Predicting the number of visceral leishmaniasis cases in Kashgar, Xinjiang, China using the ARIMA−EGARCH model

Hu−ling Li[1], Rong−jiong Zheng[2], Qiang Zheng[3], Wei Jiang[3], Xue−liang Zhang[4], Wei−ming Wang[5], Xing Feng[1], Kai Wang[4✉], Xiao−bo Lu[2✉]

[1]College of Public Health, Xinjiang Medical University, Urumqi 830011, P.R. China

[2]Department of Infectious Diseases, the first Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, P. R. China

[3]Xinjiang Center for Disease Control and Prevention, Urumqi, 830002, P. R. China

[4]Department of Medical Engineering and Technology, Xinjiang Medical University, Urumqi 830011, P.R. China

[5]School of Mathematics Science, Huaiyin Normal University, Huaian 223300 P. R. China

ABSTRACT

Objective: To forecast the visceral leishmaniasis cases using auto-regress integrated moving average (ARIMA) and hybrid ARIMA-EGARCH model, which offers a scientific basis to control visceral leishmaniasis spread in Kashgar Prefecture of Xinjiang, China.

Methods: The data used in this paper are monthly visceral leishmaniasis cases in the Kashgar Prefecture of Xinjiang from 2004 to 2016. The sample data between 2004 and 2015 were used for the estimation to choose the best model and the sample data in 2016 were used for the forecast. Time series of visceral leishmaniasis started on 1 January 2004 and ended on 31 December 2016, consisting of 1 790 observations reported in Kashgar Prefecture.

Results: For Xinjiang, the total number of reported cases were 2 187, the male-to-female ratio of cases was 1:1.42. Patients aged between 0 and 10 years accounted for 82.72% of all reported cases and the largest percentage of visceral leishmaniasis cases was detected among scattered children who accounted for 68.82%. The monthly incidences fitted by ARIMA (2, 1, 2) (1, 1, 1)$_{12}$ model were consistent with the real data collected from 2004 to 2015. However, the predicted cases failed to comply with the observed case number; we then attempted to establish a hybrid ARIMA-EGARCH model to fit visceral leishmaniasis. Finally, the ARIMA (2, 1, 2) (1, 1, 1)$_{12}$-EGARCH (1, 1) model showed a good estimation when dealing with volatility clustering in the data series.

Conclusions: The combined model has been determined as the best prediction model with the root-mean-square error (RMSE) of 7.23% in the validation phase, which means that this model has high validity and rationality and can be used for short-term prediction of visceral leishmaniasis and could be applied to the prevention and control of the disease.

1. Introduction

Leishmaniasis or black fever, which also known as kala-azar is a chronic infectious disease which can be divided into visceral leishmaniasis and cutaneous leishmaniasis[1–4]. Visceral leishmaniasis is a vector borne tropical infection provoked by protozoans which belongs to the genus *Leishmania*. The transmission vector of the disease is the sandfly (Phlebotomus and *Lutzomyia* species). The syndromes of visceral leishmaniasis are fever, weight loss, splenomegaly, hepatomegaly, skin darkening and anemia[5]. It is estimated about 50 000-90 000 new cases of visceral leishmaniasis occur worldwide annually[6], 90% of which

are mainly from 6 countries including India, Bangladesh, Sudan, South Sudan, Brazil and Ethiopia[7]. In terms of global estimation, the number of annual visceral leishmaniasis cases in Somalia is on the top two among the six countries, after South Asia[8,9] and South Asia is considered as the highest visceral leishmaniasis prone region in the world[10].

In China, visceral leishmaniasis is prevalent in Xinjiang, Gansu, Sichuan and some Midwest regions(autonomous regions). The main transmission types in Xinjiang are wildlife animals type and human type. The type of dogs usually occurs in Gansu, Sichuan and other endemic areas. China has reported a total of 3 994 visceral leishmaniasis cases and cases are distributed in 27 provinces and 336 counties nationwide from 2005 to 2015. The top three reported provinces are Xinjiang, Gansu, and Sichuan, and the reported cases account for 48.00%, 33.12%, and 14.17% of the total reported cases, respectively. The total number of cases account for 95.29% of the total number of cases has been reported in the country[11]. Xinjiang is a key epidemic area of visceral leishmaniasis in China, visceral leishmaniasis cases had increased from 377 in 1980-1990 to 1 588 in 2004-2014. At the end of 2014, 1 272 cases were reported in Kashgar for 11 years which accounted for 80.10% of the total reported cases in Xinjiang[12,13]. The Kashgar region in Xinjiang has become the area with the highest number of black fever patients and the highest incidence rate in the country. Among them, Kashi City, Shufu County and Jiashi County are main endemic areas of visceral leishmaniasis in China, and the death caused by black fever is continuing[14−16]. This is why we choose Kashgar Prefecture of Xinjiang as the research area.

In recent years, there are many studies using various models for the modeling and forecasting of diseases. Among them, time series models play an important role in disease prediction. For instance, ARIMA model in time series analysis has been employed to forecast visceral leishmaniasis[17−19]. The three studies on visceral leishmaniasis, all of whom have an identical point. The relationship between meteorological factors or influencing factors and disease has been taken into consideration in their models. However, at present, studies predicting the trend of Visceral leishmaniasis incidence with the consideration of time series models in Xinjiang are scarce. Hence, this study aims at adopting time series analysis to simulate visceral leishmaniasis cases in Kashgar Prefecture of Xinjiang and offering insights on visceral leishmaniasis and measuring the occurrence of visceral leishmaniasis by using multiplicative seasonal ARIMA model, Exponential General Autoregressive Conditional Heteroscedasticity (EGARCH) technique. This paper aims to use the hybrid ARIMA-EGARCH model for modeling and forecasting visceral leishmaniasis in Xinjiang which has not been reported yet to the best of our knowledge.

