A mathematical framework for studying rainfall intensity-duration-frequency relationships

Demetris Koutsoyiannis, Demosthenes Kozonis and Alexandros Manetas

Department of Water Resources, National Technical University, Heroon Polytechniou 5,

GR-157 80 Zografou, Athens, Greece

Abstract

A general formula for the rainfall intensity-duration-frequency (idf) relationship, consistent with the theoretical probabilistic foundation of the analysis of rainfall maxima is proposed. Specific forms of this formula are explicitly derived from the underlying probability distribution function of maximum intensities. Several appropriate distribution functions are studied for that purpose. Simple analytical approximations of the most common distribution functions are presented, which are incorporated in, and allow mathematically convenient expressions of idf relationships. Also, two methods for a reliable parameter estimation of idf relationships are proposed. The proposed formulation of idf relationships constitutes an efficient parameterisation, facilitating the description of the geographical variability and regionalisation of idf curves. Moreover it allows incorporating data from non-recording stations, thus remedying the problem of establishing idf curves in places with a sparse network of rain-recording stations, using data of the denser network of non-recording stations. Case studies, based on data of a significant part of Greece, briefly presented in the paper, clarify the methodology for the construction and regionalisation of the idf relationship.

Keywords Hydrologic statistics, Rainfall intensity, Flood design, Flood risk

1. Introduction

The rainfall intensity-duration-frequency (idf) relationship is one of the most commonly used tools in water resources engineering, either for planning, designing and operating of water resource projects, or the protection of various engineering projects (e.g., highways, etc.) against floods. The establishment of such relationships goes back to as early as 1932

(Bernard, 1932). Since then, many sets of relationships have been constructed for several parts of the globe. Since the 60s, the geographical distribution of intensity-duration-frequency relationships has been studied in several developed countries and maps have been constructed to provide the rainfall intensities or depths for various return periods and durations. For example, in USA such maps have been developed since 1961 by the U. S. Weather Bureau (Hershfield, 1961) and more recently by NOAA (Miller et al., 1973 for the western USA; Frederick et al., 1977, for the eastern and continental USA). These maps have been reproduced in many hydrological handbooks and textbooks (e.g., Chow, 1964, p. 9.51-9.56; Linsley et al., 1975, p. 358; Chow et al., 1988, pp. 446-451; Viessman et al., 1989, p. 337; Wanielista, 1990, p. 59; Smith, 1993). In the UK and Ireland, maps have been constructed by the Institute of Hydrology (NERC, 1975) and reproduced in various textbooks (e.g., Wilson, 1990, pp. 278-338). Similar maps have been constructed in other countries or parts of countries, e.g., Australia (Canterford et al., 1987); India (UNESCO, 1974; see also Subramanya, 1984, p. 40); Sri Lanka (Baghirathan and Shaw, 1978); SWA-Namibia (Pitman, 1980); region of Tuscany, Italy (Pagliara and Viti, 1993). In some cases such as for Nigeria (Oyebande, 1982) and Pennsylvania (more detailed analysis than the above referenced for USA; Aron et al., 1987) instead of constructing maps with contours, the regions of interest were divided into homogeneous subregions and one set of curves was devised for each subregion.

However, in most other countries such maps with rainfall intensity contours have not been constructed until now, and one has to retrieve the original intensity records of a nearby rain-recording station to construct the intensity-duration-frequency relationship, when needed. However, nowadays, due to the great usefulness of the map delineated information of rainfall idf curves, and the convenience provided by the expanded use of computerised databases (in storage and processing of hydrometeorological data) and geographical information systems (in regionalisation of information), it is anticipated that the development of maps will propagate to other less developed countries in the near future.

In the last decades, significant progress has been made in the statistical and stochastic modelling of hydrological time series. However, this progress is not reflected in the

procedures for formulating and constructing idf curves, which remain semi-empirical. This may be a direct consequence of the fact that construction of idf curves and related issues do not at all constitute a hot research topic; the established methodologies are adequate for the developed countries, that have gone to a conclusion for such issues decades ago. Nevertheless, the less developed countries, that are now proceeding to these issues can take advantage of the advances in statistical modelling of time series and develop new more refined methodologies.

Such methodologies must also take into account the specific problems of less developed countries, which are mainly related with data availability. A typical problem that is met in many countries is the very sparse network of rain-recording stations, whose data are the natural basis for idf calculations. As a solution to this problem, additional information from the denser network of non-recording stations can be utilised. To this aim, an appropriate methodology for incorporating data from non-recording stations must be developed.

This paper proposes a new approach to the formulation and construction of the idf curves using data from both recording and non-recording stations. More specifically, it discusses a general rigorous formula for the idf relationship whose specific forms are explicitly derived from the underlying probability distribution function of maximum intensities. Also, it proposes two methods for a reliable parameter estimation of the idf relationship. Finally, it discusses a framework for the regionalisation of idf relationships by also incorporating data from non-recording stations. The paper includes a brief presentation of an application of the developed methodology to a significant part of Greece.

The paper is organised in five sections, the first being this introduction. In section 2 we give the mathematical formulation of the idf relationship. Section 3 is devoted to parameter estimation issues and section 4 deals with the geographical variation of idf curves and regionalisation issues. In both sections 3 and 4 the proposed procedures are illustrated with applications using real-world data. Conclusions are drawn in section 5.

2. Mathematical formulation of the idf relationship

It is self-evident that the intensity-duration-frequency relationship is a mathematical relationship among the rainfall intensity *i*, the duration *d*, and the return period *T* (or, equivalently, the annual frequency of exceedance, typically referred to as "frequency" only). However, these terms may have different meaning in different contexts of engineering hydrology and this may lead to confusion or ambiguity. For the sake of a comprehensive presentation and inambiguousness in the material that follows we include in subsections 2.1 and 2.2 definitions, clarifications, and description of the general properties of the idf relationships. The reader familiar with these issues may proceed directly to subsection 2.3.

2.1 Definition of variables, notation and clarification

Let $\zeta(t)$ denote the instantaneous rainfall intensity process, where *t* denotes time. Let *d* be a selected (arbitrary) time duration (typically from a few minutes to several hours or few days), which serves as the length of a time window over which we integrate the instantaneous rainfall intensity process $\zeta(t)$. Moving this time window along time we form the moving average process, given by

$$\zeta_d(t) = \frac{1}{d} \int_{t-d}^{t} \zeta(s)\, ds \tag{1}$$

In reality, because we do not know the instantaneous intensity in continuous time, but rather have measurements of the average intensity $\zeta_\delta(t)$ for a given resolution δ (typically 5-10 min to 1 hour), (1) becomes

$$\zeta_d(t) = \frac{\delta}{d} \sum_{i=0}^{N-1} \zeta_\delta(t - i\,\delta) \tag{2}$$

where it is assumed that the duration *d* is an integer multiple of the resolution δ, i.e., $d = N\,\delta$. Given the stochastic process $\zeta_d(t)$ we can form the series of the maximum average intensities (or simply maximum intensities) $i_l(d)$ ($l = 1, \ldots, n$), which consists of *n* values, where *n* is the number of (hydrological) years through which we have available measurements of rainfall

intensities. This can be done in two ways. According to the first way, we form the series of annual maxima (or annual maximum series) by

$$i_l(d) := \max_{l^- < t < l^+} \{\zeta_d(t)\} \qquad (3)$$

where l^- and l^+ are the beginning and end time of the lth year. According to the second way, we form the series above threshold (also known as partial duration series or annual exceedance series) by selecting those values of $\zeta_d(t)$ that exceed a certain threshold φ, selected in a manner that the series $\{i_l(d)\}$ includes exactly n values. To ensure stochastic independence among $i_l(d)$, we also set a lower time limit τ (e.g., one or more days) for consecutive values, thus defining the series by

$$\{i_l(d),\ l = 1,\ \ldots,\ n\} := \left\{ \zeta_d(t_l) \mid \zeta_d(t_l) > \varphi, \quad t_l > t_{l-1} + \tau, \quad \zeta_d(t_l) = \max_{t_l - \tau < t < t_l + \tau} \{\zeta_d(t)\} \right\} \quad (4)$$

where the three conditions of the right-hand part must hold all together, otherwise the point t_l and the respective intensity $\zeta_d(t_l)$ are not selected for the series $\{i_l(d)\}$.

In practice, the construction of the series of maximum intensities is performed simultaneously for a number k of durations $d_j, j = 1,\ \ldots,\ k$, starting from a minimum duration equal to the time resolution δ of observations (e.g., from 5-10 min to 1 hour depending on the measuring device) and ending with a maximum duration of interest in engineering problems (typically 24 or 48 hours). Normally, all k series must have the same length n but, due to missing values, it is possible to have different lengths n_j for different durations d_j.

The above description of the construction of the maximum intensity allows us to observe that the duration d is not a random variable but, rather, a parameter for the intensity. It is not related to the actual duration of rainfall events, but is simply the length of the time window for averaging the process of intensity. On the contrary, the series of maximum intensities $i_l(d)$ is considered as a random sample of a random variable $I(d)$.

The return period T for a given duration d and maximum intensity $i(d)$ is the average time interval between exceedances of the value $i(d)$. It is well known (see, e.g., Kottegoda, 1980, p. 213) that for the annual series, under the assumption that consecutive values are

independent, the return period of an event is the reciprocal of the probability of exceedance of that event, i.e.,

$$T = \frac{1}{1 - F} \tag{5}$$

where F denotes the probability distribution function of $I(d)$ which of course is evaluated at the particular magnitude of interest. It is also known (see, e.g., Raudkivi, 1979, p. 411) that the return period T' for the series above threshold is related to that of the series of annual maxima by

$$T = \frac{1}{1 - \exp(-1 / T')} \Leftrightarrow T' = \frac{1}{-\ln(1 - 1 / T)} \tag{6}$$

(A good approximation of (6) with an accuracy of two decimal digits is given by the very simple relation $T = T' + 0.5$). Thus, the return period is always related to the distribution function of the series of annual maxima.

