Exploring recent spatial patterns of cutaneous leishmaniasis and their associations with climate in some countries of the Middle East using geographical information systems

Salahuddin M. Jaber[1], Jwan H. Ibbini[2], Nawal S. Hijjawi[3], Nafn M. Amdar[4], Mohammed J. Huwail[5], Khalid Al-Aboud[6]

[1]*Department of Water Management and Environment, Faculty of Natural Resources and Environment, Hashemite University, Zarqa, Jordan;* [2]*Department of Land Management and Environment, Faculty of Natural Resources and Environment, Hashemite University, Zarqa, Jordan;* [3]*Department of Medical Laboratory Sciences, Faculty of Allied Health Sciences, Hashemite University, Zarqa, Jordan;* [4]*Agricultural Scientific Research Center, Ministry of Agriculture, Hims, Syria;* [5]*Directorate of Public Health, Ministry of Health, Baghdad, Iraq;* [6]*Public Health Department, King Faisal Hospital, Meccah, Saudi Arabia*

Abstract. This study explores the spatial trends of cutaneous leishmaniasis (CL) and characterises the relationships between the observed spatial patterns and climate in Jordan, Syria, Iraq and Saudi Arabia in 2009. Areal interpolation revealed the presence of four major hotspots of relatively high incidence rates covering most parts of Syria, central parts of Iraq, and north-western, central, south-eastern and south-western parts of Saudi Arabia. The severity of these hotspots was seen to decrease from high to low latitudes. The spatial patterns could be partly linked to precipitation (the higher the precipitation, the higher the incidence rates) and to a lesser degree to temperature (the lower the temperature, the higher the incidence rates). No significant relationship was deduced between the observed spatial patterns of incidence rates and humidity. However, these three climatic factors could be used jointly as explanatory variables (*ceteris paribus*) to explain part of the spatial variations of the CL incidence rates in the study area by applying geographically weighted regression.

Keywords: cutaneous leishmaniasis, areal interpolation, geographically weighted regression, geographical information systems, Middle East, Jordan.

Introduction

Leishmaniasis, a vector-borne parasitic disease caused by *Leishmania* parasites, is spread by the bite of an infected female *Phlebotomine* sand fly, previously having fed on infected humans or reservoir hosts such as rodents. Around 30 different *Leishmania* species are capable of infecting mammals and 21 of them can cause infection in humans. Depending on the infecting species, the clinical manifestations are visceral, cutaneous or mucocutaneous. Of these, *L. donovani*, *L. infantum* and *L. chagasi* cause visceral leishmaniasis (VL or kala-azar); *L. tropica*, *L. major* and *L. aethiopica* cause cutaneous leishmaniasis (CL); and *L. mexicana*, *L. amazonesis* and *L. venezuelensis* cause mucocutaneous leishmaniasis (ML). The two main clinical forms are VL, which (if left untreated) is a generally fatal, severe systemic disease and CL, which causes skin sores. Although the latter is self-healing, scarring can result in serious, social and psychological stigma (CDC, 2011; WHO, 2011a).

Worldwide, more than 10 million people suffer from CL in 82 tropical countries and there are about 1.5 million new cases every year (WHO, 2011a). In the Middle East and North Africa (MENA), CL is regarded as an endemic disease. More than 350,000 new cases of CL occur annually and this trend is increasing in many MENA countries (WHO, 2011a). However, despite the worldwide spread and the severity of CL, it is still a neglected public health problem (Hotez et al., 2012). This motivated the World Health Organization (WHO) to launch an initiative to control the disease using an integrated approach. One recommendation was to map endemic areas and identify factors affecting the spread of the disease to facilitate interventions and predict epidemics (WHO, 2007).

In order to establish a basis for studies addressing the spatial and temporal distribution of CL and

Corresponding author:
Salahuddin M. Jaber
Department of Water Management and Environment
Faculty of Natural Resources and Environment
Hashemite University
P.O. Box 330010, Postal Code 13133, Zarqa, Jordan
Tel. +962 79 903 7361
E-mail: sjaber@hu.edu.jo

exploring the most common and relevant environmental, social and economic factors affecting the spread of the disease, we decided to first characterise the recent spatial trends of CL in Jordan, Syria, Iraq and Saudi Arabia, which – to the best of our knowledge – represents a first attempt in this direction in these four countries. Since it was hypothesised that the incidence of CL is sensitive to environmental factors affecting the distribution and behaviour of both vectors and reservoir hosts, in particular climate, the second major objective was to reveal the relationships between discerned spatial patterns of CL in the area and some common and potentially relevant climatic factors (*ceteris paribus*) represented by temperature, precipitation and humidity.

In the above mentioned four countries, CL is considered to be endemic even if there are only sporadic outbreaks at different time intervals in different localities (Adler and Theodor, 1929; Douba et al., 1997; Al-Tawfiq and AbuKhamsin, 2004; AlSamarai and AlObaidi, 2009; Mosleh et al., 2009). Two forms of CL are present (Saliba et al., 1994; Janini et al., 1995; WHO, 2008; Postigo, 2010; WHO, 2011b), the first of which is zoonotic as the parasite is transmitted by the vector from a range of animals to humans. It is caused by *L. major* where the fat sand rat (*Psammomys obesus*) is the suggested major reservoir host and the sand fly (*Phlebotomus papatasi*) the

major vector. The second form is anthroponotic with the parasite (in this case *L. tropica*) spread from human to human by the suggested major vector *P. sergenti* sand fly.

Materials and methods

The analyses were carried out in a geographic information systems (GIS) environment using the GIS software ArcGIS version 10.1 (ESRI, 2012a), the spatial analysis software SAM version 4.0 (Rangel et al., 2010) and the statistical software JMP version 8.0 (JMP, 2008a). All data were projected to the Universal Transverse Mercator (UTM) Zone 37 N coordinate system.

Study area

The study area (Figs. 1 and 2) comprises Jordan, Syria, Iraq and Saudi Arabia. It is located between latitudes 16°00' N and 38°00' N and longitudes 34°00' E and 56°00' E with a total area of about 2.7 million km² and a total population (according to 2009 estimates) of about 83.4 million (CBSSY, 2011; CDOSISA, 2011; COSIR, 2011; DOSJO, 2011). The largest country is Saudi Arabia (about 2.0 million km²) with a population of about 25.4 million, while the smallest country is Jordan (about 88,800 km²) with a population of
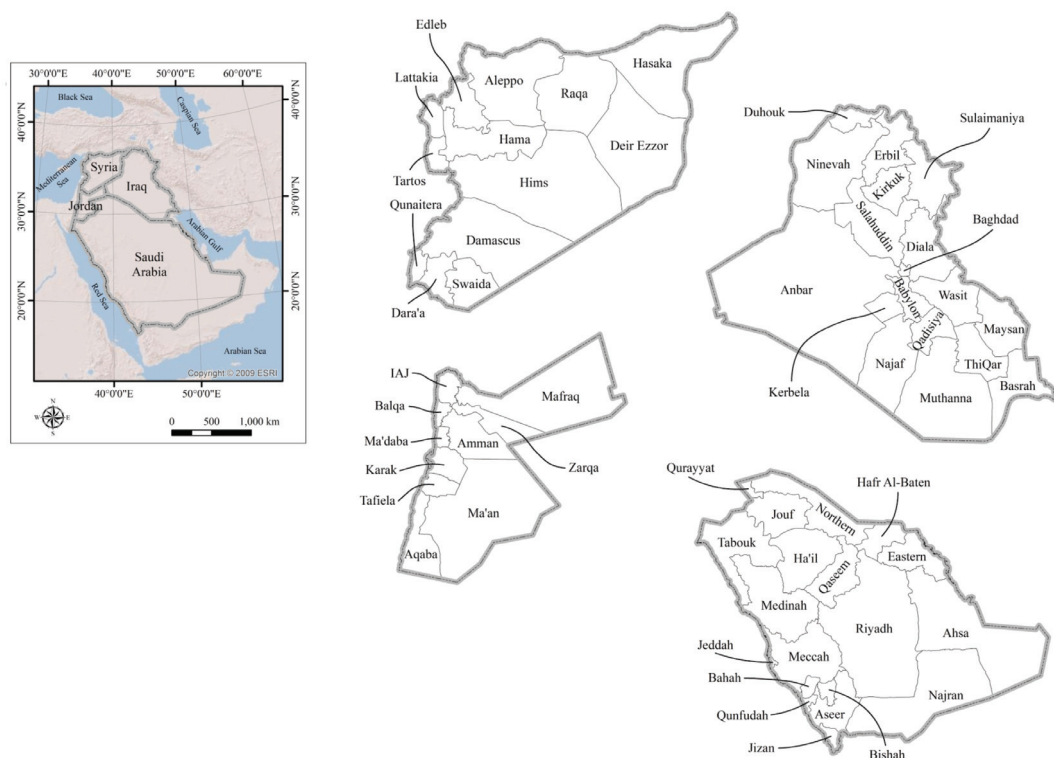

