

## Predicting CCHF incidence and its related factors using time-series analysis in the southeast of Iran: comparison of SARIMA and Markov switching models

H. ANSARI<sup>1</sup>, M. A. MANSOURNIA<sup>2</sup>, S. IZADI<sup>1</sup>, M. ZEINALI<sup>3</sup>,  
M. MAHMOODI<sup>2</sup> AND K. HOLAKOUIE-NAIENI<sup>2\*</sup>

<sup>1</sup>Health Promotion Research Center, Zahedan University of Medical Sciences, Zahedan, Iran

<sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

<sup>3</sup>Ministry of Health, Center for Disease Control and Prevention, Tehran, Iran

Received 19 December 2013; Final revision 28 February 2014; Accepted 15 April 2014;  
first published online 22 May 2014

### SUMMARY

Crimean-Congo haemorrhagic fever (CCHF) is endemic in the southeast of Iran. This study aimed to predict the incidence of CCHF and its related factors and explore the possibility of developing an empirical forecast system using time-series analysis of 13 years' data. Data from 2000 to 2012 were obtained from the Health Centre of Zahedan University of Medical Sciences, Climate Organization and the Veterinary Organization in the southeast of Iran. Seasonal autoregressive integrated moving average (SARIMA) and Markov switching models (MSM) were performed to examine the potential related factors of CCHF outbreaks. These models showed that the mean temperature (°C), accumulated rainfall (mm), maximum relative humidity (%) and legal livestock importation from Pakistan (LIP) were significantly correlated with monthly incidence of CCHF in different lags ( $P < 0.05$ ). The modelling fitness was checked with data from 2013. Model assessments indicated that the MSM had better predictive ability than the SARIMA model [MSM: root mean square error (RMSE) 0.625, Akaike's Information Criterion (AIC) 266.33; SARIMA: RMSE 0.725, AIC 278.8]. This study shows the potential of climate indicators and LIP as predictive factors in modelling the occurrence of CCHF. Our results suggest that MSM provides more information on outbreak detection and can be a better predictive model compared to a SARIMA model for evaluation of the relationship between explanatory variables and the incidence of CCHF.

**Key words:** Climate factor, Crimean-Congo haemorrhagic fever, MSM, SARIMA, time series.

### INTRODUCTION

Crimean-Congo haemorrhagic fever (CCHF) is a viral zoonotic disease with a high mortality rate in humans. CCHF is a public health problem in many regions of

the world such as Eastern Europe, Asia, the Middle East and Africa. The ecological complexity of vector-borne diseases, therapeutic controversy, and human-to-human transmission of a zoonotic infection make CCHF an interesting topic for research [1].

There has been a substantial increase in reports of CCHF virus (CCHFV) over the past 5 years and the geographical range of CCHFV is the most extensive of the tick-borne viruses that affect human health [1–3]. CCHFV can be transferred from endemic to

\* Author for correspondence: Dr K. Holakouie-Naieni, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. (Email: holakoik@hotmail.com)

non-endemic areas due to carriage of large numbers of infected ticks by migrating birds and livestock. This can cause the spread of CCHFV into uninfected areas [4]. Although there are other species of genera of ticks which act as vectors of CCHF [5], the genus *Hyalomma* has the main role in the transmission cycle of CCHF, especially in the southeast of Iran [3, 6, 7].

Several studies have discussed the relationship between climatic factors and the vector's life-cycle, ecological conditions, and consequent occurrence of CCHF in human populations [1, 8–10]. On the other hand, climatic factors may lead to livestock diseases through their effects on a number of factors including the range and abundance of vectors and wildlife reservoirs, survival of the pathogen in the environment and farming practice [8, 11]. The changes in climatic conditions have been suggested as facilitatory factors for reproduction of the tick population and the consequent increase in the incidence of tick-borne infectious diseases [1, 11]. Moreover, legal and potentially illegal animal transportation from neighbouring countries impose a risk of international spread from endemic countries to other ones [4, 11, 12]. There are both legal and illegal livestock transportation and uncontrolled population movement between Iran and its neighbouring countries, Pakistan and Afghanistan [12]. Therefore, the occurrence of CCHF appears to be affected by the import of livestock from neighbouring countries [9] and this should be considered in early warning of CCHF outbreaks.

So far the occurrence of CCHF has been reported in 23–26 out of 30 provinces of Iran. From 1999 (starting the register of CCHF in Iran) to 2012, the Sistan-va-Baluchistan province, which is located in the southeast of Iran, has been recognized as the most important endemic site of the disease in the country [3, 13, 14]. In Iran the most affected professions included jobs that involved the handling of blood and organs from infected livestock [3]. Therefore, infected livestock play a main role in the transmission cycle of this infection.

It is clear that an early warning of CCHF outbreaks based on related factors could prevent outbreaks, decrease mortality rates, and help to target preventive actions. An early warning system is a competent tool for working as a surveillance system for early diagnosis of the disease in humans and animals and for monitoring suspected outbreaks [4].

This study reports the surveillance data collected during the past 13 years in Sistan-va-Baluchistan province, and uses the climate findings and legal

livestock importation from Pakistan (LIP) details to create a basis for making early warnings of CCHF outbreaks. Specifically, this study attempts to follow two goals. First, it explores the potential impact of weather variability and LIP on CCHF incidence using valuable time-series models. Second, it examines the difference between the predictive ability of MSM and SARIMA models.

## MATERIALS AND METHODS

### Study area

Sistan-va-Baluchistan province is located in the southeast of Iran, and has common borders with Pakistan and Afghanistan. Its capital is Zahedan, which is located between 45° 32' and 48° E and 34° 47' and 35° 1' N. The economy is mainly based on agriculture and livestock. As a result, a large proportion of the population comes into close contact with livestock. This provides a high risk of exposure to CCHF virus. Sistan-va-Baluchistan province today accounts for one of the driest regions of Iran with a slight increase in rainfall from east to west and an obvious rise in humidity in winter. The climate is tropical with two distinct seasons: a dry season from April to October, characterized by relatively low rainfall from November to March. The climate condition of this province is similar to the border provinces of Pakistan and Afghanistan [15]. About 10% of the population of this province consists of immigrants from Afghanistan and Pakistan. There is also some nomadic population in sparse and scattered villages and their usual occupation is tenting and trading livestock [16].