Given the above clarifications we observe that the problem of the construction of idf curves is somehow idiosyncratic. It is not a problem of statistical analysis of a single random variable, as it includes two variables i and d. Nor it is a problem of two random variables, because d is not a random variable. In fact, it consists of the study of a family of random variables $I(d)$, where d takes (theoretically an infinite number of) values from a real interval. This family of random variables $I(d)$ does not form a typical stochastic process as I does not represent an intensity at a certain time, nor d represents time but, rather, a time interval. Nevertheless, invoking the theory of stochastic processes is not necessary in this problem because we are not interested in a multidimensional (i.e., of order greater than 1) distribution of $I(d)$. We are rather interested in the first-order distribution of $I(d)$, i.e., the function $F(i; d) = P(I(d) < i)$, which is the target of the construction of the idf curves. Indeed, the function $F(i; d)$ can be directly transformed into a relationship among the quantities i, d, T.

2.2 The idf relationship for a specified return period

All typical idf relationships of the literature for a specific return period are special cases of the generalised formula

$$i = \frac{\omega}{(d^v + \theta)^\eta} \tag{7}$$

where ω, v, θ, and η are non-negative coefficients with $v\,\eta \le 1$. The latter inequality is easily derived from the demand that the rainfall depth $h = i\,d$ is an increasing function of d. Equation (7) is not obtained by any theoretical reasoning, but is an empirical formula, encapsulating the experience from several idf studies. In the bibliography, we find simplified versions of (7), which are derived by adopting one or two of the restrictions $v = 1$, $\eta = 1$, and $\theta = 0$.

It should be noted that considering $v \ne 1$ and $\eta \ne 1$ results in overparameterisation of (7). Indeed, the quantity $1\,/\,(d^v + \theta)$ can be adequately approximated by $1\,/\,(d + \theta')^{\eta^*}$ where θ' and η^* are coefficients depending on v and θ, which can be determined numerically in terms of minimisation of the root mean square error. Consequently, $1\,/\,(d^v + \theta)^\eta$ is approximated by $1\,/\,(d + \theta')^{\eta'}$, where $\eta' = \eta\,\eta^*$. A numerical investigation was done to show how adequate the approximation of $1\,/\,(d^v + \theta)$ by $1\,/\,(d + \theta')^{\eta^*}$ is. The duration d was restricted between the values $d_{\min} = 1/12$ hours (= 5 min) and $d_{\max} = 120$ hours, an interval much wider than the one typically used. The parameter θ varied between 0 and $\theta_{\max} = 12\,d_{\min}$ (= 1 hour), and the parameter v between 0 and 1. The root mean square standardised error (rmsse) of the approximation took a maximum value of 2.3% for $v = 0.55$ and $\theta = \theta_{\max}$; the corresponding maximum absolute standardised error (mase) was 4.3%. For the most frequent case that $\theta \le d_{\min}$, the corresponding errors are 0.7% (rmsse) and 1.3% (mase). These errors are much less than the typical estimation errors and the uncertainty due to the limited sizes of the typical samples available. In conclusion, the parameter v in the denominator of (7) can be neglected and the remaining two parameters suffice. Hence, hereafter we will assume that $v = 1$.

Initially, the coefficients ω, θ, and η can be considered as dependent on the return period. However, their functional dependence cannot be arbitrary, because the relationships (7) for any two return periods T_1 and $T_2 < T_1$ must not intersect. If $\{\omega_1, \theta_1, \eta_1\}$ and $\{\omega_2, \theta_2, \eta_2\}$ are the two parameter sets for T_1 and T_2, respectively, then it can be shown that there exist at least two sets of constraints leading to feasible (i.e., not intersecting) idf curves. These are

$$\theta_1 > 0, \quad \theta_2 > 0, \quad \frac{\varsigma_1}{\varsigma_2} \leq 1, \quad \frac{\grave{e}_1}{\grave{e}_2} \geq \frac{\varsigma_1}{\varsigma_2}, \quad \frac{\grave{u}_1}{\grave{u}_2} > \frac{\grave{e}_1^{\varsigma_1}}{\grave{e}_2^{\varsigma_2}} \tag{8}$$

and

$$\theta_1 \geq 0, \quad \theta_2 \geq 0, \quad \varsigma_1 = \varsigma_2 = n, \quad \frac{\grave{u}_1}{\grave{u}_2} > 1, \quad \frac{\grave{e}_1^{\varsigma_1}}{\grave{u}_1} \leq \frac{\grave{e}_2^{\varsigma_2}}{\grave{u}_2} \tag{9}$$

To both these sets, the following obvious inequalities are additional constraints

$$\omega_1 > 0, \quad \omega_2 > 0, \quad 0 < \eta_1 < 1, \quad 0 < \eta_2 < 1 \tag{10}$$

The essential difference between the sets of constraints (8) and (9) is that the former does not allow θ to take zero value, while the latter does allow this special value. Furthermore, it can be shown that, if θ is allowed to take zero value, then the exponent η in (7) must be constant and independent of the return period. Because the case $\theta = 0$ must not be excluded, it is reasonable to adopt the set of constraints (9) for the subsequent analysis. For convenience, it is reasonable to consider θ independent of the return period, as well, thus leading to the following final set of restrictions

$$\theta_1 = \theta_2 = \theta \geq 0, \quad 0 < \eta_1 = \eta_2 = n < 1, \quad \omega_1 > \omega_2 > 0 \tag{11}$$

In this final set of restrictions, the only parameter that is considered as an (increasing) function of the return period T is ω. This leads, indeed, in a strong simplification of the problem of construction of idf curves. This theoretical discussion is empirically verified, as numerous studies have shown that real world families of idf curves can be well described with constant parameters θ and η.

2.3 The general idf relationship

After the above discussion we can formulate a generalised idf relationship in the form

$$i = \frac{a(T)}{b(d)} \tag{12}$$

which has the advantage of a separable functional dependence of i on T and d. The function $b(d)$ is

$$b(d) = (d + \theta)^\eta \tag{13}$$

where θ and η are parameters to be estimated ($\theta > 0$, $0 < \eta < 1$). The function $a(T)$ (which coincides with ω of the previous subsection) is given in the bibliography (e.g., Raudkivi, 1979, p. 85; Shaw, 1983, p. 236; Subramanya, 1984, p. 205; Chow et al., 1988, p. 459; Wanielista, 1990, p. 61; Singh, 1992, p. 904) by the following alternative relations

$$a(T) = \lambda \, T^\kappa \tag{14}$$

$$a(T) = c + \lambda \ln T \tag{15}$$

The first is the oldest (Bernard, 1932) yet the most common until recently (see, e.g. Kothyari and Garde, 1992; Pagliara and Viti, 1993). These relations are rather empirical and their use has been dictated by their simplicity and computational convenience rather than their theoretical consistency with the probability distribution functions which are appropriate for the maximum rainfall intensity. Chen (1983) applied a more theoretical analysis to obtain similar relationships. Koutsoyiannis (1994) reported that (14) is inappropriate for certain issues such as simulation (e.g., it tends to underestimate the variance). Koutsoyiannis (1996, p. 265) has demonstrated empirically that if the maximum rainfall intensity has Gumbel distribution then the parameters κ and λ of (14) are not in fact constant but they depend on the return period T (this is also demonstrated briefly below).

In fact, there is no need to introduce $a(T)$ as an empirical function, as it can be completely determined, in a theoretically consistent manner, from the probability distribution function of the maximum rainfall intensity $I(d)$. Indeed, if the intensity $I(d)$ of a certain duration d has a particular distribution $F_{I(d)}(i; d)$, this will also be the distribution of the variable $Y := I(d) \, b(d)$, which is no more than the intensity rescaled by $b(d)$ (with the parameters of the latter distribution being properly rescaled). This has been also reported by Koutsoyiannis (1994, Appendix A) for the Gumbel distribution, but it can be generalised for any distribution. Mathematically, this is expressed by

$$P\{I(d) \le i\} = P\{I(d) \, b(d) \le i \, b(d)\} = P\{Y \le y\} \tag{16}$$

where $P\{\ \}$ denotes probability, or

$$F_{I(d)}(i; d) = F_Y(y_T) = 1 - \frac{1}{T} \tag{17}$$

Hence, if y_T is the $(1 - 1/T)$-quantile of the distribution function F_Y, then

$$y_T \equiv a(T) = F_Y^{-1}(1 - 1/T) \tag{18}$$

which proves our claim that $a(T)$ is completely determined from the distribution function of intensity.

We point out that the inverse of a distribution function $F_Y^{-1}(\ )$ appearing in (18) generally does not have as simple expression as those of the empirical functions (14) and (15), and in some cases $F_Y^{-1}(\ )$ cannot be expressed with an explicit analytical equation. However, as we show below, we can always get approximate analytical expressions adequately simple and more accurate than the empirical functions (14) and (15).

In the next subsection, we examine the most typical distribution functions of maximum intensities and obtain for each distribution function the corresponding function $a(T)$. Notably, we show that the empirical functions (14) and (15) can be obtained by our general methodology, but they correspond to distribution functions that may not be appropriate for maximum rainfall intensities.