2. Materials and methods

2.1. Study area and data source

Kashgar Prefecture is located in the southwestern of Xinjiang Uygur Autonomous Region between the latitude of 35°28' to 40°16′N and the longitude of 71°39′ to 79°52′E with the area of about 162 000 square kilometers (Figure 1). The monthly visceral leishmaniasis cases from 2004 to 2016 in Kashgar Prefecture were retrieved from the Xinjiang Uygur Autonomous Region Centers for Disease Control and Prevention (CDC), age, sex, occupation and other demographic information were also collected and anlyzed. The visceral leishmaniasis cases were divided into development group (January 1, 2004, through December 31, 2015) and validation group (January 1, 2016, through December 31, 2016) for model fitting testing and empirical study, respectively.
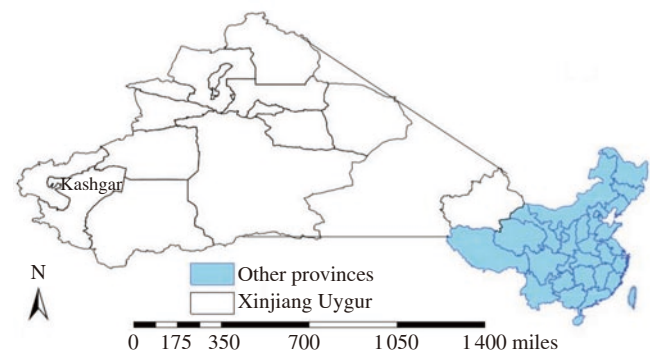

Figure 1. The geoloction of Kashgar Prefecture in Xinjiang, China (map is created by ArcGIS software, v.10.4.1).
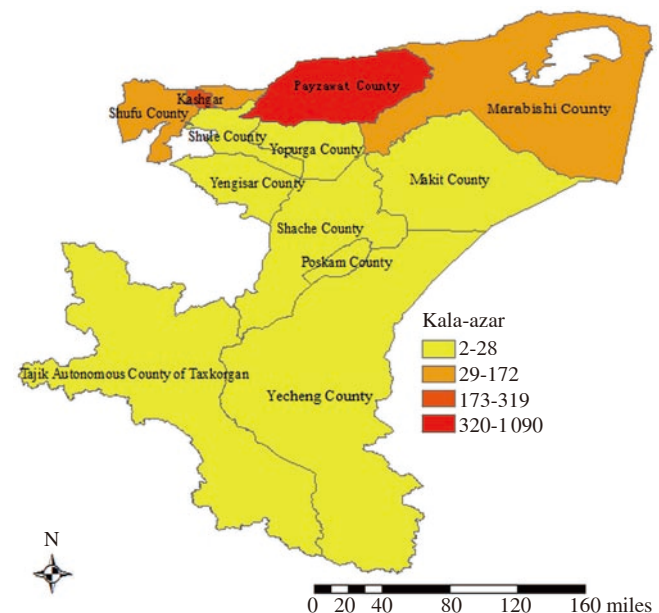

Figure 2. Geographical distribution of visceral leishmaniasis cases in Kashgar Prefecture from 2004 to 2016.

2.2. Descriptive data

The data covers 156 months, starting from January 2004 to December 2016 in Xinjiang Uygur Autonomous Region. The total number of visceral leishmaniasis cases reported throughout the period is 2 187, among which, cases in Kashgar Prefecture accounted for 81.85% (1 790/2 187). A total of 58.62% were male and 41.38% were female, with the sex ratio was 1.42:1 (1 282/905); the high-prevalence group occurred among those aged 0-10 years accounting for 82.72% (1 809/2 187) of all reported cases. Cases reported in people aged 0-75 years decreased gradually with age, While the incidence of children aged 0-2 years was the highest, accounting for 77.34% (1 399/1 809) of the incidence of children under 10 years old and 63.97% (1 399/2 187) of the total incidences. With regard to occupation, the highest percentage of visceral leishmaniasis cases were detected in scattered children accounting for 68.82% (1 505/2 187), followed by farmers (11.52%, 1 505/2 187), students (10.11%, 1 221/2 187) and kindergarten children (4.71%, 103/ 2 187).

In Kashgar Prefecture, Serious endemic districts were mainly Payzawat County (60.89%, 1 090/1 790), Kashi City (17.82%, 319/1 709), Shufu County(9.61%, 172/1 709) and Marabishi County (4.86%, 87/1 709) (Figure 2). As for the total number of reported cases in the remaining counties was fewer than 30 cases. In 2008-2009 and 2014-2016, an outbreak of visceral leishmaniasis occurred in Payzawat County, Kashgar Prefecture, which resulted in a sharp increase in the number of people suffering from visceral leishmaniasis, and at the same time, the number of patients peaked in Xinjiang.

2.3. Time series analysis

2.3.1. ARIMA model

Mostly, we applied the descriptive epidemiology approach to depict the epidemic distribution of visceral leishmaniasis, comprising the spatial distribution and demographic characteristics (sex ratio, age group, and occupation). The ARIMA model originally was conceived for economics applications, then this model has been widely used in the infectious disease for an amount of different time varying events,which is the most common used time series prediction model[20]. An ARIMA (p, d, q) model is determined by three types of parameters (p, d, q): p is the number of autoregressive (AR) terms, d is the number of times the model differenced, and q is the number of moving average (MA) terms. The ARIMA model usually is termed as multiplicative seasonal ARIMA (p, d, q) (P, D, Q) S , which is an extension of the ARIMA model to time series. In the expression, the seasonal parameters area: P is the seasonal order of autoregressive; D is the order of seasonal differencing; Q is the seasonal order moving average; and S is the length of the seasonal period, defined as 12[21,22]. To get an optimal ARIMA model, the procedure comprises three iterative steps[23–25]: identification, estimation, and diagnostic checking. Prior to fitting the ARIMA model, Augmented Dickey-Fuller (ADF) method has been adopted to identify whether the series is stationary. Appropriate difference could transform the series into a stationary one. Identification is the process of determining seasonal and non-seasonal orders of a model using the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the transformed data. The maximum likelihood estimation method has been used for parameters estimation after the identification step. At the diagnosis phase, 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. The best fitting model has been determined by comparing values of Akaike Information Criterion (AIC) and Schwartz Bayesian Criterion (SBC). Lower AIC and SBC values can demonstrate better model from models developed by different lags. In addition, we have used the root-mean-square error (RMSE) to evaluate the general performance for each model, the calculation of RMSE between actual cases and predicted cases utilize the formula:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y-\hat{Y}_t)^2}$$

where is observed values and   is predicted values at time t and N is the number of observations. Finally, the fitted ARIMA model applied to predict the monthly visceral leishmaniasis incidences between January and December 2016.

