 2004, which made the surveillance data for HFRS more accurate and comprehensive. A better understanding of the spatial distribution patterns and social demographic distribution characteristics of HFRS would help to identify

areas and populations at high risk. Early warnings are also essential for controlling or reducing the risk of outbreaks [8], epidemic modeling and forecasting can be essential tools to prevent and control HFRS [9]. In epidemiology, Autoregressive integrated moving average (ARIMA) models have been successfully applied to predict the incidence of infectious diseases, such as HIV [10], influenza [11], malaria incidence [12], and other infectious diseases [13–15].

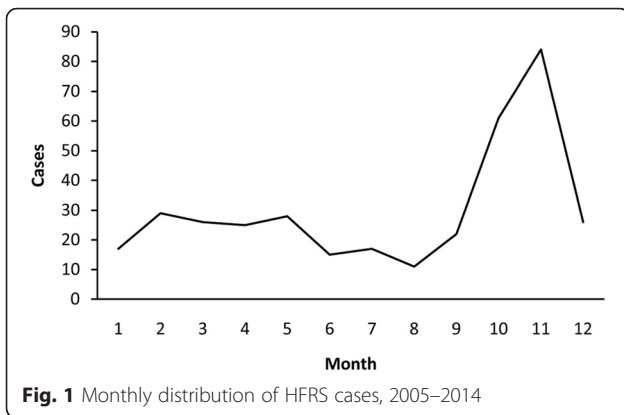
This study aimed to establish the current situation of endemic HFRS in Yiyuan, and characterize its spatio-temporal distribution and demographic distribution characteristics. Furthermore, we fit ARIMA models and predict the HFRS epidemic trend by using SAS version 9.2 (SAS Institute, Cary, NC, USA). Our study was based on HFRS epidemic data from Yiyuan County, China, where it could provide a basis for HFRS prevention and control.

\* Correspondence: lucywl120@sina.com; zibocdc@163.com

<sup>†</sup>Equal contributors

Department of Infectious Disease Control and Prevention, Zibo Center for Disease Control and Prevention, Zibo, Shandong Province, P. R. China





