MATHEMATICS IN ENGINEERING, SCIENCE AND AEROSPACE
MESA - www.journalmesa.com
Vol. 6, No. 4, pp. 607-614, 2015
© CSP - Cambridge, UK; I&S - Florida, USA, 2015

# Analysis and modeling of rainfall events

K. Ioannidis[1,2], A. Karagrigoriou[1⋆], D. F. Lekkas[1]

[1] Analysis and Simulation of Environmental Systems (ASES), University of the Aegean

[2] Tinbergen Institute, University of Amsterdam

⋆ *Corresponding Author* alex.karagrigoriou@aegean.gr

**Abstract.** The purpose of this study is the statistical analysis of rainfall events to explore patterns and dependencies that would allow the generalization in cases of missing or truncated data. More specifically, in this paper we estimate intensity–duration–frequency (IDF) curves, which are widely used to model rainfall. We use data from a meteorological station in Eresos, Greece and estimate the parameters of the model for fixed return periods. Sensitivity analysis is conducted to check whether the estimates are optimal. Finally, a more general model is applied that allows for simultaneous modeling of rainfall duration, intensity and frequency (via return periods).

## 1 Introduction

It is critical for hydrological design and water engineering to develop models that explain the pattern under which rainfall events occur and accurately predict the future inflow of rainfalls. More specifically, it is necessary to find relations that connect rainfall intensity and duration while simultaneously taking into account the frequency of occurrence. Surprisingly, a model proposed almost a century ago by Bernard in [1] is still one of the most widely used for rainfall modeling. A more general formula for the same purpose was proposed and thoroughly examined more recently by Koutsoyiannis et al. in [2].

The tools used for the modeling of such characteristics are curves commonly known as intensity–rainfall–frequency (IDF) curves. An Intensity-Duration-Frequency curve (IDF curve) is a graphical representation of the probability that a rainfall of given duration and intensity will occur. The parameters that make up the axes of the graph are:

- Rainfall duration (in hours)
- Rainfall intensity (in mm/hr)

2015 **Mathematics Subject Classification:** 62J02, 65C05, 65C60, 62P12
**Keywords**: rainfall modeling, non–linear regression, intensity-duration-frequency curves, sensitivity analysis, Monte Carlo simulation

- Rainfall frequency (via return period) (in years)

Studies based using this type of modeling have been conducted in various areas (see Stern and Coe [3]; Lee [4]; Rao and Kao [5]). We should note here that apart from this model, a more general one could be used connecting duration, intensity and frequency of rainfall events simultaneously.

In the present paper, we model the rainfall process using the previously mentioned models and data from Eresos (Lesvos, Greece) meteorological station covering a period between 17.11.2009 and 16.12.2012. In the following section, basic definitions are given that are necessary for understanding rainfall events and analyzing the proposed models. In Section 3 the methodology is applied to Eresos data and estimates (both point and interval estimation) for the parameters of the model are given. Finally, in Section 4 the validity of the estimates is studied via a sensitivity analysis to check and confirm the optimality of estimated values.

## 2 Basic Definitions and Fitting Models

### 2.1 Basic Definitions

Daily registries of rainfall activity are usually collected by a meteorological station and processed under the assumption that two rainfall events are considered independent if two non–empty registries are more than $k$ hours apart, for some value of $k$. For each event the following variables are recorded:

- $(R)$ Total rainfall amount (in mm)
- $(D)$ Rainfall event duration (in hrs)
- $(I)$ Rainfall event intensity (in mm/hr)

The above are defined as follows:

**Definition 2.1.** Total rainfall amount $(R)$ is defined as the sum of all registries of rainfall activity during a particular rainfall event and it is measured in mm.

**Definition 2.2.** Rainfall event duration $(D)$ is defined as the time distance between the first and the last non–empty registry within a particular rainfall event and is measured in hrs.

**Definition 2.3.** Rainfall event intensity $(I)$ is defined as the quotient of total rainfall amount over rainfall event duration $(I = R/D)$ and is measured in mm/hr.

### 2.2 Model Description and Estimation

Let $F(r)$ be the cumulative distribution function (cdf) of total rainfall amount $(R)$.