2.3.2. GARCH family models

In the above model, the variance of innovations (often referred to as volatility) is constant over time (homogeneity of variance). This often proves to be too restrictive of an assumption for real data. Under such an assumption, features like volatility clustering cannot be modeled. In order to solve this question, the Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models developed by Engle and extended by Baillie and Nelson have been proven to be a useful tool to capture the momentum in conditional variance. The GARCH models are commonly used in modeling financial time series that exhibit time-varying volatility clustering. GARCH models attempt to address volatility clustering in an innovations process. Volatility clustering occurs when an innovations process does not exhibit significant auto-correlation, but the variance of the process changes with time. If a series exhibits volatility clustering, it suggests that past variances might be predictive of the current variance. GARCH family models including Integrated Generalized Autoregressive Conditional Heteroscedasticity (IGARCH) is a restricted version of the GARCH model; A model which accepts the asymmetric effect of the news is the exponential GARCH model (EGARCH); A TGARCH (p, q) model proposed by Glosten *et al.* which can also handle leverage effect[26–31]. In recent years, more hybrid forecasting models have been proposed including

an ARIMA model with GARCH to predict time series data in various fields for their good performances. This article used three different criteria, named Root Mean Squared Error (RMSE), AIC and SBC to compare the performance efficiency of the ARIMA and ARIMA-GARCH family models in the forecasting of visceral leishmaniasis. The model with a smaller amount would be considered as a better and more appropriate model. All analyses, modelings and data visualizations were programmed in *R* (version 3.4.1) and ArcGIS (version 10.4.1).

3. Results

3.1. Multiplicative seasonal ARIMA model

3.1.1. Data pre-processing

The trend of monthly visceral leishmaniasis time series data from January 2004 to December 2015 can be seen in Figure 3A. The sequence identified by ADF test, which was stationary ($P=0.022$). The seasonal characteristics of visceral leishmaniasis cases revealed

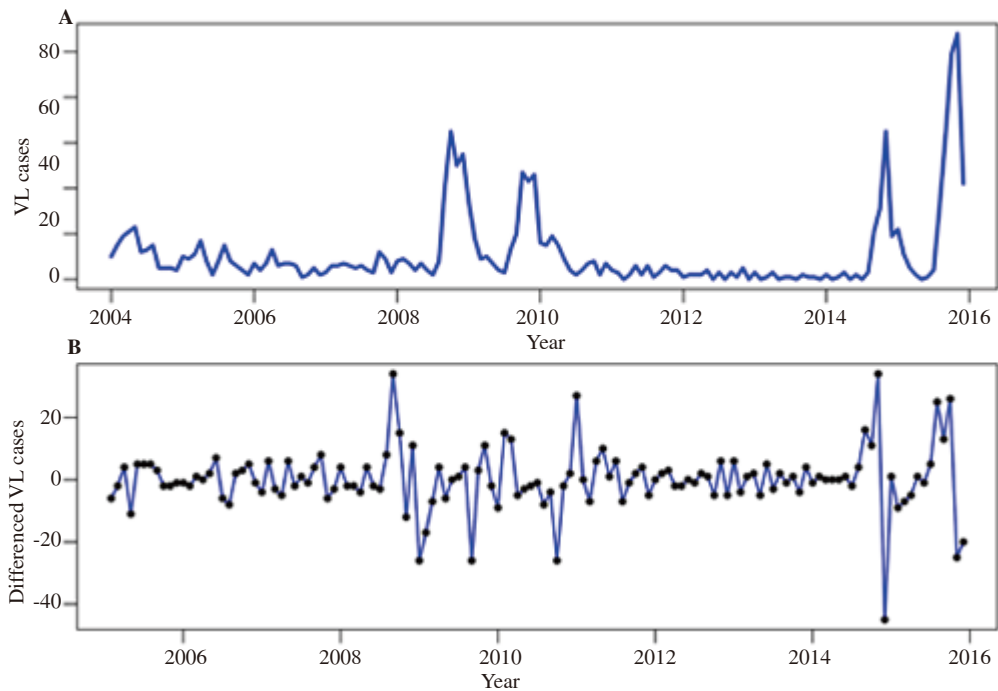**Figure 3.** (A) The monthly visceral leishmaniasis (VL) cases from 2004 to 2015 in Kashgar Prefecture. (B) Monthly visceral leishmaniasis cases after the seasonal difference and 1-order difference in Kashgar Prefecture.
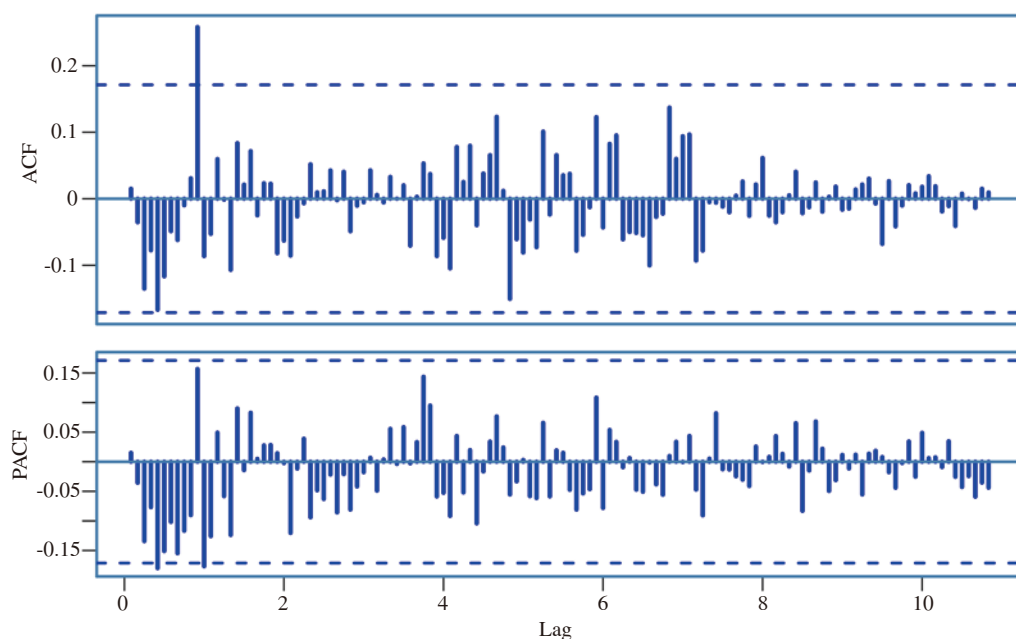

Figure 4. Autocorrelation and partial autocorrelation functions calculated using the seasonal differenced and 1-order differenced number of monthly visceral leishmaniasis cases from 2004 to 2015.

that the data series had a seasonal cycle per 12 months. Looking at the seasonal distribution of cases over the years, the peak of visceral leishmaniasis cases was between September and November, most diseases burst out in the autumn. To eliminate the effect of seasonal trends, we took the first difference and 1-order seasonal difference. The series was stationary which can be seen from Figure 3B and Augmented Dickey-fuller (ADF) test (*P*=0.01) the series has been.

