

Volume 65 | Issues 1-4 | 2020

ISSN 0262-6667

Hydrological Sciences **JOURNAL**



International Association of
Hydrological Sciences



**A stochastic framework for rainfall intensity-timescale-
return period relationships. Part II: Point modelling and**

regionalization over Greece

Journal:	<i>Hydrological Sciences Journal</i>
Manuscript ID	HSJ-2023-0524.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	16-Mar-2024
Complete List of Authors:	Iliopoulou, Theano; National Technical University of Athens, Water Resources and Environmental Engineering Koutsoyiannis, Demetris; National Technical University of Athens, Malamos, Nikolaos; University of Patras, Department of Agriculture Koukouvinos, Antonis; National Technical University of Athens Dimitriadis, Panayiotis; National Technical University of Athens, Water Resources and Environmental Engineering; National Observatory of Athens, Mamassis, Nikos; NTUA, Tepetidis, Nikos; National Technical University of Athens, Water Resources and Environmental Engineering Markantonis, David; National Technical University of Athens
Keywords:	

SCHOLARONE™
Manuscripts

A stochastic framework for rainfall intensity–timescale–return period relationships. Part II: Point modelling and regionalization over Greece

^{1*}Theano Iliopoulou, ¹Demetris Koutsoyiannis, ²Nikolaos Malamos, ¹Antonis Koukouvinos,
¹Panayiotis Dimitriadis, ¹Nikos Mamassis, ¹Nikos Tepetidis, ¹David Markantonis

¹Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechniou 5, GR-157 72 Zographou, Greece

²Department of Agriculture, University of Patras, GR-30200 Messolonghi, Greece; nmalamos@upatras.gr

* Corresponding author. *E-mail address:* tiliopoulou@hydro.ntua.gr

Abstract

In this work, we formulate a regionalization framework for rainfall intensity–timescale–return period relationships which is applied over the Greek territory. The methodology for single-site estimation is based on a stochastic framework for multi-scale modelling of rainfall intensity which is outlined in the companion paper. Five parameters are first fitted independently for each site and the resulting parameter variability is assessed. Following a systematic investigation of uncertainty and variability patterns, two parameters, i.e., the tail-index and a timescale parameter, are identified as constant in space and estimated using data pooling techniques. The other three parameters are regionalized over Greece by means of spatial interpolation and smoothing techniques that are assessed through cross-validation in a multi-model framework. The regionalization scheme is implemented in a sequential order that allows exploiting rainfall information both from rainfall stations with sub-daily resolution as well as from the higher reliability network of daily raingauges.

Keywords: intensity–duration–frequency curves; rainfall regionalization; extreme rainfall; spatial interpolation; spatial smoothing; design rainfall; geostatistical modelling

1. Introduction

One of the most essential requirements for designing and operating water-related infrastructure, such as dams, bridges, water conveyance structures and urban drainage systems, is the ability to obtain reliable rainfall estimates for different timescales and return periods. This information is usually accessible to engineers in the form of mathematical relationships linking the rainfall intensity (or depth) to the reference timescale and the return period, often misnamed as ‘duration’ and ‘frequency’ respectively. In this work, we use the term ‘ombrian relationships’ (or curves), as in the companion paper by Koutsoyiannis et al. (2023a). The identification of these relationships is based on some form of probabilistic analysis of rainfall, typically performed using extreme value distributions, coupled with a model preserving the important scaling properties of rainfall to retain consistency of the model estimates across the different timescales. There are numerous works in the hydrological literature dealing with the construction of such curves (Svensson and Jones 2010, Lanciotti et al. 2023) starting from early empirical approaches (Sherman 1931, Bernard 1932) to generalized parametric ones, as the probabilistic approach proposed by Koutsoyiannis et al. (1998) and the approach based on the rainfall simple- or multi-scaling assumption (Burlando and Rosso 1996), as well as other regression-type (‘data-driven’) approaches (Overeem et al. 2008, Haruna et al. 2023). A discussion of existing theoretical concepts for point modelling is provided in the companion paper by Koutsoyiannis et al. (2023a).

Aside from the derivation of reliable point estimates, a critical requirement for design rainfall is the ability to generalize this information over large scales, also covering the ungauged regions, a process known as ‘regionalization’. Several methods have been proposed in the literature for regionalization of design rainfall with degrees of complexity varying usually in relation to the spatial scale of interest. For instance, small regions are often treated as homogenous and common

relationships are identified through data pooling approaches (e.g., Hailegeorgis et al. 2013). As the spatial scale and the degree of spatial variability increases for larger regions, it is a common hydrological practice to identify clusters or sub-regions based on measures of hydrological similarity, and subsequently apply different regional distributions (see e.g. Aron et al. 1987, Trefry et al. 2006, Burn 2014, Forestieri et al. 2018, Iliopoulou and Koutsoyiannis 2022; and national-scale applications described in Svensson and Jones, 2010); a type of spatial approach which is often combined with the so-called ‘index-method’ and the L-moments regional framework (Hosking and Wallis 1997). In this respect, various cluster analysis techniques of different algorithmic complexity have been proposed regarding the possibility to consider some parameter(s) of extreme value distributions as invariant or not in the study area (e.g., Lin and Chen 2006, Rao and Srinivas 2008, Cassalho et al. 2019, De Luca and Napolitano 2023). For larger regions and more complicated rainfall regimes, explicit spatial modelling may be employed for one or more of the involved parameters (e.g., Iliopoulou et al. 2022, Shehu et al. 2023, Deidda et al. 2021, Faulkner and Prudhomme 1998, Perica et al. 2009). The latter approach exploits the advances in geographical information systems and regionalization techniques and allows producing estimates of the required parameters at any point over a given grid. In this way, several limitations of traditional approaches are resolved, including the difficulty of identifying homogenous regions and the occurrence of inconsistencies at the sub-regions’ borders. Typically, the employed spatial modelling methods include regression analysis (e.g., Ulrich et al. 2020, Faulkner and Prudhomme 1998, Madsen et al. 2017), the Inverse Distance Weighing (IDW) method (Barbosa et al. 2022, Ahrens 2006, Perica et al. 2009, Szolgay et al. 2009, Mascaro 2020), the kriging method and variations thereof (Watkins et al. 2005, Forestieri et al. 2018, Ceresetti et al. 2012, Blanchet et al. 2016, Libertino et al. 2018, Shehu et al. 2023, Szolgay et al. 2009), bilinear

1
2
3 surface smoothing methods (Malamos and Koutsoyiannis 2016b, Iliopoulou et al. 2022),
4
5 smoothing splines (Hutchinson 1995), neural networks (Ceresetti et al. 2012), Bayesian
6
7 hierarchical models (Jalbert et al. 2022) and nearest neighbors (e.g. Szolgay et al. 2009), among
8
9 others (see also Claps et al. 2022 for an extended review).

10
11
12 Despite the existence of many interpolation and geostatistical methods, there are various
13
14 open methodological questions regarding the implementation of the regionalization scheme. For
15
16 instance, the most straightforward and widely used procedure is the exact interpolation of
17
18 independently estimated local parameters (e.g., Blanchet et al. 2016, Mascaro 2022, Barbosa et al.
19
20 2022, Szolgay et al. 2009). Despite being simple, this approach implemented through an
21
22 uncontrolled interpolation might propagate measurement errors in the final product and give rise
23
24 to unrealistic spatial patterns and abrupt changes (Uboldi et al. 2014, Jalbert et al. 2022).
25
26 Independent estimation also cannot benefit from the opportunity to increase the reliability of
27
28 uncertain parameters exploiting information from more records; an issue that is particularly
29
30 relevant for the highly uncertain shape parameter (tail-index) of the extreme value distribution. On
31
32 the other hand, for the model to reflect the extreme rainfall climatology and retain the physically
33
34 based rainfall structure in space, it is necessary to identify the robust local ('at-site') properties
35
36 which should be regionalized. Therefore, the preliminary assessment of parameter uncertainty and
37
38 spatial patterns as well as prior knowledge of the region's climatology are prerequisites for the
39
40 regionalization.
41
42
43
44
45

46
47 Data availability in the region of interest also determines the implementation of the
48
49 regionalization scheme. As pointed out by Shehu et al. (2022), the requirements for reliable
50
51 regionalization, i.e., good data density, fine temporal resolution, and long record length, are seldom
52
53 satisfied together in any given dataset. The same holds true for the rainfall network in Greece, in
54
55
56
57

1
2
3 which the uncertainty due to limited availability of sub-daily data is exacerbated by the complex
4 topography of the mainland and the numerous distanced islands. To compensate for the short
5 length of sub-daily records and their limited spatial coverage in Greece, Koutsoyiannis et al. (1998)
6 proposed to estimate the distribution parameters employing rainfall maxima from the longer daily
7 raingauges. To augment the useful rainfall information, other works suggest the use of
8 disaggregation/rainfall scaling properties to obtain sub-daily estimates from existing daily series
9 (Shehu et al. 2022, Bara et al. 2009, Courty et al. 2019), utilization of covariates available at a
10 denser spatial network (Madsen et al. 2017), typically elevation (Goovaerts 2000), pooling
11 approaches (e.g. Burn 2014, Iliopoulou et al. 2022, Iliopoulou and Koutsoyiannis 2022), spatial
12 resampling techniques (Uboldi et al. 2014), spatial interpolation of the actual rainfall series
13 (Libertino et al. 2018), while recently the use of radar rainfall, reanalysis and satellite data has
14 been proposed as a means to improve the spatial representation of the rainfall regime (e.g. Ombadi
15 et al. 2018, Courty et al. 2019, Lanciotti et al. 2023). Therefore, regardless of the specifications of
16 each modelling framework, common choices to be made are the parameters to be regionalized as
17 well as the type of data and the order of their utilization in the regionalization scheme.

18
19 In this paper, we revisit the existing design rainfall curves for Greece, which are estimated
20 on a point basis, and develop for the first time a regional model valid over the entire Greek territory.
21 By revisiting the existing relationships, we intend to benefit from newer data with better temporal
22 and spatial coverage over Greece and explore advanced methodological frameworks both for
23 point-modelling and regionalization. Namely, the point modelling methodology has been revisited
24 by Koutsoyiannis (2023) under a stochastic framework, providing better physical meaning for the
25 parameters, theoretical consistency, and links to all-scale stochastic modelling (Koutsoyiannis and
26 Iliopoulou 2022). The framework can also be coupled with a new approach for reliable high-order

1
2
3 quantile estimation (K-moments, Koutsoyiannis 2019). The theoretical background for the point
4 modelling is presented in detail in the companion paper by Koutsoyiannis et al. (2023a). The
5 specific focus of this work is the formulation of the regionalization approach informed by the
6 stochastic behavior of extreme rainfall, practical data limitations and uncertainty consideration, as
7 well as empirical insights from the point investigation of the parameters. The latter forms the basis
8 of the regional investigation. This is the culmination of a national-scale effort to produce a gridded
9 product for design rainfall parameters all over Greece for the first time (Koutsoyiannis et al.
10 2023b).

11
12 The study is structured as follows. Rainfall data sources and their analysis are described in
13 Section 2. The employed methods are presented in Section 3; the stochastic methodology for the
14 construction of rainfall intensity–timescale and return period relationships is briefly outlined in
15 Section 3.1, while the regionalization models and their implementation are presented in Section
16 3.2; at-site performance metrics are presented in Section 3.3. Results are presented in Section 4
17 together with the final product, the performance assessment, and the design rainfall maps.
18 Discussion is provided in Section 5 and Conclusions are outlined in Section 6. Results from the
19 point investigation as well as complementary findings from the regionalization are presented in
20 Appendices A and B, respectively.

2. Data

2.1 Study area

21
22 The study area is the Greek territory, comprising a complex topography with an area of 131 957
23 km² including thousands of islands (**Figure 12**). Greek climate is generally characterized as
24 Mediterranean, yet it shows marked spatial diversity in terms of its rainfall regime, with average
25 daily precipitation ranging geographically within an order of magnitude, from 0.6 mm/d (219
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 mm/year) to 7.3 mm/d (2666 mm/year) (Koutsoyiannis et al. 2023c). The wettest part is
4
5 northwestern Greece, while the driest parts include Athens, some of the Aegean islands, and parts
6
7 of central Macedonia and Thessaly.
8
9

10 **2.2 Data processing and quality control**

11
12
13 The construction of rainfall intensity-timescale-return period relationships requires rainfall data at
14
15 a range of timescales, from fine scales, i.e., 5 to 60 min up to the 24 or 48 h scales. Such data are
16
17 usually not publicly available in Greece but are managed independently by different private and
18
19 public operating services. After laborious national-scale efforts (Koutsoyiannis et al. 2023b), two
20
21 types of ground rainfall data were collected: (a) data from rainfall recorders (tipping bucket rain
22
23 recorders and automatic sensors) having more than 10 years of fine-scale data, and (b) data from
24
25 daily rainfall gauges with more than 15 years of measurements, totalling to an initial dataset of
26
27 940 raingauge records.
28
29
30

31
32 The methodology is based on the analysis of annual maxima series instead of peaks-over-
33
34 threshold or even the complete data series for the estimation of the extremes since many historical
35
36 records are available only in this form. The original series were aggregated at a range of timescales
37
38 k from 5 min to 48 h, with $k = 0.083, 0.167, 0.25, 0.5, 1, 2, 6, 12, 24, 48$ h (depending on data
39
40 availability at the finest scale), and the maximum rainfall depth at each scale $h^{(k)}$ was extracted for
41
42 all hydrological years. Accordingly, the corresponding rainfall intensity at the given scale is
43
44 computed as $x^{(k)} = h^{(k)} / k$, thereby deriving the empirical rainfall intensities corresponding to the
45
46 annual maxima of the hydrological years. It is noted that following the stochastic reasoning
47
48 explained in Koutsoyiannis et. al. (2023a), a fixed, rather than a moving, time window is used to
49
50 extract the maximum for each scale while a correction factor (usually termed the Hershfield
51
52 coefficient) for the annual daily maxima is not used.
53
54
55
56
57
58
59
60

1
2
3 To ensure a high-quality dataset, we undertook extended quality checks starting from basic
4 consistency checks across timescales to ensure that annual maximum rainfall depths are not
5 decreasing with increasing timescale, and respective intensities are not increasing. In addition, we
6 performed hydrological consistency checks, i.e., to ensure that single-site empirical maximum
7 rainfall is consistent with hydrological experience worldwide, suggesting an unbounded right tail
8 of sub-exponential type (e.g., Koutsoyiannis and Papalexiou 2017, Courty et al. 2019). We note
9 that records from poorly maintained daily raingauges (e.g., in remote mountainous areas)
10 sometimes exhibit maximum rainfall recordings of the same (or nearly the same) amount due to
11 spillage effects during storm events. In this case, a bounded GEV distribution might falsely
12 emerge. We also performed preliminary statistical analyses for all recorded maxima values per
13 station in Greece, and we identified extremely high rainfall values, as the ones significantly
14 deviating from all other maxima values in the same region. We excluded the extreme values that
15 were not in agreement with available local rainfall information recorded at the same period from
16 other ground stations, and satellite data from the IMERG product (Integrated Multi-satellitE
17 Retrievals for GPM, Huffman et al. 2019). For instance, in a few cases, neighboring stations and
18 satellite data both recorded rainfall events of low to moderate magnitude, and therefore distinctly
19 high values could be attributed to typographic errors during digitization of the daily records (e.g.,
20 decimal place errors). Finally, we performed spatial consistency checks rejecting stations with
21 systematic deviations from neighbouring ones.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46
47 After the above meticulous quality control processing, from the initial set of 940 stations we
48 compiled a final dataset of 783 stations (Table 1), comprising:
49

- 50 • 503 daily rain gauges, 130 of which at locations where there is also a rain recorder
 - 51 • 280 rain gauges (rain recorders) with sub-daily resolution.
- 52
53
54
55
56
57
58
59
60

1
2
3 The selected stations are distributed over 651 geographical locations (**Figure 12**). The longest
4
5 available record is the daily raingauge station in Athens which covers the period from 1863 to
6
7
8 2022. As shown in
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

Table 2, the daily raingauges are generally characterized by longer lengths compared to the sub-daily ones.

