International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:9, No:4, 2015

New Hybrid Method to Model Extreme Rainfalls

Y. Laaroussi, Z. Guennoun, A. Amar

***Abstract*—**Modeling and forecasting dynamics of rainfall occurrences constitute one of the major topics, which have been largely treated by statisticians, hydrologists, climatologists and many other groups of scientists. In the same issue, we propose, in the present paper, a new hybrid method, which combines Extreme Values and fractal theories. We illustrate the use of our methodology for transformed Emberger Index series, constructed basing on data recorded in Oujda (Morocco).

The index is treated at first by Peaks Over Threshold (POT) approach, to identify excess observations over an optimal threshold u. In the second step, we consider the resulting excess as a fractal object included in one dimensional space of time. We identify fractal dimension by the box counting. We discuss the prospect descriptions of rainfall data sets under Generalized Pareto Distribution, assured by Extreme Values Theory (EVT). We show that, despite of the appropriateness of return periods given by POT approach, the introduction of fractal dimension provides accurate interpretation results, which can ameliorate apprehension of rainfall occurrences.

***Keywords*—**Extreme values theory, Fractals dimensions, Peaks Over Threshold, Rainfall occurrences.

I. Introduction

The analysis and the modeling of rainfall occurrences is typical problem in applied hydrometeorology [1]-[4]. The development of a rainfall occurrence model is increasingly in demand; not only for data generation purposes, but also to provide some useful information in various applications which includes water resource planning and management [5], [6].

Literature offers two categories of rainfall modeling. The first one concerns occurrences [7], while the second category is devoted to extreme nature of rainfall using Extreme Value Theory [8], especially the Peaks Over Threshold (POT) approach [9]. This category has received an increasing attention, especially with rising multiple types and occurrences of climatic extreme events [10], [11]. The intermittency of those events is one reason to use multifractal models which can accommodate the spatial patterns at long accumulation sizes [12].

In this paper, we develop a hybrid approach based on fractals and Extreme value theories. To analyze performances of this approach we elaborate a monthly climatic composite index based on Emberger index. Observations from the elaborated time series have both the advantage to meet the validity conditions of extreme values theory and also the combination of the most representative factors of climate, namely rainfalls.

Y. Laaroussi, Z. Guennoun, and A. Amar are with the Faculty of Sciences Rabat, University Mohammed V-Agdal, Department of Mathematics and Computer Science, 4 Avenue Ibn Battuta BP 1014 RP, Rabat, Morocco (e-mail: laaroussi.youness@gmail.com, guennoun@fsr.ac.ma, amar.abd@gmail.com).

The L. Emberger [13] pluviothermic quotient is expressed as:

$$Q = \frac{1000P}{\frac{M+m}{2}(M-m)} \tag{1}$$

Rainfalls are represented by the annual average of rainfall (P). Temperature is characterized by the average of the minimum temperature of coldest month (m) and the average of the maximum temperature of warmest month (M). Evaporation is evaluated by using the extreme amplitude (M-m).

We identify excess of monthly climatic composite index (see (9)), determined by POT approach. On this set of observations, we use fractal theory to model the chaotic behavior and we compare results with assumed asymptotic distribution, named Generalized Pareto Distribution. The use of POT avoids drawbacks of empirical threshold, used to distinguish rainy and non rainy days. Rain-day may be given as a day with measurable rain or as a day with a total rainfall greater than a selected threshold [14]-[16].

The definition of a rain-day or what is considered the daily rainfall threshold (DRT) is a critical decision, because it has a fundamental influence on any analysis of the rainfall regime such as the rainy season length (RSL), the number, duration, and yield of rain-spells [17]. In fact, the selection of a threshold should be chosen as a compromise. If the threshold is set too low, many events with minute rainfall will not be screened out and will remain as irrelevant data in an analysis of extreme. On the other hand, if the threshold is set too high, there will be few data for analysis and the results may be highly sensitive to the depths recorded in just one or two events [18].