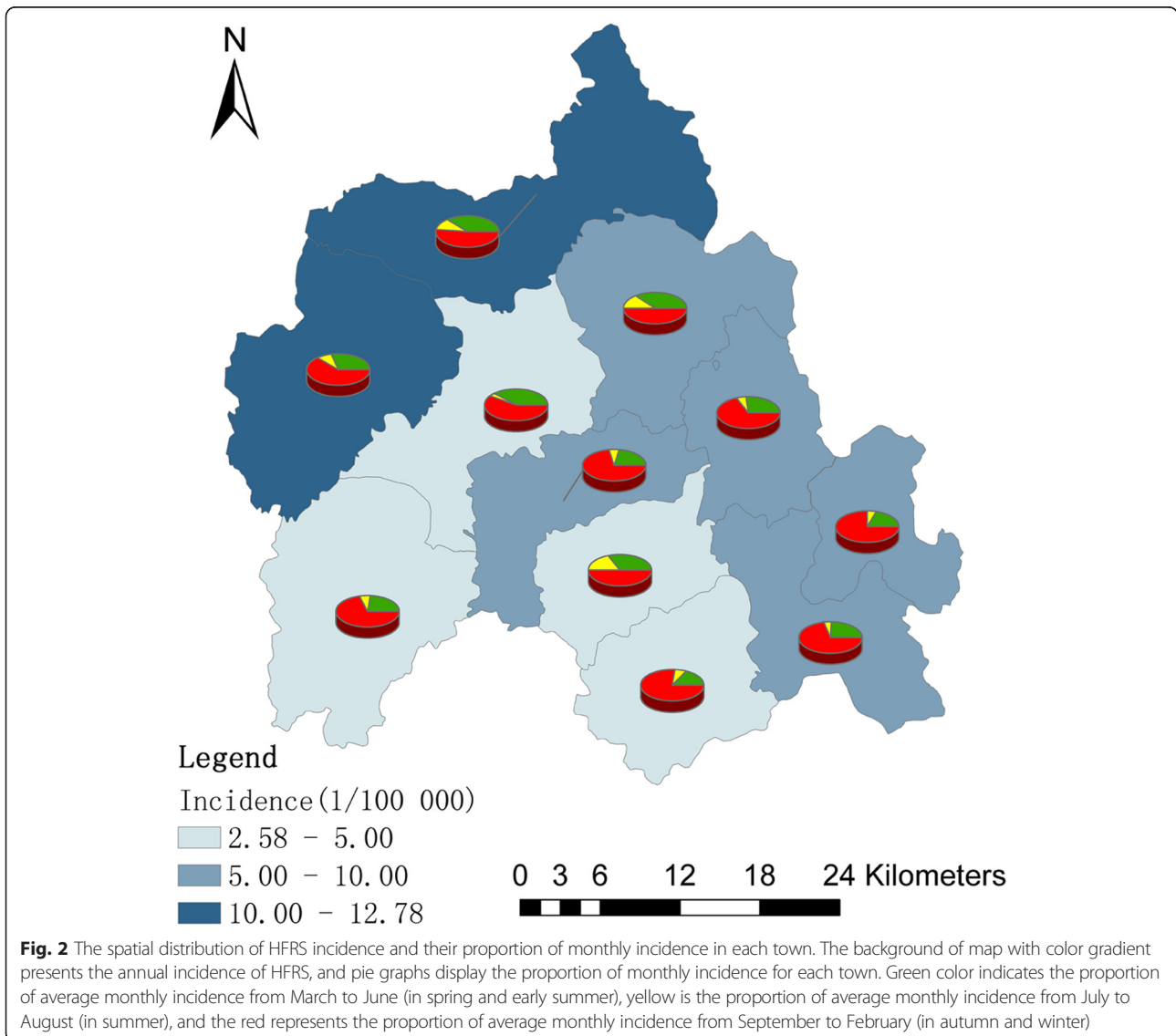
**Methods**

**Study area**

The study site is located in Yiyuan County (latitude 35° 55' ~ 36°23' N and longitude 117°54' ~ 118°31' E), in the central part of Shandong province. Monthly HFERS cases reported during 2005–2014 were provided by Zibo CDC, and we were permitted to use the data. In China, HFERS is a nationally notifiable disease and hospital physicians must report every case of HFERS to the local health authority within 24 h.

**Ethics statement**

The ethical approval was given by Ethics Review Committee of the Zibo Center for Disease Control and Prevention, and the study was conducted in compliance with the principles of the Declaration of Helsinki. Written informed consents for the use of



their clinical samples were obtained from the patients and all analyzed data were anonymized. Besides, consents of participants who were under 16 years have been obtained from their parents/guardians.

