Regression Analysis on Scorpion Envenomation and Climate Variables in M'Sila Province, Algeria from 2001 to 2010

Schehrazad Selmane [#1], Mohamed L'hadj [*2]

[#]Faculty of Mathematics, University of Science and Technology Houari Boumediene, Algeria

[*]Ministère de la sante et de la reforme hospitalière. Algeria

***Abstract*— Scorpionism represents a serious public health problem in Algeria. More than 68% of the national population is at risk of scorpion stings. M'Sila ranks among the endemic provinces of the country and records every year a high incidence of scorpion stings. A survey on epidemiological characteristics of scorpion stings was established. Using the monthly recorded scorpion sting data for the period 2001-2010 for M'Sila province, the linkage between scorpion stings and weather conditions was demonstrated through time series analysis and regression analysis considering the number of scorpion stings as dependent variable and climatic conditions as independent variables. The temperature, precipitation and wind are the retained climate factors, and the temperature has the higher effect. The model predicted the number of scorpion stings in 2011 with a good accuracy. The model could be used by public health makers of the province to anticipate the demand for antivenoms and symptomatic drugs so that they can be distributed in advance. This raises optimism for forecasting scorpion stings provided the availability of appropriate climate information.**

***Keywords*— Climate, Forecasting, Regression Analysis, Scorpion, Temperature.**

I. INTRODUCTION

Scorpion stings represent a serious public health threat in many tropical, subtropical and desert habitats. Seven areas are identified as at risk: north-Saharan Africa, Sahelian Africa, South Africa, Near and Middle East, South India, Mexico and South Latin America, East of the Andes. Although geographically limited (both north and south no species of scorpions exceeds the 50^{o} latitude); scorpionism concerns an at risk population of almost 2 billion and half people. Ecologically, scorpions are important components of arid, semi-arid, and saharian ecosystems, but they are not limited to these areas. Scorpion species distribution depends on a range of climatic and environmental variables such as temperature, humidity, rainfall, elevation, soil properties, vegetation type and land cover [1]. Studies devoted to the scorpionism are insufficient or even rare. The epidemiology of scorpion envenomation is especially determined by three main factors: the man, the environment and scorpion. The man who increases the possibilities of that accident by its reconciliation more and more to scorpions' shelters, particularly because of the expansion of towns and cities without land reclamation, thus creating ideal habitats near houses. The risk exists in city down-towns and in the suburbs, even if it is definitely higher in rural areas [4]. Epidemiologic data are scarce in many areas affected by scorpionism, due to under-declaration. The incidence is underestimated while the mortality is better known; the number of yearly deaths exceed 3250 [1].

The geographic location, climate, and diverse ecosystems, make Algeria a conducive environment for scorpions. The country houses a diverse scorpion fauna. More than 28 species, belonging to 13 genera and three families (*Buthidae, Euscorpiidae and Scorpionidae*) are described for the country. The dangerous scorpion species to humans are *Androctonus australis* (south highlands, Saharan Atlas), *Buthus occitanus tunetanus* (Septentrional edge of the Sahara), *Androctonus Aeneas* (highlands, Saharan Atlas), and *Androctonus Crassicauda* (Tindouf) [13]. However, huge gaps exist in the knowledge of this fauna in the north of the country. The epidemiological situation of the scorpion envenomation revealed that 77.1% of provinces are affected by the scorpion envenomation accidents and hence 68.91% of the national population is at risk of scorpion stings. The incidence varies between less than 7 scorpion stings per 100,000 inhabitants in the Northern provinces and more than 1000 scorpion stings per 100,000 inhabitants in those of the South. A total of 903,461 scorpion sting cases and 1996 deaths were recorded by health services between 1991 and 2012. Fourteen provinces of Highlands and Sahara together account alone for almost 90% of patients stung and the entire deaths [11].

Ranked among the poor provinces in Algeria, M'Sila is faced each year with serious health problems as a result of the degradation of the environment and public hygiene; the region records every year a high incidence of scorpion envenomation and a very high incidence of cutaneous leishmaniasis; both related to climate. The climatic conditions and environmental deterioration make of M'Sila a conducive environment to scorpions. Furthermore, the liability of human in the increase of possibilities scorpion sting accidents through its negligence and/or its ignorance and through its reconciliation more and more to scorpions' shelters make things worse. Thus, the public health makers of the province are faced to scorpionism, and consequently, they are required to establish prevention and control strategies. An early warning system remains an essential tool for preparedness and effectiveness of scorpion stings control; in this way it could help determine the appropriate number of antivenom vials necessary in health facilities and to anticipate the demand for antivenoms and symptomatic drugs so that they can be distributed in advance in this endemic province; this is the aim this work.

