Implementation and validation of a multi-site daily precipitation generator over a Swiss river catchment

D. E. Keller[1, 2, 3], A. M. Fischer[1], C. Frei[1], M. A. Liniger[1, 3], C. Appenzeller[1, 3], R. Knutti[2, 3]

[1] {Federal Office of Meteorology and Climatology MeteoSwiss, Operation Center 1, 8085 Zürich-Flughafen, Switzerland}

[2] {Institute for Atmospheric and Climate Science, ETH Zürich, Universitaetstrasse 16, 8092 Zurich, Switzerland}

[3] {Center for Climate Systems Modeling (C2SM), ETH Zurich, Universitaetstrasse 16, 8092 Zurich, Switzerland}

Correspondence to: D. E. Keller (Denise.Keller@meteoswiss.ch)

ABSTRACT

Many climate impact assessments require high-resolution precipitation time-series that have a spatio-temporal correlation structure consistent with observations, for simulating either current or future climate conditions. In this respect, weather generators (WGs) designed and calibrated for multiple sites are an appealing statistical downscaling technique to stochastically simulate multiple realizations of possible future time-series consistent with the local precipitation characteristics and its expected future changes. In this study, we present the implementation and validation of a multi-site daily precipitation generator following ideas of Wilks (1998). The generator consists of several Richardson-type WGs run with spatially correlated random number streams. We investigate the applicability of the generator for the current climate by analysing systematic biases and stochastically generated variability and assess the added value of a multi-site generator compared to multiple single-site WGs. Results are presented for the Swiss hydrological catchment *Thur* in the Swiss Alpine region for current climate condition.

The calibrated multi-site WG is skilful at individual sites in representing the annual cycle of the precipitation statistics, such as mean wet day frequency and intensity as well as monthly

precipitation sums. It reproduces realistically the multi-day statistics such as the frequencies of dry and wet spell lengths and precipitation sums over consecutive wet days. Substantial added value is demonstrated in simulating daily areal precipitation sums in comparison to multiple WGs that lack the spatial dependency in the stochastic process. Limitations are seen in reproducing daily and multi-day extreme precipitation sums, observed variability from year to year and in reproducing long dry spell lengths. Given the performance of the presented generator, we conclude that it is a useful tool to generate precipitation series consistent with the mean aspects of the current and future climate.

1 Introduction

In Switzerland, precipitation is a key weather variable with high relevance for sectors such as energy production, infrastructure, tourism, agriculture and ecosystems. Owing to a complex topography, daily precipitation varies strongly in space and time (Frei and Schär, 1998; Isotta et al., 2013). The spatial distribution of daily precipitation frequency and intensity depends on the topography, with higher frequencies and intensities along the North-Alpine ridge during summer, and a strong north-south gradient with heavier intensities in southern Switzerland from spring to autumn. The most prominent weather situations causing these precipitation patterns are shallow pressure systems favouring convective precipitation, orographically induced precipitation (e.g. Föhn situations), and frontal passages. Precipitation amounts and frequencies are typically largest in summer, mainly due to convective processes (Frei and Schär, 1998).

Given the expected changes in the hydrological cycle over the 21st century (Allen and Ingram, 2002; Held and Soden, 2006), the need for reliable and quantitative future local precipitation projections in Switzerland is continuously growing. To effectively assess the impacts related to changes in precipitation, often highly localized daily data are needed that are ideally both consistent in time and in space (e.g. Köplin et al., 2010). Currently, in Switzerland various impact assessment reports rely on the statistically downscaled precipitation change data derived from regional climate models by the well-known and simple delta change approach, which shifts an observed time series by a model-derived change in the mean climate (BAFU, 2012; Bosshard et al., 2011; CH2014-Impacts, 2014). The delta change approach accounts for changes in the mean annual cycle, but potential changes in inter-annual variability, changes in wet-day frequency and intensity or of spell lengths are not taken into account. Hence, the data are also not suitable for the analysis of future changes in extreme events (Bosshard et al., 2011). It is our aim here to develop a statistical downscaling method for Switzerland that overcomes some of these limitations and that subsequently can be easy applied to climate model output.

Over recent years a vast number of statistical downscaling methods have been developed that go far beyond a simple delta change approach (Maraun et al., 2010). These include bias-correction methods (e.g. Themeßl et al., 2011), regression-based methods (e.g. Hertig and Jacobeit, 2013) or weather generator (WG) approaches (e.g. Chandler and Wheater, 2002; Mezghani and Hingray, 2009). For our purposes, the latter method is especially appealing, since it includes a stochastic component. This is a major improvement compared to a (deterministic) delta change approach, allowing to investigate multiple time-series and uncertainty at the local scale that are consistent with a given (current or future) mean climate. Moreover, it allows the incorporation of changes in the temporal correlation structure and consequently alterations of the dry-wet sequences. From an agricultural impact's perspective this is a key aspect of future precipitation change (e.g. Calanca, 2007).

