Extreme rainstorms: Comparing regional envelope curves to stochastically generated events

A. Viglione,[1] A. Castellarin,[2] M. Rogger,[1] R. Merz,[3] and G. Blöschl[1]

Received 3 February 2011; revised 4 November 2011; accepted 21 November 2011; published 12 January 2012.

[1] The depth-duration envelope curves (DDECs) are regional upper bounds on observed rainfall maxima for several durations. Recently, a probabilistic interpretation has been proposed in the literature in order to associate a recurrence interval T to the DDECs and, consequently, to retrieve point rainfall quantiles for ungauged sites. Alternatively, extreme rainfall quantiles can be retrieved from long synthetic rainfall series obtained with stochastic rainfall generators calibrated to local time series of rainfall events. While DDECs are sensitive to outliers and data errors, the stochastic rainfall generator performance is affected by the limited record lengths used for calibration. The objective of this study is to assess the reliability of the two alternative methods by verifying if they give consistent results for a wide study region in Austria. Relative to previous studies, we propose some generalizations of the DDEC procedure in order to better represent the Austrian data. The comparison of rainfall quantiles estimated with the two methods for large T shows an excellent agreement for intermediate durations (from 1 to 6 h), while the agreement worsen for very short (15 min) and long (24 h) durations. The results are scrupulously analyzed and discussed in light of the exceptionality of rainfall events that set the regional envelopes and the characteristics of the stochastic generator used. Our study points out that the combined use of these regional and local methods can be very useful for estimating reliable point rainfall quantiles associated with large T within regions where many rain gauges are available, but with limited record lengths.

Citation: Viglione, A., A. Castellarin, M. Rogger, R. Merz, and G. Blöschl (2012), Extreme rainstorms: Comparing regional envelope curves to stochastically generated events, *Water Resour. Res.*, *48*, W01509, doi:10.1029/2011WR010515.

1. Introduction

[2] The main focus of our study is on the accurate prediction of rainfall quantiles, that is rainfall depths associated with a given duration and recurrence interval. Nevertheless, the study should be viewed in a broader perspective and placed in a more general context. Its key motivation is in fact improving the accuracy of the design flood estimates. Producing a reliable estimate of the design flood, herein defined as the discharge associated with a given probability of occurrence (or recurrence interval T), is a critical yet recurring task in many problems related to flood hazard and flood risk assessment and mitigation. In addition, hydrologists and practitioners are frequently required to predict the design flood estimates in ungauged river basins or for T values that are large, or very large, compared to the length of the available observed record [*Stedinger et al.*, 1993; *Pilgrim and Cordery*, 1993; *Klemeš*, 1993]. A viable strategy in these circumstances is the application of an indirect approach to design flood estimation by using conceptual rainfall-runoff models [e.g., design-storm method, *Pilgrim and Cordery*, 1993, p. 9.13] or by deriving the flood frequency curve from the rainfall frequency curve through simplified methods [e.g., *Eagleson*, 1972; *Fiorentino and Iacobellis*, 2001; *Sivapalan et al.*, 2005; *Gaume*, 2006; *Merz and Blöschl*, 2008; *Viglione and Blöschl*, 2009; *Viglione et al.*, 2009]. This approach shifts the original problem to the selection and parameterization of a suitable rainfall-runoff model and to the representation of the frequency regime of rainfall extremes. The former is a topical issue in hydrology, and the scientific community proposes a number of approaches to rainfall-runoff model calibration in ungauged basins [see, e.g., *Wagener and Wheater*, 2006; *Bárdossy*, 2007; *Parajka et al.*, 2007; *Hundecha et al.*, 2008; *Zhang et al.*, 2008; *Castiglioni et al.*, 2010]. Our study focuses on the latter issue and, in particular, considers the problem of estimating design rainstorms, defined herein as rainfall-depth quantiles associated with given storm durations, for large and very large recurrence intervals.

[3] We consider two tools for retrieving rainfall quantiles for large recurrence intervals that are profoundly different in nature, namely regional depth-duration envelope curves (DDECs) and locally calibrated stochastic rainfall generators.

[4] DDECs were recently proposed in the literature and reconsider the concepts of regionalization of rainstorms and definition of a statistical upper bound on the observed rainfall extremes [*Castellarin et al.*, 2009]. The curves

[1]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria.
[2]DISTART, School of Civil Engineering, University of Bologna, Bologna, Italy.
[3]Helmholtz Centre for Environmental Research, Halle, Germany.

0043-1397/12/2011WR010515

adapt the idea of probabilistic envelope curves of flood flows [see, e.g., *Castellarin et al.*, 2005; *Castellarin*, 2007] to extreme rainfall depths and are defined as the regional upper bound on all observed maximum rainfall depths for a given time scale (conventionally referred to as duration or aggregation level, here identified by τ). The probabilistic interpretation of the curves enables one to associate an estimate of the nonexceedance probability (i.e., recurrence interval) to the curves themselves. The return period is associated at a regional basis on the basis of the equivalent number of independent observations of the regional sample, similarly to the plotting positions used for local samples. This feature characterizes the curves as easy-to-use graphical tools to identify a plausible value of the extreme point rainfall depth at any (gauged and ungauged) location of the study area and to attach to this rainfall depth an estimate of the recurrence interval.

[5] *Castellarin et al.* [2009], proposed and tested DDECs for an Italian case study. An assessment of the general validity of the curves for other geographical and climatic contexts is definitely of interest for further endorsing their construction and practical utilization. Also, given the large or very large recurrence intervals associated with regional DDECs [see *Castellarin et al.*, 2009], testing the accuracy of rainfall quantiles retrieved from the curves is not an easy task, which is further complicated by the difficulties associated with the identification of reliable reference terms against which to perform the test.

[6] For many hydrological applications (e.g., estimation of long-term water yield, water resources planning, estimation of design floods) models for the synthetic simulation of rainfall time series are required, which incorporate the variability of rainfall at different time scales: within storm, between storm, seasonal, interannual, interdecadal, etc. It is important for these models to correctly reproduce the statistical characteristics of individual storms, of aggregated volumes (e.g., monthly rainfall depths), and of extreme rainfall as well, so that they can be applied for flood estimation purposes. To achieve a good fit of all these characteristics at the same time is not straightforward and depends on how the model is calibrated, i.e., which objective is of major interest.

[7] In this paper we use a slightly modified version of the stochastic rainfall model presented by *Sivapalan et al.* [2005]. Essentially, the model generates discrete independent rainfall events whose arrival times, durations, average rainfall intensity, and the within-storm intensity patterns are all random, governed by specified distributions whose parameters are assumed to be seasonally dependent. The calibration procedure is as follows: (1) the rainfall data are analyzed and the storm events are separated; (2) characteristics of the events (e.g., duration, average intensity, mass curve) are collected and their statistics (i.e., mean, variance, and their seasonal variability) are estimated; and (3) the parameters of the stochastic rainfall model are fitted to these statistics. The model is calibrated to the events and therefore can be fitted also to quite short time series (i.e., few years), provided that the number of events is still reasonably high. Specially in these cases, when short time series are used, it is important to test the reliability of simulated extreme rainfall events. One way to assess if simulations and observations are consistent is to compare the statistics of the simulated extreme rainfall events to regional information on extreme precipitation, which in this work is represented by the regional DDECs.

[8] Our analysis focuses on the comparison of rainfall quantiles that can be retrieved from regional DDECs and long stochastically generated rainfall series for the Tyrolean region, in Austria. The study mainly aims at addressing the following science questions: (1) how general and realistic is the formulation of depth-duration envelope curves? (2) How credible are extreme rainfall events extracted from synthetic long rainfall time series? Since the two methods are profoundly different in nature and are based on completely independent characteristics extracted from the data, their comparison represent a mutual validation of both methods against rainfall extremes. The paper is structured as follows: in section 2 the depth-duration envelope curves and the structure of the stochastic rainfall model are presented; in section 3 the two methods are applied using the available data in Tyrol; and in section 4 the results are compared.

2. Methods

2.1. Depth-Duration Envelope Curves

[9] Depth-duration envelope curves [see *Castellarin et al.*, 2009] are graphical representations of the maximum observed point rainfall depth (record rainfall depth) for a given duration over a region. *Castellarin et al.* [2009] propose a probabilistic interpretation of the curves, which enables one to estimate the exceedance probability of the curves for design purposes.

[10] DDECs represent an adaptation of probabilistic regional envelope curves of flood flows [see *Castellarin et al.*, 2005; *Castellarin*, 2007]. The probabilistic interpretation of the curves relies on the main assumption that the spatial variability of the frequency regime of rainfall annual maxima for a given duration τ can be accurately described by the variability of mean annual precipitation (MAP), used as a surrogate of location. The validity of this assumption has been shown for various parts of the world [e.g., *Ferreri and Ferro*, 1990; *Schaefer*, 1990; *Alila*, 1999; *Brath et al.*, 2003; *Di Baldassarre et al.*, 2006].

[11] In particular, *Castellarin et al.* [2009] show that if (1) the relationship between L-coefficients of variation and skewness (i.e., L-CV and L-CS, for a definition of L-moments see, e.g., *Hosking* [1990]) of rainfall extremes and MAP can be formalized through an exponentially decaying curve, and (2) a nondecreasing relationship holds between the mean annual maximum rainfall depth m_τ (for duration τ) and MAP [see, e.g., *Herschfield*, 1961; *Bell*, 1969; *Ferreri and Ferro*, 1990; *Alila*, 1999; *Di Baldassarre et al.*, 2006], an approximate regional representation of the dependence on MAP of the ratio between the T-year rainfall quantile $h_{\tau,T}$ (for duration τ) and MAP can be expressed as

$$\ln \frac{h_{\tau,T}}{\mathrm{MAP}} = A_{\tau,T} + B_\tau \ln(\mathrm{MAP}), \quad (1)$$

where $A_{\tau,T}$ and B_τ are regional coefficients.

[12] Hence, under the same hypotheses, *Castellarin et al.* [2009] propose to adopt equation (1) to bound all record rainfall depths observed over the study region for a given

duration τ, named $h_{\tau,\mathrm{MAX}}$. The authors recommended estimating the slope B_τ by regressing the standardized rainfall maxima against the local value of MAP, and the intercept $A_{\tau,T}$ by enveloping all rainfall maxima observed in the study region through the following equation:

$$A_{\tau,\mathrm{MAX}} = \max_{j=1,2,\ldots,M}\left\{\ln\frac{h_{\tau,\mathrm{MAX},j}}{\mathrm{MAP}_j} - B_\tau \cdot \ln(\mathrm{MAP}_j)\right\}, \quad (2)$$

where $h_{\tau,\mathrm{MAX},j}$ denotes the maximum rainfall depth observed for duration τ at site $j = 1, 2, \ldots, M$ and M is the number of sites in the region, while MAP_j is the local value of the mean annual precipitation.

[13] The intercept $A_{\tau,\mathrm{MAX}}$ estimated as in equation (2) is associated with an unknown recurrence interval T, which, in turn, is the recurrence interval associated with the envelope curve. *Castellarin et al.* [2009] propose to estimate T by using the concept of regional information content, or, analogously, equivalent number of independent annual maxima, n_{eff}. n_{eff} is generally smaller than the overall station years of data due to cross correlation among annual sequences [see, e.g., *Matalas and Langbein*, 1962; *Stedinger*, 1983; *Castellarin et al.*, 2005].