In this regard, our study intends to use Peaks Over Threshold (POT) approach, to identify excess observations over an optimal threshold u. This threshold is identified by a statistical objective approach, namely called Multiple Threshold Method (MTM) [19].

It should be noted that in our study, we use not directly rainfall but transformed Emberger Index, because rainfall data are not statistically independents. In addition, results obtained from the index allow more general conclusions and global finding, especially because this index is intimately related with rainfall and climate behavior [13]. Resulting series obtained from threshold excess index allow identifying fractal dimension of exceeding occurrence. Despite of the appropriateness of return periods given by POT approach, the introduction of fractal dimension provides accurate interpretation results, which can ameliorate modeling and apprehension of erratic characteristics of climatic variables, such as rainfalls.

II. Materials And Methods

A. Extreme Values Theory and Peaks over Threshold Approach

Climatologists have begun to use Extreme Value Theory in recent years. Reference [20] shows that EVT is entirely appropriate, to solve a range of problems related to climatic extremes. Conceptually, there are two related ways of identifying extremes in real data. The first approach considers maximum of observations in successive periods, for example, months or years. These selected observations constitute the extreme events. The second approach focuses on the realizations exceeding a given threshold u. The block maxima are the traditional method used to analyze data with seasonality, as for instance hydrological data. However, the threshold method uses data more efficiently and, for that reason, seems to become the most chosen method in recent applications. The main objective of called Peaks over threshold (POT) method is to consider the distribution function F_u of values of x above a certain threshold u, when the distribution F of X is unknown (or in the case when it is difficult to extract F_u on the basis of known distribution F).

Reference [21] confirms that for a large class of underlying distribution functions F, the conditional excess distribution function $F_u(y)$, for u large, is well approximated by:

$$F_u(y) \approx G_{\xi,\sigma}(y), y \to \infty \tag{2}$$

where:

$$G_{\xi,\sigma}(y) = \begin{cases} 1-(1+\frac{\xi}{\sigma})^{-\frac{1}{\xi}} & if \quad \xi \neq 0 \\ 1-e^{-y/\sigma} & if \quad \xi = 0 \end{cases} \tag{3}$$

For $y \in [0,(x_F - u)]$ if $\xi \geq 0$ and $y \in [0,-\frac{\xi}{\sigma}]$ if. $\xi<0$ $G_{\xi,\sigma}$ is the so called Generalized Pareto Distribution (GPD). Where σ is the scale, ξ is the shape. The tail index ξ gives an indication of the heaviness of the tail: the larger ξ, the heavier the tail.

It should be noted that the implementation of POT method involves the following steps:

- ❖ Select the threshold u.
- ❖ Fit the GPD function to the exceedances over u.
- ❖ Compute point and interval parameter estimations.

Selection threshold step is the most critical one. Theory tells us that u should be high in order to satisfy Pickands-Balkema-de Haan theorem, but the higher the threshold is, the less observations are left for the estimation of the parameters of the tail distribution function.

Numerical methods are based on objective formulas. We choose in this paper, to use a multiple threshold method (MTM) introduced by [21].

B. Fractal Dimension

Many natural phenomena are better described using a dimension between two whole numbers, i.e., a fraction. Thus, a fractal curve can have a dimension between one and two in contrast to the straight line that is one dimensional. The fractal dimension measures how much complexity is being repeated at each scale. The notion of dimension most commonly used is that of Hausdorff dimension [22]-[24]. The main problem with Hausdorff dimension is it can be fairly hard to calculate in general. So, some other notions of dimension have been developed which are easier to calculate, such as box-counting dimension:

Definition: Let $F \subset R^n$ be a bounded set, $N_\delta(F)$ is the smallest whole number of at most diameter δ covering box F. The upper and lower dimensions are defined respectively by:

$$\underline{\dim}_B(F) = \liminf_{\delta \to 0} \frac{\log N_\delta(F)}{-\log \delta} \tag{4}$$

and

$$\overline{\dim}_B(F) = \limsup_{\delta \to 0} \frac{\log N_\delta(F)}{-\log \delta} \tag{5}$$

If both limits are equal, the box dimension of F is:

$$\dim_B(F) = \lim_{\delta \to 0} \frac{\log N_\delta(F)}{-\log \delta} \tag{6}$$

In practices, we use box counting algorithm, which is intuitive and easy to apply.

