Stochastic generation of multi-site daily precipitation focusing on extreme events

Guillaume Evin, Anne-Catherine Favre, Benoît Hingray

HAL Id: hal-02912352

https://hal.inrae.fr/hal-02912352

Submitted on 12 Mar 2021

Stochastic generation of multi-site daily precipitation focusing on extreme events

Guillaume Evin, Anne-Catherine Favre, and Benoit Hingray

Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

Correspondence: Guillaume Evin (guillaume.evin@irstea.fr)

Abstract. Many multi-site stochastic models have been proposed for the generation of daily precipitation, but they generally focus on the reproduction of low to high precipitation amounts at the stations concerned. This paper proposes significant extensions to the multi-site daily precipitation model introduced by Wilks, with the aim of reproducing the statistical features of extremely rare events (in terms of frequency and magnitude) at different temporal and spatial scales. In particular, the first extended version integrates heavy-tailed distributions, spatial tail dependence, and temporal dependence in order to obtain a robust and appropriate representation of the most extreme precipitation fields. A second version enhances the first version using a disaggregation method. The performance of these models is compared at different temporal and spatial scales on a large region covering approximately half of Switzerland. While daily extremes are adequately reproduced at the stations by all models, including the benchmark Wilks version, extreme precipitation amounts at larger temporal scales (e.g., 3-day amounts) are clearly underestimated when temporal dependence is ignored.

1 Introduction

Stochastic precipitation generators are often employed in risk assessment studies to estimate the return periods of very rare flooding events (e.g., 10 000-year events). The observed series of streamflows are too short to produce reliable estimations of very rare and large floods. Typically, extreme hydrological events can be reproduced using long series of simulated precipitation data as input to hydrological models (Lamb et al., 2016).

In the last two decades, a number of precipitation models have been proposed to deal with the temporal and spatial properties of daily precipitation, for both intermittency and amount, and all have different strengths and weaknesses. Many of these models use exogenous variables to predict the statistical properties of precipitation using generalized linear models (Chandler and Wheater, 2002; Mezghani and Hingray, 2009; Serinaldi and Kilsby, 2014b), atmospheric analogs (Lafaysse et al., 2014), or modified Markov models (Mehrotra and Sharma, 2010). Introducing a link between exogenous atmospheric variables can be used to reconstruct past events, make predictions, or downscale global-climate-model-based simulations of future climate. Such models are classically referred to as statistical downscaling models (see Maraun et al., 2010, for a review.) Closely related to this approach, weather "types" or "regimes" (Ailliot et al., 2015) can be used to specifically account for different atmospheric circulation patterns. Using a hidden Markov model (HMM) with transitions between these weather states, stochastic weather generators can then simulate various aspects of the precipitation process (Rayner et al., 2016).

Alternatively, purely stochastic precipitation models can be used. These can be broadly classified into three main types.

- *Resampling methods.* The stochastic generation of precipitation fields can be performed using resampling techniques such as the *K*-nearest neighbors algorithm (Buishand, 1991; Yates et al., 2003). Unobserved precipitation amounts can be obtained using perturbation techniques (Sharif and Burn, 2007).

- *Random fields.* Spatiotemporal precipitation models can simulate precipitation fields over a regular grid. This ap-

proach is particularly useful for hydrological applications, since areal precipitation values over a basin are obtained directly. Poisson cluster-based models (Burton et al., 2008, 2010; Leonard et al., 2008; McRobie et al., 2013) randomly simulate rain disk cells, with random centers, radius and intensity, over the study area. Meta-Gaussian models (Vischel et al., 2009; Kleiber et al., 2012; Allard and Bourotte, 2015; Baxevani and Lennartsson, 2015; Bennett et al., 2017) are based on truncated and transformed random Gaussian fields. Closely related, the turning band method can be used to simulate intermittent precipitation fields with different types of advection (Leblois and Creutin, 2013). These model structures are appealing since they are able to simulate realistic precipitation fields at fine spatial scales. However, their complexity leads to numerous technical issues during parameter estimation and simulation, notably in terms of computational cost. Moreover, they are usually unable to represent large regions comprising very distinct precipitation regimes.

– *Statistical multi-site models.* In this last type of weather generator, the properties of precipitation are directly fitted at a limited number of stations using different statistical structures. This type of generator preserves the interdependency between all pairs of stations, even when the area under study exhibits different precipitation regimes. Bárdossy and Pegram (2009) and Rasmussen (2013) combine a multivariate autoregressive process and transformations (V-transform, power transformation) to simultaneously model precipitation occurrence and amount. More precisely, with these models, transformed precipitation amounts follow truncated distributions. Alternatively, Wilks (1998) proposes a multi-site model in which precipitation occurrence and amount are handled separately. Several extensions to this popular structure have been proposed in the literature. Thompson et al. (2007) reformulate the Wilks model as a hidden Markov model, inferring three precipitation states ("dry", "light", and "heavy"). Mehrotra and Sharma (2007b) apply semi-parametric techniques to add more flexibility to the spatial structure of precipitation occurrence and amount. Srikanthan and Pegram (2009) propose a modified version in which daily, monthly, and annual amounts are nested such that precipitation statistics are preserved for all these levels of aggregation.

Mehrotra et al. (2006) compare three different precipitation models, the Wilks model, a HMM and a resampling approach, and they provide strong arguments in favor of the Wilks model in terms of performance, computation time, model, and level of complexity of the model structure. Furthermore, as indicated above, this model offers a flexible structure, which can be applied to a large number of stations

with very different precipitation regimes (like in mountainous areas). This paper presents several significant extensions of the Wilks precipitation model, referred to as GWEX versions, which will be used to generate long scenarios. These extensions aim to fit the most extreme precipitation amounts at different temporal (1- and 3-day amounts) and spatial scales. Novel components are thus introduced in GWEX, including robust estimation methods (regionalization methods) for critical parameters impacting directly on the behavior of extreme precipitation at each station. Recent advances in the choice of the marginal distributions for daily precipitation amounts are also included. Using 15 029 long daily precipitation records (> 50 years) from around the world, Papalexiou et al. (2013) conclude that heavy-tailed distributions are generally in better agreement with observed precipitation extremes. Follow-up studies (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014a) apply extreme value theory to annual maxima and "peaks over threshold" (POTs) of a large subset of these records and confirm that extreme daily precipitation is not adequately represented by light-tailed distributions. Based on statistical tests on 90 000 station records of daily precipitation, Cavanaugh et al. (2015) also come to the same conclusion. These findings have important implications for precipitation models.

– Light-tailed distributions such as exponential, Gamma, and Weibull distributions, which are applied in the vast majority of the existing precipitation models, often lead to an underestimation of extreme daily precipitation amounts.

– While nonparametric densities with Gaussian kernels (Mehrotra and Sharma, 2007a, 2010) offer the flexibility of fitting the observed range of precipitation amounts, their tail also belongs to the domain of attraction of the Gumbel distribution and suffers from the same drawbacks.

Alternatively, current statistical procedures consisting of fitting a flexible distribution to the bulk of the observations and using it for extrapolation are highly questionable, as major assumptions are usually violated (Klemeš, 2000a, b). Since the tail of the distribution on precipitation amounts at each station will dictate the generation of the most extreme precipitation events, important features of GWEX are as follows:

– the application of a heavy-tailed distribution to precipitation amounts at each station (Naveau et al., 2016),

– the determination of robust estimates of the shape parameter of this distribution, which indicates the heaviness of the tail, using a regionalization approach, as in Evin et al. (2016).

Furthermore, following Bárdossy and Pegram (2009), GWEX also employs the copula theory to introduce a tail dependence between the precipitation amounts simulated at the

different stations. The second version of the GWEX model includes a disaggregation method, the observed precipitation amounts being fitted at a 3-day scale in a first step. This paper compares the performance of the different model versions and assesses the impact of the different statistical components (e.g., heavy-tailed distribution, tail dependence).

We first describe the study area in Sect. 2. The features of different multi-site precipitation models are then described in Sect. 3. The evaluation framework, presented in Sect. 4, aims to assess the performance of these models at different spatial and temporal scales. Section 5 presents an application of these daily precipitation models to 105 stations located in Switzerland, with a summary of the results focusing on the reproduction of extreme events. Finally, Sect. 6 presents our conclusions.