**Demographic distribution analysis**

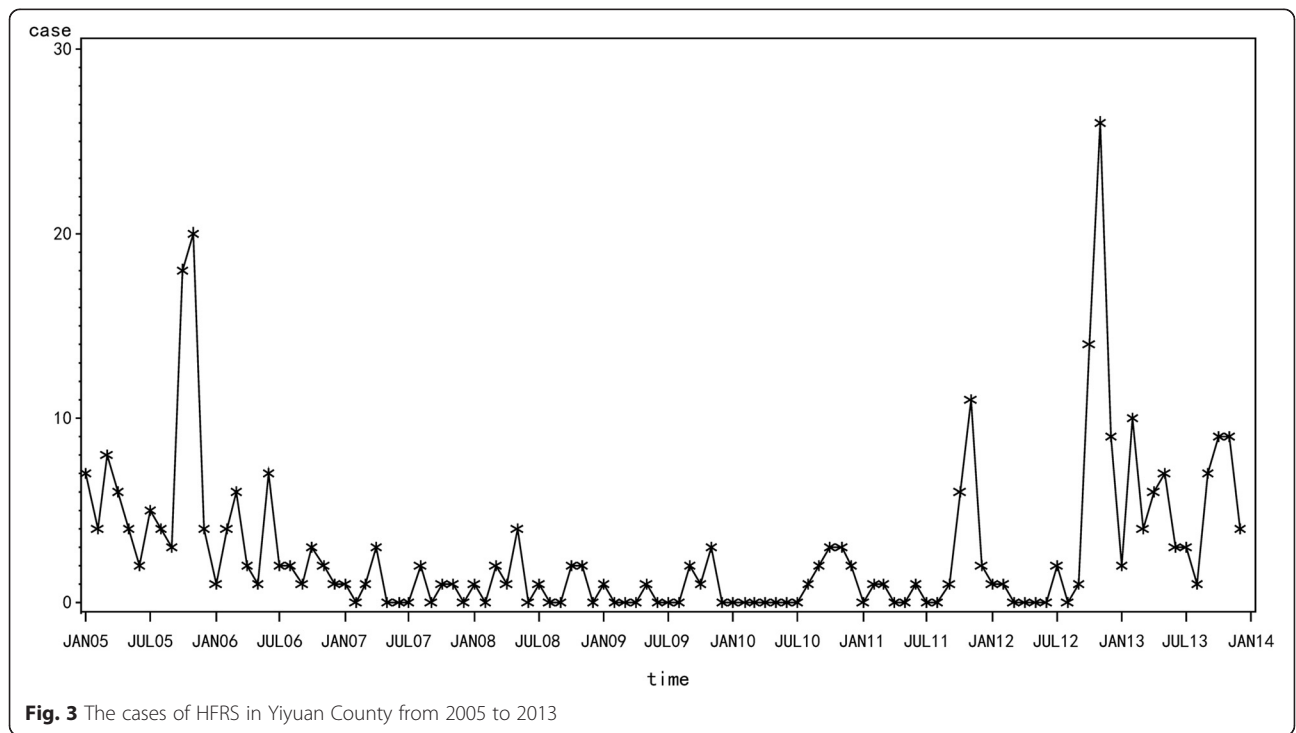
The demographic distribution characteristics including age, sex and occupation distribution of HFRS cases from 2005 to 2014 in Yiyuan County were analyzed according to surveillance data. All HFRS cases were geo-coded and matched to the town-level layers of polygon and point by administrative code using the software ArcGIS9.3 (ESRI Inc., Redlands, CA, USA). To alleviate variations of incidence in small populations and areas, annualized average incidence of HFRS per 100 000 at each town over the 10 year-period were calculated. Furthermore, annualized average incidences and the proportion of monthly average incidence for each town were mapped in gradient colors and pie charts, respectively. To approximately distinguish the dominant hantaviruses, we divided a year into three periods according to the seasonal distribution of the HFRS cases: March to June (in spring and early summer), July to August (in summer), and September to February (in autumn and winter).

**Time-series analysis**

ARIMA models are the most commonly used time series prediction models [10]. We constructed ARIMA models for monthly HFRS incidence in Yiyuan from 2005 to 2014. ARIMA was designed to deal with highly seasonal

data [16]. An ARIMA (p, d, q) model comprises three types of parameters [9, 16, 17]: the autoregressive parameters (p), number of differencing passes (d), and moving average parameters (q). The multiplicative seasonal ARIMA (p, d, q) × (P, D, Q)<sub>s</sub> model is an extension of the ARIMA method to time series in which a pattern repeats seasonally over time [15, 16, 18]. Analogous to the simple ARIMA parameters, the seasonal parameters are: seasonal autoregressive (P), seasonal differencing (D), and seasonal moving average parameters (Q). The length of the seasonal period is represented by s. For example, the incidence of infectious disease varies in the annual cycle, so s = 12 in the present study.

We used the Box-Jenkins strategy to construct models. The ARIMA model procedure consists of three iterative steps [15, 17, 18]: identification, estimation, and diagnostic checking. Prior to fitting the ARIMA model, an appropriate difference of the series is usually performed to make the series stationary. Identification is the process of determining seasonal and non-seasonal orders using the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the transformed data. Parameters in the ARIMA model(s) are estimated with the conditional least squares method after the identification step. At the diagnosis stage, the adequacy of the established model for the series is verified by employing white noise tests to check whether the residuals are independent and normally distributed. It is possible that several ARIMA models may be identified, and the selection of an optimum model is necessary. Such selection of



**Fig. 3** The cases of HFRS in Yiyuan County from 2005 to 2013



**Table 1** Parameter estimation for plausible ARIMA models<sup>a</sup>

Parameter	ARIMA (2, 1, 1) × (0, 1, 0) <sub>12</sub>			ARIMA (2, 1, 0) × (0, 1, 1) <sub>12</sub>			ARIMA (1, 1, 1) × (1, 1, 1) <sub>12</sub>			ARIMA (2, 1, 1) × (0, 1, 1) <sub>12</sub>		
	B	t	P	B	t	P	B	t	P	B	t	P
AR1,1	-0.479	-4.510	<0.001	-0.463	-4.740	<0.001	0.422	3.070	0.003	-0.458	-4.460	<0.001
AR2,1							0.449	1.800	0.075			
MA1,1	0.308	3.020	0.003				0.873	11.100	<0.001	0.275	2.650	0.009
MA2,1				0.375	3.580	<0.001	0.743	3.850	<0.001	0.314	2.910	0.005

<sup>a</sup> AR1,1 = autoregressive parameter; AR2,1 = seasonal autoregressive parameter; MA1,1 = moving average parameter; MA2,1 = seasonal moving average parameter

Fig. 1, which indicated that the occurrence of HFRS presented significant seasonality.