2.4 Alternative distribution functions

To better serve our purpose, the mathematical expressions of the alternative distribution functions $F_Y(y)$ given below may have been written intentionally in a slightly different form from that typically used in the literature. In all distributions, κ and ψ denote dimensionless parameters whereas λ and c denote parameters having same dimensions as the random variable y (or $\ln y$ in case of logarithmic transformation of the variable).

a. Gumbel distribution function

The type I distribution of maxima, also termed the Gumbel distribution function (Gumbel, 1958), is the most widely used distribution for idf analysis due to its suitability for

modelling maxima. Given that the rainfall intensity $I(d)$ has Gumbel distribution for any duration d, so will have Y and thus

$$F_Y(y) = \exp(-e^{-y/\lambda + \psi}) \tag{19}$$

where λ and ψ are the scale and location parameters, respectively, of the distribution function. Combining (18) and (19) we directly get

$$y_T \equiv a(T) = \lambda \left\{ \psi - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\} \tag{20}$$

which is an exact yet simple expression of $a(T)$.

b. Generalised extreme value (GEV) distribution

This general distribution, which incorporates type I, II, and III extreme value distributions of maxima can be written in the form

$$F_Y(y) = \exp \left\{ -\left[1 + \kappa \left(\frac{y}{\lambda} - \psi \right) \right]^{-1/\kappa} \right\} \qquad y \geq \lambda \, (\psi - 1 / \kappa) \tag{21}$$

where $\kappa > 0$, $\lambda > 0$, and ψ are shape, scale, and location parameters, respectively. For $\kappa = 0$ the GEV distribution turns into the Gumbel distribution; the case where $\kappa < 0$ is not considered here because it implies an upper bound of the variable, which is not the case in maximum rainfall intensity. We directly obtain from (21) that

$$y_T \equiv a(T) = \lambda \left\{ \psi + \frac{\left[-\ln \left(1 - \frac{1}{T} \right) \right]^{-\kappa} - 1}{\kappa} \right\} = \lambda' \left\{ \psi' + \left[-\ln \left(1 - \frac{1}{T} \right) \right]^{-\kappa} \right\} \tag{22}$$

where for simplification we have set $\lambda' = \lambda / \kappa$ and $\psi' = \kappa \psi - 1$. Again we have an exact expression of $a(T)$ for the GEV distribution that remains relatively simple.

c. Gamma distribution

The two parameter gamma distribution function, sometimes used for idf analysis, is given by

$$F_Y(y) = \int_0^y \frac{1}{\lambda^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\lambda} dx, \quad y \geq 0 \tag{23}$$

where κ and λ are the shape and scale parameters of the distribution, respectively. Due to the complicated form of (23) it is not possible to get an exact explicit relationship of $a(T)$ for this specific distribution. Approximations such as the Wilson-Hilferty and the modified Wilson-Hilferty (Kirby, 1972) do not help in this case because they transform the gamma variate into a normal variate, which is still complicated and inappropriate to yield $a(T)$. Another approximation, proposed by Koutsoyiannis (1996, pp. 171-173) is more appropriate, as it leads directly to the relatively simple formula

$$y_T \equiv a(T) \approx \frac{\lambda\mu}{\alpha}\left(1-\frac{1}{T}\right)^{\alpha} + \frac{\lambda\nu}{\beta}\left[\xi-\left(\frac{1}{T}\right)^{\beta}\right], \quad \kappa \neq 1 \tag{24}$$

(In case that $\kappa = 1$, the Gamma distribution turns into the simpler exponential distribution which is studied below.) In the above formula, μ, ν, α, β and ξ are coefficients dependent on the shape parameter κ (i.e., not independent parameters), given by the following equations

$$\mu = 0.6\left(\sqrt{\kappa}-1\right) - \left(1/\sqrt{\kappa}-1\right) \tag{25}$$

$$\nu = 0.6\left(\sqrt{\kappa}-1\right) + 0.01(\kappa-1) + 1 \tag{26}$$

$$\alpha = 0.6/\sqrt{\kappa} + 0.08 \tag{27}$$

$$\beta = 0.0234\ln\kappa \tag{28}$$

$$\xi = \begin{cases} 1 & \kappa < 1 \\ 31\mathrm{e}^{-11.6(\kappa-1)^{-0.25}} & \kappa > 1 \end{cases} \tag{29}$$

Equation (24) has resulted from the approximation of the derivative of the inverse gamma distribution $y(u) = F_Y^{-1}(u)$ by

$$y'(u) = \frac{dy}{du} \approx \ddot{e}\,\grave{\imath}\,u^{\hat{a}-1} + \ddot{e}\,\acute{\imath}\,(1-u)^{\hat{a}-1} \tag{30}$$

and a systematic numerical investigation was performed to establish (25)-(29). For $0.2 \leq \kappa \leq 100$ (or, equivalently, $0.2 \leq C_s \leq 4.5$, where C_s is the coefficient of skewness) and $1.0001 \leq T \leq 10\,000$, the approximation error of (24), defined as $e = |y_T - \hat{y}_T|/\sigma_Y$, does not exceed the value 0.11 in any point of the space defined by these inequalities. This error is smaller than

that of Wilson-Hilferty approximation, both the original and modified, for $\kappa \leq 4$ $(C_s \geq 1)$ (see Figure 1).

d. Log Pearson III distribution

A very common distribution for idf analysis is the Log Pearson III distribution, which is a logarithmic transformation of the gamma distribution, given by

$$F_Y(y) = \int_{e^c}^{y} \frac{1}{x\lambda^\kappa \Gamma(\kappa)} (\ln x - c)^{\kappa-1} e^{-(\ln x - c)/\lambda} dx, \quad y \geq e^c \tag{31}$$

where c is a scale parameter, and κ and λ are shape parameters. Making use of the approximation (24) of the gamma distribution function we get

$$y_T \equiv a(T) \approx \exp\left\{ c + \frac{\lambda\mu}{\alpha}\left(1-\frac{1}{T}\right)^\alpha + \frac{\lambda v}{\beta}\left[\xi - \left(\frac{1}{T}\right)^\beta\right] \right\} \quad \kappa \neq 1 \tag{32}$$

(In case that $\kappa = 1$, the Log Pearson III distribution turns into the simpler Pareto distribution which is studied below.) As in (24), μ, v, α, β and ξ are coefficients dependent on the shape parameter κ, given by equations (25)-(29).

e. Lognormal distribution

The two parameter lognormal distribution has been used sometimes for idf analysis. It is a logarithmic transformation of the normal distribution, given by

$$F_Y(y) = \int_{0}^{y} \frac{1}{x\sqrt{2\pi}\sigma_Z} e^{-\frac{1}{2}\left(\frac{\ln x - \mu_Z}{\sigma_Z}\right)} dx, \quad y \geq 0 \tag{33}$$

where μ_Z and σ_Z are scale and shape parameters, respectively. Observing that the lognormal distribution is the limit of the log Pearson III distribution as $\kappa \to \infty$, we can use an approximation similar with (32). This is given by

$$y_T \equiv a(T) \approx \exp\left\{ \mu_Z + v\sigma_Z\left(1-\frac{1}{T}\right)^\alpha - v\sigma_Z\left(\frac{1}{T}\right)^\alpha \right\} \tag{34}$$

(Koutsoyiannis, 1996, p. 168), where $v = 5.53$ and $\alpha = 0.12$. For $1.0001 \leq T \leq 10\,000$, the approximation error of (34), defined as the absolute value of the difference of standardised normal variates, does not exceed the value 0.03. The same approximate equation with coefficients $v = 1/0.1975$ and $\alpha = 0.135$, is obtained from a formula by Stedinger et al. (1993, p. 18.11).

f. Exponential distribution

The two parameter exponential distribution function is given by

$$F_Y(y) = 1 - e^{-y/\lambda + \psi} \qquad\qquad y \geq \lambda\,\psi \qquad\qquad (35)$$

where λ and ψ are scale and location parameters, respectively. We directly obtain from (35) that

$$y_T \equiv a(T) = \lambda\,(\psi + \ln T) \qquad\qquad (36)$$

which, notably, is functionally identical to the empirical function (15).

Although the exponential distribution is not so common for idf analysis, equation (36) can be somehow connected to the Gumbel distribution in two ways. First, it can be an adequately accurate approximation of (20) for large return periods, e.g., $T \geq 50$. In that case we can write $\ln\,[1 - (1/T)] = -(1/T) - (1/T)^2 - \cdots \approx -(1/T)$, and then (20) toggles into (36). However, in that case the estimation of parameters ψ and λ should be based on the appropriate estimators of the Gumbel distribution (e.g., those of the methods of maximum likelihood, moments, L-moments, or the Gumbel's fitting method) rather than those of the exponential distribution. Second, if we analyse an annual data series using the Gumbel distribution and we want an estimate of intensity versus the return period T' for the series over threshold, then combining (20) and (6) we find the logarithmic expression (36) again, in the form

$$y_{T'} \equiv a'(T') = \lambda\,(\psi + \ln T') \qquad\qquad (37)$$

Again the parameters ψ and λ should be estimated by the appropriate estimators of the Gumbel distribution using the statistics of the annual series.

g. Pareto distribution

The generalised, three parameter, Pareto distribution function is

$$F_Y(y) = 1 - \left[1 + \kappa \left(\frac{y}{\lambda} - \psi \right) \right]^{-1/\kappa} \qquad y \geq \lambda \, \psi \qquad (38)$$

where $\kappa > 0$, $\lambda > 0$, and ψ are shape, scale, and location parameters, respectively (for $\kappa = 0$ the GEV distribution turns into the exponential distribution; the case $\kappa < 0$ is not considered because it implies an upper bounded variable). We directly obtain from (38) that

$$y_T \equiv a(T) = \lambda \left(\psi + \frac{T^\kappa - 1}{\kappa} \right) = \lambda' \, (\psi' + T^\kappa) \qquad (39)$$

where for simplification we have set $\lambda' = \lambda / \kappa$ and $\psi' = \kappa \psi - 1$. Although the Pareto distribution is not so common for idf analysis, it is connected to the GEV distribution in the way exponential distribution is connected to the Gumbel distribution. That is, (22) is very well approximated by (39) when $T \geq 50$; furthermore combining (22) and (6) we find the power expression (39) for the return period T' for the series above threshold, i.e.,

$$y_{T'} \equiv a'(T') = \lambda \left(\psi + \frac{T'^\kappa - 1}{\kappa} \right) = \lambda' \, (\psi' + T'^\kappa) \qquad (40)$$

In both those cases the parameters ψ, κ, and λ should be estimated by the appropriate estimators of the GEV distribution using the statistics of the annual series, rather than the estimators of the Pareto distribution.