[14] *Castellarin* [2007] developed an empirical estimator of n_{eff} for a real-world data set including M individual annual maximum series (AMS) that globally span n years, being characterized by missing data, different installation years for different gauges, etc. The estimator is applied as follows. First, one identifies the number of years n_1 for which the original data set includes only one observation of the annual maximum rainfall depth, that is $M-1$ observations are missing (i.e., $M-1$ gauges are not operational or not installed yet). These n_1 observations are effective (independent) by definition. Second, the data set containing the $n - n_1$ remaining years is subdivided into $N_s \leq (n - n_1)$ subsets; each one of them (say subset s) is selected in such a way that all its $L_s \leq M$ sequences are concurrent and of equal length l_s. Using this splitting criterion, the effective number of observations n_{eff} can be estimated as the summation of the effective sample years of data estimated for all N_s subsets,

$$\hat{n}_{\mathrm{eff}} = n_1 + \sum_{s=1}^{N_s}\hat{n}_{\mathrm{eff},s} = n_1 + \sum_{s=1}^{N_s}\frac{L_s \cdot l_s}{1 + [\overline{\rho}^{\beta}]_{L_s}\cdot(L_s - 1)},$$
$$\text{with } \beta = 1.4\frac{(L_s \cdot l_s)^{0.176}}{[(1-\overline{\rho})^{0.376}]_{L_s}}, \quad (3)$$

where overbars indicate average values of the corresponding functions of the correlation coefficient (i.e., $[\overline{\rho}^{\beta}]_{L_s}$ is the average of the $L_s(L_s-1)/2$ values of $\rho_{k,j}^{\beta}$, where $\rho_{k,j} = \rho_{j,k}$ is the correlation coefficient between annual maximum floods at sites k and j, with $1 \leq k < j \leq L_s$ [see also *Castellarin*, 2007; *Castellarin et al.*, 2009]). The empirical expression used for the exponent β in (3) results from the Monte Carlo study of the variance of the regional record for a correlated region conducted by *Castellarin et al.* [2005].

[15] The application of equation (3) requires the selection of a suitable cross-correlation model for representing intersite correlation. *Castellarin* [2007] showed that the selection of the cross-correlation model has limited impact on (3) and suggested to use the model introduced by *Tasker and Stedinger* [1989] to approximate the true annual peak cross-correlation function $\rho_{i,j}$ as a function of the distance $d_{i,j}$ among sites i and j,

$$\rho_{i,j} = \exp\frac{-\lambda_1 \cdot d_{i,j}}{1 + \lambda_2 \cdot d_{i,j}}, \quad (4)$$

where $\lambda_1 > 0$ and $\lambda_2 \geq 0$ are regional parameters that may be estimated by ordinary or weighted least squares procedures.

[16] Once n_{eff} has been estimated, a suitable plotting position needs to be selected for evaluating the recurrence interval associated with the envelope curve. The return period obtained from the general plotting position reads [*Cunnane*, 1978]

$$T = \frac{\hat{n}_{\mathrm{eff}} + 1 - 2\eta}{1 - \eta}, \quad (5)$$

where η is the plotting position parameter and $\hat{n}_{\mathrm{eff}}$ is the empirical estimate of n_{eff} given in (3). Among several possible options for selecting η, a quantile unbiased-plotting position should be used in this context. A traditional choice is the Hazen plotting position ($\eta = 0.5$), which *Castellarin* [2007] showed to be particularly suitable for use when the annual maxima follow a generalized extreme value (GEV) distribution [*Jenkinson*, 1955]. The GEV distribution has been shown to satisfactorily reproduce the sample frequency distribution of hydrological extremes around the world [see, e.g., *Stedinger et al.*, 1993; *Vogel and Wilson*, 1996; *Robson and Reed*, 1999; *Castellarin et al.*, 2001; *Di Baldassarre et al.*, 2006].

[17] *Castellarin et al.* [2009] applied and assessed the concept of DDECs for a wide geographical region located in northern-central Italy. The accuracy of estimated recurrence intervals was assessed through a comprehensive cross-validation procedure that compared the rainfall quantiles (rainfall depths associated with a given duration and recurrence interval) retrieved from DDECs with the corresponding quantiles computed through a regional depth-duration frequency equation proposed by the scientific literature for the region of interest. In general, the comparison showed a good agreement between the rainfall quantiles estimated through the two different methodologies, supporting the general plausibility of the probabilistic interpretation of DDECs [*Castellarin et al.*, 2009]. Nevertheless, the results of the analysis (1) were still preliminary, (2) referred to a limited geographical and climatic context, and (3) pointed out that the log linear relationship (equation (1)) may not provide an accurate representation of the envelope for long durations (i.e., 12–24 h for the considered data set) and low MAP values (i.e., $\approx$ 500–600 mm in the considered region). We address all three points here as we provide a further and independent validation of the general theoretical framework of DDECs, we focus on a different geographical and climatic context, and we further test the general hypothesis on the mathematical expression of the envelope.

2.2. Stochastic Rainfall Generator

[18] Different approaches to rainfall modeling have been proposed in literature and used in applied hydrology [see,

e.g., *Verhoest et al.*, 2010]. Among them, point stochastic models, which make simple assumptions with respect to the physical processes, enable the description of the hierarchical structure of rainfall with a minimal set of model parameters. Popular stochastic rainfall models are the pulse-based ones, such as the Neyman-Scott and the Bartlett-Lewis models [see, e.g., *Rodriguez-Iturbe et al.*, 1987; *Cowpertwait*, 1994; *Verhoest et al.*, 1997; *Onof et al.*, 2000; *Cameron et al.*, 2001; *Fowler et al.*, 2005; *Burton et al.*, 2010, among others], which characterize random storms as summations of individual pulses whose interarrival times, intensities, and durations are described statistically. Other stochastic rainfall models are based on temporal downscaling of rainfall intensities through "fractal" techniques [see, e.g., *Menabde et al.*, 1997; *Harris et al.*, 1998; *Menabde and Sivapalan*, 2000; *Sivapalan et al.*, 2005, who use bounded random cascades for the temporal downscaling]. The model considered here belongs to the latter category. We use a slightly modified version of the stochastic rainfall model presented by *Sivapalan et al.* [2005]. Essentially, the model consists of discrete rainfall events whose arrival times (or interstorm periods) t_b, durations t_r, average rainfall intensity i_m, and the within-storm intensity patterns are all random, governed by specified distributions (see Figure 1 for an example of the generated rainfall series $i(t)$). In addition, the parameters of these distributions are assumed to be seasonally dependent. For storm durations t_r we consider the three parameter weibull distribution

$$f_{T_r}(t_r) = \frac{\beta_r}{\gamma_r}\left(\frac{t_r-\xi_r}{\gamma_r}\right)^{\beta_r-1}\exp\left(-\frac{t_r-\xi_r}{\gamma_r}\right)^{\beta_r} \tag{6}$$

with parameters ξ_r (position), γ_r (scale), and β_r (shape). The scale parameter is linked to $\mu_r = \mathrm{E}[t_r]$, i.e., the mean storm duration, by the relationship

$$\gamma_r = (\mu_r - \xi_r)\left\{\Gamma\left(1+\frac{1}{\beta_r}\right)\right\}^{-1}, \tag{7}$$

while the shape parameter is linked to the coefficient of variation

$$\mathrm{CV}[t_r-\xi_r] = \sqrt{\frac{\Gamma(1+2/\beta_r)}{\left(\Gamma(1+1/\beta_r)\right)^2}-1}\,. \tag{8}$$

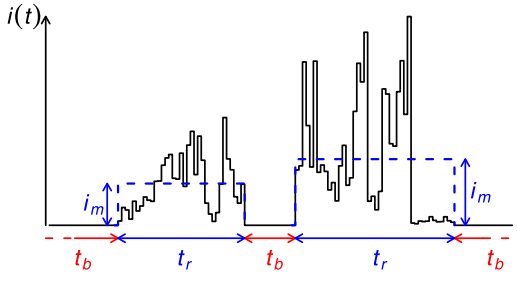

Figure 1. Example of part of a generated rainfall series $i(t)$. The interstorm periods t_b, storm durations t_r, and average rainfall intensities i_m are indicated.

The scale parameter is allowed to vary with time of year as a two-component Fourier series:

$$\mu_r = \xi_r + \delta_r + \alpha_{1r}\cos\left\{\frac{2\pi}{\omega}(\phi-\phi_{1r})\right\} + \alpha_{2r}\cos\left\{\frac{4\pi}{\omega}(\phi-\phi_{2r})\right\}, \tag{9}$$

where $\xi_r + \delta_r$ is the annual average storm duration, ϕ_{1r} and ϕ_{2r} determine the seasonal phase shift, α_{1r} and α_{2r} give the amplitude of the seasonal variation of t_r, ϕ describes the time of year, and ω is the total number of time units in a year (i.e., $\omega = 12$ months if ϕ, ϕ_{1r}, and ϕ_{2r} are in month units). The shape parameter β_r is assumed to be constant throughout the year. For interstorm periods t_b we use the same model type (equations (6), (7), and (9)) with parameters ξ_b, δ_b, α_{1b}, ϕ_{1b}, α_{2b}, ϕ_{2b}, and β_b. With respect to *Sivapalan et al.* [2005], we add here the parameters ξ_r and ξ_b (i.e., the minimum rainfall length and interstorm period, which do not need calibration since they are chosen a priori when analyzing the rainfall data series) and we consider two-component Fourier series because the one component used by *Sivapalan et al.* [2005] are not able to reproduce well the seasonal variations observed in the Austrian data.

[19] The distribution of the average rainfall intensity i_m within the storm depends on t_r, according to the gamma distribution

$$f_{I_m|T_r}(i_m|t_r) = \frac{\lambda}{\Gamma(\kappa)}(\lambda i_m)^{\kappa-1}\exp(-\lambda i_m)\,, \tag{10}$$

where parameters λ and κ are functions of t_r as

$$\kappa = \frac{t_r^{-b_2}}{a_2} \quad \text{and} \quad \lambda = \frac{t_r^{-b_1-b_2}}{a_1 a_2}\,, \tag{11}$$

that means

$$E[i_m|t_r] = a_1 t_r^{b_1} \quad \text{and} \quad \mathrm{CV}^2[i_m|t_r] = a_2 t_r^{b_2}\,. \tag{12}$$

The parameter a_1 is allowed to vary sinusoidally with time of year τ as follows:

$$a_1 = \delta_a + \alpha_{1a}\cos\left\{\frac{2\pi}{\omega}(\phi-\phi_{1a})\right\} + \alpha_{2a}\cos\left\{\frac{4\pi}{\omega}(\phi-\phi_{2a})\right\}, \tag{13}$$

where δ_a is the annual average value for a_1, ϕ_{1a} and ϕ_{2a} define the seasonal phase shift, α_{1a} and α_{2a} give the amplitude of the seasonal variation of a_1, ϕ describes the time of year, and ω is the total number of time units in a year.

