Leishmaniasis is caused by protozoan parasites that belong to the genus *Leishmania*. The disease is transmitted to humans by the bite of infected female sand flies: *Phlebotomus* species in the Old World and *Lutzomyia* species in the New World. The leishmaniases are grouped into two broad categories: anthroponotic leishmaniases in which the reservoir host is human and zoonotic leishmaniases in which the reservoir host is wild or domestic animal. The four main types of the disease are cutaneous leishmaniasis (CL), the most common form of leishmaniasis; diffuse cutaneous leishmaniasis; a chronic form of leishmaniasis that is difficult to treat; mucocutaneous leishmaniasis, the most feared form because it produces destructive and disfiguring lesions of the face; and visceral leishmaniasis, also known as Kala-Azar, which if left untreated can have a fatality rate as high as 100% within two years. Leishmaniasis occurs on four continents and is considered endemic in 98 countries or territories with more than 350 million people at risk. In the Old World, the disease is found mainly in dry, semi-arid areas whereas in the New World, it occurs mainly in tropical forests and savannas. About 95% of CL cases occur in the Americas, the Mediterranean basin, the Middle East and Central Asia. Over two thirds of new CL cases occur in Afghanistan, Algeria, Brazil, Colombia, Iran, and Syria. An estimated 0.7 million to 1.3 million new cases occur worldwide annually.

Cutaneous leishmaniasis, also called the Biskra boil, represents a real public health problem in Algeria, being the second largest focus in the world after Afghanistan. Two pronounced CL outbreaks occurred in 2005 and 2010 with 25,511 and 21,043 cases, respectively. The disease used to be mainly endemic in the sub-Saharan steppe until the last few years; a geographical extension towards the north and west has taken place. Human
infection is caused Phlebotomus papatasi, Phlebotomus perniciosus, Phlebotomus sergenti, and Phlebotomus perfiliewi. The notification of leishmaniasis became mandatory in 1979 and the disease has been under surveillance since 1985. In 2006, a national leishmaniasis control program was created. Care for leishmaniasis is provided for free in high incidence provinces.2,3

Leishmaniasis surveillance in Algeria is based on a passive system. The recording of leishmaniasis cases and the standardization of treatment are the main actions undertaken by health authorities, in addition to sand fly control, which consists of insecticide campaigns. However, the analysis was not undertaken for the purpose of designing intervention strategies. The present study aimed to analyse and interpret the recorded data to develop a forecasting model to perceive changes in incidence early enough to adopt measures to improve preparedness. For the 15 years from 2000 to 2014, the province of Biskra has ranked first in numbers of recorded CL, thereby justifying our choice of this province for this first statistical study.3

The author applied the first mathematical approach to leishmaniasis in Algeria, developing a deterministic model for transmission dynamics that showed that control measures should be directed to reservoir hosts.4 The purpose of designing intervention strategies. The present study aimed to analyse and interpret the recorded data to develop a forecasting model to perceive changes in incidence early enough to adopt measures to improve preparedness. For the 15 years from 2000 to 2014, the province of Biskra has ranked first in numbers of recorded CL, thereby justifying our choice of this province for this first statistical study.3

METHODS
Study area
The province of Biskra is located in central and eastern Algeria at the "gates of the Sahara". The province stretches over 20,986 km² and is made up of 33 municipalities distributed over 12 districts.5 As of 2014, the province accommodated an estimated population of 843,683.5 Its Saharan climate is characterized by weak and irregular precipitation, intense evaporation and wide variations in temperature reaching a monthly average of 34.8°C in July and a monthly average of 11.5°C in January. The dry period in the province lasts almost the entire year, being more pronounced in summer.6

Data
The daily recorded CL cases from 2000 to 2014 were aggregated into months; the series of monthly cases generated 180 data points. Anonymous data were provided by Biskra Department of Public Health. The climate data for the study period involving temperature (T) in °C, relative humidity (RH) in %, evaporation (E) in mm, wind speed (W) in m/s, and precipitation (P) in mm were provided by the National Meteorological Office.6

Modelling method
Time series analysis based on the Box-Jenkins method or an autoregressive moving average (ARMA) models recorded cases over time and allows forecasts to be made of expected numbers of recorded cases.7 Given a stationary time series of data $X_t$ $(t=1,..., n)$ an autoregressive moving average model of orders p and q, ARMA(p,q), is defined by the following difference equation:

$$A(L)X_t = B(L)e_t,$$  \hspace{1cm} (1)