Figure 2 showed annualized average incidence and the proportion of monthly average incidence for each town. Annualized average incidence at the town-level ranged from 2.58 to 12.78 per 100 000. Among the total 11 towns in Yiyuan, 7 towns were medium-endemic with an incidence between 5 and 15 per 100 000, an epidemic peak in the fall-winter season was mapped by the red color in the pies.

**Model identification**

Time series data for HFRS covering 2005–2013 in Yiyuan were used as the training set and monthly data for 2014 were used as the test set (Fig. 3). The sequence diagrams and the seasonal characteristics of HFRS incidence indicated that the data series had a seasonal cycle every 12 months. On the basis of these characteristics, we eliminated the effect of seasonal trends by taking 1-order trend difference and 1-order seasonal difference. The transformed series showed far less dispersion than original series (Fig. 4). Plausible models, i.e., (the ARIMA (2, 1, 1) × (0, 1, 0)<sub>12</sub>, ARIMA (2, 1, 0) × (0, 1, 1)<sub>12</sub>, ARIMA (1, 1, 1) × (1, 1, 1)<sub>12</sub> and ARIMA (2, 1, 1) × (0, 1, 1)<sub>12</sub>), were identified on the basis of autocorrelation functions (ACF) and partial autocorrelation functions (PACF) (Fig. 5), and were used for further analysis.

**Parameter estimation and model testing**

On the basis of parameter estimation and goodness of fit test statistics (Tables 1 and 2), we confirmed that the best model was ARIMA (2, 1, 1) × (0, 1, 1)<sub>12</sub>. The goodness-of-fit analysis showed that there was no significant autocorrelation between residuals at different lags.

**Table 2** Goodness of fit statistics for plausible ARIMA models<sup>a</sup>

Models	AIC	SBC
ARIMA (2, 1, 1) × (0, 1, 0) <sub>12</sub>	519.854	524.962
ARIMA (2, 1, 0) × (0, 1, 1) <sub>12</sub>	520.812	525.920
ARIMA (1, 1, 1) × (1, 1, 1) <sub>12</sub>	522.897	533.113
ARIMA (2, 1, 1) × (0, 1, 1) <sub>12</sub>	516.855	524.517

<sup>a</sup> AIC = Akaike information criterion; SBC = Schwarz Bayesian criterion