2 Data and study area

The Aare River basin covers the northern part of the Swiss Alps and has an area of $17\,700\,\mathrm{km}^2$. Basin elevations approximately range from $310\,\mathrm{m.a.s.l.}$ in Koblenz (next to the German border in the north) to $4270\,\mathrm{m.a.s.l.}$ at the Finsteraarhorn summit (in the south of the basin). The mean annual precipitation for the basin as a whole is $1300\,\mathrm{mm}$. The basin can be divided into five main sub-basins, with different hydrometeorological regimes highly governed by regional terrain features (Jura Mountains in the northwest, northern Alps in the south of the basin, and lowlands in the middle).

Figure 1 shows the location of the 105 precipitation stations used for the development and evaluation of weather generators. Located within or close to the Aare River basin, they correspond to the stations for which long daily time series of observations with less than 3 years of missing data are available over the period 1930–2014. The 105 precipitation stations cover the Aare River basin relatively well.

The proposed precipitation models are designed to simulate flood scenarios, via a conceptual hydrological model, for the whole Aare River basin and for its different sub-basins. For Switzerland, Froidevaux et al. (2015) show that the generation of floods is mainly influenced by areal precipitation amounts accumulated over short periods (e.g., 1 to 3 days). These results are obtained by analyzing a wide variety of basins, their areas ranging from 10 to $12\,000\,\mathrm{km}^2$. Therefore, the properties of the weather scenarios must be evaluated at different spatial and temporal scales, from the high resolutions required to simulate the hydrological behavior of the system (e.g., sub-daily, $100\,\mathrm{km}^2$) to lower resolutions relevant at the scale of the entire basin (e.g., n-days, $17\,700\,\mathrm{km}^2$). In this study, the performance of the different precipitation models is evaluated at the station scale, at the scale of 15 and 5 sub-basins partitioning the Aare River basin, and at the scale of the entire study area (see Sect. 5). Note that for those evaluations, areal estimates of precipitation are obtained from the precipitation amounts at the stations using the Thiessen polygon method.

3 Multi-site precipitation model

As indicated above, GWEX refers to multi-site precipitation models that rely strongly on the structure proposed by Wilks (1998). At each location k, let $P_t(k)$ be a random variable representing the accumulated precipitation over day t. The structure proposed by Wilks considers a hidden occurrence process $X_t(k)$ that can be represented by a two-state Markov chain as follows:

$$X_t(k) = \begin{cases} 0, & \text{if day } t \text{ is dry at location } k. \\ 1, & \text{if day } t \text{ is wet at location } k. \end{cases} \tag{1}$$

Precipitation amount $P_t(k)$ is then defined as

$$P_t(k) = Y_t(k)X_t(k), \tag{2}$$

where $Y_t(k)$ is a random variable describing the non-zero precipitation amounts. Non-zero precipitation amounts $Y_t(k)$ are thus modeled independently of precipitation occurrences $X_t(k)$, which act as a mask.

3.1 Precipitation occurrence process

3.1.1 At-site occurrence process

At each location, the temporal persistence of dry and wet events is introduced with a p-order Markov chain model for $X_t(k)$ so that

$$\begin{aligned} \Pr\{X_t(k) = 1 | X_{t-1}(k), \dots, X_1(k)\} \\ = \Pr\{X_t(k) = 1 | X_{t-1}(k), \dots, X_{t-p}(k)\}; \end{aligned} \tag{3}$$

i.e., the probability of having a wet day at time t depends only on the p previous states, for days $t-1, \dots, t-p$. While many authors suppose that a first-order Markov is sufficient (e.g., Wilks, 1998; Keller et al., 2015), Srikanthan and Pegram (2009) apply a fourth-order Markov chain and show that it improves the reproduction of dry/wet period lengths. In this study, different orders for this Markov chain are considered.

At each site, the probability of having a wet day on day t is given by the transition probability $\Pr\{X_t(k) = 1 | X_{t-1}(k) = i_1, \dots, X_{t-p}(k) = i_p\}$, where $i_1, \dots, i_p$ are equal to 0 or 1. This Markov chain is thus fully characterized by a transition matrix $\boldsymbol{\Pi}$ with dimension 2^p.

3.1.2 Spatial occurrence process

The spatial dependence of the precipitation states $X_t(k)$ is modeled using an unobserved Gaussian stochastic process $\mathbf{U}_t = \{U_t(1), \dots, U_t(K)\}$, where K is the number of stations. Here, Gaussian random variables $U_t(k), k = 1, \dots, K$, are
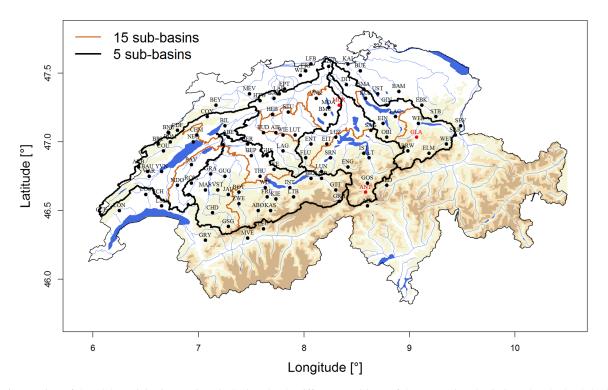
Figure 1. Location of the 105 precipitation stations in Switzerland. Different partitions of the Aare River basin into 5 and 15 sub-basins are shown.

temporally independent and $\mathbf{U}_t$ follows a multivariate normal distribution:

$$\mathbf{U}_t \sim N(0, \mathbf{\Omega}_X), \qquad (4)$$

where $\mathbf{\Omega}_X = \{\omega_{kl}\}$ is a positive-definite correlation matrix. At any location k, the precipitation state $X_t(k)$ is assumed to be completely determined by $U_t(k)$ and the previous p states at the same location. Specifically, if $X_{t-1}(k) = i_1$, ..., $X_{t-p}(k) = i_p$, and $p_1 = \Pr\{X_t(k) = 1 | X_{t-1}(k) = i_1, ..., X_{t-p}(k) = i_p\}$, then

$$X_t(k) = \begin{cases} 1, & \text{if } U_t(k) \leq \Phi^{-1}(p_1). \\ 0, & \text{otherwise,} \end{cases} \qquad (5)$$

where $\Phi[.]$ indicates the standard Gaussian cumulative distribution function.

Let $\rho_{kl} = \text{Corr}(X_t(k), X_t(l))$ denote the inter-site correlation between the states $X_t(k)$ and $X_t(l)$. Following Srikanthan and Pegram (2009), ρ_{kl} can be expressed as

$$\rho_{kl} = \frac{\pi_{00}(k,l) - \pi_0(k)\pi_0(l)}{\sqrt{\pi_0(k)\pi_1(k)}\sqrt{\pi_0(l)\pi_1(l)}}, \qquad (6)$$

where $\pi_0(s) = \Pr\{X_t(s) = 0\}$ and $\pi_1(s) = \Pr\{X_t(s) = 1\}$ denote the probabilities of having dry and wet states at location s, respectively, and $\pi_{00}(k,l) = \Pr\{X_t(k) = 0, X_t(l) = 0\}$ denotes the joint probability of having dry states at both locations k and l.
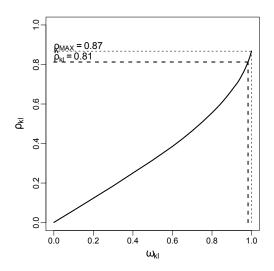

Figure 2. Illustration of the relationship between ω_{kl} and ρ_{kl} for the month of July and for stations GOS and ANT. A Markov chain of order 4 is considered in this example. The correlation between the observed states is $\hat{\rho}_{kl} = 0.81$ and can be reproduced using a bivariate Gaussian distribution with a correlation parameter of $\omega_{kl} = 0.98$. The maximum correlation ρ which can be obtained if $\omega_{kl} = 1$ is $\rho_{\text{MAX}} = 0.87$.

The relationship between ω_{kl} and ρ_{kl} is not direct since the temporal persistence of dry and wet events introduced at each station with a Markov chain also influences ρ_{kl} (Wilks,

1998). Figure 2 illustrates this relationship, obtained for the month of July via Monte Carlo simulations, for two close stations, GOS and ANT. In a first step, transition probabilities with a Markov chain of order 4 are estimated for these two stations. Given these transition probabilities, stochastic simulations of occurrence are then generated for different values of ω_{kl}, leading to different values of ρ_{kl}. Since this relationship is monotonic (see Fig. 2), it can be used to identify the value ω_{kl} leading to a specific $\hat{\rho}_{kl}$, namely the empirical value obtained from the observed time series of occurrence. The estimate of ω_{kl} is found by iterating until the evaluation of the correlation between the simulated precipitation states, ρ_{kl}, matches $\hat{\rho}_{kl}$. Note that a very high value for $\hat{\rho}_{kl}$ cannot always be reached, even if $\omega_{kl} = 1$. This is, however, a situation which rarely occurs in practice.

