Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis

P.E. Naill M. Momani

King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia

Abstract: Problem statement: Time series analysis and forecasting has become a major tool in different applications in hydrology and environmental management fields. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins, ARIMA (Autoregressive Integrated Moving Average). **Approach:** In this study we used Box-Jenkins methodology to build ARIMA model for monthly rainfall data taken for Amman airport station for the period from 1922-1999 with a total of 936 readings. **Results:** In this research, ARIMA $(1, 0, 0) (0, 1, 1)^{12}$ model was developed. This model is used to forecasting the monthly rainfall for the upcoming 10 years to help decision makers establish priorities in terms of water demand management. **Conclusion/Recommendations:** An intervention time series analysis could be used to forecast the peak values of rainfall data.

Key words: Time series analysis, rainfall model, forecasting, ARIMA

INTRODUCTION

Many methods and approaches for formulating forecasting models are available in the literature. This research exclusively deals with time series forecasting model, in particular, the Auto Regressive Integrated Moving Average (ARIMA). These models were described by Box and Jenkins[1] and further discussed in some other resources such as Walter[2-4].

The Box-Jenkins approach possesses many appealing features. It allows the manager who has only data on past years' quantities, rainfall as an example, to forecast future ones without having to search for other related time series data, for example temperature. Box-Jenkins approach also allows for the use of several time series, for example temperature, to explain the behavior of another series, for example rainfall, if these other time series data are correlated with a variable of interest and if there appears to be some cause for this correlation

Box-Jenkins (ARIMA) modeling has been successfully applied in various water and environmental management applications. The followings are examples where time series analysis and forecasting are effective:

- Water resources: Time-series analysis has become a major tool in hydrology. It is used for building mathematical models to generate synthetic hydrologic records, to forecast hydrologic events, to detect trends and shifts in hydrologic records and to fill in missing data and extend records

- Staff scheduling: A manager of an environment department would need forecast of an hourly volume and type of waste generated to be processed in order to schedule staff and equipment efficiently

- Process control: Forecasting can also be an important part of a process control system through monitoring key processes. It may be possible to determine the optimal time and extent of control action; for example, a chemical processing unit may become less efficient as hours of continuous operation increase. Forecasting the performance of the unit will be useful in planning the shutdown time and overhaul schedule

Chiew *et al.*[5] conducted a comparison of six rainfall-runoff modeling approaches to simulate daily, monthly and annual flows in eight unregulated catchments. They concluded that time-series approach can provide adequate estimates of monthly and annual yields in the water resources of the catchments. Kuo and Sun[6] employed an intervention model for average 10 days stream flow forecast and synthesis which was investigated by to deal with the extraordinary phenomena caused by typhoons and other serious abnormalities of the weather of the Tanshui River basin in Taiwan.

Time series analysis was used by Langu[7] to detect changes in rainfall and runoff patterns to search for significant changes in the components of a number of rainfall time series.

Solid waste management is another field where time series could be employed. Anastasia *et el.*[8] used Box-Jenkins methodology used for data analysis and stochastic modeling of daily municipal solid waste production. The data sets examined are the daily quantities of municipal solid wastes for consecutive day and for each day separately.

MATERIALS AND METHODS

The main stages in setting up a forecasting ARIMA model includes model identification, model parameters estimation and diagnostic checking for the identified model appropriateness for modeling and forecasting. Model Identification is the first step of this process. The data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required to make series stationary, as well as specifying the order of the regular and seasonal auto regressive and moving average polynomials necessary to adequately represent the time series model. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k. The PACF plot helps to determine how many auto regressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, trend either in the mean level or in the variance of the series.

The general model introduced by Box and Jenkins includes autoregressive and moving average parameters as well as differencing in the formulation of the model. The three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d) and moving average parameters (q). Box-Jenkins model are summarized as ARIMA (p, d, q). For example, a model described as ARIMA (1,1,1) means that this contains 1 autoregressive (p) parameter and 1 moving average (q) parameter for the time series data after it was differenced once to attain stationary

In addition to the non-seasonal ARIMA (p, d, q) model, introduced above, we could identify seasonal ARIMA (P, D, Q) parameters for our data. These parameters are: Seasonal autoregressive (P), seasonal Differencing (D) and seasonal moving average (Q). For example, ARIMA $(1,1,1)(1,1,1)^{12}$ describes a model that includes 1 autoregressive parameter, 1 moving average parameter, 1 seasonal autoregressive parameter and 1 seasonal moving average parameter.

