Simple Scaling of extreme precipitation in North America

Silvia Innocenti[1], Alain Mailhot[1], and Anne Frigon[2]

[1]Centre Eau-Terre-Environnement, INRS, 490 de la Couronne, Québec, Canada, G1K 9A9
[2]Consortium Ouranos, 550 Sherbrooke Ouest, Montrèal, Canada, H3A 1B9

Correspondence to: S. Innocenti (silvia.innocenti@ete.inrs.ca)

Abstract. Extreme precipitation is highly variable in space and time. It is therefore important to characterize precipitation intensity distributions at several temporal and spatial scales. This is a key issue in infrastructure design and risk analysis, for which Intensity-Duration-Frequency (IDF) curves are the standard tools used for describing the relationships among extreme rainfall intensities, their frequencies, and their durations. Simple Scaling (SS) models, characterizing the relationships among

5   extreme probability distributions at several durations, represent a powerful means for improving IDF estimates. This study tested SS models for approximately 2700 stations in North America. Annual Maxima Series (AMS) over various duration intervals from 15 h to 7 days were considered. The range of validity, magnitude, and spatial variability of the estimated scaling exponents were investigated. Results provide additional guidance for the influence of both local geographical characteristics, such as topography, and regional climatic features on precipitation scaling. Generalized Extreme Value (GEV) distributions

10   based on SS models were also examined. Results demonstrate an improvement of GEV parameter estimates, especially for the shape parameter, when data from different durations were pooled under the SS hypothesis.

1 Introduction

Extreme precipitation is highly variable in space and time as various physical processes are involved in its generation. Characterizing this spatial and temporal variability is crucial for infrastructure design and to evaluate and predict the impacts of

15   natural hazards on ecosystems and communities. Available precipitation records are however sparse and cover short time periods, making a complete and adequate statistical characterization of extreme precipitation difficult. The resolution of available data, whether observed at meteorological stations or simulated by weather and climate models, often mismatches the resolution needed for applications (e.g., *Blöschl and Sivapalan*, 1995; *Maraun et al.*, 2010; *Willems et al.*, 2012), thus adding to the difficulty of achieving complete and adequate statistical characterizations of extreme precipitation.

20   The need for multi-scale analysis of precipitation has been widely recognized in the past (*Rodriguez-Iturbe et al.*, 1984; *Blöschl and Sivapalan*, 1995; *Hartmann et al.*, 2013; *Westra et al.*, 2014, among others) and much effort has been put into the development of relationships among extreme precipitation characteristics at different scales. The conventional approach for characterizing scale transitions in time involves the construction of Intensity-Duration-Frequency (IDF) or the equivalent Depth-Duration-Frequency (DDF) curves (*Bernard*, 1932; *Burlando and Rosso*, 1996; *Sivapalan and Blöschl*, 1998; *Kout-*

25   *soyiannis et al.*, 1998; *Asquith and Famiglietti*, 2000; *Overeem et al.*, 2008; *Veneziano and Yoon*, 2013). These curves are a standard tool for hydraulic design and risk analysis as they describe the relationships between the frequency of occurrence of

extreme rainfall intensities (depth) X_d and various durations d (e.g., *CSA*, 2012). Analysis is usually conducted by separately estimating the statistical distributions of X_d at the different durations (see *Koutsoyiannis et al.*, 1998; *Papalexiou et al.*, 2013, for discussions about commonly used probability distributions). The parameters or the quantiles of these theoretical distributions are then empirically compared to describe the variations of extreme rainfall properties across temporal scales.

5   Despite its simplicity, this procedure presents several drawbacks. In particular, it does not guarantee the statistical consistency of precipitation distributions, independently estimated at the different durations, and it limits IDF extrapolation at non-observed scales or ungauged sites. Uncertainties of estimated quantiles are also presumably larger because precipitation distribution and IDF curve parameters are fitted separately.

Scaling models (*Lovejoy and Mandelbrot*, 1985; *Gupta and Waymire*, 1990; *Veneziano et al.*, 2007) based on the concept of

10   scale invariance (*Dubrulle et al.*, 1997), have been proposed to link rainfall features at different temporal and spatial scales. Scale invariance states that the statistical characteristics (e.g., moments or quantiles) of precipitation intensity observed at two different scales d and λd can be related to each other by a power law of the form:

$$f(X_{\lambda d}) = \lambda^{-H} f(X_d) \tag{1}$$

where $f(.)$ is a function of X with invariant shape when rescaling the variable X by a multiplicative factor λ and for some values of the exponent $H \in \mathbb{R}$. In the simplest case, a constant multiplicative factor adequately describes the scale change.

15   The corresponding mathematical models are known as *Simple Scaling* (SS) models (*Gupta and Waymire*, 1990). SS models are attractive because of the small number of parameters involved, as opposed to multiscaling (MS) models which involve more than one multiplicative factor in Eq. (1) (e.g., *Lovejoy and Schertzer*, 1985; *Gupta and Waymire*, 1990; *Burlando and Rosso*, 1996; *Veneziano and Furcolo*, 2002; *Veneziano and Langousis*, 2010; *Langousis et al.*, 2013). A single *scaling exponent* H is used to characterize the extreme rainfall distribution at all scales over which the scale invariance property holds. As a

20   consequence, a consistent and efficient estimation of extreme precipitation characteristics is possible, even at non-sampled temporal scales, and a parsimonious formulation of IDF curves based on analytical results is available (e.g., *Menabde et al.*, 1999; *Burlando and Rosso*, 1996; *De Michele et al.*, 2001; *Ceresetti*, 2011).

Theoretical and physical evidence of the scaling properties of precipitation intensity over a wide range of durations has been provided by several studies. MS has been demonstrated to be appropriate for modeling the temporal scaling features of the

25   precipitation process (i.e., not only the extreme distribution) and for the extremes in event-based representations of rainfall (stochastic rainfall modeling) (e.g., *Veneziano and Furcolo*, 2002; *Veneziano and Iacobellis*, 2002; *Langousis et al.*, 2013, and references therein). These multifractal features of precipitation last within a finite range of temporal scales (approximatively between 1 hour and 1 week) and concern the temporal dependence structure of the process. They have been connected to the large fluctuations of the atmospheric and climate system governing precipitation which are likely to produce a "cascade of

30   random multiplicative effects" (*Gupta and Waymire*, 1990).

At the same time, many studies confirmed the validity of SS for approximating the precipitation distribution tails in IDF estimation (for examples of durations ranging from 5 min to 24 h see *Menabde et al.*, 1999; *Veneziano and Furcolo*, 2002; *Yu et al.*, 2004; *Nhat et al.*, 2007; *Bara et al.*, 2009; *Ceresetti et al.*, 2010; *Panthou et al.*, 2014). This type of scaling is substantially

different from the temporal scaling since it only refers to the power law shape of the marginal distribution of extreme rainfall. Application of the SS models to precipitation records showed that the scaling exponent estimates may depend on the considered range of durations (e.g., *Borga et al.*, 2005; *Nhat et al.*, 2007) and the climatological and geographical features of the study regions (e.g., *Menabde et al.*, 1999; *Bara et al.*, 2009; *Borga et al.*, 2005; *Ceresetti et al.*, 2010). However, the application

5 of the SS framework has been mainly restricted to specific regions and small observational datasets. A deeper analysis of the effects of geoclimatic factors on the SS approximation validity and on estimated scaling exponent is thus needed.

The present study aims to deepen the knowledge of the scale-invariant properties of extreme rainfall intensity by analyzing SS model estimates across North America using a large number of station series. The specific objectives of this study are: a) asses the ability of SS models to reproduce extreme precipitation distribution; b) explore the variability of scaling exponent estimates

10 over a broad set of temporal durations and identify possible effects of the dominant climate and pluviometric regimes on SS; c) evaluate the possible advantages of the introduction of the SS hypothesis in parametric models of extreme precipitation.

The article is structured as follows. In Sect. 2 the statistical basis of scaling models is presented, while data and their preliminary treatments are described in Sect. 3. Sect 4 presents the distribution-free estimation of SS models and their validation using available series. Section 5 focusses on to the spatial variability of SS exponents and discusses the scaling exponent variation

15 from a regional perspective. Finally, the SS IDF estimation based on the Generalized Extreme Value (GEV) assumption is discussed in Sect. 6, followed by a discussion and conclusions [Sect. 7].

2 Simple Scaling models for precipitation intensity

When the equality in Eq. (1) holds for the cumulative distribution function (cdf) of the precipitation intensity X, considered at two different durations d and $D = \lambda d$, Simple Scaling can be expressed as (*Gupta and Waymire*, 1990; *Menabde et al.*, 1999):

$$X_D \overset{d}{=} \lambda^{-H} X_d, \tag{2}$$

where $H \in \mathbb{R}$ and $\overset{d}{=}$ means that the same probability distribution applies for X_d and X_D, up to a dilatation or contraction of size $\lambda^{-H} = (D/d)^{-H}$. An important consequence of the SS assumption is that, if X_D has finite moments $E[X_D^q]$ of order q, then X_D^q and $\lambda^{-H} X_d^q$ have the same distribution. Their moments are thus linked by the following relationship (*Gupta and Waymire*, 1990; *Menabde et al.*, 1999):

$$E[X_D^q] = \lambda^{-Hq} E[X_d^q]. \tag{3}$$

25 This last relationship is usually referred to as the *wide sense* simple scaling property (*Gupta and Waymire*, 1990) and signifies that simple scaling results in a simple translation of the log-moments between scales:

$$\ln\{E[X_D^q]\} = \ln\{E[X_d^q]\} - Hq\ln\lambda \tag{4}$$

Therefore, the SS model can be estimated and validated over a set of durations $d_1 < d_2 < .. < d_D$ by simply checking the linearity of the X moments versus the scale ratio λ in a log-log plot (*Gupta and Waymire*, 1990; *Burlando and Rosso*, 1996).

If H estimated for the first moment equals the exponents (slopes) for the other moments, the precipitation intensity X can be considered scale invariant under SS in the range of durations d_1 to d_D.

More sophisticated methods have also been proposed for detecting and estimating scale invariance [for instance, dimensional analysis, *Lovejoy and Schertzer* (1985); *Tessier et al.* (1993); *Bendjoudi et al.* (1997); *Dubrulle et al.* (1997); spectral analysis and wavelet estimation *Olsson et al.* (1999); *Venugopal et al.* (2006) *Ceresetti* (2011); and empirical probability distribution function (pdf) power law detection *Hubert and Bendjoudi* (1996); *Sivakumar* (2000); *Ceresetti et al.* (2010)]. However, estimation through the moment scaling analysis is by far the simplest and most intuitive tool to check the SS hypothesis for a large dataset. For this reason, the presented analyses are based on this method.

According to the literature, the values of the scaling exponents H generally range between 0.4 and 0.8 for precipitation intensity considered at daily and shorter time scales (e.g., *Burlando and Rosso*, 1996; *Menabde et al.*, 1999; *Veneziano and Furcolo*, 2002; *Bara et al.*, 2009) (note that $H_{intensity} = 1 - H_{depth}$). Values from 0.3 to 0.9 have also been reported for some specific cases (e.g., *Yu et al.*, 2004; *Panthou et al.*, 2014, for scaling intervals defined within 1 h and 24 h).

Higher H values have been generally observed for shorter-duration intervals, and regions dominated by convective precipitation (e.g., *Borga et al.*, 2005; *Nhat et al.*, 2007; *Ceresetti et al.*, 2010; *Panthou et al.*, 2014, and references therein). Nonetheless, some studies performing spatio-temporal scaling analysis reached a different conclusion. For instance, *Eggert et al.* (2015), analyzing extreme precipitation events from radar data for durations between 5 min and 6 h and spatial scales between 1 km and 50 km, indirectly showed that stratiform precipitation intensity generally displays higher temporal scaling exponents than convective intensity. For short-duration intervals (typically less than one hour), previous studies have also reported more spatially homogeneous H estimates than for long-duration intervals (e.g., *Alila*, 2000; *Borga et al.*, 2005, and references therein). This suggests that processes involved in the generation of local precipitation are comparable across different regions.

More generally, higher H values are associated with larger variations in moment values as the scale is changed (i.e. a stronger scaling), while H close to zero means that the X_d distributions for different durations d more closely match each other.

2.1 Simple Scaling GEV models

Annual Maximum Series (AMS) are widely used to select rainfall extremes from available precipitation series. Various theoretical arguments and experimental evidences support their use for extreme precipitation inference (e.g., *Coles et al.*, 1999; *Katz et al.*, 2002; *Koutsoyiannis*, 2004a; *Papalexiou et al.*, 2013).

Based on the asymptotic results of the Extreme Value Theory (*Coles*, 2001), the AMS distribution of a random variable X is well described by the Generalized Extreme Value (GEV) distribution family. If we represent the AMS by $(x_1, x_2, ..., x_n)$, the GEV cdf can be written as (*Coles*, 2001):

$$F(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{5}$$

where $\xi \neq 0$, $-\infty < x \leq \mu + \sigma/\xi$ if $\xi < 0$ (bounded tail), and $1/\mu + \sigma\xi \leq x < +\infty$ if $\xi > 0$ (heavy tail). $\mu \in \mathbb{R}$, $\sigma > 0$ and ξ respectively represent the location, scale, and shape parameters of the distribution. The shape parameter describes the charac-

teristics of the distribution tails and the frequency of the extremes generated by $F(x)$. Thus, high order quantile estimation is particularly affected by the value of ξ. If $\xi = 0$ (light-tailed shape, Gumbel distribution), Eq. (5) reduces to:

$$F(x) = \exp\left\{-\exp-\left\{\frac{x-\mu}{\sigma}\right\}\right\} \tag{6}$$

where $-\infty < x < +\infty$.

5   In applications, the GEV distribution is frequently constrained by the assumption that $\xi = 0$ (i.e., to the Gumbel distribution), due to the difficulty of estimating significant values of the shape parameter when the recorded series are short (e.g., *Borga et al.*, 2005; *Overeem et al.*, 2008; *CSA*, 2012). However, based on theoretical and empirical evidence, many authors have shown that this assumption is too restrictive for extreme precipitation, and may lead to important underestimations of the extreme quantiles (e.g., *Koutsoyiannis*, 2004a, b; *Overeem et al.*, 2008; *Papalexiou et al.*, 2013). Instead, approaches aimed at increasing series

10 length may be used to improve the estimation of the GEV distribution shape parameter (for instance, the Regional Frequency Analysis (RFA), *Hosking and Wallis*, 1997). Among these approaches, SS models constitute an appealing way to pool data from different samples (durations) and reduce uncertainties in GEV parameters.