The first mathematical modeling approach on scorpions was presented by Chowell and al; they analyzed the significance of climatological variables to predict the incidence of scorpion stings in humans in the state of Colima (Mexico) using multiple linear regression [2]. In this paper, in addition to an epidemiological survey on scorpion stings data of the province of M'Sila and analysis of time series of variables, we performed a regression analysis to estimate the relationship between scorpion sting cases (the dependent variable) and climate conditions (the independent variables).

The paper is organized as follows: in section 2, we present materials and methods used. The epidemiological panorama of scorpion stings, the analysis of the built predictive model, and discussion are presented in section 3.

II. Materials and Methods

A. The Region of study

M'Sila is located in the Highlands of east-central Algeria and is situated at $35^{\circ}42^{'}7^{''}N$ and $4^{\circ}32^{'}49^{''}E$ of the equator. The province is made up of 47 municipalities distributed between 15 districts over a land size of 18,175 square kilometers. As 2013, the province accommodates a population of 1,107,821.

Its climate is mostly arid, and partly semi-arid and Saharian. The province experiences high temperature between June and September, and rainfalls between September and December. The average temperature is $34^{\circ}C$ in summer and $10^{\circ}C$ in winter. The precipitations are low, irregular and not exceeding 180 *mm* per year. The province is dry during the whole year and August is the driest month.

B. Data

Two different monthly data sets were used for the study period: epidemiological and meteorological data. These data comprise 120 months from January 2001 to December 2010.

1) Epidemiological Data

The monthly recorded scorpion sting cases from 2001 to 2011 as well as the total scorpion sting cases by gender, by age groups, by anatomical sites, and by place for the year 2010, were provided by the Department of Public Health of M'Sila.

2) Meteorological Data

The monthly mean temperature (T) in $^{\circ}C$, the monthly accumulated precipitation amount (P) in *mm*, the monthly mean relative humidity (RH) in %, the average monthly evaporation (E) in *mm*, and the monthly mean wind speed (W) in *m/s* for the study period were provided by the national office of meteorology [8].

C. Methods

1) Epidemiological survey of scorpion envenomations

In the aim to trace the epidemiological profile and to specify some epidemiological characters of scorpion stings, the distribution of the scorpion sting cases by gender, by anatomical sites, by age groups, by location, and by daily hour of recorded cases are established for the year 2010 from the monthly synthesis sheet supplied by the Department of Public Health of M'Sila.

2) Descriptive statistics and time series analysis of variables

Descriptive statistics is used to quantitatively describe the main features of the data. In order to find any significantly relationship between the scorpion sting variable and the climate variables we use scatter plots as well as Pearson product-moment correlation coefficient. Time series analysis of data is also performed in order to extract meaningful statistics and other characteristics of the data.

3) Regression model

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. A regression model relates the dependent variable, Y to a specified function of independent variables, X, and unknown parameters, β,

$$Y \approx f(X, \beta)$$

The approximation is usually formalized as $E(Y|X) = f(X,\beta)$, where $E(Y|X)$ is the average value of the dependent variable when the independent variables are fixed. One method of estimation is ordinary least squares. This method obtains parameter estimates that minimize the sum of squared residuals [3,6].

The features of a best regression model are

- Value of R-square should be more than 60 percent. Higher the R-square value, better the data fitted.
- Most of the independent variables should be individually significant to influence the dependent variable (this matter can be checked using t-test).
- The independent variables should be jointly significant to influence or explain dependent variable (This can be checked using F-test).
- No serial correlation in the residual (can be tested using Bruesch-Godfrey serial correlation LM test).
- No heteroscesticty in the residual (can be tested using Bruesch-Pegan-Godfrey Test).
- Residuals should be normally distributed (can be tested using Jarque Bera statistics).

When all these features are fulfilled then we can use the model for forecasting [3,6].

All performed computations and generated figures were carried out with Eviews 7 software.