These parameters were computed after the series was differenced once at lag 1 and differenced once at lag 12.

The general form of the above model describing the current value X_t of a time series by its own past is:

$$(1-\phi_1 B)(1-\alpha_1 B^{12})(1-B)(1-B^{12})\, X_t = (1-\theta_1 B)(1-\gamma_1 B^{12}) e_t \qquad (1)$$

Where:

$1-\phi_1 B$	= Non seasonal autoregressive of order 1
$1-\alpha_1 B^{12}$	= Seasonal autoregressive of order 1
X_t	= The current value of the time series examined
B	= The backward shift operator $BX_t = X_{t-1}$ and $B^{12}X_t = X_{t-12}$
$1-B$	= 1st order nonseasonal difference
$1-B^{12}$	= Seasonal difference of order 1
$1-\theta_1 B$	= Non seasonal moving average of order 1
$1-\gamma_1 B^{12}$	= Seasonal moving average of order 1

This model can be multiplied out and used for forecasting after the model parameters were estimated, as we discussed below.

After choosing the most appropriate model (step 1 above) the model parameters are estimated (step 2) by using the least square method. In this step, values of the parameters are chosen to make the Sum of the Squared Residuals (SSR) between the real data and the estimated values as small as possible. In general, nonlinear estimation method is used to estimate the above identified parameters to maximize the likelihood (probability) of the observed series given the parameter values[*]. The methodology uses the following criteria in parameter estimation:

- The estimation procedure stops when the change in all parameters estimate between iterations reaches a minimal change of 0.001
- The parameters estimation procedure stops when the SSR between iterations reaches a minimal change of 0.0001

In diagnose checking step (step three), the residuals from the fitted model shall be examined against adequacy. This is usually done by correlation analysis through the residual ACF plots and the goodness-of-fit test by means of Chi-square statistics χ^2. If the residuals are correlated, then the model should be refined as in step one above. Otherwise, the autocorrelations are white noise and the model is adequate to represent our time series.

After the application of the previous procedure for a given time series, a calibrated model will be developed which has enclosed the basic statistical properties of the time series into its parameters (step four). For example, the developed model, as shown in Eq. 1 above can be multiplied out and the general model is written in terms of X_t.

Case study: According to Jordanian Ministry of Water and Irrigation[9] Jordan is located 80 kilometers east of the eastern coast of the Mediterranean Sea. Its location between 29°11'N and 33°22'N and between 34°19'E and 39°18'E with an area of 89329 km². In Jordan, more than 80% of the country is classified as arid areas with an average of rainfall ranges from 600 mm years^{-1} in the north to less than 50 mm year^{-1} in the south. The precipitation pattern is both latitude and altitude dependent. In addition, water resources in Jordan are limited and with deteriorating quality due to urban development. Therefore, it is important to know the future water resources budget in order to help decision makers improve their decisions with taking consideration the available and future water resources. Additionally, using modeling and forecasting for future water resources becomes possible with advances in forecasting methodologies such as time series analysis.

The rainy season is between October and May where 80% of the annual rainfall occurs through December to March. Jordan witnessed rainy seasons above average for the years 1970/1971 and 1991/1992 where the last one considered the highest in the last 75 years.

The climate in Jordan is predominantly of the Mediterranean type. Hot and dry summer and cool wet winter with two short transitional periods in autumn and spring. Four climatic regions are distinguishable in Jordan. They are:

- The Jordan rift valley (Al-Ghor): The climate of Al-Ghor is classified as tropical. It is very hot in summer and warm in winter with an annual rainfall of 150-250 mm. The elevation of the Ghor is below Mean Sea Level ranging from 200-400 m. Its width ranges from 15 km in the North to 30 km in the South