For the GEV distribution it is straightforward to verify that, if $X \overset{d}{=} GEV(\mu, \sigma, \xi)$ then $\lambda X \overset{d}{=} GEV(\lambda\mu, \lambda\sigma, \xi)$ for any $\lambda \in \mathbb{R}$. This means that the GEV family described by Eq. (5) and (6) satisfies Eq. (2) and thus complies with scale invariance for any

15 constant multiplicative transformation of X. Under this assumption the wide sense SS definition [Eq. (3)] gives:

$$\mu_d = d^{-H}\mu_* \ , \sigma_d = d^{-H}\sigma_* \ , \ \text{and} \ \xi_d = \xi_* \tag{7}$$

where μ_*, σ_*, and ξ_* represent the GEV parameters for a reference duration d^* chosen, for simplicity, as $d^* = 1$, so that $\lambda = d$.

2.2   SS GEV estimation

Taking advantage of the scale invariant formulation of the GEV distribution, many authors have proposed simple scaling IDF and DDF models for extreme precipitation series (e.g., *Yu et al.*, 2004; *Borga et al.*, 2005; *Bougadis and Adamowski*, 2006;

20 *Bara et al.*, 2009; *Ceresetti*, 2011). In these cases, the scaling exponent and the GEV parameters are generally estimated in two separate steps: first, the H value is empirically determined through a log-log linear regression, as described above; then, GEV parameters μ_*, σ_*, and ξ_* for the reference duration d_* are estimated on the pooled sample of all available durations. Classical estimation procedures, such as GEV Maximum-Likelihood (ML) (*Coles*, 2001) or Probability Weighted Moment (PWM) (*Greenwood et al.*, 1979; *Hosking et al.*, 1985), can be used.

25 ## 3   Data and study region

Four station datasets were used for the construction of intensity AMS at different durations: the Daily Maxima (DM) and the Hourly (H) datasets provided by Environment Canada (*ECCC*) and the *MDDELCC* [in french Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques] for Canada, and the Hourly Precipitation

Data (HPD) and 15-Min Precipitation Data (15PD) datasets made available by the National Oceanic and Atmospheric Admin-istration (*NOAA*) agency [http://www.ncdc.noaa.gov/data-access/land-based-station-data] for United States. The total number of stations was approximately 3400, with roughly 2200 locations having both DM and H series, or both HPD and 15PD series. The majority of stations are located in the United States and in the southern and most densely populated areas of Canada. In

5   northern regions the station network is sparse and the record length does not generally exceed 15 or 20 years. Moreover, for most of DM and H stations, the annual recording period does not cover the winter season and available series generally include precipitation measured from May to October. For this reason, the year was defined as the period from June to September for the stations located north of the 52nd Parallel [122 days a year were used], while the period from May to October was used for remaining stations [184 days a year].

10   Data were collected through a variety of instruments [e.g., standard, tipping-bucket, and Fischer-Porter rain gauges] and pre-cipitation values were processed and checked using both automated and manual methods (*CSA*, 2012, *HPD and 15PD online documentation*). Most often, observations were recorded by tipping-bucket gauges with tip resolution from 0.1 mm to 2.54 mm (*CSA*, 2012; *Devine and Mekis*, 2008). 15 min series usually present the coarser resolution, with a minimum non-zero value of 2.54 mm, observed for about 80.5% of 15PD stations. The effects of such a coarse resolution on simple scaling estimates

15   could be important leading to empirical X_d cdfs becoming step-wise functions with a low number of steps. Some preliminary analyses aiming at evaluating these effects on SS estimates are presented in the supplementary material [see Fig. S2 and S3]. However, the 15PD dataset is important considering the associated network density and its fine temporal resolution, and thus it has been retained for our study. The main characteristics of the available datasets are summarized in Table 1.

The scaling AMS datasets were constructed according to the following steps:

20   (i) Three duration sets were defined: a) 15 min to 6 h with a 15min step; b) 1 h to 24 h with a 1h step; c) 6 h to 168 h (7 days) with a 6h step. These duration sets are hereinafter referred to as Short-Duration (SD), Intermediate-Duration (ID), and Long-Duration (LD) datasets, respectively.

(ii) Meteorological stations that were included in each final dataset were selected according to the following criteria: 1) pre-cipitation series must have at least 85% of valid observations each year, otherwise the year was considered as missing; 2) each

25   station must have at least 15 valid years; 3) for each station, it was possible to compute AMS for all durations considered in the scaling dataset (e.g., H and HPD stations were not included in the SD dataset because only hourly durations were available).

(iii) A moving window was applied to 15PD, H, and HPD series to estimate aggregated series at each duration. For DM series, a quality check was also implemented in order to guarantee that precipitation intensities recorded each day at different durations were consistent with each other (for instance, each pair of DM intensity (x_{d_1}, x_{d_2}) observed at durations $d_1 < d_2$ must respect

30   the condition $d_1/d_2 \leq x_{d_2}/x_{d_1} \leq 1$, otherwise the day was considered as missing).

(iv) For each selected station, annual maxima were extracted for each valid year and duration. Then, for stations having both DM and H series, or 15PD and HPD series, over a common time period, annual maximum values between these two series were retained.

Major characteristics of each scaling AMS dataset are reported in Table 2.

4 SS estimation through Moment Scaling Analysis (MSA)

Moment Scaling Analysis (MSA) for the SD, ID, and LD datasets was carried out to empirically validate the use of SS models for intensity AMS. Assessing the validity of the SS hypothesis for various duration intervals also aimed at determining the presence of different scaling regimes for precipitation intensity distributions. The spatial distribution of the scaling exponents was then analyzed to assess the dependence of SS on local geoclimatic characteristics.

In order to identify possible changes in the SS properties of AMS distributions, various scaling intervals were defined for the MSA. All possible subsets with 6, 12, 18 and 24 contiguous durations were considered for each dataset. Figure 1 and Figure 2 show the 136 scaling intervals thereby defined: 40 scaling intervals for SD and IS, and 56 for LD. For instance, the first matrix on the left of Fig. 1(a) presents the 6-duration scaling intervals 15min - 1.5h, 30min - 1.75h, ..., 4.75h - 6h from the SD dataset (with 15min step). This procedure was defined in order to evaluate the variability of the SS estimates when changing the position of the scaling interval [hereinafter identified by its first duration d_1; see examples in Figure 1(d)] and the number of durations considered.

For each scaling interval (for simplicity, their index has been omitted), the validity of the SS hypothesis was verified according to the following steps:

1. *MSA regression:* for $q = 0.2, 0.4, \ldots, 2.8, 3$, the slopes K_q of the log-log linear relationships between the empirical $q-$moments $\langle X_d^q \rangle$ of $X_{d_1}, X_{d_2}, \ldots, X_{d_D}$ and the corresponding durations $d_1, d_2, \ldots, d_D$ in the scaling interval were estimated by Ordinary Least Squares (OLS). Order $q \geq 3$ were not considered because of the possible biases affecting empirical high order moment estimates.

2. *Slope test:* Regressing K_q on q, the hypothesis that the estimated K_q-exponents vary linearly with the order of moment q was verified. To this end, a t-Student test was used to test the null hypothesis $\boldsymbol{H}_0$: $\beta_1 = K_1$, where β_1 is the slope coefficient of the simple regression model $K_q = \beta_0 + \beta_1 q$. If $\boldsymbol{H}_0$ was not rejected at the significance level $\alpha = 0.05$, the SS assumption for the scaling interval was considered appropriate and the simple scaling exponent $H = K_1$ was retained.

3. *Goodness-of-Fit (GOF) test:* for each duration d, the goodness of fit of the X_d distribution under SS was tested using the Anderson-Darling (AD) and the Kolmogorov-Smirnov (KS) tests. These tests aim at validating the appropriateness of the scale invariance property for approximating the X_d cdf by the distribution of $\lambda^{-H} X_{d^*}$. To this end, the pooled sample

$$\boldsymbol{x}_{d^*} = \left(\boldsymbol{x'}_{d_1}, \ldots, \boldsymbol{x'}_{d_j}, \ldots, \boldsymbol{x'}_{d_D} \right) \tag{8}$$

of the rescaled AMS, $\boldsymbol{x'}_{d_j}$, for all durations d_j, $j = 1, \ldots, D$, in the scaling interval, was used to define X_{d^*} under the SS assumption. Using $d^* = 1$ h, the rescaled sample $\boldsymbol{x'}_{d_j}$ of the annual maxima $x_{d_j, i}$, $i = 1, \ldots, n$, observed for d_j was:

$$\boldsymbol{x'}_{d_j} = \left(d_j^H x_{d_j, 1}, d_j^H x_{d_j, 2}, \ldots, d_j^H x_{d_j, i}, \ldots d_j^H x_{d_j, n} \right) \tag{9}$$

where n represents the number of observations (years) available for each duration. Hence, $n \times D$ rescaled observations were included in $\boldsymbol{x}_{d^*}$.

The tests were then applied at significance level $\alpha = 0.05$ to compare $\boldsymbol{x}_d$ and $\boldsymbol{x}_{d*}$ empirical distributions (Cunnane plotting formula, *Cunnane*, 1973) as in previous precipitation scaling applications (e.g., *Panthou et al.*, 2014). In fact, despite the low power of KS and AD tests for small sample tests, they represent the only suitable solution to the problem of comparing empirical cdfs when the data do not follow a normal distribution. Because both AD and KS are affected by the presence of ties in the samples (e.g., repeated values due to rounding or instrument resolution), a permutation test approach (*Good*, 2013) was used to estimate test p-values. According to this approach, pooled data of $\boldsymbol{x}_d$ and $\boldsymbol{x}_{d*}$ were randomly reassigned to the SS and non-SS samples. Then, the test statistic distribution under the null hypothesis of equality of the $d^{-H} X_{d*}$ and X_d distributions was approximated by computing its value over a large set of random samples. Finally, the test p-value was obtained as the proportion of random samples presenting a test statistic value larger than the value observed for the original sample.

The mean error resulting from approximating the X_d distribution by the SS model was then evaluated in a cross-validations setting. For this analysis, each duration was iteratively excluded from each scaling interval and the scaling model re-estimated at each station. Predictive ability indices, such as the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) between empirical and SS distribution quantiles, were estimated for highest quantiles. In particular, to focus on return periods of practical interest for IDF estimation, only quantiles larger than the median were considered (i.e., only return periods greater than 2 years).

The average over all stations of the normalized RMSE, $\overline{r}_{x_d}$, for each scaling interval and duration was used:

$$\overline{\overline{r}}_{x_d} = \frac{1}{n_s} \sum_{s=1}^{n_S} \overline{r}_{x_{d,s}} \tag{10}$$

where n_s is the number of valid SS stations in the dataset and

$$\overline{r}_{x_{d,s}} = \frac{r_{x_{d,s}}}{\overline{x}_{d,s}} \tag{11}$$

$r_{x_{d,s}}$ and $\overline{x}_{d,s}$ are, respectively, the RMSE and the mean value of all X_d quantiles of order $p > 0.5$ at station s.

4.1 Model estimation and validation

Figure 1 presents the results of points 1 to 3 of the procedure for the evaluation of the SS validity. It shows, for each scaling interval and duration, the proportion of valid SS stations [Fig. 1(a)-(c)]. As showed in the example in Fig. 1(e), for each scaling interval, valid SS stations were defined as stations having not rejected both the Slope test for the scaling interval and the GOF tests for each duration included in this scaling interval.

As expected, the proportion of valid SS stations decreased when the number of durations within the scaling interval increased and with decreasing d_1. This is particularly evident for short d in SD and ID datasets. More GOF test rejections were observed for longer scaling intervals [not shown], due to the higher probability of observing large differences between $\boldsymbol{x}_d$ and $\boldsymbol{x}_{d*}$ quantiles when $\boldsymbol{x}_{d*}$ was larger and included data from more distant durations. However, several factors can impact GOF test results when shorter d_1 are included in the scaling intervals. First, the SS hypothesis could be rejected due to the presence of very large values in short-duration samples, to which GOF tests are particularly sensitive. Second, when considering durations close to the

temporal resolution of the recorded series, stronger underestimations could affect the measure of precipitation because intense rainfall events are more likely to be split between two consecutive time steps. Finally, preliminary analyses [Fig. S2 and S3 in the supplementary material] showed that the largest GOF test rejections could also be connected to the coarse tip resolution of 15PD series, which, similar to the temporal resolution effect, induces larger measurement errors in the shortest duration

5 precipitation series. Note that comparable resolution issues were previously reported by some authors while estimating fractal and intermittency properties of rainfall processes (e.g., *Veneziano and Iacobellis*, 2002; *Mascaro et al.*, 2013).

Valid SS station proportions between 0.99 and 1 were always observed for GOF tests in ID and LD datasets, except for some durations shorter than 3 h (ID dataset) or 6 h (LD dataset). For all three datasets, no particular pattern was observed for slope test results [not shown], with at most 2% of the stations within each scaling interval displaying a non linear evolution of the

10 scaling exponent with the moment order.

When considering both GOF and Slope test, with the exception of some durations ≤1 hour, the proportion of stations satisfying SS was higher than 0.9, and the majority of scaling intervals [65%, 90%, and 98% of the scaling intervals in SD, ID, and LD, respectively] included at least 95% of valid SS stations. For each scaling interval, only valid SS stations were considered in the rest of the analysis.

15 Figure 2 presents, for each scaling interval and duration, the station average, $\overline{r}_{x_d}$, of the normalized RMSE. These graphics show that mean relative errors on intensity quantiles did not generally exceed 5% of the precipitation estimates for 6-duration scaling intervals [Fig. 2, first col.]. As for the valid SS station proportion, the performances of the model deteriorated with decreasing d_1 and with increasing scaling interval length, especially for durations at the border of the scaling intervals. However, for more that 70% of 12-, 18-, and 24-duration scaling intervals, $\overline{\overline{r}}_{x_d} \leq 0.1$ for each duration included in the scaling interval.

20 $\overline{\overline{r}}_{x_d} \geq 0.25$ were observed for 15 min in 12-duration or longer scaling intervals, pointing out the weaknesses of the model in approximating short duration extremes when the scaling interval included durations ≥ 3 h.

4.2 Estimated scaling exponents

The median, Interquantile Range (IR), and quantiles of order 0.1 and 0.9 of the H distribution across stations, are presented in Fig. 3(i) for each 6-duration scaling interval. Figures 3(ii) - (iv) show the distribution of the scaling exponent variation $\Delta_{H_{(j)}}$

25 observed over stations when each scaling interval is lengthened from 6 to 12, 18, and 24 durations. For each station and each d_1, $\Delta_{H_{(j)}}$ was defined as:

$$\Delta_{H_{(j)}} = H_{(j)} - H_{(6)} \tag{12}$$

where $j = 12, 18$, or 24 represent the number of durations considered in the specified scaling interval, $H_{(j)}$ is the corresponding scaling exponent, and $H_{(6)}$ is the scaling exponent estimated for the corresponding 6-duration scaling interval (i.e., the

30 6-duration interval having the same d_1).