A serious limitation of many WGs is that they are often calibrated to observations at single sites only, therefore lacking the spatial correlation structure that is required for many applications, particularly in the context of hydrological impact modelling in a topographically complex terrain such as the Alps. A number of sophisticated approaches in time-space precipitation simulation have been put forward in the literature to address this issue, such as copula based approaches (e.g. Bárdossy and Pegram, 2009), Hidden Markov models (e.g. Hughes et al., 1999) Poisson cluster models (e.g. Cowpertwait, 1995; Fatichi et al., 2011) or more sophisticated field generators (e.g. Paschalis et al., 2013). K-nearest neighbor resampling approaches represent a further possibility to ensure the spatial coherence (e.g. Buishand and Brandsma, 2001). Each of these time-space WGs come with method-specific benefits and limitations for the reproduction of the daily precipitation statistics and consequently its use in impact models. For instance, some of them do better in simulating more realistically longer-term variability (e.g. generalized linear model (GLM) based multi-site WGs, Chandler, 2014), while some are explicitly adapted to deal with extreme precipitation (e.g. Huser and Davison, 2014).

The main purpose of our precipitation generator is its use as a downscaling tool in a climate change context. It should be easily transferable to different climatological regions and time-periods and its generated time-series should serve several impact applications that have different needs in terms of time-space consistency. For these reasons we opt for a precipitation generator whose degree of complexity and associated calibration requirements are still sufficiently easy to handle. This is accomplished with the multi-site precipitation generator proposed by Wilks (1998) that is based on a Richardson-type WG (Richardson, 1981) run with spatially correlated random number streams.

In this study, we implement and validate this multi-site generator for the Swiss catchment *Thur* in the Swiss Alpine region under current climate conditions to document the specific challenges encountered during the setup. The Thur catchment serves as an ideal testbed with different precipitation characteristics mainly due to the complex topography. Understanding its capabilities and systematic biases in current climate is key to later interpret the climatic changes in the simulated time-series for a future climate. Of particular relevance is the actual amount of stochastically generated variability. A second goal of the study is to assess the added value of a multi-site model against multiple single-site models. To accurately quantify these aspects, we choose a rather long calibration period that minimizes the effect of sampling uncertainties.

The structure of this paper is as follows: Sect. 2 introduces the hydrological catchment of the river *Thur* together with the used station data. In Sect. 3 we first describe the statistical models for simulating precipitation occurrence and amount and show how these models are combined to multi-site simulation. The validity of our generated multi-site precipitation series and the comparison to single-site generators is presented in Sect. 4. Sect. 5 includes a discussion and finally, Sect. 6 provides a summary and an outlook.

2 Data

This study focuses on the hydrological catchment of the river *Thur*, which is located in the north-eastern part of Switzerland (Figure 1a). The river *Thur* is a feeder river of the Rhine with a length of about 135 km and a catchment area of approximately 1696 km^2. It represents the largest Swiss river without a natural or artificial reservoir and therefore exhibits discharge fluctuations similar to unregulated Alpine rivers. Its flow regime is nivo-pluvial that is heavily influenced by snowmelt (BAFU, 2007). Owing to the complex topography over this catchment area, precipitation exhibits a large variability both in space and in time. This is illustrated in Figure 1b based on gridded observational data from Frei and Schär (1998). Over 1961-2011 and for a winter and summer month, the data clearly show larger precipitation frequencies and intensities over higher-elevated regions compared to the lowlands.

Additionally, this catchment lies in one of the Swiss regions featuring well above-average precipitation. A large portion of these precipitation characteristics can be explained by a north-east to south-west lying mountain range (*Alpstein*) extracting precipitation from westerly flows and triggering convective storms.

For the purpose of this study, we selected eight evenly distributed measurement stations (Figure 1a) of MeteoSwiss that all provide homogenized time-series covering a 51-year period from 1961-2011 (Begert et al., 2003), and that sufficiently cover the elevation profile of the catchment area from *Andelfingen* lying at 382 meters a.s.l. to *Saentis* lying at 2502 meters a.s.l.