Fig. 1. The study area showing the 60 administrative units of Jordan, Syria, Iraq and Saudi Arabia.
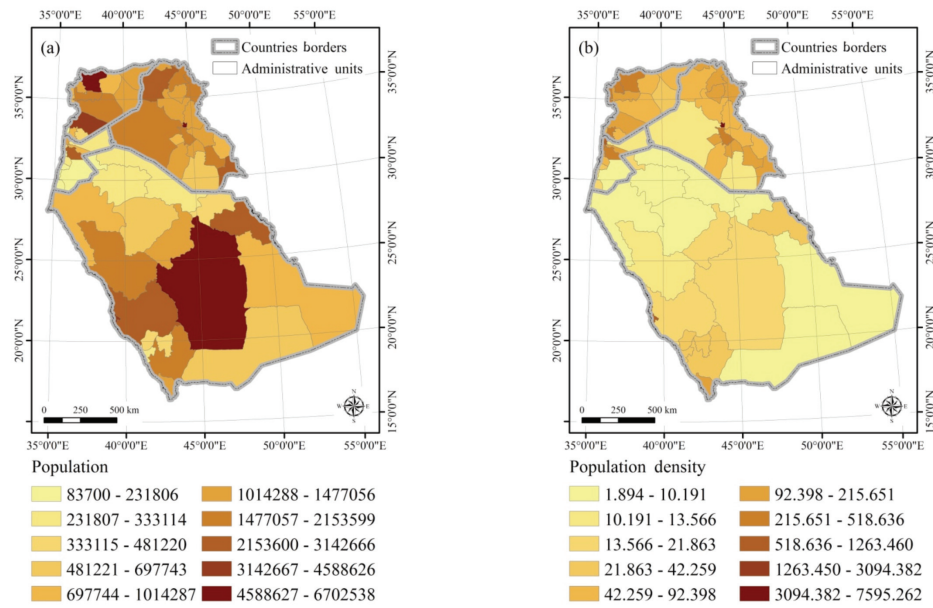
Fig. 2. Choropleth thematic maps showing the spatial distribution of (a) the population and (b) the population density of the 60 administrative units of Jordan, Syria, Iraq and Saudi Arabia in 2009.

about 6.0 million. Iraq and Syria occupy about 438,000 km² and 187,400 km² and embrace about 31.7 million and 20.4 million inhabitants, respectively. Syria has the largest population density (about 108.7 inhabitants per km²) followed by Iraq (about 72.3 inhabitants per km²). Jordan has about 67.3 inhabitants per km² and Saudi Arabia has the smallest population density (about 13.0 inhabitants per km²).

At the governorate level, the study area consists of 60 polygon administrative units (AUs): 10 in Jordan,

13 in Syria, 18 in Iraq and 19 in Saudi Arabia. The largest AU is Riyadh in Saudi Arabia with an area of about 380,000 km², while Baghdad in Iraq with approximately 882 km² is the smallest. Baghdad has not only the highest population (about 6.7 million) but also the highest population density (about 7,600 inhabitants per km²). With about 83,700 inhabitants Tafiela in Jordan has the lowest number of people, while Najran in Saudi Arabia has the lowest population density (about 1.9 inhabitants per km²).
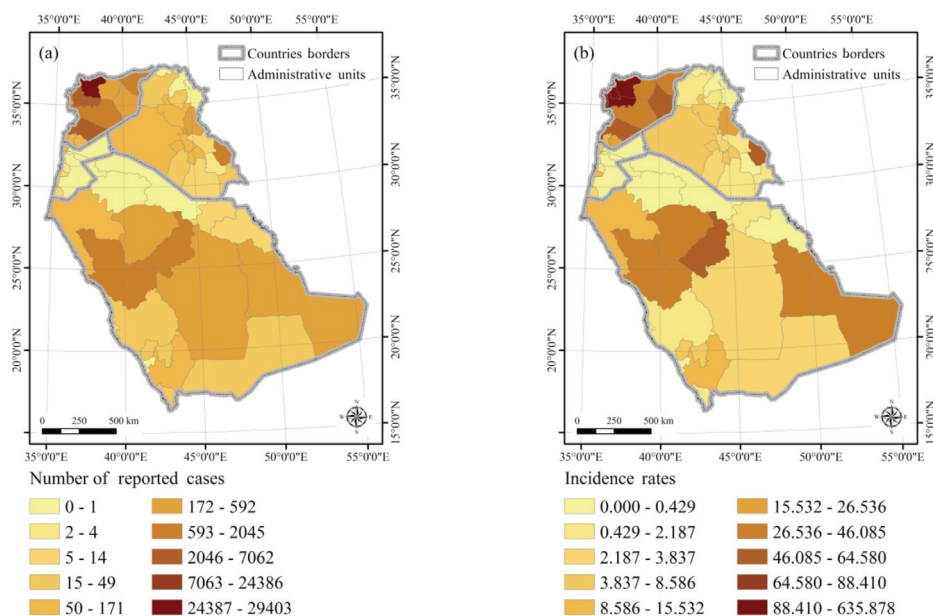

Fig. 3. Choropleth thematic maps showing the spatial distribution of (a) the number of reported cases of CL and (b) the incidence rates of CL of the 60 administrative units of Jordan, Syria, Iraq and Saudi Arabia in 2009.

CL data

Reported CL cases counts in the 60 AUs for the year 2009 (Fig. 3a) were obtained from the published governmental reports of the Ministry of Health (MoH) of Jordan (MOHJO, 2011), MoH of Syria (MOHSY, 2011), MoH of Iraq (MOHIR, 2011) and MoH of Saudi Arabia (MOHSA, 2011). A total of 51,119 CL cases were reported in Jordan, Syria, Iraq and Saudi Arabia in 2009. This represented about 0.06% of the total population of these four countries. The majority of the cases (about 90.7%) were reported in Syria (46,348 cases), which represented about 0.23% of the total population of Syria. The lowest number of cases (148) was reported in Jordan, which represented about 0.002% of the total population of Jordan and about 0.30% of all cases reported in the four countries. In Saudi Arabia and Iraq, similar numbers were reported, i.e. 2,549 in Saudi Arabia and 2,074 in Iraq representing about 5.0% (for Saudi Arabia) and about 4.1% (for Iraq) of the total number of cases reported in the four countries and about 0.01% of the total population of Saudi Arabia and about 0.01% of the total population of Iraq. Furthermore, the highest number of CL cases (29,403) was reported in Aleppo in north Syria. In Saudi Arabia and Iraq, the highest numbers of cases (654 in Saudi Arabia and 596 in Iraq) were reported in Qaseem in the central parts of Saudi Arabia and Maysan in the south-eastern parts of Iraq. In Jordan, the highest number of cases (81) was reported in IAJ along the northern borders of the country. However, 12 AU in the study area reported zero cases; five of them (Ma'an, Ma'daba, Amman, Zarqa and Mafraq) in Jordan, four of them (Qurayyat, Northern, Jouf and Qunfudah) in Saudi Arabia and the remaining three (Duhouk, Erbil and Sulaimaniya) in Iraq. In Syria, no AUs reported zero cases.

The reported CL cases counts were used to calculate the incidence rates for each AU (Fig. 3b) using equation 1 below:

$$IR09 = CL09 / P09 \times 100000$$
$$(equation\ 1)$$

where IR09 is the incidence rate in 2009, CL09 the number of CL cases reported in 2009 and P09 the population in 2009. The incidence rates of CL varied from zero in the 12 AUs, which reported zero cases as has been noticed earlier, to about 635.9 in Aleppo in Syria with mean and standard deviation (SD) of about 37.4 and 101.5, respectively. The highest incidence rates in Iraq and Saudi Arabia were found in Maysan (about 64.6) and Qaseem (about 57.6), respectively. In Jordan, the highest incidence rate was found in Tafiela (about 15.5).

Climate data

Data with respect to temperature, precipitation and specific humidity for the year 2009 in all four countries (Fig. 4) were obtained from the NCEP/NCAR Reanalysis 1 Project available at the Climate Research Unit at the University of East Anglia, United Kingdom (CRU, 2013). First, daily air temperature at 2 m above the ground (in Kelvin degrees), daily precipitation rate at the surface (in mm/sec) and daily specific humidity at 2 m above the ground (in kg/kg) for the Eurasian Quarter-Sphere (i.e. between latitudes 00°00' N and 90°00' N and longitudes 90°00' W and 90°00' E) for
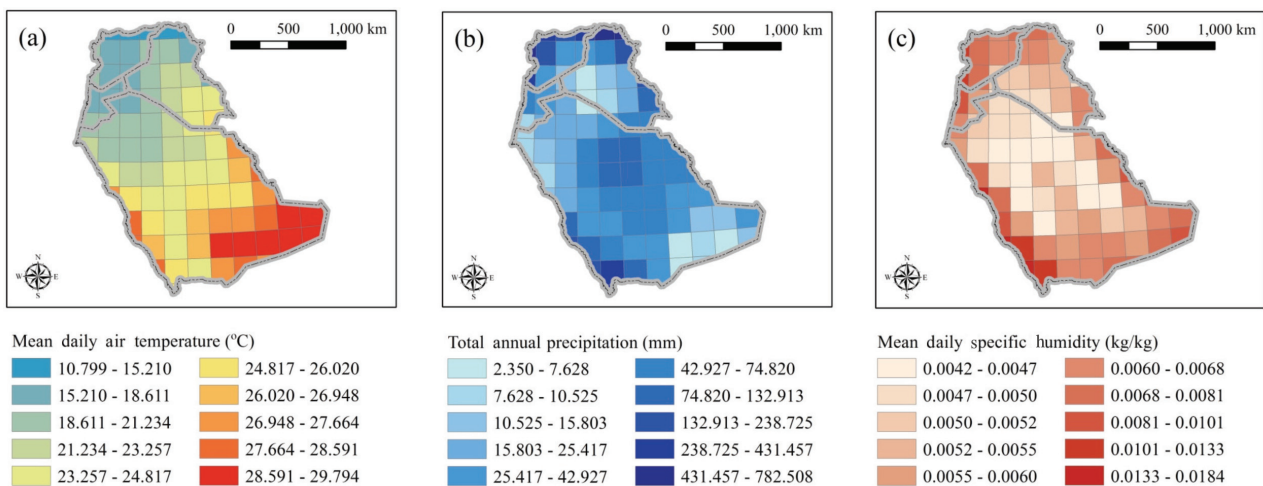

Fig. 4. Choropleth thematic maps showing the spatial distribution of (a) the mean daily air temperature (°C), (b) the total annual precipitation (mm) and (c) the mean daily specific humidity (kg/kg) of the study area in 2009.

the year 2009 were obtained as vector layers for a Gaussian grid based on spatial units of regular discrete polygons with sizes of about 1.9° × 1.9°, here called climate factor units (CFUs). For every CFU, the daily air temperature at 2 m above the ground was transformed into centigrades (°C) and their annual means calculated; the daily precipitation rate at the surface converted into mm/day and the total annual precipitation in mm obtained; and the means of the daily specific humidity at 2 m above the ground in kg/kg calculated. Finally, the three processed climatic factor layers were projected into the UTM Zone 37 N coordinate system and clipped according to the borders of the study area.