Given a sequence (ε_n) decreasing and tends to 0 slowly enough (as a geometric sequence, for example). The fractal object (E) is covered with a mesh network square of side (ε_n), and includes the number Ω_n of square meeting point E.
The fractal dimension is then:

$$D(E) = \lim_{n \to \infty} [\frac{\log(\Omega_n)}{\log(\frac{1}{\varepsilon_n})}] \tag{7}$$

The corresponding log-log diagram is defined as:

$$\left(\log(\frac{1}{\varepsilon_n}), \log(\Omega_n) \right) \tag{8}$$

For different value of ε_n, we find the number Ω_n. After this we draw this plot, and the fractal dimension is the opposite of the slope.

This method is very simple to use, but it has some serious drawbacks. In particular, if $\frac{1}{\varepsilon_n}$ is not an integer, the square of side ε_n overflow will generally left and right of the graph of series, which distorts results and introduces irregularities in the diagram (8), especially when ε_n is great.

C. The Proposed Framework

To implement our methodology, we suggest the following algorithm steps:

Step 1. Construct transformed Embreger Index ($I_{transformed}$) observations based on:

$$I_{Transformed} = \frac{1000P}{\frac{M+m}{2}(M-m)} \tag{9}$$

where P is the monthly maximum rainfall, M is the monthly maximum temperatures and m is the monthly minimum temperatures.

Step 2. Apply POT approach combined with MTM Threshold selection method to data obtained from step1 and adjust appropriate GPD. In this step, we have the value of threshold (u), observations which exceed the value (u) (excess over u) and GPD adjusted to excess.

Step 3. Assign the value 1, if the observation (from step1) exceeds the threshold u (from step2), 0 if not. Then the representation of a new constructed process is obtained by Fig. 1:

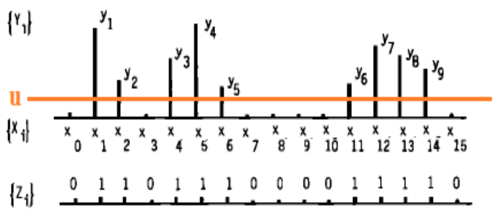

Fig. 1 Schematic representation of climatic process (transformed Emberger index) and definition of the X_i , Y_i and Z_i series.

Note that X_i is constructed Emberger index series, Y_i is the series of excxess over thresheold u and Z_i is the binary series of occurred and no occurred excess.

Step 4. Calculate fractal dimension of Z_i series (step3) by Box counting method and characterize its chaotic behavior.

Step 5. Interpret obtained results, especially return periods and fractal dimension.

D. Data

The studied series is based on climate index, which contains variables climate (rainfalls and temperature). Corresponding climate index data are recorded in Oujda.

The choice of this region is justified by intensity of extreme events which occur frequently in this region. The Oriental Region has a strategic location in the Mediterranean, because of its proximity to Europe and its immediate neighborhood with Algeria and the rest of the Maghreb.

Rainfalls recorded annually in the eastern region are low, random and sparse. Rainfall varies between years and during the year. The maximum temperature is stable or it varies little during the year.