### Case definition

A confirmed CCHF case was defined as one with a positive IgM or IgG serological test (ELISA method) and/or positive by RT-PCR detection of viral RNA in the serum sample sent to the Laboratory of Arboviruses and Viral Haemorrhagic Fevers, Pasteur Institute of Iran, Tehran, Iran [17]. All cases had been reported and registered according to the surveillance and control programme of CCHF in Iran. An individual was considered to be a suspect case when exhibiting sudden onset of fever, myalgia, and different haemorrhagic manifestations with an epidemiological background such as a history of tick bite, handling animal or human blood or tissue. These suspected cases were screened as probable

cases whose symptoms include thrombocytopenia (platelets  $<150\,000/\text{mm}^3$ ), leucopenia [white blood cell count (WBC)  $<3000/\text{mm}^3$ ] or leucocytosis (WBC  $>9000/\text{mm}^3$ ). Finally, any CCHF probable case whose serum was positive for immunoglobulin M (IgM) antibodies and/or who was positive by RT-PCR detection for viral RNA was considered a CCHF confirmed case. The samples of cases classified as a probable CCHF are sent to the National Reference Laboratory [18].

### Data collection

The data of all confirmed CCHF cases have been registered from 2000 to 2012 by the surveillance system of the Province Health Centre of Zahedan University of Medical Sciences (ZUMS). This databank is under the supervision of the Centre for Management of Communicable Diseases in Iran. However, data from 2013 was used for checking the model's validity.

The monthly temperature data ( $^{\circ}\text{C}$ ), the monthly accumulated rainfall (in  $\text{mm}^3$ ), and the monthly relative humidity (percentage) were collected from the meteorological organization of Sistan-va-Baluchistan province. We used the records of two synoptic centres located in the east (Saravan station) and northeast (Zabol station) of the province. As the occurrence of CCHF in Sistan-va-Baluchistan province appeared to be affected by the import of livestock from neighbouring countries, the data for LIP were collected from the veterinary organization of Sistan-va-Baluchistan province and also via searching the documents related to quarantine during these years. As the importation of legal livestock requires the authorization of the Veterinary Organization of Sistan-va-Baluchistan province, this organization supervised all legal livestock importation from Pakistan and registered the number of imported livestock each day.

### Statistical analyses

Simple (unadjusted) analyses were conducted for response and each explanatory variable as univariate analysis. Cross-correlation coefficients were used to compute a series of correlations between explanatory variables (climate factors and LIP) and the incidence of CCHF over a range of time lags. A time lag was defined as the time span between explanatory-variable observation and the incidence of CCHF (e.g. in this study the correlation between CCHF incidence and mean temperature at lag 1 is the correlation between

number of CCHF cases in this month and mean temperature in the previous month).

### SARIMA model

As both CCHF incidence and weather variables exhibited strong seasonal variation and fluctuations in their annual means, the seasonality was adjusted by first seasonally differencing the series, replacing each observation by the difference between itself and the observation a year ago (i.e. seasonal component was removed by a seasonal differencing:  $Z_t - Z_{t-s}$  (where  $Z_t$  = values of the time series at time  $t$  and  $Z_{t-s}$  = values of the time series at time  $t - 12$  months). Four steps were undertaken in the modelling as follows.

First, the variance of the series was stabilized by natural logarithm transformation. The dependent variable, LIP and accumulated rainfall were log transformed. Then, prior to modelling, both CCHF incidences and explanatory variables were transformed into a stationary input series [19]. Second, SARIMA models were developed using the log-transformed monthly incidence of CCHF. The log transformation facilitated the assumption of normally distributed responses. Seven main parameters were selected when fitting the SARIMA  $(p,d,q)(P,D,Q)_s$  model: the order of autoregression ( $p$ ) and seasonal autoregression ( $P$ ), the order of integration or regular differencing ( $d$ ) and seasonal integration or seasonal differencing ( $D$ ), and the order of moving average ( $q$ ) and seasonal moving average ( $Q$ ), and the length of seasonal period ( $s$ ). To identify the order of moving average and autoregressive parameters, the structure of temporal dependence of stationary time series is assessed by the analysis of autocorrelation (ACF) and partial autocorrelation (PACF) functions, respectively. The selection of SARIMA processes was conducted using Akaike's Information Criterion (AIC). Of all the models tested, a SARIMA (1,0,1)(0,1,1)<sub>12</sub> model was found to best fit the data [19].

The explanatory variables with different lags (delayed effects) were found to correlate with CCHF cases using cross-correlation coefficients as univariate analysis. The regression SARIMA model was fitted with climate variables and LIP as external regressors to CCHF incidence.

Third, the ACF of residuals and the Ljung-Box test were used to check if the assumption were met [19]. Finally, the best final model was selected using AIC, which measures how well the model fits the series and the validity of the models was checked by fitting

with data from 2013. The root mean square error (RMSE) for models was also assessed. The RMSE is equal to the square root of the mean of the differences between true and predicted values [19].

**Markov switching model (MSM)**

Although we can control for seasonality and trend, when the occurrence of outcome is not linear (such as CCHF occurrence in this study) or the research is analysing surveillance data from small geographical areas with health conditions that are rare, the Box–Jenkins SARIMA model is not recommended [20]. On the other hand, MSM is one of the time-series models that is used for modelling nonlinear outcomes.

The MSM [21] belongs to the family of state-space models. There are two types of equations in this model: measurement equations and transition equations. The measurement equation defines how hidden states affect the observable random variables. The transition equation defines how the state variables evolve over time. The observable random variables in the MSM depend on their historical values as well as the hidden state variables. This setting makes the MSM more suitable for time-series-related problems. A simple MSM can be written as:

$$y_t = a_{0,0} + a_{0,1}S_t + (a_{1,0} + a_{1,1}S_t)y_{t-1} + e_t, \tag{1}$$

$$P(S_t = j | S_{t-1} = i) = P_{ij}, \tag{2}$$

$$S_t \equiv (0, 1), \tag{3}$$

$$e_t \sim N(0, \sigma^2), \tag{4}$$

Equation (1) defines how the hidden state variable  $S_t$  controls the dynamics of the observable random variable  $y_t$ . In a non-outbreak period ( $S_t=0$ ),  $y_t$  is determined by a drift term  $a_{0,0}$  and the autoregressive parameter  $a_{1,0}$ . If an outbreak occurs ( $S_t=1$ ), the drift term increases to  $a_{0,0}+a_{0,1}$  and the autoregressive parameter increases to  $a_{1,0}+a_{1,1}$  (assuming  $a_{0,1} \geq 0$  and  $a_{1,1} \geq 0$ ). Equation (2) indicates that the hidden states evolve following a Markov process with transition probability  $P_{ij}$ .  $P_{ij}$  is the probability of state  $j$  at time  $t$  conditional to state  $i$  at time  $t-1$ .