A critical issue in the regionalization analysis is the specification of its spatial resolution, i.e., the size of the cell of which it is composed (grid size resolution). In this respect, it is important that the cell size for the regionalization grid reflects the given spatial information, e.g., the spatial density of known points (Hengl 2006). Taking into account the number and the geographical distribution of the stations, while considering the needs of hydrological design in Greece, a cell size of 5 km × 5 km was chosen for the regionalization analysis.

3. Methods

3.1 Stochastic framework for rainfall intensity–timescale–return period modelling

3.1.1 Ombrian model

The mathematical framework for design rainfall relationships applied herein is a simplified version of the stochastic model for rainfall intensity, valid at any scale supported by the data and termed ‘ombrian model’, as proposed by Koutsoyiannis (2023). Here we apply the framework only for small time scales, of the order of minutes to a few days, for which a Pareto distribution for the non-zero rainfall intensity is justified. In this case, by virtue of some simplifications detailed in the companion paper (Koutsoyiannis et al. 2023a), the rainfall intensity x is linked to the timescale k and return period T via the following relationship which can also be expressed as the ratio of the distribution function $b(T)$ to the timescale function $a(k)$:

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

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 time (conventionally, years to correspond to the standard unit for return period), α a timescale parameter in units of time (conventionally, hours to correspond to the standard unit for timescale) with $\alpha > 0$, η a dimensionless parameter with $0 < \eta < 1$, and $\xi > 0$ the tail index of the process. In the case that the return period is determined based on series of annual maxima (AM) of rainfall intensity where $\Delta = 1$ year, the respective relationship is:

$$x = \frac{b(T)}{a(k)} = \lambda \frac{(-(\beta/\Delta)\ln(1 - \Delta/T))^{-\xi} - 1}{(1 + k/\alpha)^\eta}, \quad \xi > 0 \quad (2)$$

The latter equation is used for calibrating the model, since annual maxima are available, but once the parameters are identified the former and simpler relationship is used for design purposes, following the rationale explained in Koutsoyiannis et al. (2023a).

3.1.2 Parameter estimation procedures

The advantage of the simplified version of the model is the separability of $a(k)$ and $b(T)$ functions that enables a two-step procedure for the parameter estimation. This turns out to be convenient for practical applications utilizing various sources of rainfall data. Namely, the estimation of the parameters of the timescale function (of the expression $a(k)$) is performed using sub-daily or even sub-hourly data, available from tipping-bucket rain recorders and automated sensors. The estimation of the distribution parameters (of the expression $b(T)$) may be performed using in priority (if available at the same location) the daily rainfall records due to (a) the longer record lengths of the daily rain gauge observations compared to those of the sub-daily stations, and (b) the greater reliability of rainfall measurement from daily raingauges during storm events (Molini et al. 2005).

1
2
3 The parameters of the timescale function, α and η , are estimated following the optimization
4 procedure detailed in Koutsoyiannis et al. (2023a), under the assumption that the stochastic
5 variables $y_j := a(k_j)x$ follow the same distribution for the different scales k_j . In order to improve
6 the fit to the higher quantile region, we calibrate the parameters to the highest 1/2 of the data for
7 each timescale.
8
9

10
11
12 The parameters of the distribution function, ξ , λ and β , are also estimated through optimization
13 techniques minimizing the root-mean-square error between the logarithms of the theoretical return
14 periods and the logarithms of the empirical return periods which are assigned through the K-
15 moments framework, according to the methodology outlined in Koutsoyiannis et al. (2023a). In
16 general, we use only a portion of the estimated K-moments (here, the lower orders up to the 20%
17 of the record length, with a minimum of 5 and a maximum of 20 orders) for calibration and employ
18 the higher-order moments for validation. The regional estimation of the tail-index ξ is performed
19 following a separate K-moments procedure that also considers the dependence structure and is
20 detailed later (Section 4.2.1).
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 To achieve robust modeling of the distribution function parameters λ and β in space and avoid
36 boundary effects, it is useful to employ alternative expressions of these parameters. For instance,
37 the parameter β which is lower bounded (here with an assumed minimum value of 0.01 years, see
38 Koutsoyiannis et al. 2023a) and takes values close to the lower bound at many stations, is not a
39 convenient parameter for spatial interpolation as many spatial estimates below the bound are likely
40 to appear. It is preferable to express such a parameter as a function of another characteristic
41 variable, with a smoother spatial distribution, as described below.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Let T_1 and T_2 denote two characteristic return periods, here $T_1 = 2$ years and $T_2 = 100$ years. We define a timescale, e.g., $k = 24$ h and let $b := (1 + k/\alpha)^\eta$ for the given k . Assuming known rainfall intensities x_1 and x_2 for T_1 and T_2 respectively, then from Equation (1) follows:

$$b x_1 = \lambda((T_1/\beta)^\xi - 1), \quad b x_2 = \lambda((T_2/\beta)^\xi - 1), \quad \frac{x_2}{x_1} = \frac{(T_2/\beta)^\xi - 1}{(T_1/\beta)^\xi - 1} \quad (3)$$

Setting $(T_1/\beta)^\xi = :c$ (unknown), $x_2/x_1 = :r_x$ (known), $(T_2/T_1)^\xi = :r_T$ (known), we get:

$$r_x = \frac{r_T c - 1}{c - 1}, \quad c = \frac{r_x - 1}{r_x - r_T}, \quad \beta = \left(\frac{r_x - r_T}{r_x - 1}\right)^{1/\xi} T_1, \quad \lambda = \frac{b x_1}{c - 1} \quad (4)$$

and finally, the parameters β and λ can be expressed as:

$$\beta = \left(\frac{r_x - r_T}{r_x - 1}\right)^{1/\xi} T_1, \quad \lambda = b \frac{r_x - r_T}{r_T - 1} x_1 \quad (5)$$

As the ratio r_T results directly from the choice of the characteristic return periods, parameters β and λ are then derived as functions of the intensity x_1 for the chosen timescale (here, 24 h) and return period ($T_1 = 2$ years) and the intensity ratio r_x . Therefore, the intensity x_1 and the intensity ratio r_x are estimated first to be used in the regionalization and then their grids are converted to the desired parameters grids β and λ via Equation (5). It is noted that the parameter set (x_1, r_x) is preferred to the one of (x_1, x_2) as the former is found uncorrelated to each other and thus the pair's information content is greater, which is beneficial to the regionalization.

3.2 Regionalization

3.2.1 Investigation of point variability

Before proceeding to the regionalization of the five parameters, an independent, at-site estimation thereof is conducted for all locations. The observed variability of the parameters is evaluated in

1
2
3 terms of the existence of random vs systematic geographic variation including the dependence on
4 other geophysical and hydroclimatic variables, such as the elevation and the average daily rainfall.
5
6 Parameters that are characterized by pronounced uncertainty and do not show robust physically
7
8 reasoned patterns are identified and treated as common over the entire area. For the calibration of
9
10 common parameters, we employ appropriate optimization procedures, namely a simultaneous
11
12 combined optimization procedure for the α parameter using multiple high-quality fine-scale
13
14 records, and a pooled estimation of the shape parameter ξ using the longest daily high-quality
15
16 records available and corroborated by stochastic simulations. These procedures along with the
17
18 variability investigation and the respective findings are described in detail in Sections 4.1.1 and
19
20 4.2.1, for the α and ξ parameters, respectively. For each of the three remaining parameters, we
21
22 calibrate a set of spatial models that are evaluated against each other as described below.
23
24
25
26
27
28
29

30 **3.2.2 Spatial models**

31
32 The regional estimation of distribution parameters of the ombrian model has been accomplished
33
34 by Iliopoulou et al. (2023, 2022), by implementing the Bilinear Surface Smoothing (BSS)
35
36 framework (Malamos and Koutsoyiannis, 2016a, 2016b). This incorporates smoothing terms with
37
38 adjustable weights, defined by means of the angles formed by consecutive bilinear surfaces, into
39
40 a piecewise surface regression model with known break points. There are two variants, in terms of
41
42 applying the framework with an explanatory variable, available from measurements in a
43
44 considerably denser dataset than the initial main variable, (BSSE) or not (BSS).
45
46
47

48 In the context of the present study, the BSSE variant was evaluated using the elevation of
49
50 the stations, extracted from the SRTM data (Jarvis et al. 2008), as the additional explanatory
51
52 variable. A brief overview of the mathematical framework of BSS is presented since it is detailed
53
54 in the aforementioned publications. The general idea behind both method variants is to
55
56
57

1
2
3 compromise the trade-off between the objectives of minimizing the fitting error and the roughness
4 of the fitted bilinear surface, therefore termed bilinear surface smoothing (BSS). The larger the
5 weight of the first objective, the rougher the surface will appear, while the opposite is true for a
6 larger weight of the second objective.
7
8
9
10
11

12 The mathematical framework of BSS suggests that fitting is meant in terms of minimizing
13 the generalized cross-validation error (GCV; Wahba and Wendelberger 1980) between the set of
14 the given data points and the corresponding estimates.
15
16
17
18

19 BSSE introduces an additional explanatory variable at a denser dataset compared to that of
20 the main variable, as follows. We assume that at the locations of the given data points, we also
21 know the value of an explanatory variable t , and therefore for each point z there corresponds a
22 value t . In this case, the general estimation function for point u is:
23
24
25
26
27

$$\hat{z}_u = d_u + t_u e_u \quad (6)$$

28 where d_u , e_u are the values of two fitted bilinear surfaces at that point, namely d and e , while t_u is
29 the value of the explanatory variable at that point. This is not a global linear relationship but a local
30 linear one as the quantities d_u and e_u change in space.
31
32
33
34
35
36
37

38 In the case of the BSSE, there are six adjustable parameters: the numbers of intervals along
39 the horizontal and vertical direction, respectively, i.e., m_x , m_y , and the corresponding smoothing
40 parameters $\tau_{\lambda x}$ and $\tau_{\lambda y}$ for surface d along with the smoothing parameters $\tau_{\mu x}$ and $\tau_{\mu y}$ corresponding
41 to surface e . The values of all the smoothing parameters are restricted in the interval $[0, 1)$ for both
42 directions (Malamos and Koutsoyiannis, 2016a). When the smoothing parameters are close to 1,
43 the resulting bilinear surfaces exhibit greater smoothness, whereas, for small values of these
44 parameters, interpolation among the known points is obtained. The method is proven reliable even
45 in the case of few and scarce data, in contrast to common geostatistical methods that require a
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

denser data network to be applied reliably (Iliopoulou et al. 2022, Malamos and Koutsoyiannis 2018).

We also evaluated the performance of Ordinary Kriging (OK) and Inverse Distance Weighting (IDW). Ordinary kriging (OK) is similar to simple kriging and the only difference is that OK estimates the local constant mean, then performs simple kriging on the corresponding residuals, and only requires the stationary mean of the local search window (Goovaerts 1997). The method performed well in spatial modelling of complex environmental variables (Malamos and Koutsoyiannis, 2018).

The method of OK calculates the mean of samples within the search window and approximates the value of the estimate of the unknown value \hat{z}_u for point u as the sum of the local mean μ_0 and a stochastic term ε_u :

$$\hat{z}_u = \mu_0 + \varepsilon_u \quad (7)$$

OK requires fitting a theoretical variogram to the empirical one. However, the actual process of fitting a model to an empirical variogram is highly challenging as it involves evaluation of several types of models, a procedure that is time consuming and to some extent subjective with different authorities suggesting different methods and protocols (Bohling, 2005).

On the other hand, IDW is a straightforward and non-computationally intensive method, quite effective in many aspects (Malamos and Koutsoyiannis 2018, Tegos et al. 2017, 2015). In this case, the estimate of the unknown value \hat{z}_u for point u on the (x, y) plane, given the observed z_i values at sampled locations (x_i, y_i) is acquired in the following manner:

$$\hat{z}_u = \sum_{i=1}^n w_i z_i(x_i, y_i) \quad (8)$$

where the estimated value \hat{z}_u is a linear combination of the weights (w_i) and observed z_i values.

The weights w_i are defined as:

$$w_i = \frac{d_{ui}^{-\theta}}{\sum_{i=1}^n d_{ui}^{-\theta}}, \quad \sum_{i=1}^n w_i = 1 \quad (9)$$

In equation (9), the numerator is the inverse of distance (d_{ui}) between point u and the sampled locations (x_i, y_i) with a power θ , and the denominator is the sum of all inverse-distance weights for all locations i so that the sum of all w_i 's for an unsampled point will be unity (equation 9). The parameter θ is specified as a geometric form for the weight while other specifications are possible. This specification implies that if θ is larger than 1, the so-called distance-decay effect will be more than proportional to an increase in distance, and vice versa. In the present study the parameter θ was decided based on the leave-one-out cross validation (LOOCV) results (Malamos and Koutsoyiannis 2018), as discussed in the following section.

3.2.3 Model evaluation criteria

The criteria used for the evaluation of the regionalization performance are the mean bias error (MBE), the mean absolute error (MAE), the root mean square error (RMSE), the square of the correlation coefficient (R^2) and the Nash-Sutcliffe modelling efficiency (EF) (Malamos and Koutsoyiannis 2018, 2016b). However, the evaluation of spatial interpolation methods using different statistical metrics may not be representative with respect to the validity of the interpolation results in other locations, except for those incorporated in the interpolation procedure. Furthermore, using the entire dataset for comparison between the results of methods that respect the data points exactly such as IDW against inexact-smoothing methods such as BSS and OK could be misleading. To tackle this, a LOOCV procedure was implemented for the evaluation of the four methods (BSS, BSSE, IDW, OK) efficiency, based on the already presented criteria.

1
2
3 The LOOCV procedure is one of the commonly used methods for spatial interpolation
4 methodologies evaluation, with several researchers reporting various applications in the discipline
5 of water resources (Burrough and McDonnell 1998, Li and Heap 2008, Malamos and
6 Koutsoyiannis 2018) and it has been implemented in almost every GIS software package.
7
8
9
10
11
12

13 **3.2.4 Steps of implementation**

14
15 After characterizing the type of variability (i.e., random or physically reasoned) for each
16 parameter, the spatial generalization methodology is implemented in a sequential order which is
17 necessary due to data availability. In particular, the daily raingauges cannot be used for the
18 estimation of the parameters of the timescale function, since the latter requires data at a range of
19 sub-daily, including sub-hourly, scales for its determination. Therefore, the regionalization of the
20 parameters of the timescale function from the raingauges with sub-daily resolution is implemented
21 first (Section 4.1), followed by the regionalization of the parameters of the distribution function
22 (Section 4.2), which also incorporates the daily raingauges. The following steps are performed:
23
24
25
26
27
28
29
30
31
32
33

- 34 1. Parameter α is estimated as a common value from an optimization procedure which utilizes
35 all high quality sub-daily stations simultaneously.
36
37
- 38 2. With the α value fixed, the η parameter is re-estimated from all stations with sub-daily
39 resolution.
40
41
42
- 43 3. The four spatial models (BSS, BSSE, IDW, OK) are calibrated to the local estimates of the
44 η parameter and the best model is identified through LOOCV.
45
46
47
- 48 4. Regionalization of the η parameter is performed using the selected model and a grid of η
49 parameters is produced. This concludes the fitting of the parameters of the timescale
50 function.
51
52
53
54
55
56
57
58
59
60

5. Parameter ξ is estimated as a common value resulting from a stochastic investigation of the longest records.
6. Conditional on the already regionalized parameters (α, η, ξ) the auxiliary parameters x_1 and r_x (or equivalently, parameters β and λ , see Section 3.1.2) are estimated for all locations, i.e. both from daily and sub-daily records.
7. The spatial models are fitted for both parameters and the best model per parameter is selected based on LOOCV performance.
8. Regionalization is performed for the x_1 and r_x parameters based on the best model and the resulting parameter grids are transformed to the original parameters β and λ . This concludes the fitting of the distribution parameters and the regionalization of all parameters.

