


Article

# Calibration of NSRP Models from Extreme Value Distributions

Davide Luciano De Luca \*  and Luciano Galasso

Department of Informatics, Modelling, Electronics and System Engineering, University of Calabria, 87036 Arcavacata di Rende (CS), Italy; luciano.galasso@unical.it

\* Correspondence: davide.deluca@unical.it; Tel.: +39-0984-4966-16

Received: 11 August 2019; Accepted: 14 October 2019; Published: 16 October 2019



**Abstract:** In this work, the authors investigated the feasibility of calibrating a model which is suitable for the generation of continuous high-resolution rainfall series, by using only data from annual maximum rainfall (AMR) series, which are usually longer than continuous high-resolution data, or they are the unique available data set for many locations. In detail, the basic version of the Neyman–Scott Rectangular Pulses (NSRP) model was considered, and numerical experiments were carried out, in order to analyze which parameters can mostly influence the extreme value frequency distributions, and whether heavy rainfall reproduction can be improved with respect to the usual calibration with continuous data. The obtained results were highly promising, as the authors found acceptable relationships among extreme value distributions and statistical properties of intensity and duration for the pulses. Moreover, the proposed procedure is flexible, and it is clearly applicable for a generic rainfall generator, in which probability distributions and shape of the pulses, and extreme value distributions can assume any mathematical expression.

**Keywords:** high-resolution continuous rainfall data; Neyman–Scott Rectangular Pulses model; extreme value distributions

## 1. Introduction

Long continuous high-resolution series of rainfall are more and more necessary for many hydrological purposes, in particular for applications related to small urban catchments.

However, for many locations of interest, the sample sizes of continuous data with a high resolution are usually short, or they are not available, and only annual maximum rainfall (AMR) series exist, but they are often not so long at the finest time scales (1 or 5 min).

These usual data deficiencies, together with the necessity to obtain perturbed time series, that can be representative of future rainfall fields at hydrological scales, induced the development of stochastic rainfall generators (SRGs), which are characterized by simple mathematical formulations and low computational costs, thus allowing for the quick generation of large ensembles of long precipitation time series.

A review of all these SRGs is provided in [1,2]. In particular, many works [3–9] highlighted the capacity of a specific class of SRGs, i.e., the Poisson cluster models, to satisfactorily reproduce observed summary statistics, such as mean, variance, skewness, proportion of dry/wet periods, and k-lag autocorrelation values, of the continuous rainfall process at several fine time scales. Poisson cluster models mainly comprise Neyman–Scott and Bartlett–Lewis families, which perform equally well [10]. Many applications of these models regarded England [5,8,11–13], Scotland [14], Belgium [7], Switzerland [15], Germany [9], Spain [16], Ireland [16], South Africa [17], New Zealand [18], Australia [19–21], Greece [22], Italy [23,24], United States [3,4,23,25–27], and Korean Peninsula [28].

On the contrary, other works stated the inability of these models to reconstruct extreme values, as they usually underestimate these rainfall heights at hourly and sub-hourly scales (see e.g., [29]). Consequently, many variants were proposed, related to (i) a change of model structure [9,30–32], (ii) inclusion of disaggregation techniques [33–36], (iii) model fitting with more information about the variability of precipitation [37]. However, many of these proposals implied an increasing parameterization, which limited their adoption, in particular for locations with very short samples.

In order to overcome these problems, censored versions [38] were introduced. They only model rainfall values over a prefixed threshold (i.e., there is only one more parameter to be estimated) and allow for a better reproduction of heavy portions of the hyetographs, because skewness (which usually assumes very high values at fine scales) and the proportion of dry/wet periods are both considered as unimportant. However, due to these latter aspects, these censored approaches are unsuitable for generating continuous time series as well, because a rainfall value below the censor is automatically set to zero, and then it is impossible to obtain a good reproduction of the proportion of dry/wet periods, which can be of interest for some applications.

From the above briefly discussed state-of-art for SRGs, two main key aspects emerge:

1. The sample size of continuous rainfall series with a high resolution plays a crucial role for a robust calibration of SRGs. In many locations, this sample size is very short and/or only AMR series are available.
2. The basic versions of SRGs usually underestimate extreme values at fine scales. Many attempts at improving the modeling were carried out, but they implied an unsuitable over parameterization for short sample sizes or an inability to reconstruct other features, such as dry/wet ratios.

In this context, the goal of this work is to answer to the following questions, which could arise from the above considerations:

- (a) Is it possible to calibrate a SRG by only using AMR series, which are usually more and more longer than continuous data with a high-resolution for many locations? Even, AMR series are the unique available data set for some locations.
- (b) If so, for a specific SRG are there some parameters which mostly influence the extreme value distributions?

In detail, the authors focused on the basic version of the single-site Neyman–Scott Rectangular Pulses model (NSRP [3,11,39]) and they carried out numerical experiments in order to reply to the previous questions.

The paper is organized as follows:

- In Section 2, the authors provided a brief overview of the NSRP model and described the adopted calibration and validation procedures, in order to derive possible relationships among extreme value distributions and NSRP parameters.
- The obtained results, from calibration and validations procedures, are illustrated in Section 3.
- Section 4 regards discussion about results and future developments.

## 2. Materials and Methods

### 2.1. Brief Theoretical Description of NSRP Model

In this work, the authors focused their attention on the single-site Neyman–Scott Rectangular Pulses (NSRP [3,11,39]) model, which is based on a clustering approach: rainfall is associated with clusters of rain cells making up storm events. In detail, the basic formulation is described below (see also Figure 1):

1. It is assumed that the number of storms is a homogeneous Poisson random variable. This means that the inter-arrivals,  $T_s$ , between the origins of two consecutive storms are independent and identically distributed, and they follow an exponential distribution:

$$P_{T_s}(t_s) = P[T_s \leq t_s] = 1 - e^{-\lambda \cdot t_s}, \quad (1)$$

where  $P[T_s \leq t_s]$  is the probability for the event  $T_s \leq t_s$  and  $1/\lambda$  represents the mean value for the inter-arrivals.

2. A number,  $M$ , of rain cells (also named bursts or pulses) is associated with each storm origin;  $M$  is usually considered as geometric or Poisson distributed. In this work, a geometric distribution is assumed and, with the aim of having at least one burst for each storm, the random variable  $C = M - 1$  is used, with  $E[C] = \theta - 1$ , so that  $E[M] = \theta$  and:

$$P_C(c) = \frac{1}{\theta} \cdot \left(1 - \frac{1}{\theta}\right)^c. \quad (2)$$

3. With respect to the associated storm origin, each burst origin occurs after a waiting time,  $W$ , which is assumed as an exponentially distributed variable with parameter  $\beta_W$  and  $E[W] = 1/\beta_W$ :

$$P_W(w) = P[W \leq w] = 1 - e^{-\beta_W \cdot w}. \quad (3)$$

4. Each burst has a rectangular shape, with an intensity  $I$  and a duration  $D$  which are both considered as exponential distributed, with parameters  $\beta_I$  and  $\beta_D$ , respectively, and  $E[I] = 1/\beta_I$ ,  $E[D] = 1/\beta_D$ :

$$P_I(i) = P[I \leq i] = 1 - e^{-\beta_I \cdot i}, \quad (4)$$

$$P_D(d) = P[D \leq d] = 1 - e^{-\beta_D \cdot d}. \quad (5)$$

5. The total precipitation intensity,  $Y(t)$ , at time  $t$  is then calculated as the sum of the intensities which are related to the active bursts at time  $t$ :

$$Y(t) = \int_0^{+\infty} I_{t-u}(u) dM(t-u), \quad (6)$$

where  $I_{t-u}(u)$  is the intensity of a single rectangular pulse at time  $u$  and  $M(t)$  is the counting process of the burst occurrences. Then, the aggregated process, i.e., the rainfall height  $H_j^{(\tau)}$  cumulated on the temporal  $\tau$  resolution and related to the time interval  $j$  is:

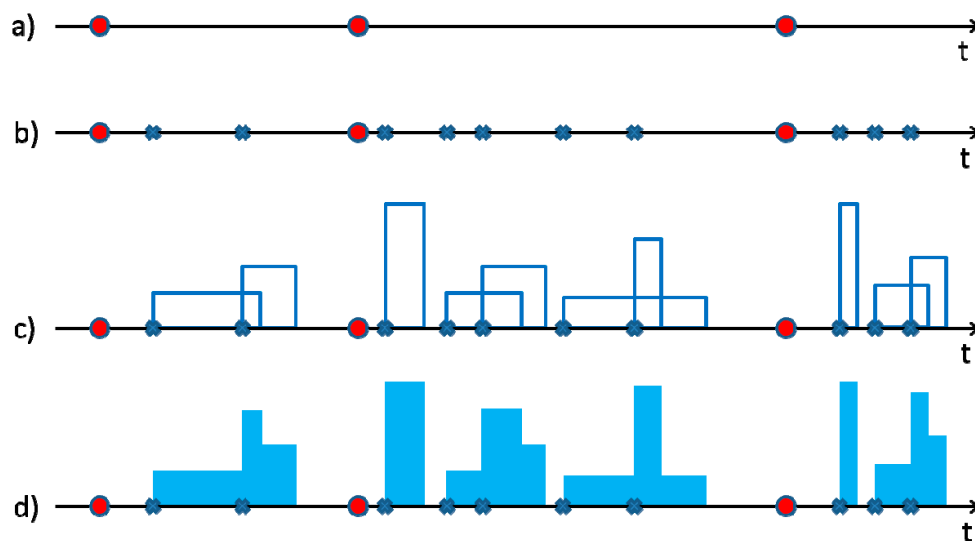
$$H_j^{(\tau)} = \int_{(j-1) \cdot \tau}^{j \cdot \tau} Y(t) \cdot dt. \quad (7)$$

Therefore, the basic version of a NSRP model has five parameters. Calibration is usually carried out by minimizing an objective function, which is defined as the sum of residuals (normalized or not) between the considered (by users) statistical properties of the observed data at chosen resolutions and their theoretical expressions. The statistical properties are typically referred to high-resolution continuous time series (e.g., 5 min rainfall time series): mean, variance, k-lag autocorrelation for  $H_j^{(\tau)}$  at several values of  $\tau$  [40] can be mentioned as examples. However, the sample size of these datasets is usually short (at most 15–20 years of records), and it is not very suitable for obtaining robust estimations, even more so when a specific five-parameter set is considered for each month or season, in order to take into account the seasonality of the process. Moreover, as already mentioned in

Section 1, currently many high-resolution time series from rain gauges usually regard only annual maximum rainfall (AMR) data; from these, it is impossible to obtain information about the above-listed statistical properties of the continuous process, and then it could seem unthinkable to obtain a NSRP calibration from AMR data.