[20] The mean storm intensity is disaggregated further to subhourly intensity patterns using bounded random cascades as by *Menabde and Sivapalan* [2000]. To reproduce the scaling properties of rainfall, we use the technique of breakdown coefficients (BDC), which was developed by *Novikov* [1966, 1969] for application to the theory of turbulence and was adopted for application to rainfall modeling by *Menabde et al.* [1997], *Harris et al.* [1998], *Menabde and Sivapalan* [2000], among others. The BDC method is applicable to both self-similar and self-affine fields and

allows the difference in their scaling behaviors to be easily visualized. The BDC is defined as

$$u(\tau_1, \tau_2) = \frac{R_{\tau_1}(t_1)}{R_{\tau_2}(t_2)} \quad \text{with} \quad \tau_1 < \tau_2 , \tag{14}$$

where $R_{\tau_1}(t_1)$ and $R_{\tau_2}(t_2)$ are the rainfall totals accumulated over the durations τ_1 and τ_2 and centered at t_1 and t_2, respectively. The interval τ_1 defined above is assumed to be completely included in the interval τ_2. The BDC is a random variable and ranges from a minimum of zero to a maximum of unity. For self-similar random fields the probability density function of $u(\tau_1, \tau_2)$ will depend only on the scale ratio τ_1/τ_2 and not on either of the two scales τ_1 and τ_2 separately. However, previous analyses of many rainfall time series suggest that there is, indeed, a scale dependence: for instance a decreasing variance with decreasing time scale [*Menabde et al.*, 1997; *Harris et al.*, 1998]. We model the BDC for different time scales with the symmetric beta distribution

$$f_U(u) = \frac{1}{B(a)} u^{a-1}(1-u)^{a-1} , \tag{15}$$

where $B(a)$ is the incomplete beta function with the sole parameter a changing with the time scale of observation τ. This dependence is parameterized by the following scaling law:

$$a(\tau) = a_0 \tau^{-H} . \tag{16}$$

This parameterization provides a natural way for the modeling of temporal disaggregation of individual storms by the bounded random cascade model. Note that the model simplifies to the normal self-similar random cascade model if the exponent H is taken to be zero, allowing a to remain invariant with scale.

2.3. Mutual Validation of the Methods

[21] Not enough data are available at one single station to validate the stochastic model described in section 2.2 against extreme events with high return periods. At the same time, also the DDEC method described in section 2.1 needs validation. Testing the accuracy of rainfall quantiles for large or very large recurrence intervals associated with regional DDECs needs the identification of reliable reference terms against which to perform the test. In this paper we perform a mutual validation of the two methods by comparing their results, i.e., are the estimated local and regional rainfall extremes consistent with each other? Since the methods are based on different pieces of information extracted from the data (record rainfall peaks and cross-correlation structure for DDECs and individual rainfall events characteristics for the stochastic rainfall generator), the consistency between the two methods will increase our confidence in the estimated rainfall quantiles, while we have to further investigate possible reasons of disagreement otherwise. We proceed as follows: (1) the DDEC technique is used to estimate $\hat{n}_{\text{eff}}$ and the return period of the envelope curves for rainfall depths associated to different durations (1/4, 1, 3, 6, 12, and 24 h) for the region Tyrol in Austria (see following section 3.1); (2) the stochastic model is calibrated to the available data for a number of rainfall stations in the same region (see section 3.2); (3) for each one of these stations, the stochastic model is used to produce very long time series of synthetic rainfall (e.g., 1,000,000 years); (4) for every synthetic time series, the quantiles of the rainfall depths for durations 1/4, 1, 3, 6, 12, and 24 h are extracted corresponding to the return periods estimated by the DDEC procedure; (5) from each of the synthetic time series, 1000 repetitions of $\hat{n}_{\text{eff}}$ years are extracted and, for the durations of interest, the maximum rainfall depths are collected in order to assess the influence of sampling variability on the envelope curves, under the hypotheses of the DDEC method; (vi) finally, for each duration (1/4, 1, 3, 6, 12, and 24 h) we compare the envelope curve obtained at point (1) with the quantiles obtained at point (4) and the 90% confidence bounds of the maximum rainfall depths obtained at point (5) (see section 4).

[22] If the two methods are consistent, the quantiles obtained by stochastic modeling should lay around the envelope curve and the 90% confidence bounds should contain this curve on average 90% of the times. If most of the quantiles lay below (above) the envelope curve, this indicates that the extremes produced locally with the stochastic model are low (high) respect to what the regional analysis suggests or, analogously, that the return period associated to the regional envelope curve is low (high) respect to what the local analysis suggests. In any case, if the 90% confidence bounds from the local stochastic models contain the regional envelope curve approximately 90% of the time, this indicates that the differences between the two methods may be explained by sampling variability and, therefore, they can be considered to be consistent.

3. Application

[23] The chosen study area is Tyrol, which is located in the western part of Austria within the Alpine region and has an extension of about 10,600 km^2. The climatic situation of Tyrol is strongly determined by the steep mountain ranges acting as meteorological divides and separating the region into more humid and drier areas. The region is mainly influenced by precipitation from western weather conditions and hence the northern rim of the Alps is the most humid and snow-prone area of Tyrol with a mean annual precipitation up to 2000 mm. This is also true for the western border of the study area. The inneralpine valleys on the other hand have a milder and drier climate with a mean annual precipitation of only 600 mm in the Upper Inn valley. Nevertheless convective events with high intensities can also occur in the drier valleys due to orographic effects.

[24] Figure 2 shows the location of the rain gauges with high resolution data (5, 10, and 15 min) which were included in the analysis. Precipitation data of 73 rain gauges with a record length of at least one year up to more than 20 years were available (see Figure 3 and Table 1).

3.1. Depth-Duration Envelope Curves In Tyrol

[25] As mentioned in section 2.1, several assumptions on m_τ, L-CV and L-CS allowed *Castellarin et al.* [2009] to postulate that the relationship between MAP and the DDEC is approximately linear in the log-log space in north-central

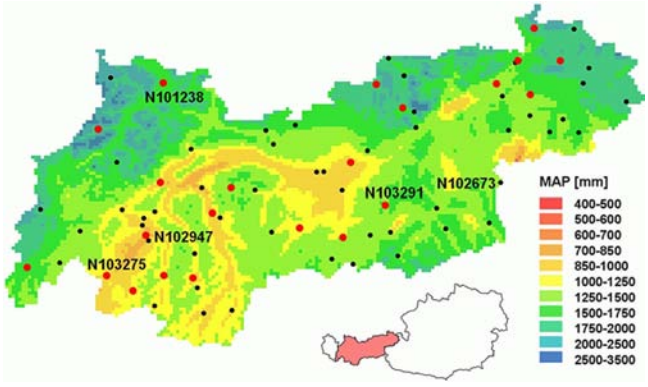

Figure 2. The 73 rain gauges in Tyrol (Austria) used in the study (black points). The 22 stations used for the comparison are highlighted in red. The color scale shows the mean annual precipitation in Tyrol. The station numbers indicate those stations mentioned in the paper: Reutte (N101238), Durlaßboden (N102673), Prutz (N102947), Pfunds (N103275), and Wattener Lizum (N103291).

Italy (see equation (1) here and Figure 3 by *Castellarin et al.* [2009]). Empirical evidence do not support for Tyrol the classical results from *Schaefer* [1990], e.g., decreasing L-CV and L-CS of the AMS with increasing MAP. As shown in Figure 4, no significant trends of L-CV and L-CS can be identified in relation to MAP for any duration τ. We would have expected a difference between the growth curves of rainfall extremes in northern and southern Tyrol because of shading effects (i.e., more and more extreme storms in the north, where MAP is higher, see Figure 2). Apparently, instead, these spatial differences are captured by the mean annual maximum rainfall depth m_τ. In that case, if we assume that the scatter in Figure 4 is just due to sample variability, the region can be assumed to be homogeneous (in the sense of *Hosking and Wallis* [1997]) for any duration τ, i.e., the frequency distributions of the annual maximum rainfall depths at different sites are the same except for the scale factor m_τ. To check this assumption, the heterogeneity measure of *Hosking and Wallis* [1993] has been calculated for the different considered durations, as shown in Table 2.

[26] For long durations, i.e., $\tau = 3$ to 24 h, the assumption of homogeneity can be accepted, i.e., the values of the heterogeneity measure are below 2, above which *Hosking and Wallis* [1993] suggest to consider the region as heterogeneous. For short durations, i.e., $\tau = 1/4$ and 1 h, the measure is above 2. Specially for $\tau = 1/4$ h, the spread of the at-site L-CVs is much higher than what would be expected for a homogeneous region. In any case, as suggested by Figure 4, the L-moments ratios seem to be independent from MAP. Therefore, differently from *Castellarin et al.* [2009], we can assume that L-CV and L-CS do not vary spatially, at least for durations $\tau > 1$ h. This means that the ratio $h_{\tau,T}/m_\tau$, hereafter called $k_{\tau,T}$, vary with duration τ and return period T but not in space (being MAP is a surrogate for space).

[27] We use the mean annual precipitation MAP to explain the spatial variability of m_τ. We are aware of the limitations of adopting this scaling law for representing m_τ, which could be avoided, for example, by adopting the empirical values of m_τ to standardize observed rainfall maxima and constraining the analysis to the regional rainfall growth curve $k_{\tau,T}$ only. Nevertheless, if we decide for this latter choice, the regional values of $h_{\tau,T}$ should be obtained, for example, by multiplying $k_{\tau,T}$ to spatial interpolations of the mean extreme precipitations m_τ, which are not as robust as the spatial interpolation of MAP. We assume that $m_\tau = a_\tau \cdot \mathrm{MAP}^{b_\tau}$. This log linear relationship is different from the linear one bt *Castellarin et al.* [2009]. Given these assumptions, instead of equation (1), we have that

$$\frac{h_{\tau,T}}{\mathrm{MAP}} = k_{\tau,T} \cdot \frac{m_\tau}{\mathrm{MAP}} = k_{\tau,T} \cdot a_\tau \cdot \mathrm{MAP}^{(b_\tau - 1)}. \qquad (17)$$

Figure 5 shows the log linear relationships $m_\tau/\mathrm{MAP} = a_\tau \cdot \mathrm{MAP}^{(b_\tau - 1)}$ fitted to the mean AMS for three durations. A large portion of the noise that characterizes Figure 5 originates from sampling variability. To assess this quantitatively, in Figure 5 we show also the 90% confidence bounds evaluated assuming that the mean AMS follows a normal distribution with mean m_τ and standard deviation equal to $\hat{\sigma}_\tau/\sqrt{n}$, where n is the record length and $\hat{\sigma}_\tau$ is an estimate of the "population" standard deviation different from the sample standard deviation. Since we assume L-CV to be constant over the region, which is a reasonable assumption for most of the durations as shown in Figure 4 and Table 2, the same can be said for CV. Therefore we calculate the regional CV as weighted mean of the sample CVs of the 73 samples and we estimate the spatially varying population standard deviation as the regression line between m_τ and MAP multiplied by this regional CV.

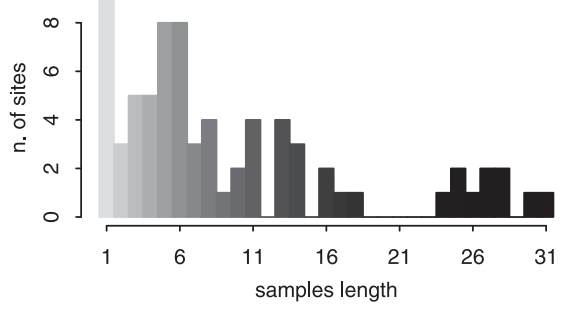

Figure 3. Histogram of the samples length (in years) for the 73 stations. The grayscale used in the following figures refers to the samples length as in this figure.