Figures 3(ii)-(iv) represent the changes observed in H values when the scaling interval length and d_1 increased. Median $\Delta_{H_{(j)}}$, as well as its IR, increased with the number of durations added to the scaling interval for all d_1. For the 24-duration scaling interval "1h - 24h" (ID dataset), for instance, median $\Delta_{H_{(24)}} = 0.047$ was observed. For the interval "15min - 6h" (SD dataset),

$\Delta_{H_{(24)}}$ was even larger, with a median scaling exponent variation approximately equal to 0.087 and with 25% of stations having $\Delta_{H_{(24)}} \geq 0.11$. These results indicate that, for some stations, a dramatic difference could exist in IDF estimations obtained with the different definitions of the scaling interval. Changes in H values were also important when comparing 6- and 12-duration scaling intervals when $d_1 \leq 1$ h (SD and ID datasets) and in LD dataset [Fig. 3 (ii)].

5 Nonetheless the median scaling exponent variation was generally smaller than 0.05, except for a relatively small proportion of stations. Equally important, $|\Delta_{H_{(j)}}|$ was generally centered on 0 and for all $d_1 \geq 1$ h more than 50% of stations had $|\Delta_{H_{(12)}}| \leq$ 0.025 (SD dataset) and $|\Delta_{H_{(18)}}| \leq 0.03$ (ID dataset) [Fig. 3 (ii)-(iii)].

The smallest median H values were observed for the shortest d_1 ($d_1 \leq 30$ min) in Fig. 3 (a-i), and for the longest d_1s in Fig. 3 (c-i). Scaling intervals beginning at 15 and 30 min also displayed the smallest variability across stations. Although fewer

10 stations were available for these intervals (only 15PD stations were used and the number of valid SS stations was smaller), this result is consistent with previous reports in the literature demonstrating that H values are spatially more homogeneous for short durations.

A larger dispersion of H values was observed when d_1 ranged between approximately 1 h and 5 h, in particular in the SD dataset, for which the 10th-90th percentile difference almost covered the entire range of observed H values [Fig. 3 (i)]. This

15 result could be in part explained by the fact that, if the scaling interval length is fixed, then the variance $V[\ln(d)]$ of the MSA regression covariate decreases as d_1 increases. In fact, the use of a logarithmic scale for the MSA regression implies that the mean distance between durations in the scaling interval decreases as d_1 increases Thus, regression errors of the same magnitude in short and long d_1 scaling intervals differently affect the OLS variance of H, especially when scaling intervals are short. This may result in larger uncertainty of H for longer d_1 scaling intervals of SD. Moreover, as showed in next sections, H variability

20 across stations may be effectively larger due to the greater spatial variability of the scaling exponent for d_1 longer than a few hours.

Largest median H were observed for d_1 greater than 10 hours [Fig. 3 (b-i)] and lower than 2 days [Fig. 3 (c-i)], with approximately half of the stations having $H \geq 0.8$. This means that a stronger scaling (i.e., larger H values) is needed to relate extreme precipitation distributions at approximately 12-hours to distributions at daily and longer scales. It may therefore be expected

25 that the stations characterized by H closer to 1 are located in geographical areas where differences in precipitation distributions are important among temporal scales included in these scaling intervals.

Examples of the spatial distributions of the scaling exponent are given in Fig. 4 and 5 for the first and last d_1 for each interval length and dataset, respectively. Since only one 24-duration scaling interval was defined for both the SD and ID datasets, only scaling intervals containing 6, 12, and 24 (Fig. 4) or 18 (Fig. 5) durations are presented. This avoids the redundancy of showing

30 twice the "15min - 6h" (SD dataset) and "1h - 24h" (ID dataset) scaling intervals.

Generally, the scaling exponent displayed a strong spatial coherence and varied smoothly in space, although a more scattered distribution of H characterizes maps in Fig. 5. In this last figure, the local variability of H may be attributed to the larger estimation uncertainties, as previously mentioned. Meaningful spatial variability and clear spatial patterns emerged for $d_1 \geq 1$ h. In fact, for stations located in the interior and southern areas of the continent, a shift from weaker scaling regimes (smaller H)

35 to higher H values was observed as d_1 increases [e.g., second and third rows of Fig. 4]. On the contrary, a smoother evolution

of H over the scaling intervals characterized the northern coastal areas, especially in north-western regions, and the Rockies, where $H > 0.75$ values were rarely observed even for greater d_1 values.

5 Regional analysis

Regional differences in scaling exponents were investigated. Only the results for the 6-duration scaling intervals are presented, similar results having been obtained for longer scaling intervals [see the supplementary material, Fig. S5 and S6 for 12- and 18-duration scaling intervals].

Stations were pooled into six climatic regions based on a previous classification suggested by *Bukovsky* (2012) [see Fig. 6]. Stations outside the domain covered by the Bukovsky regions were attributed to the nearest region. Regions with less than 10 stations were not considered (regions without colored borders in Fig. 6) and region A1 (W_Tun) was kept separated from region A2 (NW_Pac) because only 14 stations were available in region A1 (W_Tun) for ID and LD datasets.

To provide deeper insights about regional features of precipitation associated with specific scaling regimes two variables related to the precipitation events observed within AMS were also analyzed: the mean number of events per year, $\bar{N}_{eve}$, and the mean wet time per event, $\bar{T}_{wet}$, contributing to AMS within each scaling interval. For a given year and station, annual maxima associated to different durations of a given scaling interval were considered to belong to the same precipitation event if the time intervals over which they occurred overlapped [see Fig. 7 (g); in this example 3, 4, and 5 h annual maxima are associated with the first event while 1, 2, and 6 h annual maxima are associated to the second event]. The mean number of events at each station was then computed:

$$\bar{N}_{eve} = \frac{1}{n}\sum_{i}^{n} N_{eve,i} \tag{13}$$

with $N_{eve,i}$ the number of non-overlapping time intervals, i.e. the number of different events contributing to AMS during the ith year of record. The distribution of $\bar{N}_{eve}$ values within each region is presented in Fig. 7.

The mean wet time per event contributing to AMS, $\bar{T}_{wet}$, was defined as the mean number of hours with non-zero precipitation within each event. Details on the calculation of $\bar{T}_{wet}$ and corresponding results are presented in the supplementary material [Sect. S1 and Fig. S4].

5.1 Regional variation of the scaling exponents.

Figure 8 shows the distribution of H within each region. Three types of curves can be identified. First, curves in Fig. 8 (a) to (c) have a characteristic smooth S shape. Conversely, Fig. 8 (d) displays a rapid increase of H for scaling intervals defined in ID and LD datasets until $d_1 = 2$ days, preceded and followed by two plateaus, one for the longest d_1 with remarkably high H values, and one for the shortest d_1 with small H values. Finally, an inverse-U-shaped curve can be seen in Fig. 8 (e) and (f), with globally high H values already reached at sub-daily durations in dry regions (E).

The difference between Fig. 8 (a) and (e)-(f) can be partially explained by the weaker impact of convection processes in generating very short duration extremes in regions A1 and A2 with respect to southern areas (regions E and F). For northern

regions, the transition between short and long duration precipitation regimes may be smoothed by cold temperatures which moderate short-duration convective activity, especially for W_Tun (region A1). The topography characterizing the northern pacific coast may explain the smoothing effect for the curve of region NW_Pac (A2): precipitation rates at daily and longer scales are enhanced by the orographic effect acting on synoptic weather systems coming from the Pacific Ocean (*Wallis et al.*,

5 2007).

Similarly, mountainous regions in C [Fig. 8 (c)] displayed the smallest variations of H over d_1, indicating that analogous scaling regimes characterize both short- and long-duration scaling intervals. Again, this may be related to the important orographic effects of precipitation in these regions that are involved in the generation of extremes for both sub-daily and multi-daily time scales.

10 The mean number of events per year in regions A and C was higher than in regions E-F, in particular for SD scaling intervals, and displayed steeper decreases with increasing d_1 [Fig. 7 (a) and (c)].

Main differences between regions B and A were the stronger scaling regimes observed in B, which were mainly due to contributions from stations located in the south-eastern part of the E_Bor region (not shown). For scaling intervals in the ID dataset, region B was also characterized by the highest mean number of events per year, with most of the stations presenting $\bar{N}_{eve} > 2$

15 for $d_1 = 1$ h and $d_1 = 2$ h and sharp decreases of $\bar{N}_{eve}$ with increasing d_1 [Fig. 7 (b)]. Moreover, a remarkably large range of $\bar{N}_{eve}$ was observed for $1\,\text{h} \leq d_1 \leq 6\,\text{h}$, suggesting that B may be highly heterogeneous.

Two distinct scaling regimes can be observed for SW_Pac (region D) at, respectively, $d_1 \leq 3$ h (SD dataset) and $d_1 \geq 2$ days (ID dataset) [region D in Fig. 8 (d)]. These plateaus may be interpreted by recalling that $1 - H_{intensity} = H_{depth}$. On the one hand, the low and constant H observed for $d_1 \leq 3$ h indicates that the average precipitation depth increases with duration at

20 the same growth rate for all these intervals. On the other hand, H approximately equal to 0.9 at daily and longer durations demonstrates that the average precipitation depth associated with long-duration annual maxima remained roughly unchanged when the duration increased from 1.5 to 7 days ($\lambda^{H_{depth}} \approx 1$ in Eq. (3)). This, along with the fact that the scaling exponent increased almost monotonically for $1\,\text{h} \leq d_1 \leq 24\,\text{h}$ (ID and LD datasets), suggests that extremes at durations shorter than ~ 3 h (SD dataset) drive annual maxima precipitation rates at longer scales, with the rapid and continuous decay in mean intensity

25 caused by the increasing size of the temporal scale of observation.

For SW_Pac (region D), the relative absence of long-lasting weather systems able to produce important extremes for long durations, was confirmed by the analysis of $\bar{N}_{eve}$ and $\bar{T}_{wet}$ [for results on $\bar{T}_{wet}$ see Fig. S4 of the supplementary material]. In fact, the mean number of events per year was relatively high for short durations (the median $\bar{N}_{eve}$ is equal to 1.82 for $d_1 = 15$ min and to 1.4 for $d_1 = 1$h), while it rapidly decreased below 1.1 events per year for $d_1 \geq 6$ h (ID dataset) and for $d_1 \geq 18$ h

30 (LD dataset). With the exception of $d_1 = 6$ h (LD dataset), at least 90% of SW_Pac stations had $\bar{N}_{eve} \leq 1.25$ for all $d_1 > 3$ h. In other regions, median $\bar{N}_{eve}$ were never smaller than 1.1 for the SD and ID datasets, except for $d_1 \geq 12$h in region E.

Median H values displayed inverse-U shapes for the remaining regions with very small IR, despite the high number of valid SS stations: a slow transition from lower to higher H is observed approximately between 1 h and 12 h (region E) or 30 h (region F). The strongest scaling regimes were observed for $1\,\text{h} \leq d_1 \leq 2$ days in arid western regions [Fig. 8 (e)], while median H values

35 greater than 0.8 were only observed for approximately $6\,\text{h} \leq d_1 \leq 2$ days in more humid areas [8 (f)]. In both region E and F,

very short-duration extremes are typically driven by convective processes, while a transition to different precipitation regimes may be expected between 1 h and a few hours. However, the smoother increase of H visible in Fig. 8 (f) with respect to (e) may also indicate that, in eastern areas, the occurrence of sub-daily duration extremes are more likely associated to embedded convective and stratiform systems, or to mesoscale convective systems less active in western dry areas (*Kunkel et al.*, 2012). On

5 the contrary, for south-western dry regions [Fig. 8 (e)], where less intense summer extremes are expected compared to eastern areas [see supplementary material, Fig. S1], differences between short- and long-duration extreme precipitation intensity seem stronger since H tended to scatter in a range of higher values: precipitation intensity moments strongly decrease as the duration increases for approximately $1 \text{ h} \leq d_1 \leq 12 \text{ h}$.

In summary, these results suggest a regional effect on precipitation scaling of both local geographical characteristics, such as

10 topography or coastal effects, and general circulation patterns. Weak scaling regimes were observed for short d_1 and along the west coast of the continent and seem to be connected to scaling intervals and climatic areas characterized by homogeneous weather processes. Low H values correspond in fact to small variations in AMS distribution moments. On the contrary, stronger scaling regimes, which indicate important changes occurring in AMS moments across duration and, thus, in extreme precipitation features, were observed for longer d_1 in the other regions of the study area. According to these results, it would

15 be important to take into account the climatological information included in the scaling exponent to improve SS and IDF estimation. Even more important, these results could help for the definition of IDF relationships at non-sampled locations by the construction of spatial models for the IDF parameter H.

6 SS GEV etimation

Results presented in this section are limited to a descriptive analysis of GEV parameter estimates, and to an assessment of the

20 potential improvements carried out by SS GEV models with respect to PWM estimates of non-SS GEV models, for 6-duration scaling intervals. Similar results were generally obtained 12-, 18-, and 24-duration intervals [see supplementary material, Fig. S9 to S15].

In our application, GEV parameters μ_*, σ_*, and ξ_* [Eq. (7)] were estimated for x_{d^*} [Eq. (8)] using the PWM procedure. Preliminary comparisons among several estimation methods [e.g., PWM, classical ML estimators, and Generalized Additive

25 Model ML - see *Coles* (2001), *Katz* (2013) - in which the joint estimation of H, μ_*, σ_*, and ξ_* is obtained by the introduction of the duration as model covariate], showed that PWM slightly outperformed other methods.

Quantiles estimated from the SS and the non-SS GEV were compared with empirical quantiles. Global performance measures, such as RMSE, were computed to evaluate the overall fit of the estimated GEV to the empirical X_d distributions. In particular, mean errors between SS and non-SS quantile estimates and empirical quantiles were compared using the relative total RMSE

30 ratio, $R_{\overline{rmse}}$, defined as:

$$R_{\overline{rmse}} = \frac{[\overline{R}_{ss} - \overline{R}_{non-ss}]}{\overline{R}_{non-ss}} \tag{14}$$

where

$$\overline{R}_{mod} = \sum_{d=d_1}^{D} \frac{r_{d,mod}}{\bar{x}_d} \tag{15}$$

represents the normalized mean square difference between model and empirical quantiles of order $p > 0.5$ for all the durations
included in the scaling interval.

