Article

Regional Ombrian Curves: Design Rainfall Estimation for a Spatially Diverse Rainfall Regime

Theano Iliopoulou [1,*]**, Nikolaos Malamos** [2] **and Demetris Koutsoyiannis** [1]

1   Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zografou, Greece; dk@itia.ntua.gr
2   Department of Agriculture, University of Patras, Theodoropoulou Terma, GR-272 00 Amaliada, Greece; nmalamos@upatras.gr
*   Correspondence: tiliopoulou@hydro.ntua.gr

Abstract: Ombrian curves, i.e., curves linking rainfall intensity to return period and time scale, are well-established engineering tools crucial to the design against stormwaters and floods. Though the at-site construction of such curves is considered a standard hydrological task, it is a rather challenging one when large regions are of interest. Regional modeling of ombrian curves is particularly complex due to the need to account for spatial dependence together with the increased variability of rainfall extremes in space. We develop a framework for the parsimonious modeling of the extreme rainfall properties at any point in a given area. This is achieved by assuming a common ombrian model structure, except for a spatially varying scale parameter which is itself modeled by a spatial smoothing model for the 24 h average annual rainfall maxima that employs elevation as an additional explanatory variable. The fitting is performed on the pooled all-stations data using an advanced estimation procedure (K-moments) that allows both for reliable high-order moment estimation and simultaneous handling of space-dependence bias. The methodology is applied in the Thessaly region, a 13,700 km^2 water district of Greece characterized by varying topography and hydrometeorological properties.

1. Introduction

Ombrian (from the Greek word 'ὄμβρος' meaning rainfall) curves are a standard engineering tool in the form of a mathematical relationship linking rainfall intensity to timescale and return period, usually known as 'intensity-duration-frequency' curves. This term, albeit widely used, appears to be a misnomer, considering that 'duration' does not refer to the actual duration of a rainfall event but rather to the (arbitrary) time scale of averaging the rainfall intensity, while 'frequency' is not meant to be frequency but its reciprocal, i.e., return period. To oppose this common confusion (and having in mind the Aristotelian principle that science presupposes clarity—or saphenia [1]), the term 'ombrian curves' has been used as an alternative name in the past [2–5] and has been adopted here as well.

Ombrian curves have been used in hydrology since the works of Sherman [6] and Bernard [7] and own their popularity to their practical benefits when design problems affected by rainfall extremes are of interest. Although the typical curves have been constructed mostly following an empirical fashion, over the past decades, there have been several attempts to provide a theoretical basis for their modeling, e.g., [8–10], with the most recent advance being their full upgrade to multi-scale models of rainfall intensity [2]. At this point, it could be argued that the issue of deriving the curves for single sites has been efficiently tackled both at the practical level, with diverse methodologies providing satisfactory results for small scales, of the order of minutes to a few days (see the review by

Svensson and Jones [11]), and at the theoretical level as well, achieving their full modeling validity over any scale of interest [2,3].

At the same time, there have been various efforts to produce regional ombrian curves, a considerably more demanding task that is essential when hydrological analyses for large areas are to be performed, and design rainfall estimates are required for multiple sites. This is the case when regional flooding is studied, e.g., when large-scale flood protection works and urban stormwater networks are to be constructed. Even more, the construction of regional curves is critical when rainfall data for at-site analysis are missing, as is often the case. As such, the construction of ombrian curves at regional scales and even at the national level has been a priority for various countries, with a notable early example being the construction of regional curves in the US [12].

Broadly, the construction of regional ombrian curves can follow two different approaches:

(a)   the at-site, independent fitting approach, which consists of separately fitting the curves to multiple gauged sites and using spatial interpolation methods to map the parameters over the whole region.

(b)   the regional, simultaneous fitting approach, which consists of appropriately pooling the data together and obtaining a single model valid over the entire area, which is, in essence, the inverse approach to (a).

Frameworks based on the first approach are older yet still in use, e.g., [12–14], as this approach is easier to apply. Indeed, the task of constructing the curves for a single site is more straightforward, and the same is true for the mapping of the parameters in space, given the ample availability of geostatistical software. However, an important limitation of this approach is that results are very sensitive to data from single stations, which are often short and fragmented and can be impacted by large sampling uncertainty. A further issue with automated spatial interpolation methods is that their structure is obscure and cannot be easily modified, making their results not always interpretable. For instance, it is known that both the number and spatial distribution of available stations impact the reliability of the geostatistical method, but the former is not easy to assess. In this respect, Malamos and Koutsoyiannis [15] note that kriging requires a large number of available data points (at least 100, according to Oliver and Webster, [16], or 50–100 according to other studies [17]), in order to produce a reliable estimation of the variogram. The same authors propose a bilinear surface smoothing (BSS) model [18,19] that is shown to have superior performance to Universal kriging in terms of bias and heteroscedastic behaviors of the data, as well as a more interpretable theoretical structure.

On the other hand, having a single spatial model is a theoretically more powerful approach since it allows simultaneously using all observations to limit uncertainty by 'substituting space for time'; i.e., the principle behind the well-known 'regional frequency analysis' [20]. Nonetheless, the simultaneous use of all stations in the estimation process is challenging, for it requires a rigorous estimation framework to handle the underlying assumptions that control the information content of the pooled records. These are related to the different record lengths of the stations and the presence or not of spatial dependence among them. The most well-known framework of this type is the regional frequency estimation based on the L-moments [20,21]. Yet the latter is founded on the inter-site independence assumption and has been shown to decrease in accuracy as the level of dependence increases [22]. To explicitly address the effect of the spatial dependence in the data, a new regional frequency estimation framework has been proposed that allows high-order properties estimation from spatially correlated data by means of 'knowable' moments' (K-moments of high order [2,23]). Separate space dependence models have been employed in other approaches (e.g., [24]). Evidently, the issue of regional frequency analysis is still an evolving research subject.

Aside from the choice of the theoretical framework for the regional estimation, the formulation of a regionalization approach for the extreme rainfall properties is an additional demanding task that has to be performed in approach (b). The literature on the different types of regionalization techniques for frequency analysis is vast and extensively

covered by previous works [11,25]. The simpler approach considered is the estimation of a common frequency distribution by directly pooling together all stations of a homogenous region [26]. This is a reasonable approach for small areas or areas with limited spatial variability but is less efficient for complex regimes. In such cases, an extension of the same concept is to delineate the region by identifying homogenous sub-regions and pool the data within them, allowing for the possible variation only of a site-specific scale parameter, known as the 'index-rainfall'. This is a widely used approach [11,20,25,27,28], inspired by a similar method in flood frequency analysis, the so-called 'index-flood' method [29]. Considerable limitations of the approach relate to the subjectivity of detecting clusters in complex regimes and the arising of spatial discontinuities at the clusters' boundaries. Remediation to these issues is the 'region of influence approach', which, instead of using fixed boundaries regions, employs individual regions of varying boundaries centered at the site of interest [30]. Irrespective of the arrangements of each method, the need to model the spatial variability of the site-specific parameter(s) still emerges and is often treated through standard interpolation and geostatistical methods as in approach (a).

In this sense, several frameworks resort to a combination of the two approaches (e.g., [28,31,32]). Namely, the data are pooled together for the distribution fitting after being proportionally adjusted by a site-specific parameter, which itself is obtained for a given grid by interpolation methods. This approach bypasses the difficulty of identifying sub-regions based on hydrological similarity and is characteristically called 'regionless' or 'boundaryless' [25,32].

The scope of this work is to construct regional ombrian curves by exploiting and combining recent advances in the different methodological components of the analysis and integrating them into a new framework for regional frequency analysis of rainfall extremes. The latter is based on:

1.   An advanced estimation method of extreme values based on high-order moment estimation while respecting space dependence [2].
2.   A couple of flexible spatial smoothing models [18,19] to describe the spatial variability of extreme rainfall without resorting to uncontrolled interpolation.
3.   A formulation of ombrian curves, which is revisited through recent theoretical developments in the field [2].

This is the first time that the K-moments framework, by now used in frequency estimation for several hydrometeorological variables [23,33,34], has been put into practice in regional frequency estimation of rainfall extremes. This is also the first time that the BSS model framework has been employed for the regionalization of extreme rainfall. As a proof-of-concept, the framework is applied in the region of Thessaly (Greece), utilizing data from 55 stations over a 13,700 km^2 basin area. The large spatial scale, together with the hydrological complexity of the case study, form a challenging test for the methodology, from which new insights into regional rainfall frequency analysis are gained.

2. Methodology

2.1. Mathematical Form of the Ombrian Relationship

Koutsoyiannis ([2]; Chapter 8) recently developed a framework advancing typical ombrian curves to stochastic models of rainfall intensity, valid over any scale supported by the data. This type of ombrian model arises directly from the stochastic properties (dependence structure and marginal distribution) of rainfall intensity and their (non-simple) scaling behavior, and, as such, it can also be applied for simulation of the rainfall process [35]. The framework in [2] can be applied at any time scale, arbitrarily large, to produce the ombrian relationship linking rainfall intensity x to any timescale k and return period T. We note, though, that for large time scales the mathematics becomes somewhat involved. Here we apply the framework only for small time scales, for which a Pareto distribution for the non-zero rainfall intensity is justified. (For larger scales, this should be

replaced by a Pareto–Burr–Feller distribution.) In this case, the Pareto distribution quantile is given as [2]:

$$x = \lambda(k) \frac{\left(P_1^{(k)} T/k \right)^{\xi} - 1}{\xi}, \quad \xi > 0 \tag{1}$$

where $P_1^{(k)}$ is the probability wet, $\lambda(k)$ is a scale parameter and ξ is the tail-index of the Pareto distribution. Both $P_1^{(k)}$ and $\lambda(k)$ are functions of the timescale obtained as [2]:

$$P_1^{(k)} = \frac{1 - \xi}{1/2 - \xi} \frac{\mu^2}{\gamma(k) + \mu^2} \tag{2}$$

$$\lambda(k) = \frac{\mu(1 - \xi)}{P_1^{(k)}} = \frac{(1/2 - \xi)(\gamma(k) + \mu^2)}{\mu} \tag{3}$$

where μ is the mean intensity (constant at all time scales) and $\gamma(k)$ is the climacogram of the process, i.e., the function of the variance across timescale, which can follow different models [2]. A simplification of Equations (1)–(3) is possible based on the following assumptions stemming from the fine-scale behavior of rainfall. For small time scales, of the order of minutes to a few days:

- $P_1^{(k)} \propto k$, and hence we can set the quantity $\beta(k) := k/P_1^{(k)} = \beta = $ constant in Equation (1).
- $\gamma(k) \gg \mu^2$, and thus we can neglect the latter term in their sum.
- we select the generalized Cauchy-type model for the climacogram:

$$\gamma(k) = \lambda_1^2 \left(1 + \left(\frac{k}{\alpha} \right)^{2M} \right)^{\frac{H-1}{M}} \tag{4}$$

where α and λ_1 are scale parameters, with dimensions of time $[t]$ and $[x]$, respectively, and H, M are dimensionless parameters in the interval $(0, 1)$, controlling the long-range (Hurst-Kolmogorov dynamics) and local scaling of the process (fractal behavior) of the process, respectively. For M we take the neutral value $M = 1/2$ as default. We note though that if the focus is on even smaller temporal scales, this value ($M = 1/2$) can be inappropriate.