**Definition 2.4.** Return period $T_r$ is defined for every value of $r$ as

$$T_r = \frac{1}{1-F(r)} = \frac{1}{S(r)}, \tag{2.1}$$

where $S(r)$ is the survival function and it is measured in years.

Because by definition, $S(r)$ is defined as $S(r) = 1 - F(r) = \bar{F}(r)$, it represents the probability of total rainfall amount exceeding $r$ mm (in the unit of time scale), it is obvious that the inverse of this probability represents *time units of* $T_r$ until reoccurence of the required event (rainfall amount exceeding $r$ mms). Thus, the return period $T_r$ is an alternative way to express frequency. For every return period, an IDF curve is to be applied and estimated, describing the relation between duration and intensity. All IDF curves for a specific return period are special forms of the formula:

$$I = \frac{c}{(D^{\nu} + \theta)^{\eta}}, \tag{2.2}$$

where all coefficients are non–negative. The above formula is not the result of a theoretical approach, but an empirical formula, derived out of accumulated experience studying IDF curves. Simpler forms of the formula appear in the literature for $\nu = 1$, $\eta = 1$ & $\theta = 0$.

For values $\nu \neq 1$ and $\eta \neq 1$, the fraction $\frac{1}{(D^{\nu}+\theta)^{\eta}}$ is well approximated by $\frac{1}{(D+\theta^*)^{\eta^*}}$. Numerical studies (Koutsoyiannis et al. [2]) have demonstrated the validity of this approach, so that by setting $a = \theta^*$ and $b = \eta^*$, the above formula simplifies to:

$$I = \frac{c}{(D + a)^{b}}, \tag{2.3}$$

where $a, b$ and $c$ are the parameters of the model.

It is common parameter estimates for $a$ and $b$ to be similar for every return period, especially when data come from the same or neighboring stations. Furthermore, it is not hard to prove (Koutsoyiannis et al. [2]) that from the above three parameters, only $c$ depends on the return period $T_r$. In such cases, it is possible to apply a generalized model connecting simultaneously duration, intensity and frequency of rainfall events, defined as:

$$I = \frac{c(T_r)}{(D + a)^{b}}, \tag{2.4}$$

Using (2.1) we could choose as the function $c(T_r)$, the point $r$ of the cdf $F$ such that:

$$r = F^{-1}(1 - 1/T_r) \equiv c(T_r). \tag{2.5}$$

Because rainfall events with amounts above a threshold $r$ are connected with extreme events and the distribution of the maximum, typical distributions that could be used are, from the $1^{st}$ Extreme Value Theorem (Fisher-Tippett-Gnedenko Theorem, Fisher and Tippett [6]; Gnedenko [7]) Gumbel, Frechet and Weibull distributions. The choice of distribution defines, through 2.4, the final form of the IDF curve. Specifically for the Gumbel distribution, it should be noted that in hydrology it is used to analyze monthly or yearly maxima of rainfalls as well floods and droughts (Gumbel [8]; Burke et al. [9]). The convenience of this distribution lies in the fact that the scale and shape parameters $a$ and $b$, are easily estimated by the method of moments and provided in closed forms. Using the sample mean $\bar{X}$ and the sample deviation $S$, the moment estimators of the distribution parameters are given by:

$$\hat{\alpha} = \bar{X} - \frac{\gamma\sqrt{6}s}{\pi} \ \& \ \hat{\beta} = \frac{\sqrt{6}s}{\pi} \tag{2.6}$$

where $\gamma$ is Euler's constant.

Apart from this probabilistic approach, in the literature two additional formulas are quite popular (Berbard [1]; Raudkivi [10]; Singh [11]; Efstratiadis [12]):

$$I = \frac{c + K\ln(T_r^d)}{(D+a)^b}, \tag{2.7}$$

and

$$I = \frac{KT_r^d}{(D+a)^b}, \tag{2.8}$$

where $a$, $b$, $c$, and $d$ are the parameters of the model.

To estimate parameters of non–linear models, like the above, one can use the non–linear least squares method, which is a numerical method approximating the non–linear model with a linear one and using the Gauss–Newton algorithm (Seber & Wild [13]).