6.1 Estimated SS GEV parameters

Figure 9 presents the distributions over valid SS stations of the SS GEV parameters scaled at $d_* = 1$ h [Fig. 9 (a) and (b)]
and $d_* = 24$ h [Fig. 9 (c)]. For the SD dataset, even for scaling intervals which did not include the reference duration d_*, the
μ_* and σ_* distributions appeared to be similar to the non-SS μ_d and σ_d distributions [Figure 9, first row]. Conversely, in the
ID and LD datasets, both μ_* and σ_* distributions were more positively skewed than the corresponding non-SS distributions.
Moreover, the relative differences $\Delta_\mu = (\mu_* - \mu_d)/\mu_d$ and $\Delta_\sigma = (\sigma_* - \sigma_d)/\sigma_d$ were estimated for each station, duration, and
scaling interval. Two important results came out of this analysis [see Figures S10 and S11 of the supplementary material]. On
the one hand, median values of Δ_μ and Δ_σ were generally smaller than $\pm 5\%$ and $\pm 10\%$, respectively. On the other hand, Δ_σ
showed large positive values when $\xi_d = 0$ (i.e. Gumbel distributions), while small $\Delta_\sigma < 0$ were estimated when $\xi_d \neq 0$ [not
shown for conciseness]. These results are interesting since, while non-SS μ_d values are generally considered to be accurate
estimates of the X_d location parameter, small uncertainties are expected for the scale parameter only when the ξ_d value is
correctly assessed. In fact, the scale parameter σ_d may be strongly biased when the shape parameter is spuriously set to zero
($\xi_d = 0$).

In addition, μ_* and σ_* displayed strong coherence in their spatial distributions, which were characterized by an obvious North-
West to South-East gradient [Fig. 10 shows examples for the scaling intervals 15min - 1.5h, 1h - 6h, and 6h - 36h].

Notable differences between SS GEV and non-SS estimates were observed for the shape parameter [third column of Fig. 9].
Firstly, SS ξ_* were closer to 0 than non-SS ξ_d, for both positive and negative shape values. Secondly, ξ_* distributions were
generally more peaked around their median value than non-SS estimates.

Note that, the majority of stations had non-SS shape parameters ξ_d non-significantly different from zero according to asymp-
totic test proposed by Hosking et al. [1985] for PWM GEV estimators applied at level 0.05. In particular, for each duration,
non-SS models estimated light-tailed distributions (i.e., $\xi_d = 0$) for more than 85% of the stations, except that for $d = 15$ min
and $d = 30$ min [Fig. 11, first col.]. Conversely, for all scaling intervals with $d_1 > 15$ min, SS GEV shape parameters were
significantly different from zero for 40% to 45% of valid SS stations [Fig. 11, second col.]. Moreover, when using scaling
intervals of 12 durations or more, the proportion of $\xi_* > 0$ was always important (greater than 35%) for all 18- and 24-duration
scaling intervals [see the supplementary material, Fig. S9].

The previous results suggest that pooling data from several durations may effectively reduce the sampling effects impacting
the estimation of ξ, allowing more evidence of non-zero shape parameters, and, in many cases, of heavy tailed ($\xi > 0$) AMS
distributions. This conclusion is consistent with previous reports, namely that 100- to 150-year series are necessary to unam-
biguously assess the heavy-tailed character of precipitation distributions (e.g., *Koutsoyiannis*, 2004b; *Ceresetti et al.*, 2010). In

general, typical values of $\xi \approx 0.15$, close to the estimated ξ_* for cases in which $\xi_* > 0$, have also been reported (e.g., *Koutsoyiannis*, 2004b).

However, uncertainties on ξ_* estimates remain important. Support for this comes from the spatial distribution of ξ_*, which was still highly heterogeneous, with local variability dominating at small scales [e.g., Fig. 10, third col.].

6.2 Improvement with respect to Non-SS models

The proportion of series for which the SS model RMSE, $r_{d,ss}$, was smaller than the non-SS GEV RMSE, $r_{d,non-ss}$, was analyzed [see the supplementary material, Fig. S12]. For cases with non-zero ξ_*, the fraction of stations with $r_{d,ss} < r_{d,nonss}$ was higher than 60% for most of the scaling intervals and durations. On the contrary, $r_{d,ss} > r_{d,nonss}$ was observed for the majority of stations (generally more than 70%) with $\xi_* = 0$.

Figure 12 presents the $R_{\overline{rmse}}$ distribution over valid SS stations. When the SS distribution shape parameters were not significantly different from zero [Fig. 12, second col.], the relative increases in total RMSE were usually smaller than 0.1 in SD dataset, with only scaling intervals with $d_1 < 1$ h having greater $R_{\overline{rmse}}$. For the ID and LD datasets, the medians of the total relative RMSE ratio distributions were smaller than 0.05 for $d_1 \geq 4$ h and $d_1 \geq 24$ h, respectively. Furthermore, more than 90% of stations had $R_{\overline{rmse}} < 0.125$ for $d_1 \geq 6$ h (ID dataset) and $d_1 \geq 30$ h (LD dataset). When $\xi_* \neq 0$, an increase of the mean error in high order quantile estimates was observed for $d_1 = 15$ min (SD dataset) and $d_1 = 1$ h (ID dataset) for at least half of the stations [Fig. 12, first col.; note the different scale on the y-axis]. However, for all other d_1, negative $R_{\overline{rmse}}$ values were observed for the majority of stations for all scaling intervals, with a median reduction up to 30% of the mean error. Note that also for 12- and 18-duration scaling intervals the median $R_{\overline{rmse}}$ where generally negative for $d_1 > 1$ h and $\xi_* \neq 0$ [Fig. S13 and S14 of the supplementary material]. Conversely, $R_{\overline{rmse}}$ increased for the majority of stations in all 24-duration scaling intervals having $d_1 < 12$ h [Fig. S15 of the supplementary material]. Note also that no particular spatial pattern characterized the $R_{\overline{rmse}}$ estimates.

7 Discussion and conclusion

This study investigated simple scaling properties of extreme precipitation intensity across Canada and the United States. The ability of SS models to reproduce extreme precipitation intensity distributions over a wide range of sub-daily to weekly durations was evaluated. The final objective was to identify duration intervals and geographical areas for which the SS model can be used for the production of IDF curves.

The validity of SS models was empirically confirmed for the majority of the scaling intervals. In particular, the hypothesis of a scale-invariant shape of the X_d distribution held for all duration intervals spanning from 1 h to 7 days based on the comparison of SS distributions to empirical quantiles. Less convincing results were obtained for durations shorter than 1 h, especially for the longest scaling intervals (24-duration intervals). One possible explanation is that the coarse measurement resolution of the available 15 min series may strongly impact both the validation tools (for instance, GOF tests) and SS estimates. These results provide important operative indications concerning the inner and outer cut-off durations for AMS scaling.

The majority of the estimated scaling exponents ranged between 0.35 and 0.95, showing a smooth evolution over the scaling intervals and a well-defined spatial structure. Six geographical regions, initially defined according to a climatological classification of North America into 20 regions, displayed different features in terms of scaling exponent values. Specifically, distinct median values of H were observed for the various geographical regions, each characterized by a different precipitation regime.

5 This is consistent with results reported in the literature for some specific regions and smaller observational datasets (e.g., *Borga et al.*, 2005; *Nhat et al.*, 2007; *Ceresetti et al.*, 2010; *Panthou et al.*, 2014, and references therein). Moreover, while small and smooth changes of H over the scaling intervals were observed in regions containing the majority of stations, one region, SW_Pac, displayed two dramatically distinct scaling regimes separated by a steep transition occurring between a few hours and 24 h. These results limit the applicability of SS models in SW_Pac, and were connected to the local features of

10 intense precipitation events by the analysis of the mean number of events per year and the mean wet time of these events.

Weak scaling regimes, characterized by relatively small H values (H close to 0.5), were generally observed for scaling intervals containing very short durations (e.g, less than 2 h) and for regions on the west coast of the continent [regions A1, A2, and D; see Fig. 8]. For these scaling intervals and regions, we can expect that extreme precipitation events observed at various durations will have similar statistical characteristics, being governed by homogeneous weather processes.

15 The interpretation of high H values (e.g., $H > 0.8$), observed between 1 and several days, depending on the region, is more complex. These scaling regimes correspond to mean precipitation depth that varies little with duration. This suggests an important change in precipitation regimes occurring at some durations included in the scaling interval. One interesting example was region SW_Pac (region D) for scaling intervals of durations longer than 1 day . In this case, the analysis of the mean number of events per year sampled in AMS suggested that very few long-duration extreme events were produced by large-scale

20 dynamic precipitation systems.

For scaling intervals of durations longer than 4 days, scaling exponents seemed to converge to approximately 0.7 for all regions, except west coast regions (regions A1, A2, and D).

These results suggest that SS represents a reasonable working hypothesis for the development of more accurate IDF curves. Besides, the spatial distribution of the scaling exponent and its dependency on climatology should be taken into account when

25 defining SS duration intervals since the accuracy of the SS approximation may depend on the range of considered temporal scales. Equally critical, estimated H values were found to gradually evolve with the considered scaling intervals. In this respect, interesting extensions of the analysis should consider methods for the quantification of the uncertainty in H estimations as well as the possibility of modeling the scaling exponent as a function of both the observational duration and the AMS distribution quantile/moment order, i.e. by the use of a multiscaling (MS) framework for IDFs. Equally important, the events sampled by

30 the AMS also showed different statistical features within different geographical regions and some specific results [e.g., for the SW_Pac region] stimulate the interest for an analysis of the scaling property of extreme precipitation by the use of a temporal stochastic scaling approach.

The evaluation of SS model performances under the assumption of GEV distributions for AMS intensity was then performed. Results indicate that the proposed SS GEV models may lead to a more reliable statistical inference of extreme precipitation

35 intensity than that based on the conventional non-SS approach. In particular, a better assessment of the GEV shape parameter

seems possible when pooling data from several durations under the scaling hypothesis. The use of the SS approximation may introduce biases in high quantile estimates when AMS distributions move drastically away from perfect scale invariance (short durations and/or longest scaling intervals). Nonetheless, decreases in the SS GEV $RMSE$ with respect to non-SS GEV models for d_1 longer than a few hours and/or scaling intervals shorter than 24 durations indicate that quantile errors in IDF estimates

5   can be generally reduced.

Caution is advised when interpreting these results due to the fact that high order empirical quantiles were used as reference estimates of true X_d quantiles, which could be a misleading assumption especially when available AMS are short. Considering this limitation and our general results, any future extension of this study should investigate the possibility of introducing spatial information in scaling models as well as improvements of scaling GEV estimation procedures.

10   **8   Data availability**

The 15min Precipitation Data (15PD) and Hourly Precipitation Data (HPD) were freely obtained from NOAA/Climate Prediction Center (CPC) [http://www.ncdc.noaa.gov/data-access/land-based-station-data]. Houly (H) and Maximum Daily (DM) data for Canada were acquired from Environment and Climate Change Canada (ECCC) and from the MDDELCC of Québec [data available upon request by contacting *Info-Climat@mddelcc.gouv.qc.ca*].

15   *Acknowledgements.* Silvia Innocenti's scholarship was partly provided by the Consortium OURANOS. Financial support for this project was also provided by the Collaborative Research and Development Grants program from the Natural Sciences and Engineering Research Council of Canada (AM). They also thank Guillaume Talbot for preliminary data processing and Dikra Khedhaouiria for useful discussions.

References

Alila, Y. (2000), Regional rainfall depth-duration-frequency equations for Canada, *Water Resources Research*, *36*(7), 1767–1778, doi:10.1029/2000WR900046.

Asquith, W. H., and J. S. Famiglietti (2000), Precipitation areal-reduction factor estimation using an annual-maxima centered approach, *Journal of Hydrology*, *230*(1-2), 55–69, doi:10.1016/S0022-1694(00)00170-0.

Bara, M., S. Kohnová, L. Gaál, J. Szolgay, and K. Hlavcová (2009), Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia, *Contributions to Geophysics and Geodesy*, *39*(3), 187–206.

Bendjoudi, H., P. Hubert, D. Schertzer, and S. Lovejoy (1997), Interprétation multifractale des courbes intensité-durée-fréquence des précipitations, *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, *325*(5), 323–326.

Bernard, M. M. (1932), Formulas for rainfall intensities of long duration, *Transactions of the American Society of Civil Engineers*, *96*(1), 592–606.

Blöschl, G., and M. Sivapalan (1995), Scale issues in hydrological modelling: a review, *Hydrological Processes*, *9*(3-4), 251–290.

Borga, M., C. Vezzani, and G. Dalla Fontana (2005), Regional rainfall depth-duration-frequency equations for an alpine region, *Natural Hazards*, *36*(1-2), 221–235.

Bougadis, J., and K. Adamowski (2006), Scaling model of a rainfall intensity-duration-frequency relationship, *Hydrological Processes*, *20*(17), 3747–3757, doi:10.1002/hyp.6386.

Bukovsky, M. S. (2012), Masks for the Bukovsky regionalization of North America, Regional Integrated Sciences Collective, *Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO. Downloaded (2015): 05-08.*, pp. 06–18.

Burlando, P., and R. Rosso (1996), Scaling and multiscaling models of DDF for storm precipitations, *Journal of Hydrology*, *187*, 45–64, doi:10.1016/S0022-1694(96)03086-7.

Ceresetti, D. (2011), Structure spatio-temporelle des fortes précipitations: application à la région Cévennes-Vivarais, Ph.D. thesis, Université de Grenoble.

Ceresetti, D., G. Molinié, and J.-D. Creutin (2010), Scaling properties of heavy rainfall at short duration: A regional analysis, *Water Resources Research*, *46*(9), n/a–n/a, doi:10.1029/2009WR008603, w09531.

Coles, S., J. Heffernan, and J. Tawn (1999), Dependence measures for extreme value analyses, *Extremes*, *2*(4), 339–365, doi:10.1023/A:1009963131610.

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

CSA (2012), Development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: Guideline for Canadian water resources practitioners, *Tech. Rep. Canadian Standard Association, Tech. Rep. PLUS 4013, Mississauga, Ontario, 2nd ed.*

Cunnane, C. (1973), A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, *Journal of Hydrology*, *18*(3-4), 257–271, doi:10.1016/0022-1694(73)90051-6.

De Michele, C., N. T. Kottegoda, and R. Rosso (2001), The derivation of areal reduction factor of storm rainfall from its scaling properties, *Water Resources Research*, *37*(12), 3247–3252, doi:10.1029/2001wr000346.

Devine, K. A., and E. Mekis (2008), Field accuracy of Canadian rain measurements, *Atmosphere-Ocean*, *46*(2), 213–227, doi:10.3137/ao.460202.

Dubrulle, B., F. Graner, and D. Sornette (1997), *Scale Invariance and Beyond*, les Houches Workshop, march 10-14, 1997 ed., EDP Sciences, Les Ulis, France.

ECCC , Environment Climate Change Canada. Historical Climate Data Canada; editing status 2016-08-09; re3data.org - Registry of Research Data Repositories. last accessed: 2016-11-09, doi:10.17616/R3N012.