These simplifying assumptions result in some violations of a full stochastic consistency, as detailed in [2]. However, at small scales, the violations are negligible [2,3]. By virtue of these simplifications, the ombrian relationship is given as:

$$x = \lambda_1^2 \frac{(1/2 - \xi)}{\xi \mu} \left(1 + \frac{k}{\alpha} \right)^{2H-2} \left(\left(\frac{T}{\beta} \right)^{\xi} - 1 \right) \tag{5}$$

Setting $\lambda = (1/2 - \xi)\lambda_1^2 / \xi \mu$ and $\eta = 2 - 2H$, Equation (5) can be rewritten as:

$$x = \lambda \frac{(T/\beta)^{\xi} - 1}{(1 + k/\alpha)^{\eta}}, \quad \xi > 0 \tag{6}$$

where the following five parameters are involved, λ an intensity scale parameter in units of x (e.g., mm/h), β a timescale parameter in units of the return period (e.g., years), α a timescale parameter in units of timescale (e.g., h) with $\alpha \geq 0$, η a dimensionless parameter with $0 < \eta < 1$, and $\xi > 0$ the tail index of the process.

It is easily observed that Equation (6) can be written concisely as the quotient of two separable functions $b(T)$ and $a(k)$ of the return period and the timescale, respectively, in the form:

$$x = \frac{b(T)}{a(k)} \tag{7}$$

with

$$a(k) = \left(1 + \frac{k}{\alpha}\right)^{\eta}$$

(8)

being a function of time scale, in wide use as an approximation [2,36] but here resulting as a consequence of the climacogram, and $b(T)$ a function of the return period that is analytically derived from the distribution function of the rainfall intensity.

In the case that the return period of the rainfall intensity is determined based on rainfall exceedances extracted from the full series (i.e., Peaks over Threshold, POT), a Pareto distribution can be generally assumed for modeling the rainfall intensity, as implied in Equation (6). However, if the return period is determined based on series of annual maxima (AM) of rainfall intensity, then both long-term empirical evidence and theoretical arguments support the use of the Extreme Value Type 2 (EV2) distribution from the Generalized Extreme Value (GEV) distribution family [37–39]:

$$F(y) = \exp\left(-\left(1 + \xi\left(\frac{y}{\nu} - \psi\right)\right)^{-\frac{1}{\xi}}\right), \quad y \geq \nu\left(\psi - \frac{1}{\xi}\right)$$

(9)

where ψ (dimensionless), $\nu > 0$ (units same as in y) and $\xi > 0$ (dimensionless) are location, scale, and shape parameters, respectively. It should be mentioned that the case of $\xi < 0$ is not appropriate for maximum rainfall, since it presumes the existence of an upper limit for the variable, which is inconsistent with the physical reality. Furthermore, the case of $\xi = 0$, i.e., assuming a Gumbel (Extreme Value Type 1—EV1) distribution for the maximum rainfall intensity, is also not supported by worldwide empirical evidence and is to be avoided in general [37]. Therefore, it is not developed herein, but further details for this case are given in Koutsoyiannis [2].

Equivalently, the EV2 distribution as given by Equation (9) can be re-parameterized consistently to Equation (6) as follows:

$$F(y) = \exp\left(-\frac{\Delta}{\beta}\left(\frac{y}{\lambda} + 1\right)^{-\frac{1}{\xi}}\right)$$

(10)

where $\beta = (1 - \xi\psi)^{1/\xi}\Delta$ and $\lambda = (1 - \xi\psi)\,\nu/\xi$ and $\xi > 0$.

The variable y represents either the rainfall intensity x or, equivalently, the product $x\,a(k)$ (Equation (7)). Solving Equation (10) in terms of y and substituting $F(y) = 1 - \Delta/T$, where $\Delta = 1$ year for annual maxima, yields:

$$y_T = \lambda\left((-(\beta/\Delta)\ln(1 - \Delta/T))^{-\xi} - 1\right)$$

(11)

Therefore, by substituting Equations (8) and (11) in (7), the following generalized form of ombrian curves for annual maxima is derived:

$$x = \lambda\frac{(-(\beta/\Delta)\ln(1 - \Delta/T))^{-\xi} - 1}{(1 + k/\alpha)^{\eta}}, \quad \xi > 0$$

(12)

It is easily shown that for small return periods, Equation (6) deriving from a Pareto distribution yields higher intensity than Equation (12), whereas for larger return periods ($T > 10$ years), the two are practically indistinguishable, given that for small Δ/T holds $\ln[1 - (\Delta/T)] = -(\Delta/T) - (\Delta/T)^2 - \cdots \approx -\Delta/T$. Therefore, even in the case that the model fitting is based on Equation (12), i.e., when annual maxima are used (which is also the case here), it is safer, from an engineering point of view, to express the final model used to obtain design rainfall in the form of Equation (6). Yet, the availability of either POT or AM series determines which of Equation (6) or (12), respectively, will be used for model fitting. Following a slightly different parameterization, Equations (6) and (12) were also

proposed by Koutsoyiannis et al. [8] for small scales, which nevertheless are sufficient for most engineering applications of ombrian curves, namely those involving flood analyses.

The advantage of this simplified approach is precisely the separability of $a(k)$ and $b(T)$ functions that allows for a two-step procedure for the parameter estimation. This turns out to be convenient for typical applications and even more for regional analyses, as will be shown next. It also has attractive flexibility in using different sources of data. Namely, a reliable determination of the timescale parameters α and η requires data from fine-scale records, whereas the parameters of the function $b(T)$, including the highly uncertain tail index, are better inferred from daily raingauge data, which are usually more systematic and less prone to erroneous recordings [8].

2.2. Regionalization Method: Bilinear Surface Smoothing Models for the 24 h Average Annual Rainfall Maxima

The target of the regional model is to generalize Equation (6) or (12) in space, achieving their applicability for any set of coordinates in a given area. To efficiently describe the spatial heterogeneity in a region without resorting to a great number of parameters and uncontrolled interpolation, we should make an assumption on which parameters we consider as regionally varying. It is reasonable to begin by applying diagnostics for the spatial heterogeneity of rainfall in the study area, both in terms of the $a(k)$ and $b(T)$ functions' parameters, bearing in mind that selected parameters should be reliably estimated from single-site data. Following this rationale and supported by the available data, which in our case are series of annual maxima at different temporal scales, as shown in Section 3.2, we choose to spatially model the parameter λ of Equation (12). This corresponds to the scale parameter of the EV2 distribution, which is proportional to the mean value of the process, i.e., the average annual maxima at each location for any timescale. Since the transformation between timescales is controlled by the $a(k)$ function, it suffices to model the average maxima at a single, convenient timescale. We choose the 24 h scale due to the much greater availability of daily data in the study region. Depending on the need to apply further complexity, the location parameter of the EV2 distribution could be another option for spatial modeling. Yet it is not advisable to spatially model the highly uncertain shape parameter (i.e., the tail index) based on single-site data unless long-term empirical evidence supports a spatial variation thereof. Parameters α and η, which control the timescale transformations of the curves, are very sensitive to the existence of sub-daily data [8], which are scarce in the study region, and thus, they do not constitute good choices for spatial modeling either. For these reasons, we apply common values for the rest of the parameters over the whole area. The choice of the mean of the AM distribution as the parameter to be regionalized has been proven to be a robust choice in regional frequency analysis, dating back to the index-flood method [29] and several applications thereafter (see [11]).

Having chosen to spatially model only the λ parameter related to the mean value of the annual maxima series, we have to identify a model for the spatial variation. As already discussed in the introduction, a large region with complex topography cannot be efficiently treated as a homogenous area. In this case, instead of identifying several sub-regions, which may give rise to abrupt changes in the final results with questionable physical basis, it is better to explicitly model the process at any given point in the area of interest. Towards this 'regionless' modeling approach, we apply a framework of spatial smoothing modeling that is described next.

We apply two versions of a bilinear surface smoothing model proposed by Malamos and Koutsoyiannis [18,19], generalizing in 2D a previous numerical smoothing and interpolation method [40,41]. A brief overview of the mathematical framework is presented since it is detailed in the aforementioned publications. The general idea behind both methods is to compromise the trade-off between the objectives of minimizing the fitting error and the roughness of the fitted bilinear surface, therefore termed bilinear surface smoothing (BSS). The larger the weight of the first objective, the rougher the surface will appear, while the opposite is true for a larger weight of the second objective.

The mathematical framework of BSS suggests that fitting is meant in terms of minimizing the generalized cross-validation error (GCV; [42]) between the set of the given data points and the corresponding estimates. The general estimation function, $\hat{z}_u$, for point u on a plane, according to the BSS method, is:

$$\hat{z}_u = d_u \tag{13}$$

where d_u is the value of the fitted bilinear surface d at that point.

The BSS method can be extended by the introduction of an additional explanatory variable (bilinear surface smoothing with an explanatory variable; BSSE) at a denser dataset compared to that of the main variable, as follows. We assume that at the locations of the given data points, we also know the value of an explanatory variable t, and therefore for each point z there corresponds a value t. In this case, the general estimation function for point u is:

$$\hat{z}_u = d_u + t_u e_u \tag{14}$$

where d_u, e_u are the values of two fitted bilinear surfaces at that point, namely d and e, while t_u is the value of the explanatory variable at that point. This is not a global linear relationship but a local linear one as the quantities d_u and e_u change in space.