where

$$A(L)=1-\alpha_1L-\alpha_2L^2-...-\alpha_pL^p$$

and

$$B(L)=1+\beta_1L+\beta_2L^2+...+\beta_qL^q$$

$L$ is the backward shift operator defined by $LX_t=X_{t-1}$, $\{e_t\}$ is a white noise process defined as a sequence of uncorrelated zero mean random variables with constant variance, $\alpha=(\alpha_1, \alpha_2, ... , \alpha_p)$ is a vector of autoregressive coefficients and $\beta=(\beta_1, \beta_2, ... , \beta_q)$ is a vector of moving average coefficients so that equation (1) is both stationary and invertible, that is, all roots of the polynomials $A(L)$ and $B(L)$ are outside the unit circle. The orders p and q are lags for cutting off the autocorrelation function (ACF) and partial autocorrelation function (PACF), respectively. Once orders are determined, the parameters may be estimated by a nonlinear optimization technique or the least squares procedure. Eviews software (http://www.eviews.com/), which is based on the least squares approach involving non-linear iterative techniques, is used for fitting of the model.8 In a test of goodness-of-fit, the residuals should be uncorrelated with a zero mean and follow a Gaussian distribution. The autocorrelations of the residuals should not be significantly different from zero.9

RESULTS
Between 2000 and 2014, the Department of Public Health of Biskra province recorded 42,017 CL cases. The highest yearly recorded cases occurred in 2005 and 2010 with 8,594 and 6,163 cases, respectively (Figure 1). The highest monthly CL cases were also observed...
INCIDENCE OF CUTANEOUS LEISHMANIASIS

Table 1. Descriptive statistics.

| Variables | Mean | Median | Maximum | Minimum | Std. Dev. | CV |
|-----------|------|--------|---------|---------|-----------|----|
| CL        | 233.4 | 124    | 1807    | 2       | 302.1     | 1.3|
| T         | 22.8  | 22.5   | 36.5    | 9.9     | 8.1       | 0.4|
| RH        | 42.2  | 42.8   | 71      | 23      | 11.9      | 0.3|
| E         | 205.4 | 180.2  | 425.3   | 54.7    | 100.2     | 0.5|
| P         | 10.9  | 3.5    | 91.1    | 0       | 16.5      | 1.5|
| W         | 4.4   | 4.2    | 8.0     | 2.1     | 1.1       | 0.3|

CL: cutaneous leishmaniasis, T: temperature, RH: relative humidity, E: evaporation, P: precipitation, W: wind speed.

during these two years as indicated by an asterisk in Figure 2. Table 1 shows the descriptive statistics of the monthly data over the study period 2000-2014.

In the model, the lag structure was determined by varying from 0 to 12 months between the dependent variable (CL) and the independent variables (T, RH, E, P, and W). The monthly CL cases were significantly associated with monthly mean relative humidity at a lag of 0 month ($r=0.560$, Pearson product-moment correlation coefficient) and at lag of 12 months ($r=0.582$). The monthly cases were associated with a mean monthly temperature at a lag of 5 months ($r=0.576$), and with a monthly accumulated evaporation lag of 5 months ($r=0.513$). This 5-month lag corresponded to the average incubation period for CL. The cross correlogram between CL and the other climate variables found no significant relationship.

Figure 3 shows the time series of the monthly CL cases with lagged climate variables that are significantly associated with CL data. Temperature at a lag of 5 months and relative humidity follow the same trends as the CL cases.

In the statistical analysis based on the Box-Jenkins modelling we checked that the data was stationary using the augmented Dickey-Fuller unit root test implemented in Eviews. We then identified plausible models from the ACF and PACF. The ACF showed spikes at the first three lags, suggesting a moving average component of order three. The PACF also showed spikes at the first three lags suggesting an AR component of order three. We tried several possibilities while incorporating the climate variables. We based the choice of the best fitted model on criteria such as the smallest Akaike information criterion, the smallest standard error regression, the highest adjusted $R^2$, and the invertibility condition and significance of AR and MA roots, and finally the white noise condition for residuals. In several trials, the ARMA(3,3) incorporating RH and T(-5) model was selected as appropriate for the monthly time series data of CL from 2000 to 2009. The output estimations of the model are displayed in Table 2. All the coefficients were significantly different from zero, each being larger than twice its
Figure 3. Time series of CL, relative humidity, and lagged temperature.

Table 2. Model outcomes.

| Variable | Coefficient | Std. Error | t-statistic | Prob. |
|----------|-------------|------------|-------------|-------|
| C        | -313.38     | 121.09     | -2.59       | .0110 |
| T(-5)    | 15.00       | 3.34       | 4.49        | .0001 |
| RH       | 5.59        | 2.00       | 2.79        | .0063 |
| AR(1)    | 2.48        | 0.07       | 35.43       | .0001 |
| AR(2)    | -2.37       | 0.11       | -21.54      | .0001 |
| AR(3)    | 0.84        | 0.06       | 12.98       | .0001 |
| MA(1)    | -1.41       | 0.11       | -12.82      | .0001 |
| MA(2)    | 0.55        | 0.19       | 2.89        | .0045 |
| MA(3)    | 0.24        | 0.11       | 2.18        | .0325 |
| R-squared| 0.8353      |            |             |       |
| Adjusted R-squared | 0.8226 |            |             |       |
| S.E. of regression | 135.5681 |            |             |       |
| Sum squared residuals | 1893007 |            |             |       |
| Log likelihood | -704.0911 |            |             |       |
| F-statistic | 65.3383 |            |             |       |
| ProblF-statistic | 0.0000 |            |             |       |
| Inverted AR Roots | .87 | .81-.56i | .81-.56i | .25 |
| Inverted MA Roots | .83+.52i | .83-.52i | -.25 | |