Table 1. Statistics on the Rainfall Data

No. of gauges	73
Altitude (m a.s.l)	493 (min)
	1297 (mean)
	2850 (max)
MAP (mm)	548 (min)
	1110 (mean)
	1732 (max)
Series length (years)	1 (min)
	10 (mean)
	31 (max)
Station years of data	695

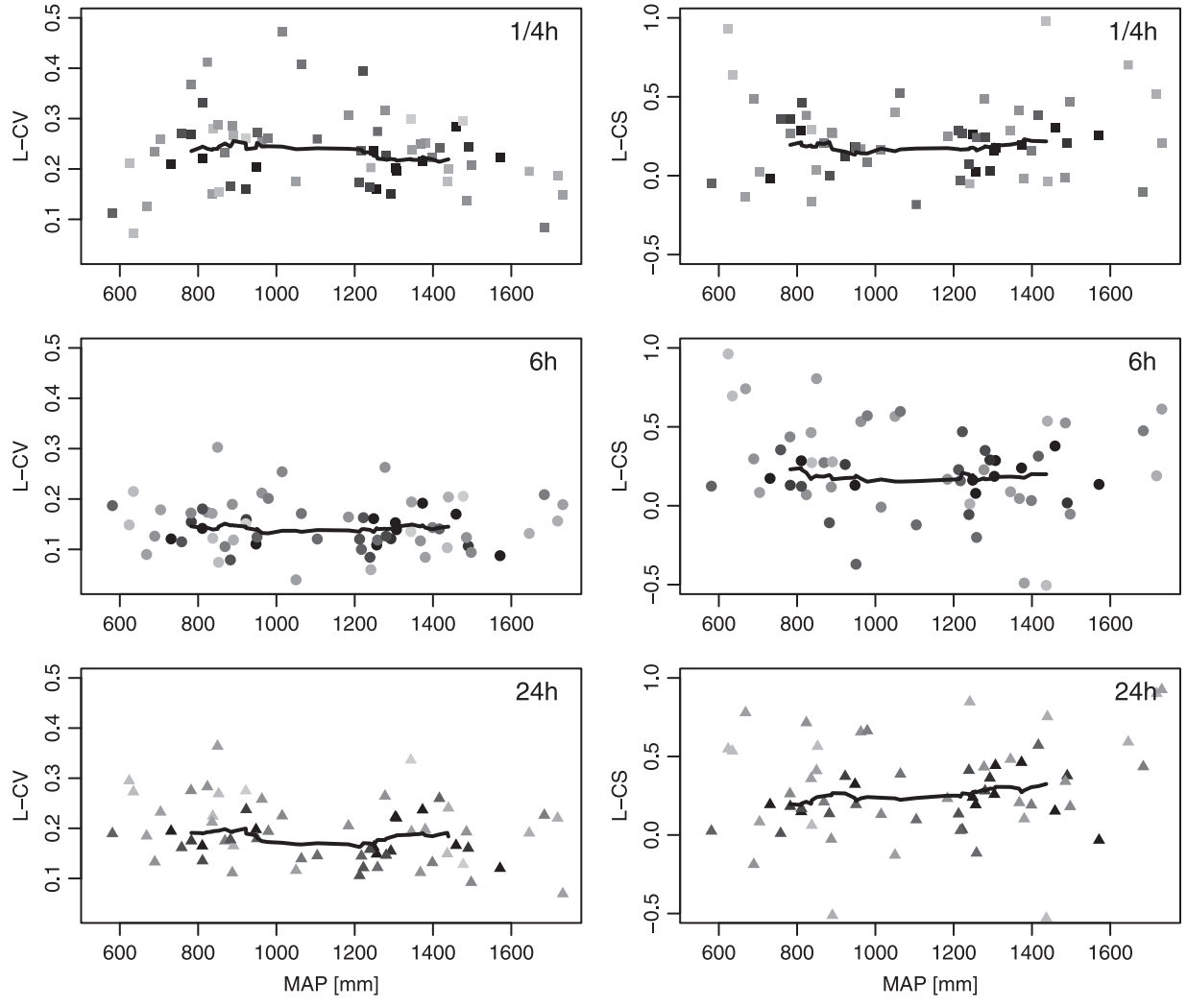

Figure 4. Relationship between sample L-CV and L-CS of the annual maximum rainfall depths for different durations (1/4, 6, and 24 h) for Tyrol. Weighted moving average curves are also indicated. The grayscale is proportional to the sample lengths, consistent with Figure 3.

[28] The computed 90% confidence bounds contain the regional regression 48 times for 1/4 h, 56 times for 1 h, 63 times for 3, 6, and 12 h, and 67 times for 24 h (over 73 samples). Due to the short or very short record lengths of many of the series used in the study, sampling variability alone explains a large amount of the noise present in the data in most of the cases. Similarly to the homogeneity test of Table 2, the scaling between m_τ and MAP is very good for long durations (3 to 24 h) while it is not for short and very short durations (e.g., 1/4 to 1 h). The limited accuracy of the scaling law for short durations is not a surprising result and may be interpreted as a consequence of the limited correlation of rainfall depth for convective storms (i.e., short duration) with the mean annual precipitation. This behavior for subhourly durations holds for various regions of the world and is well documented in the literature [e.g., *Hershfield*, 1961; *Bell*, 1969; *Ferreri and Ferro*, 1990; *Alila*, 1999; *Castellarin et al.*, 2009; *Allamano et al.*, 2009].

Table 2. Heterogeneity Measure of *Hosking and Wallis* [1993] for Different Durations[a]

1/4 h	1 h	3 h	6 h	12 h	24 h
4.3	2.5	1.7	1.2	1.4	0.7

[a]Only 43 stations with at least 6 years of data have been used for a total of 606 station years. The region should be considered heterogeneous if the measure exceeds 2.

[29] The slope of the curves in Figure 5 determine the slope of the DDECs since these latter are obtained by multiplying them by the scaling factor $k_{\tau,T}$ as in equation (17). The values of $k_{\tau,\mathrm{MAX}}$ for the envelope curves were identified by enveloping all rainfall maxima observed in the study region through the following equation:

$$k_{\tau,\mathrm{MAX}} = \max_{j=1,2,\ldots,M}\left\{\frac{h_{\tau,\mathrm{MAX},j}}{a_\tau \cdot \mathrm{MAP}_j^{b_\tau}}\right\}, \tag{18}$$

where $h_{\tau,\mathrm{MAX},j}$ denotes the maximum rainfall depth observed for duration τ at site $j = 1, 2, \ldots, M$ and M is the

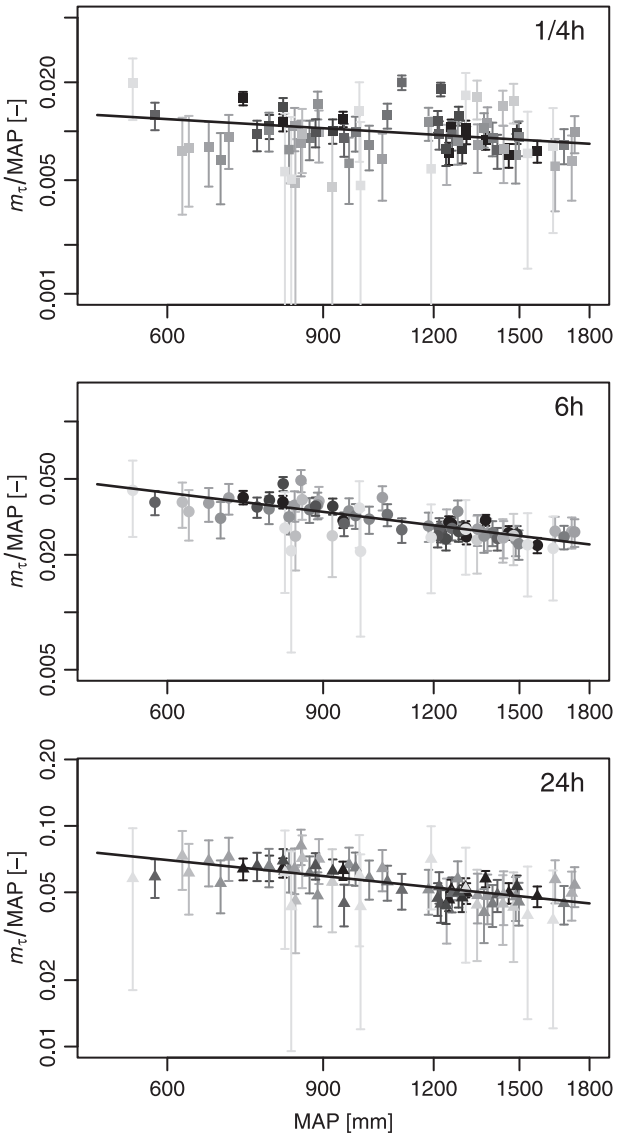

Figure 5. Empirical values of mean annual precipitation (MAP) versus scaled mean AMS rainfall depth (m_τ/MAP) for $\tau = 0.25$, 6, and 24 h. The 90% confidence bounds, accounting for sampling variability, are also indicated. The grayscale is proportional to the sample lengths, consistent with Figure 3. The regression lines are obtained with the weighted least-squares method where sample lengths are used as weight.

number of sites in the region, while MAP_j is the local value of the mean annual precipitation. This equation is analogous to equation (2) which was used by *Castellarin et al.* [2009]. The values of a_τ, b_τ, and $k_{\tau,\text{MAX}}$ for all DDECS are listed in Table 3.

[30] The schematization of the cross-correlation structure illustrated in Figure 6, and the application of the algorithm (equations (3) to (5)) described in section 2.1 to the available rainfall data, enabled us to estimate the exceedance probability of the envelope curves. The overall sample years of annual maximum rainfall depths, the estimated equivalent number of independent observation $\hat{n}_{\text{eff}}$, and the estimated value of the recurrence interval obtained by applying the Hazen plotting position to $\hat{n}_{\text{eff}}$ are listed in Table 3 for all durations of interest.

[31] Figure 7 illustrates the DDECs obtained for the study area and all durations of interest from $\tau = 15$ min to 24 h. The panels report point rainfall depths standardized by the local value of MAP and they illustrate the observed maximum rainfall depths and envelope curves. The fact that, for some stations, the regional m_τ (dashed line) is higher than the maximum observed rainfall is due to the length of the samples. All points below the regional m_τ but 1 correspond to samples of length ≤ 3. The envelope curves illustrated in Figure 7 and the corresponding estimates of the recurrence interval listed in Table 3 represent an easy-to-use graphical tool to (1) identify a plausible value of the extreme point rainfall depth at any location of the considered study area as a function of the local value of MAP (see Figure 2) for durations ranging from 15 min to 24 h and to (2) attach to this rainfall depth an estimate of the exceedance probability, expressed in terms of recurrence interval.

[32] It is important to note here that, differently from *Castellarin et al.* [2009] in which the shape of the envelope curves were checked to be approximately linear for the Gumbel and, to a lesser extent, GEV parent distributions of the rainfall extremes, the envelope curves given by equation (17) and represented in Figure 7 do not depend on the parent distribution.