Notably, in case that $\psi' = 0$ (or $\kappa \psi = 1$) equation (39) becomes identical to the empirical expression (14). Although one cannot exclude the contingency that the Pareto distribution is appropriate for idf analysis of a specific data set with its location parameter being $\psi = 1 / \kappa$ (so that $\psi' = 0$), the literature does not provide such evidence. Thus, the widespread use of (14) is not justified theoretically.

3. Parameter estimation methods

The parameters of the general idf relationship (12) fall into two categories: those of the function $a(T)$ (i.e., κ, λ, ψ, etc., depending on the distribution function adopted) and those of

the function $b(d)$ (i.e., η and θ). In this section we discuss some procedures for the estimation of parameters of both categories. We start discussing the typical procedure of the literature (subsection 3.1) and then we propose two other methods of parameter estimation (subsections 3.2 and 3.3). Finally, we give a real world application in which we test and compare the different methods (subsection 3.4). In all procedures we assume that we are given k groups each holding the historical intensities of a particular duration d_j, $j = 1, \ldots, k$. We denote by n_j the length of the group j, and by i_{jl} the intensity values of this group (samples of the random variables $I_j := I(d_j)$) with $l = 1, \ldots, n_j$ denoting the rank of the value i_{jl} in the group j arranged in descending order.

3.1 Typical procedure

The typical parameter estimation procedure for idf curves (Raudkivi, 1979, p. 85; Chow et al., 1988, p. 458; Wanielista, 1990, p. 61; Singh, 1992, p. 904) consists of three steps. The first step consists of fitting a probability distribution function to each group comprised of the data values for a specific duration d_j. In the second step the rainfall intensities for each d_j and a set of selected return periods (e.g., 5, 10, 20, 50, 100 years, etc.) are calculated. This is done by using the probability distribution functions of the first step. In the third step the final idf curves are obtained in two different ways: either (a) for each selected return period the intensities of the second step are treated and a relationship of i as a function of d (i.e., $i = i_T(d)$) is established by (bivariate) least squares, or (b) the intensities of the second step for all selected return periods are treated simultaneously and a relationship of i as a function of both d and T (i.e., $i = i(T, d)$) is established by (three-variate) least squares. In case (a) different values of the parameters ω, θ and η are obtained for each T. In case (b) unique values of the parameters θ and η are obtained while ω is determined as a function $\omega = a(T)$. The form of this function (typically (14) or (15)) is selected a priori. In case $a(T)$ is given by the power relationship (14), the estimation procedure is simplified, because (12) becomes linear by taking logarithms of both sides.

The main advantage of this parameter estimation procedure is its computational simplicity, which in fact imposes the separation of the calculations in three steps, so that the

calculations of each step are as simple as possible. However, the procedure has some flaws, which are not unavoidable. First, it bears the weakness of using an empirically established function $a(T)$ (step 3) instead of the one consistent with the probability distribution function (step 1). This has been already discussed in the previous section. Second, it is subjective, in the sense that the final parameters depend on the selected return periods in step 2. This dependence may be essential if the selected empirical function $a(T)$ departs significantly from that implied by the probability distribution function (Koutsoyiannis, 1996, p. 265). Third, it treats the three involved variables (i, d, T) as having the same nature, in spite of the fact that they are fundamentally different in nature, i.e., i represents a random variable, d is a (non-random) parameter of this random variable, and T is a transformation of the probability distribution function of the random variable.

In the following two subsections we propose two different parameter estimation methods that are free of the flaws of the above described typical procedure and harmonise with the general formulation of idf curves given in the previous section. These procedures need more complicated calculations than the typical procedure, yet remaining computationally simple. Both can be applied using a typical spreadsheet package and do not require the development of specialised computer programs.

3.2 Robust estimation

The first proposed method estimates the parameters in two steps, the first concerning the parameters of function $b(d)$ and the second those of $a(T)$. This method is based on the identity of the distribution functions of the variables $Y_j = I_j\, b(d_j)$ of all k groups, regardless of the duration d_j of each separate group. This identity leads us to the Kruskal-Wallis statistic, which is used to test whether several sample groups belong to the same population.

Let us assume that the parameters η and θ of $b(d)$ are known. Then we can find all values $y_{jl} = i_{jl}\, b(d_j)$. The overall number of data values is

$$m = \sum_{j=1}^{k} n_j \tag{41}$$

We assign ranks r_{jl} to all of the m data values y_{jl} (using average ranks in the event of ties). For each group we compute the average rank $\bar{r}_j$ of the n_j values of that group. If all groups have identical distribution then each $\bar{r}_j$ must be very close to $(m+1)/2$. This leads to the following statistic (Kruskal-Wallis) which combines the results of all groups

$$k_{\mathrm{KW}} = \frac{12}{m\,(m+1)} \sum_{j=1}^{k} n_j \left(\bar{r}_j - \frac{m+1}{2}\right)^2 \tag{42}$$

The smaller the value of k_{KW}, the greater the evidence that all groups of y values belong to the same population. Obviously, the ranks r_{jl} (and hence k_{KW}) depend on the parameters η and θ that were assumed as known. Consequently, the estimation problem is reduced to an optimisation problem defined as

$$\text{minimise } k_{\mathrm{KW}} = f_1(\eta, \theta) \tag{43}$$

Apparently, it is not possible to establish an analytical optimisation method for our case. A numerical search technique for optimisation that makes no use of derivatives (see Pierre, 1986, p. 264; Press et al., 1992, p. 394), is appropriate. However, it may be simpler to use a trial-and-error method based on a common spreadsheet computer program.

The advantages of the Kruskal-Wallis statistic are its non-parametric character and its robustness, i.e., its ability not to be affected by the presence of extreme values in the samples. We clarify, however, that the minimum value of k_{KW} determined by the minimisation process cannot be used further to perform the typical Kruskal-Wallis statistical test (actually, the testing is not really needed). The reason is that this test assumes that all k groups are mutually independent. In our case, the intensities I_j of the different groups are stochastically dependent variables, as is evident from their construction (see subsection 2.1). Thus, we do not know the distribution function of the statistic k_{KW} to perform any statistical test. Nevertheless, the minimisation of its value is achievable because the distribution function does not need to be known.

For the sake of improving the fitting of $b(d)$ in the region of higher intensities (and also to simplify the calculations) it may be preferable to use in this first step of calculations a part

of the data values of each group instead of the complete series. For example, we can use the highest 1/2 or 1/3 of intensity values for each duration.

Given the values of η and θ, we proceed to the second step of calculations, which is very easy. Assuming that, with these values, all groups have identical distribution, we append all k groups of values y_{jl} thus forming a unique (compound) sample. For this sample we choose an appropriate distribution function, such as those described in subsection 2.4, and estimate its parameters using the appropriate for that distribution estimators (e.g., those obtained by the methods of maximum likelihood, moments, L-moments, etc.; for a concise presentation of such estimators see Stedinger et al., 1993). This defines completely the form and the parameters of $a(T)$.

3.3 One-step least squares method

The second method estimates all parameters of both functions $a(T)$ and $b(d)$ in one step, minimising the total square error of the fitted idf relationship to the data. To this aim, to each data value i_{jl} we assign an empirical return period using, e.g., the Gringorten formula,

$$T_{jl} = \frac{n_j + 0.12}{l - 0.44} \tag{44}$$

So, for each data value we have a triplet of numbers (i_{lj}, T_{lj}, d_j). On the other hand, given a specific form of $a(T)$, chosen among those of subsection 2.4 from preliminary investigations of the type of the distribution function of intensity, we obtain the modelled intensity

$$\hat{i}_{jl} = \frac{a(T_{jl})}{b(d_j)} \tag{45}$$

and the corresponding error

$$e_{jl} = \ln i_{jl} - \ln \hat{i}_{jl} = \ln\left(\frac{i_{jl}}{\hat{i}_{jl}}\right) \tag{46}$$

where we have applied the logarithmic transformation to keep balance among the errors of the intensities of greater durations (which are lower) and those of lower ones. The overall mean square error is

$$e^2 = \frac{1}{k} \sum_{j=1}^{k} \frac{1}{n_j} \sum_{l=1}^{n_j} e_{jl}^2 \tag{47}$$

Again the estimation problem is reduced into an optimisation problem, defined as

$$\text{minimise} \quad e = f_2(\eta, \theta, \kappa, \lambda, \psi, \ldots) \tag{48}$$

A numerical search technique for optimisation that makes no use of derivatives, such as the Powell method (see Pierre, 1986, p. 277; Press et al., 1992, p. 412), is appropriate for this problem. However, it may be simpler to perform the optimisation using the embedded solver tools of common spreadsheet packages.

We note that the least squares method in fitting a theoretical to an empirical distribution function is not a novelty of the proposed method. Rather, the innovative element of the proposed method is the simultaneous estimation of the parameters of both the distribution function and the duration function $b(d)$.

3.4 An application

To illustrate the above described methodology we present a real-world application. A thirty year (1957-58 to 1986-87) data record of the Helliniko recording station (located at the Helliniko airport, Athens) was used. The selected durations d_j range from 5 min to 24 hours, as shown in Table 1. Due to missing values, the sample size for some durations is lower than 30. In Table 1 are also shown some summary statistics of the data.

Preliminary investigation showed that the Gumbel distribution is suitable for all groups with durations d_j. Thus, we adopted the idf relationship

$$i = \frac{a(T)}{b(d)} = \lambda \, \frac{\psi - \ln\left[-\ln\left(1 - \frac{1}{T} \right) \right]}{(d + \theta)^\eta} \tag{49}$$

An interpretation of this equation is that the variable $I(d)$ has a Gumbel distribution with its dimensionless parameter ψ constant and independent of duration d, and its scale parameter varying with d as $1/(d + \theta)^\eta$. This is (approximately) verified by fitting a Gumbel distribution

independently to each group of duration d_j. (These independently fitted Gumbel distributions are shown in Figure 2 with dotted lines.)