3 Method

The core of our multi-site WG is a Richardson-type precipitation generator (Richardson, 1981), that relies on the concept of modelling two processes at one single station: an occurrence and an amount process. Following Wilks (1998), this single-site WG is then extended in order to simultaneously generate precipitation at several sites taking into account the complex spatio-temporal correlation structure.

In the following, we explain the setup of our multi-site generator step by step: Sect. 3.1 and 3.2 present the concepts of statistically characterizing occurrence and amount at single-sites. The simulation procedure of new synthetic time-series is detailed in Sect. 3.3. In Sect. 3.4 we give a description of how we implemented the multi-site WG over the *Thur* catchment.

3.1 Precipitation occurrence process

To model occurrence at a single station we rely on a first-order two-state Markov chain (Gabriel and Neumann, 1962; Richardson, 1981). The first-order two-state Markov chain is a statistical model describing the probability to stay in the same state or switch to the other state. In this context, first-order implies the state at a given day depends only on the state at the previous day. The use of a first-order model in our WG was justified by inspecting the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978). Both the AIC and the BIC revealed a substantial improvement when going from a zero-order to a first-order model, but the additional gain at a second- or higher-order model was negligible (not shown). We used a specific wet-day threshold of 1 mm day^{-1} to discretize a given daily precipitation time-series $X(t)$ at a given site into the two states ‘dry’

($X(t)$ < 1 mm day^{-1}) and 'wet' ($X(t) \geq$ 1 mm day^{-1}) and to subsequently generate a binary series (i.e. J_t with $J_t = 0$ for a dry state and $J_t = 1$ for a wet state). Four transitions are possible: a dry day following a dry day (*00*), a wet day following a dry day (*01*), a dry day following a wet day (*10*) and a wet day following a wet day (*11*).

The first-order two-state Markov chain model can be specified by formulating the probabilities (p) of these state-transitions:

$$\begin{aligned} p_{11} &= P\{J_t = 1 \mid J_{t-1} = 1\} \\ p_{01} &= P\{J_t = 1 \mid J_{t-1} = 0\} \end{aligned} \tag{1}$$

The corresponding counterparts of transition probabilities (p_{00} and p_{10}) can then easily be derived, since the sum of two probabilities conditioned on the same state at the previous day equals one:

$$\begin{aligned} p_{11} + p_{10} &= 1 \\ p_{00} + p_{01} &= 1 \end{aligned} \tag{2}$$

The two transition probabilities of Eq. (1) suffice to fully specify the first-order two-state Markov chain model. For the remaining part of this study, we therefore concentrate on these two parameters when addressing state transitions. For an estimate we rely on their conditional relative frequencies (Wilks, 2011):

$$\begin{aligned} \hat{p}_{01} &= \frac{n_{01}}{n_{0\bullet}} \\ \hat{p}_{11} &= \frac{n_{11}}{n_{1\bullet}} \end{aligned} \tag{3}$$

where n_{01} and n_{11} are the number of transitions from dry to wet and wet to wet in the binary series and $n_{0.}$ and $n_{1.}$ are the total number of zero's and one's in the series followed by any of the two states. From the transition probabilities of Eq. (3) other important precipitation indices can be inferred. The wet day frequency (wdf, π) is defined as the ratio of the number of wet days to the total number of days over a given time period. It can be expressed in terms of the two transition probabilities (Wilks, 2011):

$$\pi = \frac{p_{01}}{1 + p_{01} - p_{11}} \tag{4}$$

Similarly, the lag -1- autocorrelation r_1 is defined as the difference between the transition probabilities (Wilks, 2011):

$$r_1 = p_{11} - p_{01} \tag{5}$$

Since day-to-day precipitation generally exhibits positive serial correlation (i.e. r_1 greater than 0), p_{11} is usually larger than p_{01} and the wdf is between the two. Note, that a first-order two-state Markov chain does not imply independence for lags greater than one. The autocorrelation r_L (6) decays exponentially with larger lags L:

$$r_L = \left(p_{11} - p_{01}\right)^L \tag{6}$$

3.2 Precipitation amount process

As will be detailed in Sect. 3.3, precipitation amounts at wet days are drawn from probability density functions (PDFs) fitted at single stations. Many studies use either an exponential (Richardson, 1981) or a gamma distribution (Buishand, 1978; Katz, 1977) to model non-zero precipitation amounts ($X(t) \geq 1$ mm day^{-1}). Both distribution types, however, do not appropriately characterize the frequency of the heavily right skewed precipitation amounts: they underestimate either light precipitation (exponential distribution) and / or heavy precipitation (exponential and gamma distribution). As an alternative, a mixture model of two exponential distributions has been proposed to provide better overall fits and to better represent precipitation extremes (Wilks, 1999a). The PDF can be formulated as:

$$f(x) = \frac{w}{\beta_1}\exp\left(-\frac{x}{\beta_1}\right) + \frac{1-w}{\beta_2}\exp\left(-\frac{x}{\beta_2}\right) \tag{7}$$

f(x) is a weighted average (weight w) of two exponential distributions with means β_1 and β_2. Its quantile function exists in a closed form. Consequently, random samples from this distribution can easily be obtained by inversion (Wilks, 2011). The parameters w, β_1 and β_2 are estimated by using the concept of maximum-likelihood (Tallis and Light, 1968). Note that the estimation of PDF parameters is subject to sampling uncertainty from the available number of wet days in a given calendar month.

3.3 Stochastic modelling of daily precipitation

3.3.1 Single-site

In this section, we demonstrate how the occurrence (Sect. 3.1) and amount model (Sect. 3.2) are applied to stochastically simulate daily precipitation at a single site. The simulation process is based on Richardson (1981) with the five above-introduced parameters serving as input in Figure 2: i.e. the transition probabilities p_{11} and p_{01} as well as w, β_1 and β_2. The simulation of precipitation at a given day and a given station (say A) is accomplished in four main steps (see yellow circles in Figure 2):

1. A random number $u_{t,A}$ is drawn from a standard normal distribution.
2. The conditional wet day probability p_A is determined depending on the state of the previous day. It is set to $p_{11,A}$ or $p_{01,A}$, depending on whether the previous simulated day was wet or dry, respectively.
3. The random number $u_{t,A}$ is compared to the standard normal quantile function Q, evaluated at p_A: if $u_{t,A}$ is larger than $Q[p_A]$, a dry day ($J_{t,A}'=0$) is simulated and else a wet day ($J_{t,A}'=1$) is set.

4.1 In case of a dry day, the simulated amount $X_{t,A}'$ is set to zero.

4.2.1 In case of a wet day, a second random number $v_{t,A}$ (independent from $u_{t,A}$) is drawn from a standard normal distribution.

4.2.2 The corresponding quantile of the random number $v_{t,A}$ is then inserted into the quantile function of the mixture model yielding the corresponding precipitation amount (x_A) at a given day.

Note that this simulation procedure could be simplified by taking random uniform [0,1] numbers instead of Gaussian random numbers. We use the latter here in order to be consistent with the multi-site extension introduced later (Sect. 3.3.2).

Steps 1-4 are repeated over all remaining days within a certain simulation period. Based on this procedure time-series of arbitrary length can be generated that resemble observed climatological precipitation statistics, both in terms of frequency and intensity. For a more realistic reproduction of the annual cycle of precipitation the WG is calibrated on a monthly basis (see Sect. 3.4).

3.3.2 **Multi-site**

So far, the procedure to generate precipitation consists of multiple single-site WGs only. Specifically, no spatial dependence in the simultaneous simulation of precipitation at several sites was taken into account. To close this gap several single-site WGs are driven simultaneously with spatially correlated but serially independent random numbers (Wilks, 1998). For simplicity, the concept is illustrated in Figure 2 for the example of two fictitious sites (*A* and *B*) only. The extension to several sites is straightforward. One of the main hurdles in simultaneously generating precipitation at several sites is the prescription of the spatial correlation matrices such that the dependence is also preserved in the final generated time-series (Wilks and Wilby, 1999; Wilks, 1998). This difficulty mainly arises from the stochastic process that partly destroys the initially imposed correlation structure again (Wilks, 1998). We will come back to this problem later. For the moment, let us assume that the optimal correlation matrices for both, occurrence and amount (i.e. $\phi_{AB,\ optim}$ and $r_{AB,\ optim}$), are known. In this case, the main extensions to single-site WGs are two spatially correlated but serially independent random number streams (dashed boxes in Figure 2): one for the occurrence (u) and the other for the amount (v) process. They are determined prior to the simulation process (see below) and contain the same number of days as the simulation period. Once these correlated random number streams are generated, the simulation proceeds as in Sect. 3.3.1 for all stations simultaneously. In practice, the multi-site WG implies the handling of three main methodological hurdles that are the following:

1) Calculating spatial correlation coefficients ϕ_{AB} and r_{AB}

Spatial dependence in binary series at site A and B is inferred by the phi-coefficient (ϕ_{AB}). Similarly as the Pearson correlation coefficient, the phi-coefficient ϕ_{AB} is bounded by -1 and 1. For the precipitation amounts, the spatial correlation coefficient (r_{AB}) is determined by the conventional Pearson product-moment correlation coefficient. The correlation is calculated over the whole precipitation series that also include time-steps with zero amounts. From a statistical point of view, this is not an optimal procedure, since the correlation coefficients could be strongly affected by the number of zeros in the time-series. However, the purpose here is to use this spatial similarity measure rather as a tool to compare the observed spatial dependencies with those in artificial data. It is assumed that the statistical limitations in the calculation apply similarly to observations and generated data. The spatial correlations between different sites are determined pair-wise. Note that the pair wise estimation of the inter-station correlation can result in matrices that are not positive definite, especially when the number of station number is large or when there are incomplete station records.

2) Finding optimal spatial correlation coefficients $\phi_{AB,\,optim}$ and $r_{AB,\,optim}$

As mentioned above, imposing observed inter-site correlations as input to our WG does not guarantee its reproduction in the generated series. This is due to a randomization process through transition probabilities calibrated at each site separately. In general, the imposed correlation is reduced by the stochastic process, both in terms of occurrence and amount process. This characteristic is illustrated at an artificial example of two fictitious sites A and B in Supplementary Fig. 1. While the random number streams (u_A and u_B) perfectly incorporate the observed spatial correlation in occurrence between A and B, it is essentially the two distinct transition probabilities at the two sites that lead to a final correlation in the binary series that is much reduced ($\phi_{AB,\,sim} = 0.6$ compared to $\phi_{AB,\,obs} = 0.8$). In case of precipitation amounts the mismatch in correlation magnitude is also present ($r_{AB,\,sim} = 0.38$ compared to $r_{AB,\,obs} = 0.5$) and can be mainly explained by two factors. First, precipitation amount is only simulated at wet days (i.e. where J_t'=1), while the correlated random number streams ($v_A(t)$ and $v_B(t)$) are representative for the full time-series. Hence, the number of zeros introduced by distinct transition probabilities impact on the generated correlation coefficient. Second, if the two fitted PDFs at the two sites are markedly different, the correlation of the observed and simulated precipitation time-series will deviate, even in absence of any zeros.

To overcome this inherent problem of a multi-site WG after Wilks (1998), an optimization procedure was proposed to find an input spatial correlation that ultimately yield the target correlation of the observations. This has to be done first for the occurrence process ($\phi_{AB,optim}$) and then in a subsequent step for the amount process ($r_{AB,\,optim}$). The optimization procedure iterates over an interval of input correlations, thereby running at each iteration the full occurrence and amount model of the multi-site WG (see Supplementary Fig. 2). After each iteration, the resulting correlation is compared to the target correlation of observations. To find an optimal correlation, we use a bisection method (Burden and Faires, 2010) as non-linear root finding algorithm. The iteration is repeated until the generated correlation equals the one of observations with a precision of 0.005 (see Supplementary Fig. 2). Note that this estimation procedure is done prior to the simulation and has to be repeated for each station pair and month.

3) Generation of correlated random number streams

There are several approaches to generate spatially correlated random numbers streams (e.g. Monahan 2011). For the study at hand we applied a Cholesky decomposition (e.g. Higham 2009):

1. Sample for each station a random number stream from a standard Gaussian distribution.
2. Apply a Cholesky decomposition to the optimized correlation matrix to get a lower triangular matrix and its transposed.
3. Multiply the resulting lower triangular matrix with the matrix of random number streams.

Cholesky decomposition requires matrices that are positive definite, i.e. that contain no negative eigenvalues. However, in case of the applied pairwise optimization process (see section (2) above) this is not always fulfilled. In absence of positive definite matrices, a fall-back solution based on the nearest positive correlation matrix was chosen. The nearest positive definite matrix was found by using the algorithm proposed by Higham (1989), which uses a weighted version of the Frobenius norm. This problem occurred in our study only a few times. Note, that the temporal correlation structure of the precipitation time-series at one specific site is not altered by the imposed spatial correlation, since the spatially correlated random number streams exhibit no serial correlation.