In the case of the BSS, there are four adjustable parameters for surface d: the numbers of intervals along the horizontal and vertical direction, respectively, i.e., mx, my, and the corresponding smoothing parameters $\tau_{\lambda x}$ and $\tau_{\lambda y}$. The incorporation of the explanatory variable for the BSSE case adds two more adjustable parameters: the smoothing parameters $\tau_{\mu x}$ and $\tau_{\mu y}$ corresponding to surface e. The values of all the smoothing parameters are restricted in the interval $[0, 1)$ for both directions [18]. When the smoothing parameters are close to 1, the resulting bilinear surfaces exhibit greater smoothness, whereas, for small values of these parameters, interpolation among the known points is obtained.

A desirable feature of the method for regional analyses is the fact that it is proven reliable even in the case of few and scarce data, in contrast to common geostatistical methods that require a denser data network to be applied reliably (e.g., to estimate a semivariogram) [15]. It is important to note that the method is also parsimonious in terms of the number of parameters and the choices involved in the modeling. This is evident when compared to the standard kriging framework, in which one has to decide among the n different types of the method, among the $4n$ standard variogram types (i.e., spherical, exponential, Gaussian and power), and also identify the values for the range, sill, and nugget, resulting in a total of $12n$ choices, depending on the selected number of kriging methods. This increases the complexity of the approach and may increase the subjectivity as well, given that an objective framework for basing these decisions is lacking.

2.3. Bilinear Surface Smoothing Model Parameters Estimation

As mentioned, the parameter estimation methodology for both the BSS and BSSE methods is based on the minimization of GCV error, and therefore, there is no effect in terms of heteroscedasticity. For a given combination of the bilinear surface segments, mx and my, the minimization of GCV error results in the optimal values of $\tau_{\lambda x}$, $\tau_{\lambda y}$ and $\tau_{\mu x}$, $\tau_{\mu y}$. This can be repeated for several trial combinations of mx and my values until the global minimum of GCV is reached.

Both variants of the method are applied, one taking into account only the coordinates of the stations (BSS) and the second one exploiting as well the elevation of the stations (BSSE) as an additional explanatory variable. In the second case, a digital elevation model of the wider area (Thessaly and neighboring areas) with 90 m resolution at the equator (SRTM; [43]) is employed both for extracting the elevation at the stations' coordinates and for making estimations for any point in the given grid.

For the objective evaluation of the two methods, two statistical indices, the root mean square error (RMSE) and the mean absolute error (MAE), are compared in terms of (a) performance using all data and (b) the leave-one-out cross-validation performance, i.e.,

when the estimation at each coordinate set is performed by omitting the known value at that position. This analysis is presented in Section 4.1.

2.4. Timescale Parameters Estimation

The simplified version of the ombrian model also allows for a simplified fitting procedure. By utilizing the separability of functions $a(k)$ and $b(T)$ in this version, an independent, two-step fitting approach can be used, as introduced by Koutsoyiannis et al. [8]. Namely, Equation (12) can be written as:

$$\left(1 + \frac{k}{\alpha}\right)^{\eta} x = \lambda\left((-(\beta/\Delta)\ln(1 - \Delta/T))^{-\xi} - 1\right) \tag{15}$$

From this expression, it is easy to see that for the different timescales k_j the stochastic variables $\underline{y}_j := a(k_j)\underline{x} = (1 + k/\alpha)^{\eta}\underline{x}$ have a common distribution function, with the $\underline{y}_j$ for the different k_j being samples of it. Let then, $\underline{y}_{ji} := a(k_j)\underline{x}_{ji}$ of length $n = \sum_j n_j$ denote the merged sample of all sub-samples $\underline{x}_{ji}$ of size n_j corresponding to timescale k_j. Let also $\underline{r}_{ji}$ denote the rank of each sub-sample $\underline{x}_{ji}$ in the merged sample $\underline{y}_{ji}$ so that the mean rank of each sub-sample is given as $\underline{r}_j = \sum_i \underline{r}_{ji}/n_j$. Replacing all $\underline{r}_{ji}$ with the mean rank value $\underline{r}_j$ we obtain a sample of n values, with n_1 equal to r_1, n_2 equal to r_2 etc. Then the mean and variance estimators are, respectively:

$$\bar{\underline{r}} := \frac{1}{n}\sum_j n_j \underline{r}_j \tag{16}$$

$$\underline{\gamma}_r := \frac{1}{n}\sum_j n_j\left(\underline{r}_j - \bar{\underline{r}}\right)^2 \tag{17}$$

If no ties are present among the different ranks, then $\bar{\underline{r}} = (n+1)/2$.

Following the assumption that the samples are from the same distribution, given by the right-hand side of Equation (15), then each $\underline{r}_j$ should be close to the mean $\bar{\underline{r}}$ while the variance should be minimal. Therefore, we can find the parameters α and η as the values that minimize the estimate of the variance γ_r from the observations x_{ji}. The original values $\underline{y}_{ji}$ could be used as well instead of the ranks, yet the use of the ranks makes the estimation process more robust to outliers. In order to improve the fit to the higher quantile region, we could also use a part of the data of each sample belonging to the highest $1/2$ or $1/3$ of the data [8]. In this study, the highest $1/2$ is used.

The limited availability of sub-daily stations, in combination with their short records, hinder the reliable estimation of the time scale function parameters at each station separately. It is also pointed out that parameter α is very sensitive to small scales intensities and ideally requires sub-hourly data to be reliably estimated [8]. Given that few stations have data at such scales, to limit uncertainty, we apply the methodology described above simultaneously to the sample of all fine-scale raingauges. In particular, we identify the set of the time scale parameter values as the one that minimizes the sum of all stations' variances, each of which is given by Equation (17).

2.5. Regional Estimation of Distribution Parameters through K-Moments

Having estimated the α and η parameters, it remains to specify the parameters of the $b(T)$ function through the following procedure. We form the pooled sample comprising all stations' annual rainfall maxima at the 24 h scale after first standardizing (dividing) them by their theoretical mean value, given by the spatial smoothing model. To the pooled standardized sample of annual maxima, we fit the EV2 distribution using the method of the non-central K-moments [2,23]. K-moments are newly proposed moments developed with the aims of being knowable for very large orders (depending on the sample size) and

also interpretable in terms of order statistics. They are closely related to the probability-weighted moments [44], but their formulation is simpler and more intuitive, which are attractive qualities similar to the ones of the L-moments [21]. The distinctive feature of K-moments, though, is that they are tailored to perform extreme-oriented analyses, as they enable reliable estimation of very high-order moments. What is more, each high-order K-moment estimate can be directly assigned a return period, which provides a direct means to empirical estimation of probability, alternative to order statistics. Furthermore, this estimation can be appropriately modified in the case that there is bias due to dependence [2]. In particular, K-moments allow straightforward estimation of high-order moments even for spatially dependent data, which is a rare property. Namely, typical regional applications of L-moments do not go beyond the 4th moment (L-kurtosis) estimation. This advantage of K-moments is exploited in the regional frequency analysis, which is particularly sensitive to high-order moments.

Koutsoyiannis [2,23] has introduced several variants of K-moments, of which here we use the simplest non-central variant, defined as:

$$K'_p := p\mathrm{E}\left[(F(\underline{x}))^{p-1}\underline{x}\right] \tag{18}$$

for the moment order $p \geq 1$. K'_p has the important property that it equals the expected value of the maximum of p independent stochastic variables identical to $\underline{x}$, i.e.,

$$K'_p = \mathrm{E}\left[\max\left(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_p\right)\right] \tag{19}$$

The estimators of the non-central *K*-moments are given by the following formulae:

$$\hat{\underline{K}}'_p = \sum_{i=1}^{n} b_{inp}\, \underline{x}_{(i:n)} \tag{20}$$

$$b_{inp} = \begin{cases} 0, & i < p \\ \dfrac{p}{n}\, \dfrac{\Gamma(n-p+1)}{\Gamma(n)}\, \dfrac{\Gamma(i)}{\Gamma(i-p+1)}, & i \geq p \geq 0 \end{cases} \tag{21}$$

where $\underline{x}_{(i:n)}$ is the *i*th smallest variable in a sample $\underline{x}$, of size n, (the *i*th item of the sample in ascending order) and p is the order of the moment, which can be any positive number $p \leq n$. In addition, the following holds:

$$\sum_{i=1}^{n} b_{inp} = 1 \tag{22}$$

The fact that $b_{inp} = 0$ for $i < p$ means that as the moment order increases, fewer data are used in the estimation, until only one is left, the maximum, when $p = n$, and $b_{nnn} = 1$. For $p > n$, $b_{inp} = 0$ for every i, $1 \leq i \leq n$, and the estimation becomes impossible. The first order non-central *K*-moment is the mean value of the sample.

K-moment values can be assigned a return period as follows [2]:

$$\frac{T\left(K'_p\right)}{\Delta} = p\Lambda_p \approx \Lambda_\infty p + (\Lambda_1 - \Lambda_\infty) \tag{23}$$

where Λ_1, Λ_∞ are coefficients depending on the distribution function. For the EV2 distribution it is shown [2] that the Λ coefficients are functions of the shape parameter ξ:

$$\Lambda_1 = \frac{1}{1 - \exp\left(-(\Gamma(1-\xi))^{-\frac{1}{\xi}}\right)} \tag{24}$$

$$\Lambda_\infty = \Gamma(1-\xi)^{\frac{1}{\xi}} \tag{25}$$

For validation purposes, the following relationship of empirical return periods based on order statistics is also used, which is shown to provide an unbiased estimate of the logarithm of the return period [2]:

$$\frac{T_{(i:n)}}{\Delta} = \frac{n + e^{1-\gamma} - 1}{n - i + e^{-\gamma}} = \frac{n + 0.526}{n - i + 0.561} \tag{26}$$

The procedure outlined above could be directly applied for assigning return periods to the K-moments of any single station, and the parameters of the EV2 distribution could be obtained by minimizing an error metric (e.g., MAE or RMSE) between the theoretical quantiles and the empirical K-moments, or between the corresponding return periods. However, for the pooled sample, the resulting information gain, and thus the maximum return period that can be estimated from the data, is a function of the sample's dependence structure. It is well known that in the case of cross-correlated variables, the quantity of information for the variable corresponds to a smaller sample of length compared to the case of independence but still greater than that of an individual station.