The application of the robust estimation method was done in two steps, as described in subsection 3.2. In the first step (estimation of η and θ) we used the highest 1/3 intensities of each group (i.e. 10 data values for durations 5 min - 12 hours and 7 data values for duration 24 hours). The minimisation of the Kruskal-Wallis statistic k_{KW} (equation (42)) was easily performed by the MS-EXCEL spreadsheet (by a trial-and-error procedure) and resulted in a minimum value of $k_{KW} = 3.33$ and parameter values $\eta = 0.796$ and $\theta = 0.189$. (The corresponding values when we use all data values of each group are $\eta = 0.776$ and $\theta = 0.139$.)

For the second step, i.e., the estimation of the distribution function parameters λ and ψ, we adopted the more robust method of L-moments, which unlike the other methods does not overemphasise an occasional extreme event, as it does not involve squaring or cubing of the data. This method results in the following estimators (Stedinger et al., 1993, p. 18.17):

$$\lambda = \hat{\lambda}_2 / \ln 2 \qquad \psi = \overline{y} / \lambda - 0.577 \tag{50}$$

Here $\hat{\lambda}_2$ is the estimate of the second L-moment given by

$$\hat{\lambda}_2 = 2 \sum_{l=1}^{n} \frac{(n-l)\, y_l}{n(n-1)} - \overline{y} \tag{51}$$

where the sample of the observations y_l is arranged in decreasing order, so that l is the rank of y_l. Other parameter estimation methods, such as those reviewed (among others) by Kite (1988, p. 96) and Koutsoyiannis (1996), could be used here instead of the L-moments method.

Transforming all intensities i into y values and unifying all groups (228 values in total) we find that $\overline{y} = 25.701$ and $\hat{\lambda}_2 = 5.761$. With these values, (50) results in $\psi = 2.515$ and $\lambda = 8.31$ (the estimates of the method of moments for $s_Y = 10.208$ are $\psi = 2.652$ and $\lambda = 7.96$). Thus, the idf relationship is

$$i = 8.31 \; \frac{2.515 - \ln\left[-\ln\left(1 - \frac{1}{T}\right)\right]}{(d + 0.189)^{0.796}} \qquad (d \text{ in h, } i \text{ in mm/h}) \tag{52}$$

The one-step least squares method, aiming at the minimisation of the total error e (equation (47)) was also easily performed using the MS-EXCEL spreadsheet. More specifically, the embedded Solver utility of this spreadsheet performed the optimisation directly (writing no code at all) and resulted in a minimum value of $e = 0.078$ and parameter values $\eta = 0.778$, $\theta = 0.143$, $\psi = 2.615$ and $\lambda = 7.59$. The idf relationship is slightly different, i.e.,

$$i = 7.59 \, \frac{2.615 - \ln\left[-\ln\left(1 - \frac{1}{T}\right)\right]}{(d + 0.143)^{0.778}} \quad (d \text{ in h}, i \text{ in mm/h}) \tag{53}$$

Graphical comparisons of (52) and (53) are shown in Figure 2 and Figure 3. Figure 2 is a plot of the idf curves (52) and (53) in the form of probability distribution functions, i.e., i versus the Gumbel reduced variate $k = -\ln[-\ln(1 - 1/T)]$. Each curve corresponds to a particular duration d_j. Apart from the curves resulting from equations (52) (continuous lines) and (53) (dashed lines) we have also plotted the empirical distribution functions of the samples using the Gringorten plotting position (points), and the Gumbel distributions for the intensities of each duration d_j fitted independently of the other durations (dotted lines). We observe that all three sets of curves of (52), (53), and the independently fitted Gumbel distributions are in good agreement with each other (in most cases indistinguishable from each other). They also agree with the empirical distribution functions.

Figure 3 is a logarithmic plot of idf curves (52) and (53) in the form of i versus d for a wide range of return periods, $T = 5, 50, 500$ and 5000 years. Apart from the curves resulting from (52) (continuous lines) and (53) (dashed lines), we have also plotted (points) the intensities obtained directly from the independently fitted Gumbel distribution of each duration. Both sets of curves are practically indistinguishable from each other and from the corresponding series of points. For comparison, a third set of curves (dotted lines), obtained using the empirical equation (14) are also shown. The idf relationship in this case was obtained by the typical procedure of the literature described in subsection 3.1 using returns periods 2, 5, 10, 20 and 50 years (second step of the typical procedure), and is given by

$$i = \frac{20.978 \, T^{0.237}}{(d + 0.167)^{0.784}} \quad (d \text{ in h}, i \text{ in mm/h}) \tag{54}$$

We observe that for return periods $T = 5$ and 50 years (which belong to the interval used for the fitting of (54)) these curves are indistinguishable from the other two sets. However, for $T = 500$ and 5000 the curves obtained by (54) (marked as Empirical in the figure) depart significantly from the other two sets and from the points obtained directly from the independently fitted Gumbel distribution of each duration. This verifies our claim (subsection 2.3) that the parameters κ and λ of (14) are not in fact constant but they depend on the return period T. In other words, this demonstrates that one cannot use the Gumbel distribution to model the maximum rainfall intensities and simultaneously use approximation (14) for idf curves.

4. Geographical variation of idf curves

The above general framework provides a good basis for studying the geographical variation of idf curves and, more specifically, the construction of maps that can be used to infer idf curves at any point of a particular area. The general idea is to study the variation of the parameters of the idf relationships, instead of the variation of rainfall intensities. The study of parameters can be separated in two phases: first, study of the parameters of the function $b(d)$, and second, study of the parameters of the function $a(T)$. This separation makes possible the incorporation of data from the more dense network of non-recording stations (e.g., 24 or 48 hour depths) in the second phase, thus providing more detailed information of the geographical variation of idf curves. Such data are not appropriate for the first phase because, apparently, the determination of $b(d)$ requires intensities of small durations to be available. The application of these ideas using an extensive data set of a large part of Greece is demonstrated below.

The study area for this demonstration is the Sterea Hellas region (central Greece) with an area of approximately 25 000 km^2 (about 1/5 of the total area of Greece; Figure 4). This region includes five important and many smaller rivers providing water for hydropower, irrigation, and water supply. The Pindus mountain chain on the west side of this region causes heavy orographic rainfall and therefore a wetter rainfall regime, as compared to that of the

east side. Thus, the annual rainfall varies from about 2000 mm in the northwestern part of the region to about 400 mm in the southeastern part (Athens).

Records of maximum intensities at 13 recording stations uniformly distributed in the study area were used. The time resolution in most of the records was 1 hour and thus the durations examined were 1, 2, 6, 12, 24 and 48 hours. In addition, annual series of maximum daily and 2-day rainfall depths were available for the 13 above stations and other 58 non-recording rain gauges (71 stations in total; Kozonis, 1995). Summary data of the rain gauges are shown in Table 2. To the values of the daily and 2-day series, adjustments were made to account for the fact that they are fixed-interval rainfall amounts. The adjusting factors of the bibliography (e.g., Linsley et al., 1975, p. 357) were used, which are 1.13 and 1.04 for the daily and 2-day maxima, respectively. (A more recent research by Dwyer and Reed, 1994, has resulted in a slightly higher value 1.167 for the daily rainfall in UK).

At the first phase we have used the records of maximum intensities of the 13 recording stations, to which we fitted equations of the form (49). This form was suitable for all stations, as the Gumbel distribution was found (using the χ^2 test) to be appropriate for all records. Because the minimum duration was 1 hour, accurate estimation of the parameter θ was not possible, and thus we assumed that $\theta = 0$. As shown in Figure 3, the idf curves become approximately straight lines in the logarithmic plot for $d \geq 1$ hour, even if $\theta \neq 0$. This indicates that the assumption $\theta = 0$ is adequate, if we are interested in durations greater than 1 hour. Consequently, for each station we have three unknown parameters, namely the η, ψ and λ, which have to be estimated. Using the one-step least squares method, we found that the geographical variation of parameters η and ψ is very slight and that of λ is significant. More specifically, η and ψ can be considered as constant within each of three "homogeneous" subregions of the study area. The boundaries of these subregions almost coincide with the divides of the water districts of Western Sterea Hellas, Eastern Sterea Hellas, and Attica, which had been defined in the past using several topographical, climatological and hydrological factors.

At the end of this first phase we compared the series of 24-hour and-48 hour intensities, derived from the recording devices at each of the 13 stations, with the 1-day and 2-day

intensities derived from the daily observations at the same stations. To this aim we fitted Gumbel distributions to the latter series and compared these distributions with those of the former series. Remarkably, in most cases the series of the daily observations resulted in higher intensities for each recording station, even if their data values were not corrected with the above mentioned adjusting factors. This is possibly explained by the different features of the two measuring devices. It seams that the recording devices with their vulnerable mechanisms are more sensible to erroneous recordings, whereas the standard non-recording rain gauges are more reliable due to their simpler structure. This leads us to the conclusion that the daily observations of non-recording devices must never be ignored, even in the case of coexistence of recording devices at the same station.

At the second phase we entered into the calculations the 24-hour and 48-hour maximum intensities of the 71 non-recording stations. As mentioned above, the estimation of the parameters of η and θ of the function $b(d)$ is unattainable using these data. However, if we adopt some values of those parameters, inferred from the previous estimation phase (using data of the recording stations), then we can estimate the parameters ψ and λ of the function $a(T)$ using the data of the non-recording stations. The results of the first estimation phase, which point out that η can be considered as constant for each of the three subregions (and $\theta = 0$ for the entire study area), facilitate the second estimation phase. The analysis of this phase affirmed the outcome of the first phase that ψ is approximately constant for each subregion and allowed a more detailed representation of the variation of λ (71 points).