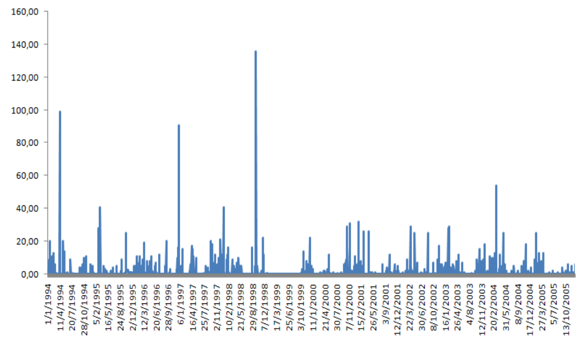

Fig. 2 Evolution of monthly extreme rainfall in Oujda, during 1994 to 2005

In summer, Temperature Max is always above 30°C, recorded mainly during the months of July and August. The minimum temperature is usually recorded in winter and especially during the month of January. Fig. 2 shows the evolution of rainfall during the period 1994-2005. The evolution presents two different patterns: The first is from 1994 to 1999 and characterized by three notable pics (March 1994, November 1996 and August 1998). Rainfalls in the second period are very volatile and not exceed 55 mm. We can note also seasonal characteristics of rainfalls.

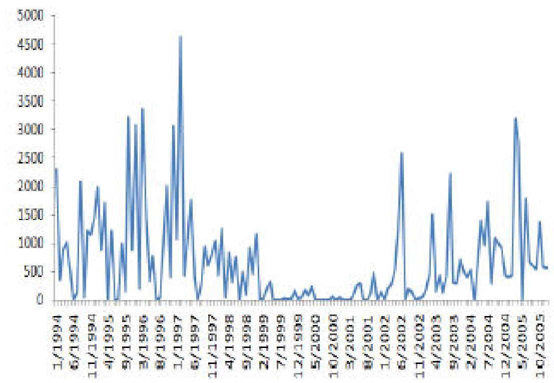

Fig. 3 Evolution of elaborated index (I)

Fig. 3 shows that, the representation of ($I_{transformed}$) is very close to rainfalls one. $I_{transformed}$ has the same structure of rainfall evolution. In addition, we can note that $I_{transformed}$ is very volatile so, this justifies the study of extreme values of this index. In the following, we will focus on the application of our methodology to $I_{transformed}$ series. One objective is to compare POT approach and fractal theory results.

III. Results and Discussions

The application of MTM algorithm leads to a threshold value about 440. The estimation of shape and scale parameters is obtained by maximizing the likelihood function. This approach is asymptotically the best for large samples. So, we can fit the following distribution to transformed Embreger Index excess:

$$F_{440}(x, 834.58; 0.03) = 1 - \left(1 - 0.03\frac{x - 440}{834.58}\right)^{-\frac{1}{0.03}} \qquad (10)$$

These results allow assuming that ($I_{transformed}$) distribution belongs to a family of Frechet, with a tail index equal to 0.03. On the statistical level, this finding illustrates lack of normality, excess skewness and leptokurtosis ($I_{transformed}$).

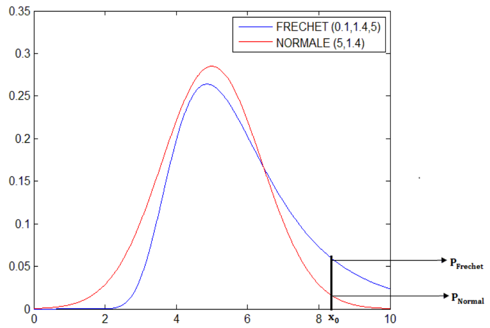

Fig. 4 Fréchet and normal distributions

Fig. 4 shows for the same quintile x_0, that the probability calculated by Normal distribution is less than the one calculated by Fréchet distribution. So, Normal distribution underestimates the probability of realization of an extreme event. In This case, the use of Fréchet distribution allows alerting decision makers in the good times to guard against risks.

Fréchet distribution is largely used. Applications range from accelerated life testing through to earthquakes, floods, horse racing, rainfall, queues supermarkets, sea currents, winds speeds and track race. In our case, Fréchet distribution confirms that ($I_{transformed}$) exceeds very often (high probability) the large values. This, characterize climate instability which is illustrated by extreme events occurred in eastern region of Morocco, and which can be related to precipitations and temperatures. This statement is corroborated by several climate statistics, which confirm that the absolute duration of summer drought in the region, can greatly exceed two months. To achieve our hybrid method, we proceed in the following, to model the chaotic behavior of $I_{transformed}$ excess occurrences. This will be done by the determination of its fractal dimension based on box counting method.