In this context, the authors investigated the possibility/feasibility of calibrating the basic version of a NSRP model (to be preferred for reasons concerning parametric parsimony) from AMR series, also in order to improve the performance regarding reconstruction of extreme events, which are generally underestimated when continuous data are used (see e.g., [29]).

With this goal, Sections 2.1 and 2.2 regard the description of the adopted calibration and validation procedures, respectively.



**Figure 1.** Representation of the Neyman–Scott Rectangular Pulses (NSRP) stochastic process for at-site rainfall modeling: (a) occurrences of storms (Equation (1)); (b) for each storm, determination of the number of pulses (Equation (2)) and their temporal occurrences (Equation (3)); (c) estimation of intensity and duration for all the pulses (Equations (4) and (5)); (d) calculus of total intensity at time  $t$  (Equation (6)).

## 2.2. Calibration Procedure

Numerical experiments, which are below described, were carried out. Specifically, the authors used long synthetic continuous time series; this choice allows for obtaining more reliable estimations for possible relationships among (in this specific case) statistical properties of continuous process and those of the associated extreme values. Obviously, this analysis could be affected by high uncertainties if observed data were used, because of their limited sample sizes (mainly of the continuous time series).

First of all, as a simple hypothesis, the rainfall process was assumed as stationary, postponing to future works the modeling of seasonality (thus considering the process as cyclostationary) and trend analyses (for climate change contexts). This simple hypothesis allows for avoiding, in a framework of parametric parsimony, the use of a specific five-parameter set for each month or season. Moreover, although it could seem too strong and unrealistic, it is coherent with the extreme value theory [41], from which the probability distributions (EV1, EV2, EV3, GEV, etc.) are derived by considering, inside a fixed temporal block (usually a year), a sequence of values (whence the maximum is extracted) as independent and identically distributed random variables, i.e., without any seasonality feature.

In detail, the adopted calibration procedure can be explained in the following steps:

1. Each NSRP parameter was considered as a uniform random variable, with a specific variation range, according to [40,42] (see Table 1), and all the parameters were assumed as independent among them, for simplicity.

2. Then, 1000 five-parameter sets were generated with a Monte Carlo approach [43], and, for each set, a 500-year continuous time series of 1 min rainfall heights was simulated, from which AMR series, related to 5, 15, 30, and 60 min, were extracted. The choice of generating only one long realization is according to the ergodicity property for a stationary stochastic process [43,44].
3. Although analytical expressions of probability distributions do not exist for AMR series from the basic version of a NSRP model, they can be approximated with an EV1 distribution [45]:

$$P_H(h) = P[H \leq h] = e^{-e^{-\alpha(h-\varepsilon)}} \tag{8}$$

with  $\alpha > 0$ ,  $\varepsilon > 0$  and  $H$  is the annual maximum rainfall height at a specific time-resolution. The parameters  $\alpha$  and  $\varepsilon$  are related to the theoretical mean,  $\mu$ , and standard deviation,  $\sigma$ , of  $H$  in the following way [43]:

$$\mu = \varepsilon + \frac{n_e}{\alpha}, \tag{9}$$

$$\sigma = \frac{\pi}{\sqrt{6} \cdot \alpha}, \tag{10}$$

where  $n_e$  denotes the Euler constant, approximately equal to 0.5772. For each considered duration (5, 15, 30, and 60 min), the EV1 parameter set  $\underline{T} = (\alpha, \varepsilon)$  was estimated by using the maximum-likelihood (ML) method [43].

4. The final step regarded the investigation of possible relationships among each NSRP parameter from the set  $\underline{\Theta} = (\lambda, \theta, \beta_W, \beta_I, \beta_D)$  with the sets  $\underline{T}$ , estimated into Step 3 for the considered durations (5, 15, 30, and 60 min).

**Table 1.** Assumed parameter ranges of variation for the adopted basic version of NSRP model.

NSRP Parameter	Min. Value	Max. Value
$1/\lambda$ (h)	120	360
$\theta$ (-)	2	10
$1/\beta_W$ (h)	5	24
$1/\beta_I$ (mm/h)	5	15
$1/\beta_D$ (h)	0.1	0.6

Concerning Step 4, several performance measures can be adopted (e.g., see [46]). In this case, the authors used the minimization of the mean square error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2, \tag{11}$$

where  $N$  is the number of pairs regarding observed and modeled data,  $y_i$  is the  $i$ -th observed value and  $y_i^*$  is the  $i$ -th modeled value. In this work  $N = 1000$ , while  $y_i$  and  $y_i^*$  correspond to, respectively, the sample estimates and the reconstructed values (by using relationships with the sets  $\underline{T}$ ) for each NSRP parameter.  $MSE$  is usually associated with the coefficient of determination,  $R^2$ :

$$R^2 = \left[ \frac{\sum_{i=1}^N (y_i - \bar{y})(y_i^* - \bar{y}^*)}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^N (y_i^* - \bar{y}^*)^2}} \right]^2 \tag{12}$$

where  $\bar{y}$  and  $\bar{y}^*$  are the mean values for  $y_i$  and  $y_i^*$ , respectively. The ranges of variation are  $(0, +\infty)$  for  $MSE$  and  $(0, 1)$  for  $R^2$ . Moreover, low values of  $MSE$  are associated with high values of  $R^2$  and vice versa.

In an equivalent way, a user can carry out the maximization of the Nash–Sutcliffe Efficiency coefficient (*NSE*):

$$NSE = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{MSE}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}, \quad (13)$$

that can be clearly rewritten, as reported in Equation (13), as a linear function of *MSE*. The range of variation for *NSE* is  $(-\infty, 1)$  and, obviously, low values of *MSE* are associated with high values of *NSE* and vice versa.

Minimization of other performance measures, such as root mean square error (*RMSE*) or mean absolute error (*MAE*) do not produce significant difference from adoption of *MSE* (see [46]), because of their particular mathematical structure. In fact, *RMSE* is simply the root square of *MSE*, while *MAE* adopts the sum of absolute values of residuals, instead of their squares: both measures present the same minimum of *MSE*.

### 2.3. Validation Procedure

In this work, validation is aimed at testing whether NSRP series from (observed or generated) EV1 data provide extreme value distributions which are compatible with the starting ones.

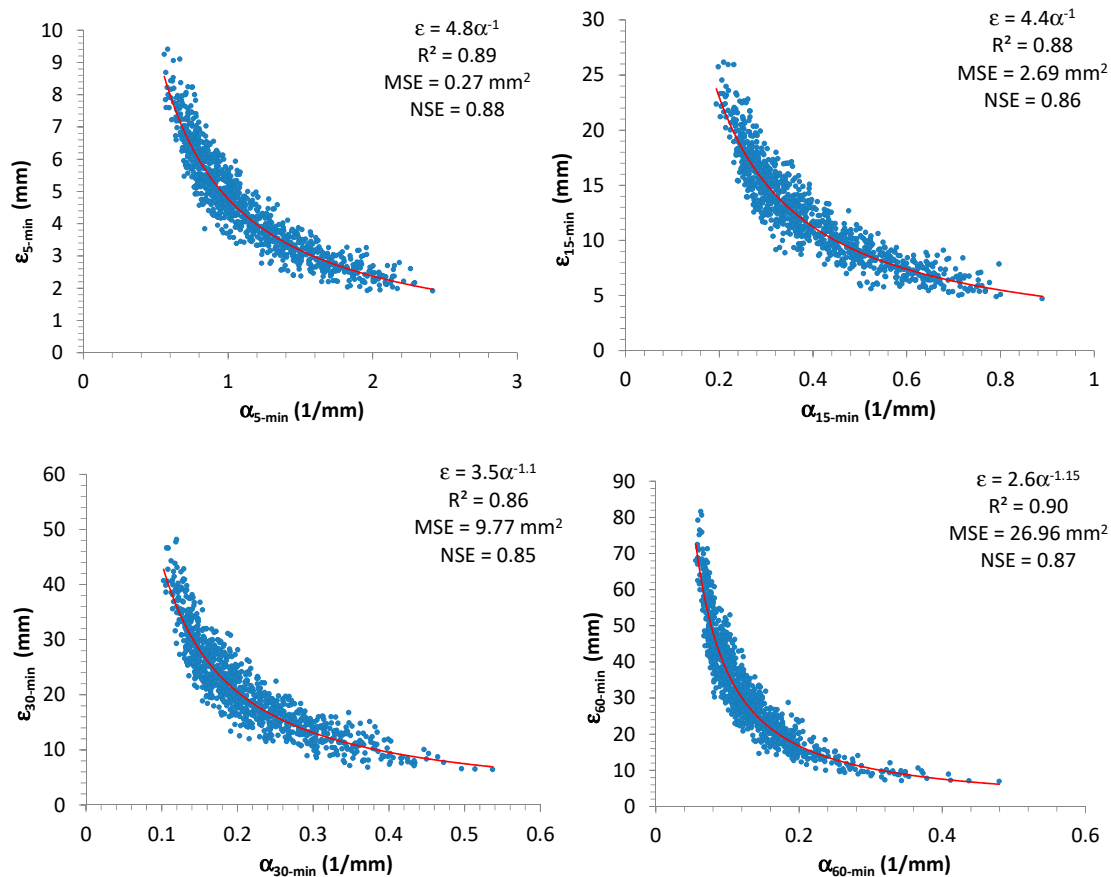
In detail, the proposed procedure can be validated by considering some 500-year AMR synthetic samples, derived from NSRP generations, and some observed (and exhibiting a clear EV1 behavior) AMR time series, with a resolution between 5 and 60 min. In detail, for each AMR series, the following steps can be carried out:

- Estimation of EV1 parameters (with the ML technique or others).
- Generation of NSRP parameters, by considering the investigated relationships. If no relationship is found for one (or more) NSRP parameter, this is considered as a random uniform variable, with the variation range in Table 1, and then, for each (virtual or real) site, 1000 NSRP five-parameter sets are generated, in which only the parameters with a robust relationship with EV1 distribution present the same value, while the others vary in a uniform way.
- Reproduction of a 500-year continuous 1 min data series, for each generated NSRP five-parameter set and consequent determination of the associated AMR series.
- Evaluation of frequency distributions for each EV1 parameter, from the whole ensemble of 1000 NSRP generations, and comparison with the corresponding ML estimates of the starting sample. Analogous comparisons can be carried out by considering mean and standard deviation values.