A MSM with exogenous variables and modelling of seasonal effects can be written as:

$$Y_t = a_{0,0} + a_{0,1}S_t + (a_{1,0} + a_{1,1}S_t)y_{t-1} + b_1 \sum_{i=1}^{12} \sin\left(\frac{2\pi i}{12}\right) + b_2 \sum_{i=1}^{12} \cos\left(\frac{2\pi i}{12}\right) + \sum_{i=1}^k c_i v_{t,i} + e_t, \tag{5}$$

where  $b_1 \sum_{i=1}^{12} \sin\left(\frac{2\pi i}{12}\right)$  and  $b_2 \sum_{i=1}^{12} \cos\left(\frac{2\pi i}{12}\right)$  are seasonality control terms and  $\sum_{i=1}^k c_i v_{t,i}$  are exogenous and potential controlling factors. If necessary, lagged independent variables can also be included. For example, we can set  $v_{t,1} = x_{t-1}$ ,  $v_{t,2} = x_{t-2}$ , ...,  $v_{t,n} = x_{t-n}$ . The expectation maximization (EM) algorithm [22] was used for model estimation.

However, the probability of outbreak at period  $t+1$ , could be estimated as follows:

$$\left. \begin{aligned} P[S_{t+1} = 0] &= (1 - P_{11}) / (2 - P_{00} - P_{11}), \\ P[S_{t+1} = 1] &= (1 - P_{00}) / (2 - P_{00} - P_{11}). \end{aligned} \right\} \tag{6}$$

The validity of the MSM was also checked by fitting with data from 2013. However, the final SARIMA and MSM models were comparable based on AIC, RMSE and absolute mean number of cases (number of predicted cases minus number of observed cases in 2013). In this model, seasonality was modelled using a sinusoidal transformation of time, including both  $\sin(2\pi i/12)$  and  $\cos(2\pi i/12)$  in the regression models, where  $i$  represents the number of the month (e.g. January = 1, etc.).

The statistical software Stata v. 10 (StataCorp., USA) and OxMetrics 6.01 (oxmetrics.net/) was used for all analyses.

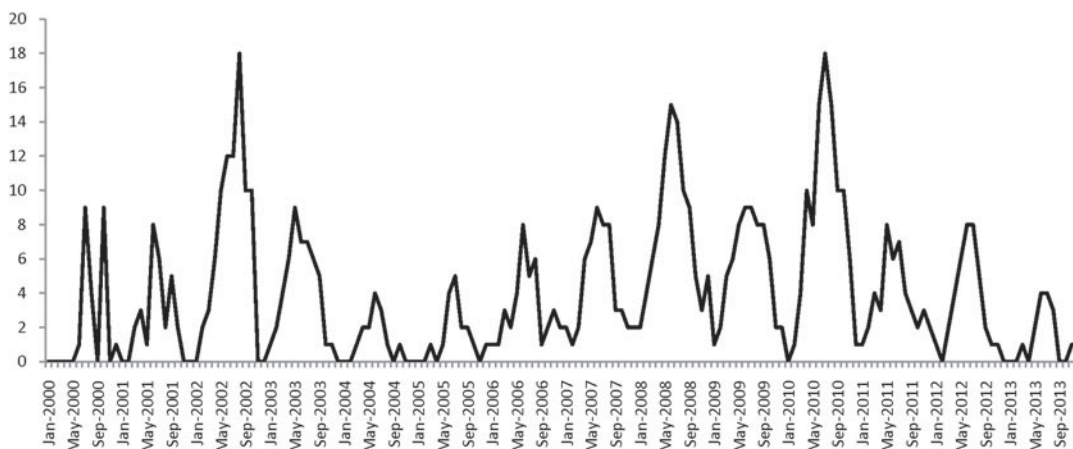
**RESULTS**

**Descriptive analysis**

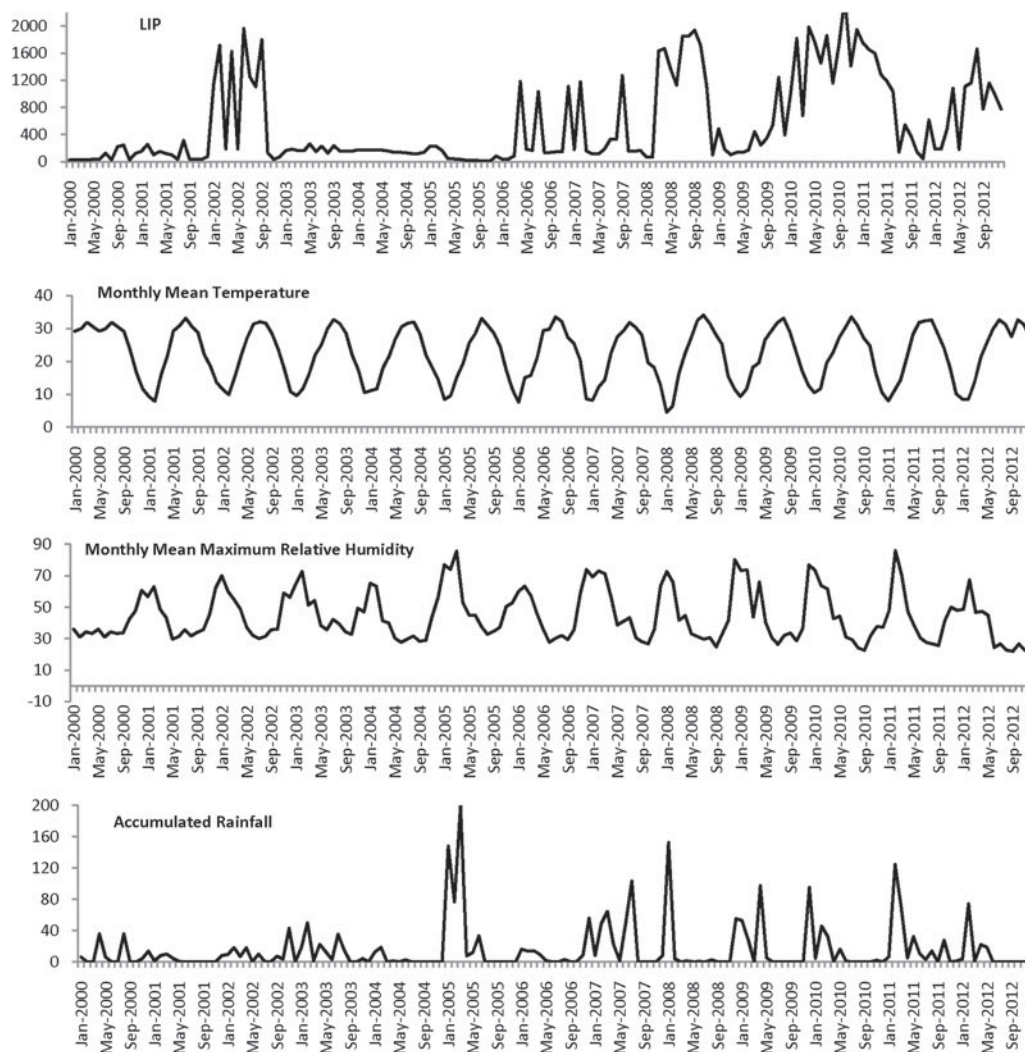
Between January 2000 and December 2012, 647 confirmed CCHF cases were reported from Sistan-va-Baluchistan province, in the southeast of Iran. The disease was most common in the months of May, June, July and August and there was no clear pattern of decline during these years. The years 2002, 2008 and 2010 were the worst with regard to the number of cases and occurrence of outbreaks. The trend of number of cases from 2000 to 2013 is displayed in Figure 1. The trend of significant climate data and LIP during the study period (the data for climate and LIP from 2000 to 2012 was used for analysing and modelling) are displayed in Figure 2.