Eggert, B., P. Berg, J. O. Haerter, D. Jacob, and C. Moseley (2015), Temporal and spatial scaling impacts on extreme precipitation, *Atmos. Chem. Phys.*, *15*(10), 5957–5971, doi:10.5194/acp-15-5957-2015.

5 Good, P. (2013), *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer Science & Business Media.

Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis (1979), Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, *15*(5), 1049–1054.

Gupta, V. K., and E. Waymire (1990), Multiscaling properties of spatial rainfall and river flow distributions, *Journal of Geophysical Research: Atmospheres*, *95*(D3), 1999–2009, doi:10.1029/JD095iD03p01999.

10 Hartmann, D. L., A. M. G. Klein Tank, M. Rusicucci, L. V. Alexander, B. Broenniman, Y. Charabi, F. J. Dentener, E. J. Dlugokencky, D. R. Easterling, A. Kaplan, B. J. Soden, P. W. Thorne, M. Wild, P. M. Zhai, and E. C. Kent (2013), Observations: Atmosphere and surface, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 159–254, Cambridge University Press, Cambridge.

15 Hosking, J. R. M., and J. R. Wallis (1997), *Regional Frequency analysis: an approach based on L-moments*, Cambridge University Press.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985), Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, *27*(3), 251–261.

Hubert, P., and H. Bendjoudi (1996), Introduction à l'étude des longues séries pluviométriques, *XIIème journées hydrologiques de l'Orstom*, pp. 10–11.

20 Katz, R. W. (2013), Statistical methods for nonstationary extremes, in *Extremes in a Changing Climate*, edited by A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, no. 65 in Water Science and Technology Library, pp. 15–37, Springer Netherlands.

Katz, R. W., M. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, *25*, 1287–1304.

Koutsoyiannis, D. (2004a), Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation, *Hydrological Sciences Journal*, *49*(4), doi:10.1623/hysj.49.4.575.54430.

25 Koutsoyiannis, D. (2004b), Statistics of extremes and estimation of extreme rainfall: II. empirical investigation of long rainfall records, *Hydrological Sciences Journal*, *49*(4), doi:10.1623/hysj.49.4.591.54424.

Koutsoyiannis, D. (2005), Uncertainty, entropy, scaling and hydrological stochastics. 1. Marginal distributional properties of hydrological processes and state scaling/Incertitude, entropie, effet d'échelle et propriétés stochastiques hydrologiques. 1. Propriétés distributionnelles marginales des processus hydrologiques et échelle d'état, *Hydrological Sciences Journal*, *50*(3).

30 Koutsoyiannis, D., D. Kozonis, and A. Manetas (1998), A mathematical framework for studying rainfall intensity-duration-frequency relationships, *Journal of Hydrology*, *206*(1-2), 118–135, doi:10.1016/S0022-1694(98)00097-3.

Kunkel, K. E., D. R. Easterling, D. A. R. Kristovich, B. Gleason, L. Stoecker, and R. Smith (2012), Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States, *Journal of Hydrometeorology*, *13*(3), 1131–1141, doi:10.1175/JHM-D-11-0108.1.

35 Langousis, A., A. A. Carsteanu, and R. Deidda (2013), A simple approximation to multifractal rainfall maxima using a generalized extreme value distribution model, *Stochastic Environmental Research and Risk Assessment*, *27*(6), 1525–1531.

Lovejoy, S., and B. B. Mandelbrot (1985), Fractal properties of rain, and a fractal model, *Tellus A*, *37*(3), 209–232.

Lovejoy, S., and D. Schertzer (1985), Generalized scale invariance in the atmosphere and fractal models of rain, *Water Resources Research*, *21*(8), 1233–1250, doi:10.1029/WR021i008p01233.

Maraun, D., F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, and others (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, *48*(3).

Mascaro, G., R. Deidda, and M. Hellies (2013), On the nature of rainfall intermittency as revealed by different metrics and sampling approaches, *Hydrology and Earth System Sciences*, *17*(1), 355–369, doi:10.5194/hess-17-355-2013.

MDDELCC , Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques, 2016. Doneeés du Programme de surveillance du climat, Direction geéneérale du suivi de l'état de l'environnement, Queébec.

Menabde, M., A. Seed, and G. Pegram (1999), A simple scaling model for extreme rainfall, *Water Resources Research*, *35*(1), 335–339.

Nhat, L. M., Y. Tachikawa, T. Sayama, and K. Takara (2007), A simple scaling charateristics of rainfall in time and space to derive intensity duration frequency relationships, *Ann. J. Hydraul. Eng*, *51*, 73–78.

NOAA , Climate Data Online; editing status 2016-06-17; re3data.org - Registry of Research Data Repositories. last accessed: 2016-11-09, doi:10.17616/R32059.

Olsson, J., V. P. Singh, and K. Jinno (1999), Effect of spatial averaging on temporal statistical and scaling properties of rainfall, *Journal of Geophysical Research: Atmospheres*, *104*(D16), 19,117–19,126, doi:10.1029/1999JD900271.

Overeem, A., A. Buishand, and I. Holleman (2008), Rainfall depth-duration-frequency curves and their uncertainties, *Journal of Hydrology*, *348*(1-2), 124–134, doi:10.1016/j.jhydrol.2007.09.044.

Panthou, G., T. Vischel, T. Lebel, G. Quantin, and G. Molinié (2014), Characterising the space–time structure of rainfall in the Sahel with a view to estimating IDAF curves, *Hydrology and Earth System Sciences*, *18*(12), 5093–5107, doi:10.5194/hess-18-5093-2014.

Papalexiou, S. M., D. Koutsoyiannis, and C. Makropoulos (2013), How extreme is extreme? An assessment of daily rainfall distribution tails, *Hydrology and Earth System Sciences*, *17*(2), 851–862, doi:10.5194/hess-17-851-2013.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire (1984), Scale considerations in the modeling of temporal rainfall, *Water Resources Research*, *20*(11), 1611–1619.

Sivakumar, B. (2000), Fractal analysis of rainfall observed in two different climatic regions, *Hydrological Sciences Journal*, *45*(5), 727–738, doi:10.1080/02626660009492373.

Sivapalan, M., and G. Blöschl (1998), Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves, *Journal of Hydrology*, *204*(1-4), 150–167, doi:10.1016/S0022-1694(97)00117-0.

Tessier, Y., S. Lovejoy, and D. Schertzer (1993), Universal multifractals: Theory and observations for rain and clouds, *J. Appl. Meteorol., 32, 223-250*, *32*, 223–250.

Veneziano, D., and P. Furcolo (2002), Multifractality of rainfall and scaling of intensity-duration-frequency curves, *Water Resources Research*, *38*(12), 1306, doi:10.1029/2001WR000372.

Veneziano, D., and V. Iacobellis (2002), Multiscaling pulse representation of temporal rainfall, *Water Resources Research*, *38*(8), 13–1, doi:10.1029/2001WR000522.

Veneziano, D., and A. Langousis (2010), Scaling and fractals in hydrology, in *Advances in data-based approaches for hydrologic modeling and forecasting.*, Sivakumar, Bellie and Berndtsson, Ronny ed., World Scientific, Singapore.

Veneziano, D., and S. Yoon (2013), Rainfall extremes, excesses, and intensity-duration-frequency curves: A unified asymptotic framework and new nonasymptotic results based on multifractal measures, *Water Resources Research*, *49*(7), 4320–4334, doi:10.1002/wrcr.20352.

Veneziano, D., C. Lepore, A. Langousis, and P. Furcolo (2007), Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall, *Water Resources Research*, *43*(10), n/a–n/a, doi:10.1029/2007wr006040.

Venugopal, V., S. G. Roux, E. Foufoula-Georgiou, and A. Arnéodo (2006), Scaling behavior of high resolution temporal rainfall: New insights from a wavelet-based cumulant analysis, *Physics Letters A*, *348*(3), 335–345.

5   Wallis, J. R., M. G. Schaefer, B. L. Barker, and G. H. Taylor (2007), Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State, *Hydrology and Earth System Sciences*, *11*(1), 415–442, doi:10.5194/hess-11-415-2007.

Westra, S., H. J. Fowler, J. P. Evans, L. V. Alexander, P. Berg, F. Johnson, E. J. Kendon, G. Lenderink, and N. M. Roberts (2014), Future changes to the intensity and frequency of short-duration extreme rainfall, *Reviews of Geophysics*, *52*(3), 2014RG000,464, doi:10.1002/2014RG000464.

10 Willems, P., K. Arnbjerg-Nielsen, J. Olsson, and V. T. V. Nguyen (2012), Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings, *Atmospheric Research*, *103*, 106–118, doi:10.1016/j.atmosres.2011.04.003.

Yu, P.-S., T.-C. Yang, and C.-S. Lin (2004), Regional rainfall intensity formulas based on scaling property of rainfall, *Journal of Hydrology*, *295*(1-4), 108–123, doi:10.1016/j.jhydrol.2004.03.003.

Table 1. List of available datasets and their main characteristics.

Dataset	Region	N. of stations	Operational period[b]	Temporal resolution	Prevalent[c] resolution [mm]
Daily Maxima[a] (DM)	Canada	370	1964-2007	1, 2, 6, 12 h	0.1 (82.25%)
Hourly (H)	Canada	665	1967-2003	1 h	0.1 (70%)
Hourly Prec. Data (HPD)	USA	2531	1948-2013	1 h	0.254 (82.5%)
15-Min Prec. Data (15PD)	USA	2029	1971-2013	15 min	2.54 (80.42%)

[a] Daily maxima depth series over a 24-hour window beginning at 8:00 AM.

[b] Main station network operational period corresponding to 25th percentile of the first recording year and the 75th percentile of the last recording year of the stations.

[c] Prevalent measurement resolution, estimated by the lowest non-zero value for each series, and corresponding percentage of stations with this resolution.

Table 2. Final datasets used in scaling analysis and corresponding AMS characteristics.

Scaling dataset	Durations	N. of Stations	Mean series length [yr]	Max series length [yr]
SD[a]	15min, 30min, ..., 6h	1083	20	36
ID	1h, 2h, ...,24h	2719	37.4	66
LD	6h, 12h, ..., 168h	2719	37.4	66