III. RESULTS

A total of 46,589 scorpion stings and 78 deaths were recorded by the Department of Public Health of the province of M'Sila between 2001 and 2010; which is equivalent to a mean annual incidence of 494 per 100,000 inhabitants. The recorded scorpion sting cases and the lethality are plotted from 2001 to 2010 (see Fig. 1), where the lethality is defined as the ratio of the number of deaths due to scorpion envenomation over the number of cases stung during the same period and expressed in percentage. The highest (resp. lowest) total yearly scorpion sting cases occurred in the year 2007 with 5,532 (resp. 2005 with 3,766) cases and the highest number of deaths were notified in 2008 with 15 deaths and in 2007 with 11 deaths.

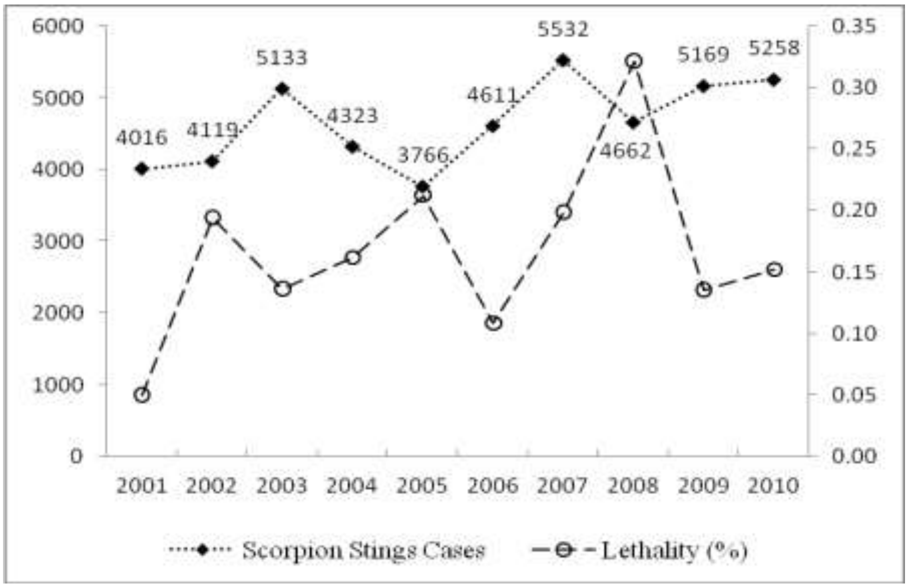

Fig.1 Evolution of annual recorded scorpion sting cases and lethality in M'Sila province from 2001 to 2010.

A. Epidemiological panorama of scorpion stings in 2010

The scorpions sting round the clock with peak between 6 *pm* and 12 *am*, all year long with peak in summer, all ages with no sex predominance, with a predominance within domiciliary. Envenomations are graded as mild, moderate, or severe. Among the 5,258 recorded scorpion sting cases by the Department of Public Health of M'Sila in 2010 approximately 95% of cases were mild, 2.5% of cases were moderate, and 2.5% of cases were severe. The total cost corresponding to the 5,258 antivenom vials used is 9,398,675DA (≈ 118,700 US dollars) which represents a heavy burden for a poor province as M'Sila.

a) Scorpion stings by anatomical sites

The more prone human body areas to scorpion stings are the upper limbs with 51.92% of cases followed by the lower limbs with 39.78% of cases. The fact that the upper and lower limbs are the most affected parts of the body (91.7% of reported victims) suggests that it is the human who has a great deal for liability in these accidents through its negligence and/or its ignorance. Thus, the development of an educational program of the population associated with the improvement of environmental hygiene could significantly reduce the incidence.

b) Distribution of envenomed cases by age groups

The most frequently affected age group ranges from 15 to 49 year, with 62.12% of stung cases. Children less than 14 years old represent 23.51% of cases, and the envenomations are more severe and mortality is dramatically higher for this category than in adults. As a whole there is no sex predominance.

c) Scorpion stings by place

Nearly two thirds of scorpion sting cases occurred within dwellings and one third occurred outside dwellings. More precisely, 3,070 cases were recorded inside dwellings and 2,188 cases were recorded outside dwellings.

d) Daily distribution of scorpion sting cases

More than half of scorpion sting cases (53.94%) occurred between 6 *pm* and 6 *am*, with 2,052 (resp. 784) scorpion sting cases between 6 *pm* and 12 *am* (resp. 12 *am* and 6 *am*). For the slot 6 *am* - 12 *pm* (resp. 12 *pm* - 6 *pm*) 1,398 (resp. 1,024) cases were recorded.