3.2 Precipitation intensity process

Given the occurrence of precipitation $X_t(k)$ at different locations k, GWEX models generate the amounts of precipitation $Y_t(k)$ using

- marginal heavy-tailed distributions,

- a tail-dependent spatial distribution,

- an autocorrelated temporal process.

3.2.1 Marginal distributions

At a given location k, daily precipitation has often been modeled by light-tailed distributions: exponential and Weibull distributions (Bárdossy and Pegram, 2009); gamma distributions (Srikanthan and Pegram, 2009; Mezghani and Hingray, 2009); a mixture of exponential distributions (Wilks, 1998; Keller et al., 2015); and a mixture of gamma distributions (Chen et al., 2014). However, as shown by many recent studies on a very large number of daily precipitation series (Papalexiou et al., 2013; Serinaldi and Kilsby, 2014a; Cavanaugh et al., 2015), exponentially decaying tails often result in a severe underestimation of extreme event probabilities. The introduction of a heavy-tailed distribution is thus crucial for the reproduction of the most extreme precipitation events (Hundecha et al., 2009).

In this work, the distribution representing the precipitation intensity at each location, $Y_t(k)$, is the E-GPD distribution. This distribution was first proposed by Papastathopoulos and Tawn (2013), who referred to it as an extended GP-Type III distribution, and it has since been shown to adequately model the whole range of precipitation intensities (Naveau et al., 2016). Compared to other heavy-tailed distributions applied to daily precipitation amounts (e.g., mixtures of GPD and gamma distribution; see Vrac and Naveau, 2007), the E-GPD is parsimonious and provides a very good compromise between flexibility and stability, which is an essential feature for extrapolation.

This distribution can be described by a smooth transition between a gamma-like distribution and a heavy-tailed Generalized Pareto distribution (GPD). This transition is obtained via a transformation function, $G(\nu)$, such that the whole range of precipitation intensities is modeled without a threshold selection (Naveau et al., 2016):

$$F_Y\{Y_t(k)\} = G\Big[H_\xi\big\{Y_t(k)/\sigma\big\}\Big], \tag{7}$$

where

$$H_\xi(z) = \begin{cases} 1 - (1 + \xi z)_+^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - e^{-z} & \text{if } \xi = 0, \end{cases} \tag{8}$$

where $a_+ = \max(a, 0)$ is the standard cumulative distribution function of the GPD, $\sigma > 0$ is a scale parameter, and $G(\nu) = \nu^\kappa, \kappa > 0$. Thus, a three-parameter set $\{\sigma, \kappa, \xi\}$ needs to be estimated at each station.

3.2.2 Spatial and temporal dependence of precipitation amounts

Spatial and temporal dependence of precipitation amounts is represented using a multivariate autoregressive model of order 1 (MAR(1)). A MAR(1) process has been used by different authors (Bárdossy and Pegram, 2009; Rasmussen, 2013) to simultaneously represent spatial and temporal dependences. Let $\mathbf{Z}_t$ denote a vector of K Gaussian random variables with mean 0 defined as

$$Z_t(k) = \Phi^{-1}\big[F_Y\{Y_t(k)\}\big]. \tag{9}$$

The stochastic Gaussian process $\mathbf{Z}_t$ is assumed to follow a MAR(1) process defined as follows:

$$\mathbf{Z}_t = \mathbf{A}\mathbf{Z}_{t-1} + \boldsymbol{\epsilon}_t, \tag{10}$$

where $\mathbf{A}$ is a $K \times K$ matrix and $\boldsymbol{\epsilon}_t$ is an innovation term described by a random $K \times 1$ noise vector. The elements of $\boldsymbol{\epsilon}_t$ have zero means and are independent of the elements of $\mathbf{Z}_{t-1}$. The covariance matrix of $\boldsymbol{\epsilon}_t$ is denoted by $\boldsymbol{\Omega}_Z$. Following Bárdossy and Pegram (2009), $\mathbf{A}$ is taken to be a diagonal matrix with diagonal elements that are the lag-1 serial correlation coefficients of the intensity process $Y_t(k)$. The matrix $\boldsymbol{\Omega}_Z$ can be expressed as

$$\boldsymbol{\Omega}_Z = \mathbf{M}_0 - \mathbf{A}\mathbf{M}_0'\mathbf{A}, \tag{11}$$

where $\mathbf{M}_0$ is the covariance matrix of $\mathbf{Z}_t$, which indicates the degree of spatial dependence between each pair of stations, and $\mathbf{M}_0'$ is its transpose.

Innovations $\boldsymbol{\epsilon}_t$ are often assumed to follow a standard multivariate normal distribution. However, the upper tail dependence of the multivariate normal distribution is 0, which means that extreme precipitation amounts simulated at the different sites are not spatially dependent. To introduce a tail dependence between at-site extremes, a possibility is to use

a Student copula to represent the dependence structure of ϵ_t, providing an additional parameter, ν, related to the tail dependence. Both dependence structures will be considered in the following.

3.3 Parameter estimation

3.3.1 Occurrence process

Following Wilks (1998), parameters related to the occurrence process $X_t(k)$ are estimated using the method of moments, i.e., using the empirical counterparts of the parameters. Observed states are first obtained using a low precipitation threshold (e.g., 0.2 mm). The matrix $\mathbf{\Pi}$ of transition probabilities is then estimated directly by the proportion of wet days $X_t(k) = 1$ following observed sequences $\{X_{t-1}(k), \ldots, X_{t-p}(k)\}$. Concerning the spatial occurrence process, $\hat{\rho}_{kl}$ estimates are obtained using the empirical counterparts of π_{00}, π_0, and π_1 (see Eq. 6), which correspond respectively to the proportion of days for which dry states are observed simultaneously at two locations ($\hat{\pi}_{00}$) and to the proportions of dry days $\hat{\pi}_0$ and wet days $\hat{\pi}_1$. The correlation matrix $\hat{\mathbf{\Omega}}_X$ is then composed of the cross-correlations $\hat{\omega}_{kl}$ obtained for all possible pairs of stations. If $\hat{\mathbf{\Omega}}_X$ is not positive-definite, the closest positive-definite matrix is considered (Rousseeuw and Molenberghs, 1993; Rebonato and Jaeckel, 2011). Furthermore, the seasonality of the occurrence process is taken into account by estimating these parameters on a monthly basis.

3.3.2 Intensity process

E-GPD distributions are first fitted to precipitation amounts available at each location k. As local estimations of the GPD tail exhibit a lack of robustness, we propose estimating the ξ parameter of the E-GPD (see Eq. 8) using a regionalization method similar to that of Evin et al. (2016), which can be summarized as follows.

1. Following Burn (1990), for each station, a region-of-influence (RoI) is delimited by a circle around the site, the radius being determined using homogeneity tests. All the stations inside this RoI are then considered homogeneous up to a scale factor.

2. The ξ parameters are then estimated with the maximum likelihood method using the precipitation observations from all the stations inside the RoI.

This regionalization method is applied to the precipitation data available from 666 stations in Switzerland, for four different seasons:

- *winter:* December, January, and February;

- *spring:* March, April, and May;

- *summer:* June, July, and August;

- *autumn:* September, October, and November.

In this work, the estimation of the ξ parameter is bounded below by 0. When $\xi < 0$, the E-GPD distribution has an upper bound. As shown by many recent studies (e.g., Serinaldi and Kilsby, 2014a), negative estimates of ξ are usually due to parameter uncertainty and are not realistic. The two remaining parameters of the E-GPD, the scale parameter σ and the parameter of the transformation κ, are estimated from the observations available at that station. Here, we use a method of moments based on probability weighted moments (see Naveau et al., 2016, for further details).