[a] Only 15PD series.
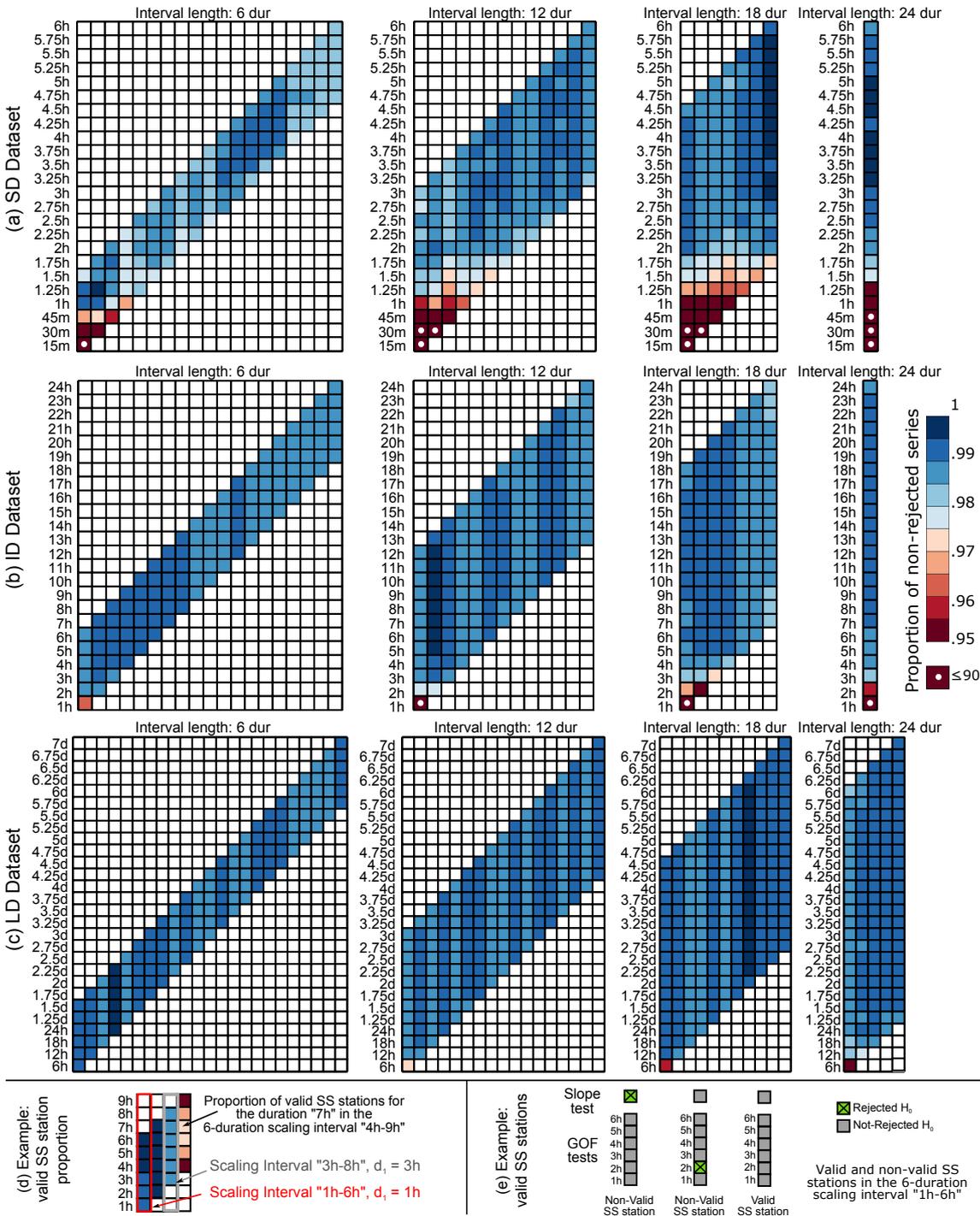
Figure 1. a) - c) Proportion of stations satisfying both the Slope and GOF tests applied at the 0.95 confidence level, for each duration (vertical axis) and scaling interval (horizontal axis) for the SD, ID, and LD datasets [row a), b), and c) respectively]. White circles indicate proportions between 0.25 and 0.90; d) Example of valid SS station proportion values and identification of durations and scaling intervals within each matrix; e) Examples of valid and non-valid SS stations.
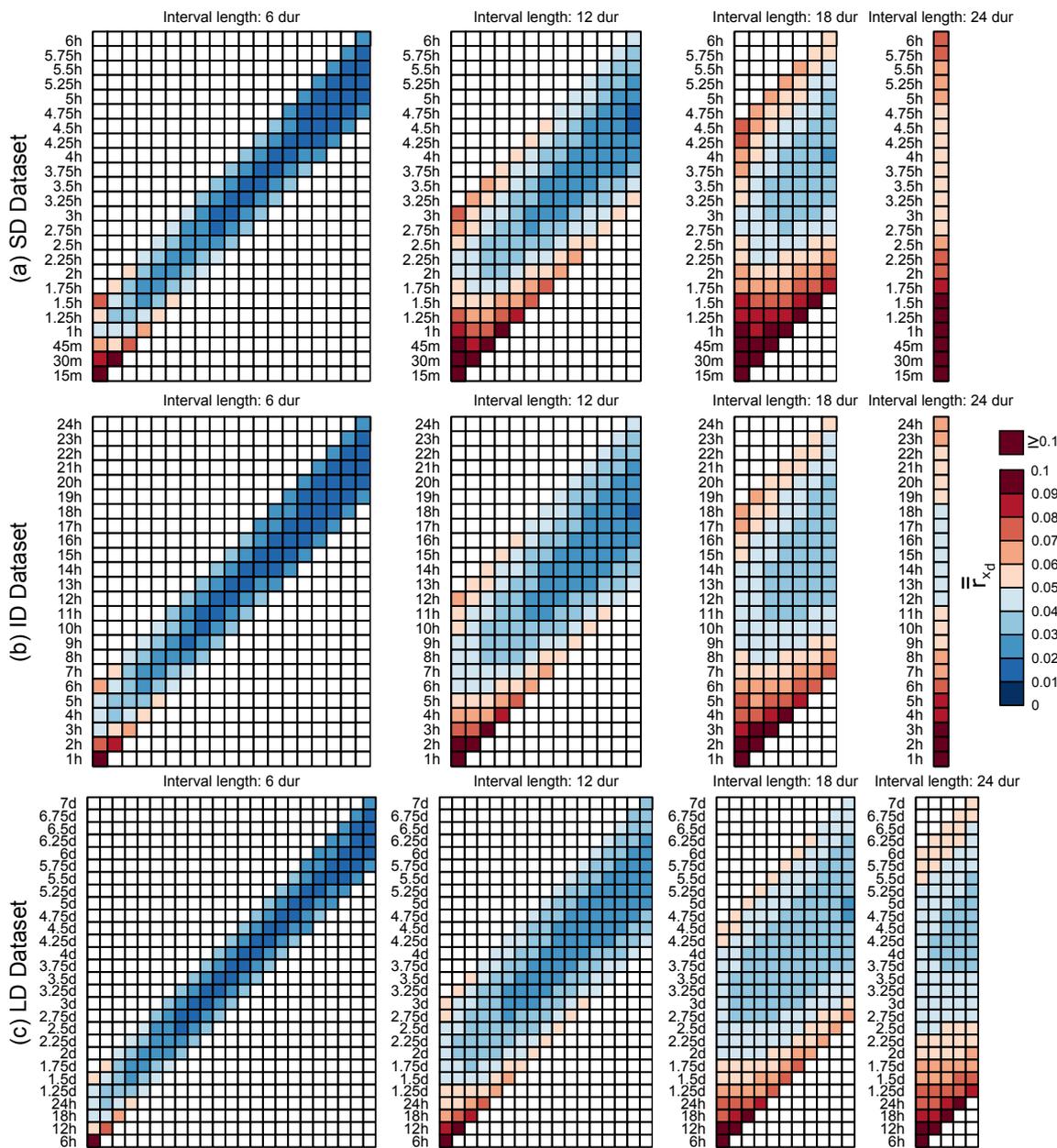
Figure 2. Cross-Validation Normalized RMSE averaged over all valid SS stations ($\overline{r}_{x_d}$) for each duration (vertical axis) and scaling interval (horizontal axis) in the SD, ID, and LD datasets [row a), b), and c) respectively]. See Fig. 1 (d) for the identification of durations and scaling intervals within each matrix. White circles indicate values between 0.15 and 0.3.
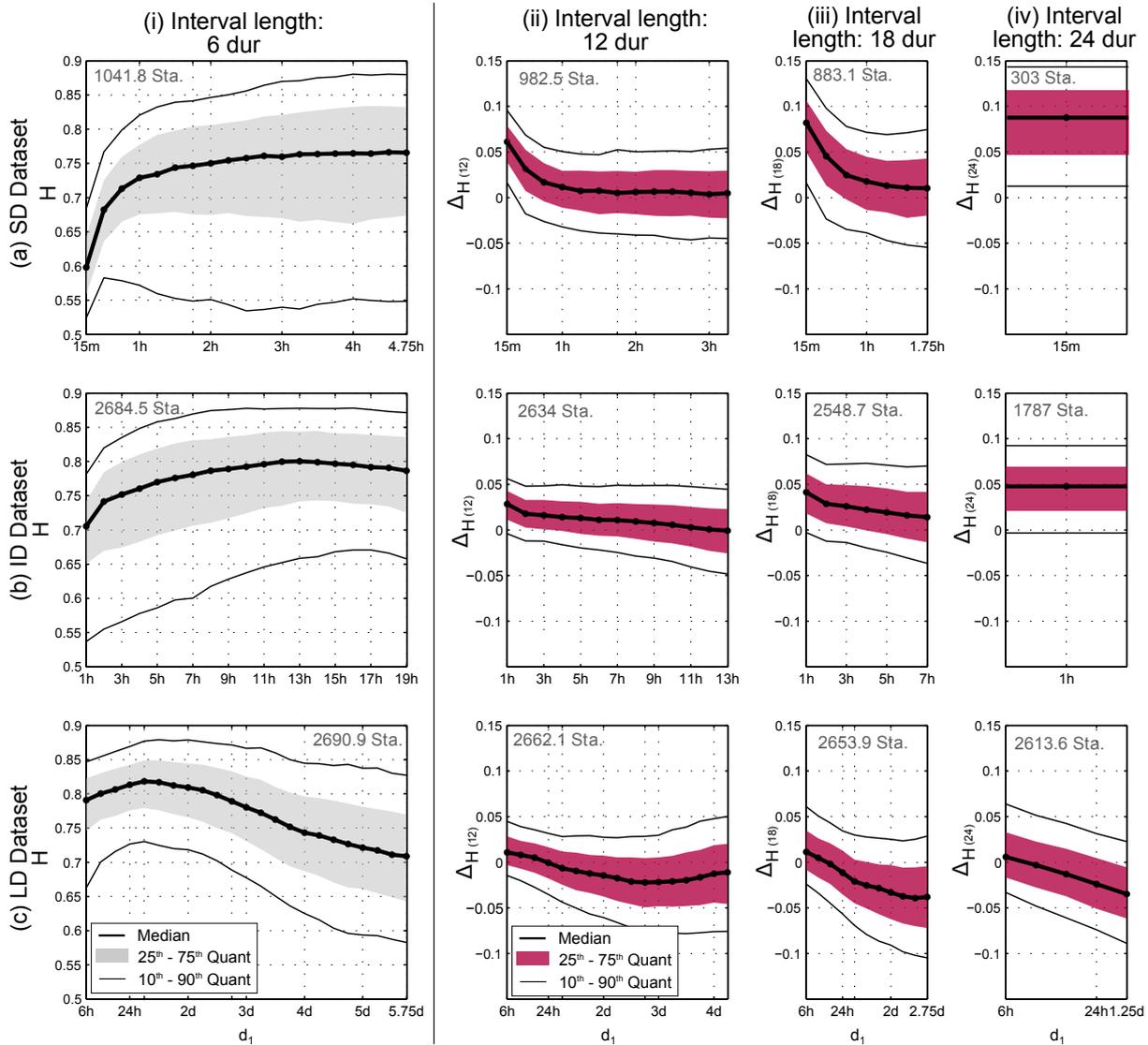
Figure 3. Col. (i): Median and relevant quantiles of the scaling exponent distribution over all valid SS stations for each 6-duration scaling interval. Col. (ii)-(iv): Median and relevant quantiles of the distribution of the scaling exponent deviation $\Delta_{H_{(j)}}$ [defined in Eq. (12)]. The average number of valid SS stations over the scaling intervals (identified by their first duration, d_1) is indicated at the top of each graph.
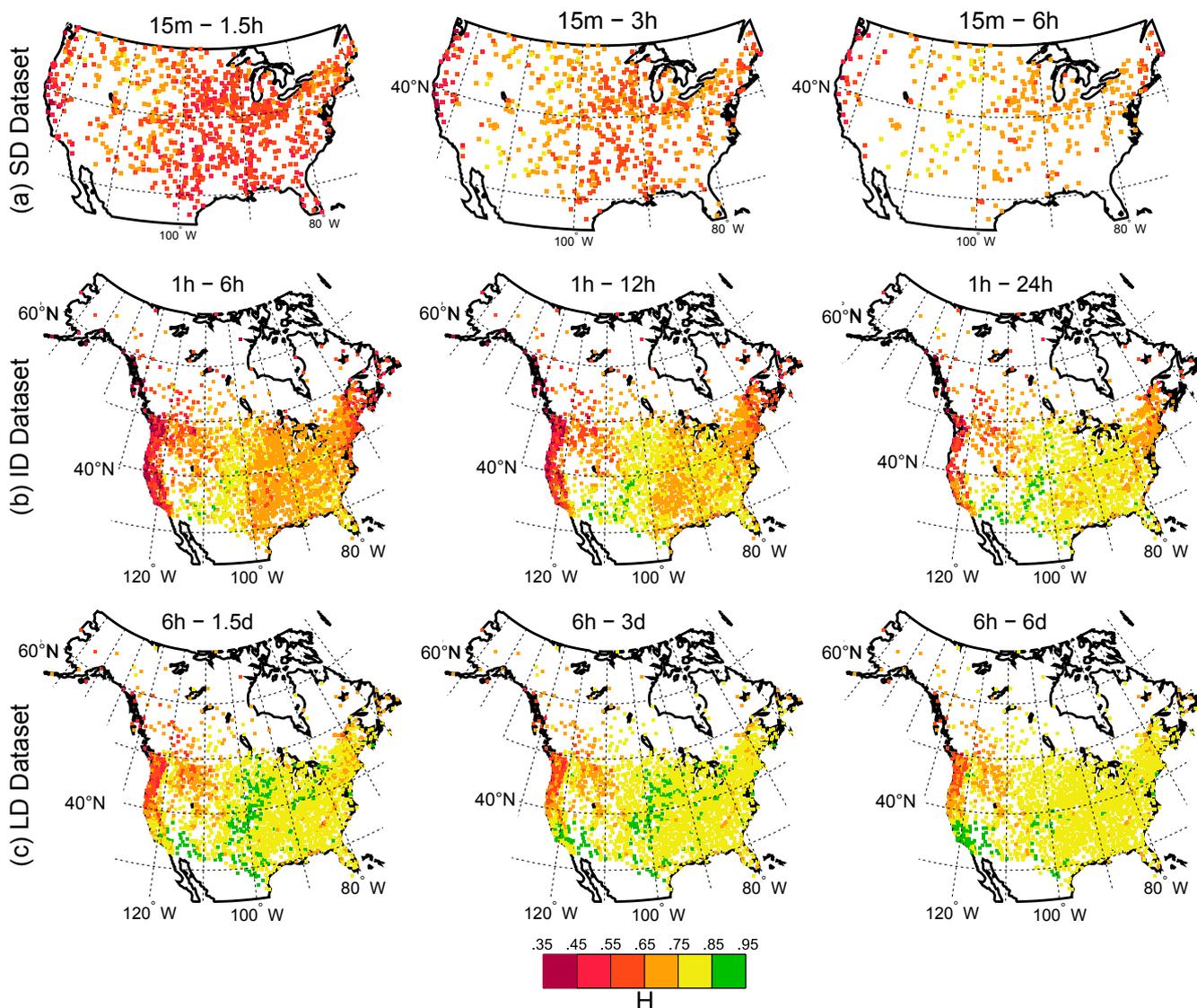