## 3. Results

### 3.1. Calibration

Analyzing the synthetic samples, Figure 2 shows the scatter plots between  $\alpha$  and  $\varepsilon$ , which were estimated by using the maximum-likelihood (ML) method [43], for each considered time-resolution (5, 15, 30, and 60 min). From Figure 2 it is clear, as expected, that there is an inverse dependence between these two EV1 parameters (see Equation (9) and impose, for simplicity, a fixed value of  $\mu$ ). Moreover, in Figure 2, the authors also reported the best regression equations representing the relationships between  $\varepsilon$  and  $\alpha$  (expressed in this case as a power formula), estimated for each duration, together with the corresponding values of performance measures, mentioned in Section 2.2. Due to the obtained high values for  $R^2$  (always greater than or equal to 0.86) and *NSE* (always greater than or equal to 0.85), i.e., close to 1, the estimated formulas clearly constitute a good fitting. Furthermore, the relationships among the various  $\alpha$  parameters are illustrated in Figure 3.



**Figure 2.** Scatter plots for estimated EV1 parameters  $\alpha$  and  $\epsilon$ , related to annual maximal rainfall (AMR) data, which were obtained from 500-year NSRP time series with a resolution of 1 min.

The authors investigated the possible relationships among NSRP and EV1 parameters by mainly focusing on  $\alpha$ , because  $\epsilon$  is strictly connected to  $\alpha$ , as illustrated in Figure 2.

This analysis highlighted that  $\lambda$ ,  $\beta_W$ , and  $\theta$  seem to not influence the extreme values distribution at high resolutions. As examples, scatter plots among  $\alpha$  estimated from 15-min AMR series and  $\lambda$ ,  $\beta_W$ ,  $\theta$  are reported in Figure 4.

These results can be justified as follows:

- It is expected that the waiting time between two consecutive storm occurrences could imply effects on monthly and annual scale, but not at finer resolutions (in particular, sub-hourly). In fact, a heavy rainfall event can fully develop itself, without being affected by a large or short waiting time with the previous and the successive storms. Greater values for  $1/\lambda$  could determine large dry periods and consequently low aggregated rainfall values only at monthly and yearly scales.
- Similar considerations can be made for  $\theta$  and  $\beta_W$ . In fact, the number of bursts (and consequently their occurrences) for each storm could mainly influence the aggregated process (see Equation (7) and Figure 1) at coarser resolutions (from daily or multi-daily), but not at finer scales (hourly and sub-hourly), especially if durations of the pulses are always sub-hourly on average, as assumed in this work (see Table 1).

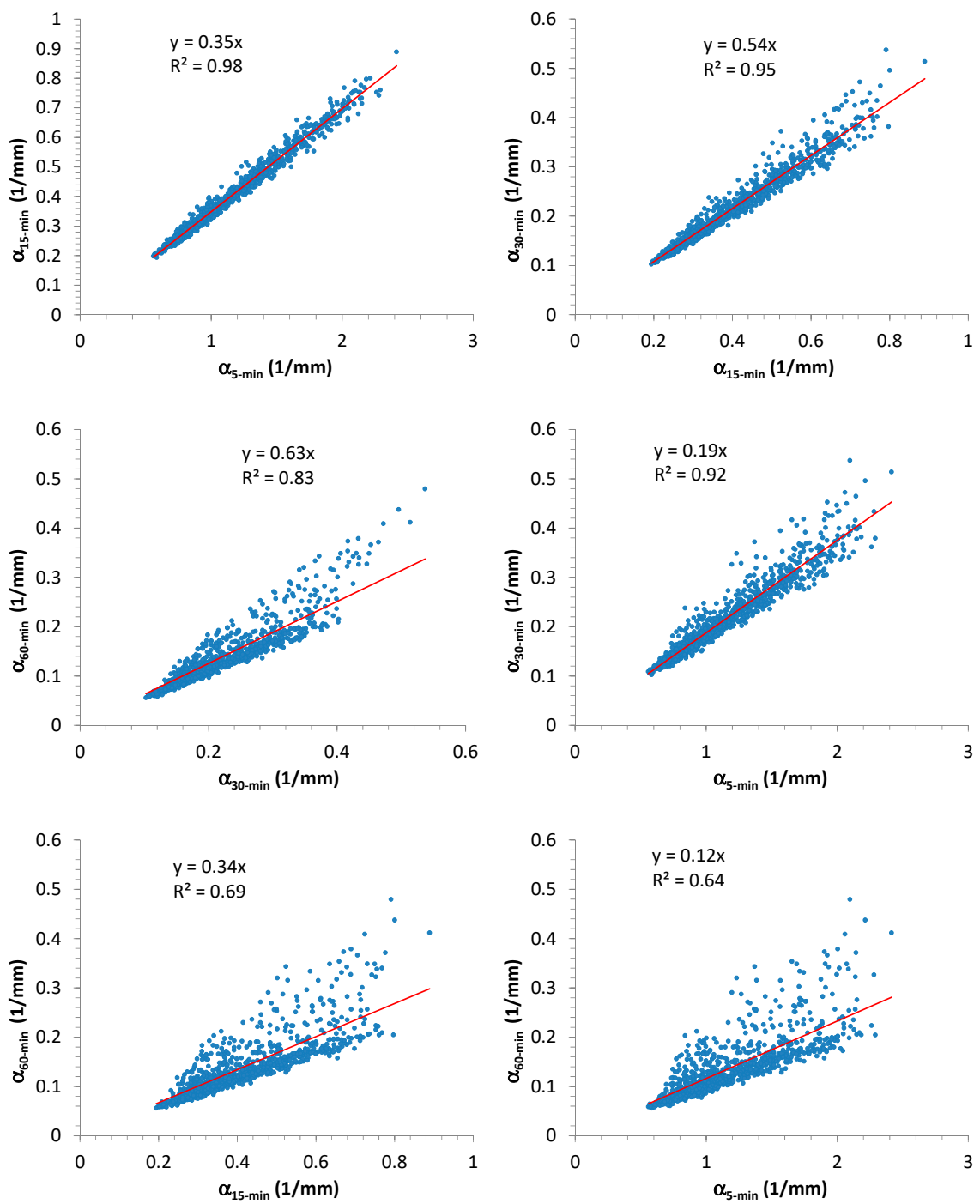
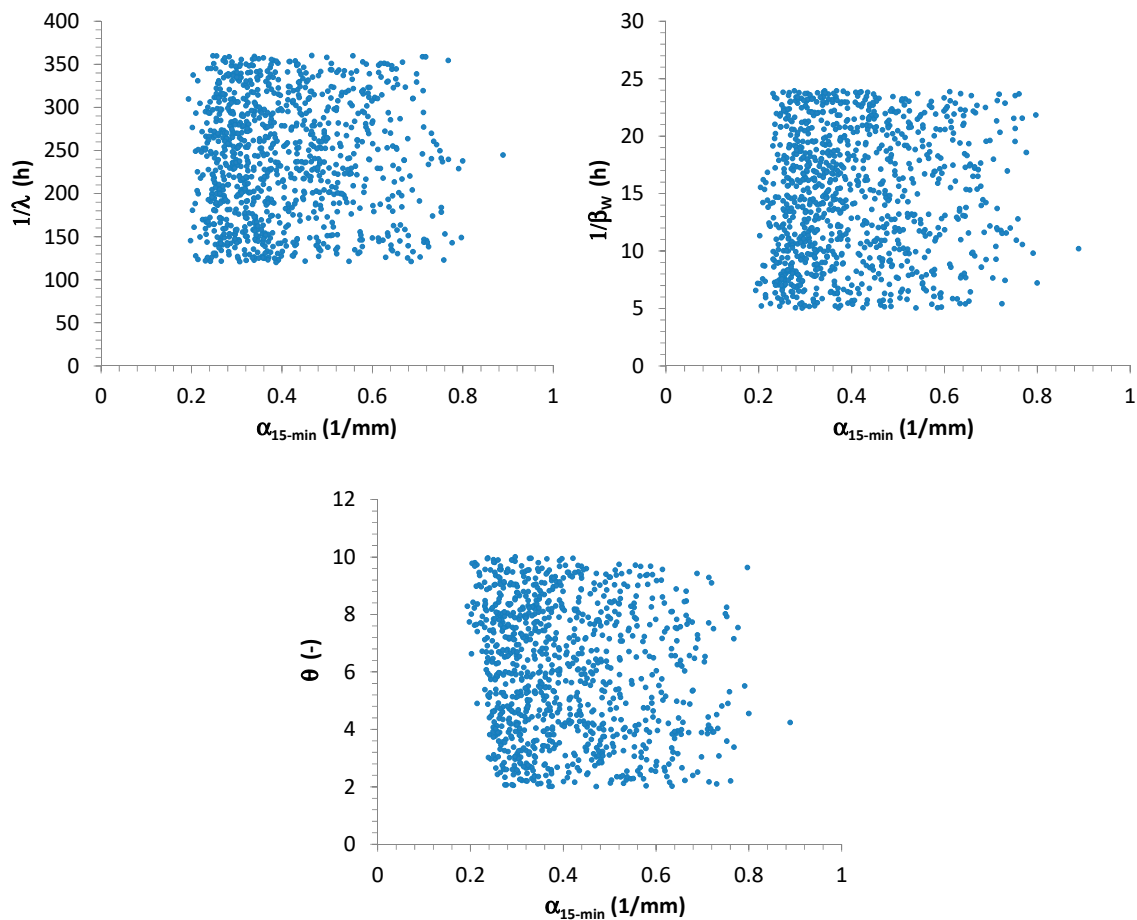


Figure 3. Scatter plots showing the relationships among the various  $\alpha$  parameters.





**Figure 4.** Scatter plots among  $\alpha$ , which were estimated from 15-min synthetic AMR series, and  $\lambda$ ,  $\beta_w$ , and  $\theta$ .

On the contrary,  $\beta_I$  and  $\beta_D$  show a more evident influence for EV1 distributions of high-resolution rainfall data.