3.2. Stochastic Rainfall Generator

[33] The stochastic rainfall model described in section 2.2 has been calibrated for each of 22 sites that are representative of the climatic conditions of the entire area (red points in Figure 2). Ranges for the parameters are given in Table 4. Figure 8 shows how some model parameters are

Table 3. Characteristics of the Annual Maximum Rainfall Depths for Different Durations; Calibrated Coefficients of the Cross-Correlation Formula (4); Empirical DDEC Parameters; and Estimated Recurrence Interval[a]

Duration τ (h)	1/4	1	3	6	12	24
No. of effective obs. $\hat{n}_{\text{eff}}$	663.9	679.5	665.0	484.1	251.3	157.6
λ_1 (km^{-1})	5.99×10^{-4}	8.87×10^{-4}	4.57×10^{-4}	5.40×10^{-4}	1.64×10^{-4}	3.26×10^{-5}
λ_2 (km^{-1})	2.05×10^{-4}	2.84×10^{-4}	1.41×10^{-4}	3.38×10^{-4}	1.64×10^{-4}	3.26×10^{-5}
a_τ	9.08×10^{-2}	4.90×10^{-1}	1.15	1.59	1.20	9.77×10^{-1}
b_τ	6.82×10^{-1}	5.18×10^{-1}	4.40×10^{-1}	4.33×10^{-1}	5.17×10^{-1}	5.88×10^{-1}
$k_{\tau,\text{MAX}}$	3.80	3.51	3.49	2.94	2.93	3.47
Recurrence interval (years)	1328	1359	1330	968	503	315

[a]The number of stations considered are 73 for a total of 695 observations.

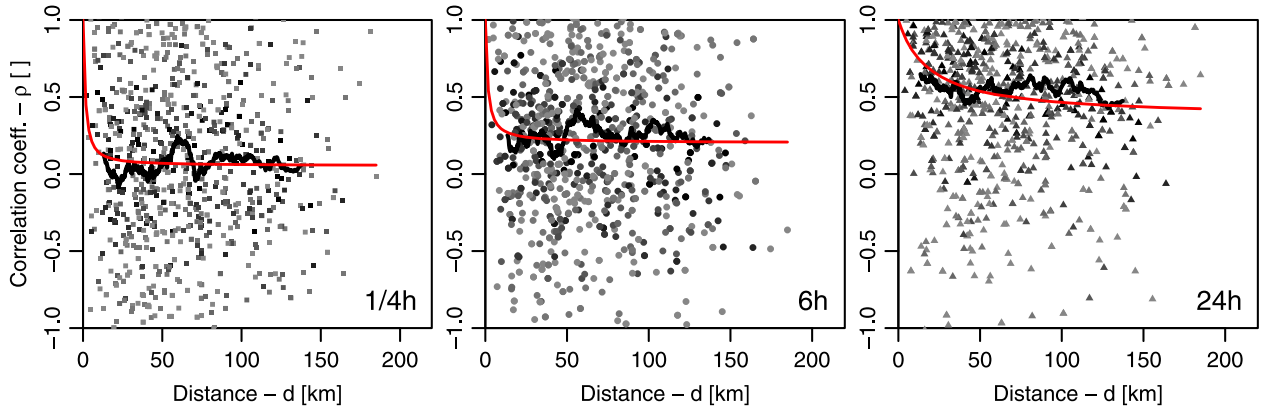

Figure 6. Sample cross-correlation coefficients (dots); weighted moving average curve (thick black line); correlation formula (4) calibrated for the whole study area (red line) for different durations (1/4, 6, and 24 h). The grayscale is proportional to the length of the overlapping series and is consistent with Figure 3.

obtained for the station Wattener Lizum, which is located at an elevation of 1970 m a.s.l. (local MAP value: 1292 mm) in the Wattenbach catchment, one of the tributary of the Inn river. For this station, about 25 years (1983–2008) of rainfall data at the resolution of 5 min are available. The time series has been processed in order to separate the storms by setting the minimum duration of dry periods as 3 h. It has been verified that the choice of a different minimum duration has an impact on the calibrated parameters of the bounded random cascade but not on the characteristics of the simulated series. Characteristics such as duration, average intensity, and mass curves were extracted from 3,761 identified storm events. The average storm intensity is about 1.4 mm h^{-1}, but storms with average intensity exceeding 10 mm h^{-1} have been observed (which are convective summer storms with short durations, i.e., less than 2 h).

[34] In Figure 8a the curve of equation (9) is fitted to the monthly average storm durations. The parameters ξ_r, δ_r, α_{1r}, ϕ_{1r}, α_{2r}, and ϕ_{2r} are obtained through a least-squares method for two-component Fourier series. The parameter β_r is obtained by inverting equation (8) numerically. Analogously, in Figure 8b, a two-component Fourier series is fitted to the monthly average interstorm periods and the estimates of ξ_b, δ_b, α_{1b}, ϕ_{1b}, α_{2b}, and ϕ_{2b} are obtained. Figure 8c represents the seasonality of the average storm intensity i_m. Since i_m depends on the storm duration t_r, the measured values have been standardized as $i_m/t_r^{b_1}$ in order to be able to represent the seasonality in one unique curve. The monthly averages ($\overline{i_m/t_r^{b_1}}$) are then used to obtain the seasonal evolution of the parameter a_1 as in equation (13). The dependence between i_m and t_r is accounted for by fitting the parameters of equation (12). Figure 8d shows the fit of a log linear relationship between CV $^2[i_m|t_r]$ and t_r where the data points have been obtained by grouping the events in classes of durations and evaluating the coefficient of variation of their average intensities. The coefficients of this regression give the estimates for a_2 and b_2. For the parameters a_0 and H, the BDC technique described by *Menabde and Sivapalan* [2000] has been used.

[35] The calibrated model has been used to generate 1,000,000 years of synthetic point rainfall time series for the 22 sites highlighted in red in Figure 2 at the 15 min time step. For Wattener Lizum, Figure 9 shows a comparison between some characteristics of the observed and simulated series.

[36] Figures 9a and 9b show the seasonal variation of mean precipitation during wet periods for short (1 h) and long (24 h) aggregation levels. Figures 9c and 9d show the standard deviation of precipitation during wet periods. The gray points are the statistics calculated from the observations (no more than 25 points for each month), the black line is their mean, while the green lines represent the mean and 5 and 95% quantiles of the simulations. For $\tau = 1$ h (Figure 9a) the simulated mean precipitation during wet periods slightly underestimates the observations both on average and in variability. Its standard deviation is overestimated during winter (Figure 9c). This is due to the fact that the bounded random cascade model is parameterized without considering the seasonal variations. For Wattener Lizum assuming constant parameters for the time patterns is apparently not optimal (for many other stations it is not the case). It is worth noting that winter precipitation at Wattener Lizum is mainly snow and that measurement errors are possible. For $\tau = 24$ h the fit between observations and simulations is better. The seasonal pattern is in both cases well represented.

[37] Figures 9e and 9f show the seasonal variation of the zero depth frequency (percentage of no rain) for short (1 h) and long (24 h) durations. The agreement between simulations and observations is generally good, very good for 1 h, and characterized by a slight overestimation for 24 h. Figure 9g shows the monthly rainfall depths. As can be seen, for Wattener Lizum the agreement between these cumulative statistics of observed and simulated rainfall is good.

[38] The behavior of extreme rainfall is shown in Figure 9h. For aggregations equal to 1, 3, 6, 12, and 24 h, the figure shows the box-plots of the observed and simulated annual maximum rainfall depths (AMS). For Wattener Lizum, the simulated rainfall extremes are slightly higher than the observed ones. This is not a general pattern, i.e., for other stations the quantiles from the data are higher

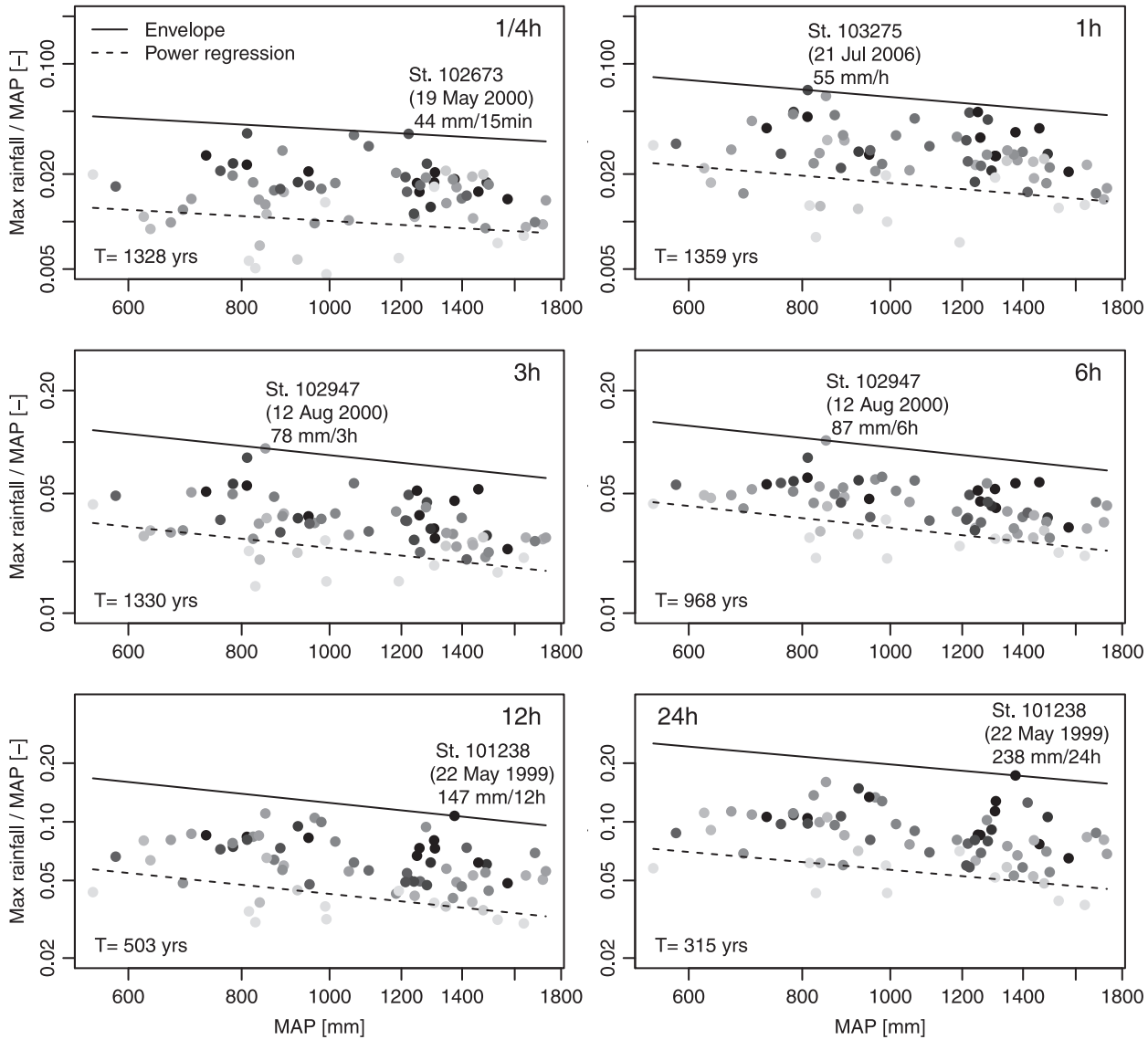

Figure 7. Empirical depth duration envelope curves for different durations (1/4 to 24 h) in Tyrol. The circles represent mean annual precipitation (MAP) versus the rescaled maximum recorded rainfall depth for the 73 rainfall stations. The grayscale is proportional to the sample lengths, consistent with Figure 3. The envelope curve of equation (17) is indicated by the continuous line. The dashed line represents the relation between (m_T/MAP) and MAP (see Figure 5).

than the simulated quantiles. It is important to stress here that the calibration was not done by optimizing an objective function built on the simulated series. Through an objective function we could have obtained a better match between observed and simulated AMS statistics, but we would have missed the opportunity of directly calibrating the parameters in equations (6) to (16). The fit to extremes is validated here using regional information, as we show in section 4 by comparing the quantiles of the simulated series to the regional DDECs.