B. Statistical analysis

1. Descriptive statistics of the variables

The descriptive statistics of the monthly data used in this study as well as the Pearson product-moment correlation coefficient (r) between scorpion sting cases and climate variables are displayed in Table 1. The climate variables with strong positive correlation coefficient are temperature (r = 0.907), and evaporation (r = 0.705). This confirms the increasing activity of scorpion with increasing the environment temperature. There is also strong negative correlation between scorpion sting cases and relative humidity (r = -0.814). The correlation between accumulated precipitation amount and the scorpion sting cases is weaker (r = -0.236), and almost zero with wind speed (r = -0.041). The coefficient of variation CV (CV = SD/Mean = 1.19 where SD is the standard deviance) is closer to 1, which means the greater the variability of scorpion data.

Table 1: Descriptive statistics of the variables

Variables	*Minimum*	*Maximum*	*Mean*	*SD*	*r*	*P value*
S	*0*	*1563*	*388.24*	*462.58*	*1*	
T	*6.2*	*33.9*	*19.58*	*8.49*	*0.907*	*0.000*
RH	*30*	*88*	*58.07*	*15.58*	*-0.814*	*0.000*
E	*22.3*	*404.4*	*153.05*	*92.97*	*0.705*	*0.000*
P	*0*	*83*	*15.25*	*16.58*	*-0.236*	*0.009*
W	*2.1*	*6.2*	*4.19*	*0.83*	*-0.041*	*0.657*

2. Time series of the variables

Scorpion stings are recorded throughout the year, even during the winter months. The monthly mean distribution of cases of envenomation for the study period 2001-2010 is plotted in Fig. 2 (a). The monthly peaks are observed July (26.8% of cases) and August (26.8% of cases), accounting alone for more than half of cases. The maximum recorded scorpion sting cases during the study period occurred in July 2009 (1,563 cases) and August 2008 (1,463 cases); in these dates highest temperature recorded was 33.8^{o}C in July 2009 and 31.8 ^{o}C in August 2008. Most of the cases (70.63%) were notified during the summer period followed by autumn period (17.06%), then spring period (12.02%) (Fig. 2 (b)). Mainly stings are distributed between the months of March and October corresponding to dry period of the region and therefore confirm the conclusions on studies in other world's region; that is, the incidence of scorpion stings is strongly related to dry and hot climate.

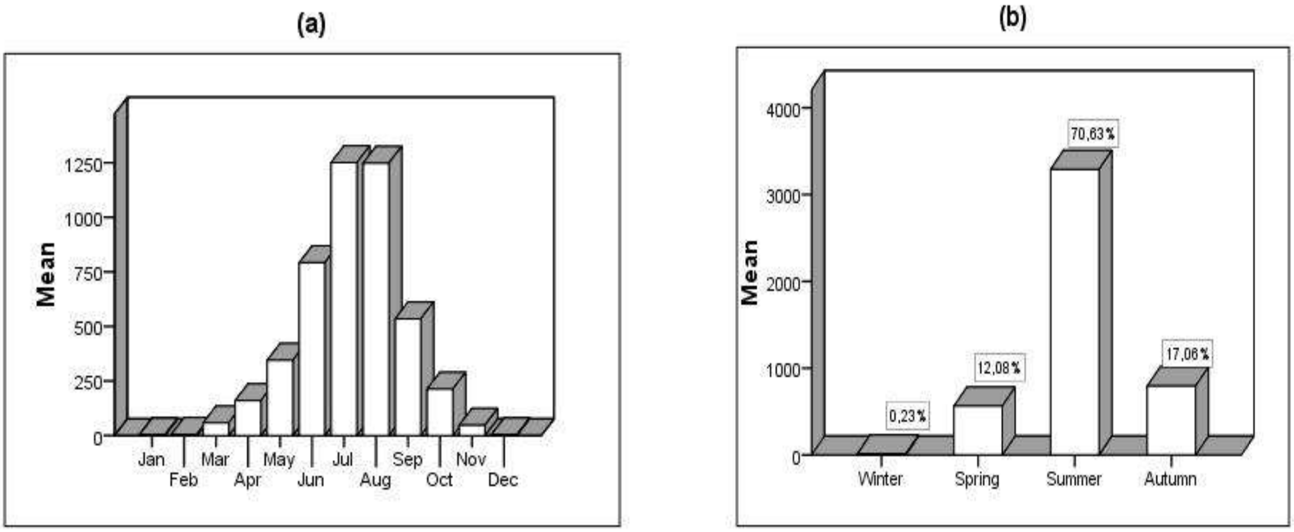

Fig 2 (a) Monthly average distributions of recorded scorpion sting cases. (b) Seasonal average distribution of recorded scorpion sting cases in the province of M'Sila for the period 2001-2010.