In particular, the scatter plots in Figure 5 remark a significant interconnection between  $\alpha$  and  $\beta_I$ , mainly for 5 and 15 min. Considering that observed 15-min AMR series are more available than 5-min ones in many locations, the authors chose to use the relationship between  $\beta_I$  and  $\alpha_{15\text{min}}$  for the successive elaborations:

$$\frac{1}{\beta_I} = b_0 \cdot \alpha_{15\text{min}}^{b_1} \quad (14)$$

Minimization of *MSE* (and simultaneous maximization of  $R^2$ ) and maximization of *NSE* provided  $b_0 = 4.1$ ,  $b_1 = -0.89$ , with  $MSE = 1.2 \text{ (mm}^2/\text{h}^2)$ ,  $R^2 = 0.88$ , and  $NSE = 0.86$ . From Equation (14) it is clear that the authors focused on  $1/\beta_I$ , and this implies that *MSE* is expressed in  $\text{mm}^2/\text{h}^2$ . In Figure 5, the estimated coefficients for Equation (14), together with the associated values of *MSE*,  $R^2$ , and *NSE* are also reported for the other durations. From result analysis, it is evident that adoption of 15-min AMR series produces similar good performances of 5-min AMR series. On the contrary, the use of 60-min AMR series (if it were the unique available data set for a specific site) could not provide a good reconstruction for  $\beta_I$  ( $R^2 = 0.58$  and  $NSE = 0.60$ ).

As regards  $\beta_D$ , the best relationship was found by linking with estimates of  $\alpha$  for 15 and 60 min (Figure 6):

$$\frac{1}{\beta_D} = c_0 \cdot \alpha_{15\text{min}}^{c_1} \cdot \alpha_{60\text{min}}^{c_2} \quad (15)$$

with  $c_0 = 0.015$ ,  $c_1 = 2.3$ ,  $c_2 = -2.5$ ; the associated values for the performance measures were  $MSE = 0.007 \text{ h}^2$ ,  $R^2 = 0.68$ , and  $NSE = 0.67$ . From Equation (15), it is clear that the authors focused on  $1/\beta_D$ , and this implies that  $MSE$  is expressed in  $\text{h}^2$ . From a visual analysis of Figure 6, the performance is clearly not comparable with that for  $\beta_I$ , however it was considered as acceptable for the successive elaborations, because the values of  $R^2$  and  $NSE$  are greater than 0.5 [47].

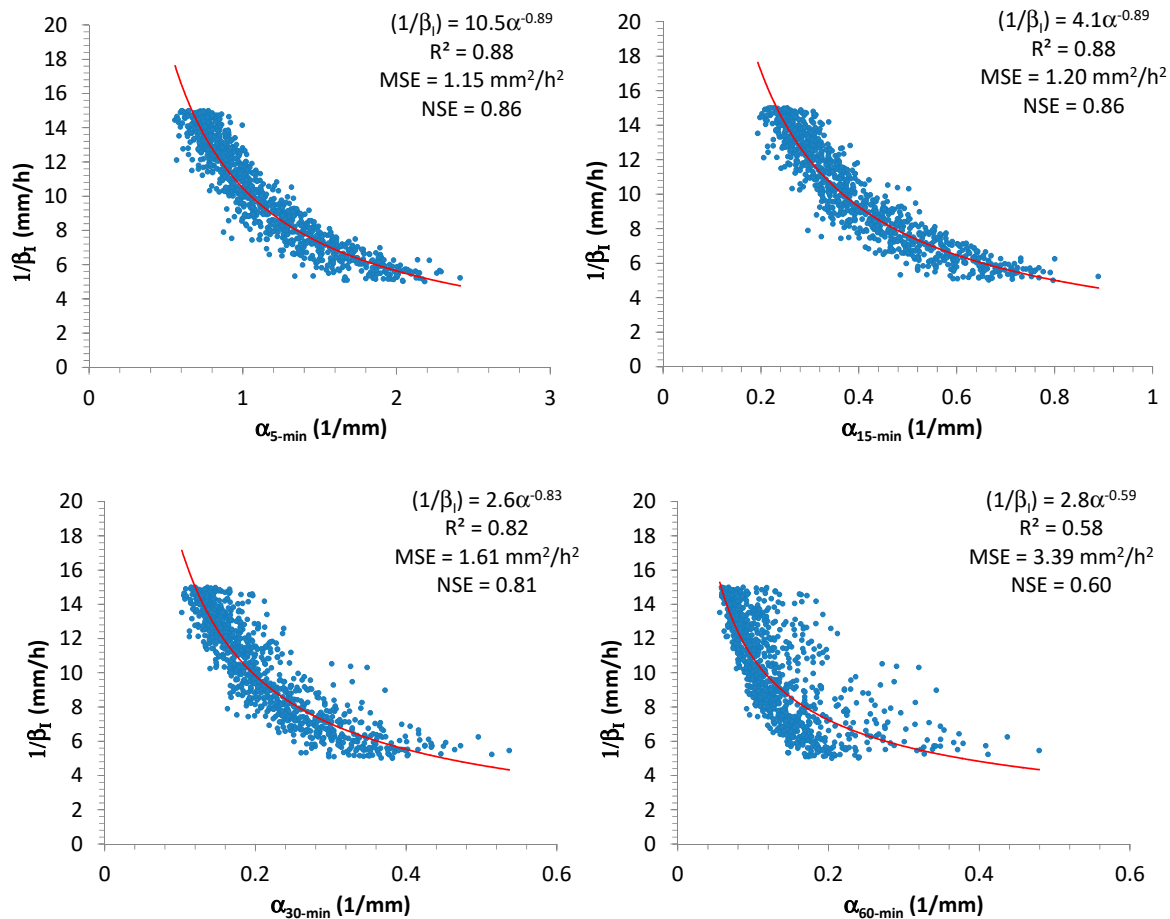


Figure 5. Scatter plots showing the relationships among  $\beta_I$  and the various  $\alpha$  parameters.

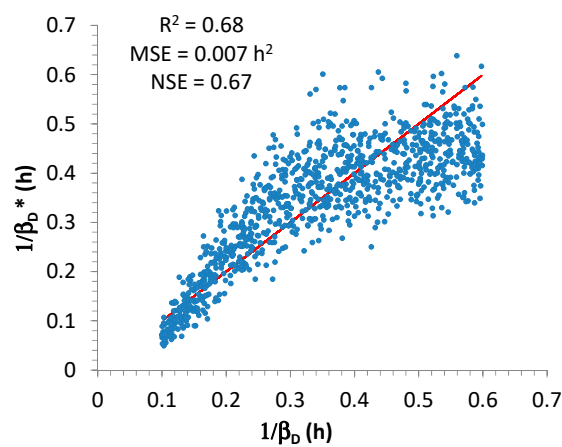


Figure 6. Scatter plots showing the relationship between starting  $\beta_D$  (horizontal axis) and estimated  $\beta_D$  (indicated with the superscript \* and reported on vertical axis).

### 3.2. Validation

First of all, the authors validated the proposed procedure by focusing on:

- two 500-year synthetic samples with a resolution of 1 min, generated from NSRP with parameter sets reported in Table 2, from which 5, 15, 30, and 60-min AMR series were derived (Figures 7–10), and;
- observed 15-min and 60-min AMR series, related to Cosenza rain gauge (Southern Italy), having a sample size  $M = 29$  years for 15-min AMR and  $M = 63$  years for 60-min AMR, and characterized by a clear EV1 behavior, as shown in EV1 probabilistic plots (Figure 11).

**Table 2.** NSRP parameters for two synthetic samples analyzed in the validation step.

NSRP Parameter	$1/\lambda$ (h)	$\theta$ (-)	$1/\beta_W$ (h)	$1/\beta_I$ (mm/h)	$1/\beta_D$ (h)
Synthetic sample 1	240	6	8	8	0.25
Synthetic sample 2	240	6	8	12	0.25

For each (virtual or real) site, as theoretically explained in Section 2.3:

- EV1 parameters were estimated with ML technique (see Tables 3 and 4) for the associated AMR series.
- 1000 NSRP five-parameter sets were generated, in which  $\lambda$ ,  $\beta_W$ , and  $\theta$  vary in a uniform way within the ranges reported in Table 1, while  $\beta_I$  and  $\beta_D$  values are always the same, calculated from Equations (14) and (15), by using the estimated  $\alpha$  parameters from the previous step (see Tables 5 and 6).
- For each generated NSRP five-parameter set, a 500-year continuous 1 min data series was reproduced, from which the associated 5, 15, 30, and 60-min AMR series were extracted, together with the correspondent EV1 parameter estimations.
- From the previous step, frequency distributions for  $\alpha$ ,  $\varepsilon$  were obtained and compared with the ML estimates related to the specific (virtual or real) starting data series (Figures 7 and 8). Comparisons were also carried out in EV1 probabilistic plots (Figures 9 and 10), among starting AMR series, their EV1 theoretical curves, and EV1 curves associated with (i) median values for  $\alpha$  and (ii)  $\varepsilon$  values which were calculated with the relationships reported in Figure 2.