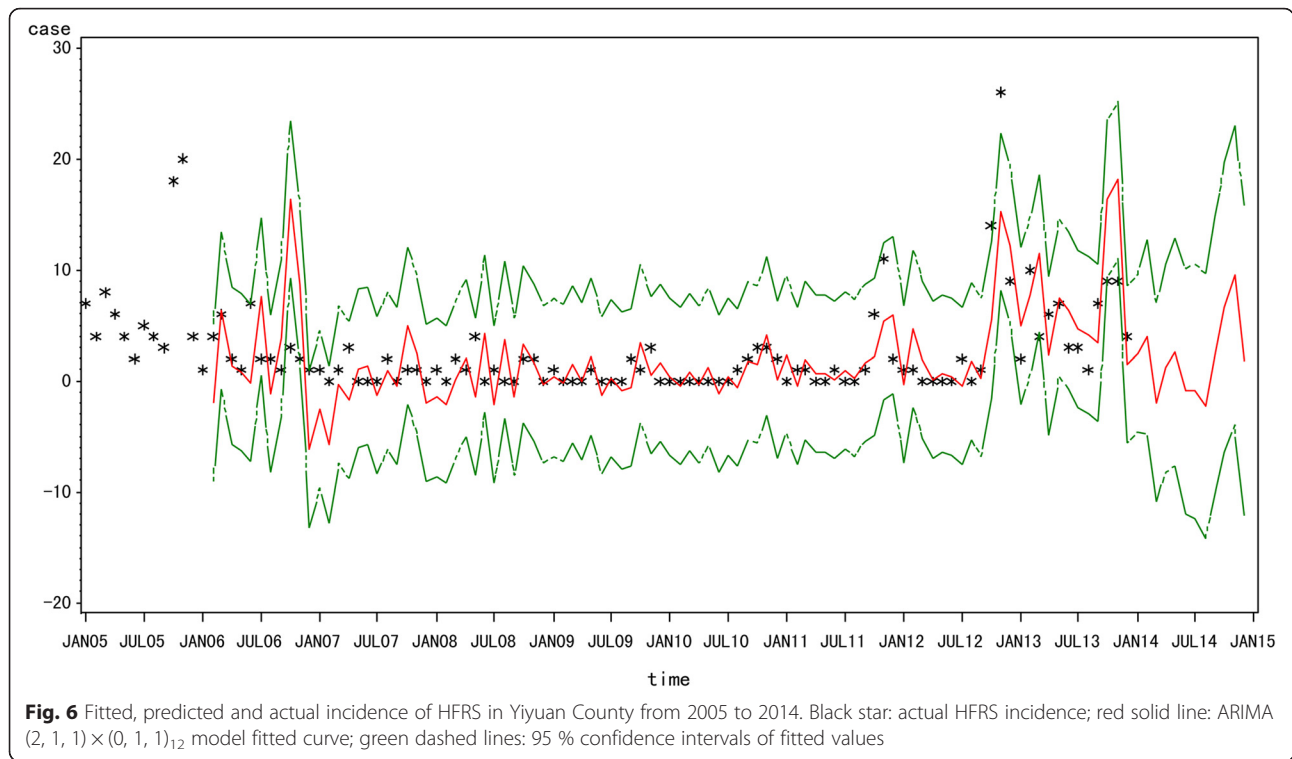
**Forecast analysis**

We used ARIMA model (2, 1, 1) × (0, 1, 1)<sub>12</sub> and time series data for 2005–2013 as the training set, and the data from January–December 2014 were used as the test set (Fig. 6, and Table 3). The predicted data for the actual data and the predicted data 95 % confidence limit for 2014 are shown in Table 3. The RMSE value was 3.56. The predicted data and the actual data were not perfectly matched, but the actual data fell within the predicted 95 % confidence interval.

**Discussion**

In this study, our results demonstrated that the HFRS incidence had been increasing from 2009 to 2014 in Yiyuan County. The proportion of monthly average incidence for each town showed that HFRS in Yiyuan was mainly caused by HTNV. We applied multiplicative seasonal ARIMA (p, d, q) × (P, D, Q)<sub>s</sub> models to analyze the surveillance data of HFRS in Yiyuan, China. According to the results above, the ARIMA (2, 1, 1) × (0, 1, 1)<sub>12</sub> model is reliable with a high validity, which can be used to predict the next one year’s HFRS incidence in Yiyuan.

Yiyuan, are mountainous and hilly with numerous *Apodemus agrarius*, and farmers have more chances to be exposed to contaminated urine and feces of infected rodents. Agricultural activities such as sleeping in the fields, irrigating, and working on the farmland during the autumn harvest season might have played a significant role in the occurrence of HFRS [20]. Previous studies reported that the transmission of HTNV through *Apodemus agrarius* peaked in the winter, while *Rattus norvegicus* associated SEOV infections mainly occurred in the spring [21–24]. Thus, the human infections in Yiyuan in the fall and winter reflect a seasonal characteristic pattern of HTNV transmission. The forecast results suggest that the HFRS incidence in China will experience a slight growth in the next one year. A rise in the number of HFRS incidence may also result from an increase in the number and size of natural foci [25], climate change, especially the increase of mean temperature [20, 23]. Therefore, knowledge of HFRS forecasts is necessary to prompt health departments to strengthen surveillance systems and reallocate resources in anticipation of increasing HFRS incidence.



Epidemiological surveillance of communicable diseases is one of the most traditional health-related activities. Time-series analysis of incidence of various infections is extremely useful in developing hypotheses to explain and anticipate the dynamics of the observed phenomena and subsequently in the establishment of a quality control system and reallocation of resources [26]. There are a number of methods applied for time series analysis including ARIMA model [8–13, 16], maximum entropy method (MEM) spectral analysis [27] and the autoregressive conditional heteroscedastic (ARCH) model [28]. The ARIMA model has its advantages in time-series analyses. The secular trend, seasonal variation, and auto-correlation could all be easily controlled by difference, auto-regression, moving average, and seasonal functions without performing complicated transformations or using extra surrogate variables [18]. Once a satisfactory model has been obtained, it can be used to forecast expected numbers of cases for a given number of future time intervals [29].