The climate data for the study area, divided into 93 CFUs, showed a relatively simple pattern of mean daily air temperatures (T) (Fig. 4a). It decreased gradually from the south-east to the north-west in the study area from about 29.8 °C to 10.8 °C with mean and SD of about 23.3 °C and 4.5 °C, respectively. However, the spatial patterns of precipitation and specific humidity were more complex. The total annual precipitation (P) (Fig. 4b) varied from 2.4 mm to 782.5 mm with mean and SD of about 99.1 mm and 156.6 mm, respectively, reaching maximum values in the northern parts of Syria and Iraq and in the south-western parts of Saudi Arabia. Areas with high annual precipitation values were also noticed in the north-eastern parts of Saudi Arabia. The mean daily specific humidity (S) (Fig. 4c) varied from 0.004 to 0.018 kg/kg with mean and SD of about 0.007 kg/kg and 0.003 kg/kg, respectively. Low values were noticed in the central parts of the study area with high values along the peripheries.

Spatial trends of CL and their association with climate

The first major objective of this study was to produce a heat map (i.e. a continuous raster surface) to better show the spatial variability and the "hotspots" (areas with high-value spatial clusters) and "coldspots" (areas with low-value spatial clusters) for CL incidence rates that might not be readily recognisable when displaying the raw data. However, the CL incidence rates were associated with the 60 AUs and thus regarded as areal data (Waller and Gotway, 2004). The second major objective was to reveal the types of relationships between the discerned spatial patterns of CL incidence rates in the area and temperature, precipitation and specific humidity. However, these three climatic variables were associated with 93 CFUs. Therefore, before commencing with the analyses, it was necessary to unify the spatial units upon which the CL incidence

rates and the three climatic variables were based and this was done by areal interpolation.

Originally, areal interpolation was devised to deal with "basis change" (Goodchild et al., 1993), which involves using known attribute values for a set of (source) areal units to predict unknown attribute values for another set of (target) areal units, where the areas are generally not nested hierarchically and the boundaries between the areas generally do not coincide (Guan et al., 2011). However, areal interpolation in the ArcGIS Geostatistical Analyst extension (ESRI, 2012b) is a geostatistical interpolation technique that extends kriging theory to data aggregated over areal units using a two-step process. First, a smooth prediction surface for the CL incidence rates was created from the 60 AUs, which were regarded as the source areal units (a surface that can often be interpreted as a risk surface). Then, the prediction surface was aggregated back to the 93 CFUs, which were regarded as the target areal units.

Although normality of the data is not a prerequisite for interpolation or regression, they work better and generate good results when the data are normally distributed. The incidence rates were not normally distributed as indicated by having a relatively high positive skewness and high positive kurtosis values of about 4.6 and 22.8, respectively. This could be accentuated by their normal quantile plot (Fig. 5a). Consequently, in order to remedy the problem a natural logarithm transformation could be applied, but it only works when the data have only positive and non-zero values and the incidence rates do have zero values. One common approach to handle this problem is to translate the data by adding a constant, which might be a very small number such as "one" before applying the transformation. Therefore, the translated incidence rates and the natural logarithm transformation of the translated incidence rates for the 60 AUs were calculated using equations 2 and 3 below:

$$TIR09 = (CL09 + 1) / P09 \times 100000$$
(equation 2)

$$LTIR09 = lnTIR09$$
(equation 3)

where TIR09 is the translated incidence rate of the AU in the year 2009, LTIR09 the natural logarithm of the translated incidence rate of the AU in the year 2009 and CL09 and P09 what has been explained earlier.

The values of LTIR09 varied from -3.1 to 6.5 with mean and SD of about 1.5 and 2.4, respectively. They
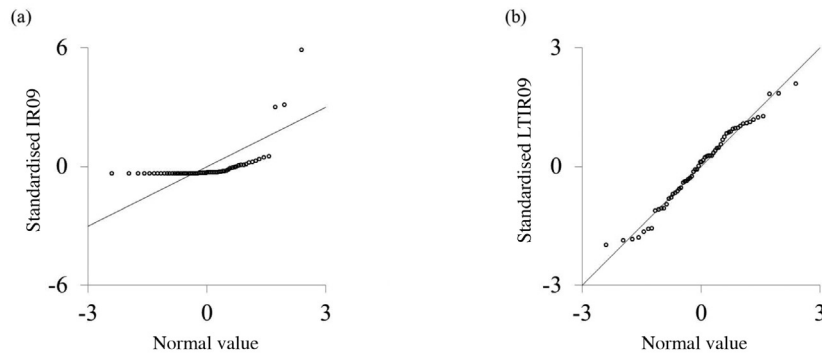
Fig. 5. Normal quantile plots for (a) IR09 and (b) LTIR09.

showed a normal distribution indicated by a close to zero skewness of about -0.2 and kurtosis of about -0.6. This could be emphasised by the almost straight line on the normal quantile plot (Fig. 5b).

Creating the LTIR09 risk surface involves developing an empirical semivariogram expressing the spatial dependence between neighbouring observations. It can be defined as one-half the variance of the difference between the attribute values at all points separated by a lag h. It was calculated using equation 4 below:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

$$\text{(equation 4)}$$

where $\hat{\gamma}(h)$ is the estimated semivariance for the lag h, $N(h)$ the total number of pairs of data points separated by the lag h and $Z(x_i)$ and $Z(x_i + h)$ the actual values of the variable Z at x_i and $x_i + h$ locations, respectively (Isaaks and Srivastava, 1989).

Then, the mathematical functions circular, spherical, tetraspherical, pentaspherical, exponential, Gaussian, rational quadratic, hole effect, K-Bessel, J-Bessel and stable were used to fit the developed empirical semivariogram. Using the leaving-one-out cross-validation method (Davis, 1987) by sequentially omitting a point, predicting its value using the rest of the data and comparing the measured and predicted values, the model producing the lowest root mean squared error (RMSE) was selected as the best and was consequently used to generate the LTIR09 risk surface. This model, in addition of possessing a strong correlation between measured and predicted values, should have normally distributed independent random errors with mean zero and constant variance. By back-transforming the developed LTIR09 risk surface, a TIR09 risk surface predicting TIR09 at each individual location in the study area was developed.

Completing the workflow of areal interpolation through the ArcGIS Geostatistical Analyst extension rendered the LTIR09 and TIR09 risk surfaces at the 93 CFUs. The developed vector layers, which were at the scale and extent required to demonstrate the spatial relationships of CL and the climate across the study area for the year 2009 by applying correlation analysis using Pearson product-moment correlation coefficient (r) and multiple linear regression analysis using ordinary least squares (OLS) and geographically weighted regression (GWR), were here called "base-changed predicted LTIR09" and "base-changed predicted TIR09", respectively.

The Pearson product-moment correlation coefficient (JMP, 2008b) measures the strength of the linear relationship between two variables. If there is an exact linear relationship, the correlation is +1 or -1, depending on whether the variables are positively or negatively related. If there is no linear relationship, the correlation tends toward zero.

Multiple linear regression (Chatterjee and Hadi, 2012) quantifies the relationship between a dependent (response) variable and multiple independent (explanatory) variables. The model for a population of interest has the form expressed by equation 5 below:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

$$\text{(equation 5)}$$

where y_i denotes the response variable, which is a function of the linear combination of the explanatory variables x_{ki}; β_0 represents the regression parameter intercept, while β_k denotes the regression parameter of the variable x_{ki}. The factor ε_i is a normally distributed independent random error term with constant variance and zero mean. The subscript k is the number of explanatory variables in the model, while the subscript i denotes that this relationship holds for each observation in the population of interest.