3.4 Implementation

3.4.1 Implementation of the multi-site WG over the *Thur* catchment

Our developed precipitation generator is calibrated on a monthly basis. First, all the single-site input parameters (p_{11}, p_{01}, β_1, β_2 and w) were estimated for each of the 8 stations within the catchment and for each month separately using a time-window of 51 years (1961-2011). In this study we chose a relatively long calibration period in order to minimize the effect of sampling uncertainties. This allows us to accurately assess the added value of a multi-site model against multiple single-site models and to better quantify systematic biases of the WG. For the two transition probabilities in a given month, the climatological mean over the 51 yearly values of p11 and p01 was taken. In the case of fitting a PDF to non-zero precipitation amounts and the estimation of β_1, β_2 and w, we used the daily data over all 51 years together. In addition, a three-month window centred at the month of interest was chosen, in order to increase sample size and the robustness. The distributional parameters were derived based on maximum-likelihood (Tallis and Light, 1968). Despite our three-month time-window, cases occurred when the maximum-likelihood algorithm did not converge. For such cases, a fall back solution was applied where the parameter estimates from the previous month were adopted. With the monthly parameters from all the calibrated single-site WGs and the monthly observed inter-station correlations (symmetric correlation matrices), the optimized correlation matrices had to be found for each month based on the procedure described in Sect. 3.3.2. Note, that by calibrating the multi-site WG on a monthly instead of a seasonal basis, additional sampling uncertainty is introduced due to the rather small time-window to estimate our parameters. This is the downside of prescribing an improved annual cycle in the WG parameters.

Once the multi-site WG was calibrated, we generated 100 ensembles of daily time-series, of 51-year length. All the results presented in Sect. 4 are calculated over the time-period 1961-2011.

3.4.2 **Reproduction and uncertainty of WG model parameters**

To test whether our developed WG is properly implemented, we evaluated the reproduction of WG input parameters extracted from the generated time-series. A correct reproduction in parameters such as wet day intensity, frequency and transition probabilities is a prerequisite for all the subsequent analyses presented in Sect. 4. The evaluation was performed for four subjectively-defined climatic regimes: a very dry, a dry, a wet and a very wet climate. The corresponding model parameters are indicated in Figure 3 with dashed vertical lines. For each of these precipitation regimes, 100 synthetic daily time-series were generated. To test the effect of sample-size, different sizes of time-windows were used: (a) 10’000 days, (b) 1000 days, (c) 100 days and (d) 30 days. The latter corresponds to the same sample-size as used to simulate monthly precipitation over the *Thur* catchment. For each of the generated time-series the WG parameters were re-estimated and the 95% interquantile range was computed across the set of 100 realizations (Figure 3). Three main results can be inferred: (a) our precipitation generator is able to correctly reproduce the key WG parameters implying that the chances for substantial coding errors are small; (b) as expected the estimate of the input parameters becomes more uncertain for smaller sample sizes; in fact, the uncertainty range increases by a factor of 18.3 when the sample size is reduced from 10000 to 30. At a sample size of 1000 the uncertainty range stays at around ± 0.03, that only marginally lowers when going to a sample of 10000. (c) the different pre-defined climate regimes affect the uncertainty, particularly in the estimated transition probabilities. In a very dry or wet climate, the wet-wet or dry-wet transition probability, respectively, exhibits large uncertainties in the estimate. This again is mainly related to a sample size problem due to very few wet-wet or dry-wet pairs. Thus, we expect that the weather generator does not work optimally in arid climates.

4 Results

An in-depth evaluation of the generated time-series with our calibrated multi-site WG is now undertaken with real observations. First, the reproduction of the daily and longer-term precipitation statistics at individual sites is analysed (Sect. 4.1). In a second step, the performance of the multi-site model is investigated regarding spatially aggregated precipitation indices in comparison to WGs without incorporating spatial dependencies (Sect. 4.2).

4.1 Validation of the precipitation generator at individual sites

Based on our ensemble of synthetic time-series, each containing 51 years, we analyse the reproduction of key precipitation characteristics. This validation goes beyond the reproduction of pure model parameters used to calibrate the WG (Sect. 3.4.2), as it includes precipitation statistics that are not directly used in the specification and calibration of the model. Note, that we present this analysis for the same time-period as used for calibrating our WG. This is justified for the study here, as long as we treat and use our WG to simulate long-term monthly precipitation statistics. In such a setup the stationarity of the model is given by definition. However, in a climate prediction or projection context, this stationarity assumption would have to be tested and hence separate calibration and validation periods are needed.