Concerning the spatial and temporal dependence of precipitation amounts, direct estimates of $\mathbf{M}_0$ and $\mathbf{A}$ cannot be obtained since non-zero precipitation amounts $Y_t(k)$ are not observed. Here, we follow the methodology proposed by Wilks (1998) and Keller et al. (2015). For each pair of stations, we generate long sequences of precipitation amounts $P_t(k)$ using the estimated parameters of the occurrence process ($\hat{\mathbf{\Pi}}$ and $\hat{\omega}_{kl}$), the parameters of the marginal distributions, and a correlation coefficient $m_0(k, l)$, indicating the degree of spatial dependence. Similarly to the occurrence process, $\hat{m}_0(k, l)$ is then found iteratively by matching the correlation between these long random streams with the observed correlation $\text{Corr}(P_t(k), P_t(l))$ (see Wilks, 1998; Keller et al., 2015, for further details). The correlation matrix $\hat{\mathbf{M}}_0$ is then composed of the cross-correlations $\hat{m}_0(k, l)$ obtained for all possible pairs of stations. For each station, the estimates of the lag-1 serial correlation coefficients of the matrix $\mathbf{A}$ are obtained using the same simulation approach.

The matrix $\hat{\mathbf{\Omega}}_Z$, i.e., the estimate of the covariance matrix of the innovations ϵ_t, is then obtained using Eq. (11). Since $\hat{\mathbf{\Omega}}_Z$ is not necessarily positive-definite (see Eq. 11), the closest positive-definite matrix is taken as the covariance matrix of ϵ_t if necessary. Given $\hat{\mathbf{\Omega}}_Z$, the parameter ν is estimated by maximizing the likelihood, as described in McNeil et al. (2005, Sect. 5.5.3.).

Similarly to the occurrence process, the seasonal aspect of the precipitation intensity is taken into account by performing the parameter estimation for each month, on a 3-month moving window.

3.4 Model versions

Different versions of the proposed multi-site precipitation model are considered in this paper, each corresponding to different extensions of the Wilks model. A flowchart summarizing the increasing complexity of these models is presented in Fig. 3.

3.4.1 Wilks

A first benchmark version of the multi-site model, referred to here as "Wilks", is considered. It closely matches the multi-site model proposed by Wilks (1998), detailed in particular as follows.
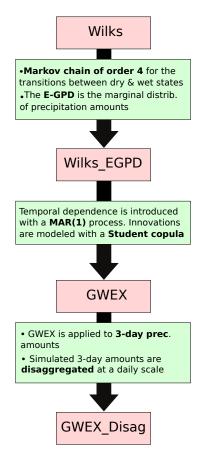
Figure 3. Flowchart of the different model versions. The differences between the models are summarized inside green boxes.

- The at-site occurrence process is a Markov chain of order 1.

- The marginal distribution on precipitation amounts is a mixture of exponential distribution, for which the probability density function is defined as

$$f(x) = \frac{w}{\beta_1} \exp\left(-\frac{x}{\beta_1}\right) + \frac{1-w}{\beta_2} \exp\left(-\frac{x}{\beta_2}\right). \qquad (12)$$

  The parameters w, β_1 and β_2 are estimated using the expectation-maximization (EM) method (Dempster et al., 1977).

- Precipitation amounts are not considered to be temporally correlated; i.e., the matrix $\mathbf{A}$ in Eq. (10) is a zero matrix. Furthermore, innovations ϵ_t follow a standard multivariate normal distribution and represent the spatial correlations.

3.4.2 Wilks_EGPD

A modified Wilks version is considered, for which the at-site occurrence process is a Markov chain of order 4 and

the mixture of exponential distributions is replaced by the E-GPD distribution. As indicated above, Srikanthan and Pegram (2009) show that a fourth-order Markov chain improves the reproduction of dry/wet period lengths. This direct extension of the Wilks model is used to illustrate the impact of using a Markov chain of order 4 compared to order 1. Differences in performance between a heavy-tailed distribution (E-GPD) and a low-tailed distribution (mixture of exponentials) will be highlighted.

3.4.3 GWEX

The initial GWEX model has the following characteristics.

- The at-site occurrence process is a Markov chain of order 4.

- The marginal distribution for precipitation amounts is the E-GPD distribution.

- Precipitation amounts follow a MAR(1) process with innovations modeled by a Student copula.

3.4.4 GWEX_Disag

In this paper, an alternative version, referred to as GWEX_Disag, is also proposed. GWEX_Disag is applied to 3-day precipitation amounts and has the same characteristics as GWEX, except the following.

- The at-site occurrence process is a Markov chain of order 1.

- A threshold of 0.5 mm separates dry and wet states.

With GWEX_Disag, daily scenarios are first generated at a 3-day scale and then disaggregated at a daily scale using a method of fragments (e.g., Wójcik and Buishand, 2003). Simulated 3-day amounts are disaggregated using the temporal structures of the closest observed 3-day amounts, in terms of similarity of the spatial fields. The same observed 3-day sequence is thus used to disaggregate the 3-day amounts simulated at the 105 stations, which ensures the spatial coherence of these disaggregated amounts. Details of the disaggregation method are provided in Appendix A. Compared to GWEX, GWEX_Disag offers the following advantages.

- The 3-day precipitation amounts are directly modeled and have a better chance of being adequately reproduced.

- The disaggregation of 3-day precipitation amounts creates an inherent link between the occurrence and the intensity processes. For very extreme precipitation events, we can expect these processes to be dependent (higher chance of being in a wet state over the whole Aare River basin, as well as large and persistent precipitation amounts).

4 Multi-scale evaluation

The proposed stochastic models intend to preserve the most critical properties of precipitation at different spatial and temporal scales, especially extreme precipitation amounts. For hydrological applications, it can be assumed that a precipitation model preserving these properties has a better chance of adequately reproducing flood properties for small sub-basins as well as for large basins. This statement is supported by empirical evidence provided by Froidevaux (2014) and Froidevaux et al. (2015) for our study area (i.e., Switzerland). Using 60 years of gridded precipitation data, Froidevaux et al. (2015) show that, in Switzerland, high discharge events are usually triggered by meteorological events with a duration of several days, in late summer and autumn. Typically, the 2-day precipitation sum before floods is most correlated with flood frequency and flood magnitude.

The performance of the different multi-site precipitation models is thus assessed for multiple spatial and temporal scales. We investigate whether or not the statistical properties of precipitation data are adequately reproduced at the scale of the stations and for different partitions of the Aare River basin (see Fig. 1). In order to achieve this, 100 daily precipitation scenarios are generated, each scenario having a length of 100 years.

For the different evaluated statistics, performance is categorized according to the comprehensive and systematic evaluation (CASE) framework proposed by Bennett et al. (2017). The CASE framework enables a systematic comparison of stochastic models and offers a consistent way of computing the performance metrics, which is important in order to obtain a fair assessment of the strengths/weaknesses of the different model versions. This approach consists in assigning one of three categories: "good", "fair", and "poor" performance, to each metric, according to the agreement between the observed metric and the simulated metrics computed from the 100 scenarios. Table 1 summarizes the tests leading to each performance category. A good performance is obtained when the observed metric is inside the 90 % probability limits of the 100 simulated metrics (case 1). It indicates that simulated metrics are in good agreement with the observed metric. However, an observed metric can obviously lie outside these limits without necessarily indicating a failure of the model. In this case, fair performance may be assigned if either of the following two rules is satisfied.

1. *Case 2.* The observed metric is outside the 90 % probability limits but within 3 standard deviations (SD) of the simulated mean, which corresponds to the 99.7 % probability limits if we assume that the uncertainty in the statistics is normally distributed. This case covers the situation where we could expect that the observed metric is outside the 90 % limits due to sampling uncertainty.

Table 1. Performance categorization criteria from Bennett et al. (2017).

Performance classification	Key	Test
Good	🟩	Observed metric inside 90 % limits (case 1)
Fair	🟨	Observed metric outside 90 % limits but within the 99.7 % limits (case 2) OR absolute relative difference between the observed metric and the average simulated metrics is 5 % or less (case 3)
Poor	🟥	Otherwise (case 4)

Table 2. Hydrological regimes and characteristics of extreme floods in Switzerland (Froidevaux, 2014).

	Mean elevation (m)	Season	Triggering events
Glacial	> 1900	summer	showers and snow melt
Nival	1200–1900	summer, spring	showers, long rain
Pluvial	< 1200	summer	long rain

2. *Case 3.* The absolute relative difference $|(S_{\mathrm{obs}} - \overline{\mathbf{S}}_{\mathrm{sim}})/S_{\mathrm{obs}}|$ between the observed metric S_{obs} and the mean of the simulated metrics $\overline{\mathbf{S}}_{\mathrm{sim}}$ is 5 % or less. If the variability of the simulated metrics is very small, it can happen that the observed metrics lie outside the 99.7 % limits without being too far from the simulated mean in terms of relative difference.