3.1.2. Model identification

First, the orders were examined for elected ARIMA (p, d, q) (P, D, Q)$_{12}$. The Auto-correlation Function (ACF) and Partial Autocorrelation Function (PACF) were regarded as an indication of order (Figure 4). Although the ACF and PACF implement an effective tool for identifying pure AR (p) or MA (q) models, however, the theoretical ACFs and PACFs have an infinite number of non-zero values for the hybrid AR-MA model, using ACF and PACF to identify the hybrid model is extremely difficult. To facilitate the determination of ARIMA model order, it could be determined that the difference order d=1, D=1; the initial model was ARIMA(p, 1, q) (P, 1, Q)$_{12}$; the remaining parameters took 0, 1, 2. The different combinations of orders leaded to 81 different ARIMA model for further analysis.

3.1.3. Parameter estimation and model diagnostic

Table 1 showed the parameter estimation and testing results of the multiplicative seasonal ARIMA models which had passed parameter verification. There were three plausible models for ARIMA in this study, including ARIMA (1, 1, 0) (2, 1, 0)$_{12}$, ARIMA (2, 1, 1) (1, 1, 1)$_{12}$ and ARIMA (2, 1, 2) (1, 1, 1)$_{12}$. Table 2 showed the ARIMA (2, 1, 2) (1, 1, 1)$_{12}$ which was considered as the most suitable one for all

tested models because of the smaller AIC, AICC and SBC (966.46, 967.37 and 986.58). RMSE (8.62%) implied that the selected model was workable. The white noise diagnostic check (χ^2=0.002, *P*=0.965) of the residuals of the best model revealed that the ARIMA (2, 1, 2) (1, 1, 1)$_{12}$ model had got whole information from the visceral leishmaniasis times series data in this paper. So the residual series was a white noise.

Table 1. Parameter estimation for plausible ARIMA models.

Models	Estimate	Standard error	t	P
ARIMA (1, 1, 0) (2, 1, 0)$_{12}$				
AR (1)	0.033	0.016	2.063	0.020
SAR (1)	-0.132	0.021	-6.286	<0.001
SAR (2)	-0.105	0.008	-13.125	<0.001
ARIMA (2, 1, 1) (1, 1, 1)$_{12}$				
AR (1)	0.904	0.108	8.370	<0.001
AR (2)	-0.175	0.095	-1.842	0.034
MA (1)	-0.945	0.074	-12.770	<0.001
SAR (1)	0.638	0.186	3.430	<0.001
SMA (1)	-0.831	0.166	-5.006	<0.001
ARIMA (2, 1, 2) (1, 1, 1)$_{12}$				
AR (1)	1.687	0.105	16.067	<0.001
AR (2)	-0.850	0.065	-13.077	<0.001
MA (1)	-1.798	0.169	-10.639	<0.001
MA (2)	0.871	0.160	5.444	<0.001
SAR (1)	0.527	0.190	2.774	0.003
SMA (1)	-0.795	0.169	-4.704	<0.001

AR: Autogressive; MA: Moving average.

Table 2. Goodness of fits for plausible ARIMA models.

Models	AIC	AICC	SBC	RMSE (%)
ARIMA (1, 1, 0) (2, 1, 0)$_{12}$	982.08	982.39	993.58	9.49
ARIMA (2, 1, 1) (1, 1, 1)$_{12}$	971.83	972.51	989.09	8.89
ARIMA (2, 1, 2) (1, 1, 1)$_{12}$	966.46	967.37	986.58	8.62

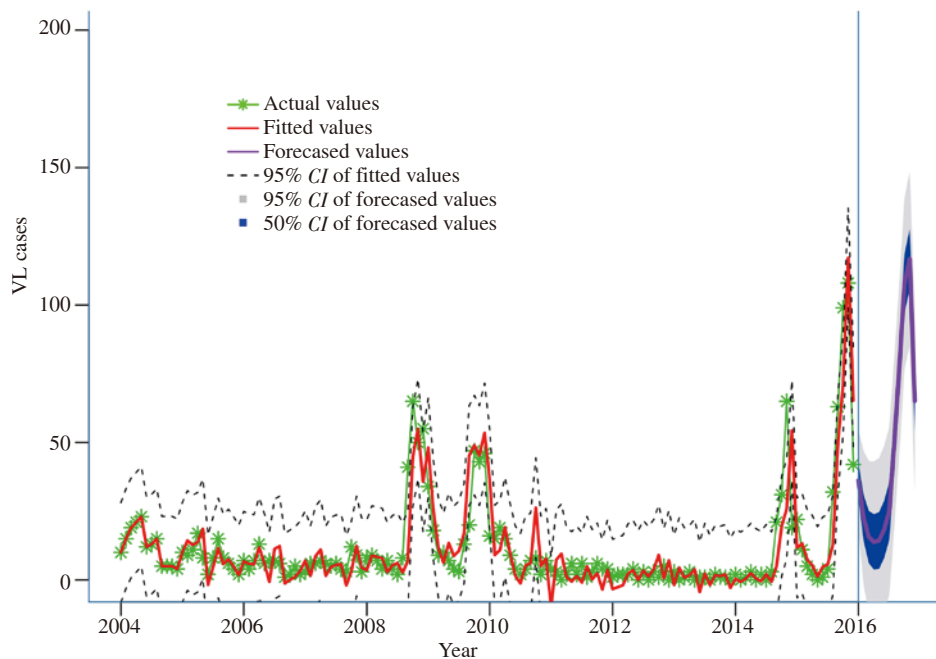**Figure 5.** Fitted, predicted versus actual visceral leishmaniasis cases between January 2004 and December 2016 in Kashgar Prefecture. Green star: Actual values; red solid line: ARIMA (2, 1, 2) (1, 1, 1)$_{12}$ model fitted curve; black dashed line: 95% confidence intervals of fitted values; purple solid line: Forecasted values; the parts of gray and blue are 95% *CI* and 50% *CI* of forecasted values, respectively.