4.1.1 Long-term mean and inter-annual variance of monthly precipitation sums

In a first step of validating our WG, we focus on the reproduction of the long-term mean in monthly precipitation sums. Figure 4 shows both the modelled (blue) and observed (black) long-term monthly precipitation sum for each of the eight investigated stations. In general, the annual cycle of precipitation sums is well reproduced. Consistently, this is also true for the long-term seasonal as well as for the annual precipitation sums (not shown). But the WG tends to slightly underestimate precipitation sums in June and August, and overestimate them in October. In addition, the two stations *Bischofszell* (BIZ) and *Herisau* (HES) show rather large positive deviations from the observed record during the winter months. In order to explain part of these deviations, we decomposed the long-term mean of monthly (T=30 days) precipitation sums ($E[S(T)]$) into the product of the mean monthly wet day frequency (wdf) and intensity (wdi) (Figure 5):

$$E[S(T)] = T \cdot wdf \cdot wdi \qquad (8)$$

Since these two climatological quantities are indirectly forced (Sect. 3.4.2), we expect from the results in Figure 3 a good match on average. As shown in Figure 5, this is true for the wet day frequency, where the deviations between generated (red) and observed (black) values are relatively small. The differences, however, are more pronounced in case of mean wet day intensities. In fact, it is the wet day intensities that explain the mismatches in precipitation sums. In case of the winter performance over *Bischofszell* and *Herisau* the deviations can be attributed to the failure of converging in case of fitting the non-zero precipitation amount. For those instances, the fallback solution had to be used (see 3.4.1).

Next we focus on the inter-annual variability of monthly precipitation sums, which is often more difficult to realistically model than the long-term mean (Wilks and Wilby, 1999). The shaded areas in Figure 4 represent the inter-quartile range of the observed (grey) and modelled (blue) monthly precipitation sums. From Figure 4 it is obvious that the variability of the WG is smaller than in observations for all of the analysed stations. This implies that the stochastic model only explains part of the observed total variability. This reduced variability is expected, as observations are subject to additional sources of variability, which our comparable simple WG is not trained for. The WG is forced with mean observed values, varying between months but not between different years. The annual cycle is assumed to be stationary, and hence interannual variability, e.g. related to the North Atlantic Oscillation

(Hurrell et al., 2003) is missing. Consequently, the ratio of simulated over observed variance accounts for approximately 33% on average. The magnitude of this result is consistent with other studies (e.g. Gregory et al. 1993). Further insights can be gained from a decomposition of the variance of monthly (*T*=30 days) precipitation sums (*Var[S(T)]*) into the variance of non-zero amount (*Var[X ≥1 mm day^{-1}]*) and the variance of the number of wet days (*Var [N(T)]*) as proposed by Wilks and Wilby (Wilks and Wilby, 1999):

$$Var[S(T)] = T \cdot wdf \cdot Var\left[X \geq 1\frac{mm}{d}\right] + Var[N(T)] \cdot wdi^2 \qquad (9)$$

Since the mean wet day frequency (wdf) and intensity (wdi) are reasonably reproduced, we expect that the reduced variability of monthly precipitation sums originate from deficiencies in correctly reproducing the inter-annual variability of the number of wet days and/or of the non-zero amount. One likely reason is the neglect of low-frequency variability in the WG parameters. It has been shown that physically based models that include large-scale circulation as a predictor could alleviate this problem (Chandler and Wheater, 2002; Furrer and Katz, 2007; Wheater et al., 2005; Yang et al., 2005).

4.1.2 **Reproduction of PDF of daily non-zero amount**

The adequate reproduction of the mean wet day intensity and frequency is a necessary but not sufficient precondition of a WG to be used for subsequent (impact) studies. Due to a large variability of precipitation amounts, it strongly matters how its frequency distribution is reproduced. For this, we compared simulated and observed quantiles of the daily non-zero precipitation distribution at each station (Supplementary Fig. 3). Generally, the mixture model of two exponential distributions captures the frequencies of the intensities reasonably well, even at the high-Alpine station *Saentis* (SAE). This is at least the case up to the 80th percentile, above which intensities are systematically underestimated at all stations. This issue could be overcome by more sophisticated amount models combining e.g. a Gamma with a Generalized Pareto distribution (Vrac and Naveau, 2007). However, this comes at the expense of fitting many parameters with a limited sample size.

4.1.3 **Reproduction of multi-day statistics**

While the frequencies of precipitation amounts and the frequencies of wet and dry days are realistically simulated, it remains unclear how the WG performs for multi-day spells. For many application studies, this is an essential information that requires a specific analysis.