Figures 3, 4, 5, and 6 show the time series of the monthly recorded scorpion sting cases with climate variables. Temperature and evaporation follow the same trends, while the relative humidity have opposite trends with scorpion sting cases. The highest accumulated precipitation amounts were recorded in January 2003 with 74.7 *mm* and without recorded scorpion sting cases, in October 2003 with 83 *mm* and with the highest number of recorded scorpion stings (352) for this month (average number of cases being 213 cases), in May 2004 with 74.8 *mm* and with the lowest number of recorded scorpion stings (210) for this month (average number of cases being 345 cases), and in April 2007 with 79.8 *mm* and with 154 recorded scorpion sting cases which is near average number of cases (160) for this month. This is against the stories of local residents who believe that precipitations decrease the activity of scorpions and therefore the risk to be stung.

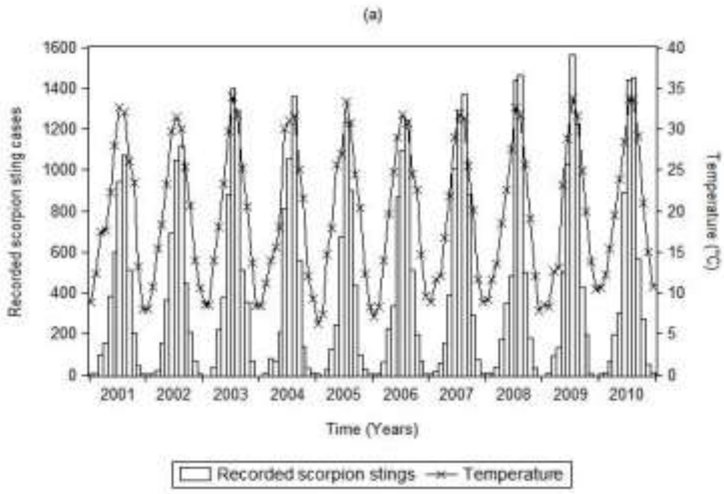

Fig 3 Time series of the monthly recorded scorpion sting cases (bars) with monthly average temperature (solid line) in M'Sila province for the period 2001-2010.

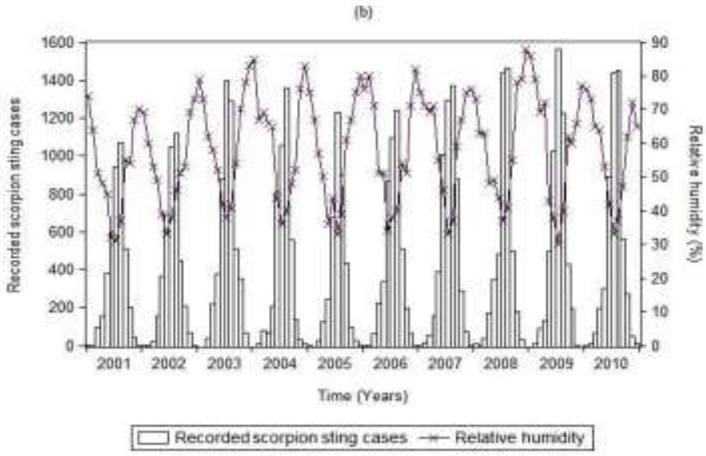

Fig 4 Time series of the monthly recorded scorpion sting cases (bars) with monthly average relative humidity (solid line) in M'Sila province for the period 2001-2010.