4. Comparison of Methods and Discussion

[39] The DDEC technique enables one to estimate the return period of the envelope curves (Table 3 and Figure 7). It is now of interest to see if these estimates of the return period of the curves are consistent with what can be obtained through the stochastic rainfall model. We follow the procedure described in section 2.3. Analyzing the 1,000,000 year series for each simulated station, the quantiles of the rainfall depths for durations of 1/4, 1, 3, 6, 12, and 24 h are extracted for the return periods estimated by the DDEC procedure of section 3 as listed in Table 3.

[40] The comparison is shown in Figure 10. The observed records of the 22 stations and the regional envelope curves are in black, while the quantiles obtained by simulation and the 90% confidence bounds, which assess the influence of sampling variability on the envelope curve, are in green. In the ideal case in which the two methods are perfectly consistent, the quantiles obtained by stochastic modeling lay around the envelope curve and the 90% confidence bounds contain the curve 90% of the times.

Table 4. Ranges for the Parameters of the Stochastic Rainfall Model Calibrated for the 22 Stations as Described in Section 2.2[a]

	0%	25%	50%	75%	100%
ξ_r (h)	0.5	0.5	0.5	0.5	0.5
δ_r (h)	3.92	6.61	7.42	8.09	9.80
α_{1r} (h)	0.460	0.805	1.33	1.81	2.51
τ_{1r} (months)	−3.91	−0.643	-3.70×10^{-3}	0.237	1.25
α_{2r} (h)	0.117	0.406	0.514	0.728	1.48
τ_{2r} (months)	−2.93	−2.61	−1.84	1.69	2.99
β_r	0.778	0.878	0.907	0.952	1.06
ξ_b (h)	3	3	3	3	3
δ_b (h)	42.0	44.7	51.9	61.4	80.7
α_{1b} (h)	5.99	14.5	15.6	26.1	40.3
τ_{1b} (months)	−0.856	−0.211	0.115	0.440	1.10
α_{2b} (h)	1.31	3.67	5.75	8.43	11.6
τ_{2b} (months)	−2.93	−2.70	−2.22	2.55	2.94
β_b	0.661	0.682	6.92	0.733	0.817
δ_a (mm h$^{-b_1-1}$)	1.06	1.33	1.41	1.54	1.83
α_{1a} (mm h$^{-b_1-1}$)	0.362	0.451	0.554	0.692	0.892
τ_{1a} (months)	−5.86	−5.41	−5.24	−5.18	−4.29
α_{2a} (mm h$^{-b_1-1}$)	0.102	0.159	0.201	0.240	0.292
τ_{2a} (months)	−0.113	0.393	0.565	0.817	2.24
a_2	0.622	1.63	1.83	2.79	7.17
b_1	−0.365	−0.234	−0.197	−0.170	−0.0253
b_2	−0.991	−0.562	−0.456	−0.337	−0.156
a_0 (h^H)	2.10	2.70	3.50	3.50	4.00
H	0.270	0.380	0.450	0.450	0.500

[a]The minimum (0%), median (50%), maximum (100%) as long as the 25% and 75% quantiles over 22 values are listed.

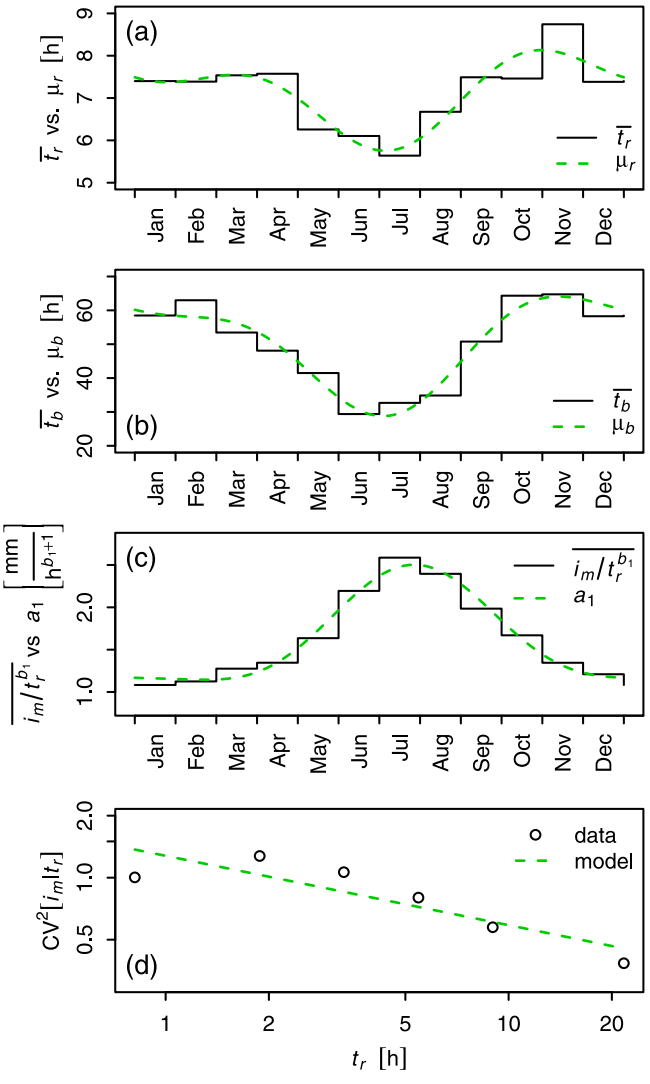

Figure 8. Calibration of some of the model parameters for the rainfall station N103291 (Wattener Lizum). Seasonal variation of: (a) mean monthly storm duration; (b) mean monthly interstorm period; (c) mean monthly standardized average storm intensity. Panel (d) shows the modeled relationship between the variability of the average storm intensity and the storm duration.

[41] Panels of Figure 10 show that the agreement between empirical DEECs and synthetic rainfall quantiles ranges from fair to excellent. This is definitely an encouraging outcome of our study. It is worth remarking here that the two procedures are profoundly different in nature and are implemented (DDECs) and calibrated (stochastic rainfall generator) on the basis of completely independent sources information: record rainfall peaks and cross-correlation structure (DDECs) and individual rainfall events characteristics (stochastic rainfall generator). In particular, Figure 10 clearly illustrates that the match between the two procedures is excellent for the intermediate durations, i.e., τ from 1 to 6 h. The 22 green box-plots lay around the black envelope curve and the 90% confidence bounds contain the curve 20, 21, and 21 times (for 1, 3, and 6 h, respectively). It is worth noting that these τ values match the empirical estimates of the expected storm duration, which ranges for the study area between 4.4 and 10.3 h (i.e., range of $\xi_r + \delta_r$ values calibrated on the 22 stations used for the stochastic rainfall generation, see Table 4).

[42] Results are poorer for very short and very long τ values (i.e., 1/4 h, and 12 and 24 h). The physical reasons for this behavior are very different in nature for the two cases. For 1/4 h, most of the green points lay below the envelope curve meaning that the (regional) return period predicted by the DDEC approach is on average lower than the (local) return period obtained through simulations. On the other hand, the 90% confidence bounds almost always include the envelope curve (21 times over 22 stations) suggesting that, being the assumptions of the DDEC method satisfied, the discordance could be attributed to sampling variability. As shown in section 3, however, the validity of the scaling law of m_τ with MAP and of regional homogeneity are questionable for $\tau = 1/4$ h. *Castellarin* [2007] clearly states that the presence of a limited number of discordant sites in the region, that is sites characterized by markedly different frequency distributions of rainfall extremes, has a significant influence on the empirical envelope, namely determining an overestimation of the curve itself. Also, regarding the stochastic rainfall model, in many cases rainfall data are of poor quality for $\tau = 1/4$ h relative to longer aggregation intervals (i.e., time series retrieved from tipping bucket rain gauges are sensitive to the bucket volume for short time intervals), thus affecting the parameter estimation of the stochastic rainfall model. Therefore, the top left panel of Figure 10 illustrating the results for $\tau = 1/4$ should be regarded as an approximate indication of the agreement between the two procedures for subhourly durations, and it is not as informative and accurate as the panels associated with larger τ values.