Otherwise, we consider that performance is poor, indicating that the model fails to reproduce this particular statistical properly.

In summary, good performance represents cases for which the observed metric is clearly well reproduced by the model, whereas fair performance indicates a reasonable match between the observed and the simulated metrics. The number of metrics for which poor performance is obtained is thus the first criterion indicating the overall performance of a model.

For illustration purposes, we also present the results of the evaluation for three precipitation stations corresponding to different hydrological regimes (see Table 2). Figure 1 shows the 3 (out of 105) selected precipitation stations. Station ANT (at Andermatt) is located in a glacial basin, station GLA (at Glarus) in a nival basin, and station MUR (at Muri) in a pluvial basin.

5 Results

This section presents the results of the multi-scale evaluation framework (see Sect. 4) for several metrics related to the occurrence process of the precipitation events, daily amounts, and precipitation extremes. Summary assessments are provided, with several statistics provided for all the spatial scales of interest.

The precipitation observations are split into two sets. (1) A total of 45 years randomly chosen among the period 1930–2014 are used to estimate the parameters, and (2) the 40 remaining years are used to evaluate the performance of the models. This separation between an estimation set and a validation set is crucial to test the ability of the model to adequately represent the statistical properties of events which have not been used during the fitting procedure. In this study, the multi-scale evaluation is only applied to the 40-year validation set.

5.1 Parameter estimation and generation of scenarios

The different model parameters are estimated with the 45-year estimation set of observations, following the methodology described in Sect. 3.3, except for the ξ parameter of the E-GPD, which is estimated using all available precipitation data in Switzerland. This approach ensures that robust estimates are obtained for this parameter, which is crucial in our context since extreme simulated precipitation amounts are highly sensitive to the ξ parameter.

For GWEX, the estimation of the ξ parameter is performed at a daily scale. In order to highlight spatial patterns of ξ over Switzerland, we show the maps of the interpolated parameter estimates in Fig. 4. Fat tails are obtained in the southern and eastern parts of the Aare River basin, particularly during spring and summer seasons. In the south of Switzerland, a region with high estimates ($\xi \sim 0.2$), highlighted in red, is obtained for the summer and autumn seasons. These high ξ estimates are consistent with the presence of strong convective storms in this mountainous region during this period of the year (Rudolph and Friedrich, 2012).

For GWEX_Disag, the regionalization method is applied at a 3-day scale (see Fig. 5). The resulting estimates are similar to the ones obtained at a daily scale. However, note that the very high estimates obtained during the summer season at a daily scale are lower at a 3-day scale. This seems to confirm the interpretation of these high ξ estimates; i.e., the relationship between summer convective storms and high ξ estimates is not as strong at a 3-day scale since storms of this type usually have a shorter duration. Note that non-zero ξ estimates in Figs. 4 and 5 (in green, yellow, and red) indicate that low-tailed distributions lead to an underestimation of extreme precipitation in these regions.

Figure 6 compares empirical and fitted distributions (mixture of exponentials and E-GPD) at a daily scale, for three illustrative stations and for the months of January, April, July, and October. Both distributions fit the observed precipitation amounts reasonably well. Concerning the highest precipitation intensities, it is hard to draw conclusions on a significant over- or underestimation. Indeed, local assessments of precipitation extremes are often inconclusive due to insufficient information on the distribution tails (Papalexiou and Koutsoyiannis, 2013).

For each multi-site precipitation model investigated in this paper (Wilks, Wilks_EGPD, GWEX and GWEX_Disag), we generate 100 daily precipitation scenarios with these parameter estimates, each scenario having a length of 100 years. These scenarios are compared to the precipitation observed for the 40-year validation period.

5.2 Occurrence process

The monthly number of wet days obtained from observed and simulated precipitation data are compared in Fig. 7. The average number of wet days is adequately reproduced by all models, with approximately 30 % of cases with poor performance. These poor performance cases seem to occur mainly during the winter and spring seasons. The SD of the monthly number of wet days indicates the inter-annual variability of this metric. While the magnitudes of the SD from the simulated precipitation roughly match the corresponding observed SD, it seems that the highest observed variabilities are underestimated by all the models, most markedly by the Wilks model.

Figures 8 and 9 show the distributions of observed and simulated dry and wet spells, respectively, for the three illustrative stations. Concerning the distributions of dry spell lengths, the Wilks_EGPD, GWEX, and GWEX_Disag models lead to adequate performance, the performance being classified as good in 48, 48, and 49 % of the cases, respectively. The performance of the Wilks model is slightly lower because of an imprecise reproduction of the frequency of the shortest dry spells. This difference in performance is explained by the order of the Markov chain used to simulate the transitions between dry and wet states, which is the only difference between the occurrence processes of Wilks and Wilks_EGPD or GWEX. The fourth-order Markov chain of the Wilks_EGPD and GWEX models seems to provide a more adequate representation of these transitions than the first-order Markov chain of the Wilks model, confirming previous findings (Srikanthan and Pegram, 2009).

The frequencies of wet spell lengths are adequately reproduced by the Wilks, Wilks_EGPD, and GWEX models, with more than 50 % of good performance. The lower overall performance of GWEX_Disag for this metric is due to a slight underestimation of the longest wet spells for some stations (which is however not the case for the stations shown in Fig. 9).
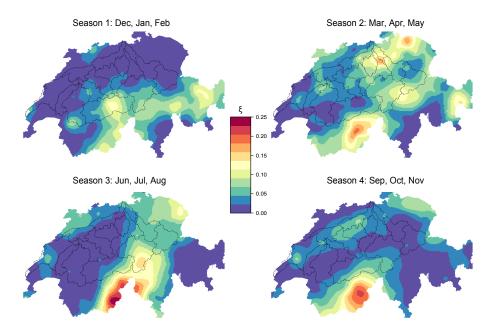
Figure 4. Regionalized ξ parameters at a daily scale, for the different seasons. Here, we present the spatial interpolation of at-site estimates for a better readability of their variability.
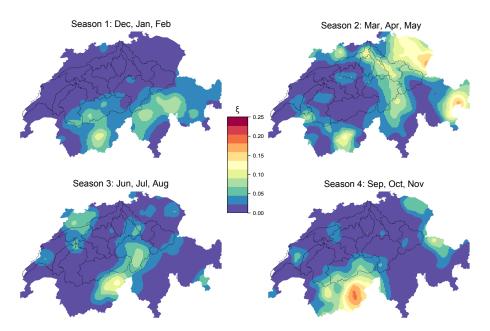

Figure 5. Regionalized ξ parameters at a 3-day scale, for the different seasons. Here, we present the spatial interpolation of at-site estimates for a better readability of their variability.

5.3 Inter-site correlations of precipitation amounts

Figure 10 compares observed and simulated inter-site correlations for the different model versions. Unlagged cross-correlations, which represent the spatial dependence, are close to the 1 : 1 diagonal line, as expected given that these correlations are explicitly taken into account by all model versions. However, a slight underestimation can be observed,

especially concerning correlations above 0.8. This underestimation is a side effect of the transformation applied to obtain a positive-definite matrix (see Sect. 3.3).

An adequate reproduction of lag-1 inter-site correlations is important for the reproduction of persistent precipitation events. Simulated lag-1 cross-correlations are close to 0 for the Wilks and Wilks_EGPD models, as expected given that these versions ignore the temporal dependence. Con-
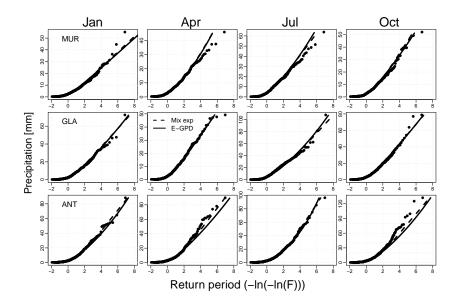
Figure 6. Empirical and fitted distributions (dashed curves for mixture of exponentials and solid curves for E-GPD) at a daily scale, for the three illustrative stations and for the months of January, April, July, and October.
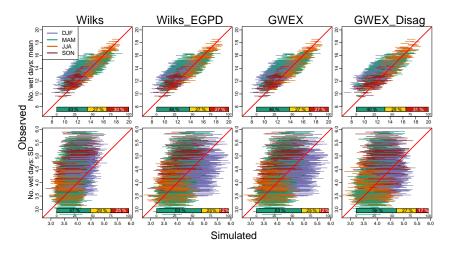

Figure 7. At-site number of wet days for all sites and months: inter-annual mean and standard deviation (SD). The 90 % probability limits are shown for the different seasons. Overall performance is represented by the indicated percentages of good, fair, and poor performance for all sites and months ($105 \times 12 = 1260$ cases).

sequently, these two model versions significantly underestimate observed lag-1 cross-correlations, which range between 0 and 0.4. Concerning GWEX, lag-1 serial autocorrelations at the stations (black points in the bottom plots) are perfectly aligned along the 1 : 1 line, as expected given that they are explicitly fitted by the MAR(1) process. Simulated and observed lag-1 cross-correlations are roughly in agreement, though the largest observed cross-correlations are underestimated. This is also the case, to a lesser extent, for GWEX_Disag. However, the agreement between observed and simulated cross-correlations is much stronger.