**Cross-correlation function**

The results of the cross-correlations adjusted for seasonality show that the incidence of CCHF was significantly associated with LIP, mean temperature and maximum monthly relative humidity at lags of up



**Fig. 1.** Trend and distribution of confirmed Crimean-Congo haemorrhagic fever cases by month and year, from 2000 to 2013 in southeast of Iran.



**Fig. 2.** The time series and trend of climate data and legal livestock importation from Pakistan (LIP) from 2000 to 2012 in the southeast of Iran.

Table 1. Cross-correlation coefficients between explanatory variables and monthly number of CCHF cases, in Sistan-va-Baluchistan province, southeast of Iran

Time-lag (months)	Average temperature (°C)	Maximum temperature (°C)	Minimum temperature (°C)	Average relative humidity (%)	Maximum relative humidity (%)	Minimum relative humidity (%)	Rainfall (mm)	Sunny hours	Barometric pressure (mb)	Livestock importation from Pakistan
-5	-0.43*	-0.15	-0.12	0.11	0.15	0.15	0.27*	-0.19	0.11	0.11
-4	-0.14	-0.13	0.15	0.12	0.13	0.15	0.04	0.01	-0.02	0.09
-3	0.09	0.1	0.083	0.14	0.39*	0.13	0.15	0.12	0.07	0.09
-2	0.36*	0.12	0.1	0.09	0.38*	0.09	-0.05	0.01	0.05	0.26*
-1	0.32*	0.11	0.09	0.11	0.5*	0.11	-0.23*	0.14	0.12	0.45*
0	0.45*	0.13	0.13	0.12	0.34*	0.12	-0.13	0.12	-0.12	0.52*

\* Statistically significant at the 5% level.

to 2 months. Moreover, significant correlation for monthly mean temperature was found at a lag of 5 months (reversed) and maximum monthly relative humidity at lag of 3 month. We also found that the monthly accumulated rainfall was inversely correlated with CCHF incidence at a lag of 1 month and directly correlated to disease incidence at a lag of 5 months. However, there was no significant correlation between CCHF incidence and the other variables in any lag (Table 1).

**SARIMA model**

The best-fit final SARIMA model (based on AIC and results of Lrtest) show that first-order autoregression, first-order moving average, first-order seasonal moving average, monthly mean temperature at a prior moving average (lag) of 2 and 5 months, maximum monthly relative humidity at a lag of 2 months, monthly accumulated rainfall at a lag of 5 months and LIP without delay (lag-0), were significantly associated with CCHF incidence. The model estimated with the explanatory variables was a better fit than the model without these variables, in terms of smaller values of AIC and RMSE (Table 2). Because of collinearity, the significant climate variables at lag-0 month, was not included in the multiple SARIMA model. The Ljung–Box test confirmed that the time-series residuals were statistically not dependent (P=0.6). Moreover, autocorrelation functions and histogram of residuals approved the independence and normality of residuals, respectively.

The selected SARIMA model fitted observed data from January 2000 to December 2012. To predict for 2013, the 12 month-step approach showed the smallest RMSE and the predictions and their confidence intervals were improved after the variables introduced remained in the final model (Fig. 3).

**MSM**

We normalized the dependent variable using log transformation. Based on results of the simple MSM fitted model, one time lag of CCHF can help to predict this disease in later time. (Table 3). As stated earlier, the simple MSM is  $y_t = a_{0,0} + a_{0,1}s_t + (a_{1,0} + a_{1,1}s_t)y_{t-1} + e_t$ . Therefore, we found that  $a_{0,0} = 0.207$ ,  $a_{0,1} = 1.24 - 0.207 = 1.033$ ,  $a_{1,0} = 0.315$ ,  $a_{1,1} = 0.454 - 0.315 = 0.139$ . Furthermore, based on the simple model, the

Table 2. Regression coefficients of SARIMA (1, 0, 1) (0, 1, 1)12 on the monthly incidence of CCHF in Sistan-va-Baluchistan, southeast of Iran, 2000–2012 (seasonality controlled with seasonal differencing)

Variables	Model without explanatory variables (simple SARIMA model)			Model with explanatory variables (multiple SARIMA model)			
	$\beta$	s.e.	<i>P</i> value	$\beta$	s.e.	95% CI	<i>P</i> value
Autoregression	0.917	0.064	<0.001	0.865	0.097	0.67 to 1.05	<0.001
Moving average	-0.667	0.101	<0.001	-0.567	0.100	-0.76 to -0.37	<0.001
Seasonal moving average	-0.744	0.098	<0.001	-0.899	0.180	-1.2 to -0.54	<0.001
Mean temperature_lag1	—			0.004	0.014	-0.023 to 0.03	0.770
Mean temperature_lag2	—			0.033	0.013	0.007 to 0.05	0.015
Mean temperature_lag5	—			-0.100	0.030	-0.15 to -0.04	0.001
Maximum relative humidity_lag2	—			0.067	0.020	0.02 to 0.16	0.001
Accumulated rainfall_lag5	—			0.014	0.007	0.0003 to 0.02	0.049
LIP_lag0				0.218	0.095	0.03 to 0.4	0.022
AIC	336.68			278.8			
RMSE	0.896			0.725			

CCHF, Crimean-Congo haemorrhagic fever; s.e., standard error; CI, confidence interval; LIP, legal livestock importation from Pakistan; AIC, Akaike’s Information Criterion; RMSE, root mean square error.