- The mountainous (hilly) region: The climate of these Regions is rather mild in summer and cold in winter. The amount of rainfall ranges from 300-600 mm year^{-1}. Snowfall occurs over the mountains. This region lies to the east of the Jordan rift valley extending from North to South. Its elevation varies from 750-1200 m with some tops exceeding 1700 m
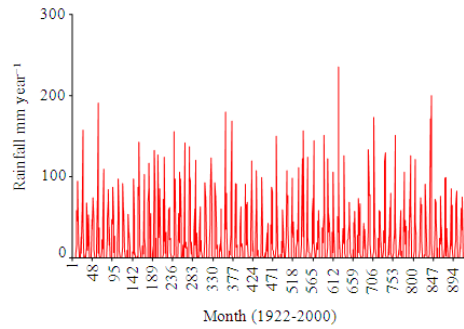

Fig. 1: Monthly rainfall data for Amman airport station

Table 1: Basic statistics for Amman station monthly rainfall data in millimeter

No of observations	Mean	St. Dev.	Variance	Min	Max
936	22.47	35.896	1288.52	0	235.2

- The badia region: A flat terrain that lies to the east of the high lands with an elevation varies from 600-700m. It is characterized by dry hot summer and relatively cold dry winters with rainfall varying from 40-100 mm year^{-1}

- The Gulf of Aqaba: This area is considered very hot in summer and warm in winter with an amount of rainfall <50 mm year^{-1}

Amman is the capital city of Jordan where more than one million people live. In this research, readings from Amman Airport Station monthly rainfall data were considered, as shown in Fig. 1. The descriptive statistics for our data is shown in Table 1. This station is operated by the Metrological Department, climate division, who operates 37 other stations distributed throughout the country. Since rainfall data for this station is around the average among these stations it is used as a case study in our analysis.

RESULTS AND DISCUSSION

Since the data is a monthly rainfall, Fig. 1, shows that there is a seasonal cycle of the series and the series is not stationary. The ACF and PACF of the original data, as shown in Fig. 2, show that the rainfall data is not stationary.

In order to fit an ARIMA model stationary data in both variance and mean are needed. We could attain stationarity in the variance could be attained by having log transformation and differencing of the original data to attain stationary in the mean. For our data, we need to have seasonal first difference, d = 1, of the original data in order to have stationary series. After that, we need to test the ACF and PACF for the differenced series to check stationary.
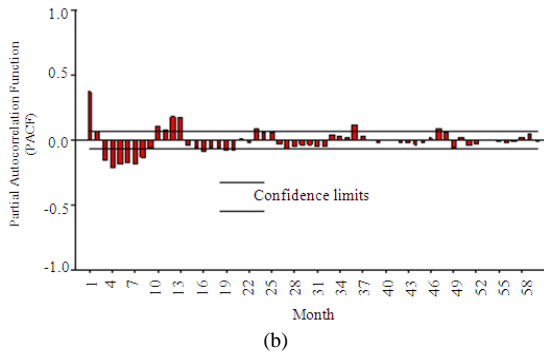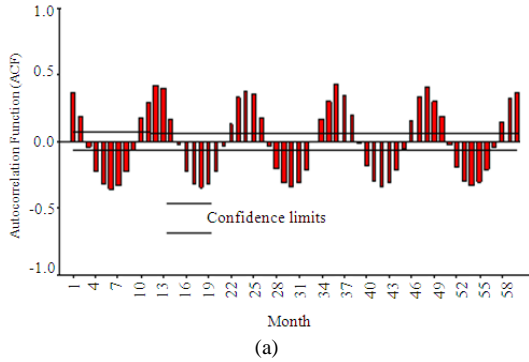
Fig. 2: (a): Autocorrelation (ACF); (b): Partial Autocorrelation (PACF) for original rainfall data
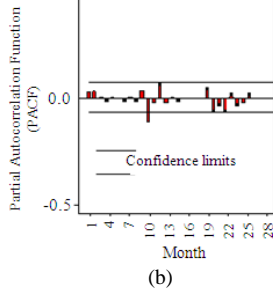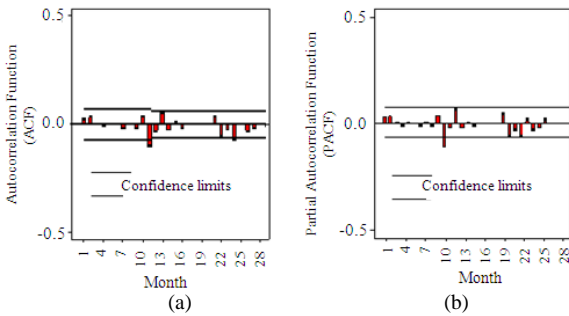