## 3 Analysis and Modeling of Eresos Data

Data used in this paper come from the meteorological station of Eresos in Lesvos island (Greece) and cover the period from 17.11.2009 to 16.12.2012 with small periods of no registries due to technical difficulties. The collected raw data were processed under the assumption that independence is ensured for $k = 2$ (in hours). In other words, if there are more than $k = 2$ hours between two non–empty registries (meaning there is no rainfall activity for at least 2 hours), then rainfall events are independent. In total, 234 rainfall events were created and for each one the three previously mentioned variables $R$, $D$ and $I$ were recorded.

Descriptive statistics of the data are presented in Table 1.

**Table 1** Descriptive statistics of rainfall events

| Statistic | Amount ($R$) | Duration ($D$) | Intensity ($I$) |
|---|---|---|---|
| Maximum | 79.8 | 12.50 | 61.8 |
| Minimum | 0.60 | 0.50 | 0.62 |
| Mean | 5.87 | 1.89 | 4.21 |
| SD | 9.31 | 2.41 | 7.61 |
| $1^{st}$ Quantile | 0.70 | 0.50 | 1.40 |
| $2^{nd}$ Quantile | 2.10 | 0.50 | 1.69 |
| $3^{rd}$ Quantile | 6.90 | 2.38 | 3.55 |

Figure 1 provides the graphical representation of the relation between rainfall duration and intensity, which seem to be inversely proportional.

We applied model (2.3) for 4 different return periods, that is for rainfalls that appear once every 2, 5, 10 and 20 years. The parameter estimates along with the corresponding 95% confidence intervals are presented in Table 2.

The resulting IDF curves in equation form are:

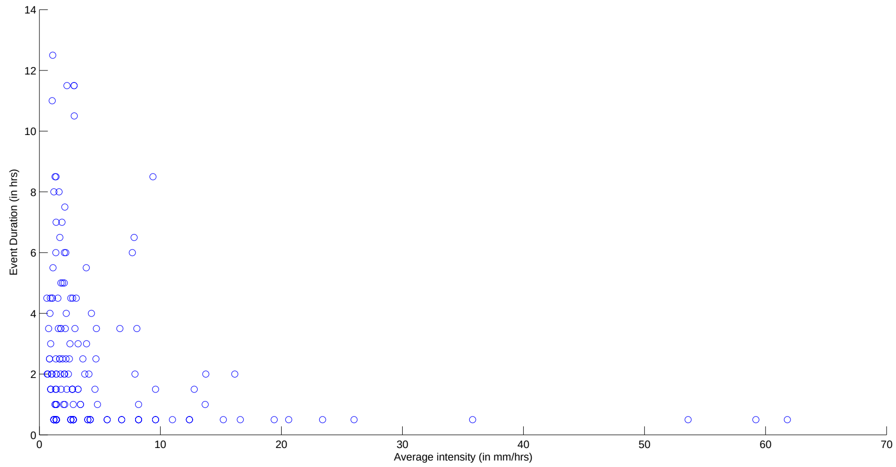


**Fig. 1** Scatter plot of rainfall duration–intensity

**Table 2** Parameter estimates with 95% confidence intervals

| $T_r$ | 2 years | 5 years | 10 years | 20 years |
|---|---|---|---|---|
| $a$ | -0.47 | -0.29 | -0.36 | -0.26 |
| 95% CI | (-0.70,-0.24) | (-0.91,0.33) | (-0.82,0.10) | (-0.67,0.16) |
| $b$ | 0.37 | 0.73 | 0.62 | 0.69 |
| 95% CI | (-0.31,1.06) | (-0.06,1.52) | (-0.04,1.29) | (0.27,1.11) |
| $c$ | 4.69 | 12.98 | 15.39 | 21.86 |
| 95% CI | (1.60,7.79) | (0.39,25.57) | (2.89,27.89) | (8.56,35.15) |

$$I = \frac{4.69}{(D-0.47)^{0.37}}, T_r = 2 \quad years$$
$$I = \frac{12.98}{(D-0.29)^{0.73}}, T_r = 5 \quad years$$
$$I = \frac{15.39}{(D-0.36)^{0.62}}, T_r = 10 \quad years$$
$$I = \frac{21.86}{(D-0.26)^{0.69}}, T_r = 20 \quad years$$

As mentioned in the previous section, an IDF curve is a graphical method with 3 axes, duration, intensity and return period. The above equations in a common plot provide the graph in Figure 2.