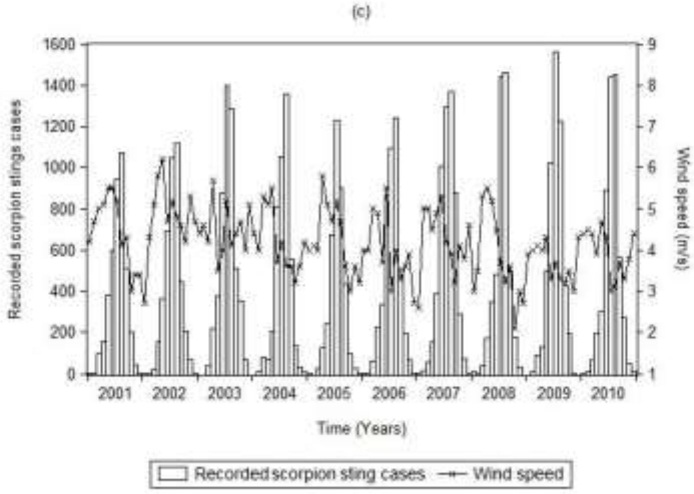

Fig 5 Time series of the monthly recorded scorpion sting cases (bars) with monthly average wind speed mean (solid line) in M'Sila province for the period 2001-2010

3. Regression analysis

The independent variables, temperature and relative humidity and evaporation, are highly correlated, thus they will impart nearly exactly the same information to a regression model. To avoid the multicollinearity and the unreliability of the regression model's regression coefficients related to these highly correlated variables, it is suggested to include only one of these variables. Knowing for a fact that the activity of scorpion increases with the environment temperature increase, we selected then temperature.

To achieve linearity for regression analysis, we used a quadratic model by applying the squared root transformation to the dependent variable S; ($S^{½}$) was best correlated with statistical significance with the T^2 (r = 0.979). We thus performed a regression analysis considering ($S^{½}$) as dependent variable, and T^2 as independent variable. However, even if this model is good, it cannot be used for forecasting; the residuals were serially correlated (the corresponding p-value = 0.0182 is less than 5%).

We then performed a quadratic regression by including all the other climate variables as well as a trend variable (Tr) to account for non-climatic factors such human behavior, degradation of the environment, and other factors that could influence the number of sting cases. A model fulfilling the six features of the best regression model for forecasting corresponds to the model incorporating the constant, the climate variables T, T^2, P, W, and the trend variable Tr, as independent variables, and the squared root of S as dependent variable. The model outcomes are displayed in Table 2.

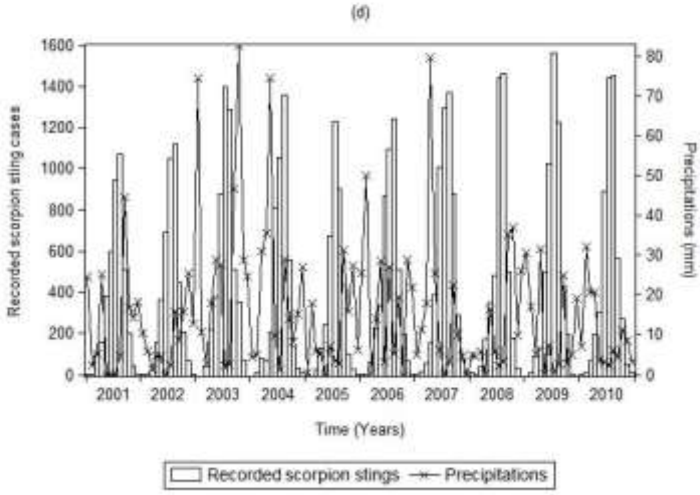

Fig 6 Time series of the monthly recorded scorpion sting cases (bars) with monthly accumulated precipitations (solid line) in M'Sila province for the period 2001-2010

Table 2: Model outcomes

Dependent Variable: SQR(S)
Method: Least Squares
Date: 07/09/14 Time: 12:50
Sample: 2001M01 2010M12
Included observations: 120

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	*-5.994819*	*1.785169*	*-3.358124*	*0.0011*
T	*0.318325*	*0.162133*	*1.963354*	*0.0520*
T^2	*0.027283*	*0.003980*	*6.855115*	*0.0000*
P	*0.038382*	*0.013816*	*2.778065*	*0.0064*
W	*0.290149*	*0.291055*	*0.996887*	*0.3209*
T_r	*0.016392*	*0.006763*	*2.423789*	*0.0169*

R-squared	*0.964518*	*Akaike info criterion*	*4.607757*
Adjusted R-squared	*0.962962*	*Durbin-Watson stat*	*1.579531*
S.E. of regression	*2.364548*	*F-statistic*	*619.7872*
Log likelihood	*-270.4654*	*Prob(F-statistic)*	*0.000000*