The implementation of Box counting method consists to identify the slope (D) of the graph represented by:

$$\left(\log(\frac{1}{\varepsilon_n}), \log(N(\varepsilon_n))\right) \qquad (11)$$

ε_n is the size of boxes; $N(\varepsilon_n)$ is the minimum of number of boxes, needed to cover the hole of represented series (in our case, the series is $I_{transformed}$ excess occurrences).

To determine number and size of boxes, we elaborated a computing program on VBA. Results are given in Table I:

TABLE I
NUMBER OF BOXES FOR DIFFERENT SIZES

Size of boxes (Months)		Number of boxes	LOG	
			Size of boxes (Months)	Number of boxes
2^0 =	1	68	0,00	4,22
2^1 =	2	41	0,69	3,71
2^2 =	4	26	1,39	3,26
2^3 =	8	14	2,08	2,64
2^4 =	16	8	2,77	2,08
2^5 =	32	5	3,47	1,61
2^6 =	64	3	4,16	1,10
2^7 =	128	2	4,85	0,69
2^8 =	256	1	5,55	0,00

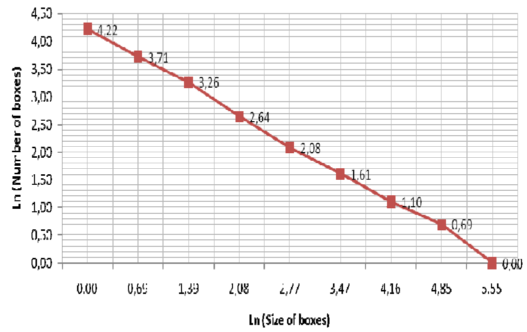

Fig. 5 Representation of Ln (size of boxes) and their corresponding Ln (Number of boxes)

Fig. 5 shows that the relation between Log (size of boxes) and Log (number of boxes) can be represented by a linear relation. As practically considered, an object shows fractal behavior when the log (unit of measurement) plotted against the log (number of units, N) for several scales is a straight line. In this case, the fractal dimension is given by the opposite of the slope, D, in:

$$\log(N(\varepsilon_n)) = -D.\log(\varepsilon_n) + c \qquad (12)$$

where c is a constant. To determine D, we regress $\log(N(\varepsilon))$ on $\log(\varepsilon)$, Tables II and III show the obtained results:

TABLE II
REGRESSION RESULTS (MODEL SUMMARY)

R Square	Adjusted R Square	Std. Error of the Estimate
0,998	0,998	0,06059

TABLE III
REGRESSION RESULTS (COEFFICIENTS)

Model	Unstandardized Coefficients		Standardized Coefficients
	B	Std. Error	Beta
(Constant)	4,231	0,037	
Size_Boxes	-0,752	0,011	-0,999

Regression results characterize the chaotic behavior of $I_{transformed}$ excesses occurrences by fractal dimension of value about 0,75.

$I_{transformed}$ excesses occurrences related to Oujda location, defined as a set of excess periods observed, is regarded as a fractal object belonging to one-dimensional space of time (a time series is 1-dimensional). So necessarily, the dimension of our index is bounded by 0 and 1. This means that the index occupy a fraction of the available space greater than a single point (dimension 0), and lower than a line (dimension 1).

As defined in methodology section, fractal dimension provides a useful tool to quantify systems complexity. In fact, fractal dimension counts the effective number of degrees of freedom in the dynamical system. In our case study, the value of 0.7, close to 1, confirms that Oujda region is characterized by a predominance of extreme events, given the fact that we privileged the study of excess above a given threshold.

In addition, we can evoke an important conclusion related to resulting fractal dimension. The value about 0.7 near to one and which means that the constructed index, exceeds the given threshold 68 times from 143 months, confirms the high frequency of extreme events, in addition to high variability rainfall amounts, in Mediterranean climate.