The workflow of the methodology is presented in **Figure 13**.

3.3 At-site deviations metrics

After the regionalization procedure, we aim to evaluate the deviations between the regionalized model estimates and the empirical rainfall quantiles. To investigate the impact of the regionalization on estimation for different return periods, we also estimate the deviations between the local (at-site) independently fitted parameters and the ones obtained after the regionalization.

The following three deviation statistics are used for each scale:

- a. The dimensionless (%) mean deviation of the model from the empirical K-moments, for a given timescale k :

$$D_a^{(k)} := \frac{100}{n_k} \sum_{p=1}^{n_k} \frac{x_{T_p}^{(k)} - \hat{K}_p^{r(k)}}{\hat{K}_p^{r(k)}} \quad (10)$$

where \hat{K}'_p the empirical K-moment of order p corresponding to return period T_p , n_k the data length for the given timescale, and x_{T_p} the model-derived rainfall intensity corresponding to the same return period T_p . The methodology to assign return periods to the empirical K-moments is detailed in the companion paper (Koutsoyiannis et al. 2023a).

- b. The dimensionless (%) root-mean-square error (RMSE) of the model estimates from the empirical K-moments, for a given timescale k :

$$\text{RMSE}^{(k)} := \frac{100}{\hat{K}'_{p_{\text{avg}}(k)}} \sqrt{\frac{1}{n_k} \sum_{p=1}^{n_k} (x_{T_p}^{(k)} - \hat{K}'_p^{(k)})^2} \quad (11)$$

where $\hat{K}'_{p_{\text{avg}}(k)}$ the average of the empirical K-moments for the given timescale, which is used to standardize and compare the RMSE of the different timescales.

- c. The dimensionless (%) deviation of the regional model from the local model:

$$D_r^{(k, T)} := 100 \frac{(x_{T, \text{reg}}^{(k)} - x_{T, \text{point}}^{(k)}) \times 2}{x_{T, \text{reg}}^{(k)} + x_{T, \text{point}}^{(k)}} \quad (12)$$

where $x_{T, \text{reg}}^{(k)}$ and $x_{T, \text{point}}^{(k)}$ the regional and at-site (point) model estimate for a given scale k and return period T . The D_r is estimated only for the set of the sub-daily raingauges since they allow full calibration of the ombrian relationships prior to regionalization (in contrast to the daily raingauges for which the parameters of the timescale function cannot be determined beforehand).

4. Results and analysis

4.1 Regionalization of parameters of the timescale function

4.1.1 Parameter α

From the preliminary point estimation of the parameters of the timescale function, we found that the estimation of the parameter α greatly depends on the temporal resolution of the measuring instrument, as shown in Figures A1 and A2 in the Appendix. Specifically, in stations with fine temporal resolution (5 or 10 min) the resulting values of the parameter α were small—and vice versa. This is interpreted as an artificial statistical effect rather than representing some physical reality. To compensate for the sensitivity of the α parameter to time resolution of the data, we identified a single value of this parameter for Greece, by the following procedure:

- (i) We selected the 53 stations with the longest records having temporal resolution 30 min or finer, distributed over all 14 water districts.
- (ii) We re-estimated the parameters of the equation $a(k)$, α and η , through optimization in which we set as a constraint that the value of the α parameter is the same among all 53 stations, while the η value was allowed to vary.

As a result of this optimization, the common value of $\alpha = 0.18$ h is obtained and used in all further analyses.

4.1.2 Parameter η

Conditional on the common α parameter, $\alpha = 0.18$ h, the η parameter is re-estimated for all stations with sub-daily resolution. The four spatial models, BSS, BSSE, IDW and OK, are fitted to the point estimates and results are evaluated through multiple error metrics for the entire dataset and the LOOCV analysis, as shown in **Table 3** and

1
2
3 **Table 4**, respectively. As explained in Section 3.2.3, the best model is chosen based on the
4 LOOCV analysis which in this case, favours the BSSE model, having as an additional explanatory
5 variable the elevation. In particular, an inverse relationship of η with the altitude is identified, i.e.,
6 lower values of the parameter are more likely at high altitudes and vice-versa, as seen in Figure
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4.2 Regionalization of parameters of the distribution function

4.2.1 Tail-index ξ

In the preliminary point investigation of the parameter variability, the parameter ξ (tail index of the distribution) is estimated individually per station and per instrument, and simultaneously with the optimization of the other parameters of the ombrian curves. As expected, a large spatial variability of the parameter estimates is obtained (Figure A3 in the Appendix), which indicates both the measurement uncertainty of maximum rainfall and the estimation uncertainty due to varying records lengths of the employed stations. To assess whether this variability might also reflect physically reasoned patterns in the regional realization of extreme rainfall, we perform the following investigations.

We form two datasets: (a) one comprising 61 daily raingauge stations which have at least 60 years of complete daily timeseries (termed complete daily long dataset, CDLD) and (b) a second one comprising 147 stations across the Greek territory which have at least 60 years of annual maxima values (termed annual maxima long dataset, AMLD) and which can support the estimation of the parameter with greater reliability. Using the CDLD, we attempt to link the variability of the tail index with (i) elevation, (ii) average daily rainfall, (iii) wet-day average daily rainfall and (iv)

1
2
3 the average of the n largest rainfall depths, where n the number of years. To reduce the uncertainty,
4 we condition the β parameter to represent the probability dry estimated from the daily series (see
5
6 Koutsoyiannis et al. 2023a) and allow only ξ and λ to vary. We find:
7
8

- 9
10 (i) a weak negative correlation with elevation, explaining only 9% of the variance,
11
12 (ii) a moderate positive correlation with the inverse of average daily rainfall, explaining
13
14 28% of the variance,
15
16 (iii) a moderate positive correlation with the inverse of the wet-day average daily rainfall,
17
18 explaining 31% of the variance, and
19
20 (iv) a weak correlation with the average of the n largest rainfall depths explaining 18% of
21
22 the variance of the estimates.
23
24
25

26 It is thus deemed that only cases (ii) and (iii) might be useful for the estimation of the parameter,
27
28 with case (ii) being practically superior because information for the average daily rainfall is already
29
30 available in a grid produced by the BSSE method (Koutsoyianis et al. 2023c).
31
32

33 We attempt to reproduce analysis (ii) using the AMLD dataset, thus making the estimation
34
35 based only on annual maxima data and extracting the average daily rainfall information from the
36
37 BSSE model grid, since otherwise this covariate should be estimated from complete daily series.
38
39 In this case, we observed (not shown) that although the tail index estimates from the two datasets
40
41 retained a satisfactory match ($y=0.88x$, $R^2=0.79$), the correlation with the average daily rainfall
42
43 weakened dramatically with respect to the CDLD analysis and the explained variance reduced to
44
45 zero. Considering the above results and the data limitations, we conclude that there is no strong
46
47 basis to model the tail index estimated from the annual maxima dataset employing other available
48
49 explanatory variables, representing a physical regime.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 We proceed the investigation by assuming that the entire variability of ζ estimates is a
4 statistical effect. Making this assumption, we can unify (merge) all CDLD records at a certain
5 timescale after standardizing with the mean and we estimate a unique value of ζ from the unified
6 record comprising 299 481 (standardized) nonzero daily rainfall values (the value of 0.1 mm is
7 used as the threshold for a wet day). We fit the Pareto distribution, which in preliminary analyses
8 proved to be a good model for nonzero daily rainfall, using the method of K-moments. Parameters
9 λ and ξ are estimated conditional on a fixed β parameter which is in turn estimated based on the
10 probability wet (see Koutsoyiannis et al. 2023a). For this estimation (which is based on the unified
11 full daily records), we exclude the record's largest K-moment to avoid potential outlier effects,
12 and we also exclude the K-moments which are lower than the mean value of the unified record, to
13 make the estimation more focused on larger events. The resulting ζ is estimated to be 0.18 if the
14 different stations are assumed independent ($\Theta = 0$) or larger if dependence is assumed ($\xi = 0.23$
15 for $\Theta = -0.04$, where Θ denotes bias; see Koutsoyiannis et al. 2023a for details), as shown in Figure
16 4.

17
18
19 The lower value $\zeta = 0.18$ is finally chosen.

20
21
22 To validate the assumption of a common ζ parameter, we perform the following six Monte
23 Carlo scenarios, in each of which we produce 70 simulations with 70 values from the Pareto
24 distribution. In the first three we assume a ζ value of 0.18 and in the rest 0.01, while the λ value is
25 1 for all cases. The scenarios are generated based on the following three assumptions:

- 26
27
28 (a) For each series, the 70 values are generated from the unbounded Pareto distribution, $P(\lambda, \zeta)$.
29
30 (b) For each series, the 70 values are assumed to constitute the upper $c = 20\%$ of a larger
31 sample (350 values) which is generated from the conditional Pareto distribution $P(\lambda, \zeta, c)$.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 (c) For each series, the 70 values are assumed to constitute the upper $c = 2\%$ of a larger sample
4
5 (3500 values) which is generated from the conditional Pareto distribution $P(\lambda, \xi, c)$.
6
7

8 The third assumption is more representative of the historical data and is also the one yielding the
9
10 smallest uncertainty among the three; relevant results are shown in **Figure 16** and **Figure 17**. The
11
12 findings show the (expected) large variability of the estimates in the case of a large tail index, even
13
14 for the third case. Specifically, when the true value is 0.01 the estimates span from -0.13 to 0.14 ,
15
16 but the respective variability of the estimated value of ξ (ξ_e), spans from ~ -0.1 to ~ 0.5 , when the
17
18 true value is $\xi = 0.18$. This finding supports the consistency of the assumption of a single $\xi = 0.18$
19
20 for the entire Greek territory, which is kept for the rest of the analysis.
21
22
23
24

25 **4.2.2 Parameters β and λ**

26
27 At this step, we estimate the final two parameters x_1 and r_x (to be transformed to β and λ),
28
29 conditional on the previously estimated α , η and ξ parameters. These are estimated as the
30
31 theoretical rainfall intensities x_1 corresponding to return periods $T_1 = 2$ years and $k = 24$ h and the
32
33 ratio of intensities $r_x := x_2/x_1$, where x_2 the theoretical intensity corresponding to $T_2 = 100$ years
34
35 and $k = 24$ h. The spatial models are fitted independently to the auxiliary parameters and the
36
37 results evaluated as in the case of the η parameter, are shown in **Table 3** and
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **Table 4.** In this case, both evaluations favor the IDW model which is applied for both parameters.
4
5 After the regionalization, the grids are transformed (via Equation (5)) to the original parameters β
6
7 and λ , the geographic distribution of which is shown in **Figure 18**.
8
9

10 **4.3 Performance assessment**

11 **4.3.1 Regionalization models**

12
13
14
15
16 **Table 3** and
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

Table 4 summarize the regionalization performance for the three spatially varying parameters η , x_1 and r_x in terms of the statistics calculated for the entire dataset and the ones calculated based on the LOOCV, respectively. It is recalled that the best regionalization method per parameter is chosen as the one showing superior performance in LOOCV mode, as this is a better index of its performance in pragmatic design conditions, in which ungauged areas are more likely. Expectedly, results as per the best model may differ in these two evaluations, i.e., for the η parameter the BSSE model does not show the best performance when the entire dataset is used, yet it is superior in LOOCV evaluation. The latter is an indication of the increased uncertainty involved in the regionalization of this parameter, which is based on records with sub-daily resolution. On the other hand, the IDW is consistently selected as the best model for the x_1 and r parameters in both evaluations. The superiority of the IDW method may be surprising given its simplicity and exactness of the interpolation but these aspects make it superior in the frame of the sequential order of the regional scheme, having previously incorporated smoothing schemes. We note that the regionalization of the x_1 and r parameters is performed simultaneously at the end of the regionalization process, after the choice of two common parameters and one regionally varying derived by the BSSE smoothing method. Therefore, a considerable amount of smoothing has been implemented in the conditional point estimates and hence, the more exact IDW method, preserving the local effects, is selected to compensate for the existing smoothing effect and retain a greater degree of accuracy in the final estimates.

The inter-comparison of the BSS and the BSSE performance, with the latter employing elevation as an additional covariate, allows to also evaluate the impact of orography in an objective manner. It is seen that elevation is useful as a covariate both for the η parameter (in this case the BSSE is also the best performing model), and in the case of the x_1 parameter which stands for a

1
2
3 low (2-year) return period estimate of rainfall intensity. On the other hand, the amelioration of the
4
5 r_x regionalization after the inclusion of elevation information is minor, which suggests that
6
7 elevation is less informative in the case of an extremeness ratio (it is recalled that r_x is the ratio of
8
9 the 100-year return period estimate to the 2-year one).
10
11

12 It is noted that Tables 3 and 4 report the global error statistics, while the spatial distribution
13
14 thereof was also visually checked, and no systematic errors were identified over Greece. However,
15
16 a few higher errors in the regionalization of the η parameter (controlling the timescale function of
17
18 the ombrian curves) were identified in orographically-complex areas, exhibiting various and
19
20 abrupt changes in elevation. These may indicate local orographic effects which are difficult to
21
22 capture given the sparse resolution of the sub-daily raingauges, which are used for the
23
24 regionalization of the η parameter.
25
26
27
28
29

30 **4.3.2 At-site deviations of rainfall quantiles**

31
32 To evaluate the performance of the regionalization with respect to the at-site rainfall quantiles, we
33
34 estimate the deviations of the regional model estimates from the empirical values for all stations,
35
36 daily and sub-daily. The latter are statistically evaluated according to the metrics defined in Section
37
38 3.3, based on the K-moments framework. The average deviation over all timescales is shown
39
40 separately for the daily raingauges (averaged over the 24 h and 48 h scales) and the sub-daily
41
42 raingauges (averaged over all available scales) also differentiating between those at unique
43
44 locations used for the calibration of all parameters (Set A) and those that were at the same location
45
46 with a daily raingauge which was preferred for the calibration of the distribution function (Set B).
47
48 It is observed that the distribution of average deviation is slightly biased towards positive values
49
50 (due to the correction of the shape parameter in the regional model) with the median being 8.14%
51
52 for the daily raingauges, 5.6% for Set A and 6.19% for Set B, while the median of the average
53
54
55
56
57
58
59
60

1
2
3 RMSE is 14.51% for the daily raingauges, 18.67% for Set A and 17.56% for Set B. Interestingly,
4 a similarly good performance is obtained for Set B, which acts as a validation set, since these
5 records were not used for the calibration of the distribution function, but only for the estimation of
6 the parameters of the timescale function. Detailed results for all timescales are shown in the
7 Appendix B (Figures B1 and B2), suggesting the increased uncertainty prevalent at the small
8 timescales for which however fewer records are available (Table 1).
9

10
11 To inspect the impact of regionalization on design rainfall estimates for various return
12 periods, we compare the deviations between the estimates using the regional parameters and the
13 ones obtained using the local (at-site) parameters, shown for two characteristic scales in **Figure**
14 **20**.
15