As a result of the above analysis, only one parameter (λ) has significant geographical variation, whereas the other parameters are constant within subregions. Consequently, one map with contours of λ (also indicating the values of the other parameters per subregion) suffices for the representation of regional analysis. Equivalently, instead of the λ contours, the map may be compiled in terms of any other variable related with λ. This is the case in the map of Figure 4, which contains isohyets of the 5-year 24-hour rainfall depth ($h_5(24)$ in mm). This type of contours was preferred to increase interpretability, as the values of the 5-year 24-hour rainfall depth are more familiar to the user that those of parameter λ. At any point, given the

value of $h_5(24)$, the value of λ is estimated from the following relationship, which is consequence of (49):

$$\ddot{e} = \frac{h_5(24)}{24^{1-\varsigma}\,[\phi - \ln(-\ln 0.8)]} = \frac{h_5(24)}{24^{1-\varsigma}\,(\phi + 1.5)} \qquad (h_5(24) \text{ in mm}) \qquad (55)$$

For the compilation of this map we used the ARC/INFO geographical information system; the contours were drawn using the TINCONTOUR method (ESRI, 1992).

For the verification of the method we used the map of Figure 4 to determine values of all parameters of the idf relationships at the locations of certain recording or non-recording stations. With these values we reconstructed the idf curves and compared them to the direct idf curves, i.e., those obtained from the historic data. Such a comparison is shown in Figure 5 for the Helliniko station, whose direct idf curves were constructed in the application presented in subsection 3.4. We found that the curves constructed indirectly using the map agree well with the direct idf curves.

In recent studies of the geographical variation of idf curves (e.g., Kothyari and Garde 1992), it was attempted to express the regional variation of the parameters of idf curves by introducing climatic descriptors such as the annual rainfall or the maximum monthly rainfall. In this study such links of idf curves to aggregated properties of rainfall were not identified. On the contrary, it was found that parts with very different climate regime may have similar idf curves. For example, in Figure 4 we observe that near Athens, where the annual rainfall is about 400 mm, the 24-hour 5-year rainfall depth is about 80 mm. The same contour of 80 mm appears also in the middle of the Sterea Hellas area, where the annual rainfall is double that of the Athens value, i.e. 800-900 mm. Similarly, the contour 140 mm for the 24-hour 5-year rainfall depth appears both in the eastern and the western Sterea Hellas, although the annual rainfall in the eastern part (about 900 mm) is half that of the western part (about 1800 mm). In conclusion, we did not detect a link between the variability of maximum rainfall intensities and that of the mean annual rainfall. In fact, there no need to do so, as the developed methodology is simple and can be easily performed without considering any annual or monthly properties of rainfall, which are sometimes used for scaling purposes to reduce the computations of regionalisation .

In addition, an attempt was made to relate for each subregion the 24-hour, 5-year depth to the station elevation. It resulted in no significant correlation. This does not mean that the orography does not affect the intense rainfall. Rather, the effect of the orography on maximum intensities is better represented by the geographical location rather than the elevation of the station. Also, we remind that the orography was already considered to divide the area into subregions.

The assumption of constant parameters η, θ and ψ is not a structural constraint of the proposed methodology. Generally, we can allow more than one parameter to vary geographically and we need to construct one map with contours for each varying parameter. For the completeness of the case study examined, we have also used the assumption that both parameters ψ and λ vary geographically, whereas η remains constant for each subregion and θ = 0. For this case a set of two maps giving contours of $h_5(24)$ and $h_{10}(24)$ was constructed (Kozonis, 1995), which can be used to determine both parameters ψ and λ at any point. However, it was found that the use of two sets of contours (and two variable parameters) does not add significant information to that obtained by a single set of contours. This, however, may not be the case if we apply the methodology to other regions.

We emphasise that the objective of the above analysis was the establishment and test of a general methodology for idf curve construction and regionalisation. More analyses and more data are needed for the construction of final maps for Greece, suitable for operational use.

5. Conclusions

The existing typical framework of rainfall idf analysis and synthesis is not free of empirical considerations, which are inconsistent, to some extent, with the theoretical probabilistic foundation of the idf relationships. This paper is an attempt to formulate a more consistent approach to that issue. More specifically, it proposes a general rigorous formula for the idf relationship whose specific forms are explicitly derived from the underlying probability distribution function of maximum intensities. Also, it proposes two methods for a reliable parameter estimation of the idf relationship. The general formulation and parameter

estimation methods are clarified with an application using real-world data. The resulting idf curves incorporate in the same relationship the distribution function of maximum intensity and its functional dependence on duration, and are superior to typical semi-empirical relationships. For a mathematically convenient yet consistent expression of idf relationships, simple approximations of certain complicated distribution functions are presented.

The proposed formulation of the idf relationship constitutes an efficient paramaterisation of this relationship using three to five parameters (depending on the type of the distribution function and the type of the functional dependence of the intensity on duration). An investigation of the geographical variability of the idf relationships performed with data of a large part of Greece, shows that the proposed framework offers a good basis for the regionalisation of idf relationships. Moreover it allows incorporating data from non-recording stations, thus remedying the problem of establishing idf curves in places with a sparse network of rain-recording stations, using data of the denser network of non-recording stations. We emphasise that the objective of the above analysis was the establishment and test of a general methodology for idf curve construction and regionalisation. More analyses and more data are needed for the construction of final maps for Greece, suitable for operational use.

Acknowledgements

application presented belong to the Hellenic National Meteorological Service (HNMS). The data of other recording and non-recording stations belong to HNMS, the MERPPW, and the Public Power Corporation. MS-EXCEL is a registered trademark of the Microsoft Corporation. ARC/INFO is a registered trademark of ESRI Inc.

References

Aron, G., D. J. Wall, E. L. White, and C. N. Dunn, Regional rainfall intensity-duration-frequency curves for Pennsylvania, *Water Resour. Bul.*, 23(3), 479-485, 1987.

Baghirathan, V. R., and Shaw, E. M., Rainfall depth-duration-frequency studies for Sri Lanka, *J. of Hydrol.*, 37(3), 223-239, 1978.

Bernard, M. M., Formulas for rainfall intensities of long durations, *Trans. ASCE*, 96, 592-624, 1932.

Bobée, B., and F. Ashkar, *The Gamma Family and Derived Distributions Applied in Hydrology*, Water Resources Publications, Littleton, Colorado, 1991.

Canterford, R. P., N. R. Pescod, N. H. Pearce, L. H. Turner, and R. J. Atkinson, Frequency analysis of Australian rainfall data as used for flood analysis and design, *Regional flood frequency analysis: Proc. of the Intern. Symp. on Flood Frequency and Risk Analyses*, 14-17 May 1986, Louisiana State University, Baton Rouge, USA, D. Reidel Publishing Co., Boston, MA., 293-302, 1987.

Chen, C.-I., Rainfall intensity-duration-frequency formulas, *J. of Hydraul. Eng.*, 109(12), 1603-1621, 1983.

Chow, V. T., *Handbook of Applied Hydrology*, McGraw-Hill, New York, 1964.

Chow, V. T., D. R. Maidment, and L. W. Mays, *Applied Hydrology*, McGraw-Hill, 1988.

Dwyer, I. J., and D. W. Reed, Effective fractal dimension and corrections to the mean of annual maxima, *J. of Hydrol.*, 157, 13-34, 1994.

Frederick, R. H., V. A. Meyers and E. P. Auciello, *Five- to 60-minute Precipitation Frequency For the Eastern and Central United States*, NOAA Tech. Mem. NWS HYDRO-35, Washington, DC, 1977.

ESRI (Environmental Systems Research Institute), ARC Command References, ESRI, Redlands, CA, 1992.

Gumbel, E. J., *Statistics of Extremes*, Columbia University Press, New York, 1958.

Hershfield, D. M. , *Rainfall Frequency Atlas of the United States for Durations From 30 Minutes to 24 hours and Return Periods From 1 to 100 Years*, U. S. Weather Bureau Technical Paper 40, Washington, DC, 1961.

Kirby, W., Computer oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type 3 distribution, *Water Resour. Res.*, 8(5), 1251-1254, 1972.

Kite, G. W., *Frequency and Risk Analyses in Hydrology*, Water Resources Publications, Littleton, Colorado, 1988.

Kothyari, U. C., and R. J. Garde, Rainfall intensity-duration-frequency formula for India, *J. Hydraul. Eng.*, 118(2) 323-336, 1992.

Kottegoda, N. T., *Stochastic Water Resources Technology*, Macmillan Press, London, 1980.

Koutsoyiannis, D., A stochastic disaggregation method for design storm and flood synthesis, *Journal of Hydrology*, 156, 193-225, 1994.

Koutsoyiannis, D., *Statistical Hydrology* (in Greek), National Technical University Press, Athens, 1996.

Kozonis, D., Construction of idf curves from incomplete data sets, Application to the Sterea Hellas area, Diploma thesis (in Greek), National Technical University, Athens, 1995.

Linsley, R. K. Jr., M. A. Kohler and J. L. H. Paulus, *Hydrology for Engineers*, McGraw-Hill, Tokyo, 2nd edition, 1975.

Miller, J. F., R. H. Frederick and R. J. Tracey, *Precipitation Frequency Analysis of the Western United States*, NOAA Atlas 2, National Weather Service, NOAA, U. S. Department of Commerce, Silver Spring, Md. 1973.

NERC (National Environmental Research Council), *Flood Studies Report*, Institute of Hydrology, Wallingford, 1975.

Oyebande, L., Deriving rainfall intensity-duration-frequency relationships and estimates for regions with inadequate data, *Hydrol. Sci. J.*, 27(3), 353-367, 1982.

Pagliara, S. and C. Viti, Discussion: Rainfall intensity-duration-frequency formula for India by U. C. Kothyari and R. J. Garde, *J. Hydraul. Eng.*, 119(8), 962-967, 1993.

Pierre, D. P., *Optimization Theory With Applications*, Dover, New York, 1986.

Pitman., W. V., A depth-duration-frequency diagram for point rainfall in SWA-Namibia, *Water SA*, 6(4), 157-162, 1980.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1992.

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

Shaw, E. M., *Hydrology in Practice*, Van Nostrand Reinhold, Berkhire, UK, 1983.