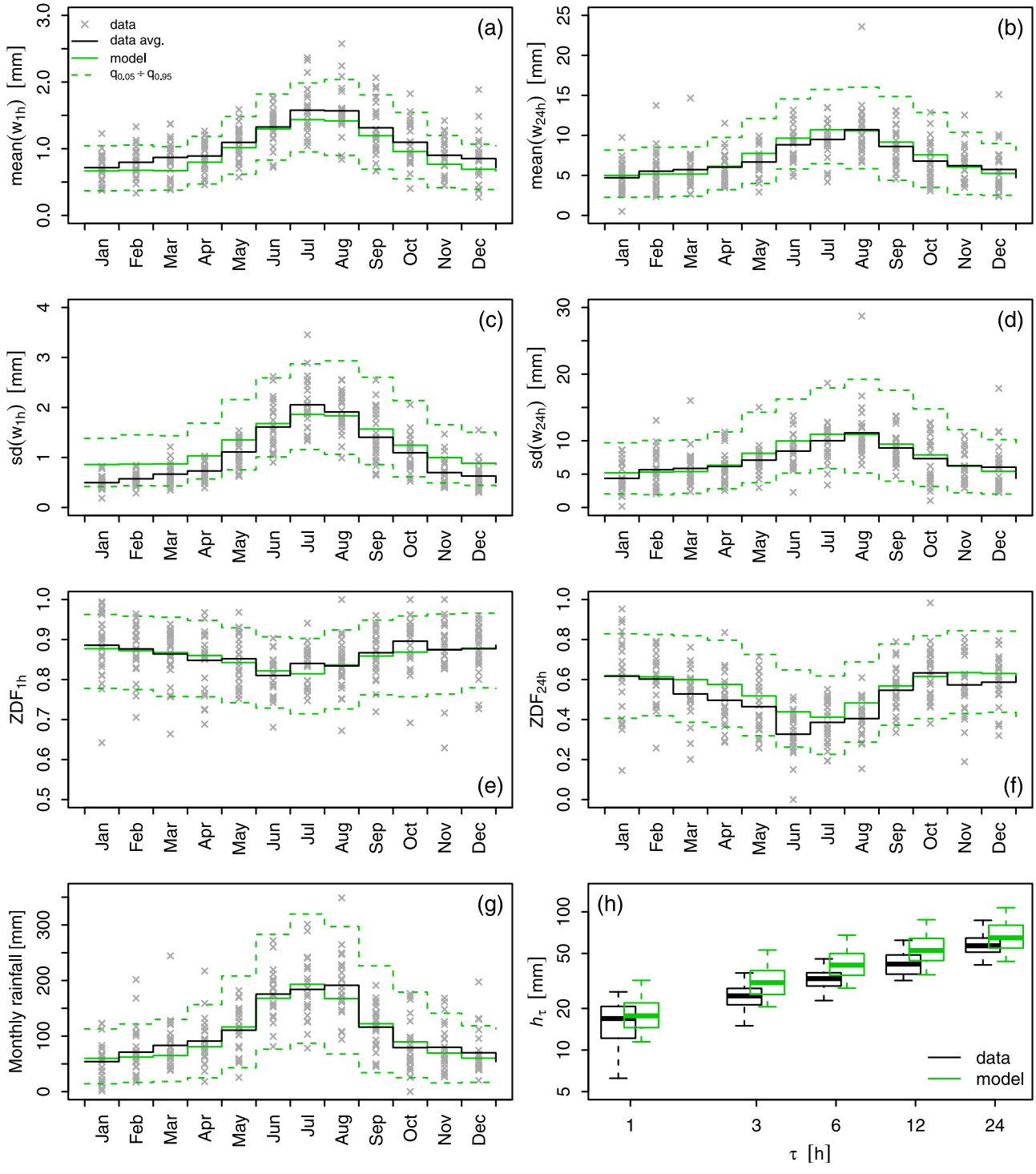

Figure 9. Validation of some of the stochastic rainfall model results for the rainfall station N103291 (Wattener Lizum): (a) seasonal variation of mean precipitation during wet periods for 1 h duration (all observations are in gray, the averaged observed is black, the modeled average and 5%–95% quantiles are green); (b) seasonal variation of mean precipitation during wet periods for 24 h duration; (c) seasonal variation of the standard deviation of precipitation during wet periods for 1 h duration; (d) seasonal variation of the standard deviation of precipitation during wet periods for 24 h duration; (e) seasonal variation of the zero depth frequency (ZDF) for 1 h duration; (f) seasonal variation of the ZDF for 24 h duration; (g) seasonal variation of the monthly total precipitation; (h) observed and modeled maximum annual rainfall depths for different durations (the box-plots show the 5%, 25%, 50%, 75%, and 95% quantiles).

[43] For $\tau = 24$ h and, to a lesser extent, $\tau = 12$ h the quantiles obtained by simulation are significantly lower than the envelope curve for the same return period. In both cases, it is the record rainfall observed at Reutte (station N101238) in May 1999 (190 mm/day), that sets the envelope curve. Dropping station N101238 results in a much better agreement, as shown in Figure 10 (thin line). In particular, a very good match can be obtained for $\tau = 12$ h (for which the envelope is contained 20 times in the 90% confidence intervals), and a much better match can also be obtained for $\tau = 24$ h even though DDEC still stays over the synthetic rainfall quantiles in this case (and only for 12 cases over 22 the 90% confidence intervals contain the envelope curve). Notice that, by dropping the maximum values, we are not evaluating the return periods of the second maximum regional events but performing a sensitivity analysis.

[44] Unfortunately, station N101238 is not included in the set of daily rain gauges listed in Austrian hydrological yearbooks [*Hydrographischer Dienst in Österreich*, 2007]. Therefore a direct validation of this observed rainfall depth was not possible. Nevertheless, during the same event similar daily rainfall depths were observed in neighboring rain gauges. For this reason we decided to keep the rainfall event and the entire record observed at Reutte.

[45] This event is functional to illustrate the remarkable sensitivity of the DDEC procedure to abnormally large

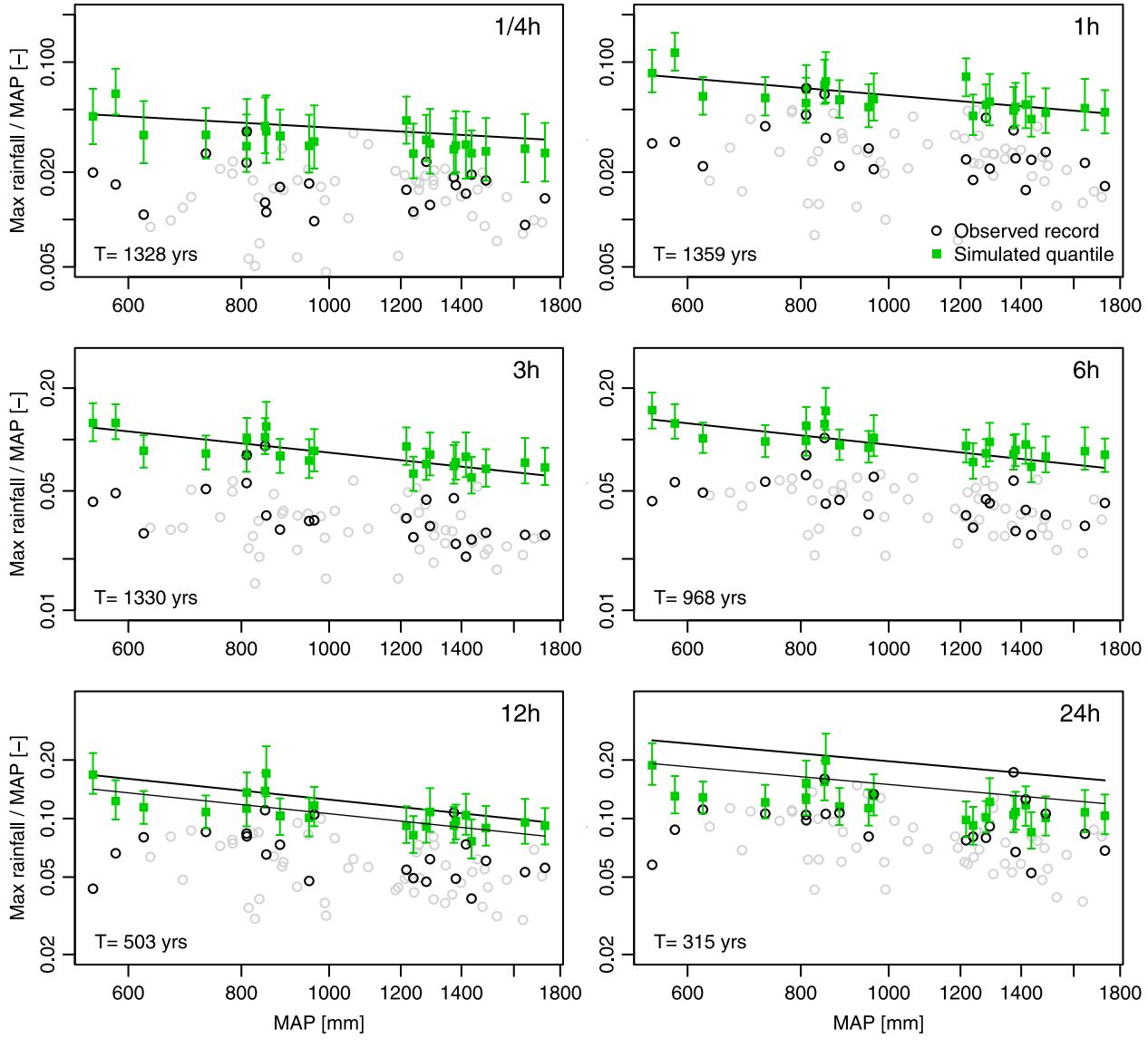

Figure 10. Comparison between empirical DDECs and simulated rainfall quantiles for the return period given by the DDEC procedure. The figure is analogous to Figure 7 but only the 22 stations used for the comparison are highlighted (black empty circles). The simulated quantiles, resulting from the stochastic generation of 1,000,000 years of rainfall, are indicated by green full squares. The 90% confidence bounds obtained by considering the maximum rainfall depths of 1000 repetitions of $\hat{n}_{\text{eff}}$ simulated years are also indicated in green. For 12 and 24 h the envelope obtained discarding station N101238 is shown as a thin line.

extreme events (outliers). One single peak sets the envelope. Therefore the assumption of it being representative of the entire region is very delicate. Underestimation of the return period of the envelope curve can be due to the fact that the event is actually an outlier, i.e., that a problem of ergodicity exists (the pooled sample is not representative of the population). Analogously but from the opposite point of view, overestimation of the position of the envelope can be due to spatial heterogeneity of the data or data errors. Regarding spatial heterogeneity, *Castellarin* [2007] shows that the impact of discordant sites can be significant when the region (or group of sites) is heterogeneous in terms of sample L-moments variability, which is not the case here, at least for durations larger than 1 h. It is worth noting that the comparison between the envelopes and the simulated quantiles, though, enabled us to detect the presence of gross errors in annual sequences of three other rain gauges. The large overestimation of simulated rainfall quantiles by empirical DDECs suggested to further check the validity of the observed rainfall data also for those cases. Gross data errors were identified by cross checks with institutions responsible for data collections (Hydrographic Service Tyrol) and rain gauges were discarded. Note that, given the nature of the comparison, errors may be detected also for stations without simulations.

[46] The reason behind the significant mismatch between the two methods for long durations is likely to result from the Reutte event being an outlier, but also from the nature of the stochastic generator used in the study. In the stochastic model, long duration intensities are controlled by the gamma distribution, while short duration intensities by the bounded random cascade for interstorm time patterns. The fact that the gamma distribution is not fat-tailed probably leads to an underestimation of the extreme quantiles for long durations and other distributions should be used (*Menabde and Sivapalan* [2000] use, for example, the Levy-stable distribution). For short durations, instead, the bounded random cascade produces fat-tailed extremes that better fit real extreme rainfall. This was checked by analyzing the tails of the simulated maximum annual rainfall intensities for short and long durations for the station Wattener Lizum (N103291). The graphical methods proposed by *Beirlant et al.* [2006] and *El Adlouni et al.* [2008] were used (for reasons of space the graphs are not shown here) suggesting that the simulated AMS for short durations (e.g., 1/4 h) have heavy right tails while the AMS for long durations (e.g., 24 h) have subexponential tails.

[47] The comparison between the regional DDEC method and the local stochastic simulations has another evident advantage. The DDEC procedure allows us to assign return periods to the envelope curves of rainfall peaks on a regional basis. In principle, one could extend the DDEC procedure to retrieve the return periods of all regional records in a nonparametric way, similarly to the plotting positions used for local samples. Anyway the extension is not straightforward. In fact, the assessment of the number of effective years n_{eff}, which indicate the informative content for the data, is referred here to the maximum only (see *Castellarin et al.*, [2005] and equation (3) here) and a different model for the exponent β in (3) should be derived for other order statistics (the second regional maximum, the third, etc.). In any case, the DDEC procedure does not allow us to evaluate peak quantiles for a given return period. The independent validation of the stochastic model through the DDECs implies that one can use the synthetic rainfall series to obtain any quantile associated with a recurrence interval. From a regional perspective, if one regionalizes the stochastic model parameters, the check with the DDECs would be particularly useful.

5. Conclusions

[48] DDECs were derived for a wide study region in Austria (Tyrol) for durations ranging from 15 min to 24 h and the corresponding recurrence intervals have been estimated. Relative to previous studies, some generalizations of the DDEC procedure was proposed that enables us to better represent the Austrian data. The curves have been then compared with rainfall quantiles estimated from long synthetic rainfall series. The two approaches have an independent nature since are based/calibrated one (DDEC) on the rainfall maxima and the other (stochastic model) on the time series of rainfall events. The comparison of the methods is twofold. First, it enables us to verify how realistic probabilistic envelope curves are for the study area. For our case study, the agreement between the two alternative methods is extremely good for intermediate durations (from 1 to 6 h), thus reassuring us about the reliability of estimation of the DDECs return periods. This check confirms the validity of the estimator of the effective number of observations (n_{eff}) in a different geographical/climatic context than in previous studies. Second, the comparison enables us to assess the reliability of rainfall quantiles obtained for large return periods from the long synthetic rainfall series. The model, and in particular the bounded random cascade for interstorm time patterns, seems to reproduce well the (fat-tailed) behavior of rainfall extremes for these durations (from 1 to 6 h).