Usually the method of OLS is used to find the estimates b_o, b_1, b_2, ..., b_k of the parameters β_0, β_1, β_2, ...,
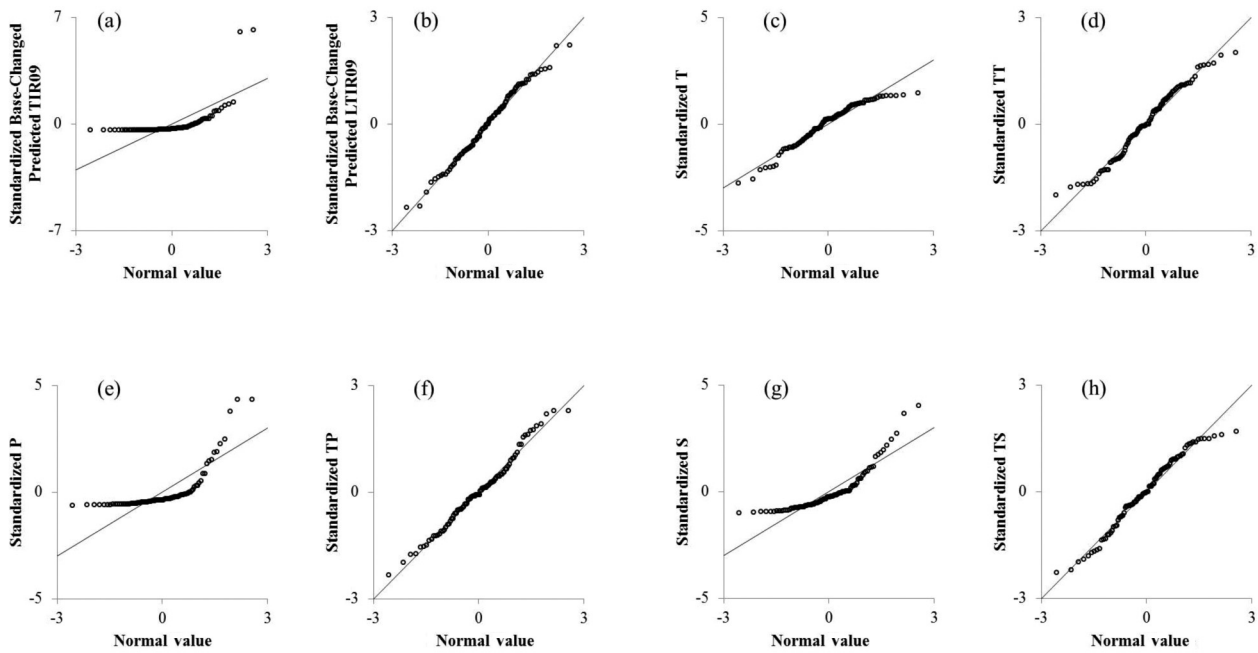
Fig. 6. Normal quantile plots for the (a) base-changed predicted TIR09 and (b) base-changed predicted LTIR09 and the climatic factors (c) mean daily air temperature (T), (d) transformed mean daily air temperature (TT), (e) total annual precipitation (P), (f) transformed total annual precipitation (TP), (g) mean daily specific humidity (S), and (h) transformed mean daily specific humidity (TS) for the 93 climate factor units of the study area in 2009.

β_k from a sample from the population such that the sum of squared residuals ($\Sigma_{i=1}^{n} e_i^2 = \Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2$) across the entire sample is minimised, where $\hat{y}_i$ is the predicted value of y_i and e_i is an estimate of the error term ε_i.

On the other hand, GWR (Fotheringham et al., 2002) is a spatial modelling technique hypothesises that the traditional multiple regression is global and aspatial and cannot explain the relationships between some sets of variables where spatial non-stationarity exists. In contrast, GWR allows different relationships to exist at different points in space so that the parameters estimates in the model rather than being global are specific to those points in space. The model has the form expressed by equation 6 below:

$$y_i = \beta_{0(u_i, v_i)} + \beta_{1(u_i, v_i)} x_{1i} + \beta_{2(u_i, v_i)} x_{2i} + ... + \beta_{k(u_i, v_i)} x_{ki} + \varepsilon_i$$
(equation 6)

where y_i denotes the response variable, which is a function of the linear combination of the explanatory variables x_{ki}s; (u_i, v_i) denotes the coordinates of the ith point in space; $\beta_{0(u_i, v_i)}$ represents the regression parameter intercept for the ith point with coordinates (u_i, v_i), while $\beta_{k(u_i, v_i)}$ denotes the regression parameter of the variable x_{ki} for the ith point with coordinates (u_i, v_i); ε_i a normally distributed independent random error term with constant variance and zero mean; and the subscript k the number of explanatory variables in the model.

Geographically weighted regression produces unique parameters estimates for all points by spatially weighting the observations according to their proximity to i. Observations closer to the ith point are given more weight than are observations further away. The weights are derived through a distance-decay function. A spatial kernel (either fixed or adaptive) is used at the ith point. The kernel bandwidth is determined by either cross-validation or corrected Akaike information criterion (AIC_c) minimization using all or sample (percentage) of the data points. Geographically weighted regression also provides local goodness-of-fit measures and local residuals.

Before conducting regression analysis, it was noticed that the distribution of the base-changed predicted TIR09 was not normally distributed while the distribution of the base-changed predicted LTIR09 was normally distributed as indicated by the normal quantile plots (Figs. 6a and 6b). Hence, the response was chosen to be the base-changed predicted LTIR09. Furthermore, the normality of the explanatory variables T, P and S were tested and it was clear that none of them were normally distributed (Figs. 6c, 6e and 6g). After a process of trial and error and taking into consideration the types and characteristics of the distributions of the explanatory variables, it was found that applying a natural logarithm transformation on the reflected version of the explanatory variable T

(equation 7), natural logarithm transformation on the explanatory variable P (equation 8) and inverse transformation of the explanatory variable S (equation 9), produced transformed versions of the explanatory variables (i.e. TT, TP and TS) that were close to normal distributed (Figs. 6d, 6f and 6h).

$$TT = \ln(34 - T) \qquad \text{(equation 7)}$$

$$TP = \ln P \qquad \text{(equation 8)}$$

$$TS = 1 / S \qquad \text{(equation 9)}$$

Hence, a pool of competing models using OLS and GWR were developed for explaining the relationships between the observed spatial patterns of the base-changed predicted LTIR09 for the 93 CFUs (as the response variable) and different combinations of the variables T, P and S and their transformed versions TT, TP and TS (as the explanatory variables). The main goal of including transformed versions of the explanatory variables was to test the effect of normalising the explanatory variables T, P and S on enhancing the explanation powers of the developed models. Using the leaving-one-out cross-validation method (as has been explained earlier), the model with the smallest cross validation AIC_c was deemed to be the best fit to the data and therefore selected for explaining the observed spatial variations of the CL incidence rates in the study area. Furthermore, the chosen model, in addition to having strong correlation between meas-

ured and predicted values, should have normally distributed independent random errors with mean zero and constant variance.

Results

As the CL risk surface was being created, it was found that sorting the models developed for LTIR09 by fitting the functions circular, spherical, tetraspherical, pentaspherical, exponential, Gaussian, rational quadratic, hole effect, K-Bessel, J-Bessel and stable to the calculated empirical semivariogram (where the default parameters for the lattice spacing (103,488 m), lag size (191,013 m), number of lags (12) and standard search neighbourhood (with 15 and 10 maximum and minimum neighbors, respectively) were used) in an ascending order according to their cross-validation RMSEs, showed that the model using the exponential function was the best (Table 1). The model had the lowest RMSE of about 1.707 and the strongest correlation between measured and predicted values of about 0.690 (P-value <0.001) and the errors associated with the model showed close to normal distribution, as indicated by having almost straight line on normal quantile plot (Fig. 7). Moreover, these errors showed no clear signs of bias or heteroscedasticity, a fact that was demonstrated by having errors mean (ME) close to zero, i.e. about -0.017 (Table 1), which varied from -4.038 to 3.713 with almost independent random distribution and constant variance (Fig. 8). Furthermore, all the developed models had

Table 1. Cross-validation of the models obtained for the LTIR09 of CL in the study area using areal interpolation (sorted in an ascending order according to the RMSE).

Model type	Error					Correlation between measured and predicted	
	MINE	MAXE	ME	RMSE	RMSSE	r	P-value
Exponential	-4.038	3.713	-0.017247	1.707109	1.100897	0.6898	<0.001
Pentaspherical	-4.922	4.226	0.008819	1.760021	1.133990	0.6722	<0.001
Tetraspherical	-4.993	4.356	0.015244	1.774631	1.145186	0.6669	<0.001
Circular	-4.835	4.788	0.000738	1.790520	1.172026	0.6644	<0.001
Rational Quadratic	-4.147	4.649	0.002950	1.794570	1.167782	0.6616	<0.001
Spherical	-5.078	4.544	0.022236	1.796037	1.162707	0.6601	<0.001
K-Bessel	-3.573	4.516	0.086311	1.805258	1.251794	0.6475	<0.001
Gaussian	-3.822	4.659	0.100102	1.886168	1.497117	0.6219	<0.001
Stable	-3.822	4.659	0.100102	1.886168	1.497303	0.6219	<0.001
J-Bessel	-5.182	6.451	0.116890	2.222089	2.826125	0.4752	<0.001
Hole Effect	-6.953	6.372	0.098221	2.637093	4.222851	0.3176	0.013

MINE = minimum error; MAXE = maximum error; ME = mean error; RMSE = root mean squared error; RMSSE = root mean squared standardised error; r = Pearson product-moment correlation coefficient.
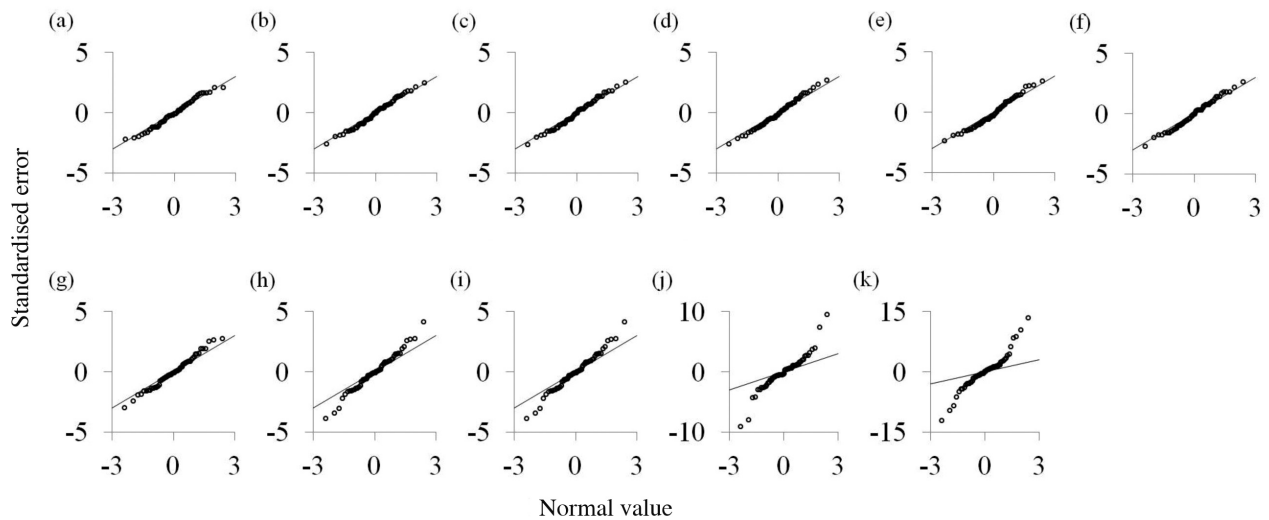
Fig. 7. Cross-validation normal quantile plots for the errors associated with the models developed for the LTIR09 of CL in the study area by applying areal interpolation using various functions. (a) exponential; (b) pentaspherical; (c) tetraspherical; (d) circular; (e) rational quadratic; (f) spherical; (g) K-Bessel; (h) Gaussian; (i) Stable; (j) J-Bessel; (k) hole effect.