16
17 For both scales, the deviations consistently tend to increase towards larger values as the return
18 period increases. This is attributed to the use of the single high value of the parameter ξ in the
19 regionalization, the influence of which is stronger in high return periods (for $T = 1000$ years,
20 median 32.35% at 1 h and 33.86% at 24 h). In the low return periods (of the order of 2 years), the
21 spatially generalized rainfall model leads to slightly smaller rainfall estimates (for $T = 2$ years,
22 median -3.55% at 1 h and -5.22% at 24 h). In general, though, a positive bias prevails in these
23 evaluations, which is consistent with the use of a common and high shape parameter (tail-index)
24 in the regional model.
25

26
27 An illustration of the fitting is shown in **Figure 21** for two stations, Limnos, a station with
28 only sub-daily data at a single location (example of Set A) and Lofos Nymfon (Hill of Nymphs) a
29 station with two gauges, a daily raingauge used for the calibration of the distribution function and
30 a subdaily raingauge used for the calibration of the timescale function (an example of Set B). It is
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 noted that the daily raingauge at the Lofos Nymfon is also the longest rainfall record available in
4
5 Greece (dating back to 1863).
6
7

8 **4.4 Design rainfall maps**

9

10
11 The design rainfall (in mm) maps for scales 1 h and 24 h are shown in Figure 22 for return periods
12
13 $T= 50, 100$ and 1000 years. The emerging spatial patterns are consistent with known Greek
14
15 climatology, namely the impact of orography and the increased spatial diversity, with lower
16
17 amounts of rainfall occurring in Eastern Greece compared to its Western counterpart, except for
18
19 some coastal areas in Eastern Greece, in which rainfall rates are very intense. Geographic patterns
20
21 are more abrupt at the 1 h scale, reflecting the erratic nature of fine-scale precipitation whereas
22
23 patterns are smoother over the 24 h scale and in accordance with the orographic effect, i.e., higher
24
25 rainfall amounts at the daily scale are more likely in the mountainous regions. On the other hand,
26
27 the orographic enhancement of precipitation is not evident at the 1 h scale; rather the rainfall depths
28
29 appear decreased at some regions with higher elevations (namely along the Pindus Mountain
30
31 range) compared to lower elevation/coastal areas.
32
33
34
35
36

37 **5. Discussion**

38

39
40 The large-scale regionalization scheme employed herein entails several methodological choices
41
42 that warrant further discussion, aside from the resulting findings.
43
44

45 A critical choice, stemming from a thorough physically reasoned and stochastic
46
47 investigation, is the implementation of a common tail-index (parameter ξ) over Greece. In the
48
49 absence of robust geographic and hydroclimatic links, this choice is dictated by the substantial
50
51 uncertainty involved in the estimation of the tail index, which is particularly relevant given the
52
53 limited record lengths available (Koutsoyiannis 2004). This uncertainty is re-estimated herein by
54
55
56
57
58
59
60

1
2
3 means of Monte Carlo simulations and found consistent to the observed variability of the (at-site)
4 point estimates. It is noted that the alternative approach of independent estimation and
5 interpolation of the ζ parameters, despite being simple, is likely to result into unrealistic spatial
6 patterns and abrupt changes, as already shown by Uboldi et al. (2014) and Jalbert et al. (2022). On
7 the other hand, a spatially invariant tail-index is a widely used approach in regionalization of
8 extreme rainfall (see e.g., Jalbert et al. 2022), even used in the global-scale investigation of IDF
9 curves by Courty et al. (2019). Shehu et al. (2022) in the regionalization of the curves for Germany
10 also concluded on keeping the tail-index constant, and simultaneously regionalizing the rest
11 parameters using kriging methods. As a result of the common and high tail index, the regional
12 model's predictions, particularly for the higher return periods, tend to be greater than the single-
13 station rainfall quantiles, as well as the single-site (independently estimated) model predictions.
14 This is to be expected as the pooling methodology used for the estimation of the tail-index in the
15 regional model aimed in the first place to compensate for the underestimation bias of the tail-index
16 due to the small sample size of the individual records.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 Another finding worth discussing is the inclusion of the IDW method in the regionalization
36 framework, which demonstrated its superiority in the case of the x_1 and r_x parameters. Bloeschl
37 and Grayson (2001) have shown that the IDW generates spurious artefacts in the case of highly
38 variable quantities and irregularly spaced data sites. We also deem that this would be true had we
39 applied IDW as the primary regionalization method. Yet by applying this method at the end of the
40 sequential regionalization, we have already smoothed out a significant degree of variation and at
41 that point of the regionalization process, greater precision is a desired fact. Indeed, the resulting
42 patterns are relatively smooth and can be efficiently interpolated by a simpler method. This is the
43 reason why IDW is often part of hybrid approaches (e.g., Perica et al. 2009) and even sophisticated
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 algorithms incorporate some sort of IDW (e.g., PRISM, Daly 2006). It is also important to consider
4
5 that the superiority of no method is guaranteed under general conditions. For instance, recently,
6
7 Sangüesa et al. (2023) suggested that IDW outperformed kriging in the case of sparse data.
8
9

10 Finally, the usefulness of elevation as a geographic covariate for extreme rainfall is an open
11
12 research question, with various methodologies and regions reporting diverse results. For instance,
13
14 Szolgay et al. (2009) did not find elevation significant as an additional explanatory variable for 2
15
16 and 100-year daily rainfall depths in Central Slovakia. In contrast, regional rainfall studies in
17
18 Greece at smaller spatial scales (e.g., Iliopoulou et al. 2022, Iliopoulou and Koutsoyiannis 2022)
19
20 identified elevation as a good explanatory variable mainly affecting the mean of the annual maxima
21
22 distribution at the daily scale. Here, elevation is found useful for the regionalization of the η
23
24 parameter (controlling the scaling of the intensity) and to a lesser extent found meaningful for the
25
26 2-year daily rainfall intensity (although this model did not rank best). Resulting maps (Section 4.4)
27
28 suggest that orography enhances precipitation at the 24 h scale, but this effect is not evident at the
29
30 1 h scale; instead a tendency of intensity to drop with elevation at the short timescales is identified.
31
32
33 Therefore, the dependence of rainfall intensity on elevation is not uniform even across scales.
34
35
36 Similar results have been reported for Italy (see e.g., Mazzoglio et al. 2022), where a decrease of
37
38 annual maxima 1 h rainfall with elevation has been identified and called ‘reverse orographic effect’
39
40 (Avanzi et al. 2015). More research is required however to disentangle the effect of elevation
41
42 across the different scales and to investigate the usefulness of other proxies of the rainfall
43
44 generating mechanisms (e.g., convection) into the regionalization of the parameters.
45
46
47
48
49

50 **6. Conclusions**

51
52
53 In this work, we have developed a sequential regionalization methodology for design rainfall in
54
55 the form of rainfall intensity–timescale–return period relationships and applied the latter over the
56
57

1
2
3 entire Greek territory (131,957 km²). This is the first time a regional design rainfall model is
4 available in gridded form (in a 5 km x 5 km grid) for Greece. The approach followed incorporates
5 an advanced framework for regional probability analysis employing knowable (K-) moments and
6 a more physically sound re-parameterization of the Koutsoyiannis et al. (1998) model. The
7 theoretical background is extensively discussed in the companion paper by Koutsoyiannis et al.
8 (2023a) whereas the regional implementation and the results are presented herein.
9
10
11
12
13
14
15
16

17 The regionalization process was informed by extensive analyses using the independent, at-
18 site estimation results as a basis. By comprehensive evaluations of the variability involved in the
19 estimation of the parameters, two parameters were considered spatially invariant, i.e., the timescale
20 parameter α and the tail-index ξ , and were estimated simultaneously, in order to increase the
21 robustness of the estimation. The calibration of the tail index has been corroborated by extensive
22 stochastic investigations. The other three parameters (η , λ and β) were regionalized following a
23 sequential implementation scheme that exploits information both from daily and sub-daily
24 raingauges. In particular, by first regionalizing the parameters of the timescale function from the
25 sub-daily raingauges, the methodology allows an efficient incorporation of the denser and more
26 reliable network of daily raingauges for the subsequent calibration of the distribution function.
27 Four regionalization methods, ranging from smoothing to exact interpolation schemes, namely the
28 BSS, BSSE, OK and IDW, were applied and evaluated for each of the three parameters, while the
29 impact of elevation was also considered. The best method was selected per parameter in terms of
30 the leave-one-out-cross-validation performance, and a final hybrid scheme emerged that combines
31 smoothing and exact interpolation techniques. In particular, the smoothing BSSE method with the
32 altitude as explanatory variable was selected for the regionalization of the η parameter of the
33 timescale function, performed first in order, while the exact IDW was selected for the remaining
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 two parameters, λ and β , the regionalization of which was performed at the end of the sequential
4
5 procedure. Therefore, the framework allowed the evaluation of the necessary degree of smoothing
6
7 at each step of the sequential regionalization based on the cross-validation performance.
8
9

10 The final product consists of a five-parameter relationship, with two constant parameters and
11
12 three spatially varying parameters available in a 5 km x 5 km grid. This forms a powerful and easy
13
14 to apply design tool covering the entire territory of Greece while the methodology can be readily
15
16 transferred to other countries or parts thereof.
17
18
19
20

21 **Funding:** This research was funded by the General Directorate of Water of the Greek Ministry of
22
23 Environment and Energy through the unit “Επιτελική Δομή ΕΣΠΑ Υπουργείου Περιβάλλοντος
24
25 και Ενέργειας Τομέα Περιβάλλοντος” in the framework of the project “Production of maps with
26
27 updated parameters of the ombrian curves at country level (implementation of the EU Directive
28
29 2007/60/EC in Greece)”.

30
31
32 **Data Availability Statement:** The data of the General Directorate of Water of the Greek Ministry
33
34 of Environment and Energy are available online for free from the Hydroscope platform
35
36 (<http://www.hydroscope.gr/> (accessed on 25 March 2023)). All other data belong to other Greek
37
38 organizations and are not publicly available.
39
40
41

42 **Acknowledgments** We are grateful to Associate Editor Elena Volpi, as well as to Professor
43
44 Pierluigi Claps and a second anonymous reviewer for their comments which helped us to improve
45
46 the manuscript.
47
48

49 **Conflicts of Interest:** We declare no conflict of interest.
50
51
52
53
54
55
56
57
58
59
60

References

- Ahrens, B., 2006. Distance in spatial interpolation of daily rain gauge data. *Hydrology and Earth System Sciences*, 10(2), pp.197-208.
- Aron, G., Wall, D.J., White, E.L. and Dunn, C.N., 1987. REGIONAL RAINFALL INTENSITY-DURATION-FREQUENCY CURVES FOR PENNSYLVANIA 1. *JAWRA Journal of the American Water Resources Association*, 23(3), pp.479-485.
- Avanzi, F., De Michele, C., Gabriele, S., Ghezzi, A. and Rosso, R., 2015. Orographic signature on extreme precipitation of short durations. *Journal of Hydrometeorology*, 16(1), pp.278-294.
- Bara M., Marta, Kohnová, S., Ladislav, G.A.Á.L., Szolgay, J. and Hlavčová, K., 2009. Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia. *Contributions to Geophysics and Geodesy*, 39(3), pp.187-206.
- Barbosa, A.D.G., Souza Neto, I.R.D., Costa, V.A.F. and Mendes, L.A., 2022. Assessing intensity-duration-frequency equations and spatialization techniques across the Grande River Basin in the state of Bahia, Brazil. *RBRH*, 27.
- Bernard, M.M., 1932. Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers*, 96(1), pp.592-606.
- Blanchet, J., Ceresetti, D., Molinié, G. and Creutin, J.D., 2016. A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis. *Journal of Hydrology*, 540, pp.82-95.
- Bohling, G., 2005. Introduction to geostatistics and variogram analysis. *Kansas geological survey*, 1(10), pp.1-20.
- Burlando, P. and Rosso, R., 1996. Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187(1-2), pp.45-64.
- Burn, D.H., 2014. A framework for regional estimation of intensity–duration–frequency (IDF) curves. *Hydrological Processes*, 28(14), pp.4209-4218.
- Burrough, P.A., McDonnell, R.A. and Lloyd, C.D., 2015. Principles of geographical information systems. Oxford University Press, USA.
- Cassalho, F., Beskow, S., de Mello, C.R., de Moura, M.M., de Oliveira, L.F. and de Aguiar, M.S., 2019. Artificial intelligence for identifying hydrologically homogeneous regions: A state-of-the-art regional flood frequency analysis. *Hydrological processes*, 33(7), pp.1101-1116.
- Ceresetti, D., Ursu, E., Carreau, J., Anquetin, S., Creutin, J.D., Gardes, L., Girard, S. and Molinie, G., 2012. Evaluation of classical spatial-analysis schemes of extreme rainfall. *Natural hazards and earth system sciences*, 12(11), pp.3229-3240.
- Claps, P., Ganora, D. and Mazzoglio, P., 2022. Rainfall regionalization techniques. In *Rainfall* (pp. 327-350). Elsevier.
- Courty, L.G., Wilby, R.L., Hillier, J.K. and Slater, L.J., 2019. Intensity-duration-frequency curves at the global scale. *Environmental Research Letters*, 14(8), p.084045.
- Daly, C., 2006. Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 26(6), pp.707-721.
- De Luca, D.L. and Napolitano, F., 2023. A user-friendly software for modelling extreme values: EXTRASTAR (EXTRemes Abacus for STATistical Regionalization). *Environmental Modelling & Software*, 161, p.105622.
- Deidda, R., Hellies, M. and Langousis, A., 2021. 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. *Stochastic Environmental Research and Risk Assessment*, 35(12), pp.2605-2628.
- Faulkner, D.S. and Prudhomme, C., 1998. Mapping an index of extreme rainfall across the UK. *Hydrology and Earth System Sciences*, 2(2/3), pp.183-194.
- Forestieri, A., Lo Conti, F., Blenkinsop, S., Cannarozzo, M., Fowler, H.J. and Noto, L.V., 2018. Regional frequency analysis of extreme rainfall in Sicily (Italy). *International Journal of Climatology*, 38, pp.e698-e716.
- Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, 228(1-2), pp.113-129.
- Goovaerts, P., 1997. Geostatistics for natural resources evaluation. *Applied Geostatistics*.
- Hailegeorgis, T.T., Thorolfsson, S.T. and Alfredsen, K., 2013. Regional frequency analysis of extreme precipitation with consideration of uncertainties to update IDF curves for the city of Trondheim. *Journal of Hydrology*, 498, pp.305-318.
- Haruna, A., Blanchet, J. and Favre, A.C., 2023. Modeling Intensity-Duration-Frequency Curves for the Whole Range of Non-Zero Precipitation: A Comparison of Models. *Water Resources Research*, 59(6), p.e2022WR033362.
- Hengl, T., 2006. Finding the right pixel size. *Computers & Geosciences*, 32(9), pp.1283-1298.
- Hosking, J.R.M. and Wallis, J.R., 1997. Regional frequency analysis (p. 240).
- Huffman, G.J., E.F. Stocker, D.T. Bolvin, E.J. Nelkin, Jackson Tan, 2019. GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), 10.5067/GPM/IMERG/3B-HH/06
- Hutchinson, M.F., 1995. Interpolating mean rainfall using thin plate smoothing splines. *International journal of geographical information systems*, 9(4), pp.385-403.
- Iliopoulou, T. and Koutsoyiannis, D. 2022. A parsimonious approach for regional design rainfall estimation: the case study of Athens, *Proceedings of 7th IAHR Europe Congress "Innovative Water Management in a Changing Climate"*, Athens, International Association for Hydro-Environment Engineering and Research (IAHR).
- Iliopoulou, T., Koutsoyiannis, D., Koukouvinos, A., Malamos, N., Tepetidis, N., Markantonis, D., Dimitriadis, P. and Mamassis, N., 2023. Regionalized design rainfall curves for Greece (No. EGU23-8740). *Copernicus Meetings*.
- Iliopoulou, T., Malamos, N. and Koutsoyiannis, D., 2022. Regional ombrian curves: Design rainfall estimation for a spatially diverse rainfall regime. *Hydrology*, 9(5), p.67.
- Jalbert, J., Genest, C. and Perreault, L., 2022. Interpolation of precipitation extremes on a large domain toward idf curve construction at unmonitored locations. *Journal of Agricultural, Biological and Environmental Statistics*, 27(3), pp.461-486.
- Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara, 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (<http://srtm.csi.cgiar.org>)
- Koutsoyiannis, D., 2004. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological sciences journal*, 49(4).
- Koutsoyiannis, D. 2023. *Stochastics of Hydroclimatic Extremes - A Cool Look at Risk*, Edition 3, ISBN: 978-618-85370-0-2, 391 pages, doi:10.57713/kallipos-1, Kallipos Open Academic Editions, Athens.
- Koutsoyiannis, D. and Iliopoulou, T., 2022. Ombrian curves advanced to stochastic modeling of rainfall intensity. *Rainfall*, pp.261-284.