It should be noted that the choice of threshold impacts the value of fractal dimension. Thus, as mention in the table below, for different values of threshold, chosen by varied methods as mean of series values or MTM, one can obtain different values of fractal dimensions (see Table IV):

TABLE IV
FRACTAL DIMENSION FOR DIFFERENT VALUES OF THRESHOLD

Selection Threshold	Threshold	Fractal dimension
Arbitrarily	1800	Invalid (non fractal behavior)
Mean of series values	669	0,69
MTM	**440**	**0,75**
Arbitrarily	300	0,78
Arbitrarily	50	0,83 (graph very close to a line)

In comparison with MTM threshold considered as an optimal value because chosen objectively, we find that a very remote values (50 or 1800) give very confused results. For 50 we obtain nearly a line graph while for 1800, we conclude that excess value has no fractal behavior.

Thus, as mention early, the choice of threshold represents a crucial step in this kind of studies. This, justify the use of MTM for the threshold selection.

To confirm the high frequency of extreme events in the studied region, we explore the relationship between the probability of occurrence of an event corresponding to p quintile and the return period T (T>=2), given as:

$$p = 1 - \frac{1}{T} \tag{13}$$

The return period and frequency are statistical descriptors of the severity of an event. The return period is the expected length of time between two events that exceed a specific magnitude. Frequency, or exceedance probability, is the inverse of the return period. For our case we obtain very small values of return periods. This corroborates the obtained fractal dimension value (0.75).

The high information quantity is obtained in the high frequency domain. By calculating fractal dimension of excess index occurrence, it was statistically proven the existence of relationship between this dimension and frequency. For our study, we have proved empirically that a low return period is associated with a high frequency and this implies a high fractal dimension, so chaotic events are usually characterized by a high frequency. These results allow concluding that the studied region is exposed to many climatic extreme events.

IV. CONCLUSION

A new hybrid method, which combines Extreme Values and fractal theories, is proposed for describing and analyzing time series behavior. Performances of this new method are analyzed on both empirical and theoretical views. It should be noted that the pretreatment of Embrger index which appears in empirical part of this article, was justified by needs to meet some necessary theoretical hypothesis.

As results shows, our methodology approach avoids drawbacks of classic threshold choice methods, provides a simple algorithm for calculating fractal dimension of extreme events and, exhibits a close relationship between this dimension and frequency.

REFERENCES

[1] A. mirataee, B., Montaseri, M. and Rezaei, H., “Assessment of goodness of fit methods in determining the best regional probability distribution of rainfall data” International Journal of Engineering-Transactions A: Basics, Vol. 27, No. 10, (2014), pp. 1537-1546.
[2] Evin, G. and Favre, A.C., “Further developments of a transient Poisson-cluster model for rainfall”, Stochastic environmental research and risk assessment, vol. 27, (2013), pp. 831-847.
[3] Koutsoyiannis, D., “Statistics of extremes and estimation of extreme rainfall”, Theoretical investigation. Hydrological Sciences Journal, vol. 49, (2004), pp. 575–590.
[4] Ceresetti, D., Anquetin, S., Molinie, G., Leblois, E. and Creutin, J. D., “Multiscale Evaluation of Extreme Rainfall Event Predictions Using Severity Diagrams”, Weather and Forecasting, vol. 27, (2012), pp. 174-188.
[5] Barkotulla, M. A. B., “Stochastic Generation of the Occurrence and Amount of Daily Rainfall”. Department of Crop Science and Technology University of Rajshahi Rajshahi-6205, Bangladesh. Pak.j.stat.oper.res, vol. 6, No. 1, (2010), pp. 61-73.
[6] Sayang, M. D. and Jemain, A. A., “Fitting the distribution of dry and wet spells with alternative probability models”. Meteorology and Atmospheric Physics, vol. 104, (2009), pp. 13-27.
[7] Masala and Giovanni, “Rainfall derivatives pricing with an underlying semi-Markov model for precipitation occurrences”. Stochastic Environmental Research and Risk Assessment, vol. 28, (2014), pp. 717 – 727.
[8] Vidal, I., “A Bayesian analysis of the Gumbel distribution: an application to extreme rainfall data”. Stochastic Environmental Research and Risk Assessment, vol. 28, (2014), pp. 571-582.
[9] Andrés, M. A., Patricia, Z. B. and Manuel, G. S., “Comparing Generalized Pareto models fitted to extreme observations: an application to the largest temperatures in Spain”. Stochastic Environmental Research and Risk Assessment, vol. 28, (2014), pp. 1221-1233.
[10] Christidis, Nikolaos, Peter A., Adam, S., Scaife, A., Alberto, A., Gareth, S. J., Dan, C., Jeff, R. K. and Warren, J. T., “A New HadGEM3-A-Based System for Attribution of Weather and Climate-Related Extreme Events”. Journal of Climate, vol. 26, (2013), pp. 2756-2783.
[11] Hristidis, N., Stott, P.A., Scaife, A., Arribas, A., Jones, G.S., Copsey, D., Knight, J. R. and Tennant, W. J., “A new HadGEM3-A-based system for attribution of weather and climate-related extreme events”, Journal of Climate, vol. 26, (2013), pp. 2756–2783.