The framework of K-moments allows for the effect of dependence to be explicitly accounted for in the estimation of the return period. This is achieved through proper modification of the order of the moments of the unified sample, p', which in turn modifies the estimation of the return period. In particular, let n_1 denote the sample length of each station, m denote the number of stations, and $n = m\,n_1$ denote the size of the merged sample, and then the following methodology is applied [2]:

- For $p \le n_1$ we set $p' = p$.
- For $p > n_1$ the following approximation is used. We estimate the equivalent Hurst parameter H, based on the spatial correlation of the stations ρ:

$$H = \frac{1}{2} + \frac{\ln(1+\rho)}{2\ln 2} \tag{27}$$

Based on the estimated H the following coefficient is used for bias correction, Θ^{HK}:

$$\Theta^{\mathrm{HK}}(n, H) \approx \frac{2H(1-H)}{n-1} - \frac{1}{2(n-1)^{2-2H}} \tag{28}$$

Then the modified orders of the moments are obtained as:

$$p' \approx 2\Theta + (1 - 2\Theta)(p - n_1 + 1)^{((1+\Theta)^2)} + n_1 - 1 \tag{29}$$

and their corresponding return periods are adjusted accordingly based on Equation (23).

It is obvious that n_1 controls the maximum moment order, which is unaffected by dependence. In the case that the stations have different lengths, n_1 can be estimated as the average record length of all stations (here, $n_1 = 42$). An increased value of n_1 suggests that the information gain is also increasing, as fewer return periods are modified downwards, while the opposite is true when n_1 decreases. In order to bypass the uncertainty regarding the modification of the return periods based on the estimated correlation structure, a good strategy is to use for model calibration only the moment orders up to n_1, and employ the higher moments for validation. In so doing, the moments used in the calibration are still much more than the ones used in regular moment fitting procedures (typically up to 3 or 4 orders), while a second-moment set is also available for validation.

To transform the parameters of the EV2 distribution (expressed as in Equation (10)) of the standardized 24 h rainfall maxima to the final ombrian $b(T)$ parameters, we use the following procedure. Let u_T denote the 24 h annual maximum rainfall value for return period T standardized by its mean value μ. Then the rainfall intensity for any station at the 24 h timescale is $x_T^{(24\,\mathrm{h})} = \mu\,u_T/24$, where μ is the mean value used in the standardization. Likewise, the quantity $y_T := x_T^{(24\,\mathrm{h})}\,(1 + 24/\alpha)^{\eta}$, whose distribution defines the function

$b(T)$ of the ombrian relationship will be $y_T = \mu \, u_T (1 + 24/\alpha)^\eta /24$. Consequently, the variable y shares the same distribution function with the variable u with the same shape and location parameters, and scale parameter proportional to the one of u by a factor of $\mu \, (1 + 24/\alpha)^\eta /24$, i.e.,

$$\xi = \xi_u, \ \ \beta = \beta_u, \ \ \lambda = \lambda_u \, \mu \, (1 + 24/\alpha)^\eta /24 \tag{30}$$

where the subscript u denotes the standardized rainfall maxima at the 24 h scale and α is expressed in h. It is recalled that the mean value μ for each location used in the standardization is derived from the BSS/BSSE models.

In this way, parameters α and η, which are estimated simultaneously from all stations, in combination with the parameters of the distribution of the standardized maximum 24 h rainfall values and the spatially modeled mean value of the process by the BSS/BSSE models fully determine the ombrian curves, as given by Equations (6) and (12).

3. Data

3.1. Study Area

The study region is the geographic area of the Water District (WD) of Thessaly (~13,700 km^2) which is one of the 14 WD of Greece. The district extends mostly within the administrative region of Thessaly, while it also includes a small part of the region of Central Greece and a small part of the Western and Central Macedonia region. The topography of the area is characterized by the existence of four mountain ranges in its perimeters, Olympus-Kamvounia in the north, Pindus in the west, Othrys in the south, and Pelion-Ossa in the east, which surround the Thessalian plain that rests in the central area (Figure 1a). The Thessalian plain contains the largest part of the water bodies of the district and is traversed by the Pineios river and its tributaries. It is also the largest agricultural area in Greece, with lowland topography making it prone to frequent and heavy flooding [45,46]. A recent flood event of the 18–19 September 2020 triggered by a Mediterranean cyclone has caused human and livestock losses and extensive structural and agricultural damages to the area, sparking a revitalization of the decade-long initiatives for improving the area's flood protection design and strategy [47]. The climate of the Western region is continental, while the Eastern region has a typical Mediterranean climate, while the rainfall pattern exhibits strong differences between the lowlands and the mountain regions [48]. These characteristics of the study region, namely its hydrometeorological diversity, vast spatial extent, and criticality of flood risk, make it a challenging case study for the application of the methodology.

3.2. Data Processing and Quality Control

The construction of ombrian curves is based on rainfall intensity data at a range of timescales, typically starting from fine scales, i.e., 5 to 60 min, and extending to the 24 or 48 h scale for common applications. To this aim, we assemble a set of 17 rainfall records from tipping buckets and telemetric stations, recording data at the 5–30 min timescale and 61 rainfall records from daily raingauges. The data are obtained from the Public Power Corporation (PPC) of Greece, the Hellenic Ministry of Environment and Energy (HMEE), the Hellenic Ministry of Agricultural Development and Food (HMADF), the Hellenic Ministry of Agriculture (HMA), the Hellenic National Meteorological Service (HNMS) and the meteo network [49] of the National Observatory of Athens. The properties of the stations are detailed in Tables S1 and S2 of the Supplementary material.

We aggregate the original series at a range of timescales from 5 min to 48 h, with $k =$ 0.08, 0.17, 0.25, 0.5, 1, 2, 6, 12, 24, 48 h (depending on data available at the finest scale), and we extract the maximum rainfall depth at each scale $h^{(k)}$ for all hydrological years. Accordingly, we compute the corresponding rainfall intensity at the given scale as $x^{(k)} = h^{(k)}/k$, thus deriving the empirical rainfall intensities corresponding to the annual maxima of the hydrological years. The reason for using AM series instead of POT or even

the full data series for the estimation of the extremes is that several historical records are available only in this form.
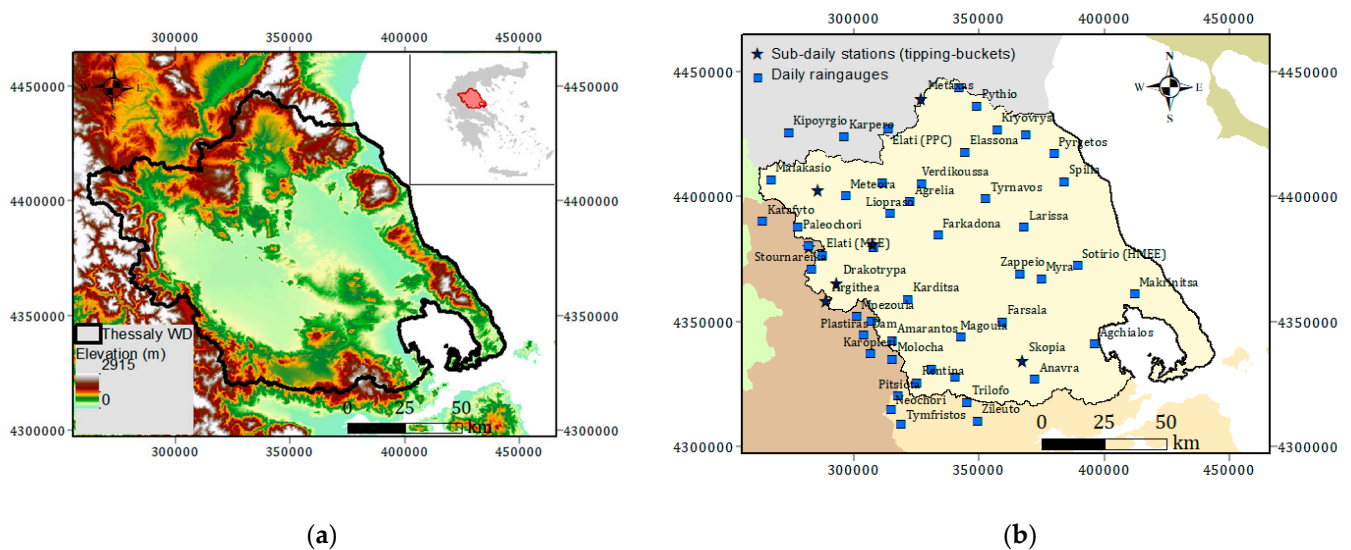
Figure 1. (**a**) Spatial extent and elevation of the study region (Thessaly's Water District in Greece). (**b**) Geographic location of the 55 rainfall records, from daily raingauges and sub-daily tipping-buckets, with data at the 24 h scale (used in the BSS/BSSE models and the EV2 distribution modeling).

We note that the choice of the starting point for the aggregation is arbitrary, and a change thereof would likely result in adifferent estimate. For this reason, it was a common hydrological practice in the past to either take the maximum estimate resulting from all possible positions of the starting point or 'inflate' the given estimate by a specific factor, known as the Hershfield coefficient [50]. Although this practice aims for safer estimates from an engineering point of view, in theory, all realizations of a stochastic process are equivalent, and there is no theoretical basis to 'correct' them. In fact, by correcting the series, we distort its stochastic properties and, instead of studying $\underline{x}_\tau^{(k)}$, the behavior of $\underline{w}_\tau^{(k)} := \max_j\left(\underline{x}_{\tau+j}^{(k)}, j = 0, \ldots k-1\right)$ is studied, which is a different stochastic process [2]. Hence, we do not apply the Hershfield coefficient.

To ensure a good quality dataset for our analysis, we use stations that have at least 12 years of data, and we undertake quality checks based on hydrological experience in the study area. In particular, we perform spatial consistency checks excluding stations with systematically lower recordings in comparison to neighboring ones. In addition, we perform hydrological consistency checks, i.e., to ensure that single-site empirical maximum rainfall is consistent with hydrological experience worldwide, suggesting an unbounded right tail of sub-exponential type [37,38]. We note that poorly maintained raingauge records (e.g., in remote mountainous areas) sometimes exhibit maximum rainfall recordings of the same (or nearly the same) amount due to spillage effects during storm events. In this case, a bounded GEV distribution might falsely emerge.