**Table 3.** EV1 parameters for two synthetic samples analyzed in the validation step.

EV1 Parameter	5 min		15 min		30 min		60 min	
	$\alpha$ (1/mm)	$\varepsilon$ (mm)	$\alpha$ (1/mm)	$\varepsilon$ (mm)	$\alpha$ (1/mm)	$\varepsilon$ (mm)	$\alpha$ (1/mm)	$\varepsilon$ (mm)
Synthetic sample 1	1.30	3.65	0.45	9.68	0.25	16.21	0.15	22.85
Synthetic sample 2	0.85	5.40	0.30	14.40	0.17	24.06	0.10	33.85

**Table 4.** EV1 parameters for AMR series of Cosenza rain gauge.

EV1 Parameter	15 min		60 min	
	$\alpha$ (1/mm)	$\varepsilon$ (mm)	$\alpha$ (1/mm)	$\varepsilon$ (mm)
Cosenza rain gauge	0.27	13	0.14	17.2

**Table 5.** NSRP parameters for two synthetic samples derived from EV1 parameters in Table 3.

NSRP Parameter	$1/\lambda$ (h)	$\theta$ (-)	$1/\beta_W$ (h)	$1/\beta_I$ (mm/h)	$1/\beta_D$ (h)
Synthetic sample 1	[120; 360]	[2; 10]	[5; 24]	8.43	0.26
Synthetic sample 2	[120; 360]	[2; 10]	[5; 24]	12.10	0.27

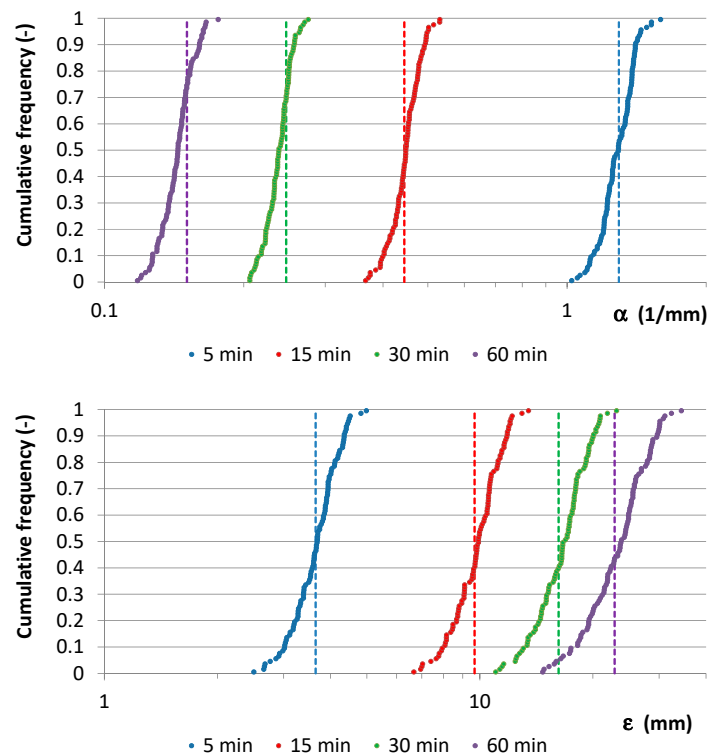
**Table 6.** NSRP parameters for Cosenza rain gauge derived from EV1 parameters in Table 5.

NSRP Parameter	$1/\lambda$ (h)	$\theta$ (-)	$1/\beta_W$ (h)	$1/\beta_I$ (mm/h)	$1/\beta_D$ (h)
Cosenza rain gauge	[120; 360]	[2; 10]	[5; 24]	13.15	0.10

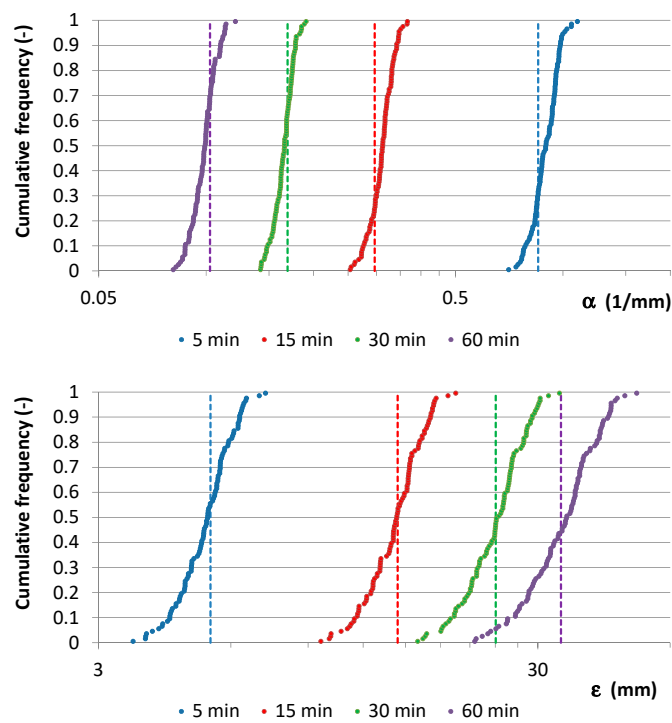
From Figures 7 and 8, it is clear that the starting value for  $\varepsilon$  is about the median value of the cumulative frequency distribution for each analyzed duration, while the starting value for  $\alpha$  is always inside the interquartile range. Moreover, EV1 probabilistic plots (Figures 9 and 10) show that median EV1 curves from NSRP generation and EV1 theoretical curves from synthetic AMR series perform equally well.

Similar results were obtained for Cosenza rain gauge. In detail, median EV1 curves from NSRP generation reconstruct in an acceptable way the sample data, with respect to EV1 theoretical curves, which were directly estimated from the observed AMR series (Figure 11).

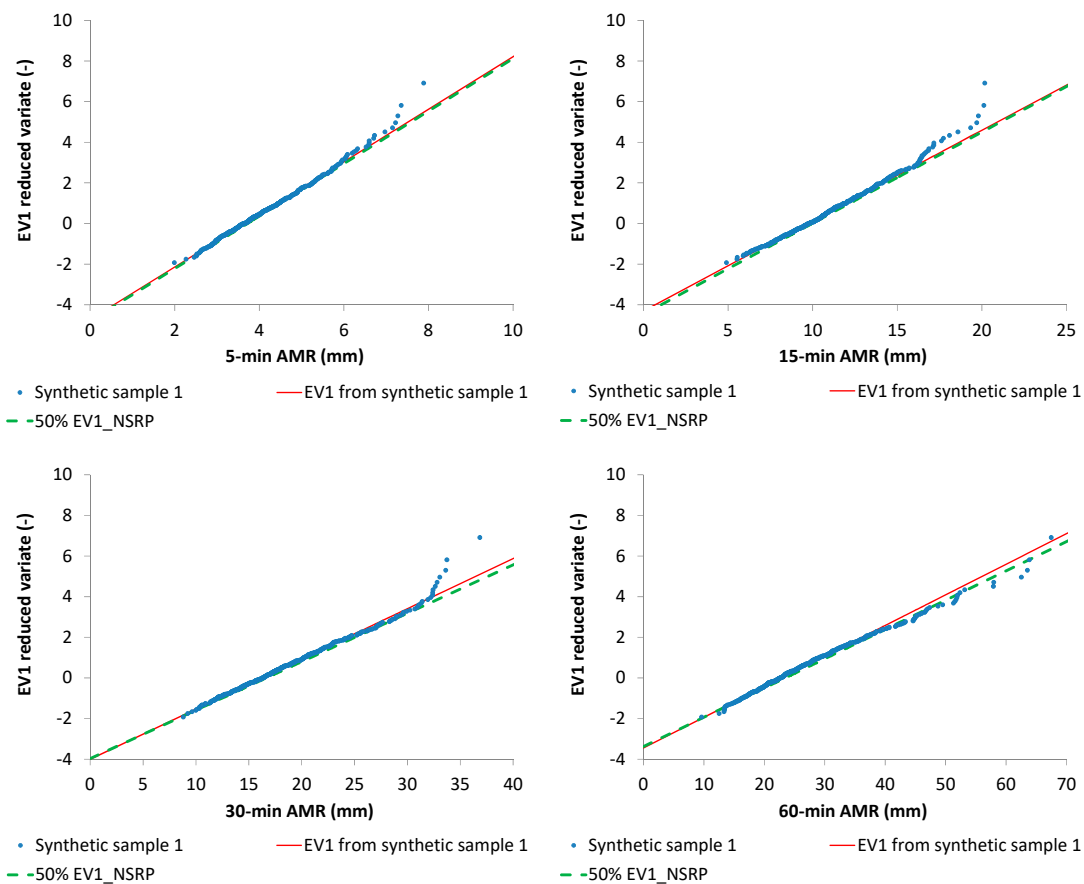
The performance of EV1 distributions, obtained (i) directly with ML parameter estimation from synthetic samples and Cosenza data set and (ii) from NSRP generation, were analyzed by comparison with the sample frequency values, also in this case in terms of *MSE*, *R<sup>2</sup>*, and *NSE* (see Table 7). The obtained results showed that reconstruction of extreme events from the proposed calibration of a NSRP model provides similar performances with respect to the theoretical frequency distributions which were directly estimated from AMR data: *R<sup>2</sup>* and *NSE* always assumed values greater than 0.9.



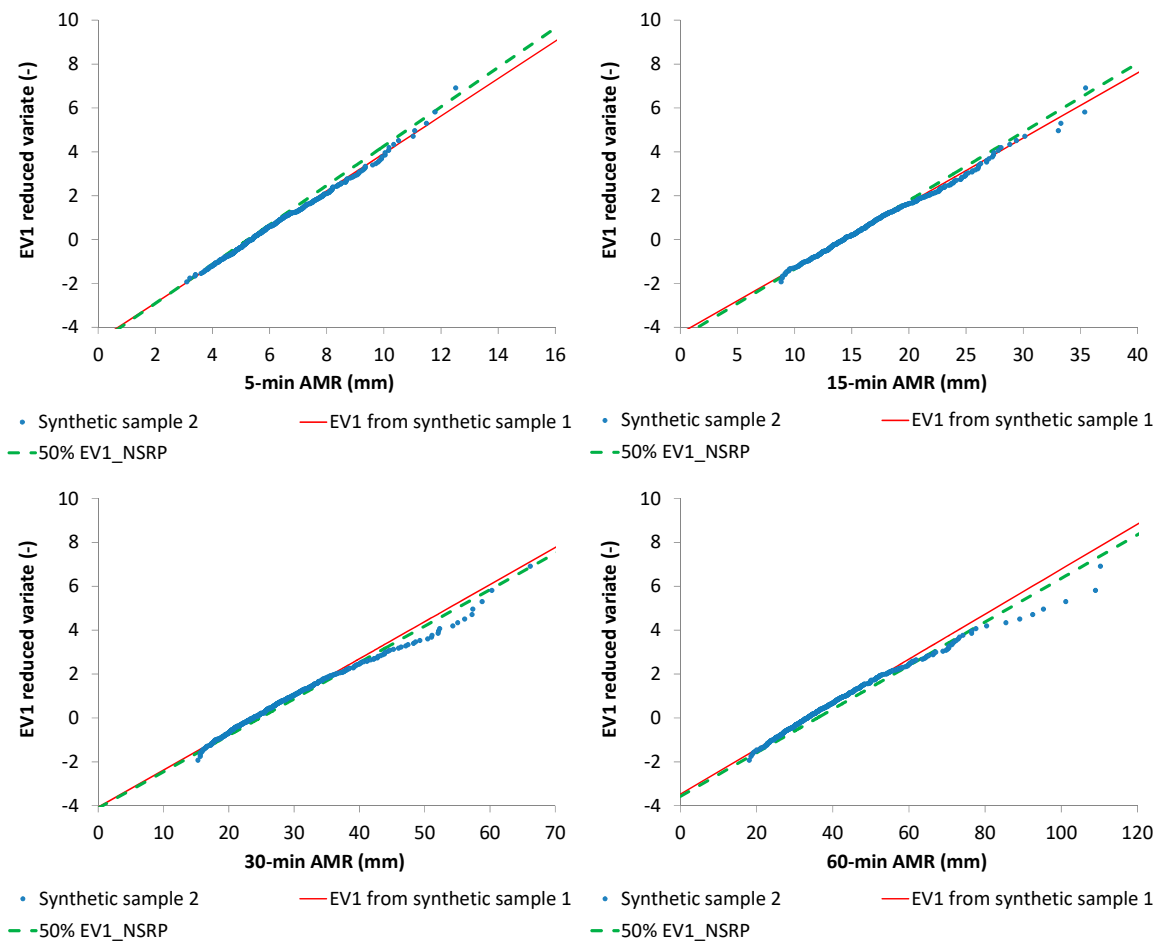
**Figure 7.** Synthetic sample 1: Cumulative frequency distributions for  $\alpha$  (top) and  $\varepsilon$  (bottom). The starting values are indicated with vertical dashed lines.