Fig. 3: (a): Autocorrelation (ACF); (b): Partial Autocorrelation (PACF) for first order seasonal differencing and de-seasonalized original rainfall data

As shown in Fig. 3, the ACF and PACF for the differenced and de-seasonalized rainfall data are almost stable which support the assumption that the series is stationary in both the mean and the variance after having 1st order non seasonal difference. Therefore, an ARIMA $(p, 0, q) (P, 1, Q)^{12}$ model could be identified for the differenced and de-seasonalized rainfall data. After ARIMA model was identified above, the p, q, P and Q parameters need to be identified for our model.
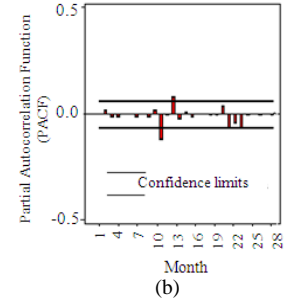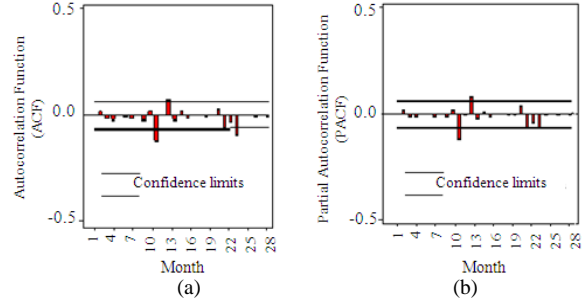

Fig. 4: (a): Autocorrelation (ACF); (b): Partial Autocorrelation (PACF) for residual errors resulted from ARIMA $(1, 0, 1) (1, 1,1)^{12}$ model

Table 2: ARIMA $(1, 0, 1) (1, 1, 1)^{12}$ model parameter characteristics

Parameter	Parameter value	T-value
ϕ_1	0.4308	0.40
α_1	-0.0267	-0.81
θ_1	0.4077	0.37
γ_1	0.9805	200.31

In Fig. 3, we have one autoregressive (p) and one moving average (q) parameters and the ACF has exponential decay starting at lag 12. Similarly, the PACF has an exponential decay starting at lag 12. These patterns suggest that we have ARIMA (1, 1) for seasonal and non seasonal rainfall data. Since we have identified the first order seasonal difference for the rainfall data our tentative model will be ARIMA $(1, 0, 1) (1, 1, 1)^{12}$. In order to make sure that this model is representative for our data and could be used to forecast the upcoming rainfall data we need to test the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) for the residual errors resulted from fitting such model for our data. Figure 4 shows the residual errors for the ARIMA fitted model. It is clear that we have no pattern was observed in the residual errors which shows that the model could be used to represent our data.

After we fitted an ARIMA $(1, 0, 1) (1, 1, 1)^{12}$ for our data we need to estimate parameters values for our model, as shown in Eq. 2. As a rule of thump, in ARIMA modeling we need to minimize the sum squared of residuals needs to be minimized between the forecasted and existing values:

$$(1-\phi_1 B)(1-\alpha_1 B^{12})(1-B^{12}) X_t = (1-\theta_1 B) (1-\gamma_1 B^{12})e_t \qquad (2)$$

The sum squared of residuals for the model was 655288 and the following parameters values, as shown in Table 2.

As shown in Table 2, the parameters t-statistics are relatively small except for γ_1 parameter. Comparing

these values with the value $t_{0.05}(932) = 1.645$ (936 is the length of time series minus the number p, q, P, Q parameters). It is important to note that not only a model to fit the data with minimum sum of squared residuals is needed, but also a model with the least parameters is needed. Therefore, the correlation matrix was tested to check if any of the parameters are correlated in order to eliminate any of the correlated ones. Table 3, shows that the ϕ_1 and θ_1 parameters have high correlation value which suggests that one of these parameters could be eliminated from our model.