3.1.4. Forecast analysis of model

The ARIMA $(2, 1, 2) (1, 1, 1)_{12}$ model was used to simulate visceral leishmaniasis incidence (January 2004 to December 2015) and forecast visceral leishmaniasis incidence (January 2016 to December 2016). The fitted and real amounts of the disease cases have been presented in Figure 5. The prediction results of the ARIMA model $(2, 1, 2) (1, 1, 1)_{12}$ with 95% confidence interval were also depicted in Figure 5. As shown in the Figure, the pattern of the fitted monthly values was very close to the actual visceral leishmaniasis cases. The achieved results of Table 3 in the validation phases (January 2016 to December 2016) predicted that the visceral leishmaniasis cases were similar to the actual cases in the first half of the year and the actual values lied in the predicted 95% confidence interval (*CI*). However, in the next six months, the forecasted cases were not consistent with the observed visceral leishmaniasis, the ARIMA model presented an RMSE of 48.53%. Unfortunately, almost all observed number of visceral leishmaniasis cases were not within the 95% confidence interval of the predicted value, which proved that the multiplicative seasonal ARIMA model failed to accurately predict the development trend of cases the number of visceral leishmaniasis cases in the future. Therefore, in order to boost the predicted precision of ARIMA model, we introduced the ARIMA-EGARCH model in the next section.

3.2. Hybrid ARIMA−EGARCH model

In this section, to enhance the precision of ARIMA $(2, 1, 2) (1, 1, 1)_{12}$ model, we analyzed the residual of the model. ARIMA combined with different models in the GARCH family models. We chose the ARIMA-EGARCH model as the hybrid model.

3.2.1. Testing for ARCH effects

From the previous introduction, since the ARIMA model was established on the assumption that the data was homoscedastic. In order to take the changes in variance for the visceral leishmaniasis data into consideration, we needed to model the heteroscedastic behavior of the data series, which was an ARCH effect. Therefore, we needed to test the auto-correlation of the square of the residual sequence. Here, we used the McLeod-Li test to obtain p values that were much smaller than the significance level (significant level

$P<0.05$), so the null hypothesis was rejected. It is considered that there was a strong correlation between the squared terms of the residuals, which was, the sequence had conditional heteroskedasticity (Figure 6).

3.2.2. Create hybrid ARIMA−EGARCH model

Since the model had conditional heteroskedasticity, the volatility equation was established for the residual of the model. We chose the EGARCH model to explain the conditional heteroskedasticity. The aim of the general EGARCH model was (1, 1) to explain most models, hence, we built the ARIMA$(2, 1, 2) (1, 1, 1)_{12}$-EGARCH (1, 1) model to boost the precision of forecast. By the model diagnostic, we found the correlation of the square of standardized residual of the combined model was weak correlation (Figure 7A, 7B), and there was no ARCH effect longer. It is considered that the combined model was workable. Table 4 presented information criterion, in the fitted phase, AIC, SBC and RMSE for the ARIMA $(2, 1, 2) (1, 1, 1)_{12}$ model were 966.46, 986.58 and 8.62%, respectively. However, the values of AIC, SBC and RMSE of the ARIMA $(2, 1, 2) (1, 1, 1)_{12}$-EGARCH (1, 1) model were 6.16, 6.39 and 11.07%, respectively. The RMSE values for predicted data of ARIMA $(2, 1, 2) (1, 1, 1)_{12}$ and combined model were 48.53% and 7.23%, respectively. This suggested that the combined model was able to achieve dramatic performance improvement. ARIMA $(2, 1, 2) (1, 1, 1)_{12}$-EGARCH (1, 1) model presented a reasonable result in forecasting visceral leishmaniasis series (Figure 8).
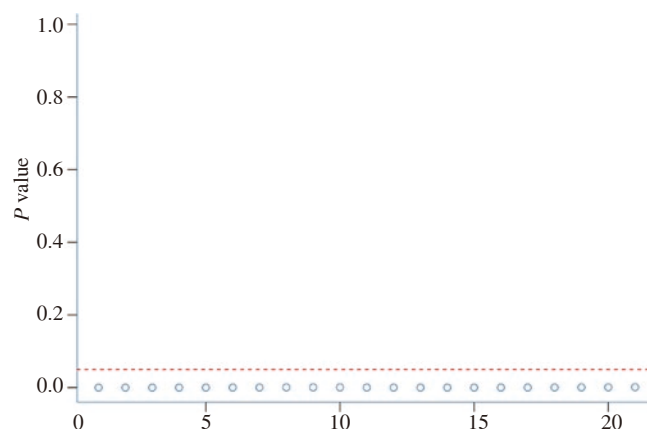

Figure 6. McLeod-Li test for residuals of ARIMA $(2, 1, 2) (1, 1, 1)_{12}$ for visceral leishmaniasis data.

Table 3. Comparison of predicted visceral leishmaniasis cases of ARIMA $(2,1,2) (1,1,1)_{12}$ model and actual values for Kashgar Prefecture from January to December 2016.

Time (Month)	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
AC	40	36	6	9	8	6	0	11	8	23	15	9
PC	36.23	22.57	15.66	13.84	14.29	18.27	25.03	50.10	77.90	107.94	116.77	65.13
95% *CI* of PC	18-54	0-47	0-43	0-43	0-44	0-49	0-55	20-81	47-109	77-139	85-148	33-98

AC: Actual cases; PC: Predictive cases.

Table 4. Information criterion for ARIMA $(2, 1, 2) (1, 1, 1)_{12}$/ARIMA $(2, 1, 2) (1, 1, 1)_{12}$-EGARCH (1, 1) models.