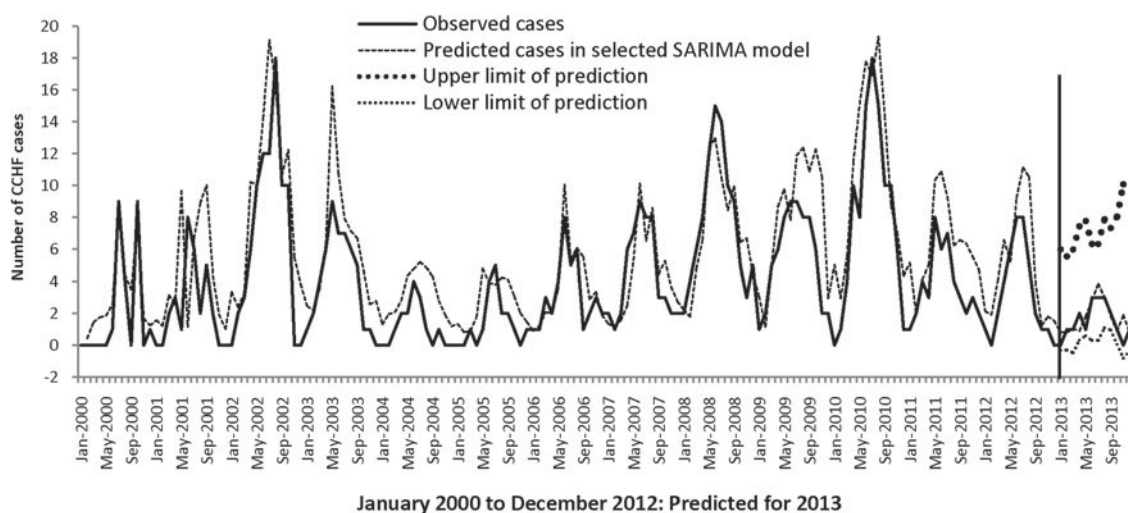


Fig. 3. Observed reported cases of Crimean-Congo haemorrhagic fever (CCHF) and predicted values based on final selected SARIMA model in the southeast of Iran. The multiple SARIMA (1,0,1)(0,1,1)12 model and 12 months-ahead values with 95% prediction intervals are presented.

transition probabilities can be estimated as follows:

$$P_{ij} = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} P_{00} & 1 - P_{00} \\ 1 - P_{11} & P_{11} \end{bmatrix} \\
 = \begin{bmatrix} 0.7 & 0.3 \\ 0.41 & 0.59 \end{bmatrix},$$

where,  $P_{00}$  is the probability of non-outbreak state at both periods  $t$  and  $t+1$ ,  $P_{11}$  is the probability of outbreak state at both periods  $t$  and  $t+1$ ,  $P_{01}$  is the probability of changing the series from non-outbreak state at period  $t$  to outbreak state at period  $t+1$  and

$P_{10}$  is the probability of changing the series from outbreak state at period  $t$  to non-outbreak state at period  $t+1$ . Based on equation (6) the  $P(S_{t+1}=1)$  is as follows:

$$P[S_{t+1} = 0] = (1 - 0.59)/(2 - 0.7 - 0.59) = 0.57, \\
 P[S_{t+1} = 1] = 0.43.$$

As our data was up to December 2012, this means that the probability of an outbreak is relatively low in January 2013 (1 month later).

Table 3. Regression coefficients of simple and multiple MSM on the monthly incidence of CCHF in Sistan-va-Baluchistan, southeast of Iran, 2000–2012 (seasonality controlled with harmonic seasonal factors and sinusoidal term included in the model). The  $\beta$  coefficients for the variables indicate the effect of explanatory variables on CCHF incidence in non-outbreak and outbreak periods, respectively

Variables	Simple MSM			Multiple MSM			
	$\beta$	s.e.	P value	$\beta$	s.e.	95% CI	P value
Constant(0), $a_{0,0}$	0.207	0.134	0.120	0.64	0.42	-0.18 to 1.4	0.120
Constant(1), ( $a_{0,0} + a_{0,1}$ )	1.24	0.221	<0.001	1.11	0.53	0.07 to 2.1	0.042
Autoregression (0) ( $a_{1,0}$ )	0.315	0.098	0.002	0.22	0.10	0.02 to 0.41	0.036
Autoregression (0), ( $a_{1,0} + a_{1,1}$ )	0.454	0.059	<0.001	0.64	0.21	0.22 to 1.05	0.032
Mean temperature_lag1(0)	—	—	—	0.07	0.35	-0.6 to 0.7	0.831
Mean temperature_lag1(1)	—	—	—	0.48	0.13	0.22 to 0.73	0.001
Mean temperature_lag2(0)	—	—	—	0.05	0.03	-0.008 to 0.1	0.112
Mean temperature_lag2(1)	—	—	—	0.16	0.05	0.06 to 0.2	0.002
Mean temperature_lag5(0)	—	—	—	0.04	0.03	-0.01 to 0.09	0.071
Mean temperature_lag5(1)	—	—	—	0.05	0.15	-0.2 to 0.31	0.701
Maximum relative humidity_lag2(0)	—	—	—	0.09	0.12	-0.14 to 0.32	0.911
Maximum relative humidity_lag2(1)	—	—	—	0.34	0.12	0.1 to 0.5	0.012
Accumulated rainfall_lag1(0)	—	—	—	0.03	0.81	-1.5 to 1.6	0.901
Accumulated rainfall_lag1(1)	—	—	—	0.02	0.02	-0.02 to 0.04	0.112
Accumulated rainfall_lag5(0)	—	—	—	0.06	0.07	-0.07 to 0.19	0.351
Accumulated rainfall_lag5(1)	—	—	—	0.01	0.02	-0.02 to 0.05	0.601
LIP_lag0(0)	—	—	—	0.04	0.19	-0.33 to 0.41	0.812
LIP_lag0(1)	—	—	—	0.72	0.08	0.56 to 0.87	<0.001
LIP_lag1(0)	—	—	—	0.01	0.02	-0.03 to 0.05	0.512
LIP_lag1(1)	—	—	—	0.66	0.06	0.54 to 0.77	<0.001
$P_{00}$	0.70	—	—	0.78	—	—	—
$P_{10}$	0.41	—	—	0.39	—	—	—
AIC	327.1311	—	—	266.339	—	—	—
RMSE	0.736	—	—	0.625	—	—	—