By eliminating any of the p or q parameters we will end up with the following models: ARIMA $(1, 0, 0) (1, 1, 1)^{12}$ model, the sum of squared residuals is 737590 and ARIMA $(0, 0, 1) (1, 1, 1)^{12}$, the sum of squared residuals is 737984. Table 4, shows the parameter values for the model identified above and their t-statistics values. These results show that both models have similar characteristics and could be representative time series models for our data. ARIMA $(1, 0, 0) (1, 1, 1)^{12}$ will be considered for further analysis.

Can any of the model parameters be eliminated in the model above without affecting model appropriateness for our data? As shown in Table 5, there is no strong correlation between parameter values. Therefore, all parameters are important to build the ARIMA model. On the other hand, by comparing the t-statistics values for the model parameters α_1 parameter

has small value compared with t-statistics value $t_{0.05}(934) = 1.645$. Therefore, it is not significant to represent the data in the model above.

As discussed previously, the best ARIMA model has the least parameters numbers with the least squared residuals. Thus, the P parameter will be eliminated from our model and see if ARIMA $(1, 0, 0) (0, 1, 1)^{12}$ is more appropriate to model our rainfall data. The sum of squared residuals from this model is 737593, which is not significantly different from the one resulted from ARIMA $(1, 0, 1) (1, 1, 1)^{12}$ discussed above. The t-statistics for ARIMA $(1, 0, 0) (0, 1, 1)^{12}$ model parameters are significant to represent our data in ARIMA model if we compare these values with t-statistics value $t_{0.05} = 1.645$, as shown in Table 6.

The ACF and PACF of the residuals resulted from our model should not show any pattern. It is clear, as shown in Fig. 5, that there is no pattern in residuals. In addition, we measured the goodness-of-fit test by means of Chi-square statistics χ^2. The value χ^2 for the autocorrelations up to lag 24 is 26.9. The value of χ^2 for 22 degree of freedom (24- p- q) is 33.9244. Therefore, the set of autocorrelations for residuals, as shown in Fig. 5, are not significant and considered white noise since the observed χ^2 value of 26.9 is less than $\chi^2_{0.05}$ of 33.9244.

Finally, this conclude that ARIMA $(1, 0, 0) (0, 1, 1)^{12}$ model identified previously is adequate to represent our data and could be used to forecast the upcoming rainfall data.

After the model parameters were estimated, they would be used to forecast the upcoming rainfall data.

Table 3: ARIMA $(1, 0, 1) (1, 1, 1)^{12}$ model parameter correlation matrix

Parameter	ϕ_1	α_1	θ_1	γ_1
ϕ_1	1.000	-0.058	1.000	-0.008
α_1	-0.058	1.000	-0.060	0.057
θ_1	1.000	-0.060	1.000	-0.008
γ_1	-0.008	0.057	-0.008	1.000

Table 4: ARIMA $(1, 0, 0) (1, 1, 1)^{12}$ and ARIMA $(0, 0, 1) (1, 1, 1)^{12}$ model characteristics

Parameter	ARIMA $(1, 0, 0) (1, 1, 1)^{12}$		ARIMA $(0, 0, 1) (1, 1, 1)^{12}$	
	Parameter value	T- value	Parameter value	T-value
ϕ_1	0.0714	2.15	-------------	-------
α_1	0.0025	0.07	0.0021	-0.06
θ_1	----------	-------	0.0037	-1.92
γ_1	0.8909	55.33	0.8902	55.04
Sum of Squares residuals	737590	737984		

Table 5: ARIMA $(1, 0, 0) (1, 1, 1)^{12}$ model parameter correlation matrix

Parameter	ϕ_1	α_1	γ_1
ϕ_1	1.000	0.100	0.105
α_1	0.100	1.000	0.442
γ_1	0.105	0.442	1.000

Table 6: ARIMA $(1, 0, 0) (0, 1, 1)^{12}$ model parameter correlation matrix

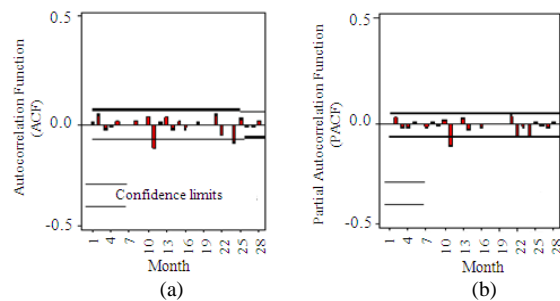| Parameter	Parameter value	T-value
ϕ_1	0.15	4.4
γ_1	0.88	44.8