Models	Fit			Pred
	Root-mean-square error (%)	Akaike information criterion	Schwartz Bayesian criterion	Root-mean-square error (%)
ARIMA $(2, 1, 2) (1, 1, 1)_{12}$	8.62	966.46	986.58	48.53
ARIMA $(2, 1, 2) (1, 1, 1)_{12}$-EGARCH (1, 1)	11.07	6.16	6.39	7.23

4. Discussion

Kashgar is located in the southern border of Xinjiang and is economically underdeveloped. It is the main epidemic area of visceral leishmaniasis in China. Under the circumstances, although the country has basically controlled visceral leishmaniasis, the epidemic of visceral leishmaniasis in Kashgar Xinjiang still has not been fundamentally controlled. With the acceleration of the development in the western region and the increase of the floating population, the control and prevention of visceral leishmaniasis in Xinjiang has become more complex and difficult[32]. Our study suggested that serious endemic areas mainly emerged in Payzawat County, Kashi City, Shufu County and Marabishi County. In 2008-2009 and 2014-2016, an epidemic of visceral leishmaniasis occurred in Payzawat County, Kashgar Prefecture, which led to a sharp increase of the number of people suffering from visceral leishmaniasis and the highest peak in the number of people suffering from the disease in Xinjiang. In other years, it has maintained stable
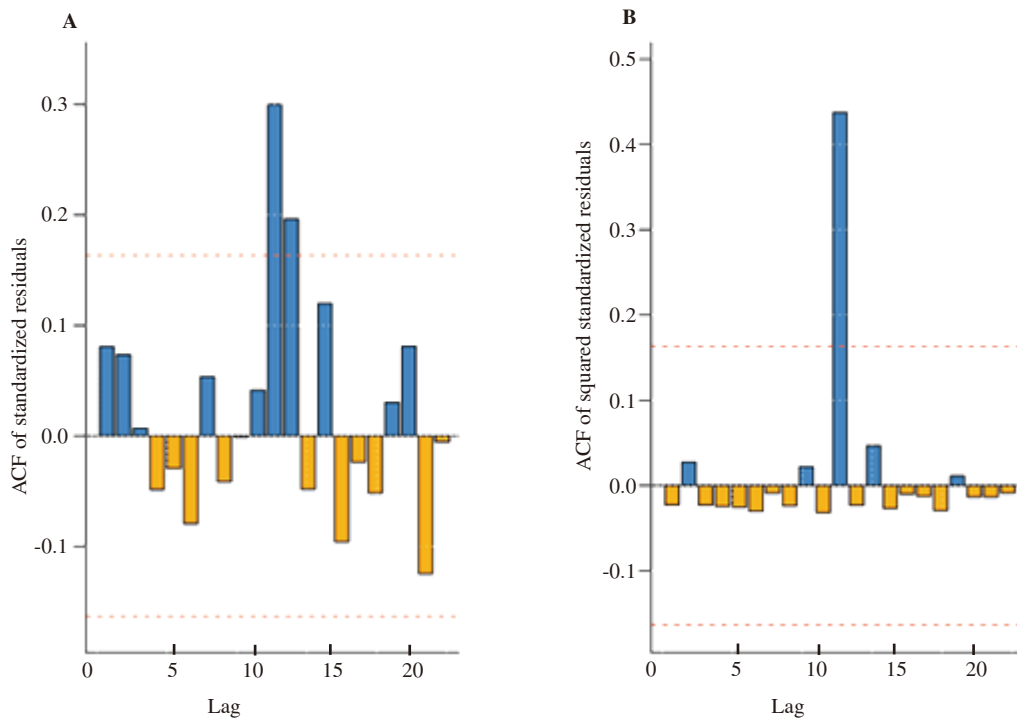

Figure 7. Standardized residual ACF of hybrid ARIMA-EGARCH model.
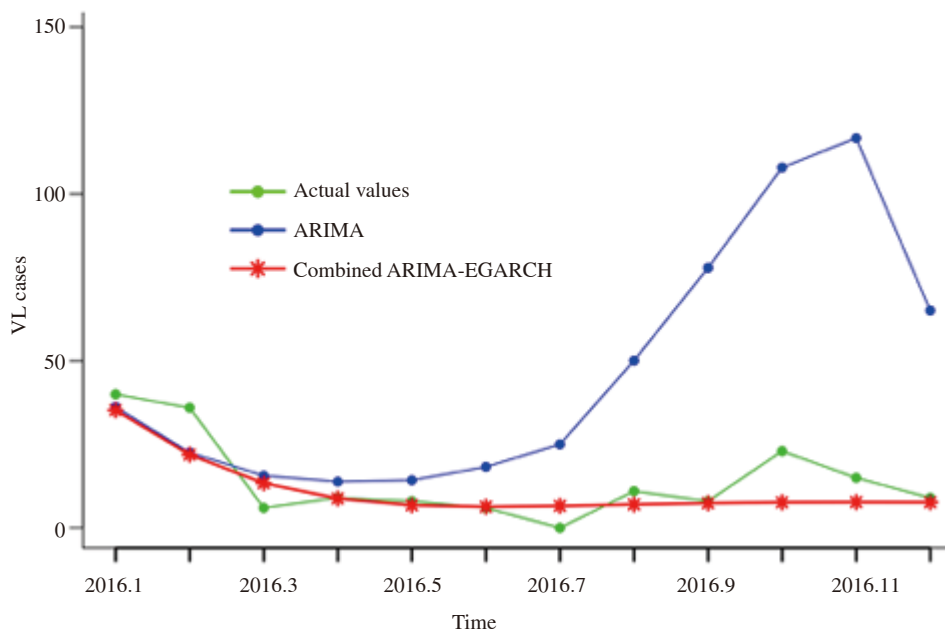

Figure 8. Comparison of predicted values of ARIMA (2, 1, 2) (1, 1, 1)$_{12}$ model and ARIMA (2, 1, 2) (1, 1, 1)$_{12}$-EGARCH (1, 1) model versus the actual visceral leishmaniasis from January to December 2016.

fluctuations. The visceral leishmaniasis cases in Xinjiang mainly were comprised of scattered children and farmers. Among them, infants and children under 2 years old accounted for more than half of the total number of cases, and they were the age-appropriate population of the natural epidemic type visceral leishmaniasis in Xinjiang. Our study is similar with the epidemiological aspects of visceral leishmaniasis in Kaleybar and Khoda-Afarin districts, north-west of Iran[33]. They collected 1 420 human(children under 12 years) samples, 101 domestic dogs samples (*Canis familiaris*), and 577 female sand fly samples were collected. Sera of human and dogs were tested using the direct agglutination test, and sand flies were identified at species level using the microscopic method. Furthermore, a structured questionnaire was applied to evaluate the correlation between the potential risk factors and the related clinical signs/symptoms with the human and dogs' seropositivity. This study showed that visceral leishmania infection is prevalent in rural areas of Kaleybar and KhodaAfar districts located in East-Azerbaijan province, therefore active detection and treatment of visceral leishmaniasis cases should not be neglected.