Dependent Variable: LC; Method: Least Squares; Sample (adjusted): 2000M09 2009M12; Included observations: 112 after adjustments; Convergence achieved after 32 iterations; MA Backcast: 2000M06 2000M08

The residuals were white noise, showing no specific pattern; the autocorrelations of the residuals were not significantly different from zero. Moreover, the model is both stationary and invertible, and these findings attested to the adequacy of the model. The estimated monthly CL cases for the period 2000-2009 computed using estimation equation were closely approximated to the monthly recorded cases and followed the actual trend (Figure 4) with an estimated R² of 83.5%, which suggests the usefulness of the model for disease forecasting.

The model was then used for predicting monthly CL cases for different time spans of predictions. The mean absolute percentage error (MAPE), given by

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \]

where \( n \) is the total number of months of data, \( Y_i \) and \( \hat{Y}_i \) are the recorded and predicted number of cases respectively at month \( t \), was used to validate the predicted ability of the model for different periods of time: 3, 6, 9, 12, 24, 36 and 48 months. For the period up from January 2010 to December 2014, the predicted values and the reported CL cases matched reasonably well (MAPE=72.4) as shown in Figure 5. However, when the different time spans of predictions are considered, the model affords better prediction for a 3-month period (MAPE=55.5) and for a 24-month period (MAPE=50.4).

DISCUSSION

Our data suggests that an early warning system can be strengthened by using a forecasting system for different time spans based on routine data collection within the existing surveillance system. Data analysis, interpretation and prediction, in addition to data collection, are important components of a surveillance system. There is a need to integrate forecasting methods into surveillance systems by identifying the best prediction model to serve as a base for making predictions.

The statistical approach using time series analysis based on the Box-Jenkins method predicted the number of CL cases in Biskra province. The fitted autoregressive moving average model, incorporating climate factors with recorded CL cases in this province from 2000 to 2009, was adequate in predicting the monthly number of CL from January 2010 to December 2014. Our intention was to examine the association between climate factors and CL incidence, and to identify the best model for predicting CL incidence using climate factors. The model demonstrated the linkage between
CL and weather conditions. The temperature and relative humidity were retained, which indicates that temperature had the higher effect. Moreover, the built model best predicted CL cases for a 3- and 24-month period.

In conclusion, time series analysis based on the Box-Jenkins method is one of the essential tools for following the evolution of CL. This study is an important step towards understanding the impact of weather factors on the CL and can support the design of intervention strategies in affected regions. The model also suggests that climate change will have important implications for human health in the study region. The model was based on recorded CL cases and climate factors assuming all other conditions such as environmental conditions, human behaviour, and reservoir hosts were constant over time. Hence, results should be carefully considered knowing that these conditions may vary over time. The impact of these conditions on disease incidence should be investigated further. Our findings are optimistic for forecasting CL by means of a surveillance system based on climate information.

Acknowledgments
The author thanks the Biskra Department of Public Health for providing data for the study, the National Meteorological Office for providing weather data, and the anonymous referees for helpful and valuable comments.

REFERENCES
1. World Health Organization. WHO Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of Leishmaniases. WHO Press, Geneva (2012). 2010:186.
2. Bachi F. Aspects épidémiologiques et cliniques des leishmanioses en Algérie. La Lettre de l’Infectiologue. 2006;1(Janvier-Février):9-15.
3. Institut national de la santé publique. REM: Relevé Épidémiologique Mensuel 2000-2014. Available from: http://www.ands.dz/insp/rem.html.
4. Selmane S. Stability Analysis of a Human-Phlebotomus Papatasi-Rodent Epidemic Model. Interdisciplinary Topics in Applied Mathematics, Modeling and Computational Science, Springer Proceedings in Mathematics & Statistics. 2015;117:1-4.
5. Office national des statistiques. Available from: http://www.ons.dz/.
6. Office national de météorologie. Available from: http://www.onm.meteo.dz.
7. Allard R. Use of time-series analysis in infectious disease surveillance. Bull World Health Organ. 1998;4(76):327-33.
8. EViews7. Available from: http://www.eviews.com/.
9. Box GEP, Jenkins GM. Time Series Analysis, Forecasting and Control. Holden-Day San Francisco. 1970.