After screening with these criteria and excluding stations with severe inconsistencies, the resulting set of stations includes 48 daily raingauge stations, 7 of which are at locations gauged by tipping buckets as well, and 14 sub-hourly tipping bucket/telemetric stations. The estimation of the extreme properties and of the average 24 h AM rainfall (Sections 2.3 and 2.5) is based on the set of daily raingauge stations due to the latter being more and of larger record lengths. Yet, maximum rainfall data at the 24 h scale are also employed from the fine-scale rainfall stations when daily raingauge data are not available at the same location. Taking the latter into account, the distribution properties of the maximum rainfall are estimated using a combined set of 55 samples of 24 h annual rainfall

maxima, whose spatial distribution is depicted in Figure 1b. The set of the 14 fine-scale rainfall stations is used in the estimation of the timescale parameters (Section 2.4).

4. Results

4.1. Fitting of the Bilinear Surface Smoothing Models

Before fitting the regional model, we explore the spatial variability of the rainfall regime by identifying the variations in the mean and standard deviations of the rainfall intensity and the possible association with the elevation of the stations. In Figure 2, it can be seen that the first two moments of the rainfall intensity across scales follow a similar statistical behavior, and, therefore, it is reasonable to apply common timescale parameters (i.e., the function $a(k)$). In terms of the average annual maxima at the 24 h scale, there also appears to be a positive association with the stations' elevation, although this is not verified in all cases (Figure 3). Therefore, it seems that the elevation might serve as an explanatory variable, but it needs to be incorporated into a more general spatial model identifying additional patterns of the rainfall maxima in space.
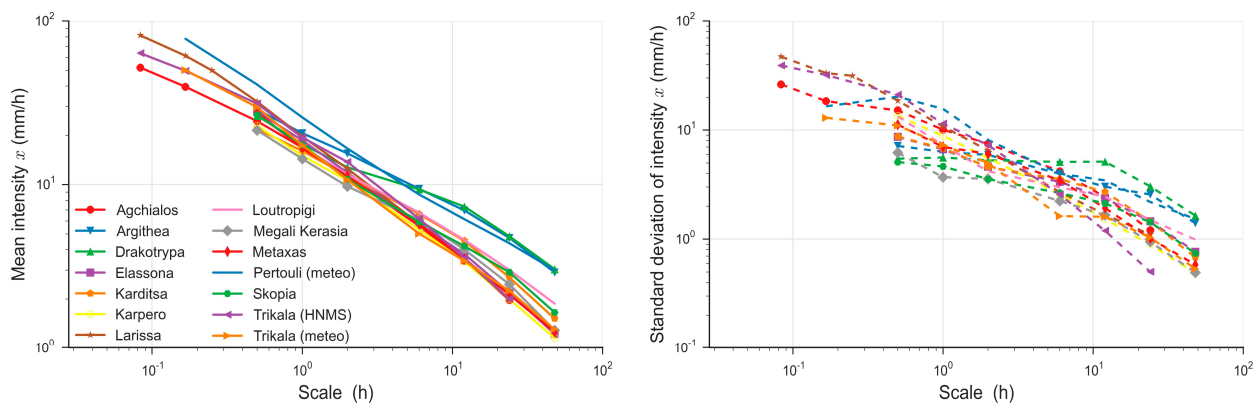
Figure 2. Mean and standard deviation of the empirical rainfall intensities for 5 min to 48 h scales for the 14 fine-scale rainfall stations.
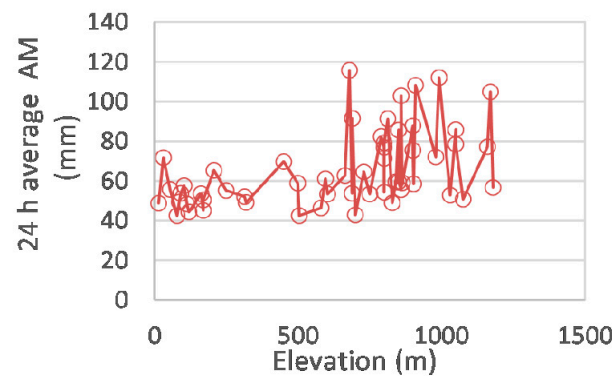
Figure 3. Average annual rainfall maxima (mm) at the 24 h scale vs. the stations' elevation.

To explore the suitability of elevation as an explanatory variable in an objective manner, we evaluate its performance within the BSS/BSSE model framework, comparing the results from both versions. In the BSSE case, the stations' altitudes are derived from a digital elevation model of the area (Thessaly and neighboring areas) with 90 m resolution at the equator (SRTM; [43]), which is also used for the estimation of the average maximum rainfall at each point in space.

The parameters deriving from the optimization are mx = 3, my = 5, $\tau_{\lambda x}$ = 0.550, $\tau_{\lambda y}$ = 0.005 for the BSS model and mx = 4, my = 12, $\tau_{\lambda x}$ = 0.068, $\tau_{\lambda y}$ = 0.001, $\tau_{\mu x}$ = 0.621, $\tau_{\mu y}$ = 0.451 for the BSSE model. In order to compare the model fits we compute RMSE

and MAE (a) for the fit of both models using all the data and (b) for the fit applying the leave-one-out cross validation method. Results are shown in Table 1.

Table 1. Results from the BSS and BSSE models fitting.

Spatial Smoothing Model	BSS	BSSE
Parameters	mx= 3, my = 5, $\tau_{\lambda x}$= 0.550, $\tau_{\lambda y}$= 0.005	mx= 4, my= 12, $\tau_{\lambda x}$= 0.068, $\tau_{\lambda y}$= 0.001, $\tau_{\mu x}$= 0.621, $\tau_{\mu y}$= 0.451
A. All data		
RMSE (mm)	9.0	7.0
MAE (mm)	7.3	5.3
B. Leave-one-out cross validation		
RMSE (mm)	12.5	10.8
MAE (mm)	9.8	8.1

It is seen that the BSSE model involving elevation as an additional explanatory variable is found superior in both comparisons, which is expected by hydrological experience in the area and also supported by previous applications for annual rainfall in Central Greece [19]. Accordingly, the BSSE model is applied to estimate the 24 h average annual maxima in the center points of a 2 × 2 km grid of the study region, as shown in Figure 4. It is observed that both the Thessaly plain and the surrounding mountain ranges are strongly identified in the resulting rainfall patterns.
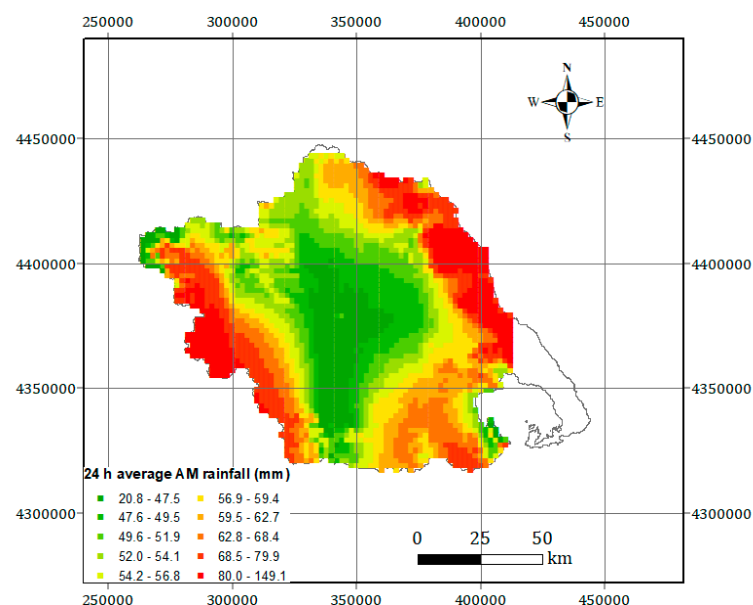
Figure 4. Spatial estimation of the 24 h average annual maximum (AM) rainfall (mm) by the BSSE model.

4.2. Construction of the Regional Ombrian Curves

To estimate a common value of the timescale parameters that would be representative of all stations, we optimize the fit for all stations by simultaneously minimizing the sum of all 14 variances as estimated from each station by Equation (17). This optimization results to parameters α = 0.03 and η = 0.64 which are considered representative for the whole region.

To estimate the distribution parameters, we first divide each annual maxima value at the 24 h scale by its modeled mean value as estimated by the BSSE model (Figure 4). To take space dependence into account, as explained in Section 2.5, we compute the spatial

correlation of the 55 standardized annual maxima series at the 24 h scale. This is estimated to be $\rho = 0.17$ corresponding to $H = 0.61$ (Equation (27))—a moderate value. The standardized series are then unified into one for the estimation of the $b(T)$ parameters.

As explained above, we determine n_1 as the mean record length of the stations ($n_1 = 42$), following the rationale outlined in Section 2.5. We recall that n_1 is equal to the maximum moment order, which need not be modified for spatial dependence bias. Higher moment orders (corresponding to higher return periods) are adapted for spatial dependence. In the approach we follow herein, we choose to fit the model only up to the moment order not impacted by dependence bias. This set of 42 moments is to be used as a calibration set, while we also use the remaining higher 2305 moments as a validation set (Figure 5).
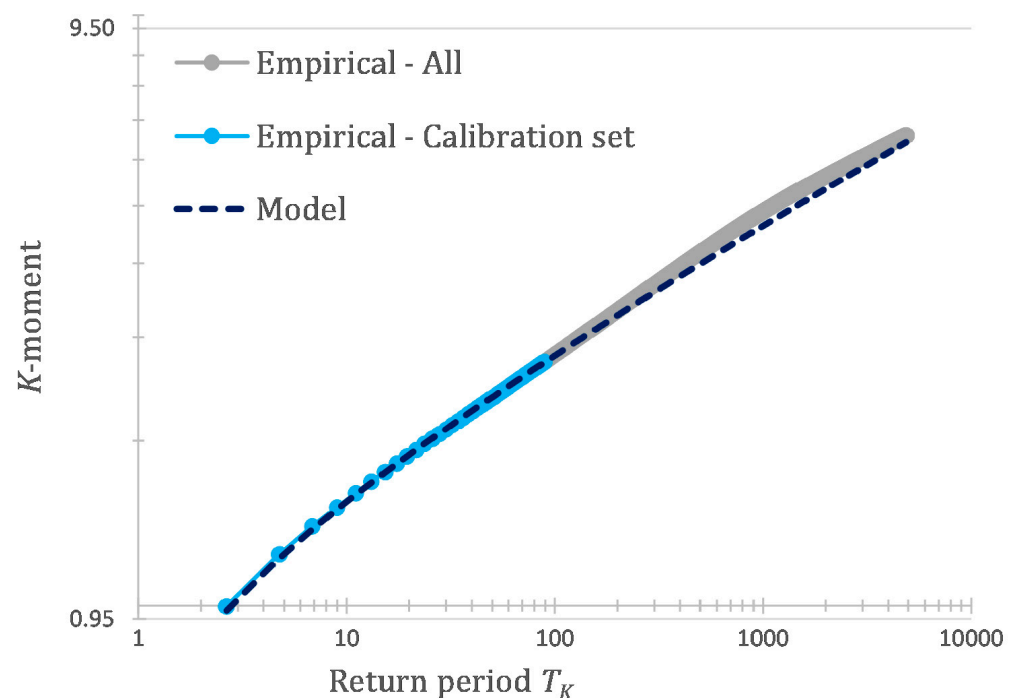
Figure 5. Empirical K-moments of the pooled standardized 24 h rainfall annual maxima sample, their theoretical values by the EV2 distribution and the corresponding return periods.