In this study, we used ARIMA (p, d, q) (P, D, Q)$_{12}$ model to analyze the visceral leishmaniasis data in Kashgar, China and the visceral leishmaniasis cases were simulated by ARIMA (2, 1, 2)(1, 1, 1)$_{12}$ model. However, the fitted cases did not reflect the actual case number. And the RMSE value for the validation phase was 48.56%. The predicted cases did not highly comply with the observed visceral leishmaniasis trend indicating that multiplicative seasonal ARIMA model can not accurately predict the occurring trend of case numbers of visceral leishmaniasis in the future. In order to improve the predicted accuracy of ARIMA (2, 1, 2) (1, 1,1)$_{12}$ model, the ARIMA (2, 1, 2) (1, 1, 1)$_{12}$-EGARCH (1, 1) model was further established to predict the visceral leishmaniasis incidence and it showed the best value of RMSE, AIC and SBC criteria equaled to 11.07%, 6.16 and 6.39, respectively. In the validation phases, the RMSE of ARIMA (2, 1, 2) (1, 1, 1)$_{12}$-EGARCH (1, 1) model was less than that of the ARIMA (2, 1, 2) (1, 1, 1)$_{12}$. So, through the RMSE, AIC and BIC criteria, ARIMA-EGARCH has been proven to be the best model.

Furthermore, through seasonal index analysis, it was shown that visceral leishmaniasis had a significant seasonality, as the most incidences occurred in Autumn and the peaked between September and November. Previous study had shown[14] that the outdoor bites and people whose neighbors have visceral leishmaniasis were the main risk factors for infecting leishmania. Avoiding outdoor bites and the universal popularity of the distribution of bed nets is an efficiency way to prevent and control outbreaks of visceral leishmaniasis infection in local populations. The epidemic factors of visceral leishmaniasis are complex, such as the influence from pathogens, vectors and floating population, which increases the difficulty of the prevention and control. Therefore, it is necessary

to raise the awareness of the long-term and arduous task of the prevention and control of visceral leishmaniasis.

Diverse models have also been developed due to the deficiencies of ARIMA, such as ARIMAX, SVM and MaxEnt ecological niche modeling. In 2014, Zhang Xingyu *et al*.[22] conducted a comparative study of four typical time series methods in the prediction of nine infectious diseases, named two decomposition methods (regression and exponential smoothing), ARIMA model and support vector machine based model. The differences in the principles and practices of these methods have been compared. The results of the study showed that no single method was found to be completely superior but the SVMs outperformed ARIMA models and decomposition methods in most cases.

A series of studies on visceral leishmaniasis have been published in Iran[17–19], and these articles all used time series methods to analyze the cases of visceral leishmaniasis. For example, Sharafi M *et al*.[17] used the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to predict the trend of cutaneous leishmaniasis and assess the relationship between the disease trend and weather variables in south of Fars province. Another research, Nikonahad A *et al*.[18] investigated the relationship between the environmental and metrological variables and cutaneous leishmaniasis transmission and its prediction in a region susceptible to this disease prevalence with using the time series model. Similarly, Tohidinik HR *et al*, in order to forecast the occurrence of zoonotic cutaneous leishmaniasis and evaluate the effect of climatic variables on disease incidence in the east of Fars province, they used the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The three studies on visceral leishmaniasis mentioned above all have an identical point. That is, variables such as earthquakes, rainfall, temperature, sunshine hours, and relative humidity might be related to the occurrence of visceral leishmaniasis. Finally, the results of these studies indicated that introducing meteorological variables into the model could improve its accuracy. Shiravand B *et al*.[34] obtained the distribution data for vector and reservoir hosts of zoonotic cutaneous leishmaniasis in Yazd province. Using MaxEnt ecological niche modeling to predict environmental suitability. BCC-CSM1-1(m) model and two climate change scenarios, RCP 4.5 and RCP 8.5 were used for horizons 2030 and 2050 climate projections. With both scenarios in 2030 and 2050, the results of jackknife test indicated that the mean temperature of wettest quarter and temperature annual range had the greatest effect on the model for the vector and the reservoir hosts, respectively. Their study demonstrated that the climate conditions are the major determinants of zoonotic cutaneous leishmaniasis incidence rate in Yazd Province. Understanding the role of environmental and bioclimatic factors in zoonotic cutaneous leishmaniasis occurrence can provide a guide for policy-makers in the creation and implementation of more effective policies for

prevention and control. In Xinjiang, China, few reports used the ARIMAX model to analyze the exist of visceral leishmaniasis. In future, we can consider applying these methods to the study of visceral leishmaniasis and seek more exact models to predict the incidence of visceral leishmaniasis in Xinjiang.

This study examined the visceral leishmaniasis cases in Kashgar Prefecture from 2004 to 2016 by using time series methodology. After series of models tried, we hold the ARIMA $(2, 1, 2) (1, 1, 1)_{12}$ as the best model because of its low AIC, AICc and BIC criteria and RMSE. As the visceral leishmaniasis time series possess volatility, an attempt was made to model this volatility using ARIMA-EGARCH models which was considered as appropriate for best fitting the model. In the future, we should apply other predictive models in the study of visceral leishmaniasis to present a richer theoretical basis for prevention and control of visceral leishmaniasis.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

Authors' contributions

Both Wang K and Lu XB are equally contributed authors who designed the work and participated the critical revision of the article and contributed to the final version of the manuscript. Zheng RJ, Zheng Q and Jiang W collected the data. Li HL performed the data analyses, interpreted the results and wrote the initial draft of the paper. Zhang XL and Wang WM revised the manuscript. Feng X help to perform the analysis with constructive discussions.

References

[1] Malaviya P, Hasker E, Singh RP, Van Geertruyden JP, Boelaert M, Sundar S.Village health workers in Bihar, India: An untapped resource in the struggle against visceral leishmaniasis.*Trop Med Int Health* 2013; **18**(2): 188-193.

[2] Nampoothiri RV, Sreedharanunni S, Chhabria BA, Jain S. *Visceral leishmaniasis: QJM* 2016; **109**(5): 347-348.

[3] Yaghoobi-Ershadi M. Phlebotomine sand flies (Diptera: Psychodidae) in Iran and their role on leishmania transmission. *J Arthropod Borne Dis* 2012; **6**(1):1-17.

[4] Masoori L, Kheirandish F, Haghighi A, Mohebali M, Akhoundi B, Taghipour N, et al. Molecular-based detection of *Leishmania infantum* in human blood samples in a new focus of visceral leishmaniasis in Lorestan Province, Iran. *J Arthropod Borne Dis* 2018; **12**(1): 67-75.