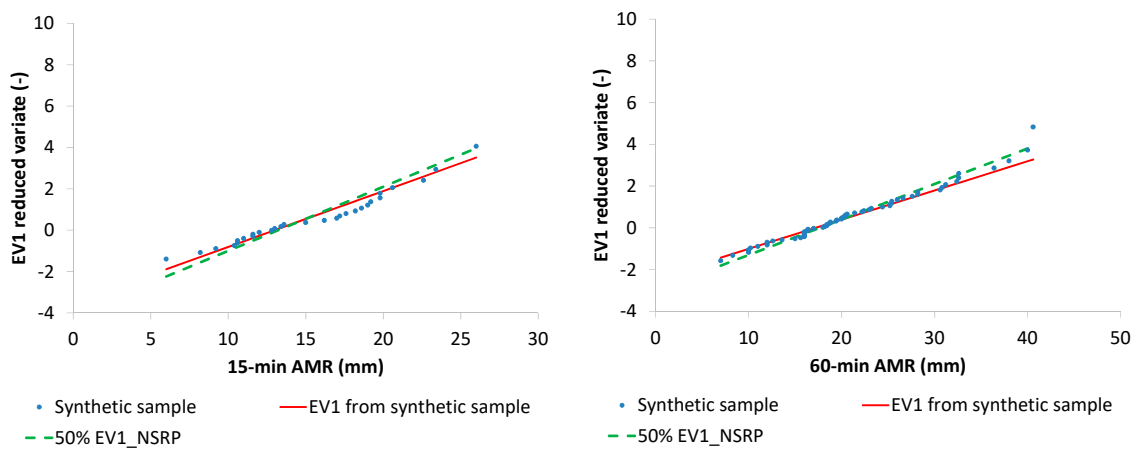
**Figure 8.** Synthetic sample 2: Cumulative frequency distributions for  $\alpha$  (top) and  $\epsilon$  (bottom). The starting values are indicated with vertical dashed lines.



**Figure 9.** Synthetic sample 1: EV1 probabilistic plots with AMR series, their EV1 theoretical curves and median EV1 curves from NSRP generation.



**Figure 10.** Synthetic sample 2: EV1 probabilistic plots with AMR series, their EV1 theoretical curves and median EV1 curves from NSRP generation.



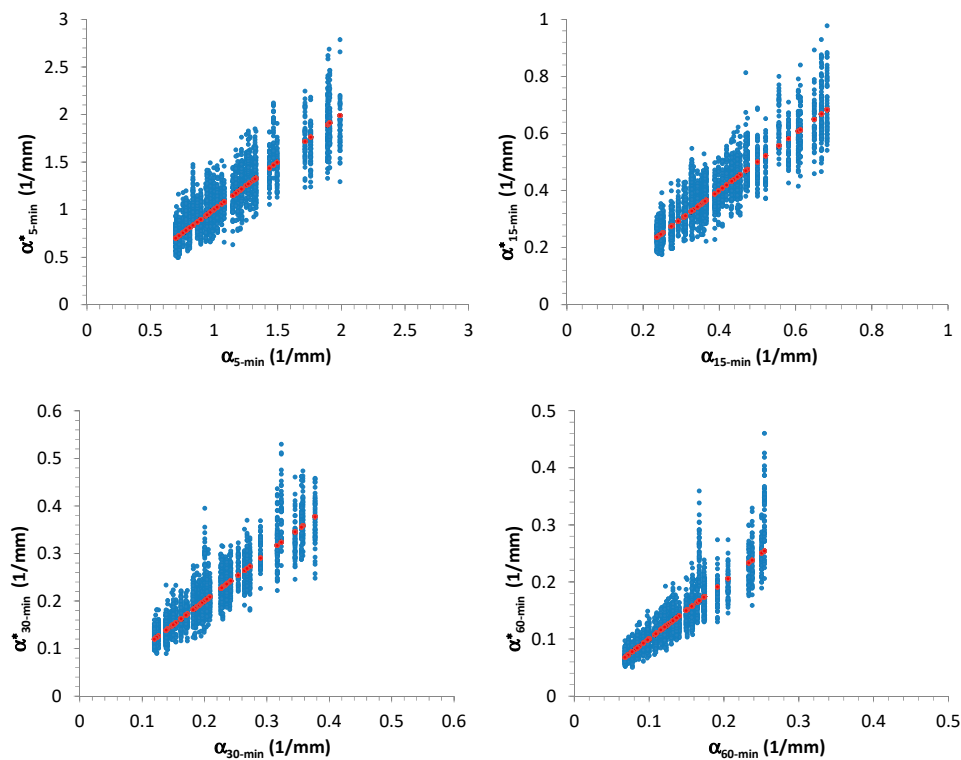
**Figure 11.** Cosenza rain gauge: EV1 probabilistic plots with AMR series, their EV1 theoretical curves and median EV1 curves from NSRP generation.

**Table 7.** Evaluation of EV1 distributions for synthetic samples and Cosenza data series.

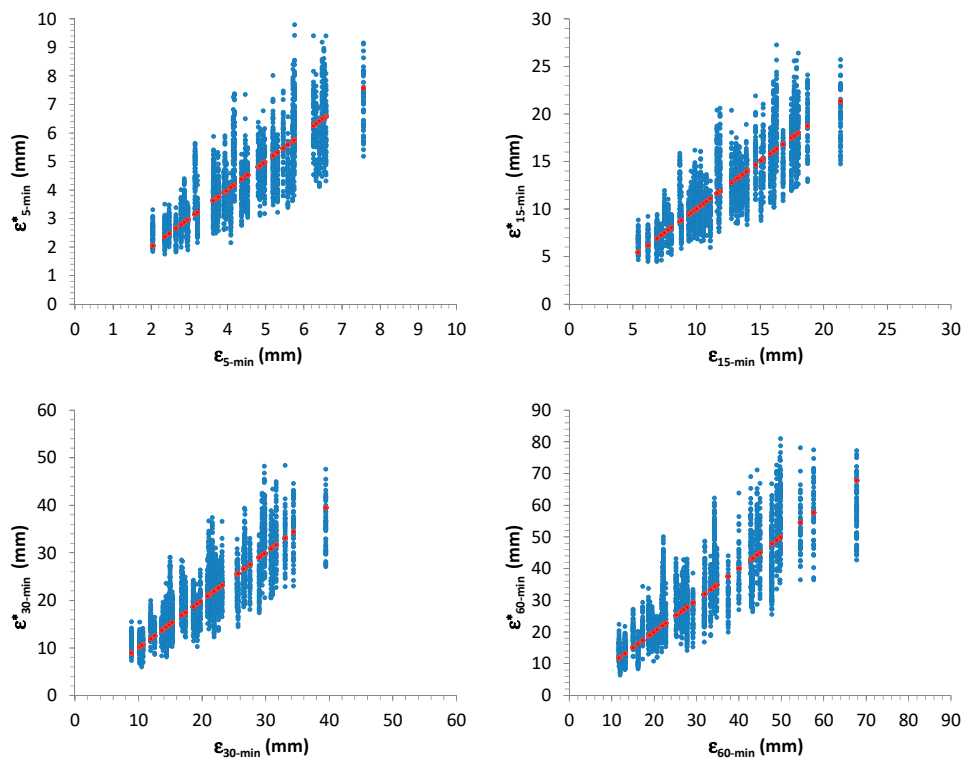
		50% EV1_NSRP				ML Parameter Estimation from Data			
	Performance Measures	5 min	15 min	0 min	60 min	5 min	15 min	30 min	60 min
Synthetic sample 1	MSE	0.017	0.042	0.037	0.019	0.010	0.024	0.018	0.018
	R <sup>2</sup>	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	NSE	0.99	0.97	0.98	0.99	0.99	0.99	0.99	0.99
Synthetic sample 2	MSE	0.041	0.023	0.036	0.069	0.006	0.008	0.033	0.018
	R <sup>2</sup>	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	NSE	0.98	0.99	0.98	0.96	0.99	0.99	0.98	0.99
Cosenza data series	MSE		0.143		0.041		0.090		0.049
	R <sup>2</sup>		0.96		0.98		0.96		0.98
	NSE		0.91		0.96		0.94		0.98

As another check, the authors repeated the validation for a further 50 synthetic samples, generated from 50 NSRP parameter sets which were randomly chosen from the ranges in Table 1.