5.4 Daily amounts

The reproduction of precipitation amounts at a daily scale is assessed in Fig. 11, for all spatial scales and months. For all models, we obtain a reasonable agreement between observed and simulated average daily amounts (90 % limits close to the 1 : 1 line), with more than 40 % of good cases and less than 30 % of poor cases. The SD of these daily amounts are also adequately reproduced (Fig. 11, bottom plots).

5.5 Extreme precipitation amounts

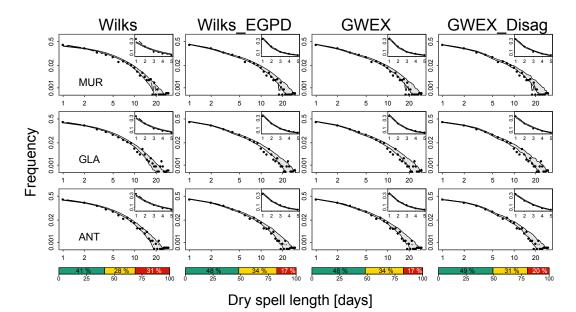Figures 12 and 13 show the relative differences, expressed as a percentage, between observed and simulated 10- and

Figure 8. Distribution of dry spell lengths at the stations: the 90 % probability limits are shown. Overall performance is represented by the indicated percentages of good, fair, and poor performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.
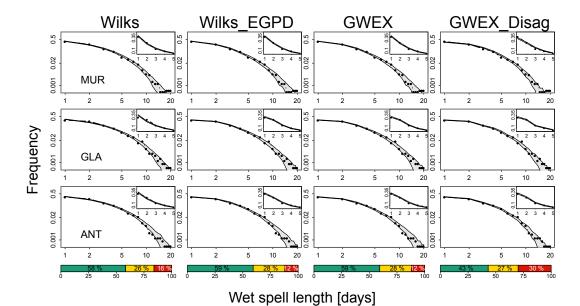

Figure 9. Distribution of wet spell lengths at the stations: the 90 % probability limits are shown. Overall performance is represented by the indicated percentages of good, fair, and poor performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.

50-year return periods, at daily and 3-day scales, respectively, for all spatial scales. The percentiles corresponding to these return periods are estimated empirically using the Gringorten formula (Gringorten, 1963). These figures provide an overview of model performance regarding extreme precipitation amounts.

At the daily scale (Fig. 12), there is no major difference in performance between the four models. For the 10-year and 50-year return periods, the number of poor perfor-

mance cases is below 20 % for all models. The relative differences are globally centered around zero, which means that the mixture of exponentials (Wilks model) and the E-GPD (Wilks_EGPD, GWEX and GWEX_Disag models) all produce a reasonable performance at this temporal scale. However, if we compare the 50-year return periods simulated by the Wilks and Wilks_EGPD models, we note an increase of 10 % of good performance cases (from 65 to 75 %), which
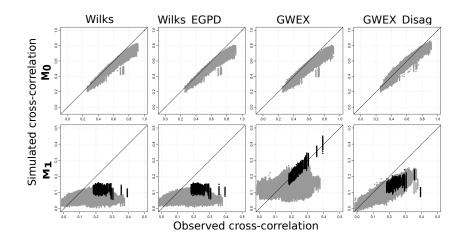
Figure 10. Comparison of unlagged inter-site correlations (M_0) and lag-1 inter-site correlations (M_1) in observed and simulated precipitation series, for the winter (DJF) and summer (JJA) seasons and for the different model versions considered. Black points indicate lag-1 serial autocorrelations at the stations.
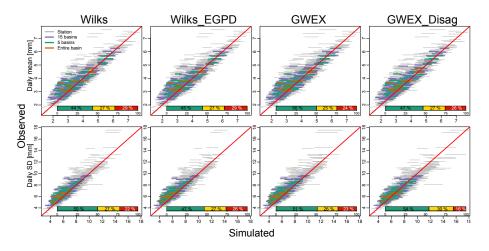

Figure 11. Daily amounts for all spatial scales and months: inter-annual mean (top) and standard deviation (SD, bottom). The 90 % probability limits are shown. Overall performance is represented by the indicated percentages of good, fair, and poor performance for all spatial scales and months.

can be explained by a slight underestimation of the largest maxima with Wilks, for some stations.

Comparing Wilks_EGPD and GWEX, the scores are almost identical, which suggests that the tail dependence introduced by the Student copula in GWEX does not produce a significant improvement for the reproduction of extremes. However, if we focus on the largest spatial scales (at the basins), and in particular on the entire Aare River basin (orange lines), it seems that the slight underestimation of the 50-year return periods obtained with Wilks_EGPD is reduced thanks to this tail dependence. GWEX_Disag also reproduces the largest precipitation amounts at all spatial scales adequately, even if a slight overestimation of the maxima at the largest spatial scales can be suspected. Nevertheless, this performance shows that the disaggregation process leads to an adequate reproduction of the daily maxima.

At the 3-day scale (Fig. 13), the underestimation of the maxima by Wilks and Wilks_EGPD is clear at all spatial scales. GWEX does not suffer from the same shortcomings, which means that the MAR(1) process (Eq. 10) improves the temporal structure of the largest 3-day precipitation amounts. As GWEX_Disag is fitted at a 3-day scale, this model logically leads to an adequate reproduction of extreme 3-day precipitation amounts. The strategy consisting of simulating 3-day precipitation amounts, which are then disaggregated at a daily scale, presents several advantages.

– As the model is fitted at a 3-day scale, 3-day maxima are adequately reproduced.

– As the method of fragments uses observed 3-day temporal structures to disaggregate 3-day amounts, the daily amounts resulting from a generated 3-day maxima are
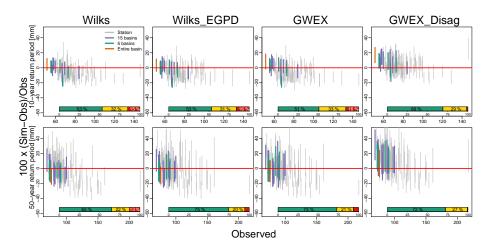
Figure 12. Daily annual maxima for all spatial scales: relative differences, expressed as a percentage, between observed and simulated 10-year (top plots) and 50-year (bottom plots) return periods. The 90 % probability limits are shown. Overall performance is represented by the indicated percentages of good, fair, and poor performance for all spatial scales.
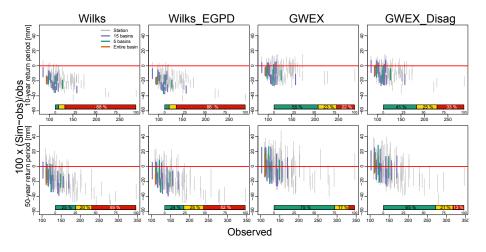

Figure 13. The 3-day annual maxima for all spatial scales: relative differences, expressed as a percentage, between observed and simulated 10-year (top plots) and 50-year (bottom plots) return periods. The 90 % probability limits are shown. Overall performance is represented by the indicated percentages of good, fair, and poor performance for all spatial scales.

physically plausible. In particular, the temporal and spatial structures of large and persistent observed precipitation events are used, which ensures consistency between the generated extreme events at the daily and 3-day scales.

GWEX and GWEX_Disag both adequately reproduce extreme precipitation amounts at daily and 3-day scales, as well as at all spatial scales. As indicated above, these models will be used to generate long precipitation scenarios, which will feed a hydrological model in order to produce flood scenarios. Ultimately, the reproduction of the flood properties using GWEX and GWEX_Disag will indicate which model is the most adequate. Since they correspond to the same model version fitted at daily and 3-day scales, respectively, we can

expect that resulting floods will have slightly different properties.