Since estimates for $a$ and $b$ were close, we proceed with the implementation of the general model given in (2.8) and the evaluation of new estimates presented in Table 3:

The previous IDF curve in equation form is:

$$I = \frac{5.893 T_r^{0.91}}{(D+0.9325)^{1.921}}$$

and can be graphically represented as shown in Figure 3.

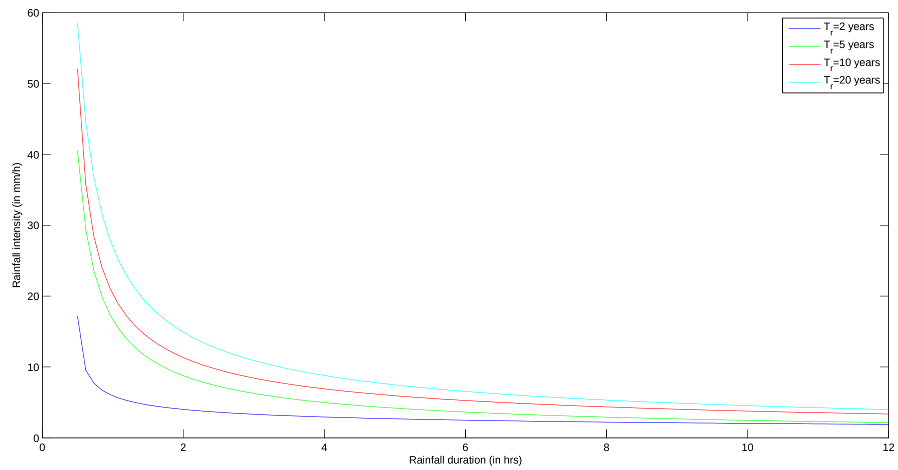


**Fig. 2** IDF curves for $T_r$ 2,5,10 and 20 years

**Table 3** Estimates and confidence intervals for the general model

| Parameter | Estimate | 95% CI |
|---|---|---|
| *a* | 0.9325 | (0.3229,1.542) |
| *b* | 1.921 | (1.449,2.393) |
| *K* | 5.893 | (0.1919,11.59) |
| *d* | 0.91 | (0.8813,0.9387) |

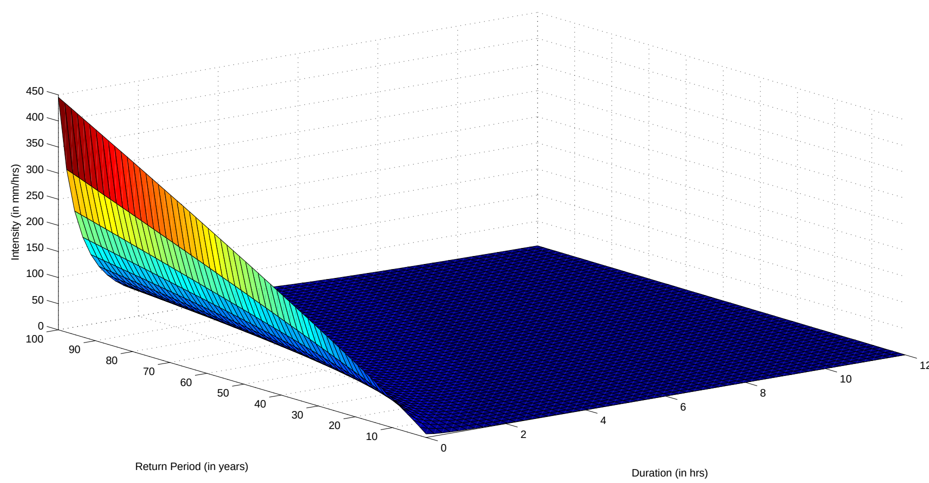


**Fig. 3** IDF curve with duration, intensity and return period

## 4 Sensitivity Analysis

Sensitivity analysis is a technique that explores the consequences on optimal (or estimated) solution of a model, as a result of changes in data or parameter values. Basically in every model, the dynamical environment of constant value changes in which the phenomenon to be studied occurs should be taken care of.

In non-linear models, there is a chance that estimation methods can result in sub-optimal estimates of the parameters. This can happen when the objective function we try to maximize (here $R^2$) gets stuck in a local optima or is not well defined for the selected model.