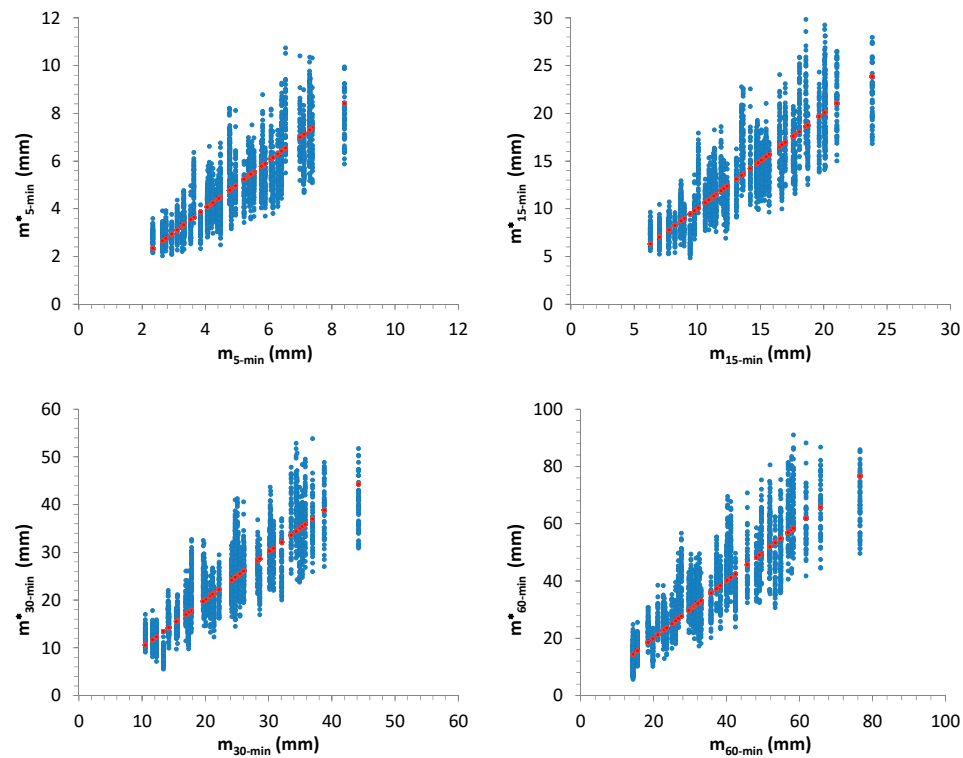
Figures 12–15 summarize the outcomes: starting values for  $\alpha$ ,  $\varepsilon$  and for sample mean,  $m$ , and standard deviation,  $s$ , related to 50 AMR series were compared with the associated distributions derived from the further 1000 generated samples for each considered one. In many cases, the starting values are very close to the median of the distributions.



**Figure 12.** Scatter plots showing the relationship among starting values for  $\alpha$  (horizontal axis) and their associated distribution (indicated with the superscript \* and reported on vertical axis). The red dots indicate the 1:1 line.

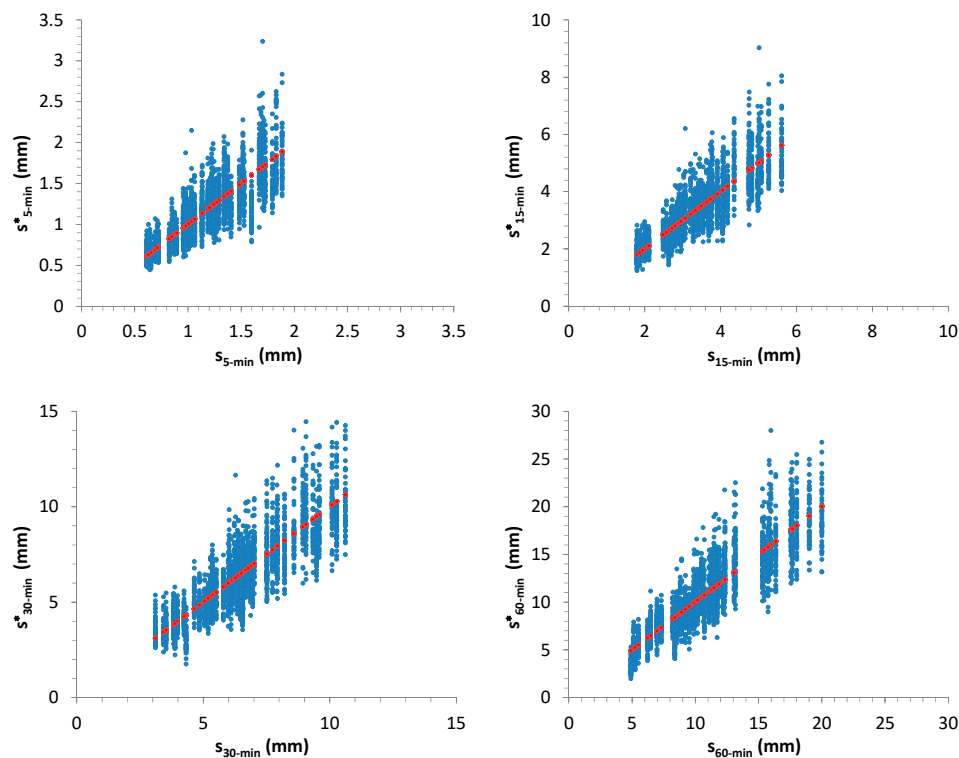


**Figure 13.** Scatter plots showing the relationship among starting values for  $\epsilon$  (horizontal axis) and their associated distribution (indicated with the superscript  $*$  and reported on vertical axis). The red dots indicate the 1:1 line.



**Figure 14.** Scatter plots showing the relationship among starting values for sample mean,  $m$ , related to 50 AMR series (horizontal axis) and their associated distribution (indicated with the superscript  $*$  and reported on vertical axis). The red dots indicate the 1:1 line.





**Figure 15.** Scatter plots showing the relationship among starting values for sample standard deviation,  $s$ , related to 50 AMR series (horizontal axis) and their associated distribution (indicated with the superscript \* and reported on vertical axis). The red dots indicate the 1:1 line.

#### 4. Discussion

The overall obtained results are highly promising, as they make possible the use of only AMR series (having very long sample sizes in many locations) for defining some features of the continuous process (for which sample sizes are shorter or absent for many sites). This is clearly an important aspect: unlike all the previous works, also mentioned in Section 1, the proposed methodology does not consider the usually limited continuous data series at high resolution for a SRG calibration, but uses the longer AMR series, and this allows for a good reconstruction of extreme events, without applying any change in model structure for a SRG (as in [9,30–32]) or introducing disaggregation techniques (as in [33–36]), censor values [38], or increasing in general the number of parameters.

In this work, the authors investigated the crucial role of sub-hourly AMR series, which can be considered as correlated with the statistical properties of intensity and duration for the rectangular pulses. Moreover, this research revalued the rectangular shape, as it was judged in many previous papers as “unsuitable for fine-scale data” [9] due to underestimation of extreme values at fine temporal scales.

It is clear that the proposed approach is highly flexible, and then it can be applied for other kinds of SRGs, in which: (i) the probability distributions are different from those adopted in Equations (1)–(5) and/or (ii) the pulses are hypothesized to have another shape, different from rectangular [9,30–32] and/or (iii) other probability distributions are adopted for AMR series.

Future interesting developments in this research can regard the study of possible importance for other data series (as examples, daily and multi-daily AMRs, monthly and annual rainfall data), of greater length than continuous data at fine scales, in order to better explain other significant features of continuous process, such as those represented by  $\lambda$ ,  $\beta_W$ , and  $\theta$  in a NRSP model, which were considered here as irrelevant for sub-hourly AMR series.

In fact, as already highlighted in Section 3.1, it is expected that inter-arrivals between two consecutive storms, and the associated number of bursts with their temporal occurrences, could mainly

influence aggregated rainfall heights at coarser resolutions (e.g., multi-daily, monthly, and yearly), while short-duration heavy rainfall events should not depend on these features. Moreover, seasonality is another important aspect to take into account, and, in particular, the research should focus on which NSRP parameters need of assumptions of a cyclostationary nature. This work remarked that seasonality is not essential for intensity and duration of the rectangular pulses, but it could be necessary for other parameters, mainly for a good modeling of dry/wet periods, not investigated here.

## 5. Conclusions

This described work constitutes the first step of an interesting research, aimed to investigate the possibility of a good calibration for models such as the NSRP, suitable for the generation of continuous high-resolution rainfall series, by using only data from annual maximum rainfall (AMR) series at fine scales, which are usually longer than continuous sub-hourly series, or they are the unique available data set for many locations.

The obtained results highlighted the crucial role of high-resolution extreme value distributions as important signatures for defining the statistical properties of intensity and duration for rectangular pulses, without considering continuous high-resolution rainfall data.

Future works will refer to the investigation of other possible signatures, deriving from further data series of greater length and available for many locations, such as multi-hourly and daily/multi-daily AMRs, or monthly and annual aggregated rainfall heights, in order to realize a comprehensive framework for the calibration of SRGs, in which it would be possible to estimate other important parameters (such as  $\lambda$ ,  $\beta_W$ , and  $\theta$  in NSRP), also taking into account the seasonality, without the need to use continuous time series at finest scales, whose sample sizes are still short for many sites.

Moreover, in the context of climate change, the discussed results and the flexible structure of a NSRP model can allow for investigation of potential trends into the continuous process at fine scales by analyzing/checking possible non-stationary behaviors in AMRs and in other longer series.

Obviously, improvements in the reconstruction of continuous rainfall series can also benefit the modeling of all those quantities which are connected with a pluviometric input, at hydrological space–time scales, such as streamflow time series in small basins and occurrences of shallow landslides.