CCHF, Crimean-Congo haemorrhagic fever; MSM, Markov switching model; s.e., standard error; CI, confidence interval; LIP, legal livestock importation from Pakistan; AIC, Akaike’s Information Criterion; RMSE, root mean square error.

The best-fit final MSM with explanatory variables based on regimen change pattern, show that first-order autoregression ( $a_{1,0} + a_{1,1}S_t$ ), monthly mean temperature at a prior moving average or lags of 1 and 2 months, maximum monthly relative humidity at a lag of 2 months, and LIP at lags of up to 1 month were significantly associated with CCHF outbreaks. The MSM estimated with the explanatory variables was a better fit than the model without these variables, in terms of smaller values of AIC and RMSE (Table 3). As the most difference between coefficient of explanatory variables in outbreak and non-outbreak functions is due to the LIP variable, this is therefore the most important factor in causing outbreaks in the southeast of Iran. The transition probabilities in multiple MSM are different from simple MSM and the explanatory variables changed the probability of outbreaks and their transitions. Similar to simple MSM (mentioned above), we can calculate the

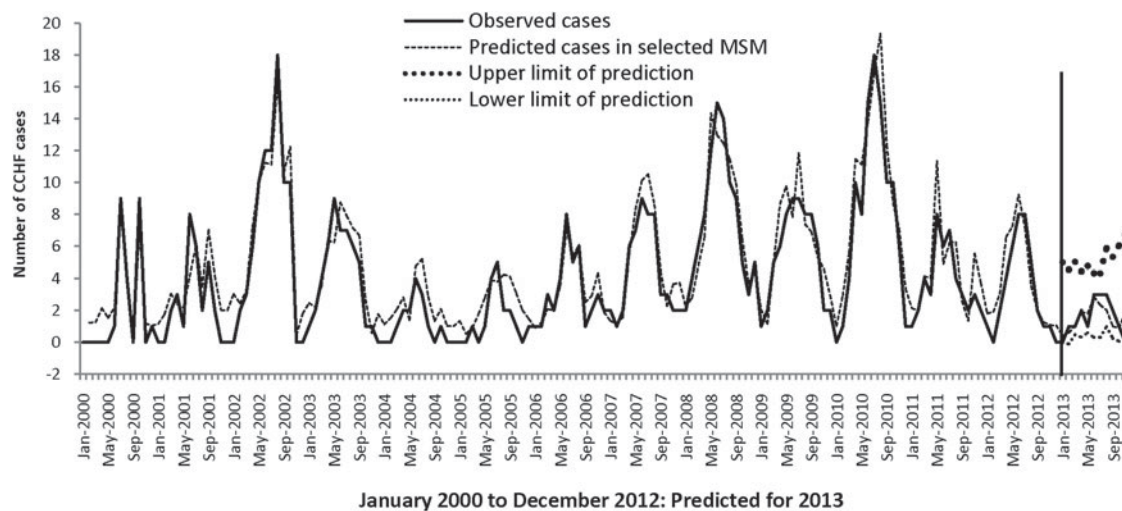
$P(S_{t+1} = 1)$  using estimated coefficients in multiple MSM in Table 3.

The selected MSM fitted observed data from January 2000 to December 2012. We also present the 12-month step approach and its confidence intervals (Fig. 4).

**Comparison of MSM and SARIMA models**

The nature and magnitude of the effect estimates is not too different for the two methods used in this study. The selection of variables was based on cross-correlation function. Both MSM and SARIMA models show that LIP without delay (lag-0), monthly mean temperature and monthly mean maximum humidity at a lag of 2 month were significantly associated with CCHF disease. Moreover, monthly mean temperature and LIP at a lag of 1 month were significantly associated with CCHF disease only in the MSM.





**Fig. 4.** Observed reported cases of Crimean-Congo haemorrhagic fever (CCHF) and predicted values based on final selected MSM in the southeast of Iran. The multiple MSM and 12 month- ahead values with 95% prediction intervals are presented.

However, in both SARIMA and MSM models, the validity of the models was checked by fitting with data from 2013. When the models were fitted to the data of 2013, the models were able to predict infections in that year in an appropriate way (Figs 3, 4); the peaks and dips of the prediction and infection curves are in the same direction in both SARIMA and MSM models. However, the mean number of cases (number of predicted cases minus number of observed cases in 2013) (regardless of mathematical sign) was 0.61 and 0.82 for MSM and SARIMA models, respectively. Although both models had reasonable accuracy over the predictive period, based on AIC and RMSE, the MSM in both simple and multiple models had slightly better fitted than the SARIMA model (Tables 2 and 3).

## DISCUSSION

Based on our findings in this study there was no clear pattern of decline in the reported number of CCHF cases during the past 13 years and a fluctuation was seen in occurrence of this disease with three peaks in years 2002, 2008 and 2010. This result is somewhat similar to previous study conducted in Bulgaria that describe the trends of CCHF between 1997 and 2009 [10]. However, in a Turkish study the number of cases increased markedly from 2004 to 2007 [23].

The results of this study (based on two reliable models) suggest that climate variability (particularly mean temperature and maximum humidity) and LIP

may have played a significant role in the incidence of CCHF in the southeast of Iran either directly or through other unmeasured variables. This result is relatively similar to previous studies conducted in Iran and Europe [9, 10]. The predominant effect of climate variables was observed after a lag of 1 and 2 or 3 months for some climate variables (accumulated rainfall, mean temperature, maximum relative humidity). However, the predominant effect of LIP was after a 1-month lag or without delay (lag-0).