[12] Zu-Guo, Y., Yee, L., Yongqin, D. C., Qiang, Z., Vo, A. and Yu, Z., "Multifractal analyses of daily rainfall time series in Pearl River basin of China", Physica A: Statistical Mechanics and its Applications, vol. 405, (2014), pp. 193-202.
[13] MORAT, Ph., "Note sur l'application à Madagascar du quotient pluviomètre d'Emberger", Cah., ORSTOM, sér.bio1., vol. 10, (1969), pp. 117-132.
[14] Shaw, E. M., "Hydrology in Practice". Second Edition, Van Nostrand Reinhold (Intemational), London, United Kingdom, (1988).
[15] Cook, G. D. and Heerdegen, R. G., "Spatial variation in the duration of the rainy season in monsoonal Australia", International Journal of Climatology, vol.21, (2001), pp. 1723–1732.
[16] Moon, S. E., Ryoo, S. B. and Kwon, J. G., "A Markov chain model for daily precipitation occurrence in South Korea", International Journal of Climatology, vol. 14, (1994), pp. 1009–1016.
[17] Reiser, H. and Kutiel, H., "Rainfall uncertainty in the Mediterranean: definitions of the daily rainfall threshold (DRT) and the rainy season length (RSL)", Theoretical and Applied Climatolology, vol. 97, (2008), pp. 151-162.
[18] Reiser, H. and Kutiel, H., "Rainfall uncertainty in the Mediterranean: definition of the rainy season – a methodological approach", Theoretical and Applied Climatolology, vol. 94, (2008), pp. 35-49.
[19] Zoglat, A., El Adlouni, S., Badaoui, F., Amar, A. and Okou, "Managing hydrological risk with extreme modeling: application of peaks over threshold model to the Lokkous basin". Journal of Hydrologic Engineering, vol. 19, (2014).
[20] Naveau, P., Nogaj, M., Ammann, C., Yiou, P., Cooley, D. and Jomelli, V., "Statistical methods for the analysis of climate extremes", Comptes Rendus Geoscience, vol. 337, (2005), pp. 1013– 1022.
[21] Deidda, R., "A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series", Hydrology and Earth System Sciences, vol. 14, (2010), pp. 2559-2575.
[22] Jibrael, F. J., "Multiband Cross Dipole Antenna Based On the Triangular and Quadratic Fractal Koch Curve", International Journal of Engineering, Vol. 4, (1991).
[23] Anisheh, S.M. and Hassanpour, H., "adaptive segmentation with optimal window length scheme using fractal dimension and wavelet transform", International Journal of Engineering, Transactions B: Applications, Vol. 22, No.3 (2009).
[24] Tricot, C., "Curves and Fractal Dimension", Springer-Verlag, (1995).