- 1
2
3 Koutsoyiannis, D., Iliopoulou, T., Koukouvinos, A., Malamos, N., 2023a. A stochastic framework
4 for rainfall intensity-timescale-return period relationships. Part I. Theory and estimation
5 strategies, *Hydrological Sciences Journal*, *submitted*.
- 6
7 Koutsoyiannis, D., Iliopoulou, T., Koukouvinos, A., Malamos, N., Mamassis, N., Dimitriadis, P.,
8 Tepetidis, N., and Markantonis, D., 2023b. Technical Report, Production of maps with
9 updated parameters of the ombrian curves at country level (implementation of the EU
10 Directive 2007/60/EC in Greece), Department of Water Resources and Environmental
11 Engineering – National Technical University of Athens.
- 12
13 Koutsoyiannis, D., Iliopoulou, T., Koukouvinos, A., Malamos, N., Mamassis, N., Dimitriadis, P.,
14 Tepetidis, N. and Markantonis, D., 2023c. In Search of Climate Crisis in Greece Using
15 Hydrological Data: 404 Not Found. *Water*, 15(9), p.1711.
- 16
17 Koutsoyiannis, D., Kozonis, D. and Manetas, A., 1998. A mathematical framework for studying
18 rainfall intensity-duration-frequency relationships. *Journal of hydrology*, 206(1-2), pp.118-
19 135.
- 20
21 Koutsoyiannis, D. and Papalexiou, S.M., 2017. Extreme rainfall: Global perspective. *Handbook*
22 *of Applied Hydrology*; McGraw-Hill: New York, NY, USA, pp.74-1.
- 23
24 Lanciotti, S., Ridolfi, E., Russo, F. and Napolitano, F., 2022. Intensity–Duration–Frequency
25 Curves in a Data-Rich Era: A Review. *Water*, 14(22), p.3705.
- 26
27 Li, J. and Heap, A.D., 2008. A review of spatial interpolation methods for environmental scientists.
28
29 Libertino, A., Allamano, P., Laio, F. and Claps, P., 2018. Regional-scale analysis of extreme
30 precipitation from short and fragmented records. *Advances in Water Resources*, 112,
31 pp.147-159.
- 32
33 Lin, G.F. and Chen, L.H., 2006. Identification of homogeneous regions for regional frequency
34 analysis using the self-organizing map. *Journal of Hydrology*, 324(1-4), pp.1-9.
- 35
36 Madsen, H., Gregersen, I.B., Rosbjerg, D. and Arnbjerg-Nielsen, K., 2017. Regional frequency
37 analysis of short duration rainfall extremes using gridded daily rainfall data as co-variate.
38 *Water Science and Technology*, 75(8), pp.1971-1981.
- 39
40 Malamos, N. and Koutsoyiannis, D., 2016a. Bilinear surface smoothing for spatial interpolation
41 with optional incorporation of an explanatory variable. Part 1: Theory. *Hydrological*
42 *Sciences Journal*, 61(3), pp.519-526.
- 43
44 Malamos, N. and Koutsoyiannis, D., 2016b. Bilinear surface smoothing for spatial interpolation
45 with optional incorporation of an explanatory variable. Part 2: Application to synthesized
46 and rainfall data. *Hydrological Sciences Journal*, 61(3), pp.527-540.
- 47
48 Malamos, N. and Koutsoyiannis, D., 2018. Field survey and modelling of irrigation water quality
49 indices in a Mediterranean island catchment: A comparison between spatial interpolation
50 methods. *Hydrological sciences journal*, 63(10), pp.1447-1467.
- 51
52 Mascaro, G., 2020. Comparison of local, regional, and scaling models for rainfall intensity–
53 duration–frequency analysis. *Journal of applied meteorology and climatology*, 59(9),
54 pp.1519-1536.
- 55
56 Mazzoglio, P., Butera, I., Alvioli, M. and Claps, P., 2022. The role of morphology in the spatial
57 distribution of short-duration rainfall extremes in Italy. *Hydrology and Earth System*
58 *Sciences*, 26(6), pp.1659-1672.
- 59
60 Molini, A., Lanza, L.G. and La Barbera, P., 2005. The impact of tipping-bucket raingauge
measurement errors on design rainfall for urban-scale applications. *Hydrological*
Processes: An International Journal, 19(5), pp.1073-1088.

- 1
2
3 Ombadi, M., Nguyen, P., Sorooshian, S. and Hsu, K.L., 2018. Developing
4 intensity-duration-frequency (IDF) curves from satellite-based precipitation: Methodology
5 and evaluation. *Water Resources Research*, 54(10), pp.7752-7766.
- 6
7 Overeem, A., Buishand, A. and Holleman, I., 2008. Rainfall depth-duration-frequency curves and
8 their uncertainties. *Journal of Hydrology*, 348(1-2), pp.124-134.
- 9 Perica, S., Martin, D., Lin, B., Parzybok, T., Riley, D., Yekta, M., Hiner, L., Chen, L.C., Brewer,
10 D., Yan, F. and Maitaria, K., 2009. *Precipitation-Frequency Atlas of the United States.*
11 *Volume 4 Version 3. Hawaiian Islands.*
- 12 Rao, A.R. and Srinivas, V.V., 2008. Introduction. In A. R. Rao, & V. V. Srinivas (Eds.),
13 *Regionalization of watersheds: An approach based on cluster analysis* (pp. 1–16).
14 Dordrecht: Springer Science+Business Media B.V. [https://doi.org/10.1007/978-1-4020-](https://doi.org/10.1007/978-1-4020-6852-2_1)
15 [6852-2_1](https://doi.org/10.1007/978-1-4020-6852-2_1).
- 16
17 Sangüesa, C., Pizarro, R., Ingram, B., Ibáñez, A., Rivera, D., García-Chevesich, P., Pino, J., Pérez,
18 F., Balocchi, F. and Peña, F., 2023. Comparing Methods for the Regionalization of
19 Intensity– Duration– Frequency (IDF) Curve Parameters in Sparsely-Gauged and
20 Ungauged Areas of Central Chile. *Hydrology*, 10(9), p.179.
- 21 Shehu, B., Willems, W., Stockel, H., Thiele, L.B. and Haberlandt, U., 2023. Regionalisation of
22 rainfall depth–duration–frequency curves with different data types in Germany. *Hydrology*
23 *and Earth System Sciences*, 27(5), pp.1109-1132.
- 24 Sherman, C.W., 1931. Frequency and intensity of excessive rainfalls at Boston, Massachusetts.
25 *Transactions of the American Society of Civil Engineers*, 95(1), pp.951-960.
- 26 Svensson, C. and Jones, D.A., 2010. Review of rainfall frequency estimation methods. *Journal of*
27 *Flood Risk Management*, 3(4), pp.296-313.
- 28 Szolgay, J., Parajka, J., Kohnová, S. and Hlavčová, K., 2009. Comparison of mapping approaches
29 of design annual maximum daily precipitation. *Atmospheric Research*, 92(3), pp.289-307.
- 30 Tegos, A., Malamos, N., Efstratiadis, A., Tsoukalas, I., Karanasios, A., Koutsoyiannis, D., 2017.
31 *Parametric Modelling of Potential Evapotranspiration: A Global Survey.* *Water* 9.
32 <https://doi.org/10.3390/w9100795>
- 33 Tegos, A., Malamos, N. and Koutsoyiannis, D., 2015. A parsimonious regional parametric
34 evapotranspiration model based on a simplification of the Penman–Monteith formula.
35 *Journal of Hydrology*, 524, pp.708-717.
- 36 Trefry, C.M., Watkins Jr, D.W. and Johnson, D., 2005. Regional rainfall frequency analysis for
37 the state of Michigan. *Journal of Hydrologic Engineering*, 10(6), pp.437-449.
- 38 Uboldi, F., Sulis, A. N., Lussana, C., Cislighi, M., & Russo, M. (2014). A spatial bootstrap
39 technique for parameter estimation of rainfall annual maxima distribution. *Hydrology and*
40 *Earth System Sciences*, 18(3), 981-995.
- 41 Ulrich, J., Jurado, O.E., Peter, M., Scheibel, M. and Rust, H.W., 2020. Estimating IDF curves
42 consistently over durations with spatial covariates. *Water*, 12(11), p.3119.
- 43 Wahba, G. and Wendelberger, J., 1980. Some new mathematical methods for variational objective
44 analysis using splines and cross validation. *Monthly weather review*, 108(8), pp.1122-
45 1143.
- 46
47 Watkins Jr, D.W., Link, G.A. and Johnson, D., 2005. Mapping regional precipitation intensity
48 duration frequency estimates 1. *JAWRA Journal of the American Water Resources*
49 *Association*, 41(1), pp.157-170.
- 50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7 **Appendix A: Results from point investigation**
8

9 [Figure A1]
10

11 [Figure A2]
12

13 [Figure A3]
14
15
16
17
18
19

20 **Appendix B: Supplementary results from regionalization**
21

22 [Figure B1]
23

24 [Figure B2]
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Number of stations with data at each timescale.

	Daily Raingauges		Sub-daily Raingauges										
	1 d	2 d	5 min	10 min	15 min	30 min	1 h	2 h	3 h	6 h	12 h	24 h	48 h
Count	503	490	38	128	47	231	273	279	269	280	280	280	223

For Peer Review Only

Table 2. Number of stations with 24-hour annual maximum time series longer than 30, 60 and 90 years.

	Length 24 h ≥30 years	Length 24 h ≥60 years	Length 24 h ≥90 years
Sub-daily raingauges	109	2	1
Daily raingauges	383	129	1

For Peer Review Only

Table 3. Entire dataset statistics for η , x_1 and r_x .

Interpolation method	MBE (mm)	MAE (mm)	RMSE	R^2	EF
η					
BSS	0.00	0.07	0.08	0.25	0.26
BSSE*	0.00	0.05	0.06	0.57	0.57
OK	0.00	0.03	0.03	0.91	0.88
IDW	0.00	0.01	0.02	0.98	0.98
x_1					
BSS	0.00	0.02	0.04	0.28	0.27
BSSE	0.00	0.38	0.54	0.63	0.43
OK	0.00	0.21	0.31	0.90	0.88
IDW*	0.01	0.13	0.24	0.93	0.92
r_x					
BSS	0.00	0.17	0.26	0.26	0.20
BSSE	0.00	0.18	0.26	0.27	0.26
OK	0.00	0.15	0.23	0.38	0.33
IDW*	0.01	0.13	0.24	0.93	0.92

*denotes best performance according to LOOCV

Table 4. Leave-one-out cross validation statistics for η , x_1 and r_x .

Interpolation method	MBE (mm)	MAE (mm)	RMSE	R^2	EF
η					
BSS	0.00	0.07	0.08	0.28	0.28
BSSE*	0.00	0.06	0.07	0.40	0.40
OK	0.00	0.07	0.08	0.30	0.28
IDW	0.00	0.06	0.08	0.28	0.28
x_1					
BSS	0.00	0.02	0.04	0.18	0.17
BSSE	-0.01	0.51	0.75	0.43	0.40
OK	0.00	0.43	0.64	0.49	0.47
IDW*	0.01	0.44	0.65	0.49	0.48
r_x					
BSS	0.00	0.18	0.27	0.16	0.12
BSSE	0.00	0.18	0.27	0.18	0.15
OK	0.00	0.18	0.28	0.08	0.07
IDW*	-0.01	0.17	0.29	0.32	0.28

*denotes best performance according to LOOCV

Figure captions

Figure 1. Elevation map of Greece along with the locations of the daily and sub-daily resolution rainfall stations used in the analysis. The coordinate reference system is the GGRS87/Greek Grid (EPSG:2100).

Figure 2. Workflow of the regionalization procedure. Yellow boxes correspond to procedures for simultaneous parameter estimation (shared parameters), green ones to the final regionalization procedures and the remaining white boxes denote procedures that are intermediate and auxiliary to the regionalization.

Figure 3. Regionalized η parameter resulting from the application of the BSSE model. The coordinate reference system is the GGRS87/Greek Grid (EPSG:2100).

Figure 4. Theoretical (Pareto) and empirical distribution of the standardized daily rainfall depth for the case where dependence bias is considered ($\theta=-0.04$) and the case where dependence is neglected ($\theta=0$).

Figure 5. Probability distribution of simulated tail index ξ_e , generated from a Pareto distribution with scale parameter $\lambda = 1$ and theoretical tail index $\xi = 0.01$ and $\xi = 0.18$ assuming that the generated values are the upper 2% of a larger sample. The results are obtained from 70 simulated series of 70 values, for each of which the empirical values λ_e and ξ_e are estimated by optimization. In addition to the empirical distributions, the theoretical normal distributions $N(\xi - \zeta, \sigma)$ are also shown, where $\zeta = -0.03$ is the bias and σ the standard deviation, with values as in the legend of the figure, both estimated with Monte Carlo.

Figure 6. Probability distribution of simulated standardized daily rainfall, generated from a Pareto distribution with scale parameter $\lambda = 1$ and tail index $\xi = 0.18$ assuming that the simulated values are the upper 2% of a larger sample. In addition to the median and 90% confidence limits of the simulated empirical distributions, two individual simulated distributions are shown with an empirical (optimized) index ξ_e approximately equal to its upper and lower confidence limits (where the high and low ξ_e are 0.35 and -0.02 , respectively).

Figure 7. Regionalized β (years) and λ (mm/h) parameters using the IDW method. The coordinate reference system is the GGRS87/Greek Grid (EPSG:2100).