Figure 4. Spatial distribution of the scaling exponent for the first (i.e. with minimum d_1) 6-, 12-, and 24-duration scaling intervals (first, second, and third col., respectively) for SD, ID, and LD datasets (first, second, and third row, respectively). These scaling intervals correspond to the first column of matrices in Fig. 1 and 2.
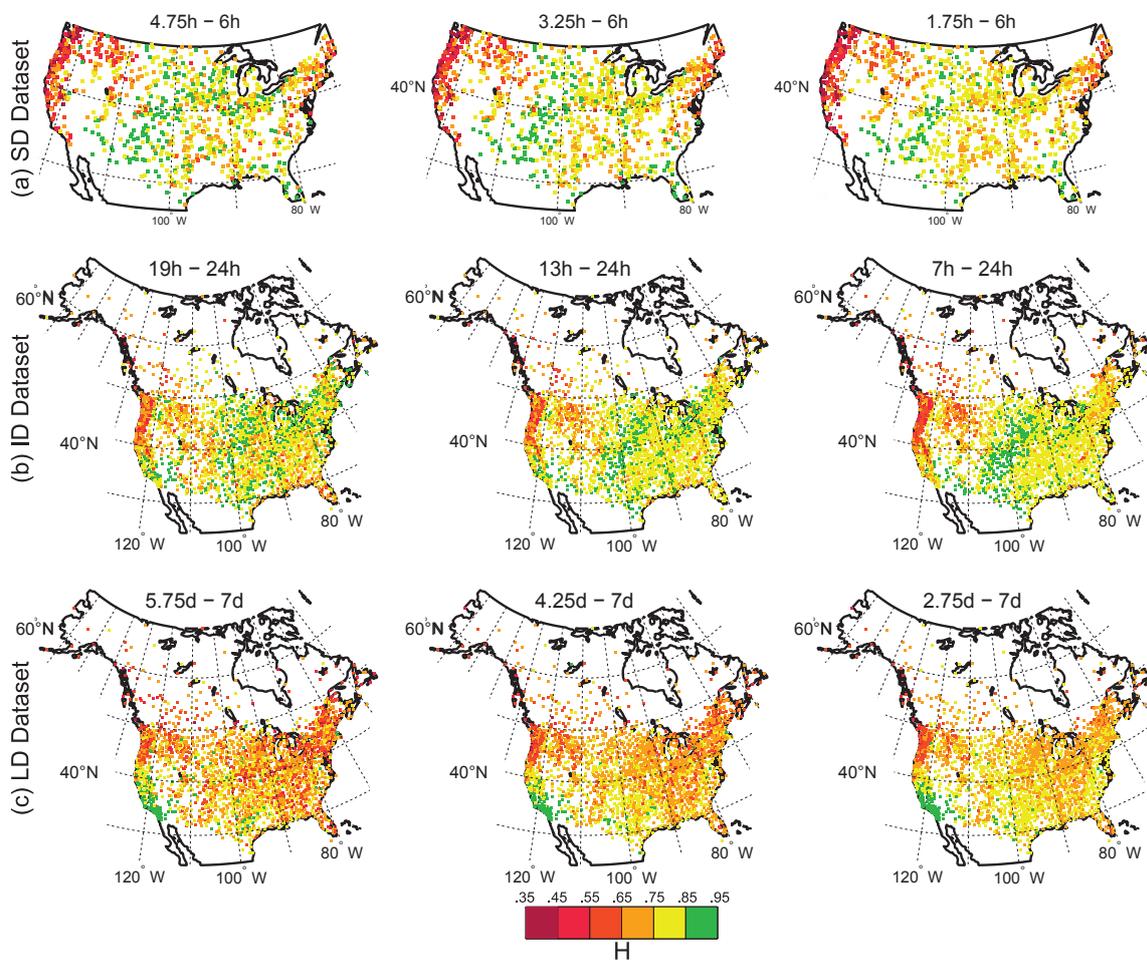
Figure 5. Spatial distribution of the scaling exponent for the last (i.e. with maximum d_1) 6-, 12-, and 18-duration scaling intervals (first, second, and third col., respectively) for SD, ID, and LD datasets (first, second, and third row, respectively). These scaling intervals correspond to the last column of matrices in Fig. 1 and 2.
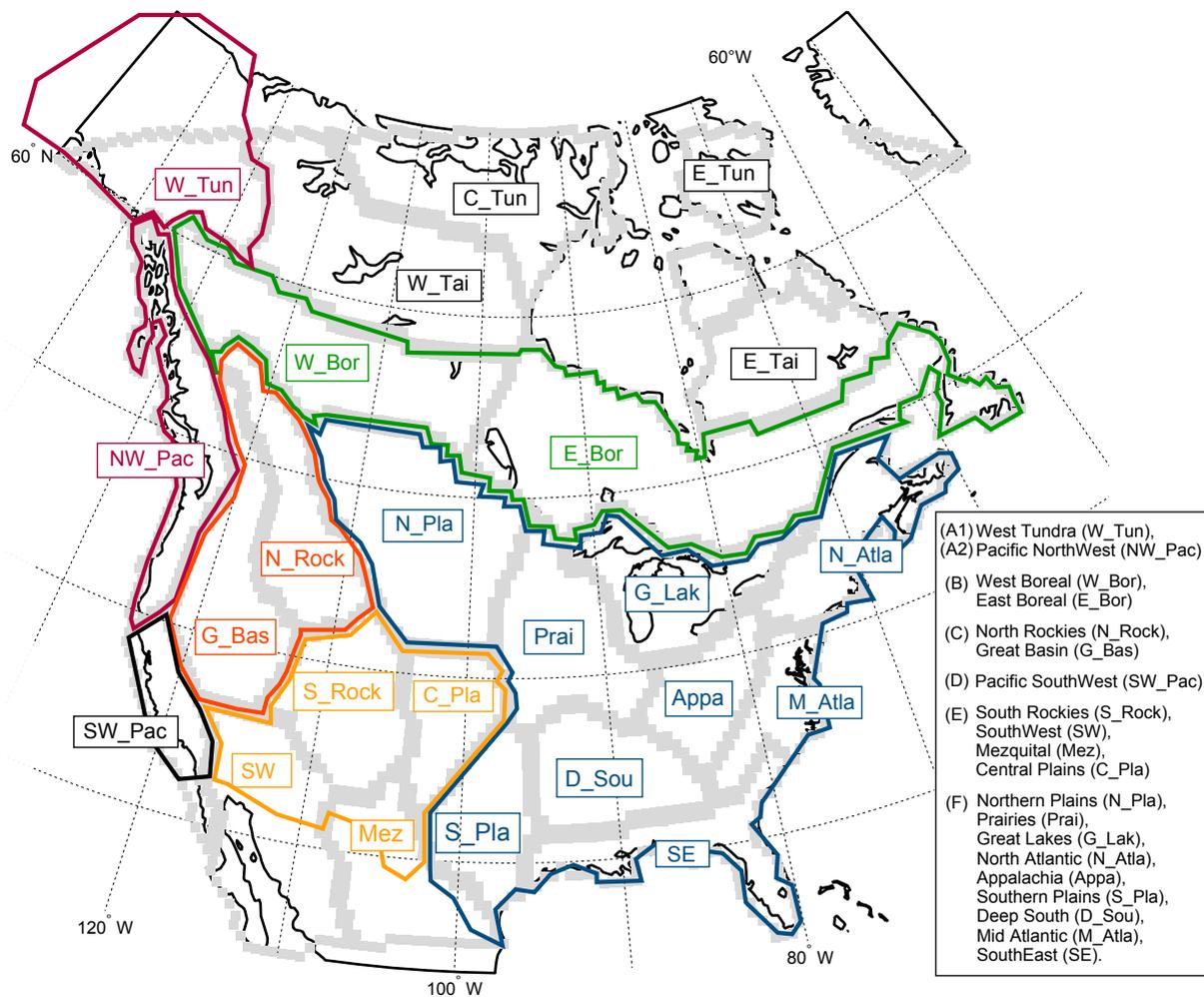
Figure 6. Climatic regions of *Bukovsky* (2012) [grey borders] and regions defined for this analysis [regions A1 to F in the legend; colored borders]. Abbreviations for each region are in parenthesis.
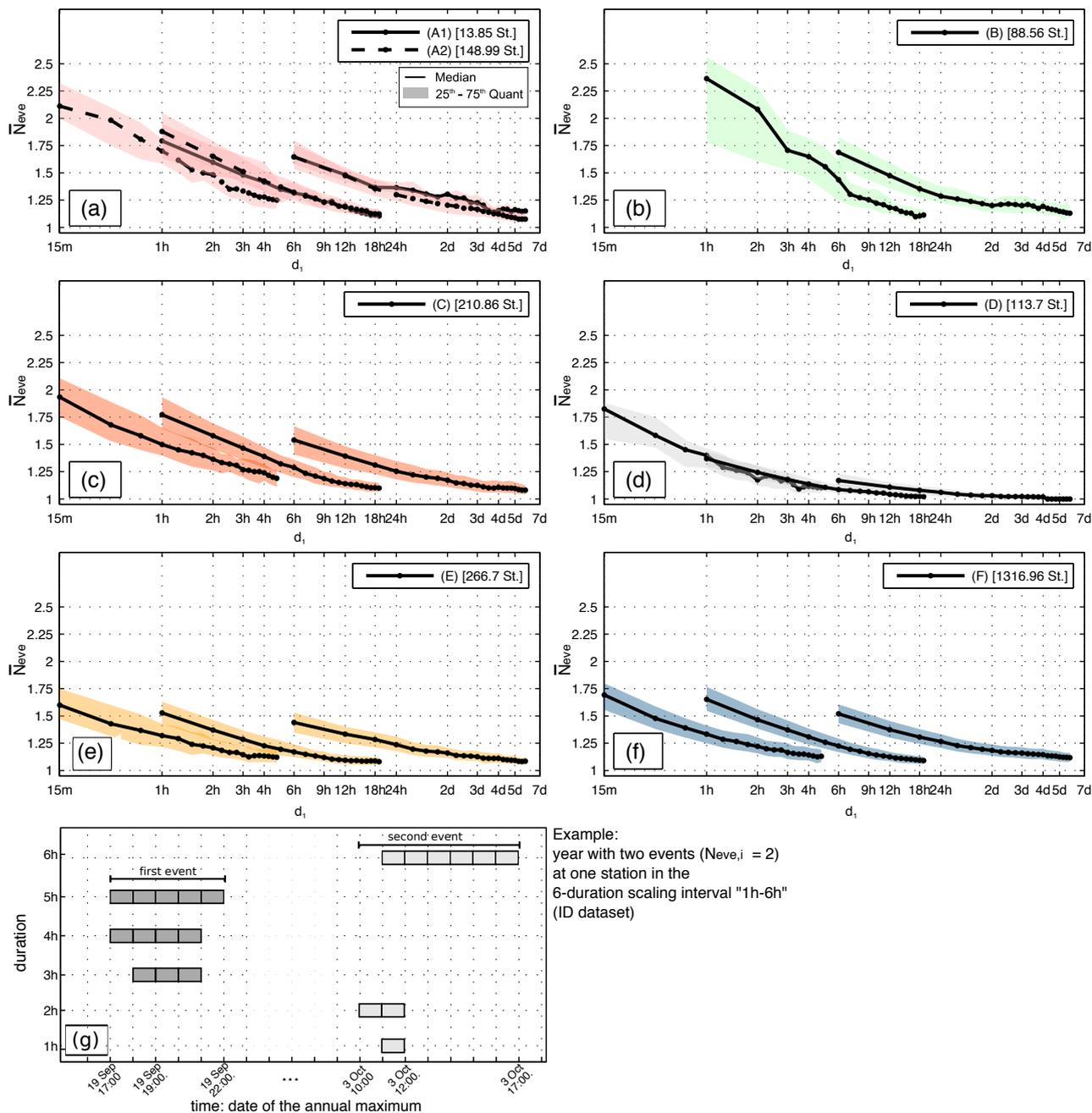
Figure 7. Median and Interquantile Range (IR) of the distribution over valid SS stations of the mean number of events per year, $\bar{N}_{eve}$, within each region of Fig. 6 for each 6-duration scaling interval for the SD (left curve), ID (central curve), and LD (right curve) datasets. For each region, the mean number of valid SS stations over the scaling intervals is indicated in brackets in the legend. Panel (g) displays an example of how $N_{eve,i}$ is estimated for year i.
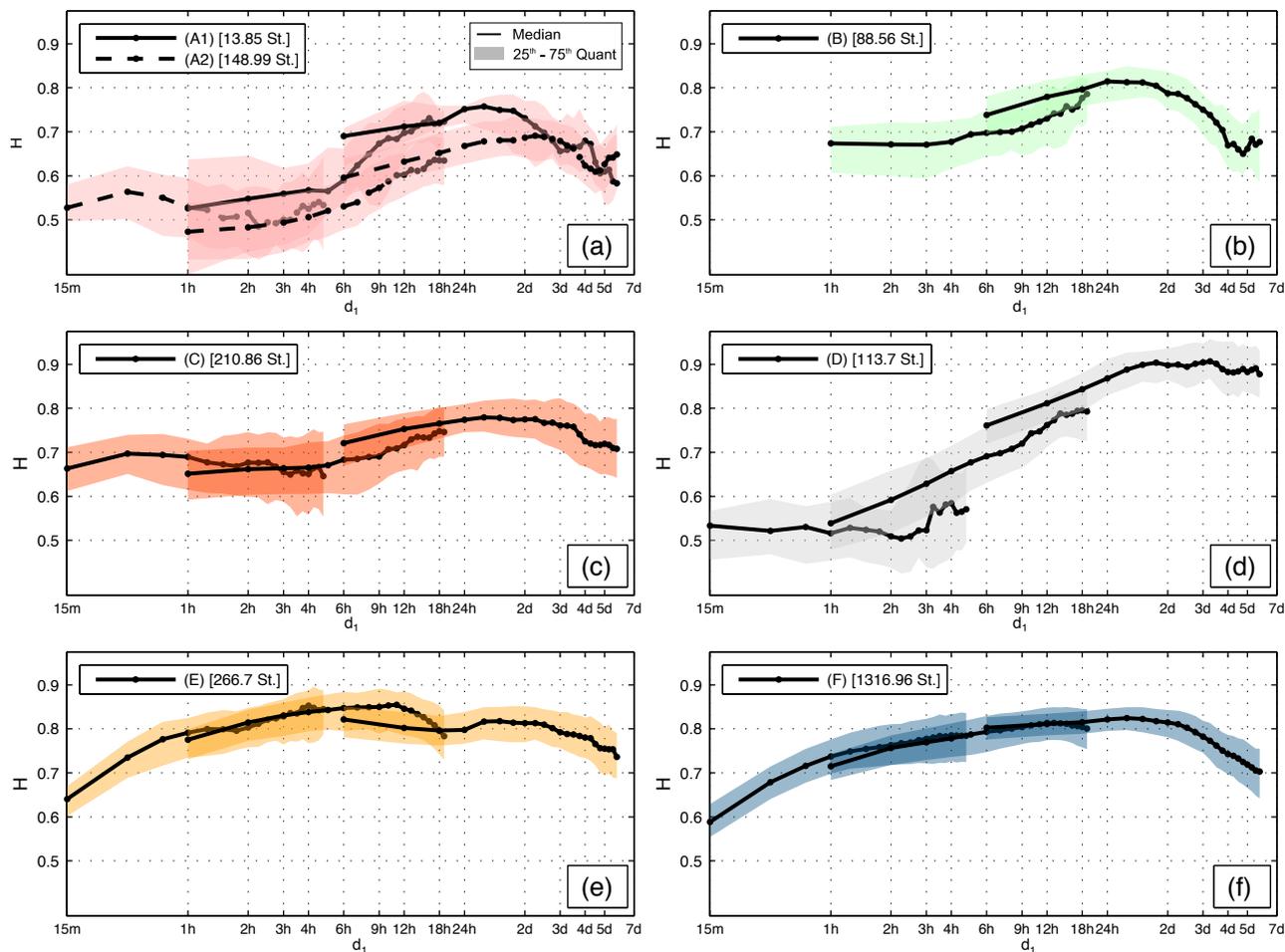
Figure 8. Median and Interquantile Range (IR) of the scaling exponent distribution over valid SS stations within each region of Fig. 6 for 6-duration scaling intervals for the SD (left curve), ID (central curve), and LD (right curve) datasets. For each region, the mean number of valid SS stations over the scaling intervals is indicated in brackets in the legend.