6 Conclusions and outlook

Precipitation models are usually developed for the purpose of risk assessment in relation to natural hazards (e.g., droughts, floods). Most existing precipitation models aim to reproduce a wide range of statistical properties of precipitation, at different scales, in order to be used as a general tool in different contexts. In this study, our main objective was to provide a precipitation generator that could be used together with a hydrological model for the evaluation of extreme flooding events in a region covering approximately half of Switzerland. As a consequence, we were especially interested in the

reproduction of extreme precipitation amounts at medium to large spatial scales. As the daily and 3-day precipitation amounts are a major determinant of flood magnitudes in large Swiss basins (Froidevaux et al., 2015), an adequate reproduction of precipitation at these timescales was also required.

In this paper, we considered different multi-site precipitation models targeting the reproduction of extreme amounts at multiple temporal (daily, 3-day) and spatial scales. Different extended versions of the model introduced by Wilks (Wilks, 1998) have been proposed. A first direct extension, Wilks_EGPD, considers a Markov chain (of order 4 instead of order 1) for the at-site occurrence process. Furthermore, taking advantage of recent advances regarding extreme precipitation, a heavy-tailed distribution (instead of a mixture of exponential distributions), the E-GPD, is applied to the precipitation intensities at each station. Two important extensions of Wilks_EGPD, named GWEX and GWEX_Disag, are then considered. In the GWEX model, temporal and spatial dependencies of the occurrence and intensity process are introduced using the copula theory and a multivariate autoregressive process. A second version, GWEX_Disag, applies the same model, but at a 3-day scale. The 3-day simulated amounts are then disaggregated using an adaptation of the method of fragments (Wójcik and Buishand, 2003).

In this study, we support the use of a systematic evaluation framework. The CASE framework proposed by Bennett et al. (2017) provides a useful tool in this respect, making it possible to compare performance between precipitation models fairly. Regarding the reproduction of extreme precipitation, evaluations until now have usually been qualitative (e.g., interpretations based on one or two examples) and limited in terms of spatial scales (often only at the stations). The evaluation of extreme precipitation amounts proposed in this paper is multi-scale in time (daily and 3-day scale) and space (at the stations, for two different divisions of the study area into sub-basins, and for the entire Aare River basin).

The different multi-site precipitation models have been applied to 105 stations located in Switzerland. A multi-scale evaluation led to the following conclusions.

– A fourth-order Markov chain outperforms a first-order Markov chain for the transitions between dry and wet states, notably for the reproduction of dry spell lengths.

– At the scale of the stations, daily amounts (average, SD, and extremes) are reasonably well reproduced by all the models.

– With only three parameters, the E-GPD provides a parsimonious and flexible representation of the whole of precipitation amounts. Its GPD tail is in agreement with recent results, showing that extreme precipitation amounts must be modeled by heavy-tailed distributions (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014a). Furthermore, robust estimates of the parameter controlling the heaviness of the distribution tail are obtained using a regionalization method. In our study area, the E-GPD does not bring a significant improvement of the performance compared to the mixture of exponential distributions. However, the general framework proposed in this paper can be applied to very distinct precipitation regimes, and the possible heavy tail of the E-GPD might be valuable in other areas.

– At a 3-day scale, precipitation extremes are severely underestimated by Wilks and Wilks_EGPD. This underestimation can be explained by an incorrect representation of the persistence by these models.

– GWEX and GWEX_Disag adequately reproduce extreme precipitation amounts at daily and 3-day scales, and at all spatial scales. These models are deemed adequate for the evaluation of extreme flood events.

Future research will investigate whether the floods simulated by a hydrological model using the generated precipitation scenarios have statistical properties in agreement with observed floods. An extensive investigation is currently underway with a distributed version of the HBV hydrological model, applied to 87 sub-basins of the whole study area and using precipitation scenarios produced by GWEX as inputs. This hydrological evaluation of our weather scenarios will be presented in future publications.

Appendix A: Temporal disaggregation from a 3-day scale to a daily scale

For a 3-day period $\mathbf{D} = \{d, d+1, d+2\}$ starting on a day d, the observed and simulated precipitation amounts at a station k are denoted by $Y_{\mathbf{D}}(k)$ and $\widetilde{Y}_{\mathbf{D}}(k)$, respectively. We want to disaggregate the simulated 3-day amount for the period $\widetilde{\mathbf{D}} = \{\tilde{d}, \tilde{d}+1, \tilde{d}+2\}$. This disaggregation is achieved in the following steps.

1. A set of observed 3-day sequences are retained as candidate periods $\mathbf{D}$ according to two criteria.

    –   *Season.* Periods $\widetilde{\mathbf{D}}$ and $\mathbf{D}$ must belong to the same season, as defined in Sect. 3.3.
    –   *Mean intensity.* Simulated and observed precipitation fields must have the same order of magnitude. Let $q_{0.5}$, $q_{0.75}$, $q_{0.9}$, and $q_{0.99}$ denote the quantiles of the mean observed precipitation intensities over all the stations associated with probabilities 0.5, 0.75, 0.9, and 0.99, respectively. Observed and simulated 3-day periods are classified into five groups according to their mean intensity $\overline{Y} = \frac{1}{n}\sum_k Y_{\mathbf{D}}(k)$: dry periods ($\overline{Y} < q_{0.5}$), moderately wet periods ($q_{0.5} \le \overline{Y} < q_{0.75}$), wet periods ($q_{0.75} \le \overline{Y} < q_{0.9}$), very wet periods ($q_{0.9} \le \overline{Y} < q_{0.99}$), and extremely wet periods ($q_{0.99} \ge \overline{Y}$).

    This first selection of candidate periods aims to increase the chance of retaining periods corresponding to similar meteorological events.

2. For each observed 3-day candidate period $\mathbf{D}$, we compute the following score:

$$
\text{SCORE}(\widetilde{\mathbf{D}}, \mathbf{D}) = \sum_k \left| \frac{\widetilde{Y}_{\tilde{d}-1}(k)}{\sum_k \widetilde{Y}_{\tilde{d}-1}(k)} - \frac{Y_{d-1}(k)}{\sum_k Y_{d-1}(k)} \right|
$$
$$
+ \left| \frac{\widetilde{Y}_{\mathbf{D}}(k)}{\sum_k \widetilde{Y}_{\mathbf{D}}(k)} - \frac{Y_{\mathbf{D}}(k)}{\sum_k Y_{\mathbf{D}}(k)} \right|.
$$

This score measures the similarity between the simulated spatial field for the period $\widetilde{Y}_{\mathbf{D}}(k)$ and the observed spatial field for the period $\widetilde{\mathbf{D}}$ and also takes into account the similarity between the spatial fields for the previous days $\tilde{d}-1$ and $d-1$.

Absolute differences between relative precipitation intensities are computed (the lowest scores are therefore obtained for spatial fields with similar shapes) among the observed periods, corresponding to the same season and order of magnitude selected in the previous step.

3. For each simulated period $\widetilde{\mathbf{D}}$, the observed precipitation fields corresponding to the 10 lowest scores are retained. For each station k, if a positive precipitation amount has been simulated ($\widetilde{Y}_{\widetilde{\mathbf{D}}}(k) > 0$), we look at the corresponding observed amount $Y_{\mathbf{D}}(k)$. If $Y_{\mathbf{D}}(k) = 0$, this observed period cannot be used to disaggregate $\widetilde{Y}_{\widetilde{\mathbf{D}}}(k)$ and we look at the next best observed field among the 10 selected fields. If the observed field contains a positive precipitation amount at this station ($Y_{\mathbf{D}}(k) > 0$), then we obtain the simulated daily amount for day $\tilde{d}$ as follows:

$$
\widetilde{Y}_{\tilde{d}}(k) = Y_d(k) \times \frac{\widetilde{Y}_{\widetilde{\mathbf{D}}}(k)}{Y_{\mathbf{D}}(k)}, \tag{A1}
$$

with similar expressions for days $\tilde{d}+1$ and $\tilde{d}+2$. Simulated daily amounts correspond to the observed daily amounts, rescaled by the ratio between the simulated and observed 3-day amounts. The 3-day simulated amounts and observed temporal structures are thus preserved.

4. While the 3-day spatiotemporal consistency is generally conserved by applying the preceding steps, it can happen that the simulated 3-day amount is positive even though there is no positive precipitation among the 10 best 3-day observed fields. In this case, we seek similar observed amounts at this station only and randomly choose one 3-day period among the 10 best 3-day periods.