To check whether the set of parameters is the optimal (the global maximum has been reached), an analysis using a Monte Carlo type method was conducted to investigate parameter behavior. For each return period $T_r$, we create 10000 random samples $(a_i, b_i, c_i)$, where $a_i, b_i, c_i$ follow the *Uniform* distribution for different choices of intervals and each time we keep the parameter–vector with the largest $R^2$. The above procedure was repeated 100 times. Results, as shown in Figure 4, validate the optimality of the parameter estimates.

From Figure 4 it is clear that parameter plots are unimodal for every return period, with modes occurring at the estimate values obtained in the previous section, and as a consequence it is confirmed that they are optimal.

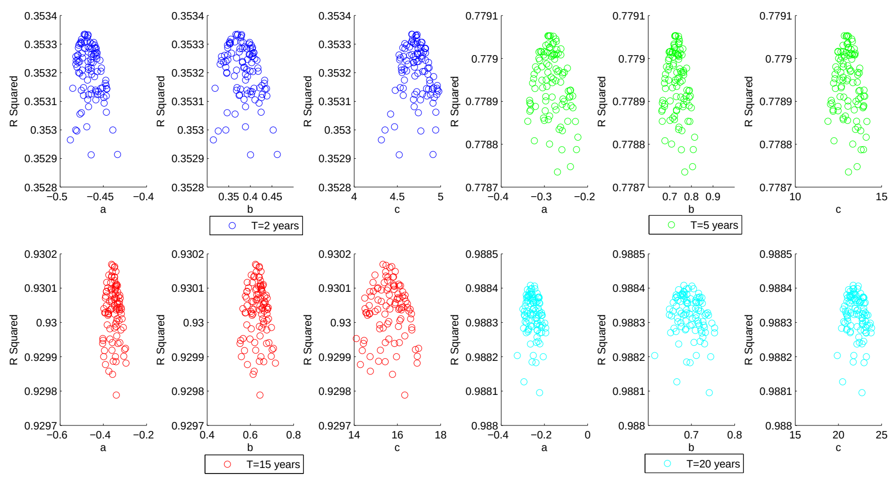


**Fig. 4** Scatter plots of parameters and $R^2$

## 5 Conclusions

We applied a model that has been extensively used in the literature to model rainfall, IDF curves. Similar studies have been conducted by Stern and Coe in [3], Lee in [4] and Rao and Kao in [5]. The parameters of the model were estimated and graphical representations of the IDF curves were created. Sensitivity analysis was conducted on the parameters to check for their optimality and the estimates were found optimal.

The present paper is the initial part of ongoing research that can be generalized in two directions. Spatially, as data from only one meteorological station were used (Eresos, Lesvos) and the model could be expanded using data from other meteorological stations in Lesvos or the broader North-Eastern Aegean area. Temporally, by comparing predicted future data with future inflows of actual data to validate the accuracy of the model to predict rainfall behavior.

## References

[1] M.M. Bernard. Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers*, 96(1):592–606, 1932.

[2] D. Koutsoyiannis, D. Kozonis and A. Manetas. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1):118–135, 1998.

[3] R.D. Stern and R. Coe. A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General)*, 147(1):1–34, 1984.

[4] C. Lee. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in southern Taiwan. *Crop Environ. Bioinf*, 2(1):31–38, 2005.

[5] A.R. Rao and S.C. Kao. Statistical analysis of Indiana rainfall data. In *Joint Transportation Research Program*, volume 1, pages 1-167, Indiana Department of Transportation, 2006.

[6] R.A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press, 1928.

[7] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

[8] E.J. Gumbel. The return period of flood flows. *Annals of Mathematical Statistics*, 12(2):163–190, 1941.

[9] S.J. Brown, E. J. Burke and R. H. J. Perry. An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology*, 388(1-2):131-143, 2010.

[10] A.J. Raudkivi. *Hydrology, An Advanced Introduction to Hydrological Processes and Modelling*. Pergamon Press, Oxford, UK, 1979.

[11] V.P. Singh. *Elementary Hydrology*. Prentice-Hall, New Jersey, 1992.

[12] A. Efstratiadis. Lecture notes on flood hydrology and design of sewage networks, 2011.

[13] G.A.F. Seber and C.J. Wild. *Nonlinear regression*. John Wiley & Sons, New Jersey, 2003.