The model equation is given by :

$$S^{1/2} = -5.991 + 0.318\,T + 0.027\,T^2 + 0.038\,P + 0.290\,W + 0.016\,Tr \qquad (1)$$

- The value of $R^2 = 0.9645$ indicates that the model is nicely fitted (R^2 is more than 60%, more, 96.45% variance in the dependent variable can be explained jointly by the selected independent variables. The rest 3.55 percent variation in the dependent variable can be explained by residuals or other variables other than the selected independent variables.
- Most of the independent variables are individually significant to influence the dependent variable; the corresponding p-value is less than to 5 percent.
- The independent variables are jointly significant to influence dependent variable; the p-value of F-statistic is less than 5%.
- The residuals are not serially correlated; the Bruesch-Godfrey serial correlation LM test shows that corresponding p-value is 0.0519.

- According to Bruesch-Pegan-Godfrey Test, the residuals are not heteroscedastic; the corresponding p-value is 0.1078; more than 5%.
- The residuals are normally distributed; Jarque-Bera value 1.117 is less than 5.99 and the corresponding p-value is 0.5719; more than 5%.

All features of a best regression model are fulfilled, thus the model can be used for forecasting.
In Fig. 7 we plotted the actual versus the fitted scorpion sting cases for the study period. The simulated number of scorpion stings for the validation period (2001-2010) is closely approximated to the recorded data.

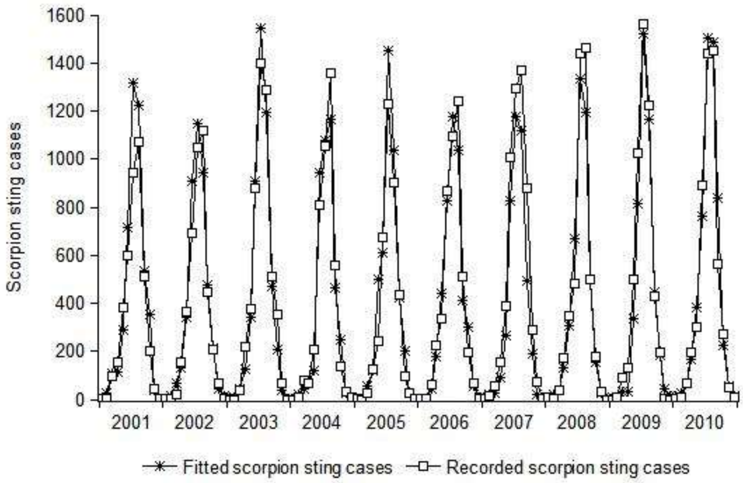

Fig 7 Fitted scorpion sting cases versus recorded scorpion sting cases in 2001-2010.

The number of scorpion stings in 2011 was predicted using temperature, precipitation, and wind conditions in 2011, and compared with recorded scorpion stings for the same year. The result is shown in Fig. 8 (a). The scatter plot (Fig. 8 (b)) of the recorded scorpion sting cases and simulated scorpion sting cases shows that the model is accurate at predicting high and low scorpion sting cases, with certain exceptions which can be considered as outliers in the recorded data.