mean squared standardised errors (RMSSEs) greater than one, which indicated that they under-estimated the variability of the actual surface of LTIR09. This under-estimation was the lowest in the best model since it had the lowest RMSSE, which was the closest to one. Hence, this model was chosen to produce a continuous raster surface predicting LTIR09 at each individual location in the study area (with a cell size of about 8.693×8.693 km²). The developed surface showed strong anisotropy as indicated by having a semivariogram with the major range (1,149,984 m) three times larger than the minor range (384,670 m). This anisotropy changed more rapidly in the northeast-southwest direction (the minor range) than the

northwest-southeast direction (the major range); the direction was found to be about 106.9°. Furthermore, in addition to the cross-validation measures of error that were calculated for the chosen model and in order to further assess the accuracy of the generated surface, a cross-validation standard error surface was developed (Fig. 9a). This surface quantified the uncertainty associated with each location in the interpolated surface. Generally, the uncertainty decreased in the areas surrounding the centroids of the 60 AUs. Also, the smaller the area of the AU, the lower the uncertainty of the developed surface; a fact that could be noticed in the AUs forming the northern parts of the study area, which included Jordan, Syria, Iraq and the
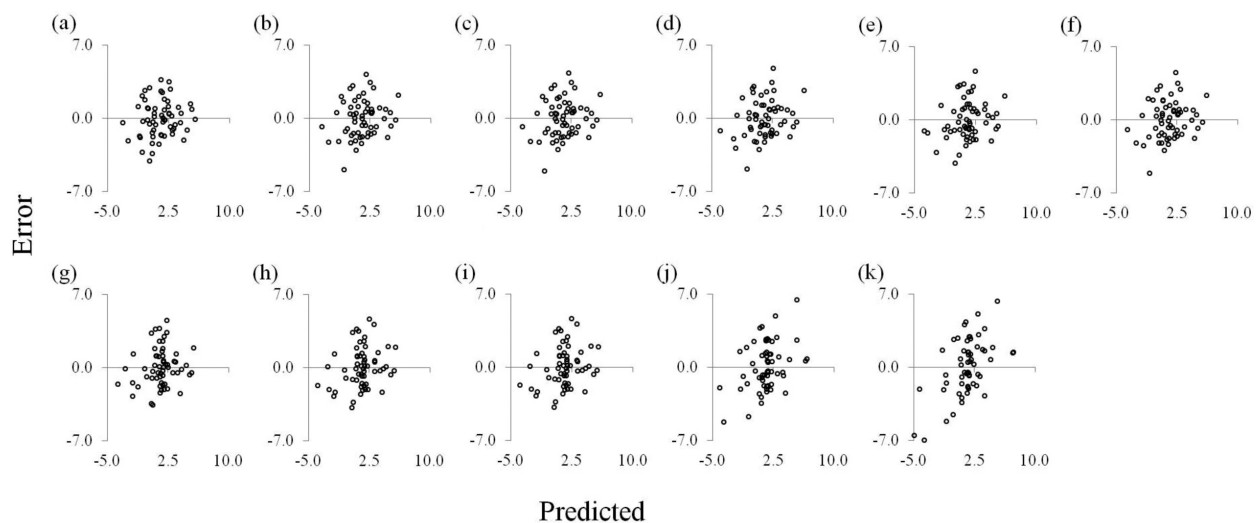

Fig. 8. Cross-validation error plots for the models that were developed for the LTIR09 of CL in the study area by applying areal interpolation using various functions. (a) exponential; (b) pentaspherical; (c) tetraspherical; (d) circular; (e) rational quadratic; (f) spherical; (g) K-Bessel; (h) Gaussian; (i) Stable; (j) J-Bessel; (k) hole effect.
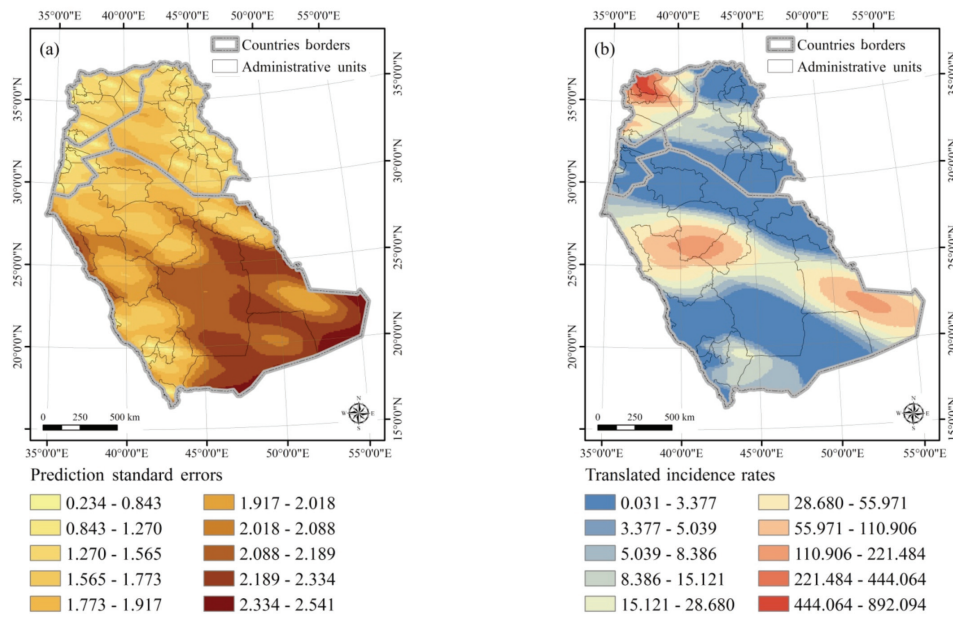
Fig. 9. Continuous raster surfaces showing the spatial distribution of (a) the prediction standard errors for the interpolated surface for the LTIR09 of CL in the study area in 2009 using the exponential function by applying areal interpolation and (b) the TIR09 of CL for the study area in 2009 after back-transforming the interpolated surface for the LTIR09 of CL.

northern and south-western parts of Saudi Arabia.

Fig. 9b shows the TIR09 risk surface after back-transforming the LTIR09 risk surface. In the study area, TIR09 was found to vary from 0.031 to 892.094 with mean and SD of about 24.909 and 55.080, respectively. Moreover, the surface revealed the presence of a spatial pattern of four major hotspot areas of TIR09. The first hotspot covered most parts of the AUs of Syria in the north-western parts of the study area. This hotspot extended (with relatively lower TIR09 severity) through the central parts of Iraq in the north-eastern parts of the study area, which included most of the Anbar and Salahuddin in addition to Diala and Maysan. The second hotspot covered the north-western and central AUs in Saudi Arabia, which included the southern parts of Tabouk and most of Medinah, Ha'il and Qaseem. The third hotspot covered most of Ahsa in the south-eastern parts of Saudi Arabia. These last two hotspots formed a stripe, which extended in a northwest-southeast direction in the central parts of the study area. In addition, a fourth hotspot (with relatively lower TIR09 severity) exited in the south-western AUs of Saudi Arabia, which included some parts of Baha, most parts of Bishah and Aseer and the western parts of Najran. In Jordan, a relatively small hotspot (with very low TIR09) appeared in Tafiela (along the Dead Sea). The remaining parts of the study area formed relatively coldspot areas. The first coldspot appeared in the north-eastern parts of the study area. It covered the Iraqi Kurdish

AUs (Dohouk, Erbil and Sulaimaniya) in addition to Ninevah and Kirkuk (and the topmost eastern parts of Syria). The second coldspot formed a stripe in the direction extending from the central and eastern parts of Jordan to the northern and north-eastern parts of Saudi Arabia in addition to the southern parts of Iraq. The third coldspot also formed a stripe approximately directed northwest-southeast, which extended from Jeddah, Meccah and Qunfudah AUs in the western parts of Saudi Arabia and covered most of the southern parts of Riyadh and eastern parts of Najran along the southern parts of the study area.

Fig. 10 shows the base-changed predicted TIR09 after completing the workflow of areal interpolation through the ArcGIS Geostatistical Analyst extension. It is clear that the base-changed predicted TIR09 for the 93 CFUs varied from 0.05 to 510.0 with mean and SD of about 31.5 and 77.4, respectively.

By applying Pearson product-moment correlation coefficient (r), it was found that P had the highest significant positive correlation with base-changed predicted TIR09 ($r = +0.414$ and P-value <0.001) followed by T, which had a significant negative correlation with base-changed predicted TIR09 ($r = -0.285$ and P-value $= 0.006$). However, no significant correlation was noticed between S and the base-changed predicted TIR09 ($r = +0.039$ and P-value $= 0.714$).