1
2
3 **Figure 8.** Dimensionless (%) mean deviation of the model from the empirical K-moments (left)
4 and dimensionless (%) RMSE (right), both averaged over all timescales available from the daily
5 and sub-daily raingauges. Set A indicates the subdaily raingauges used for the full calibration of
6 the ombrian relationships, while Set B denotes the subdaily raingauges used only for the calibration
7 of the parameters of the timescale function.
8
9

10
11
12 **Figure 9.** Dimensionless (%) deviation of the regional model from the local model for 1 h scale
13 (left) and 24 h scale (right) and different return periods.
14
15

16
17 **Figure 10.** Theoretical and empirical distributions of annual maximum intensities at 5 min to 48
18 h scales (depending on the available samples) from the sub-daily stations of Limnos (top) and
19 Lofos Nymphon-Athens (bottom). For the latter, the empirical intensities at the 24 h and 48 h
20 scales from the daily raingauge are shown as well. The empirical intensities plotted based on order
21 statistics are also shown for validation.
22
23
24

25
26 **Figure 11.** left panel: Estimated rainfall depth (mm) for 1 h and return periods 50, 100 and 1000
27 years, right panel: Estimated rainfall depth (mm) for 24 h and return periods 50, 100 and 1000
28 years. Quantile classification is employed for all legends. The coordinate reference system is the
29 GGRS87/Greek Grid (EPSG:2100).
30
31
32
33

34
35 **Figure A1.** Estimate of parameter α vs temporal resolution of the rain recorder.
36

37 **Figure A2.** Average of estimated α parameters vs the temporal resolution of the rain recorders.
38

39 **Figure A3.** Geographic distribution of the independently estimated ζ parameters.
40

41 **Figure B1.** Dimensionless (%) mean deviation of the model from the empirical K-moments (left)
42 and dimensionless (%) RMSE (right), for all timescales available from the daily raingauges.
43

44 **Figure B2.** Dimensionless (%) mean deviation of the model from the empirical K-moments (left)
45 and dimensionless (%) RMSE (right), for all timescales available from the sub-daily raingauges.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