**Author Contributions:** Conceptualization, D.L.D.L. and L.G.; data curation, D.L.D.L. and L.G.; formal analysis, D.L.D.L.; investigation, D.L.D.L.; methodology, D.L.D.L.; software, D.L.D.L.; supervision, D.L.D.L.; validation, D.L.D.L.; visualization, D.L.D.L. and L.G.; writing—original draft, D.L.D.L. and L.G.; writing—review and editing, D.L.D.L.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Onof, C.; Chandler, R.E.; Kakou, A.; Northrop, P.; Wheeler, H.S.; Isham, V. Rainfall modelling using Poisson-cluster processes: A review of developments. *Stoch. Environ. Res. Risk Assess.* **2000**, *14*, 384–411. [[CrossRef](#)]
2. Wheeler, H.S.; Chandler, R.E.; Onof, C.J.; Isham, V.S.; Bellone, E.; Yang, C.; Lekkas, D.; Lourmas, G.; Segond, M.L. Spatial-temporal rainfall modelling for flood risk estimation. *Stoch. Environ. Res. Risk Assess.* **2005**, *19*, 403–416. [[CrossRef](#)]
3. Rodriguez-Iturbe, I.; Cox, D.R.; Isham, V. Some Models for Rainfall Based on Stochastic Point Processes. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **1987**, *410*, 269–288. [[CrossRef](#)]
4. Rodriguez-Iturbe, I.; Cox, D.R.; Isham, V. A Point Process Model for Rainfall: Further Developments. *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **1998**, *417*, 283–298. [[CrossRef](#)]
5. Onof, C.; Wheeler, H.S. Improvements to the modelling of British rainfall using a modified Random Parameter Bartlett-Lewis Rectangular Pulse Model. *J. Hydrol.* **1994**, *157*, 177–195. [[CrossRef](#)]
6. Cowpertwait, P.S.P.; O’Connell, P.E.; Metcalfe, A.V.; Mawdsley, J.A. Stochastic point process modelling of rainfall. I. Single-site fitting and validation. *J. Hydrol.* **1996**, *175*, 17–46. [[CrossRef](#)]

7. Verhoest, N.; Troch, P.A.; De Troch, F.P. On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point. *J. Hydrol.* **1997**, *202*, 108–120. [[CrossRef](#)]
8. Cameron, D.; Beven, K.; Tawn, J. An evaluation of three stochastic rainfall models. *J. Hydrol.* **2000**, *228*, 130–149. [[CrossRef](#)]
9. Kaczmarek, J.; Isham, V.; Onof, C. Point process models for fine-resolution rainfall. *Hydrol. Sci. J.* **2014**, *59*, 1972–1991. [[CrossRef](#)]
10. Wheeler, H.S.; Isham, V.S.; Chandler, R.E.; Onof, C.J.; Stewart, E.J. *Improved Methods for National Spatial–Temporal Rainfall and Evaporation Modelling for BSM*; Department for Environment, Food and Rural Affairs (DEFRA), Flood management Division: London, UK, 2007.
11. Cowpertwait, P.S.P. Further developments of the Neyman–Scott clustered point process for modeling rainfall. *Water Resour. Res.* **1991**, *27*, 1431–1438. [[CrossRef](#)]
12. Entekhabi, D.; Rodriguez-Iturbe, I.; Eagleson, P.S. Probabilistic representation of the temporal rainfall process by a modified Neyman–Scott Rectangular Pulses Model: Parameter estimation and validation. *Water Resour. Res.* **1989**, *25*, 295–302. [[CrossRef](#)]
13. Onof, C.; Wheeler, H.S. Improved fitting of the Bartlett–Lewis Rectangular Pulse Model for hourly rainfall. *Hydrol. Sci. J.* **1994**, *39*, 663–680. [[CrossRef](#)]
14. Glasbey, C.A.; Cooper, G.; McGeachan, M.B. Disaggregation of daily rainfall by conditional simulation from a point-process model. *J. Hydrol.* **1995**, *165*, 1–9. [[CrossRef](#)]
15. Paschalis, A.; Molnar, P.; Fatichi, S.; Burlando, P. On temporal stochastic modeling of precipitation, nesting models across scales. *Adv. Water Resour.* **2014**, *63*, 152–166. [[CrossRef](#)]
16. Khaliq, M.N.; Cunnane, C. Modelling point rainfall occurrences with the modified Bartlett–Lewis rectangular pulses model. *J. Hydrol.* **1996**, *180*, 109–138. [[CrossRef](#)]
17. Smithers, J.C.; Pegram, G.G.S.; Schulze, R.E. Design rainfall estimation in South Africa using Bartlett–Lewis rectangular pulse rainfall models. *J. Hydrol.* **2002**, *258*, 83–99. [[CrossRef](#)]
18. Cowpertwait, P.; Isham, V.; Onof, C. Point process models of rainfall: Developments for fine-scale structure. *Proc. R. Soc. A–Math. Phys. Eng. Sci.* **2007**, *463*, 2569–2587. [[CrossRef](#)]
19. Gyasi-Agyei, Y.; Willgoose, G.R. A hybrid model for point rainfall modeling. *Water Resour. Res.* **1997**, *33*, 1699–1706. [[CrossRef](#)]
20. Gyasi-Agyei, Y. Identification of regional parameters of a stochastic model for rainfall disaggregation. *J. Hydrol.* **1999**, *223*, 148–163. [[CrossRef](#)]
21. Wasko, C.; Pui, A.; Sharma, A.; Mehrotra, R.; Jeremiah, E. Representing low-frequency variability in continuous rainfall simulations: A hierarchical random Bartlett Lewis continuous rainfall generation model. *Water Resour. Res.* **2015**, *51*, 9995–10007. [[CrossRef](#)]
22. Kossieris, P.; Efstratiadis, A.; Koutsoyiannis, D. Coupling the strengths of optimization and simulation for calibrating Poisson cluster models. In Proceedings of the Facets of Uncertainty: 5th EGU Leonardo Conference–Hydrofractals 2013–STAHY 2013, Kos Island, Greece, 17–19 October 2013.
23. Bo, Z.; Islam, S.; Eltahir, E.A.B. Aggregation–disaggregation properties of a stochastic rainfall model. *Water Resour. Res.* **1994**. [[CrossRef](#)]
24. Islam, S.; Entekhabi, D.; Bras, R.L.; Rodriguez-Iturbe, I. Parameter estimation and sensitivity analysis for the modified Bartlett–Lewis rectangular pulses model of rainfall. *J. Geophys. Res. Atmos.* **1990**, *95*, 2093–2100. [[CrossRef](#)]
25. Kim, D.; Cho, H.; Onof, C.; Choi, M. Let-It-Rain: A web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling. *Stoch. Environ. Res. Risk Assess.* **2016**. [[CrossRef](#)]
26. Kim, D.; Olivera, F.; Cho, H.; Socolofsky, S.A. Regionalization of the modified Bartlett Lewis rectangular pulse stochastic rainfall model. *Terr. Atmos. Ocean. Sci.* **2013**, *24*, 421–436. [[CrossRef](#)]
27. Velghe, T.; Troch, P.A.; De Troch, F.P.; Van de Velde, J. Evaluation of cluster-based rectangular pulses point process models for rainfall. *Water Resour. Res.* **1994**, *30*, 2847–2857. [[CrossRef](#)]
28. Kim, D.; Kwon, H.H.; Lee, S.O.; Kim, S. Regionalization of the Modified Bartlett–Lewis rectangular pulse stochastic rainfall model across the Korean Peninsula. *J. HydroEnviron. Res.* **2014**, 1–15. [[CrossRef](#)]
29. Verhoest, N.; Vandenbergh, S.; Cabus, P.; Onof, C.; MecaFigueras, T.; Jameleddine, S. Are stochastic point rainfall models able to preserve extreme flood statistics? *Hydrol. Process.* **2010**, *24*, 3439–3445. [[CrossRef](#)]

30. Cowpertwait, P. A generalized point process model for rainfall. *P. Roy. Soc. A-Math. Phys.* **1994**, *447*, 23–37. [[CrossRef](#)]
31. Cameron, D.; Beven, K.; Tawn, J. Modelling extreme rainfalls using a modified random pulse Bartlett–Lewis stochastic rainfall model (with uncertainty). *Adv. Water Resour.* **2000**, *24*, 203–211. [[CrossRef](#)]
32. Evin, G.; Favre, A. A new rainfall model based on the Neyman–Scott process using cubic copulas. *Water Resour. Res.* **2008**, *44*, W03433. [[CrossRef](#)]
33. Koutsoyiannis, D.; Onof, C. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *J. Hydrol.* **2001**, *246*, 109–122. [[CrossRef](#)]
34. Onof, C.; Townend, J.; Kee, R. Comparison of two hourly to 5-min rainfall disaggregators. *Atmos. Res.* **2005**, *77*, 176–187. [[CrossRef](#)]
35. Onof, C.; Arnbjerg-Nielsen, K. Quantification of anticipated future changes in high resolution design rainfall for urban areas. *Atmos. Res.* **2009**, *92*, 350–363. [[CrossRef](#)]
36. Kossieris, P.; Makropoulos, C.; Onof, C.; Koutsoyiannis, D. A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett–Lewis based model with adjusting procedures. *J. Hydrol.* **2018**, *556*, 980–992. [[CrossRef](#)]
37. Kim, D.; Olivera, F.; Cho, H. Effect of the inter-annual variability of rainfall statistics on stochastically generated rainfall time series: Part 1. Impact on peak and extreme rainfall values. *Stoch. Env. Res. Risk A.* **2013**, *27*, 1601–1610. [[CrossRef](#)]
38. Cross, D.; Onof, C.; Winter, H.; Bernardara, P. Censored rainfall modelling for estimation of fine-scale extremes. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 727–756. [[CrossRef](#)]
39. Rodriguez-Iturbe, I.; Febres De Power, B.; Valdes, J.B. Rectangular pulses point process models for rainfall: Analysis of empirical data. *J. Geophys. Res.* **1987**, *92*, 9645–9656. [[CrossRef](#)]
40. Cowpertwait, P.S.P. A Poisson-cluster model of rainfall: High-order moments and extreme values. *Proc. R. Soc. Lond.* **1998**, *454*, 885–898. [[CrossRef](#)]
41. Fisher, R.A.; Tippett, L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceed. Camb. Philos. Soc.* **1928**, *24*, 180–290. [[CrossRef](#)]
42. Calenda, G.; Napolitano, F. Parameter estimation of Neyman–Scott processes for temporal point rainfall simulation. *J. Hydrol.* **1999**, *225*, 45–66. [[CrossRef](#)]
43. Kottegoda, N.T.; Rosso, R. *Applied Statistics for Civil and Environmental Engineers*; Wiley-Blackwell: Hoboken, NJ, USA, 2008; p. 736. ISBN 978-1-405-17917-1.
44. De Luca, D.L.; Galasso, L. Stationary and Non-Stationary Frameworks for Extreme Rainfall Time Series in Southern Italy. *Water* **2018**, *10*, 1477. [[CrossRef](#)]
45. Marco, J.B.; Harboe, R.; Salas, J.D. *Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization*; Springer Publisher: Berlin, Germany, 1993; ISBN 978-94-010-4743-2.
46. Bennett, N.; Croke, B.; Guariso, G.; Guillaume, J.; Hamilton, S.; Jakeman, A.J.; Marsili-Libelli, S.; Newham, L.; Norton, J.; Perrin, C.; et al. Characterising Performance of Environmental Models. *Environ. Model. Softw.* **2013**, *40*, 1–20. [[CrossRef](#)]
47. Ritter, A.; Muñoz-Carpena, R. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *J. Hydrol.* **2013**, *480*, 33–45. [[CrossRef](#)]