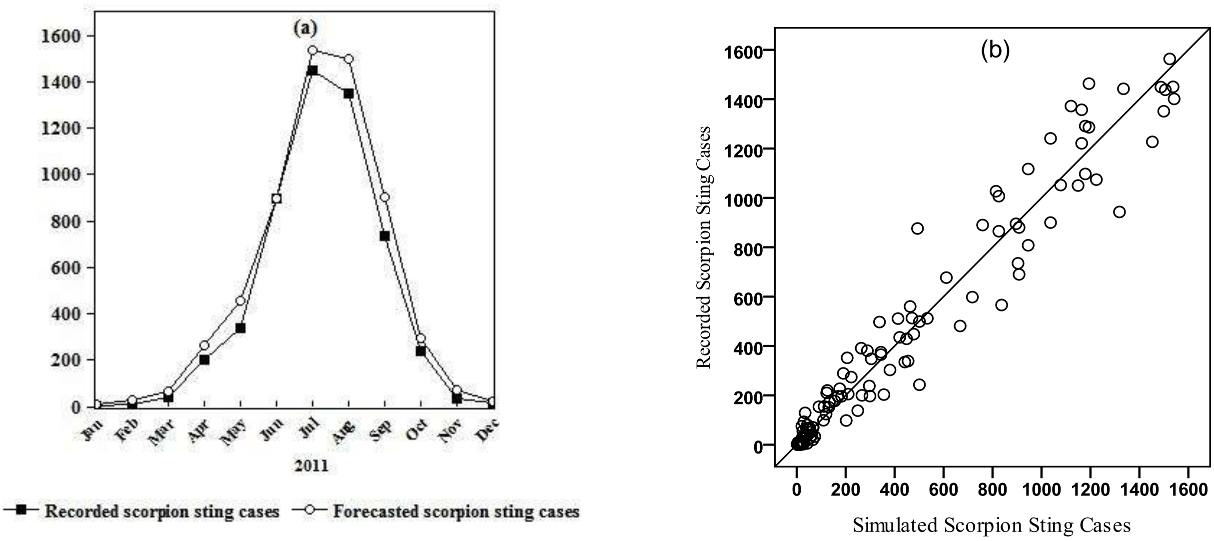

Fig 8 (a) Forecasted versus recorded scorpion sting cases in 2011. (b) The scatter plot of the recorded and simulated scorpion sting cases in M'Sila province in 2001-2011, with correlation 0.976.

C. Discussion

Using the monthly recorded scorpion sting data for the period 2001-2010 for M'Sila province, the linkage between scorpion stings and weather conditions was demonstrated using a regression analysis. The temperature, precipitation and wind are the retained climate factors; moreover, temperature has the higher effect. This raises optimism for forecasting scorpion stings provided that appropriate climate information are at our disposal.

The scorpion stings and their increases can be monitored by the regression model (1). This latest, shows how and when there will be an increase in the number of sting cases provided that appropriate climate information are at our disposal. If we know beforehand the change in a climate variable, we can use the built regression model to estimate how much the change in the value of this variable influences the number of cases of scorpion stings. This could be used to help determine the appropriate number of antivenom vials necessary for the province in advance. Therefore, this work represents an important step to find a way to help in the designing of a control strategy.

In conclusion, our study shows optimism for weather-based forecasting of scorpion stings. It represents an important support for designing of intervention strategies. However, further studies are needed to explore whether other independent variables, such as land cover index, can improve the prediction. As the epidemiology of scorpion envenomation is determined, besides scorpions, by man and environment, the modeling incorporating environmental conditions and human behavior is to be undertaken.

REFERENCES

[1] J. P. Chippaux and M. Goyffon, Epidemiology of scorpionism: A global appraisal, Acta Trop 107(2):71{79, 2008.

[2] G. Chowell, J. M. Hyman JM, P. Diaz-Duenas, and N. W. Hengartne, Predicting scorpion sting incidence in an endemic region using climatologically variables, Int J Environ health Res 15(6):425{435, 2005.

[3] J. Fox, Applied Regression Analysis, Linear Models, and Related Methods, SAGE Publications, Social Science, 1997.

[4] M. Goyffon, Le role de l'homme dans l'expansion territoriale de quelques espèces de scorpions, Bulletin de la Société Zoologique de France Evolution et Zoologie 1171: 15{19, 1992.

[5] M. Goyffon, J. P. Chippaux, Les envenimations scorpioniques en Afrique, Rev. Medicopharmaceutique 53(4):17{21, 2009.

[6] P. McCullagh and J. A. Nelder, Generalized Linear Models, London Chapman and Hall, 1989.

[7] J. Neter, M. Kutner, W. Wasserman, Applied Linear Statistical Models, 1996.

[8] Office national de météorologie. URL : http://www.onm.meteo.dz

[9] Office national des statistiques. URL : http://www.ons.dz

[10] Prise en charge de l'envenimation scorpionique, Comite national de lutte contre l'envenimation scorpionique, 2009.

[11] Relevé épidémiologique mensuel, Institut national de la santé publique. URL: http://www.ands.dz/insp/scorpionisme.html.

[12] S. Selmane, H. ElHadj, and L. Benferhat, "*The Impact of Climate Variables on the Incidence of Scorpion Stings in Humans in M'Sila's Province in Algeria*", IAENG proceedings ISBN 978-988-19252-7-5: 2014.

[13] M. Vachon, *Etude sur les scorpions*, Institut Pasteur d'Algérie. Alger, 1952.