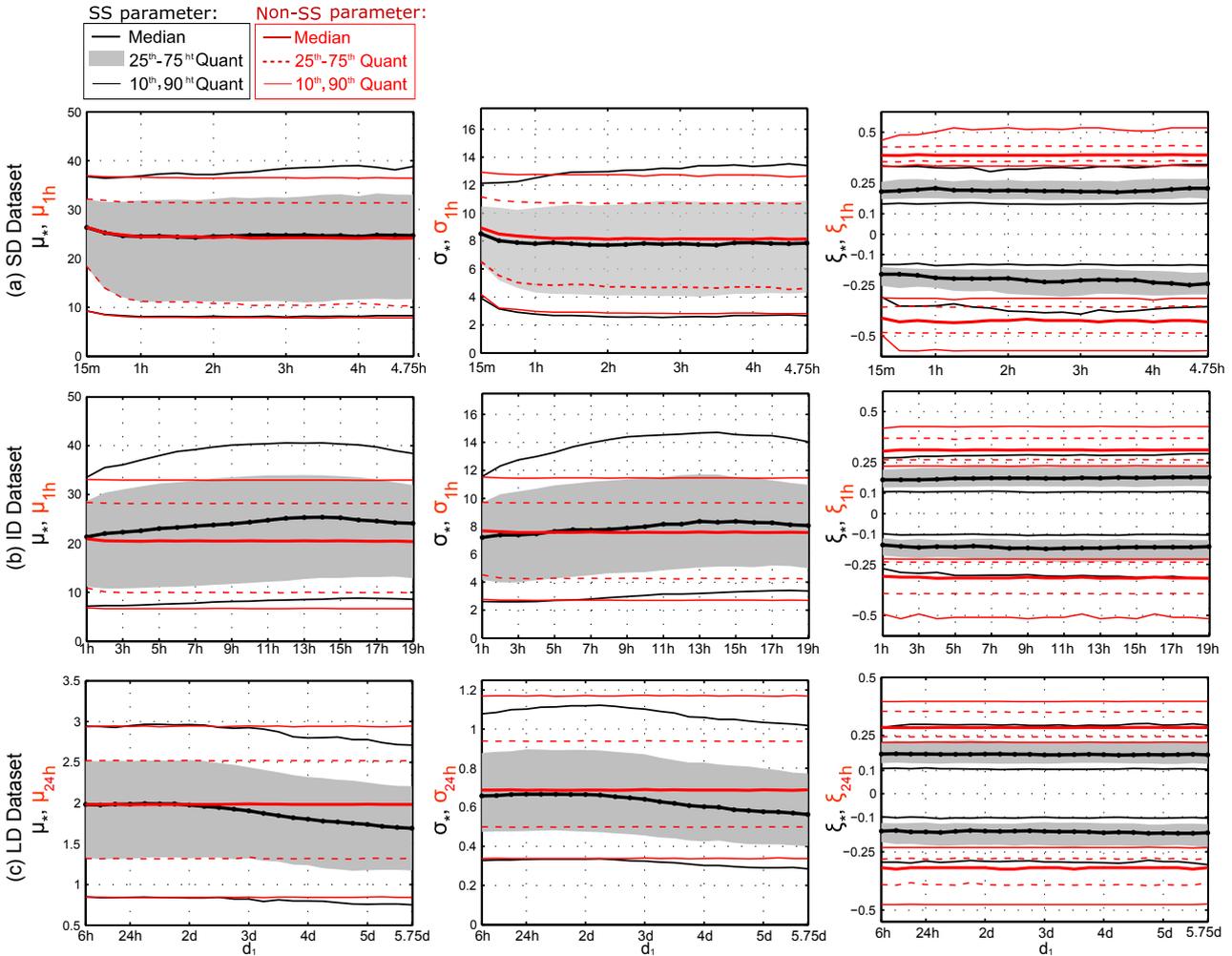
Figure 9. Distribution over valid SS stations of SS GEV (gray and black lines) and non-SS GEV (red solid and dashed lines) parameters for 6-duration scaling intervals. Location and scale parameters (first and second col., respectively) are scaled at $d_* = 1$h (SD and ID datasets) and $d_* = 24$h (LD dataset). Distributions for the shape parameter (third col.) are presented for $\xi > 0$ and $\xi < 0$, excluding cases where $\xi = 0$ (Gumbel distribution).
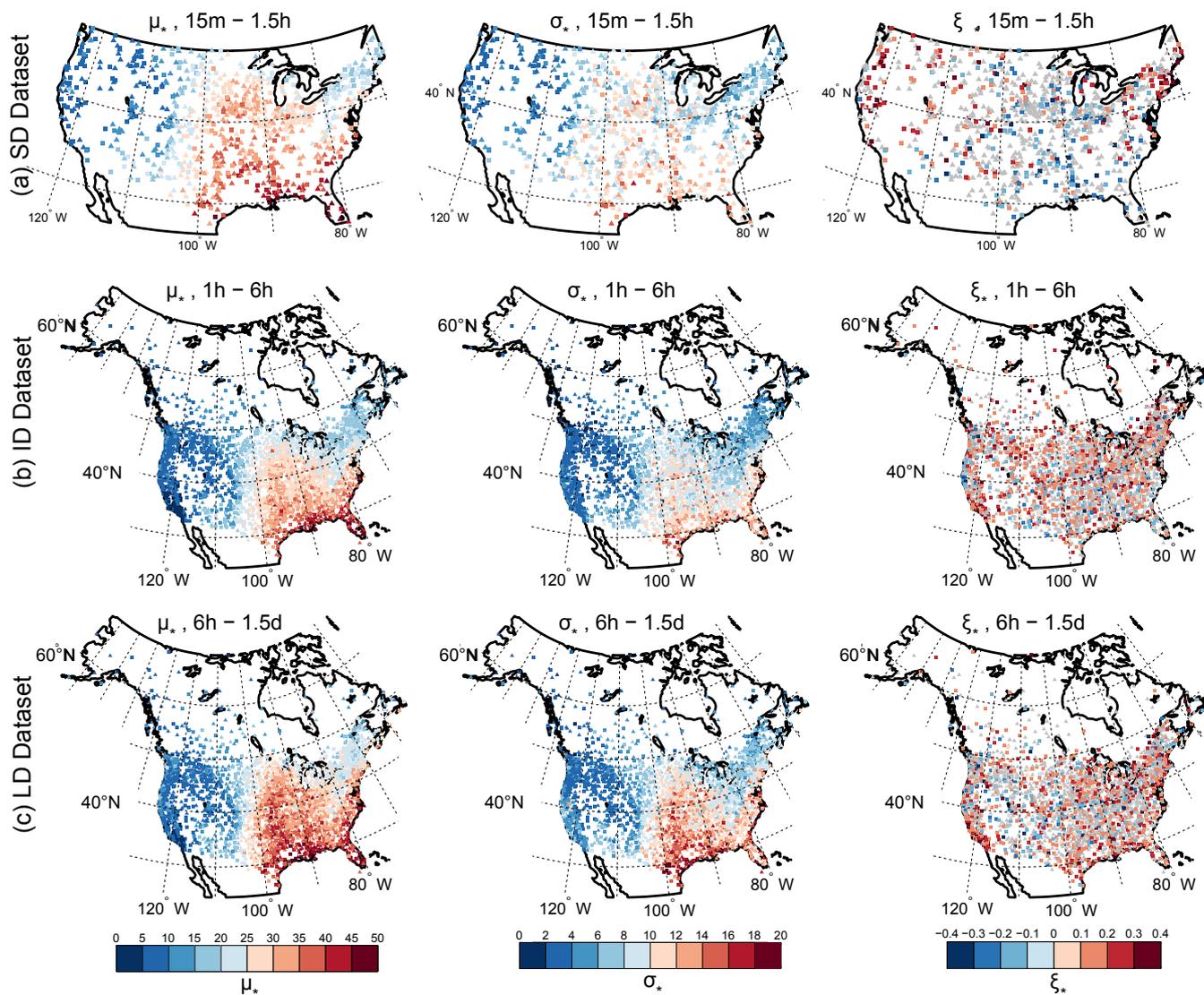
Figure 10. Spatial distribution over valid SS stations of SS GEV position (first col.), scale (second col.), and shape (3rd col.; gray symbols indicate Gumbel distributions, $\xi_* = 0$) parameters scaled at $d_* = 1$h for the first 6-duration scaling interval (i.e. interval with minimum d_1) of: SD (a), ID (b), and LD (c) datasets.
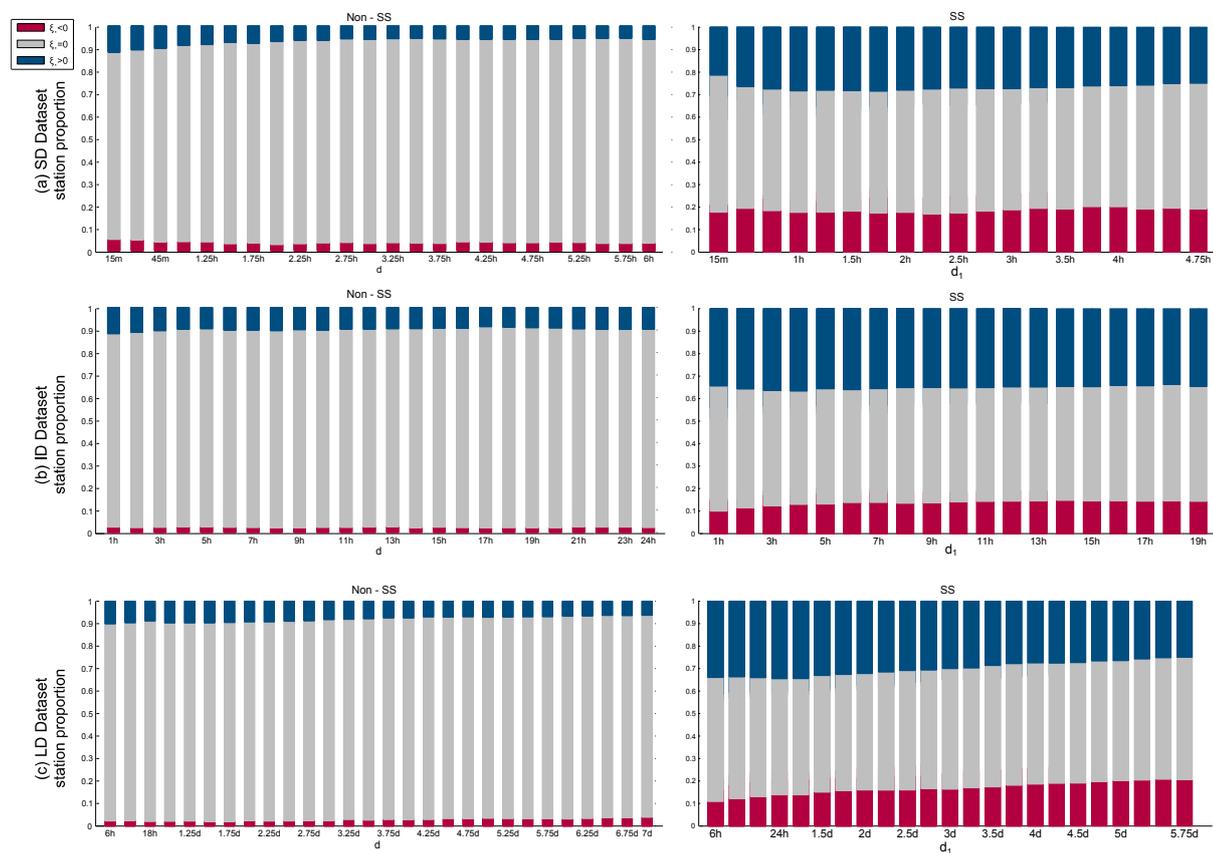
Figure 11. Stacked histograms of the fractions of valid SS stations with $\xi < 0$ (in red), $\xi = 0$ (in grey), and $\xi > 0$ (in blue) for each duration (non-SS GEV, first col.) and each 6-duration scaling interval (SS GEV, second col.) for: SD (a), ID (b), and LD (c) datasets.
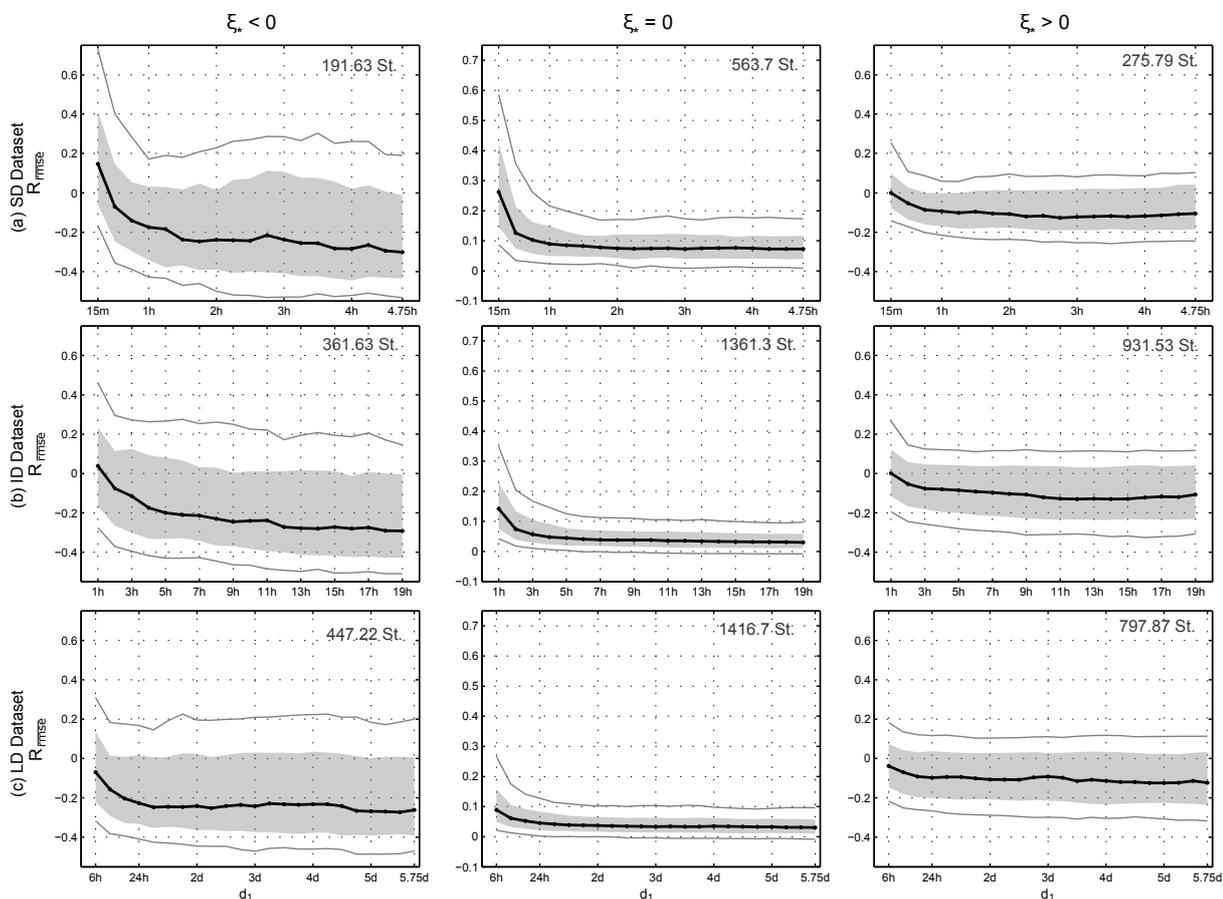
Figure 12. Distribution of the relative total RMSE ratio, $R_{\overline{rmse}}$, for $\xi_* < 0$ (first col.), $\xi_* = 0$ (second col.), and $\xi_* > 0$ (third col.) for 6-duration scaling intervals in SD (a), ID (b), and LD (c) datasets. The average number of valid SS station over the scaling intervals is indicated in the right-top corner of each graph.