Applying OLS with different combinations of the explanatory variables T, P and S and their transformed versions TT, TP and TS for explaining the base-
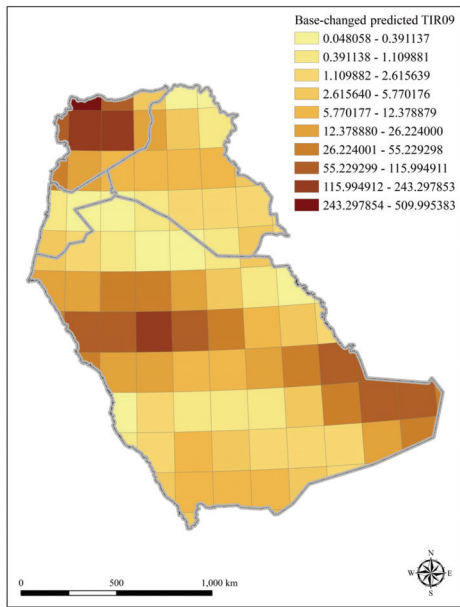
Fig. 10. Choropleth thematic map showing the spatial distribution of base-changed predicted TIR09 of CL for the 93 climate factor units of the study area in 2009 using areal interpolation.

changed predicted LTIR09 for the 93 CFUs (i.e. Models OLS01 to OLS08 in Table 2) showed that, although the developed models had cross-validation errors distributions that were close to normal (Figs. 11a-h), there were noticeable structures and patterns in their cross-validation errors plots (Figs. 12a-h) indicating non-randomness and non-independence of the errors with non-constant variances. Moreover, the explanation powers of the models were very weak (R^2 varied from 0.007 to 0.018 and adjR^2 varied from -0.004 to -0.016) and not significant (F Ratios varied from 0.194 to 0.539 and the P-values varied from 0.657 to 0.900).

Applying GWR for explaining the base-changed predicted LTIR09 for the 93 CFUs using different combinations of the explanatory variables T, P and S and their transformed versions TT, TP and TS (i.e. models GWR01 to GWR08 in Table 2), using a moving window spatial weighting function and adaptive spatial kernel with optimization to minimize AIC_c searching from 5 to 20% of neighbours, produced better results as compared to those that were obtained from applying OLS. In addition to having models cross-validation errors distributions that were close to normal as indicated by their normal quantile plots (Figs. 11i-p), the structures and patterns in the errors distributions were reduced as indicated by their errors plots (Figs. 12i-p) indicating stronger evidences of randomness and independence of the errors with almost constant variances and close to zero means. Moreover, the explanatory powers of the models were enhanced dramatically (R^2 varied from 0.743 to 0.760 and adjR^2 varied from 0.662 to 0.681) and were significant (F Ratios varied

Table 2. Cross-validation of the models developed for the base-changed predicted LTIR09 of CL for the 93 climate factor units of the study area.

Model no.	Predictors	ENP	AIC$_c$	R^2	adjR^2	F Ratio	P-value
OLS01	T, P, S	4	403.838	0.018	-0.004	0.531	0.662
OLS02	TT, P, S	4	404.459	0.011	-0.011	0.330	0.804
OLS03	T, TP, S	4	404.711	0.008	-0.014	0.249	0.862
OLS04	T, P, TS	4	403.813	0.018	-0.004	0.539	0.657
OLS05	TT, TP, S	4	404.881	0.007	-0.016	0.194	0.900
OLS06	TT, P,TS	4	404.286	0.013	-0.009	0.386	0.763
OLS07	T, TP, TS	4	404.082	0.015	-0.007	0.452	0.717
OLS08	TT, TP, TS	4	404.159	0.014	-0.008	0.427	0.734
GWR01	T, P, S	25.140	342.990	0.752	0.669	8.526	<0.001
GWR02	TT, P, S	25.250	343.909	0.751	0.667	8.428	<0.001
GWR03	T, TP, S	24.305	343.937	0.749	0.668	8.785	<0.001
GWR04	T, P, TS	24.729	339.049	0.760	0.681	9.116	<0.001
GWR05	TT, TP, S	24.368	345.604	0.745	0.662	8.560	<0.001
GWR06	TT, P,TS	24.960	341.033	0.758	0.677	8.892	<0.001
GWR07	T, TP, TS	24.017	342.043	0.749	0.670	8.936	<0.001
GWR08	TT, TP, TS	24.056	344.231	0.743	0.663	8.666	<0.001

ENP = effective number of parameters; AIC_c = corrected Akaike information criterion; R^2 = multiple coefficient of determination; adjR^2 = adjusted R^2
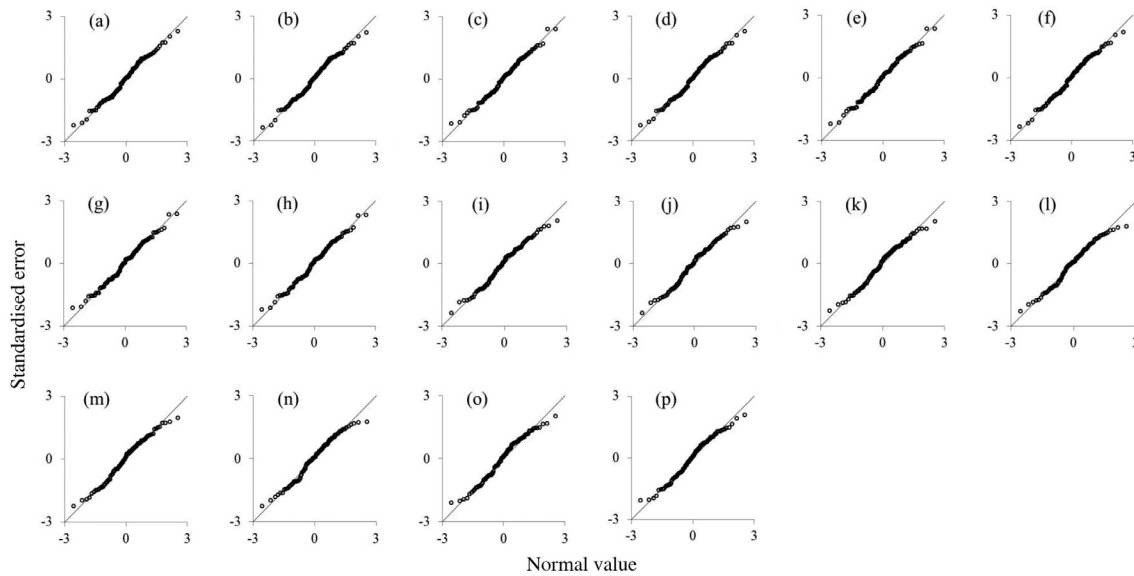
Fig. 11. Cross-validation normal quantile plots for the errors associated with various models developed for the base-changed predicted LTIR09 of CL for the 93 climate factor units of the study area. (a) OLS01; (b) OLS02; (c) OLS03; (d) OLS04; (e) OLS05; (f) OLS06; (g) OLS07; (h) OLS08; (i) GWR01; (j) GWR02: (k) GWR03; (l) GWR04; (m) GWR05; (n) GWR06; (o) GWR07; (p) GWR08.

from 8.428 to 9.116 and P-values <0.001). However, sorting the models in an ascending order according to their cross-validation AIC_c indicated that the model GWR04 was deemed to be the best fit to the data (AIC_c was found to be about 339.049 searching about 12% of the neighbours) and hence was selected as the model for explaining the base-changed predicted LTIR09 for the 93 CFUs. The explanatory variables in this model were T, P and TS. Figs. 13 and 14 show the spatial distribution of the coefficients and standard errors of these explanatory variables (in addition to the intercept), respectively. The developed coefficients had values varied from negative to positive indicating the existence of different relationships between the dependent and independent variables across the study area. Moreover, the standard errors of the variables are generally high in the central and south-eastern parts relative to the peripheries of the study area. The model had errors that varied from -2.3 to 1.8 with mean and SD of about 0.002 and 1.021, respectively.
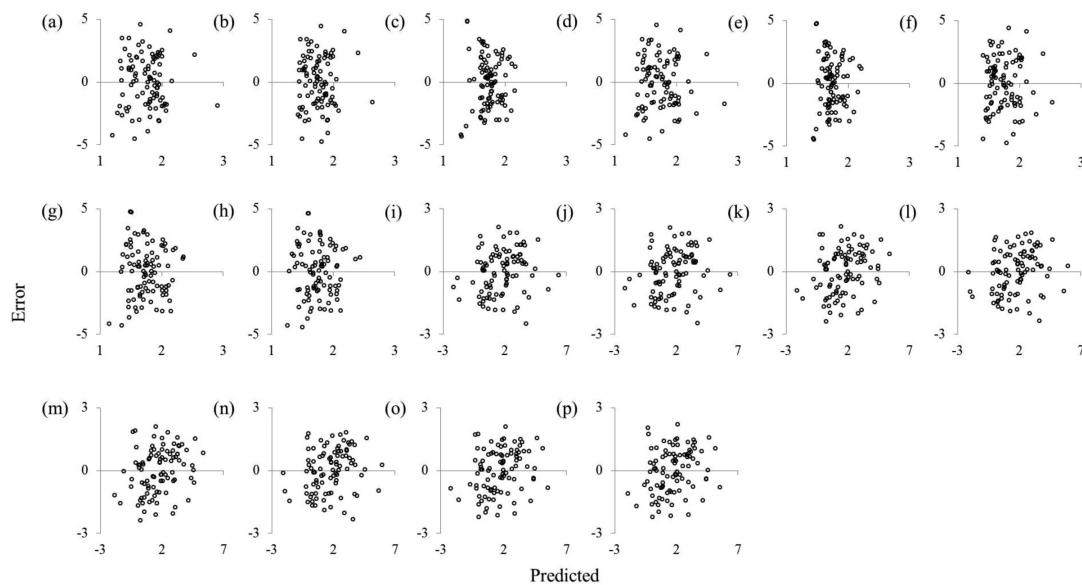