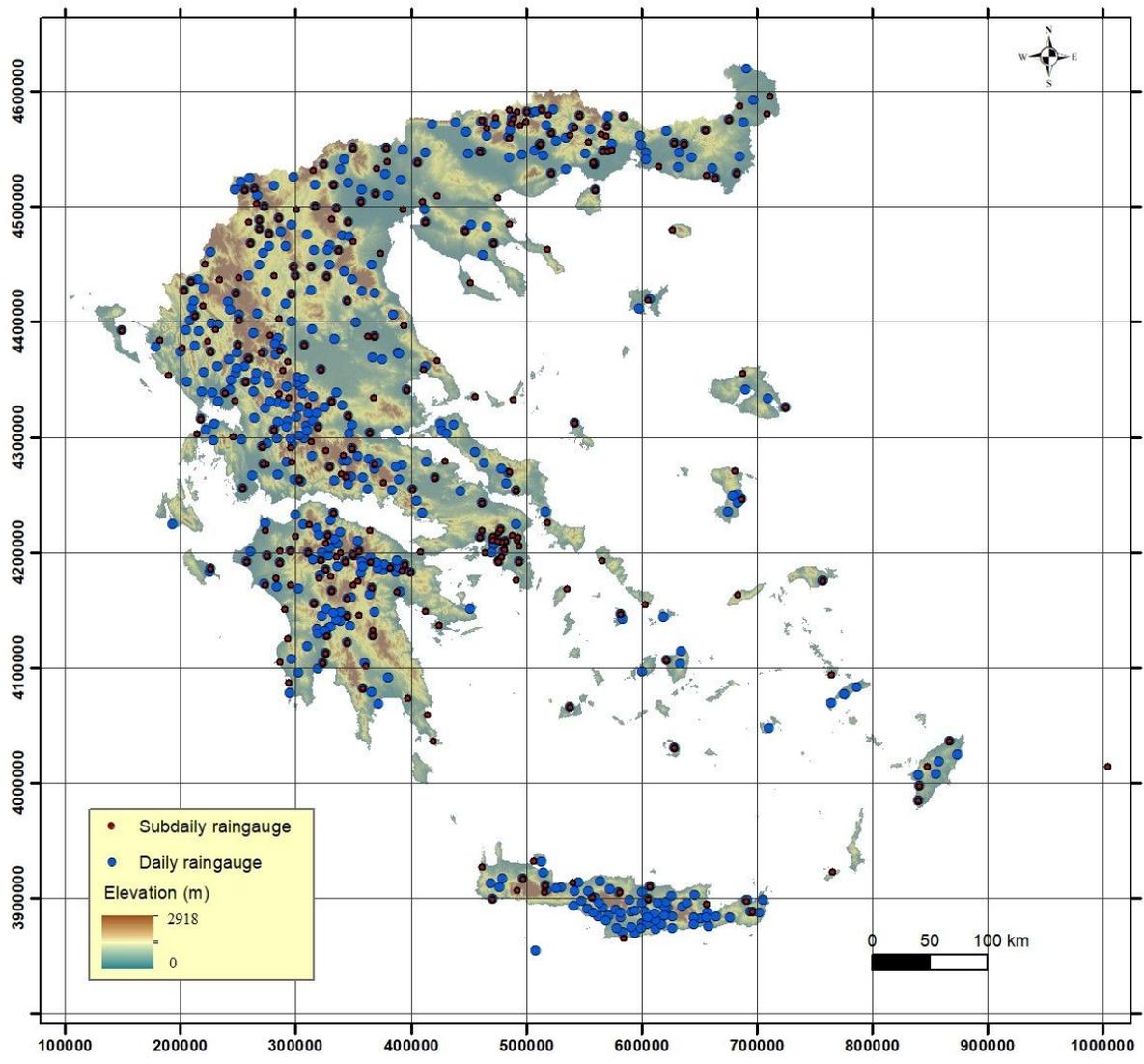


Figure 12

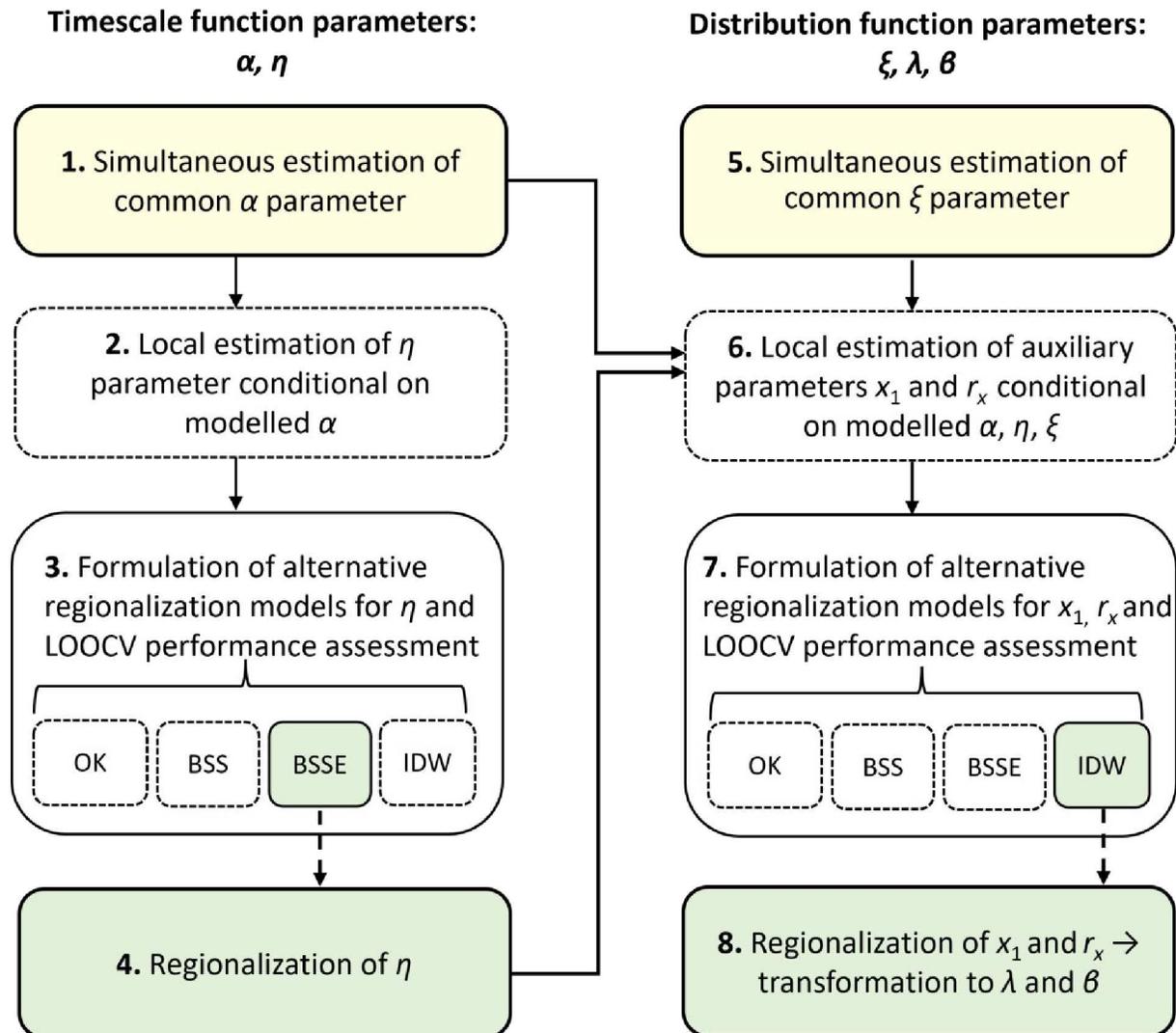


Figure 13

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

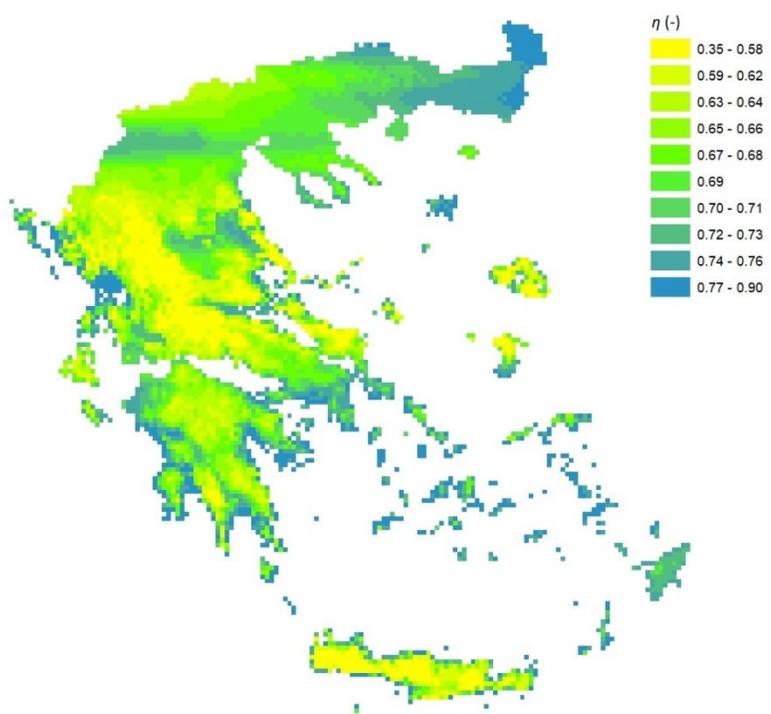


Figure 14

Review Only

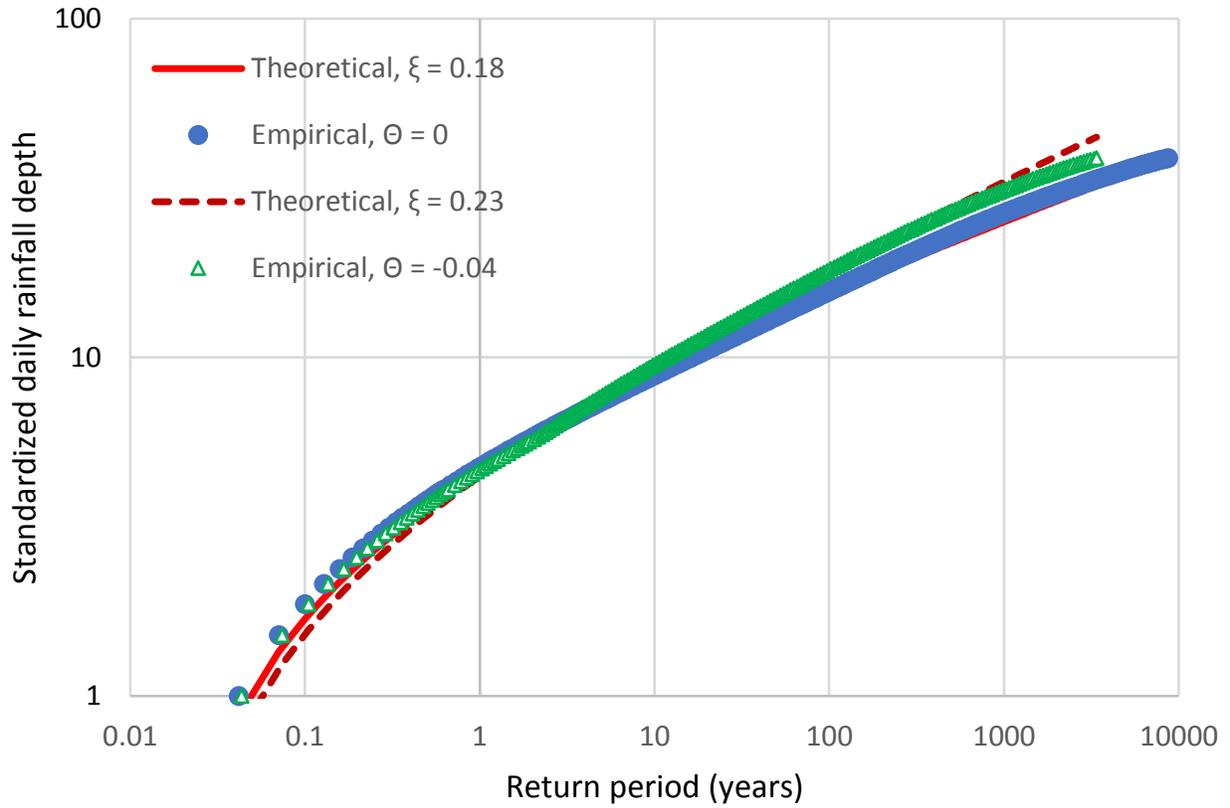


Figure 15

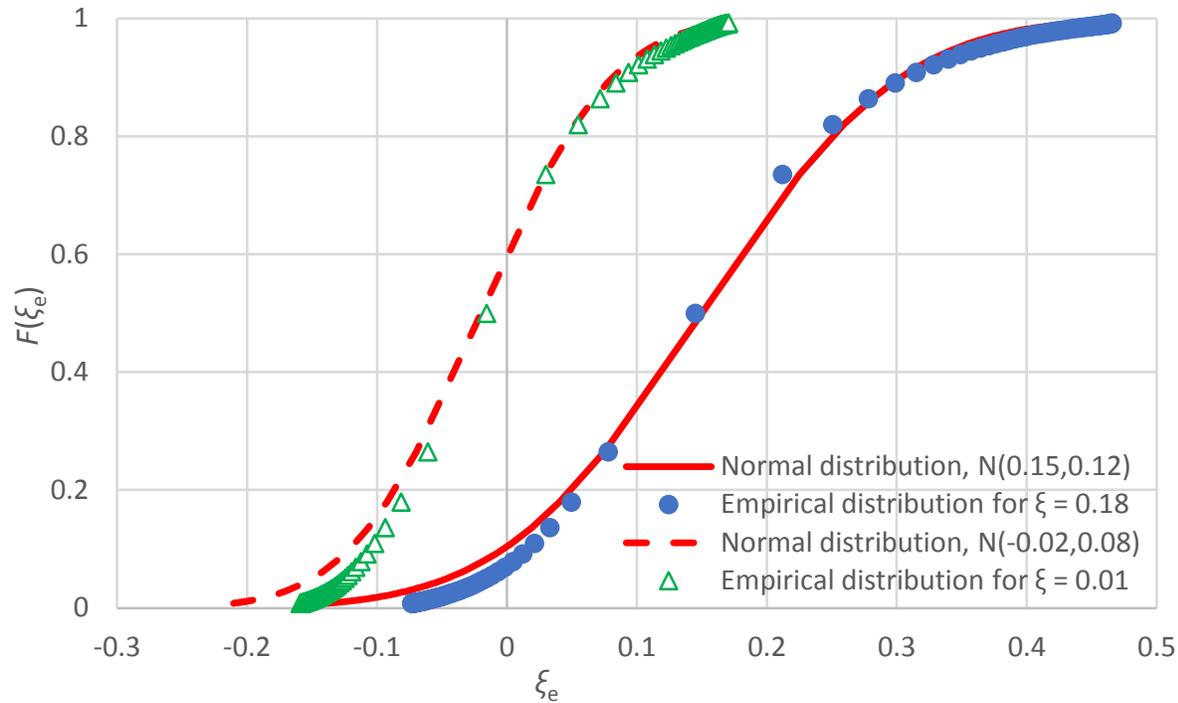


Figure 16

Review Only

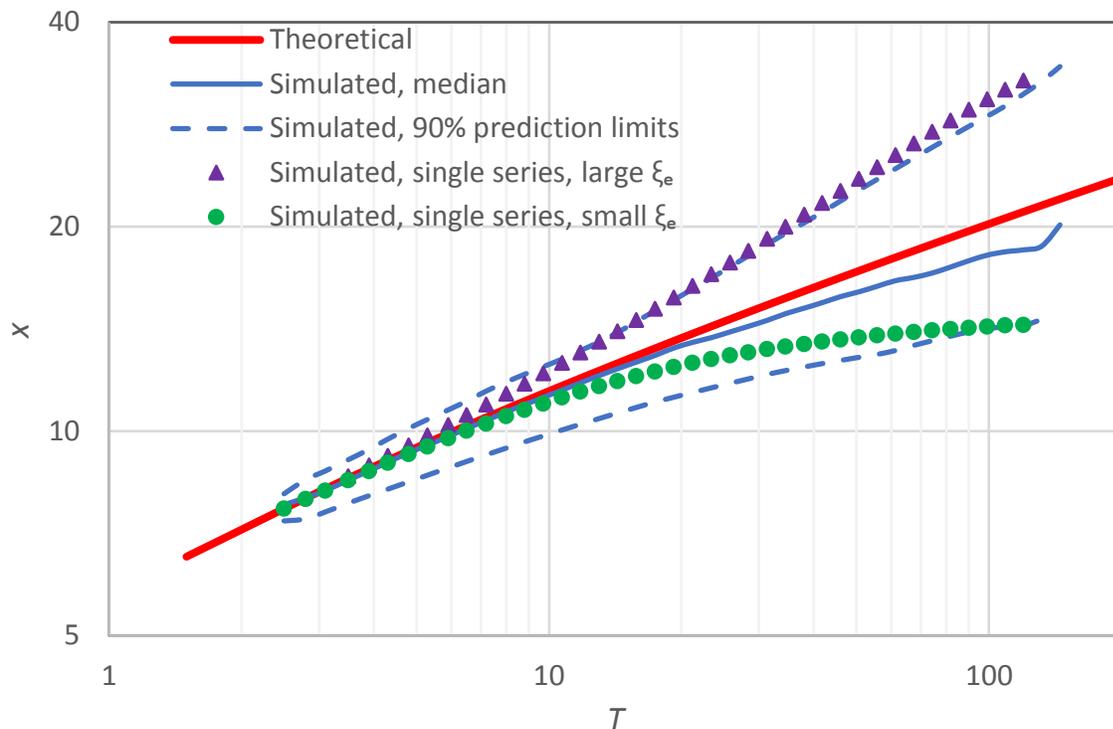


Figure 17

Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

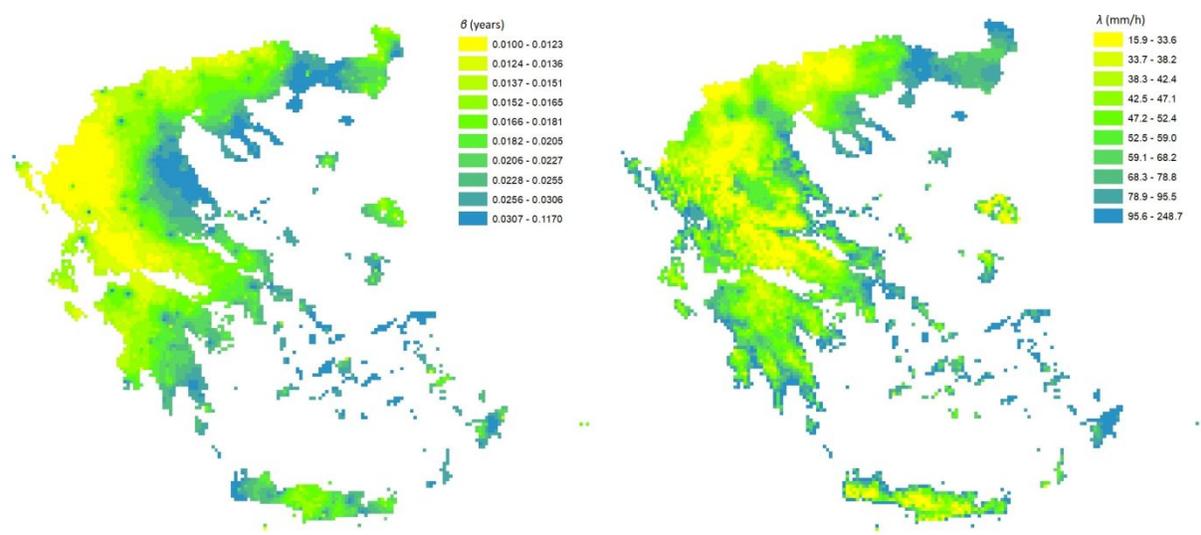


Figure 18

Peer Review Only

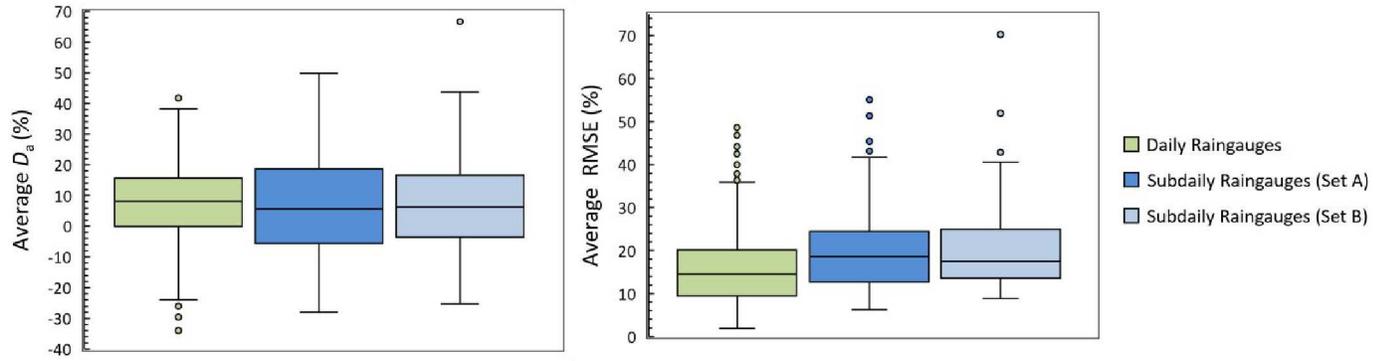


Figure 19

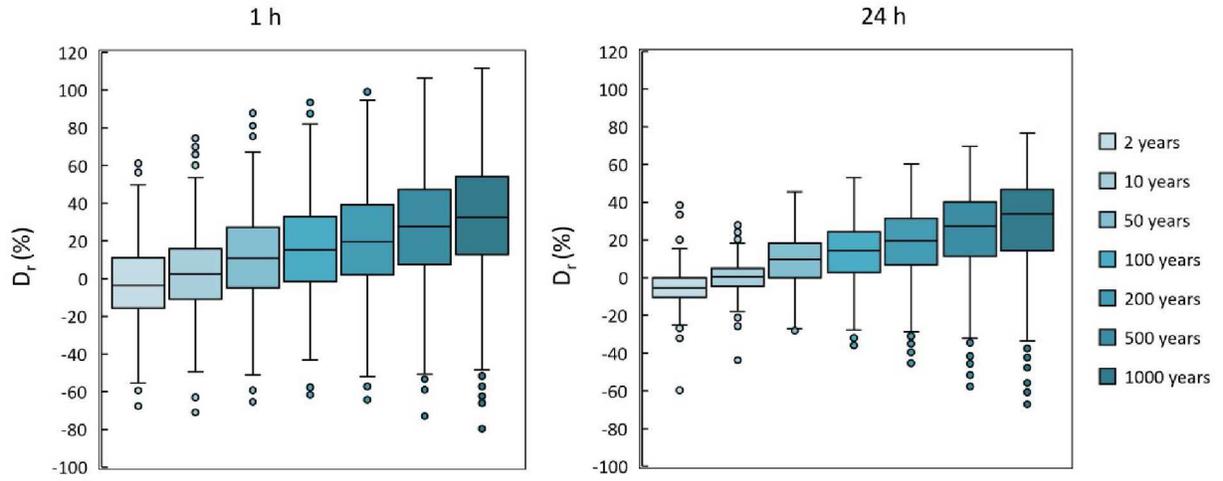
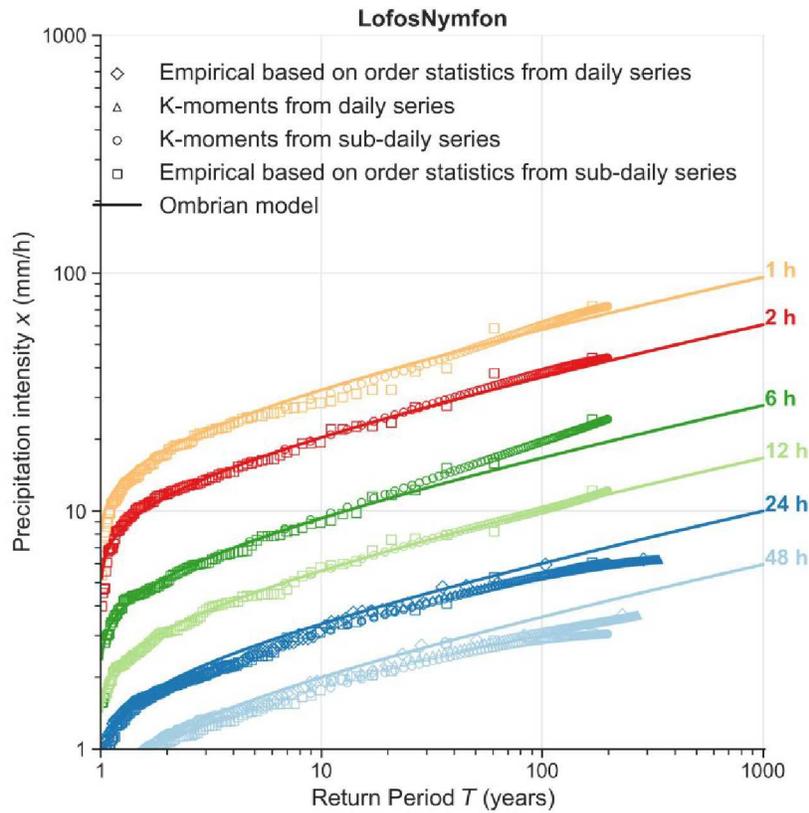
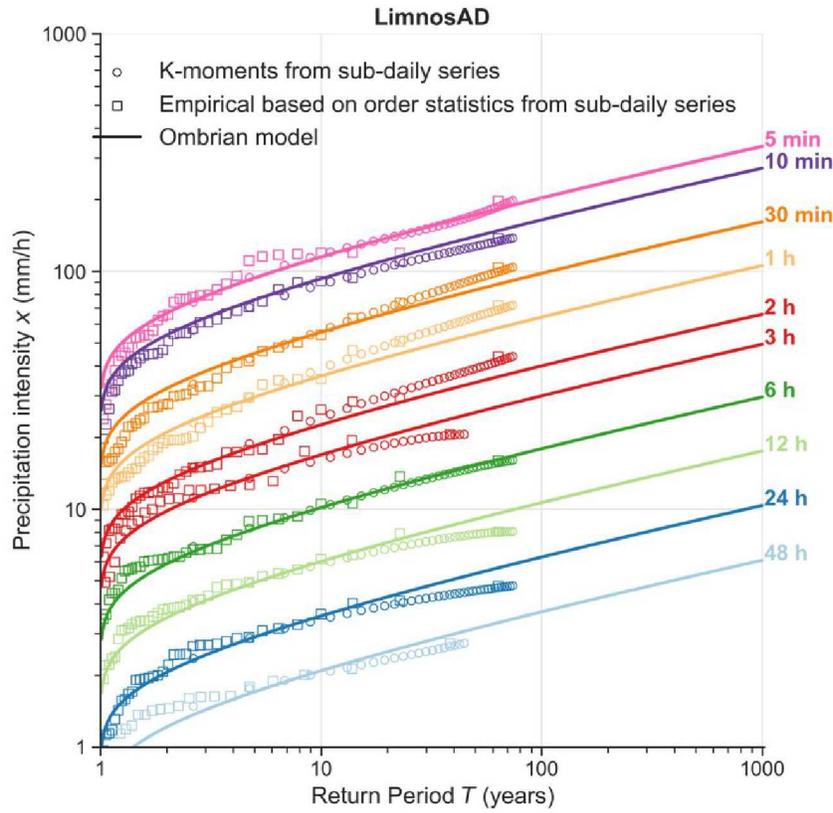