[5] McGwire BS, Satoskar AR. Leishmaniasis: Clinical syndromes and treatment. *QJM* 2014; **107**: 7-14.

[6] Bi K, Chen Y, Zhao S, Kuang Y, John Wu CH. Current visceral leishmaniasis research: A research review to inspire future study. *Biomed Res Int* 2018; **2018**: 9872095.

[7] Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* 2012; **7**(5): e35671.

[8] Hailu A, Dagne DA, Boelaert M. Leishmaniasis. In: Gyapong J, Boatin B. (eds.) *Neglected tropical diseases–Sub–Saharan Africa.* Switzerland: Springer International Publishing AG; 2016, p. 87-112.

[9] Bekele F, Belay T, Zeynudin A, Hailu A. Visceral leishmaniasis in selected communities of Hamar and Banna-Tsamai districts in Lower Omo Valley, South West Ethiopia: Sero-epidemological and leishmanin skin test surveys. *PLoS One* 2018; **13**(5): e0197430.

[10] Md YA, Md RR, Akib J, Arook T, Mst RA. Indigenous santal people sense and etiology regarding black fever illness. *Am J Health Res* 2016; **4**(5): 143-150.

[11] Zheng CJ, Xue CZ, Wu WP, Zhou XN. Epidemiological characteristics of visceral leishmaniasis disease in China, during 2005-2015. *Zhonghua Liu Xing Bing Xue Za Zhi* 2017; **38**(4): 431-434.

[12] Xiong GH. Present epidemiological situation of visceral leishmaniasis in China. *Difang Bing Tong Bao* 1992; **7**(3): 113-125.

[13] Maimaitiyiming YLK, Zhang HT, Zhang S, Zhumahong RZGL, Abuduwayiti SKDE. Epidemic analysis of visceral leishmaniasis in Xinjiang Uygur autonomous region from 2004 to 2014. *Jibing Yu Fang Kong Zhi Tong Bao* 2015; **30**(3): 5-8.

[14] Zhao GH, Wang HF, Zhong WX. Risk factors of visceral leishmaniasis outbreak in Kashi Prefecture, Xinjiang, China. *Zhongguo Ren Shou Gong Huan Bing Za Zhi* 2015; **31**(6): 592-596.

[15] Yan CY, Ai SJ, Kai S. Epidemiological analysis on visceral leishmaniasis cases during 1990-2014 in Kashi Prefecture, Xinjiang, China. *Guoji Yi Xue Ji Sheng Chong Bing Za Zhi* 2015; **42**(5): 276-279.

[16] Wang LY, Wu WP, Fu Q, Guan YY, Han S, Niu YL, et al. Spatial analysis of visceral leishmaniasis in the oases of the plains of Kashi Prefecture, Xinjiang Uygur Autonomous Region, China. *Parasit Vector* 2016; **9**(1):148.

[17] Sharafi M, Ghaem H, Tabatabaee HR, Faramarzi H. Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time series method. *Asian Pac J*

Trop Med 2017; **10**(1): 79-86.

[18] Nikonahad A, Khorshidi A, Ghaffari HR, Aval HE, Miri M, Amarloei A, et al. A time series analysis of environmental and metrological factors impact on cutaneous leishmaniasis incidence in an endemic area of Dehloran, Iran. *Environ Sci Pollut Res Int* 2017; **24**(16): 14117-14123.

[19] Tohidinik HR, Mohebali M, Mansournia MA, Niakan Kalhori SR, Ali-Akbarpour M, Yazdani K. Forecasting zoonotic cutaneous leishmaniasis using meteorological factors in eastern Fars province, Iran: A SARIMA analysis. *Trop Med Int Health* 2018; **23**(8): 860-869.

[20] 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): 358-362.

[21] Anwar MY, Lewnard JA, Parikh S, Pitzer VE. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malar J* 2016; **15**(1):566.

[22] 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.

[23] Moosazadeh M, Khanjani N, Nasehi M, Bahrampour A. Predicting the incidence of smear positive tuberculosis cases in Iran using time series analysis. *Iran J Public Health* 2015; **44**(11): 1526-1534.

[24] Lee HS, Her M, Levine M, Moore GE. Time series analysis of human and bovine brucellosis in South Korea from 2005 to 2010. *Prev Vet Med* 2013; **110** (2): 190-197.

[25] Fu KQ, Yan L, He J. Time series analysis of correlativity between pulmonary tuberculosis and seasonal meteorological factors based on theory of human-environmental inter relation. *J Tradit Chin Med Sci* 2018; **5**(2): 119-127.

[26] Babu CN, Reddy BE. Prediction of selected Indian stock using a partitioning-interpolation based ARIMA-GARCH model. *Appl Comput Informat* 2015; **11**(2): 130-143.

[27] Goyal A, Arora S. The Indian exchange rate and Central Bank action: An EGARCH analysis. *J Asian Econ* 2012; **23**(1): 60-72.

[28] 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.

[29] Bildirici M, Ersin Ö. Modeling Markov switching ARMA-GARCH neural networks models and an application to forecasting stock returns. *Sci World J* 2014; **2014**: 497941.

[30] Viorica D, Jemna D, Pintilescu C, Asandului M. The relationship between inflation and inflation uncertainty. Empirical evidence for the newest EU countries. *PLoS One* 2014; **9**(3): e91164.

[31] He J, Luo L, Jin RG, Li JM. The application of ARIMA in forecasting the cases of rabies in China different human groups. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* 2018; **36**(7): 512-515.

[32] Zhao S, Li Z, Zhou S, Zheng C, Ma H. Epidemiological Feature of visceral leishmaniasis in China, 2004-2012. *Iran J Public Health* 2015; **44**(1): 51-59.

[33] Behniafar H, Moin-Vaziri V, Mohebali M, Tabaei SJS, Zarei Z, Kazemirad E, et al. Visceral leishmaniasis among children in an endemic area of northwestern Iran between 2016 and 2017: An epidemiological study. *Asian Pac J Trop Med* 2019; **12**(7): 306-314.

[34] Shiravand B, Hanafi-Bojd AA, Tafti AAD, Abai MR, Almodarresi A, Mirzaei M. Climate change and potential distribution of zoonotic cutaneous leishmaniasis in Central Iran: Horizon 2030 and 2050. *Asian Pac J Trop Med* 2019; **12**(5): 204-215.