Fig. 5: (a): Autocorrelation (ACF) (b): Partial Autocorrelation (PACF) for residual errors resulted from ARIMA $(1, 0, 0) (0, 1, 1)^{12}$ model
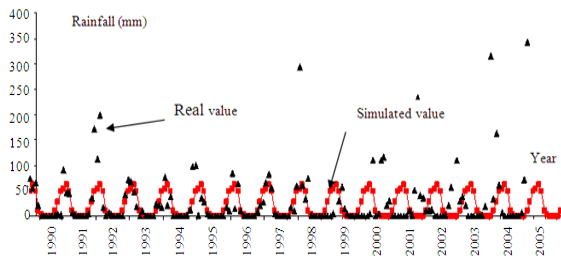
Fig. 6: Real and simulated values for 1990-2005 of overall time series by using ARIMA (1, 0, 0) (0, 1, 1)^{12} model

As we discussed before ARIMA (1, 0, 0) (0, 1, 1)^{12} model could be written in the following form, as shown in Eq. 3:

$$(1-\phi_1 B)(1-B^{12})\, X_t = (1-\gamma_1 B^{12})\, e_t \qquad (3)$$

This equation can be multiplied out and written in a form that is used in forecasting as shown in Eq. 4:

$$X_t = (1+\phi_1)X_{t-12} + \phi_1 X_{t-1} + e_t - \gamma_1 e_{t-12} \qquad (4)$$

In Eq. 4, the value of X_t could be estimated by substitution the parameter values as we estimated above. Figure 6, shows a comparison between the real values and the ones resulted from the developed ARIMA model for the period between 1990 and 1999. It is clear that the model was not able to represent the peak values. In addition, it is clear that rainfall pattern continues for the upcoming years and there is no indication that the amount of rainfall decreases with time.

CONCLUSION

Time series analysis is an important tool in modeling and forecasting. ARIMA (1, 0, 0) (0, 1, 1)^{12} model give us information that can help the decision makers establish strategies, priorities and proper use of water resources in Jordan. This piece of information was not appropriate to predict the exactly monthly rainfall data. Therefore, individual monthly rainfall data should not be used in decision making by depending on our model. However, an intervention time series analysis can be tested to see if we can improve our model performance in forecasting the peak values of rainfall data

REFERENCES

1. Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. Revised Edn., Hoden-Day, San Francisco. http://adsabs.harvard.edu/abs/1976tsaf.conf.....B
2. Walter Vandaele, 1983. Applied time series and Box-Jenkins Models. Academic Press Inc., Orlando, Florida, ISBN: 10: 0127126503, pp: 417.
3. Chatfild, C., 1996. The Analysis of Time Series-an introduction, 5th Edn., Chapman and Hall, UK.
4. Montgomery, D.C. and L.A. Johnson, 1967. Forecasting and Time Series Analysis. McGrow-Hill Book Company, http://www.abebooks.com/Forecasting-Time-Series-Analysis-Montgomery-Douglas/1323032148/bd
5. Chiew, F.H.S., M.J. Stewardson and T.A. McMahon, 1993. Comparison of six rainfall-runoff modeling approaches. J. Hydrol., 147: 1-36. http://cat.inist.fr/?aModele=afficheN&cpsidt=4788795
6. Kuo, J.T. and Y.H. Sun, 1993. An intervention model for average 10 day stream flow forecast and synthesis. J. Hydrol., 151: 35-56. http://cat.inist.fr/?aModele=afficheN&cpsidt=3832268
7. Langu, E.M., 1993. Detection of changes in rainfall and runoff patterns. J. Hydrol., 147: 153-167. http://cat.inist.fr/?aModele=afficheN&cpsidt=4788801
8. Katsamaki, A., S. Willems and E. Diamadopoulos, 1998. Time series analysis of municipal solid waste generation rates. J. Environ. Eng., 124: 178-183. http://cat.inist.fr/?aModele=afficheN&cpsidt=2143876
9. Water Authority of Jordan, 2008. Jordanian Ministry of water and irrigation. http://www.mwi.gov.jo/