Using the method of the non-central K-moments, we fit the EV2 distribution to the unified sample of all standardized annual maxima at the 24 h scale minimizing the MAE between the empirical first 42 K-moments and the respective quantiles of the EV2 distribution. In Figure 5, it is seen that the fit is excellent for all 42 K-moments (MAE = 0.00489, RMSE = 0.00471), which constitute the calibration set, and there is also good agreement between the theoretical and empirical moments for higher orders, albeit some deviations in the area of higher return periods. Still, we have to note that very high orders are impacted by the spatial dependence structure of the data, whose estimation is in turn impacted by higher uncertainty. In any case, the use of higher return periods as a validation set proves that the fitting is robust.

Taking into account the results of the BSSE model and following the parameter transformations described in Section 2.5, the values of the four common parameters are derived as shown in Table 2, whereas the spatial distribution of the regionally varying λ parameter is shown in Figure 6. Note that the λ values are analogous to the average maxima values predicted by the BSSE model (Figure 4), as implied by Equation (30).

Table 2. Ombrian parameters α, η, ξ, β of Equations (6) and (12). Parameter λ is analytically derived at any point in space through the BSSE model.

α (h)	η (-)	ξ (-)	β (years)	λ (mm/h)
0.03	0.64	0.18	0.013	Geographic distribution shown in Figure 6 for all grid points.
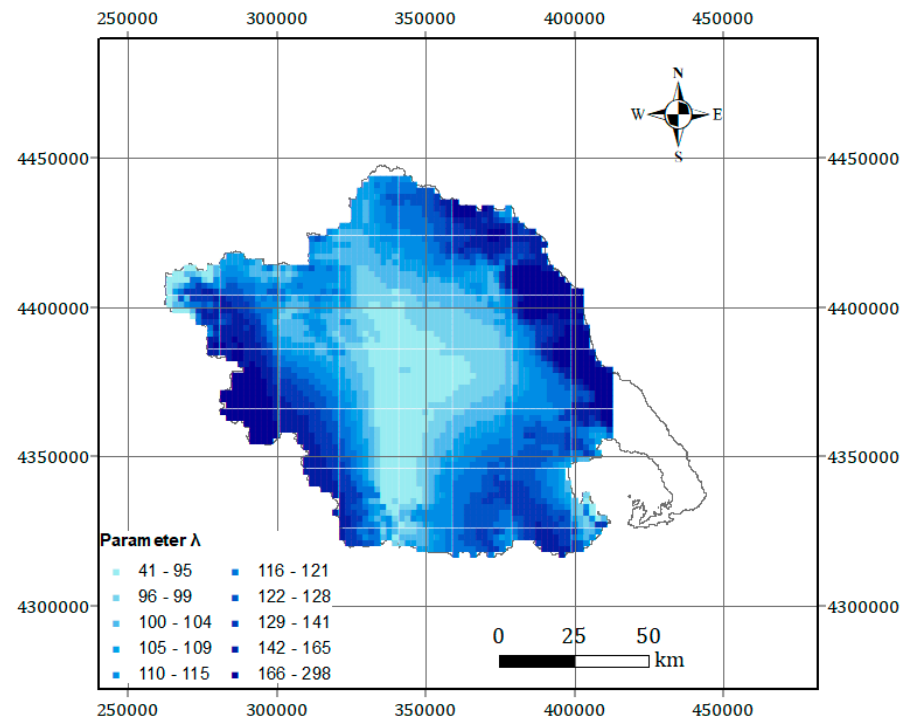
Figure 6. Spatial estimation of the scale parameter λ (mm/h).

4.3. At-Site Verification

To ensure that the spatial model is in agreement with the at-site empirical behavior, we compare the theoretical to the empirical curves as derived for various stations in characteristic locations of the WD. In Figure 7, we show the theoretical quantiles (as given by Equation (12)) of the rainfall intensities for the daily raingauges at six locations representative of the different rainfall regimes of the area, while in Figure 8, the same plots are shown for a wider range of scales that are available from the sub-daily rainfall stations. For the return period estimation of the empirical intensities, we also plot the estimates from the order statistics by Equation (26), in addition to the K-moments. Figures 7 and 8 show that the empirical distribution functions are generally in good agreement with the theoretical ones, with some notable yet non-systematic deviations in the area of higher return periods. The presence of measurement uncertainty is also evident in certain deviations, in the area of higher return periods, between the empirical intensities estimated from the daily raingauges and the sub-daily resolution gauges (Figure 8). The latter are, however, of shorter length compared to the daily gauges.