Figure 20

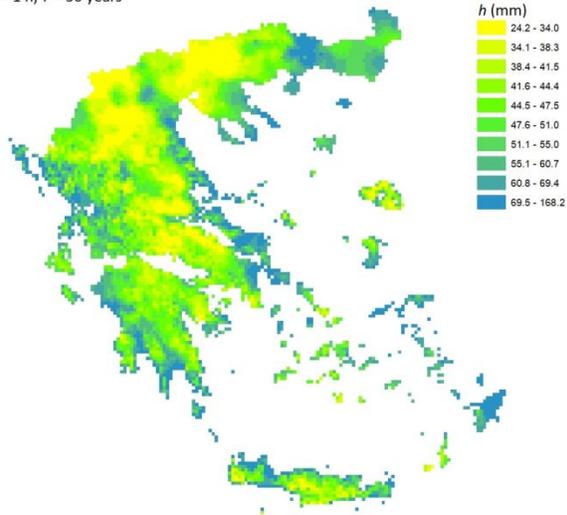


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

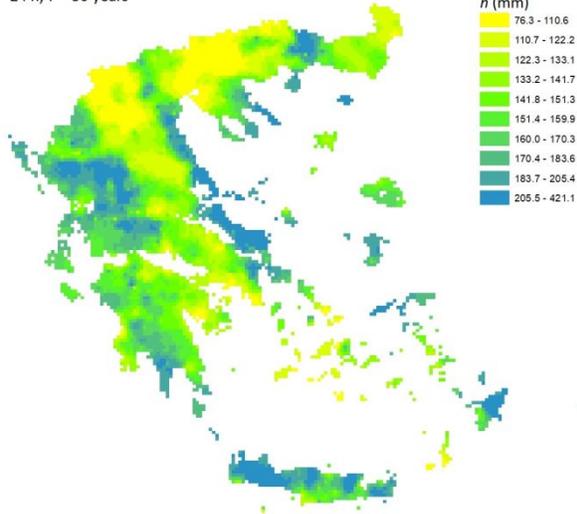
Figure 21

For Peer Review Only

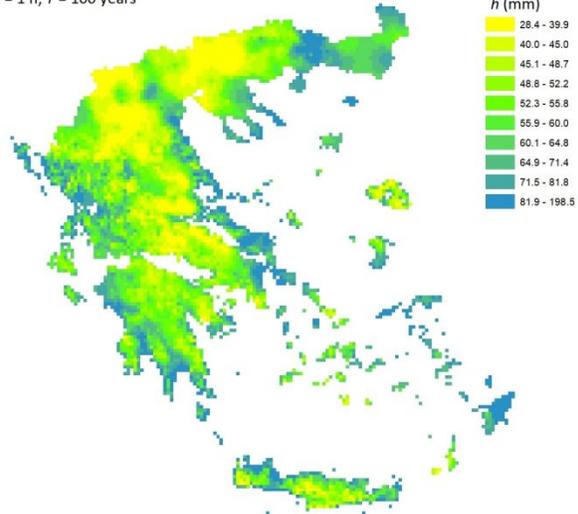
$k = 1 \text{ h}, T = 50 \text{ years}$



$k = 24 \text{ h}, T = 50 \text{ years}$



$k = 1 \text{ h}, T = 100 \text{ years}$



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

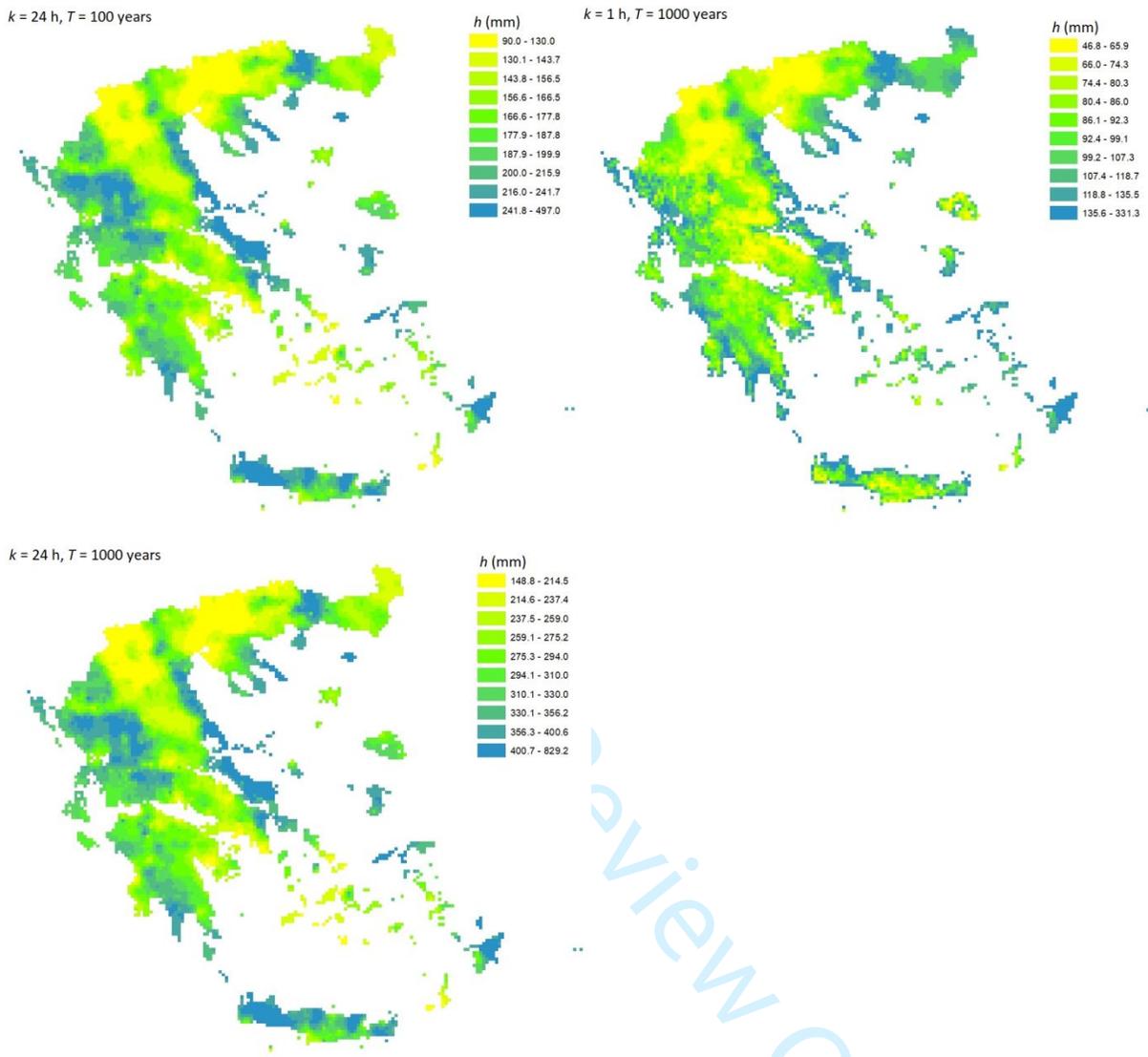


Figure 22

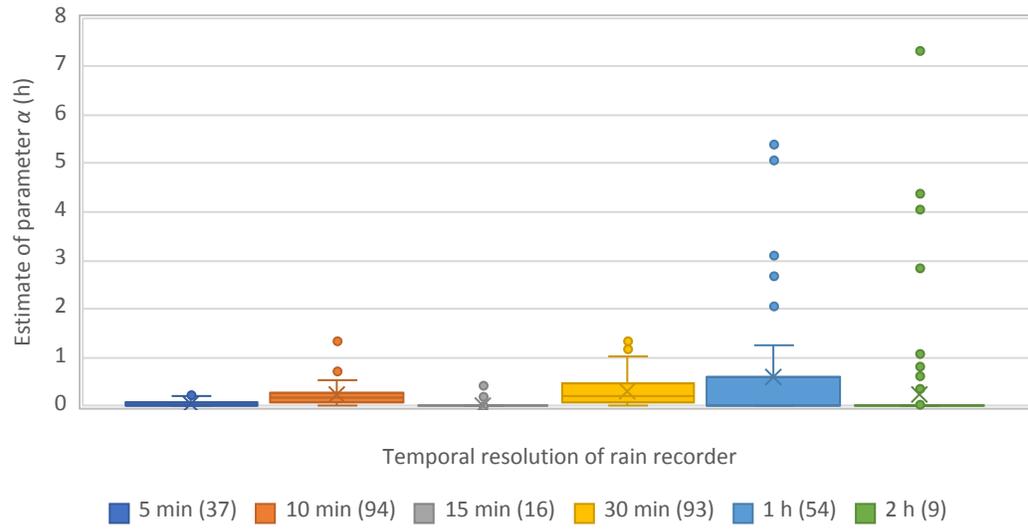


Figure A1

Peer Review Only

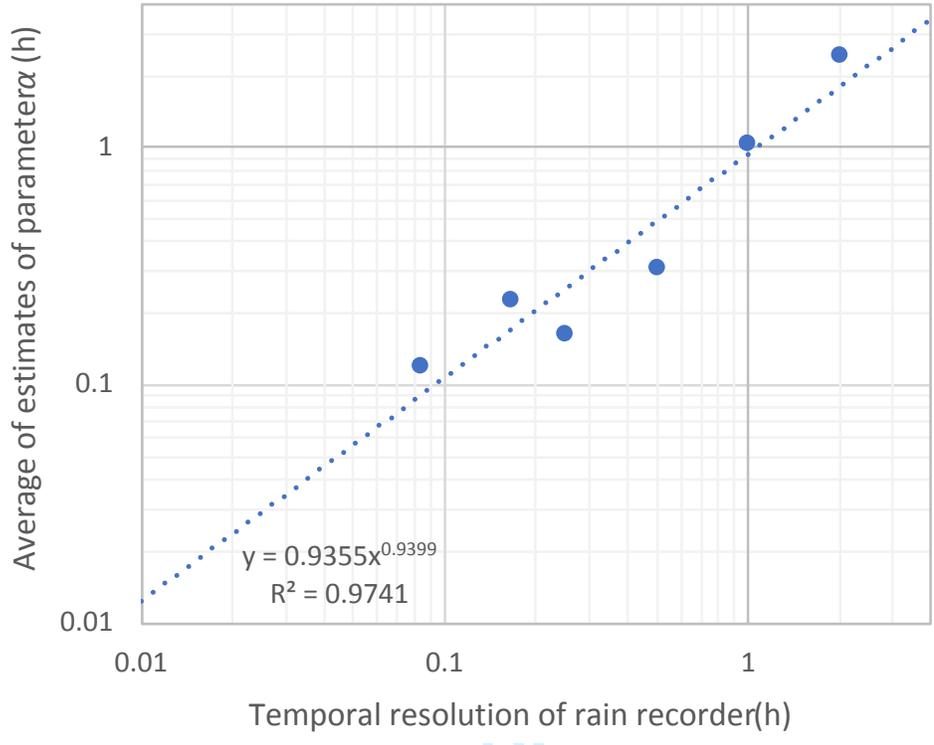


Figure A2.

Review Only

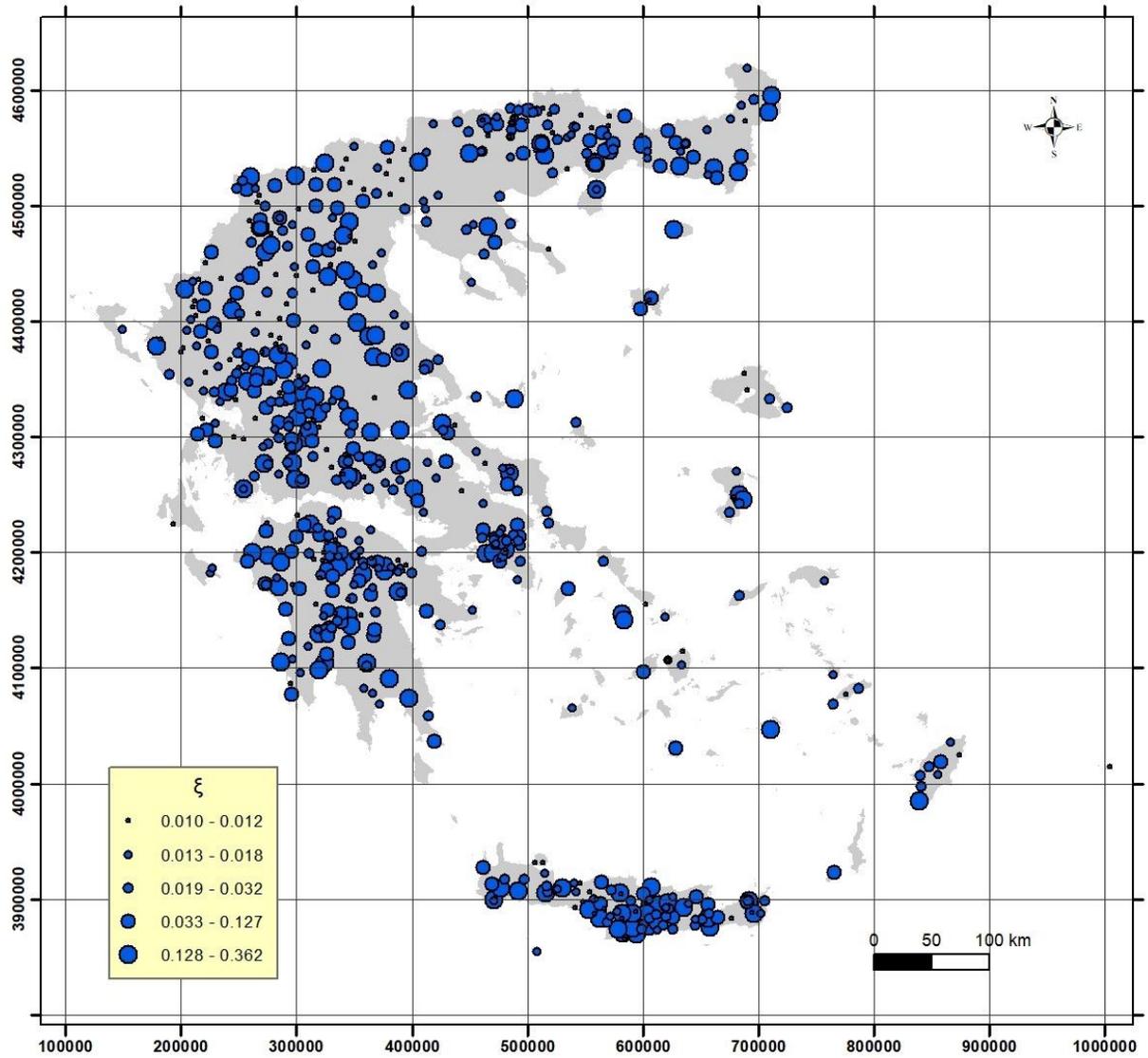


Figure A3

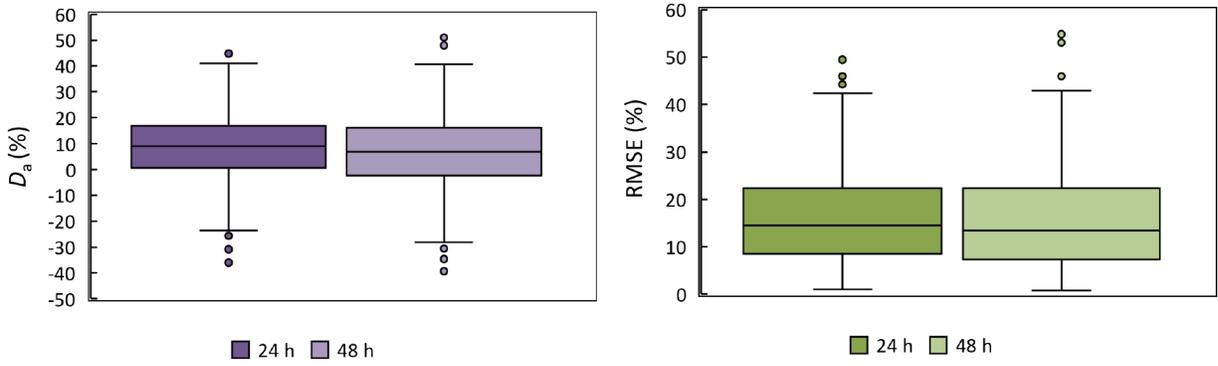


Figure B1

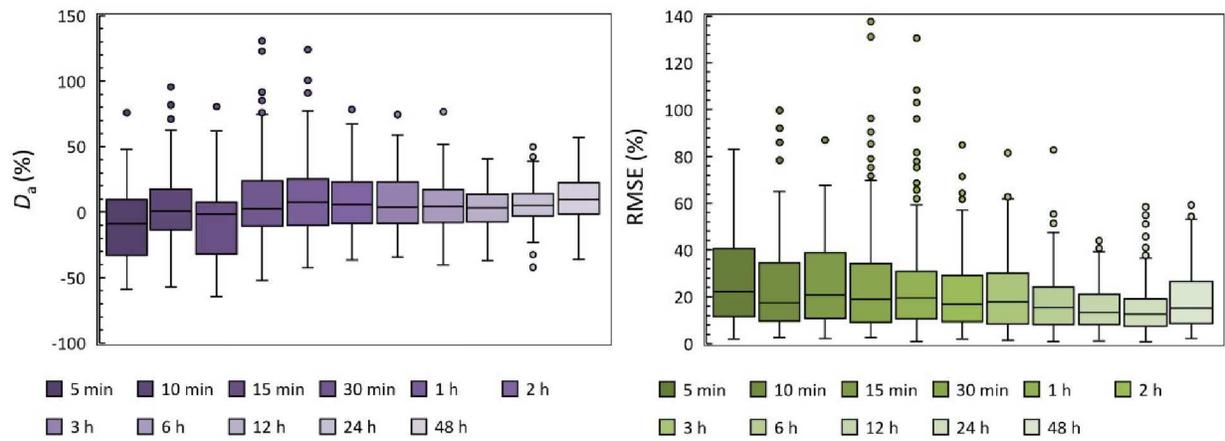


Figure B2