Besides, the application of GIS, together with time-series analyses in the present study, provides ways to quantify explicit HFRS and to further identify environmental factors responsible for the increasing disease risk. Although analyses are still preliminary, the findings can be helpful for generating hypothesis for further investigation. For example, based on the prediction results, the government can invest more health resources during high-risk periods and decrease it during low-risk periods to improve the cost-effectiveness of interventions and scheduling of resources. It can also be used to evaluate the effectiveness of public health interventions under varying assumptions by comparing actual HFRS incidence with expected incidence.

However, limitations should also be considered in this present study. The RMSE value was 3.56, and the actual data did not match the predicted data of the model perfectly. Due to a lack of time series data on the population densities of rodents, and the influencing factors, it is difficult to further uncover the probable causes and

**Table 3** Comparison of predicted HFRS cases and actual values for Yiyuan, during January-December 2014<sup>a</sup>

No cases	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actual value	3	9	4	7	11	2	4	1	5	4	7	4
Predictive value	2	4	0	1	3	0	0	0	2	7	9	2
95 % CL of PV	0–10	0–13	0–7	0–11	0–13	0–10	0–11	0–10	0–15	0–20	0–23	0–16

<sup>a</sup> CL = confidence level; PV = predicted value

shifts of the characters of HFRS. Future researches are warranted to focus on the risk factors of HFRS to modify the ARIMA model such as rodent population densities, human activities, farming patterns, various socio-economic and environmental factors in Yiyuan.

## Conclusion

In summary, the results of our study provide useful information on the prevailing epidemiological situation of HFRS in Yiyuan. We further confirmed the consensus that ARIMA model is a useful tool in monitoring and predicting changing trends in HFRS. The ARIMA model could be used to optimize HFRS prevention by providing short-term forecasting on the HFRS incidence. To control and prevent HFRS, a comprehensive preventive strategy including public health education and promotion, rodent control, surveillance, and vaccination should be implemented in Yiyuan.

## Abbreviations

ARIMA: Autoregressive integrated moving average; HFRS: Hemorrhagic fever with renal syndrome; HTNV: Hantaan virus.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

TW extracted the data, conducted the statistical analysis and drafted the manuscript. JL and YPZ conceived of the project concept, helped to interpret the results and modify the manuscript. FC helped to interpret the results. ZSH extracted the data. SYZ and LW conceived of the project concept, assisted with the data interpretation, and helped write the manuscript. All of the authors have read and approved the final manuscript.