[49] When mismatch between the methods was found, the data leading to the envelope curves were carefully analyzed. For instance, we detected the presence of gross errors in annual sequences of some of the stations in the region which were then discarded. For long durations (12 and 24 h) the mismatch found at the end of the analysis is not due to data errors because the event that sets the envelopes really happened. Is the event an outlier whose return period is much greater than the one estimated with the DDEC approach? Does the stochastic model fail to mimic extreme rainfall events for long durations? Most likely, the mismatch is due to a combination of these two causes. The usefulness of the comparison consists in the possibility to critically discuss the results.

[50] Most importantly, the two methods are complementary: their combined use (as in this case) can be very useful for estimating reliable rainfall quantiles associated with large return periods at any point within regions where many rain gauges are available, but with limited record lengths. For instance, if a regionalization of the stochastic model parameters is performed, extreme events that can be simulated by applying the regionalized rainfall generator may be compared with the regional DDEC retrieving useful indication on the reliability of synthetic rainfall extremes in ungauged locations. The results presented in this paper are promising and further tests in different geographical/climatic contexts are envisaged.

[51] **Acknowledgements.** Financial support for the project "Mountain floods – regional joint probability estimation of extreme events" from the Austrian Academy of Sciences is acknowledged. We would like to thank the reviewers and specially Eric Gaume, whose comments enabled us to significantly improve the paper.

References

Alila, Y. (1999), A hierarchical approach for the regionalization of precipitation annual maxima in Canada, *J. Geophys. Res. Atm.*, *104*(D24), 31,645–31,655.

Allamano, P., P. Claps, F. Laio, and C. Thea (2009), A data-based assessment of the dependence of short-duration precipitation on elevation, *Phys. Chem. Earth*, *34*(10–12), 635–641, doi:10.1016/j.pce.2009.01.001.

Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*(2), 703–710.

Beirlant, J., T. De Wet, and Y. Goegebeur (2006), A goodness-of-fit statistic for Pareto-type behaviour, *J. Comput. Appl. Math.*, *186*(1), 99–116, doi:10.1016/j.cam.2005.01.036.

Bell, F. (1969), Generalized rainfall depth-duration-frequency relationships, *J. Hydraul. Div.*, *95*(1), 327–331.

Brath, A., A. Castellarin, and A. Montanari (2003), Assessing the reliability of regional depth-duration-frequency equations for gaged and ungaged sites, *Water Resour. Res.*, *39*(12), 1367, doi:10.1029/2003WR002399.

Burton, A., H. Fowler, C. Kilsby, and P. O'Connell (2010), A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts, *Water Resour. Res.*, *46*, W11501, doi:10.1029/2009WR008884.

Cameron, D., K. J. Beven, and J. A. Tawn (2001), Modelling extreme rainfalls using a modified random pulse Bartlett–Lewis stochastic rainfall model (with uncertainty), *Adv. Water Resour.*, *24*(2), 203–211, doi:10.1016/S0309-1708(00)00042-7.

Castellarin, A. (2007), Probabilistic envelope curves for design flood estimation at ungauged sites, *Water Resour. Res.*, *43*, W04406, doi:10.1029/2005WR004384.

Castellarin, A., D. H. Burn, and A. Brath (2001), Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, *J. Hydrol.*, *241*, 270–285.

Castellarin, A., R. M. Vogel, and N. Matalas (2005), Probabilistic behavior of a regional envelope curve, *Water Resour. Res.*, *41*, W06018, doi:10.1029/2004WR003042.

Castellarin, A., R. Merz, and G. Blöschl (2009), Probabilistic envelope curves for extreme rainfall events, *J. Hydrol.*, *378*(3–4), 263–271.

Castiglioni, S., L. Lombardi, E. Toth, A. Castellarin, and A. Montanari (2010), Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, *Adv. Water Resour.*, *33*(10), 1235–1242, doi:10.1016/j.advwatres.2010.04.009.

Cowpertwait, P. (1994), A generalized point process model for rainfall, *Proc. Roy. Soc. London A*, *447*, 23–37.

Cunnane, C. (1978), Unbiased plotting positions—A review, *J. Hydrol.*, *37*, 205–222.

Di Baldassarre, G., A. Castellarin, and A. Brath (2006), Relationships between statistics of rainfall extremes and mean annual precipitation: An application for design-storm estimation in northern central Italy, *Hydrol. Earth Syst. Sci.*, *10*(4), 589–601.

Eagleson, P. S. (1972), Dynamics of flood frequency, *Water Resour. Res.*, *8*(4), 878.

El Adlouni, S., B. Bobée, and T. B. Ouarda (2008), On the tails of extreme event distributions in hydrology, *J. Hydrol.*, *355*(1–4), 16–33, doi:10.1016/j.jhydrol.2008.02.011.

Ferreri, G., and V. Ferro (1990), Short-duration rainfalls in Sicily, *J. Hydraul. Eng.*, *116*(3), 430–435.

Fiorentino, M., and V. Iacobellis (2001), New insights about the climatic and geologic control on the probability distribution of floods, *Water Resour. Res.*, *37*(3), 721–730, doi:10.1029/2000WR900315.

Fowler, H., C. Kilsby, P. O'Connell, and A. Burton (2005), A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change, *J. Hydrol.*, *308*(1–4), 50–66, doi:10.1016/j.jhydrol.2004.10.021.

Gaume, E. (2006), On the asymptotic behavior of flood peak distributions, *Hydrol. Earth Syst. Sci.*, *10*(2), 233–243.

Harris, D., M. Menabde, A. Seed, and G. Austin (1998), Breakdown coefficients and scaling properties of rain fields, *Nonlin, Processes Geophys.*, *5*(2), 93–104.

Herschfield, D. (1961), Estimating the probable maximum precipitation, *J. Hydraul. Div.*, *87*(HY5), 99–106.

Hosking, J. (1990), L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc.*, *52*, 105–124.

Hosking, J., and J. Wallis (1993), Some statistics useful in regional frequency analysis, *Water Resour. Res.*, *29*(2), 271–281.

Hosking, J., and J. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge.

Hundecha, Y., T. B. Ouarda, and A. Bárdossy (2008), Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the 'spatial' structures of the parameters within a canonical physiographic-climatic space, *Water Resour. Res.*, *44*(1), W01427, doi:10.1029/2006WR005439.

Hydrographischer Dienst in Österreich (2007), Hydrographische Jahrbuch von Österreich, *Tech. rep.*, Hydrographiscen Zentralbüro im Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien.

Jenkinson, A. (1955), The frequency distribution of the annual maximum (or minimum) of meteorological elements, *Q. J. R. Meteorol. Soc.*, *81*, 158–171.

Klemeš, V. (1993), Probability of extreme hydrometeorological events—A different approach, in *Extreme Hydrological Events: Precipitation, Floods and Droughts*, Vol. 213, edited by Z. Kundzewicz, pp. 167–176, IAHS, Yokohama, Japan.

Matalas, N., and W. Langbein (1962), Information content of the mean, *J. Geophys. Res.*, *67*(9), 3441–3448.

Menabde, M., and M. Sivapalan (2000), Modeling of rainfall time series and extremes using bounded random cascades and Levy-stable distributions, *Water Resour. Res.*, *36*(11), 3293–3300, doi:10.1029/2000WR900197.

Menabde, M., A. Seed, D. Harris, and G. Austin (1997), Self-similar random fields and rainfall simulation, *J. Geophys. Res. Atm.*, *102*(D12), 13,509–13,515.

Merz, R., and G. Blöschl (2008), Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, *44*, W08432, doi:10.1029/2007WR006744.

Novikov, E. (1966), A mathematical model of intermittency of turbulent flow, *Doklady Akademii Nauk SSSR*, *168*(6), 1279–1282.

Novikov, E. (1969), Scale similitude for random fields, *Doklady Akademii Nauk SSSR*, *184*(5), 1072–1075.

Onof, C., R. Chandler, A. Kakou, P. Northrop, H. Wheater, and V. Isham (2000), Rainfall modelling using Poisson-cluster processes: A review of developments, *Stochastic Environ. Res. Risk Assess.*, *14*(6), 384–411, doi:10.1007/s004770000043.

Parajka, J., G. Blöschl, and R. Merz (2007), Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, *43*(6), W06406, doi:10.1029/2006WR005271.

Pilgrim, D. H., and I. Cordery (1993), Flood runoff, in *HandBook of Hydrology*, edited by D. R. Maidment, International ed., Chap. 9, p. 42, McGraw-Hill, New York.

Robson, A. J., and D. W. Reed (1999), Statistical procedures for flood frequency estimation, in *Flood Estimation Handbook*, Vol. 3, p. 338, Institute of Hydrology Crowmarsh Gifford, Wallingford, Oxfordshire.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London A*, 410, 269–288.

Schaefer, M. (1990), Regional analyses of precipitation annual maxima in Washington state, *Water Resour. Res.*, *26*(1), 119–131.

Sivapalan, M., G. Blöschl, R. Merz, and D. Gutknecht (2005), Linking flood frequency to long-term water balance: Incorporating effects of seasonality, *Water Resour. Res.*, *41*(6), W06012, doi:10.1029/2004WR003439.

Stedinger, J. R. (1983), Estimating a regional flood frequency distribution, *Water Resour. Res.*, *19*, 503–510.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. R. Maidment, International ed., Chap. 18, pp. 18.1–18.66, McGraw-Hill, New York.

Tasker, G., and J. R. Stedinger (1989), An operational GLS model for hydrologic regression, *J. Hydrol.*, *111*, 361–375.

Verhoest, N. E. C., P. A. Troch, and F. P. De Troch (1997), On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point, *J. Hydrol.*, *202*(1–4), 108–120, doi:10.1016/S0022-1694(97)00060-7.

Verhoest, N. E. C., S. Vandenberghe, P. Cabus, C. Onof, T. Meca-Figueras, and S. Jameleddine (2010), Are stochastic point rainfall models able to preserve extreme flood statistics?, *Hydrol. Processes*, *24*(23), 3439–3445, doi:10.1002/hyp.7867.

Viglione, A., and G. Blöschl (2009), On the role of storm duration in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, *13*(2), 205–216, doi:10.5194/hess-13-205-2009.

Viglione, A., R. Merz, and G. Blöschl (2009), On the role of the runoff coefficient in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, *13*(5), 577–593, doi:10.5194/hess-13-577-2009.

Vogel, R. M., and I. Wilson (1996), Probability distribution of annual maximum, mean, and minimum streamflows in the united states, *J. Hydrol. Eng.*, *1*(2), 69–76.

Wagener, T., and H. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, *320*(1–2), 132–154, doi:10.1016/j.jhydrol.2005.07.015.

Zhang, Z., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resour. Res.*, *44*, W00B04, doi:10.1029/2008WR006833.

G. Blöschl, M. Rogger, and A. Viglione, Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Karlsplatz 13/E222, 1040 Vienna, Austria. (viglione@hydro.tuwien.ac.at)

A. Castellarin, DISTART, School of Civil Engineering, University of Bologna, Viale Risorgimento, 2, 40136, Bologna, Italy.

R. Merz, Helmholtz Centre for Environmental Research, Theodor-Lieser-Str. 4, 06120, Halle, Germany.