In the southeast of Iran, the main routes of transmission for CCHF is contact with blood or tissue of infected livestock [3]. However, it should be mentioned that an increase in reservoir (ticks) activity and population cause an increase in the source (infected livestock) of the virus. Therefore, it might be concluded that climatic variables do not influence directly the incidence of the disease, but only indirectly, and through their effect on the life-cycle dynamics of both vector and virus and the consequent infection in livestock. On the other hand, the several successive phases from tick hatching to appearance of human cases led to global cumulative lags in our study (especially in univariate analysis using cross-correlations).

As stated earlier, although the genus *Hyalomma* is not the only vector of CCHF, it has, however, the main role, especially in the southeast of Iran [6, 7]. As the activity of *Hyalomma* is not limited to one specific month and they are found throughout the year [24], there is a positive relationship between

maximum humidity and temperature with CCHF incidence in different lags. It could be concluded that increases in heat and moisture in a specific month could not predict the number of cases in a later month.

Moreover, this study has demonstrated a reverse relationship between CCHF incidence and average temperature at a lag of 5 months in both univariate and multiple regression SARIMA models. This means that the increasing average temperature in October–November decreased the likelihood of CCHF cases being reported in the following year. A similar finding has previously been observed in Turkey and Bulgaria [10, 25]. The warmer temperatures in winter and shorter annual cold periods limit development and increase mortality of tick stages [26].

Some studies have observed that a decrease in rainfall might produce a more suitable condition for increased tick activity [26, 27]. We also observed such a finding in the univariate analysis of the relationship between CCHF incidence and rainfall with a lag of 1 month. The numbers of cases in the months (even from March to September where rainfall is negligible) with low rainfall were more than in the months with high rainfall.

In the southeast of Iran, most of the patients were reported from March to August and there are few cases in the rest of the year. Therefore, the MSM showed that in both simple and multiple models,  $P_{00}$  and  $P_{11}$  were more than  $P_{10}$  and  $P_{01}$ , respectively. This means the regimen (i.e. outbreak or non-outbreak periods) is not intended to change. On the other hand, in both simple and multiple models,  $P_{10}$  was more than  $P_{01}$ . This means the prolongation of non-outbreak periods, lead to a decrement in outbreak probability during the year. This study showed that LIP is the most important factor in outbreaks in the southeast of Iran. Moreover, a previous study showed that the pattern of CCHF distribution on the other side of the border, i.e. in the Baluchistan region of Pakistan, is somehow different from results of our study, with two annual surges in April and August [28]. On the other hand, the yearly peak in the number of cases in Sistan-va-Baluchistan province of Iran follows the first surge of the disease in Baluchistan province, Pakistan. This is probably due to transmission of infected immature ticks by rodents, small animals and birds [1, 29] and infected mature ticks by livestock imported from Pakistan to Iran [9]. Therefore, we conclude that, if the livestock importation from Pakistan ceased from March to July,

the probability of CCHF incidence and outbreaks would decrease and consequently the outbreak period would be short.

The climate factors do not affect the incidence of CCHF in humans directly, but the effect of climate factors on CCHF outbreaks is through their effect on the vector's (*Hyalomma* ticks) life-cycle. On the other hand, in the southeast of Iran, CCHF is transmitted mainly through contact with blood and body fluids of infected animals during the viraemic phase of disease [30, 31]. There is illegal animal trade and uncontrolled population movements between Pakistan and Afghanistan (through Quetta city) [12] and consequently between Iran and Afghanistan (through Nimroz province); therefore, LIP could influence CCHF incidence in human populations directly in this region. Since the health system is not able to intervene in weather variability, ceasing LIP in spring and summer, or more suitable quarantine in the border areas, could help to prevent outbreaks. It should be noted that, in accord with the results of previous studies, the CCHFV genome isolated from Iranian patients is similar to that from Afghanistan and Pakistan that has a close relationship with the CCHF Matin strain [2, 3, 12, 32, 33]; therefore the migration of CCHFV is free and unrestricted between Iran and Pakistan [34].

We obtained relatively similar results from two different time-series models: SARIMA and MSM. Both methods show a clear association of weather variables and LIP with CCHF disease in the southeast of Iran. However, we found that the MSM allowed for more information about the series and outbreak detection regarding transition probabilities. On the other hand, with respect to goodness of fit and predictive accuracy, the MSM was better than SARIMA. Therefore, the MSM has several advantages compared to the SARIMA model, in particular, its forecasting capability and its richer information on time-related changes; but generally, both MSM and SARIMA modelling are useful for interpreting and applying surveillance data in disease control and prevention.

This study has two major strengths. First, to our knowledge, this is the first time-series study to examine the relationship between major weather variables and LIP with CCHF incidence in a most prevalent area in Iran using valuable time-series models. Our data demonstrate that in addition to LIP, of all climate variables, only mean temperature, maximum relative humidity and rainfall are associated with

CCHF. Second, we compared two common time-series analysis methods and found that a MSM forecasting model appeared to be more suitable than a SARIMA model in the assessment of the relationship between CCHF and some explanatory variables.

Some limitations of this study should also be acknowledged. The occurrence of CCHF is complex; CCHF is not only influenced by weather, but by many other biological, social, and environmental factors such as change in agricultural activities, ranching, nomadic population and illegal livestock trading that might also lead to bias in causal inference in this study.

However, near the borders, the weather conditions in neighbouring countries is the same as in Sistan-Baluchistan province in Iran (data not shown) and the results of this study could be generalized to Pakistan (Baluchistan province) and Afghanistan (Nimroz province).

The findings of this study may assist local public health authorities to utilize the model developed in this study to identify the communities that require particular attention and to mobilize limited resources to effectively control and prevent outbreaks of CCHF during epidemic seasons. These findings may have applications as a decision support tool in planning disease control and risk-management programmes.

## ACKNOWLEDGEMENTS

The authors thank Mr Hasanzehi, Zahedan University of Medical Sciences, Iran, and Mr Mohammadzadeh, Zabol University of Medical Sciences, Iran, for their help in data registering. This work was supported by the research deputy of Tehran University of Medical Sciences (grant no. 12/234).

## DECLARATION OF INTEREST

None.