## Acknowledgements

This work was supported by Zibo Center for Disease Control and Prevention

Received: 27 January 2015 Accepted: 2 February 2016

Published online: 06 February 2016

## References

- Schmaljohn CS, Dalrymple JM. Analysis of Hantaan virus RNA: evidence for a new genus of bunyaviridae. *Virology*. 1983;131(2):482–91.
- Liu YX, Feng D, Zhang Q, Jia N, Zhao ZT, De Vlas SJ, et al. Key differentiating features between scrub typhus and hemorrhagic fever with renal syndrome in northern China. *Am J Trop Med Hyg*. 2007;76(5):801–5.
- Cui F, Wang T, Wang L, Yang S, Zhang L, Cao H, et al. Spatial analysis of hemorrhagic fever with renal syndrome in Zibo City, China, 2009–2012. *PLoS ONE*. 2013;8(6):e67490.
- Song G. Epidemiological progresses of hemorrhagic fever with renal syndrome in China. *Chin Med J (Engl)*. 1999;112(5):472–7.
- Yan L, Fang LQ, Huang HG, Zhang LQ, Feng D, Zhao WJ, et al. Landscape elements and Hantaan virus-related hemorrhagic fever with renal syndrome, People's Republic of China. *Emerg Infect Dis*. 2007;13(9):1301–6.
- Zuo SQ, Fang LQ, Zhan L, Zhang PH, Jiang JF, Wang LP, et al. Geo-spatial hotspots of hemorrhagic fever with renal syndrome and genetic characterization of Seoul variants in Beijing. *China PLoS Negl Trop Dis*. 2011;5(1), e945.
- Simmons JH, Riley LK. Hantaviruses: an overview. *Comp Med*. 2002;52(2):97–110.
- Li Q, Guo NN, Han ZY, Zhang YB, Qi SX, Xu YG, et al. Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. *Am J Trop Med Hyg*. 2012;87(2):364–70.
- Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infect Dis*. 2011;11:218.
- Yu HK, Kim NY, Kim SS, Chu C, Kee MK. Forecasting the number of human immunodeficiency virus infections in the Korean population using the autoregressive integrated moving average model. *Osong Public Health Res Perspect*. 2013;4(6):358–62.
- Reichert TA, Simonsen L, Sharma A, Pardo SA, Fedson DS, Miller MA. Influenza and the winter increase in mortality in the United States, 1959–1999. *Am J Epidemiol*. 2004;160(5):492–502.
- Gaudart J, Toure O, Dessay N, Dicko AL, Ranque S, Forest L, et al. Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area. *Mali Malar J*. 2009;8:61.
- Luz PM, Mendes BV, Codeco CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro. *Brazil Am J Trop Med Hyg*. 2008;79(6):933–9.
- Yi J, Du CT, Wang RH, Liu L. [Applications of multiple seasonal autoregressive integrated moving average (ARIMA) model on predictive incidence of tuberculosis]. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2007;41(2):118–21.
- Zhang X, Zhang T, Young AA, Li X. Applications and comparisons of four time series models in epidemiological surveillance data. *PLoS ONE*. 2014;9(2), e88075.
- Lin H, Lu L, Tian L, Zhou S, Wu H, Bi Y, et al. Spatial and temporal distribution of falciparum malaria in China. *Malar J*. 2009;8:130.
- Sato RC. Disease management with ARIMA model in time series. *Einstein (Sao Paulo)*. 2013;11(1):128–31.
- Liu X, Jiang B, Bi P, Yang W, Liu Q. Prevalence of haemorrhagic fever with renal syndrome in mainland China: analysis of National Surveillance Data, 2004–2009. *Epidemiol Infect*. 2012;140(5):851–7.
- Chatfield C. *The analysis of time series: theory and practice*. London: Chapman and Hall; 1975.
- Bi P, Wu X, Zhang F, Parton KA, Tong S. Seasonal rainfall variability, the incidence of hemorrhagic fever with renal syndrome, and prediction of the disease in low-lying areas of China. *Am J Epidemiol*. 1998;148(3):276–81.
- Zhang YZ, Zhang FX, Wang JB, Zhao ZW, Li MH, Chen HX, et al. Hantaviruses in rodents and humans, Inner Mongolia Autonomous Region, China. *Emerg Infect Dis*. 2009;15(6):885–91.
- Kim YS, Ahn C, Han JS, Kim S, Lee JS, Lee PW. Hemorrhagic fever with renal syndrome caused by the Seoul virus. *Nephron*. 1995;71(4):419–27.
- Fang LQ, Wang XJ, Liang S, Li YL, Song SX, Zhang WY, et al. Spatiotemporal trends and climatic factors of hemorrhagic fever with renal syndrome epidemic in Shandong Province, China. *PLoS Negl Trop Dis*. 2010;4(8):e789.
- Chen HX, Qiu FX, Dong BJ, Ji SZ, Li YT, Wang Y, et al. Epidemiological studies on hemorrhagic fever with renal syndrome in China. *J Infect Dis*. 1986;154(3):394–8.
- Li Q, Zhao W, Wei Y, Han X, Han Z, Zhang Y, et al. Analysis of incidence and related factors of hemorrhagic fever with renal syndrome in Hebei province, China. *PLoS ONE*. 2014;9(7):e101348.
- Kuhn L, Davidson LL, Durkin MS. Use of Poisson regression and time series analysis for detecting changes over time in rates of child injury following a prevention program. *Am J Epidemiol*. 1994;140(10):943–55.
- Sumi A, Kamo K, Ohtomo N, Mise K, Kobayashi N. Time series analysis of incidence data of influenza in Japan. *J Epidemiol*. 2011;21(1):21–9.
- Techie Quaicoe M, Twenefour FB, Baah EM, Nortey EN. Modeling variations in the cedi/dollar exchange rate in Ghana: an autoregressive conditional heteroscedastic (ARCH) models. *Springerplus*. 2015;4:329.
- Allard R. Use of time-series analysis in infectious disease surveillance. *Bull World Health Organ*. 1998;76(4):327–33.