Taking into account the large sampling variability of rainfall and the spatial extent of the area, the results are deemed acceptable. Yet there are also a few cases in which the model does not capture well the single stations' behavior due to spatial differences between neighboring sites. Such an example is shown in Figure 9 for two stations in the Pertouli area, where it becomes apparent that the approach favors modeling the spatially average behavior between the two stations. It is obvious that in such cases of spatial uncertainty, a model perfectly capturing single stations' behaviors becomes less relevant; rather, the importance of robustness in the regional framework is highlighted.
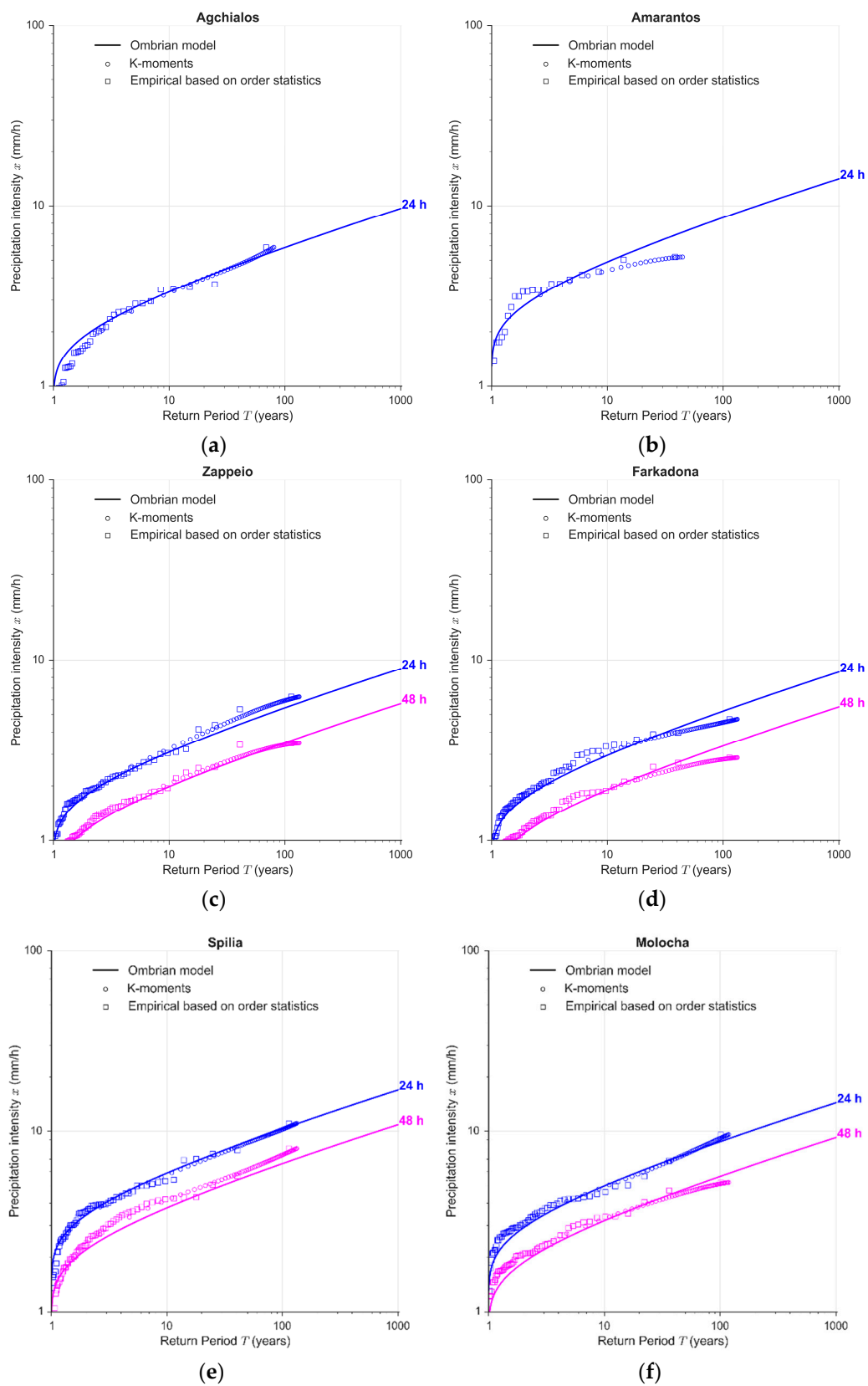
Figure 7. Theoretical and empirical distributions of 24 h and 48 h annual maximum intensities (depending on the available samples) in characteristic stations of Thessaly's WD: (**a**) Agchialos, (**b**) Amarantos, (**c**) Zappeio, (**d**) Farkadona, (**e**) Spilia and (**f**) Molocha. The empirical intensities plotted based on order statistics are also shown for validation.
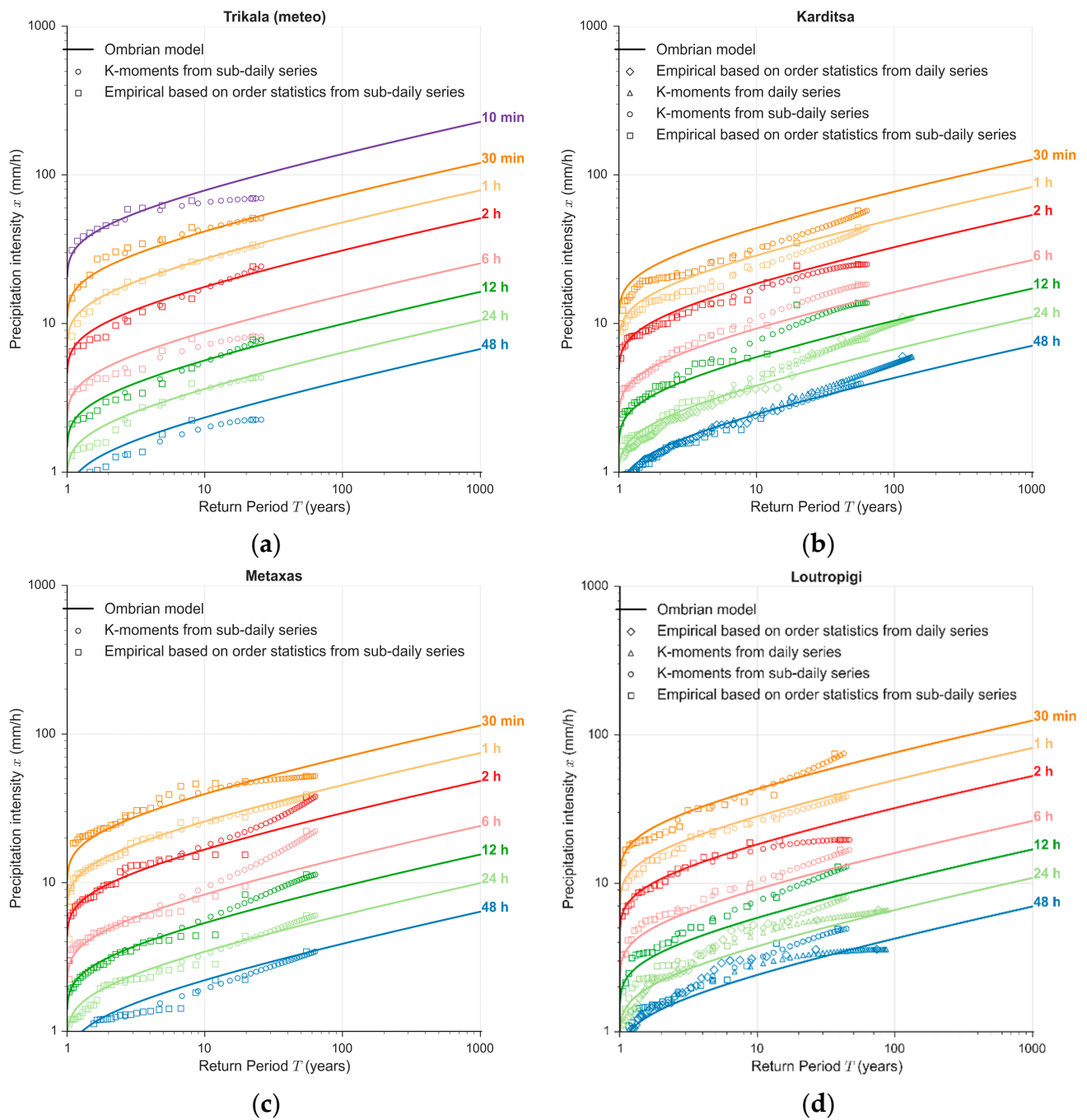
Figure 8. Theoretical and empirical distributions of annual maximum intensities at 10 min to 48 h scales (depending on the available samples) from sub-daily stations of Thessaly's W.D.: (**a**) Trikala (meteo), (**b**) Karditsa, (**c**) Metaxas and (**d**) Loutropigi. When available the empirical intensities at the 24 h and 48 h scales from the daily raingauges are shown as well. The empirical intensities plotted based on order statistics are also shown for validation.
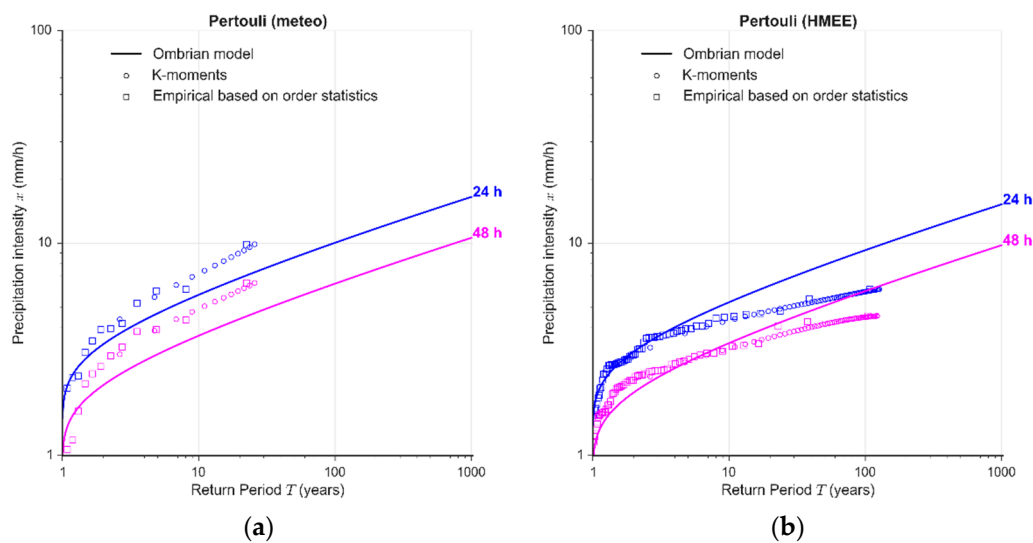
Figure 9. Example of the model fitting to two neighboring stations (from (**a**) meteo and (**b**) HMEE) at the Pertouli area with significant deviations in the stations' recordings.

5. Discussion and Conclusions

Ombrian curves have been around in hydrological engineering for approximately a century and are considered a standard task for at-site modeling. In this work, we address the more complex problem of constructing ombrian curves at regional scales that are of practical interest to the hydrologist in regional flood studies and in the common case that single-site data are either not available for the catchment of interest or the catchment is too large to characterize based on single-site data, e.g., [51]. Even more, constructing the curves by regional fitting is a powerful approach to limit estimation/sampling uncertainty resulting from short-length single records. As such, regional rainfall modeling has been an active research field in hydrological literature. The approach devised herein aims to create a framework for regional ombrian curves that incorporates recent advances in the field of regional frequency analysis and regionalization approaches within a theoretically consistent formulation of ombrian curves. The approach is tested in a challenging case study of the Thessaly WD in Greece, which shows a large variability of rainfall patterns stemming from its complex topography and large extent (~13,700 km^2).

The curves are constructed following the newly revisited mathematical formulation of single-site curves by Koutsoyiannis [2] coupled with a new regionalization approach that is developed herein. Four common parameters are identified, and one spatially varying scale parameter is employed. The site-specific scale parameter is explicitly modeled by a spatial smoothing model (BSSE) that employs elevation as an additional explanatory variable and produces a continuous 2D surface for the average 24 h annual maxima regime. This 2D surface suffices to model the spatial heterogeneity of the curves without involving cluster analysis for delineation of homogenous regions and avoiding related discontinuities and abrupt changes in the parameter space. The result is a model explicitly describing maximum rainfall at any point in a given space, which simplifies hydrological design. The BSSE model is selected for being more interpretable and less involved in parametric choices than common geostatistical software, while its robustness in cases of high spatial uncertainty has been documented [15]. We note that a map of the average maximum regime could also be used instead if already available. Still, the explicit incorporation of a surface smoothing model into the framework guarantees the consistency of the final spatial estimates to the point data.

The approach is also based on recent advances in the analysis of extremes, namely the use of reliable high-order moment estimators that account for the effect of the spatial dependence structure in assigning return periods (K-moments; [2]). This enables a rigorous fitting procedure aiming at minimizing the estimation uncertainty while respecting spatial

dependence. To our knowledge, this is the first time that such a high number of moments (42 for estimation and 2305 for validation) is estimated from a sample of spatially correlated data with a provision for space dependence bias.

Results show that the model efficiently captures the spatial variability of extreme rainfall in the area covering scales from 5 min to 48 h, and its estimates are robust even under increased spatial uncertainty due to inconsistencies among the point data, which were present in a few cases.

A few modifications to the present approach would be required if one were to generalize the model over greater time scales, above the order of a few days [2]. While this task is not of direct use to flood estimation and typical applications, it is to be considered in view of a multi-purpose rainfall model at the regional scale. Further research is also required on the effect that spatial dependence exerts on the estimation of high return periods and the accompanying uncertainty bounds. This task is demanding as results are expected to depend on the assumed type and magnitude of the spatial dependence structure. Yet this research represents a first step toward significantly increasing the number of moments that can be justifiably employed in regional analyses of extremes.