Fig. 12. Cross-validation error plots for the models developed for the base-changed predicted LTIR09 of CL for the 93 climate factor units of the study area. (a) OLS01; (b) OLS02; (c) OLS03; (d) OLS04; (e) OLS05; (f) OLS06; (g) OLS07; (h) OLS08; (i) GWR01; (j) GWR02: (k) GWR03; (l) GWR04; (m) GWR05; (n) GWR06; (o) GWR07; (p) GWR08.
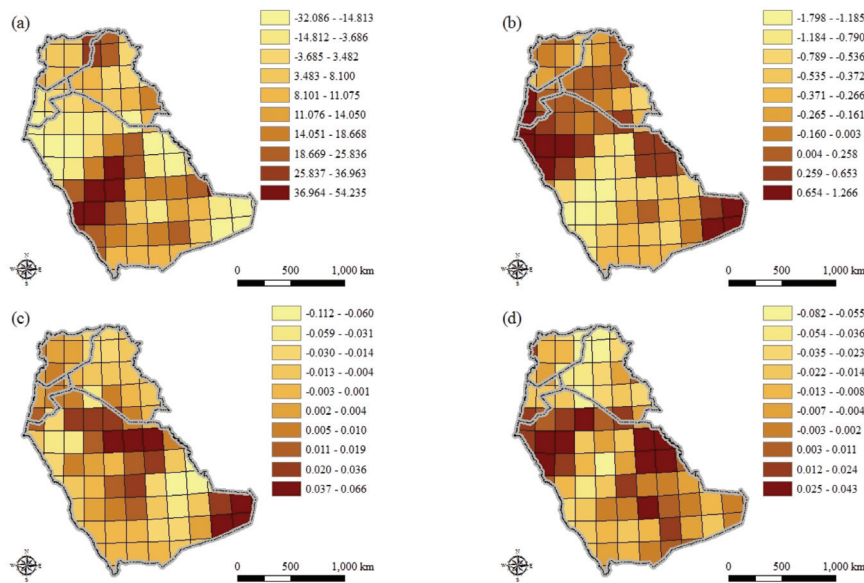
Fig. 13. Choropleth thematic maps showing the spatial distribution of the coefficients developed by applying GWR on the 93 climate factor units of Jordan, Syria, Iraq, and Saudi Arabia in 2009. (a) Intercept; (b) mean daily air temperature (T); (c) total annual precipitation (P); (d) transformed mean daily specific humidity (TS).

The model had the highest explanatory power (R^2 equalled about 0.760 with an adjR^2 of about 0.681 (F Ratio = 9.116 and P-value <0.001)) among all the developed models. However, the local explanation powers using this model in the study area varied from 0.006 to 0.773 (Fig. 15a); also indicating the presence of different relationships between the dependent and independent variables across the study area. The highest explanatory power was found in the north-eastern parts of Iraq, the southern and western parts of Jordan and the western parts of Saudi Arabia along the Red Sea, in addition to the north-eastern parts of Saudi Arabia along the Arabian Gulf. Thus, applying this model revealed almost similar interesting spatial patterns of TIR09 as those that were uncovered by applying areal interpolation (Fig. 15b). The GWR predicted TIR09 of CL for the study area in 2009 varied from 0.1 to 388.0 with mean and SD of about 19.7 and 51.1, respectively. Four major hotspot areas could be identified that covered most parts of Syria and extend-
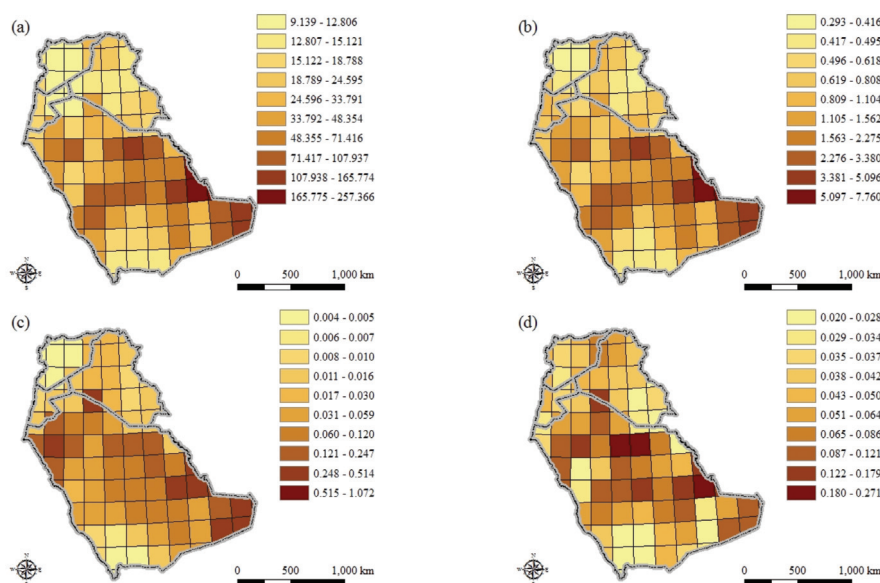

Fig. 14. Choropleth thematic maps showing the spatial distribution of the standard errors of the coefficients developed by applying GWR on the 93 climate factor units of Jordan, Syria, Iraq and Saudi Arabia in 2009. (a) Intercept; (b) mean daily air temperature (T); (c) total annual precipitation (P); (d) transformed mean daily specific humidity (TS).
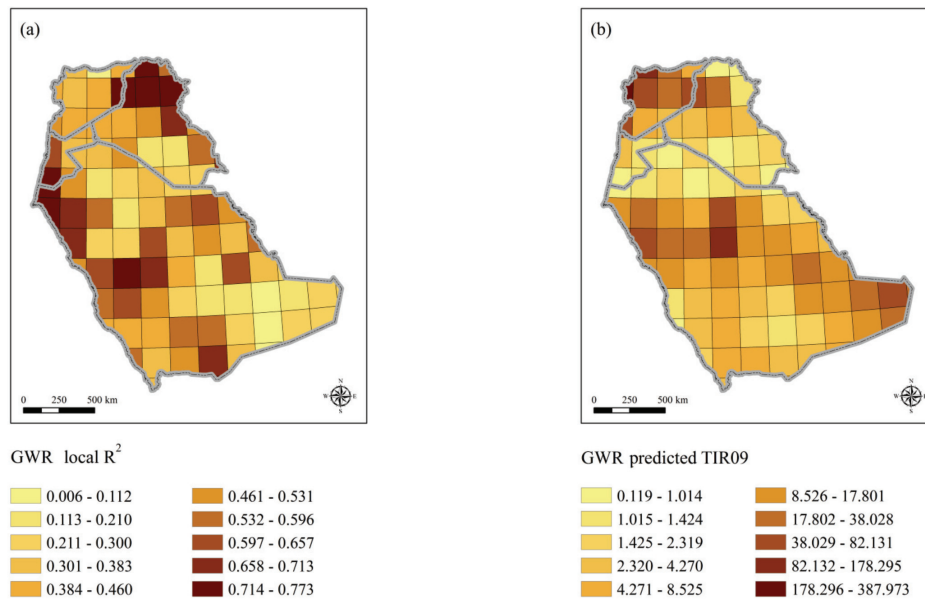
Fig. 15. Choropleth thematic maps showing the spatial distribution of (a) local R^2 and (b) predicted TIR09 of CL for the 93 climate factor units of the study area in 2009 using GWR.

ed (with relatively lower severity) through the central parts of Iraq. They also covered the north-western, central and south-eastern parts of Saudi Arabia. In addition, they covered (with relatively lower severity) the south-western parts of Saudi Arabia.

Discussion

The present study revealed quantitatively some characteristics of the spatial patterns of CL in the Middle East shedding light on the relationships between the spatial patterns of CL incidence rates and precipitation, temperature and specific humidity. However, a number of issues might have affected the results. First, CL data used in this study were obtained from the published reports of the MoH of the four countries. Of course, these are not the only sources of information about CL; however they are the most reliable and readily available ones that can be used for conducting scientific research about the disease in the study area. One consequence of this is the potential under-estimation of the disease, which could be attributed to several factors, including self-healing. Other factors include the absence of active case-detection, which means that the number of reported cases mainly reflects patients seeking medical care in the MoH public clinics. The probable occurrence of CL cases in remote and rural areas with limited resources for treatment and inadequate diagnosis adds to the problem and may have contributed to the possibly false impression that the disease burden in some AUs is low and not as high as in the neighbouring ones.

Second, the level of the spatial aggregation of the reporting systems by the four countries was based on AUs having different areas and different population sizes and densities. Although, this was accounted for during the development of CL incidence rates risk surface by implementing areal interpolation, the uncertainty of the surface varied with the density of the AUs forming the study area, i.e. the smaller the areas of the AUs and the larger the number of the AUs, the lower the uncertainty of the developed surface. Furthermore, the developed surface under-estimated the actual variability of CL incidence rates in the study area. Hence, there is a need to unify and enhance the national reporting systems of Jordan, Syria, Iraq and Saudi Arabia through the incorporation of data that are geocoded based on their addresses, adopting smaller spatial units and using GIS-based multiuser geodatabases that are connected to each other and to other countries in the world classified as areas of high-risk of CL.