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

Smith, J. A., Precipitation, Chapter 3, in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, New York, 1993.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou, Frequency analysis of extreme events, Chapter 18 in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, 1993.

Subramanya, K., *Engineering Hydrology*, Tata McGraw-Hill, New Delhi, 1984.

UNESCO, *National Resources of Humid Tropical Asia - National Resources Research XII*, UNESCO, 1974.

Viessman, W. Jr., G. L. Lewis and J. W. Knapp, *Introduction to Hydrology*, 3rd edition, Happer & Row, New York, 1989.

Wanielista, M., *Hydrology and Water Quality Control*, John Wiley & Sons, New York, 1990.

Wilson, E. M., *Engineering Hydrology*, 4th edition, Macmillan, London, 1990.

Tables

Table 1 Summary of maximum intensity data (in mm/h) of the Helliniko station.

Duration (d_j)	5 min	10 min	30 min	1 h	2 h	6 h	12 h	24 h
Sample size (n_j)	29	29	30	30	30	30	30	20
Minimum value	36.00	24.60	16.40	10.20	6.550	2.400	1.567	0.833
Maximum value	141.60	120.00	74.00	40.90	26.900	11.933	7.242	3.846
Average value	76.22	58.41	35.17	22.04	13.325	5.823	3.520	2.058
Std deviation	29.14	20.32	13.88	8.89	5.660	2.433	1.464	0.786

Table 2 Summary data of rain gauges used in the application of geographical variation of idf curves.

Subregion	Western Sterea Hellas	Eastern Sterea Hellas	Attica	Total area
Number of recording stations	5	4	4	13
Number of non-recording stations	33	18	7	58
Total number of stations	38	22	11	71
Station elevation (range; m)	2-1160	4-830	10-333	2-1160
Station elevation (average; m)	726	313	138	507
Record length (range)	15-37	10-39	10-33	10-39
Record length (average)	25.5	26	24.4	25.5

List of Figures

Figure 1 Comparison of three approximations of the gamma distribution function with scale parameter $\lambda = 1$ and shape parameter (a) $\kappa = 0.25$ ($C_s = 4$), and (b) $\kappa = 16$ ($C_s = 0.5$) (adapted from Koutsoyiannis, 1996).

Figure 2 Empirical (points) and Gumbel (lines) distribution functions of maximum intensities at Helliniko for durations (a) 5 min - 1 hour and (b) 2-24 hours. The continuous and dashed lines (in most cases indistinguishable from each other) correspond to the Gumbel distributions fitted by the robust estimation method, and the one-step least squares method, respectively. The dotted lines (also indistinguishable from the other lines in most cases) correspond to the Gumbel distribution fitted separately to the data of each duration.

Figure 3 Idf curves of Helliniko for return periods 5-5000 years, as obtained using both the robust estimation method (continuous lines) and the one-step least squares method (dashed lines; almost indistinguishable from continuous lines). The points correspond to the intensities obtained directly from the Gumbel distribution of each duration. Dotted lines represent curves obtained using empirical equation (14) (see text).

Figure 4 Map of the Sterea Hellas: (a) morphology; (b) isohyets of the 5-year 24-hour rainfall depth in mm (continuous lines). Dashed lines in (b) are boundaries of the three subregions each having approximately constant parameters η and ψ (see text) whereas circles and squares indicate locations of recording and non-recording stations, respectively (the triangle is the Helliniko station).