## REFERENCES

1. Ergönül Ö. Crimean-Congo haemorrhagic fever. *Lancet Infectious Diseases* 2006; **6**: 203–214.
2. Chinikar S, et al. Crimean-Congo hemorrhagic fever in Iran and neighboring countries. *Journal of Clinical Virology* 2010; **47**: 110–114.
3. Chinikar S, et al. Geographical distribution and surveillance of Crimean-Congo hemorrhagic fever in Iran. *Vector-Borne and Zoonotic Diseases* 2010; **10**: 705–708.
4. Vorou R. Crimean-Congo hemorrhagic fever in south-eastern Europe. *International Journal of Infectious Diseases* 2009; **13**: 659–662.
5. Shepherd AJ, et al. Viremic transmission of Crimean-Congo hemorrhagic fever virus to ticks. *Epidemiology and Infection* 1991; **106**: 373–382.
6. Mehravaran A, et al. Molecular detection of Crimean-Congo haemorrhagic fever (CCHF) virus in ticks from southeastern Iran. *Ticks and Tick-borne Diseases* 2013; **4**: 35–38.
7. Yaser SA, et al. Crimean-Congo hemorrhagic fever: a molecular survey on hard ticks (Ixodidae) in Yazd province, Iran. *Asian Pacific Journal of Tropical Medicine* 2011; **4**: 61–63.
8. Gale P, et al. The feasibility of developing a risk assessment for the impact of climate change on the emergence of Crimean-Congo haemorrhagic fever in livestock in Europe: a Review. *Journal of Applied Microbiology* 2010; **108**: 1859–1870.
9. Mostafavi E, et al. Temporal modeling of Crimean-Congo hemorrhagic fever in eastern Iran. *International Journal of Infectious Diseases* 2013; **17**: 524–528.
10. Vescio FM, et al. Environmental correlates of Crimean-Congo haemorrhagic fever incidence in Bulgaria. *BMC Public Health* 2012; **12**: 11–16.
11. Gray J, et al. Effects of climate change on ticks and tick-borne diseases in Europe. *Interdisciplinary Perspectives on Infectious Diseases* 2009; **2009**: 1–12.
12. Alam MM, et al. Genetic analysis and epidemiology of Crimean Congo hemorrhagic fever viruses in Baluchistan province of Pakistan. *BMC Infectious Diseases* 2013; **13**: 1–8.
13. Mardani M, et al. The efficacy of oral ribavirin in the treatment of Crimean-Congo hemorrhagic fever in Iran. *Clinical Infectious Diseases* 2003; **36**: 1613–1618.
14. Keshkar-Jahromi M, et al. Crimean-Congo hemorrhagic fever in Iran. *Antiviral Research* 2013; **100**: 20–28.
15. Razi T, et al. Investigation of meteorological droughts in the Sistan and Baluchistan province, Using the standardized precipitation index and Markov chain model. *Iran-Water Resource Research Journal* 2007; **3**: 25–35.
16. Alavi-Naini R, et al. Crimean-Congo hemorrhagic fever in Southeast of Iran. *Journal of Infection* 2006; **5**: 378–382.
17. Chinikar S, et al. Phylogenetic analysis in a recent controlled outbreak of Crimean-Congo haemorrhagic fever in the south of Iran, December 2008. *Eurosurveillance* 2010; **15**: 1–4.
18. Chinikar S, et al. Surveillance and Laboratory Detection System of Crimean-Congo Haemorrhagic Fever in Iran. *Transboundary and Emerging Diseases* 2008; **55**: 200–204.
19. Helfenstein U. Box-Jenkins modelling in medical research. *Statistical Methods in Medical Research* 1996; **5**: 3–102.
20. Nobre FF, et al. Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Statistics in Medicine* 2001; **20**: 3051–3069.

21. **Lu H-M, Zeng D, Chen H.** Prospective infectious disease outbreak detection using Markov switching models. *Knowledge and Data Engineering, IEEE Transactions* 2010; **22**: 565–577.
22. **McLachlan G, Krishnan T.** *The EM Algorithm and Extensions*. New York: John Wiley & Sons, 2008, pp. 30–45.
23. **Yilmaz GR, et al.** The epidemiology of Crimean-Congo hemorrhagic fever in Turkey, 2002–2007. *International Journal of Infectious Diseases* 2009; **13**: 380–386.
24. **Rahbari S, Nabian S, Shayan P.** Primary report on distribution of tick fauna in Iran. *Parasitology research* 2007; **101**: 175–177.
25. **Estrada-Peña A, et al.** The trend towards habitat fragmentation is the key factor driving the spread of Crimean-Congo haemorrhagic fever. *Epidemiology and Infection* 2010; **138**: 1194–1203.
26. **Estrada-Peña A, et al.** Phenology of the tick, *Ixodes ricinus*, in its southern distribution range (central Spain). *Medical and Veterinary Entomology* 2004; **18**: 387–397.
27. **Estrada-Peña A.** Climate, niche, ticks, and models: what they are and how we should interpret them. *Parasitology Research* 2008; **103**: 87–95.
28. **Sheikh AS, et al.** Bi-annual surge of Crimean-Congo haemorrhagic fever (CCHF): a five-year experience. *International Journal of Infectious Diseases* 2005; **9**: 37–42.
29. **Nateghpour M, et al.** Wild rodents and their ectoparasites in Baluchistan area, southeast of Iran. *Tropical Biomedicine* 2013; **30**: 72–77.
30. **Izadi S, et al.** Crimean-Congo hemorrhagic fever in Sistan and Baluchestan Province of Iran, a case-control study on epidemiological characteristics. *International Journal of Infectious Diseases* 2004; **8**: 299–306.
31. **Naieni KH, et al.** Seroprevalence, incidence and risk factors of Crimean-Congo hemorrhagic fever in Sistanva-Baluchestan province, Iran. *Iranian Journal of Public Health* 2004; **33**: 1–7.
32. **Chinikar S, et al.** Genetic analysis of Crimean-Congo hemorrhagic fever virus in Iran. *Journal of Medical Virology* 2004; **73**: 404–411.
33. **Öschlärer S, et al.** Complete sequence and phylogenetic characterisation of Crimean-Congo hemorrhagic fever virus from Afghanistan. *Journal of Clinical Virology* 2011; **50**: 90–92.
34. **Mild M, et al.** Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. *Journal of General Virology* 2010; **91**: 199–207.