Third, as a demonstration of the possibility of using climate factors for developing models for explaining incidence rates of CL in the study area, the readily available climatic factors at the Climate Research Unit at the University of East Anglia, United Kingdom, were used. Although these variables were obtained from a reputable international public source and the reliability of these data were tested and confirmed by the producing agency, these data work better for global-scale studies. Thus, there is a need to conduct more detailed and elaborated regional and local studies (based on unified fine spatial areal arrangements) in

order to produce data about the climatic factors that might affect the spread of the disease in the region. Furthermore, during applying areal interpolation to produce a smooth prediction surface for the CL incidence rates from the 60 AUs, which was aggregated back to the 93 CFUs, the best model chosen (i.e. the exponential function) using 'leave-one-out' validation method produced a correlation of about 0.690 between the measured and predicted data. Though this model has the strongest prediction ability among the 11 mathematical functions that were tested, it still leaves about 0.310 of the data not predicted. Therefore, issues of spatial uncertainty in predicting outcome CL to the CFU level might have arisen. Therefore, there is a need to enhance the national reporting systems of Jordan, Syria, Iraq and Saudi Arabia by adopting smaller spatial units. This should increase the sample size and thus, in addition to testing other mathematical functions, might enhance the predictability of the developed models.

Fourth, though applying Person product-moment correlation coefficient revealed the presence of significant positive correlation between CL incidence rates and precipitation spatial patterns and significant negative correlation between CL incidence rates and temperature spatial patterns but failed to find a relationship between CL incidence rates and specific humidity spatial patterns in the study area, GWR showed that the relationships between CL incidence rates and these three climatic factors are more complex and non-stationary and vary spatially across the study area. It is well known that CL does not respect the political borders of the countries as different environmental, social and economic factors affect its presence and spread. In addition, movement of non-immune populations to areas of high CL endemicity, possibly exacerbated by political instability and influx of refugees between neighbouring countries, might affect the spatial patterns of the disease. It is well-known that human populations inhabiting *Leishmania*-endemic areas exhibit attenuated *Leishmania* infections and are less likely to contract disease as compared to newcomers. Historically, this phenomenon has been attributed to a gradual onset of immunity against *Leishmania* parasites in endemic areas. This could be in part related to the exposure to immunogenic sand fly salivary proteins that induce an overall protective immunity against *Leishmania* infection (Valenzuela et al., 2001). Although, this study included four neighbouring countries of the Middle East and concentrated on exploring the types of the relationships

between the spread of the disease and three climatic variables, there is a need to expand the study to include other countries in the MENA region and include other environmental, social and economic factors that might affect the spread of the disease. This could be further enhanced by adjusting incidence rates for age and other socio-demographic factors and including information about CL vectors and reservoir hosts ecological niche and specific habitat requirements, which are (at the present time) not readily available.

Conclusions

Support has been found for the CL incidence rates in Jordan, Syria, Iraq and Saudi Arabia forming four major hot spot areas covering most parts of Syria, central parts of Iraq and north-western, central, south-eastern and south-western parts of Saudi Arabia. Moreover, the severity of these hot spot areas decreases from high to low latitudes. The spatial patterns of the incidence rates seem to have a significant, moderate, direct relationship with precipitation and significant, moderate but inverse relationship with temperature. Though, no significant relationship was inferred between disease incidence rates and specific humidity spatial patterns, the three climatic factors together can be used as explanatory variables for explaining part of the spatial variations of CL incidence rates in the four countries.

References

Adler S, Theodor O, 1929. The distribution of sandflies and leishmaniasis in Palestine, Syria and Mesopotamia. Ann Trop Med Parasit 23, 269-306.

Al-Tawfiq JA, AbuKhamsin A, 2004. Cutaneous leishmaniasis: a 46-year study of the epidemiology and clinical features in Saudi Arabia (1956-2002). Int J Infect Dis 8, 244-250.

AlSamarai AM, AlObaidi HS, 2009. Cutaneous leishmaniasis in Iraq. J Infect Dev Ctries 3, 123-129.

CBSSY, 2011. Estimated population of Syria at end-year 2009. Central Bureau of Statistics of Syria. Available at: http://www.cbssyr.org/index-EN.htm (accessed on December 2011).

CDC, 2011. Parasites, Leishmaniasis. Centers for Disease Control and Prevention. Available at: http://www.cdc.gov/parasites/leishmaniasis/gen_info/faqs.html (accessed on December 2011).

CDOSISA, 2011. Estimated population of Saudi Arabia at end-year 2009. Central Department of Statistics and Information of Saudi Arabia. Available at: http://www.cdsi.gov.sa/english/

index.php (ccessed on December 2011).

Chatterjee S, Hadi AS, 2012. Regression analysis by example. Hoboken: John Wiley and Sons.

COSIR, 2011. Estimated population of Iraq at end-year 2009. Central Organization for Statistics of Iraq. Available at: http://cosit.gov.iq/english/index.php (accessed on December 2011).

CRU, 2013. NCEP/NCAR Reanalysis 1 Project, Climate Research Unit, University of East Anglia, United Kingdom. Available at: http://www.cru.uea.ac.uk/cru/data/ncep/ (accessed on January 2013).

Davis BM, 1987. Uses and abuses of cross-validation in geostatistics. Math Geol 19, 241-248.

DOSJO, 2011. Estimated population of Jordan at end-year 2009. Department of Statistics of Jordan. Available at: http://www.dos.gov.jo/dos_home_e/main/index.htm (accessed on December 2011).

Douba M, Mowakeh A, Wali A, 1997. Current status of cutaneous leishmaniasis in Aleppo, Syrian Arab Republic. Bull World Health Organ 75, 253-259.

ESRI, 2012a. ArcGIS software version 10.1. Redlands: ESRI Inc.

ESRI, 2012b. ArcGIS software version 10.1 Help. What Is Areal Interpolation? Redlands: ESRI Inc.

Fotheringham AS, Brunsdon C, Charlton M, 2002. Geographically weighted regression: the analysis of spatially varying relationships. Chichester: John Wiley and Sons Ltd.

Goodchild MF, Anselin L, Deichmann U, 1993. A framework for the areal interpolation of socioeconomic data. Environ Plann 25, 383-397.

Guan Q, Kyriakidis P, Goodchild M, 2011. A parallel computing approach to fast geostatistical areal interpolation. Int J Geogr Inf Sci 25, 1241-1267.

Hotez PJ, Savioli L, Fenwick A, 2012. Neglected tropical diseases of the Middle East and North Africa: review of their prevalence, distribution and opportunities for control. PLoS Negl Trop Dis 6, e1475.

Isaaks EH, Srivastava RM, 1989. An introduction to applied geostatistics. New York: Oxford University Press.

Janini R, Saliba E, Khoury S, Oumeish O, Adwan S, Kamhawi S, 1995. Incrimination of *Phlebotomus papatasi* as vector of *Leishmania major* in the southern Jordan Valley. Med Vet Entomol 9, 420-422.

JMP, 2008a. JMP software version 8.0. Cary: SAS Institute Inc.

JMP, 2008b. JMP software version 8.0 help, statistics and graphics guide. Cary: SAS Institute Inc.

MOHIR, 2011. Ministry of Health Annual Report 2009. Ministry of Health of Iraq. Iraq. Available at:

http://www.moh.gov.iq/ (accessed on December 2011).

MOHJO, 2011. Ministry of Health annual statistical book 2009. Ministry of Health of Jordan. Jordan. Available at: http://www.moh.gov.jo/MOH/En/publications.php (accessed on December 2011).

MOHSA, 2011. Statistical book for the year 2009. Ministry of Health of Saudi Arabia. Saudi Arabia. Available at: http://www.moh.gov.sa/en/Ministry/Statistics/book/Pages/default.aspx (accessed on December 2011).

MOHSY, 2011. Syrian epidemiological bulletin for the year 2009. Ministry of Health of Syria. Syria. Available at: http://www.moh.gov.sy/en/Home/tabid/38/Default.aspx (accessed on December 2011).

Mosleh IM, Gieth E, Schonian G, Kanani KA, 2009. Two recent but temporally distinct outbreaks of cutaneous leishmaniasis among foreign workers in the Dead Sea area of Jordan. Ann Trop Med Parasit 103, 393-400.

Postigo JAR, 2010. Leishmaniasis in the World Health Organization Eastern Mediterranean Region. Int J Antimicrob Agent 36, S62-S65.

Rangel TF, Alexandre J, Diniz-Filho F, Bini LM, 2010. SAM: a comprehensive application for spatial analysis in macroecology. Ecography 33, 46-50.

Saliba EK, Disi AM, Ayed RE, Saleh N, Oumeish O, Al-Ouran R, 1994. Rodents as reservoir hosts of cutaneous leishmaniasis in Jordan. Ann Trop Med Parasit 88, 617-622.

Valenzuela JG, Belkaid Y, Garfield M, Mendez S, Kamhawi S, Rowton ED, Sacks DL, Ribeiro JMC, 2001. Toward a defined anti-*Leishmania* vaccine targeting vector antigens: characterization of a protective salivary protein. J Exp Med 194, 331-342.

Waller LA, Gotway CA, 2004. Applied spatial statistics for public health data. Hoboken: John Wiley and Sons, Inc.

WHO, 2007. Cutaneous leishmaniasis, why are you neglecting me? A WHO initiative to control cutaneous leishmaniasis in selected old world areas. Geneva: World Health Organization.

WHO, 2008. Report of the consultative meeting on cutaneous leishmaniasis. Geneva: World Health Organization.

WHO, 2011a. Cutaneous leishmaniasis. World Health Organization. Available at: http://www.who.int/leishmaniasis/cutaneous_leishmaniasis/en/index.html (accessed on December 2011).

WHO, 2011b. Report on the program managers' review meeting on cutaneous leishmaniasis control in the Eastern Mediterranean Region. Marrakech, Morocco, 18-21 September 2010, Document WHO-EM/CTD/062/E/01.11/91. Available at: http://applications.emro.who.int/docs/intercountry_meeting_ctd_062_14278.pdf (accessed on December 2011).