Figure 5 Idf curves of Helliniko for duration ≥ 1 hour and return periods 5-5000 years, as inferred from Figure 4 (continuous lines), in comparison with the idf curves obtained from (53) (dashed lines; also shown in Figure 3).
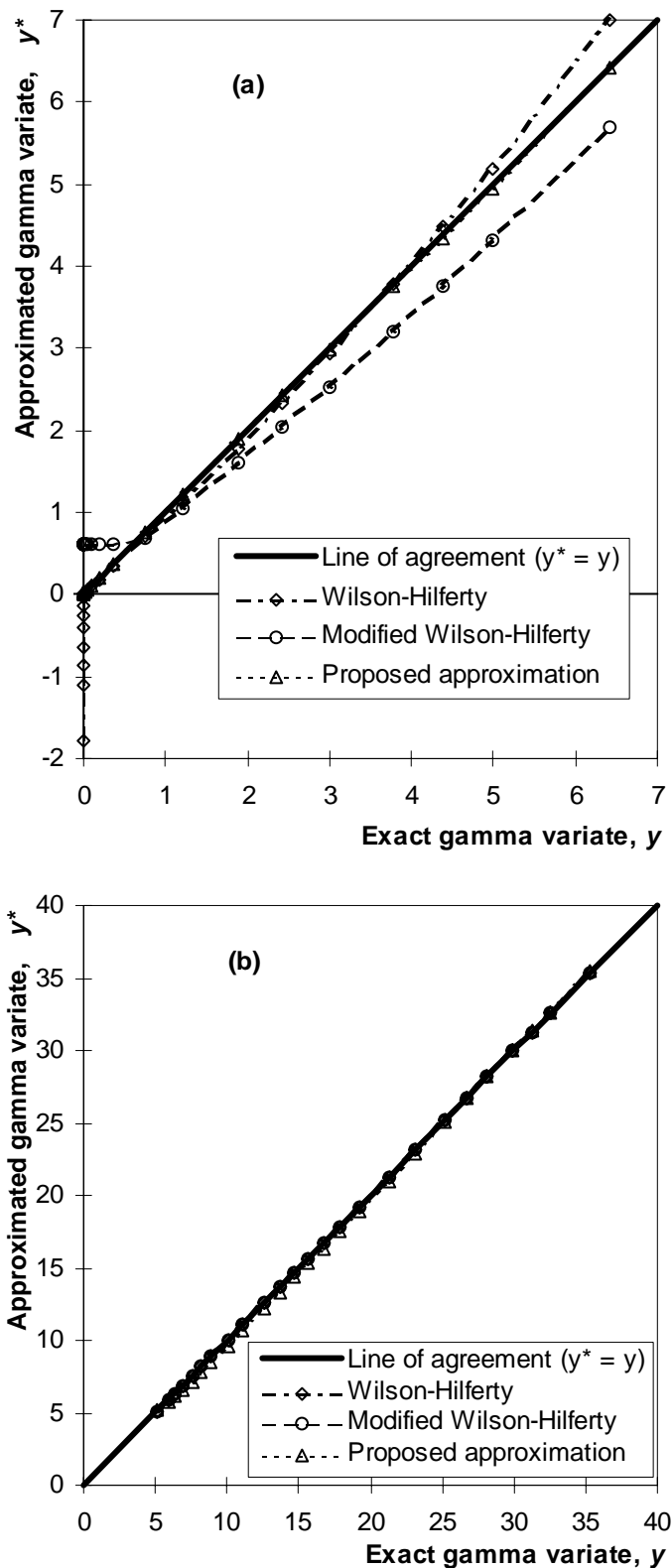
Figure 1 Comparison of three approximations of the gamma distribution function with scale parameter $\lambda = 1$ and shape parameter (a) $\kappa = 0.25$ ($C_s = 4$), and (b) $\kappa = 16$ ($C_s = 0.5$) (adapted from Koutsoyiannis, 1996).
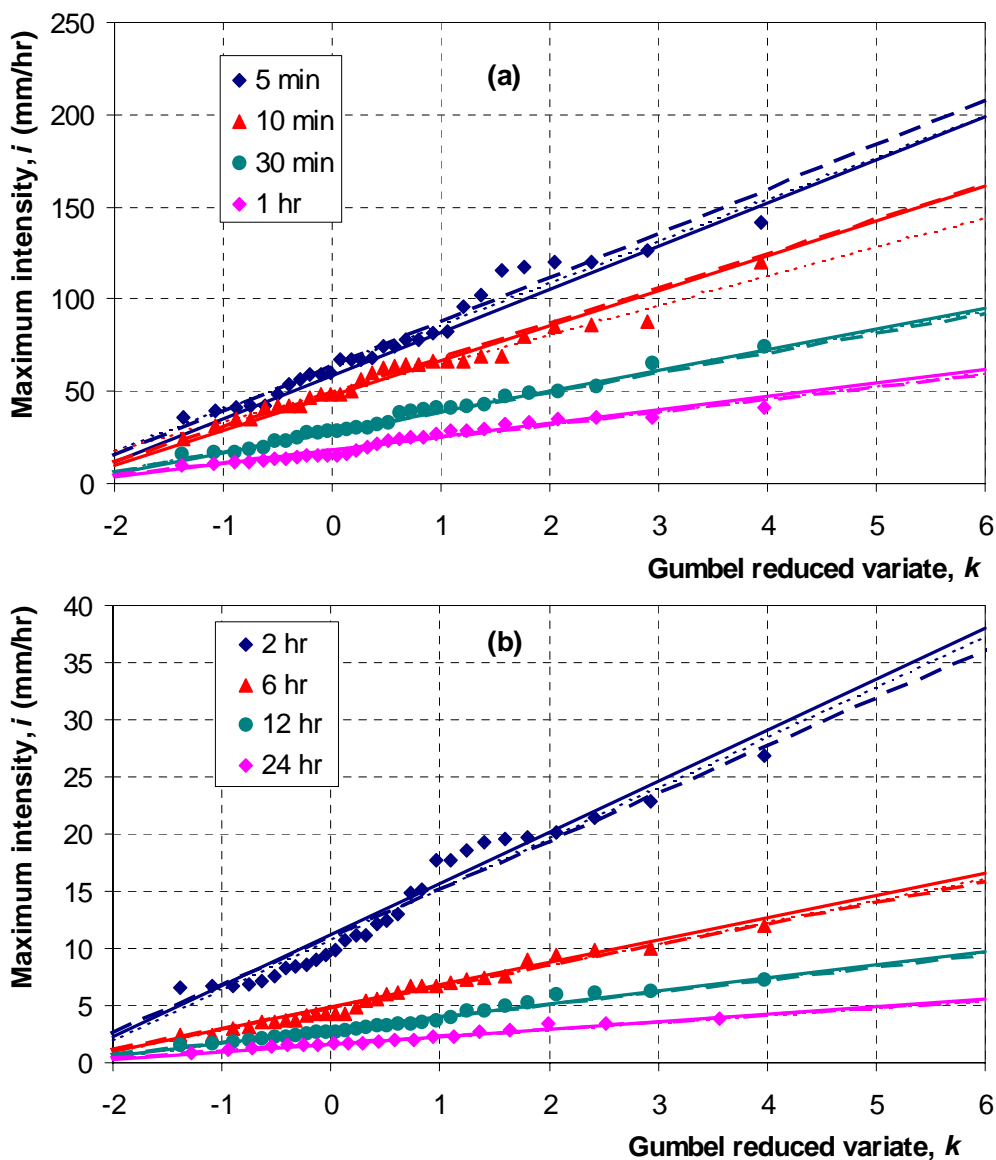
Figure 2 Empirical (points) and Gumbel (lines) distribution functions of maximum intensities at Helliniko for durations (a) 5 min - 1 hour and (b) 2-24 hours. The continuous and dashed lines (in most cases indistinguishable from each other) correspond to the Gumbel distributions fitted by the robust estimation method, and the one-step least squares method, respectively. The dotted lines (also indistinguishable from the other lines in most cases) correspond to the Gumbel distribution fitted separately to the data of each duration.
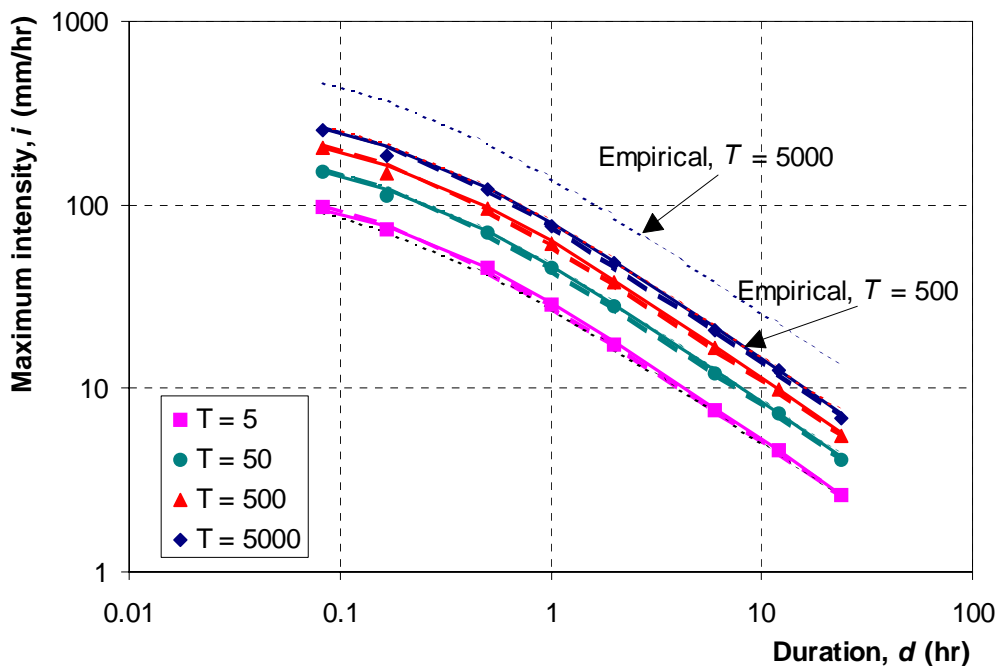
Figure 3 Idf curves of Helliniko for return periods 5-5000 years, as obtained using both the robust estimation method (continuous lines) and the one-step least squares method (dashed lines; almost indistinguishable from continuous lines). The points correspond to the intensities obtained directly from the Gumbel distribution of each duration. Dotted lines represent curves obtained using empirical equation (14) (see text).
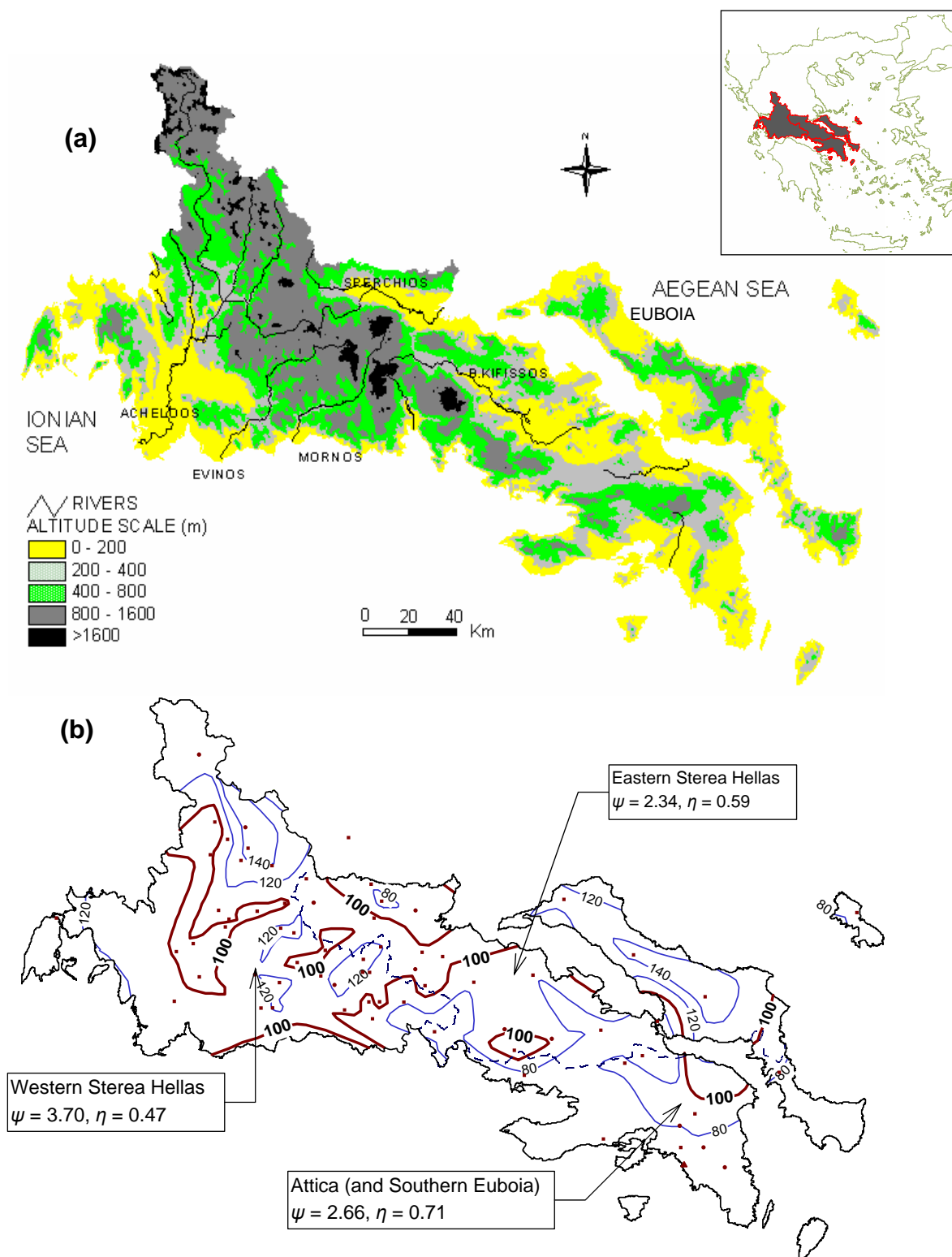
Figure 4 Map of the Sterea Hellas: (a) morphology; (b) isohyets of the 5-year 24-hour rainfall depth in mm (continuous lines). Dashed lines in (b) are boundaries of the three subregions each having approximately constant parameters η and ψ (see text) whereas circles and squares indicate locations of recording and non-recording stations, respectively (the triangle is the Helliniko station).
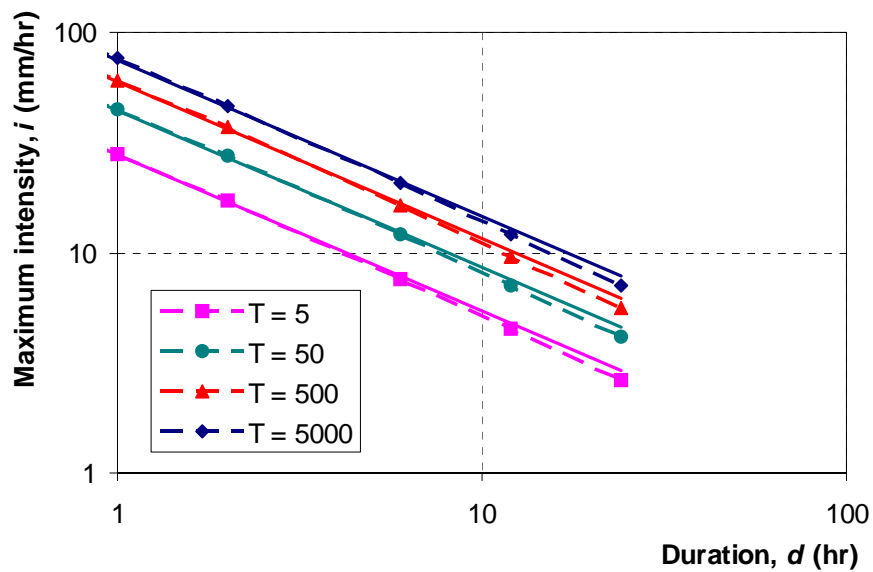
Figure 5 Idf curves of Helliniko for duration ≥ 1 hour and return periods 5-5000 years, as inferred from Figure 4 (continuous lines), in comparison with the idf curves obtained from (53) (dashed lines; also shown in Figure 3).