References

Ailliot, P., Allard, D., Monbet, V., and Naveau, P.: Stochastic weather generators: an overview of weather type models, Journal de la Société Française de Statistique, 156, 101–113, 2015.

Allard, D. and Bourotte, M.: Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process, Stoch. Environ. Res. Risk Assess., 29, 453–462, 2015.

Baxevani, A. and Lennartsson, J.: A spatiotemporal precipitation generator based on a censored latent Gaussian field, Water Resour. Res., 51, 4338–4358, 2015.

Bennett, B., Thyer, M., Leonard, M., Lambert, M., and Bates, B.: A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model, J. Hydrol., 556, 1123–1138, https://doi.org/10.1016/j.jhydrol.2016.12.043, 2018.

Buishand, T. A.: Extreme rainfall estimation by combining data from several sites, Hydrol. Sci. J., 36, 345–365, 1991.

Burn, D. H.: Evaluation of regional flood frequency analysis with a region of influence approach, Water Resour. Res., 26, 2257–2265, 1990.

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., and O'Connell, P. E.: RainSim: A spatial-temporal stochastic rainfall modelling system, Environ. Modell. Softw., 23, 1356–1369, 2008.

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

Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation, Hydrol. Earth Syst. Sci., 13, 2299–2314, https://doi.org/10.5194/hess-13-2299-2009, 2009.

Cavanaugh, N. R., Gershunov, A., Panorska, A. K., and Kozubowski, T. J.: The probability distribution of intense daily precipitation, Geophys. Res. Lett., 42, 1560–1567, doi:10.1002/2015GL063238, 2015.

Chandler, R. E. and Wheater, H. S.: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, Water Resour. Res., 38, 1192, doi:10.1029/2001WR000906, 2002.

Chen, J., Brissette, F. P., and Zhang, J. X.: A Multi-Site Stochastic Weather Generator for Daily Precipitation and Temperature, Trans. ASABE, 57, 1375–1391, 2014.

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, J. Roy. Stat. Soc.-B, 39, 1–38, 1977.

Evin, G., Blanchet, J., Paquet, E., Garavaglia, F., and Penot, D.: A regional model for extreme rainfall based on weather patterns subsampling, J. Hydrol., 541, 1185–1198, 2016.

Froidevaux, P.: Meteorological characterisation of floods in Switzerland, Ph.D. thesis, Geographisches Institut, University of Bern, 2014.

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O.: Flood triggering in Switzerland: the role of daily to monthly preceding precipitation, Hydrol. Earth Syst. Sci., 19, 3903–3924, https://doi.org/10.5194/hess-19-3903-2015, 2015.

Gringorten, I. I.: A plotting rule for extreme probability paper, J. Geophys. Res. , 68, 813–814, 1963.

Hundecha, Y., Pahlow, M., and Schumann, A.: Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes, Water Resour. Res., 45, W12412, doi:10.1029/2008WR007453, 2009.

Keller, D. E., Fischer, A. M., Frei, C., Liniger, M. A., Appenzeller, C., and Knutti, R.: Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment, Hydrol. Earth Syst. Sci., 19, 2163–2177, https://doi.org/10.5194/hess-19-2163-2015, 2015.

Kleiber, W., Katz, R. W., and Rajagopalan, B.: Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, Water Resour. Res., 48, W01523, doi:10.1029/2011WR011105, 2012.

Klemeš, V.: Tall Tales about Tails of Hydrological Distributions. I., J. Hydrol. Eng., 5, 227–231, 2000a.

Klemeš, V.: Tall Tales about Tails of Hydrological Distributions. II., J. Hydrol. Eng., 5, 232–239, 2000b.

Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L.: Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, Water Resour. Res., 50, 3317–3341, 2014.

Lamb, R., Faulkner, D., Wass, P., and Cameron, D.: Have applications of continuous rainfall-runoff simulation realized the vision for process-based flood frequency analysis?, Hydrol. Proc., 30, 2463–2481, 2016.

Leblois, E. and Creutin, J.-D.: Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method, Water Resour. Res., 49, 3375–3387, 2013.

Leonard, M., Lambert, M. F., Metcalfe, A. V., and Cowpertwait, P. S. P.: A space-time Neyman-Scott rainfall model

with defined storm extent, Water Resour. Res., 44, W09402, doi:10.1029/2007WR006110, 2008.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, doi:10.1029/2009RG000314, 2010.

McNeil, A. J., Frey, R., and Embrechts, P.: Quantitative Risk Management – Concepts, Techniques, and Tools, Princeton University Press, Princeton, N.J, 2005.

McRobie, F. H., Wang, L.-P., Onof, C., and Kenney, S.: A spatial-temporal rainfall generator for urban drainage design, Water Science and Technology: A J. Int. Assoc. Water Pollut. Res., 68, 240–249, 2013.

Mehrotra, R. and Sharma, A.: Preserving low-frequency variability in generated daily rainfall sequences, J. Hydrol., 345, 102–120, 2007a.

Mehrotra, R. and Sharma, A.: A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, J. Hydrol., 335, 180–193, 2007b.

Mehrotra, R. and Sharma, A.: Development and Application of a Multisite Rainfall Stochastic Downscaling Framework for Climate Change Impact Assessment, Water Resour. Res., 46, W07526, doi:10.1029/2009WR008423, 2010.

Mehrotra, R., Srikanthan, R., and Sharma, A.: A comparison of three stochastic multi-site precipitation occurrence generators, J. Hydrol., 331, 280–292, 2006.

Mezghani, A. and Hingray, B.: A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin, J. Hydrol., 377, 245–260, 2009.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A.: Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, Water Resour. Res., 52, 2753–2769, 2016.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resour. Res., 49, 187–201, 2013.

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, Hydrol. Earth Syst. Sci., 17, 851–862, https://doi.org/10.5194/hess-17-851-2013, 2013.

Papastathopoulos, I. and Tawn, J. A.: Extended generalised Pareto models for tail estimation, J. Stat. Plan. Infer., 143, 131–143, 2013.

Rasmussen, P. F.: Multisite precipitation generation using a latent autoregressive model, Water Resour. Res., 49, 1845–1857, 2013.

Rayner, D., Achberger, C., and Chen, D.: A multi-state weather generator for daily precipitation for the Torne River basin, northern Sweden/western Finland, Adv. Clim. Change Res., 7, 70–81, 2016.

Rebonato, R. and Jaeckel, P.: The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes, SSRN Scholarly Paper ID 1969689, Social Science Research Network, Rochester, NY, 2011.

Rousseeuw, P. J. and Molenberghs, G.: Transformation of non positive semidefinite correlation matrices, Communications in Statistics – Theory and Methods, 22, 965–984, 1993.

Rudolph, J. V. and Friedrich, K.: Seasonality of Vertical Structure in Radar-Observed Precipitation over Southern Switzerland, J. Hydrometeorol., 14, 318–330, 2012.

Serinaldi, F. and Kilsby, C. G.: Rainfall extremes: Toward reconciliation after the battle of distributions, Water Resour. Res., 50, 336–352, 2014a.

Serinaldi, F. and Kilsby, C. G.: Simulating daily rainfall fields over large areas for collective risk estimation, J. Hydrol., 512, 285–302, 2014b.

Sharif, M. and Burn, D. H.: Improved K-Nearest Neighbor Weather Generating Model, J. Hydrol. Eng., 12, 2007.

Srikanthan, R. and Pegram, G. G. S.: A nested multisite daily rainfall stochastic generation model, J. Hydrol., 371, 142–153, 2009.

Thompson, C. S., Thomson, P. J., and Zheng, X.: Fitting a multisite daily rainfall model to New Zealand data, J. Hydrol., 340, 25–39, 2007.

Vischel, T., Lebel, T., Massuel, S., and Cappelaere, B.: Conditional simulation schemes of rain fields and their application to rainfall-runoff modeling studies in the Sahel, J. Hydrol., 375, 273–286, 2009.

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, Water Resour. Res., 43, W07402, doi:10.1029/2006WR005308, 2007.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, J. Hydrol., 210, 178–191, 1998.

Wójcik, R. and Buishand, T.: Simulation of 6-hourly rainfall and temperature by two resampling schemes, J. Hydrol., 69–80, 2003.

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K.: A technique for generating regional climate scenarios using a nearest-neighbor algorithm, Water Resour. Res., 39, 1199, doi:10.1029/2002WR001769, 2003.