References

1. Lesher, J.H. Saphêneia in Aristotle: "Clarity", "Precision", and "Knowledge". *Apeiron* **2010**, *43*, 143–156. [CrossRef]
2. Koutsoyiannis, D. *Stochastics of Hydroclimatic Extremes—A Cool Look at Risk*; Kallipos: Athens, Greece, 2021; 333p, ISBN 978-618-85370-0-2. Available online: https://repository.kallipos.gr/handle/11419/6522 (accessed on 1 April 2021).
3. Koutsoyiannis, D.; Iliopoulou, T. Ombrian curves advanced to stochastic modelling of rainfall intensity. In *Rainfall: Modeling, Measurement and Applications*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 261–284. [CrossRef]
4. Iliopoulou, T.; Koutsoyiannis, D. PythOm: A python toolbox implementing recent advances in rainfall intensity (ombrian) curves. *Eur. Geosci. Union Gen. Assem.* **2021**, EGU21-389. [CrossRef]
5. Papalexiou, S.M.; Koutsoyiannis, D. Ombrian curves in a maximum entropy framework. *Eur. Geosci. Union Gen. Assem.* **2008**, *10*, 00702.
6. Sherman, C.W. Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Trans. Am. Soc. Civ. Eng.* **1931**, *95*, 951–960. [CrossRef]
7. Bernard, M.M. Formulas for rainfall intensities of long duration. *Trans. Am. Soc. Civ. Eng.* **1932**, *96*, 592–606. [CrossRef]
8. Koutsoyiannis, D.; Kozonis, D.; Manetas, A. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *J. Hydrol.* **1998**, *206*, 118–135. [CrossRef]
9. Veneziano, D.; Furcolo, P. Multifractality of rainfall and scaling of intensity-duration-frequency curves. *Water Resour. Res.* **2002**, *38*, 42-1–42-12. [CrossRef]
10. Langousis, A.; Veneziano, D. Intensity-duration-frequency curves from scaling representations of rainfall. *Water Resour. Res.* **2007**, *43*, 1306. [CrossRef]
11. Svensson, C.; Jones, D.A. Review of rainfall frequency estimation methods. *J. Flood Risk Manag.* **2010**, *3*, 296–313. [CrossRef]
12. Hershfield, D.M. *Rainfall Frequency Atlas of the United States for Durations From 30 Minutes to 24 Hours and Return Periods From l to 100 Years*; US Weather Bureau Technical Paper 40; U.S. Weather Bureau: Washington, DC, USA, 1961; Volume 40, pp. 1–61.
13. Hogg, W.D.; Carr, D.A.; Routledge, B. *Rainfall Intensity-Duration Frequency Values for Canadian Locations*; Environment Canada, Atmospheric Environment Service: Downsview, ON, Canada, 1985.
14. Borga, M.; Vezzani, C.; Dalla Fontana, G. Regional rainfall depth–duration–frequency equations for an alpine region. *Nat. Hazards* **2005**, *36*, 221–235. [CrossRef]
15. Malamos, N.; Koutsoyiannis, D. Field survey and modelling of irrigation water quality indices in a Mediterranean island catchment: A comparison between spatial interpolation methods. *Hydrol. Sci. J.* **2018**, *63*, 1447–1467. [CrossRef]
16. Oliver, M.A.; Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* **2014**, *113*, 56–69. [CrossRef]
17. Li, J.; Heap, A.D. Spatial interpolation methods applied in the environmental sciences: A review. *Environ. Model. Softw.* **2014**, *53*, 173–189. [CrossRef]
18. Malamos, N.; Koutsoyiannis, D. Bilinear surface smoothing for spatial interpolation with optional incorporation of an explanatory variable. Part 1: Theory. *Hydrol. Sci. J.* **2016**, *61*, 519–526. [CrossRef]
19. Malamos, N.; Koutsoyiannis, D. Bilinear surface smoothing for spatial interpolation with optional incorporation of an explanatory variable. Part 2: Application to synthesized and rainfall data. *Hydrol. Sci. J.* **2016**, *61*, 527–540. [CrossRef]
20. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on L-Moments*; Cambridge University Press: Cambridge, UK, 1997. [CrossRef]
21. Hosking, J.R. L-moments: Analysis and estimation of distributions using linear combinations of order statistics, J.R. *Stat. Soc. Ser. B Stat. Methodol.* **1990**, *52*, 105–124. [CrossRef]
22. Hosking, J.R.M.; Wallis, J.R. The effect of intersite dependence on regional flood frequency analysis. *Water Resour. Res.* **1988**, *24*, 588–600. [CrossRef]
23. Koutsoyiannis, D. Knowable moments for high-order stochastic characterization and modelling of hydrological processes. *Hydrol. Sci. J.* **2019**, *64*, 19–33. [CrossRef]
24. Weinmann, P.E.; Nandakumar, N.; Siriwardena, L.; Mein, R.G.; Nathan, R.J. Estimation of rare design rainfalls for Victoria using the CRC-FORGE methodology. In *Water 99: Joint Congress; 25th Hydrology & Water Resources Symposium, 2nd International Conference on Water Resources & Environment Research; Handbook and Proceedings*; Institution of Engineers: Barton, Australia, 1999; p. 284.
25. Claps, P.; Ganora, D.; Mazzoglio, P. Rainfall regionalization techniques. In *Rainfall: Modeling, Measurement and Applications*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 327–350. [CrossRef]
26. Hailegeorgis, T.T.; Thorolfsson, S.T.; Alfredsen, K. Regional frequency analysis of extreme precipitation with consideration of uncertainties to update IDF curves for the city of Trondheim. *J. Hydrol.* **2013**, *498*, 305–318. [CrossRef]
27. Aron, G.; Wall, D.J.; White, E.L.; Dunn, C.N. Regional rainfall intensity-duration-frequency curves for Pennsylvania 1. *JAWRA J. Am. Water Resour. Assoc.* **1987**, *23*, 479–485. [CrossRef]
28. Trefry, C.M.; Watkins, D.W., Jr.; Johnson, D. Regional rainfall frequency analysis for the state of Michigan. *J. Hydrol. Eng.* **2005**, *10*, 437–449. [CrossRef]
29. Dalrymple, T. *Flood-Frequency Analyses. Manual of Hydrology: Part 3. Flood-Flow Techniques*; USGPO: Washington, DC, USA, 1960; Volume 80. [CrossRef]
30. Burn, D.H. An appraisal of the "region of influence" approach to flood frequency analysis. *Hydrol. Sci. J.* **1990**, *35*, 149–165. [CrossRef]
31. Faulkner, D.S.; Prudhomme, C. Mapping an index of extreme rainfall across the UK. *Hydrol. Earth Syst. Sci.* **1998**, *2*, 183–194. [CrossRef]

32. Deidda, R.; Hellies, M.; Langousis, A. A critical analysis of the shortcomings in spatial frequency analysis of rainfall extremes based on homogeneous regions and a comparison with a hierarchical boundaryless approach. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2605–2628. [CrossRef]

33. Dimitriadis, P.; Koutsoyiannis, D.; Iliopoulou, T.; Papanicolaou, P. A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes. *Hydrology* **2021**, *8*, 59. [CrossRef]

34. Glynis, K.G.; Iliopoulou, T.; Dimitriadis, P.; Koutsoyiannis, D. Stochastic investigation of daily air temperature extremes from a global ground station network. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1585–1603. [CrossRef]

35. Koutsoyiannis, D.; Dimitriadis, P. Towards generic simulation for demanding stochastic processes. *Sci* **2021**, *3*, 34. [CrossRef]

36. Koutsoyiannis, D. An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence. *Water Resour. Res.* **2006**, *42*, W01401. [CrossRef]

37. Koutsoyiannis, D. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrol. Sci. J.* **2004**, *49*, 591–610. [CrossRef]

38. Koutsoyiannis, D.; Papalexiou, S.M. Extreme rainfall: Global perspective. In *Handbook of Applied Hydrology*; McGraw-Hill: New York, NY, USA, 2017; pp. 74.1–74.16.

39. Iliopoulou, T.; Koutsoyiannis, D.; Montanari, A. Characterizing and modeling seasonality in extreme rainfall. *Water Resour. Res.* **2018**, *54*, 6242–6258. [CrossRef]

40. Koutsoyiannis, D. Broken line smoothing: A simple method for interpolating and smoothing data series. *Environ. Model. Softw.* **2000**, *15*, 139–149. [CrossRef]

41. Malamos, N.; Koutsoyiannis, D. Broken line smoothing for data series interpolation by incorporating an explanatory variable with denser observations: Application to soil-water and rainfall data. *Hydrol. Sci. J.* **2015**, *60*, 468–481. [CrossRef]

42. Wahba, G.; Wendelberger, J. Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation. *Mon. Weather. Rev.* **1980**, *108*, 1122–1143. [CrossRef]

43. Jarvis, A.; Reuter, H.I.; Nelson, A.; Guevara, E. Hole-Filled SRTM for the Globe Version 4. 2008. Available online: http://srtm.csi.cgiar.org (accessed on 27 November 2021).

44. Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C.; Wallis, J.R. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resour. Res.* **1979**, *15*, 1049–1054. [CrossRef]

45. Mimikou, M.; Koutsoyiannis, D. Extreme floods in Greece: The case of 1994. In Proceedings of the US-ITALY Research Workshop on the Hydrometeorology, Impacts, and Management of Extreme Floods, Perugia, Italy, 13–17 November 1995.

46. Bathrellos, G.D.; Skilodimou, H.D.; Soukis, K.; Koskeridou, E. Temporal and spatial analysis of flood occurrences in the drainage basin of pinios river (thessaly, central greece). *Land* **2018**, *7*, 106. [CrossRef]

47. Koutsoyiannis, D.; Mamassis, N.; Efstratiadis, A.; Zarkadoulas, N.; Markonis, Y. *Floods in Greece, in Changes of Flood Risk in Europe*; Wallingford—International Association of Hydrological Sciences; IAHS Press: Wallingford, UK, 2012; pp. 238–256.

48. Loukas, A.; Vasiliades, L. Probabilistic analysis of drought spatiotemporal characteristics in Thessaly region, Greece. *Nat. Hazards Earth Syst. Sci.* **2004**, *4*, 719–731. [CrossRef]

49. Lagouvardos, K.; Kotroni, V.; Bezes, A.; Koletsis, I.; Kopania, T.; Lykoudis, S.; Mazarakis, N.; Papagiannaki, K.; Vougioukas, S. The automatic weather stations NOANN network of the National Observatory of Athens: Operation and database. *Geosci. Data J.* **2017**, *4*, 4–16. [CrossRef]

50. Hersfield, D.M.; Wilson, W.T. Generalizing of rainfall-intensity-frequency data. *AIHS. Gen. Ass. Tor.* **1957**, *1*, 499–506.

51. Pasculli, A.; Longo, R.; Sciarra, N.; Di Nucci, C. Surface Water Flow Balance of a River Basin Using a Shallow Water Approach and GPU Parallel Computing; Pescara River (Italy) as Test Case. *Water* **2022**, *14*, 234. [CrossRef]

52. Koutsoyiannis, D. An open letter to the Editor of Frontiers. *Researchgate* **2021**. [CrossRef]