Figure 6 displays observed and modelled cumulative frequencies of dry and wet spells lengths at the example of two months and two stations. The two stations *Saentis* and *Andelfingen* are selected for display since they represent the stations with the highest and lowest elevation in the catchment. For both stations a clear seasonal difference in the probability of dry spells toward more short and less long dry spells during summer compared to winter is found. A plausible explanation are the more intermittent (convective) precipitation systems during summer. In contrast to dry spells, no seasonal differences in wet spell length probabilities can be inferred. This is likely related to the fact that the dry-dry transition probability p_{00} exhibits a more distinct annual cycle than the wet-wet transition probability p_{11}. Figure 6 also shows that the frequency at shorter spell lengths (up to 3 days) is more realistically reproduced by the model than the frequency at longer spell lengths. Generally, a better reproduction of wet spell probabilities is seen compared to the dry spell counterpart. Long dry spell lengths are more frequently underestimated by the model than longer wet spell lengths. The underestimation of long wet and dry spells is a common shortcoming of the Richardson-type weather generator and has been reported by many studies before (e.g. Racsko et al. 1991). This deficiency mainly arises due to the fast exponential decay of the autocorrelation function with larger lags (see Eq. (6)). Similar to the underestimation of variability in precipitation sums, higher-order Markov chains (Wilks, 1999b) or GLMs with additional predictors might improve this aspect, which is out of scope in this study here.

Given that the frequency of wet spell lengths is realistically simulated, the question arises whether this also holds for multi-day precipitation sums. Multi-day periods of rain is a common phenomenon over Switzerland, especially during prevailing weather situations that favour orographic uplift. We compared observed and simulated cumulative distribution functions (CDFs) of precipitation sums over multiple consecutive wet days (Figure 7). Overall, we found that the differences between generated and observed time-series are largest for the higher quantiles and for long lasting wet spells (5-day wet spells) where the WG tends to underestimate large multi-day sums. This reduced skill in simulating longer wet spell sums can be explained by the fact that our WG is only prescribed with the temporal structure of precipitation occurrence but not in amount. In other words, the WG has memory to realistically reproduce multi-day wet spell lengths (Figure 6), while the combined analysis of multi-day occurrence and accumulated amount loses somewhat this memory again. Two further noticeable features in Figure 7 are that intense one-day precipitation sums are often overestimated by the model compared to the observations, while a relatively good match is obtained for three-day sums. Although the deficiency in correctly simulating multi-day sums of consecutive wet days is to be expected by construction of the WG, it could be improved by more sophisticated precipitation models, such as multi-states Markov-chains with different probability density distributions at each state (Buishand, 1978; Katz, 1977). This, however, comes at the expense of fitting many additional parameters with a limited sample size.

4.2 Performance of spatial precipitation indices

Up to this point we evaluated the generator at individual sites only. One of the key issue of this study though is the potential added value of incorporating inter-station dependencies. Similarly as in the previous section, we analyse the performance first in terms of occurrence-related statistics and second in terms of the combined occurrence and amount statistics.

4.2.1 **Dry and wet spell statistics for the whole catchment**

Based on the eight stations in our catchment with each being either in a wet or dry state at a given day, theoretically 2^8 (=256) different dry-wet patterns in space are possible. In observations, though, it turns out that 70% of the investigated days over 1961-2011 are in fact either completely dry (45%) or completely wet (25%) and the remaining 254 dry-wet-patterns are subject to far smaller frequencies (around 10^{-5}- 10^{-3} %). The pre-dominance of a dry or a wet catchment makes sense given that the catchment is relatively small and given that precipitation is to a large degree circulation-triggered. Analysing the synthetic time-series from our multi-site WG reveals an almost perfect match with observations (Table 1), a consequence of prescribing the spatial dependency structure in the occurrence process. Indeed, when re-doing the same experiments with multiple single-site WGs without inter-site dependencies, only about 2% of all days are completely dry in the catchment and none of the days are simulated as completely wet (Table 1). In a single-site WG setup, the chances for all stations being dry or wet ultimately depend on the calibrated wet day frequencies at the eight stations that remain below 0.5 in almost all months (see Figure 5). This implies that the likelihood for dry conditions over the catchment is higher than for wet conditions.

Those days with complete dry or wet catchment conditions were further investigated in terms of the temporal structure. Table 1 presents observed and multi-site simulated spell length statistics for the catchment. In general, remarkably good agreement between observations and the multi-site model is found. This is also true for longer spell lengths, where the spatio-temporal correlation structure is only indirectly given as input to the WG. All of these results imply that the calibrated multi-site WG not only captures the frequencies of spatially aggregated binary series very well, it also does a surprisingly good job in reproducing multi-day dry/wet spells of the *Thur* catchment.

4.2.2 Daily non-zero precipitation sums over the catchment

The above findings on the spatio-temporal correlation structure in the occurrence process also give confidence that daily precipitation sums aggregated over the catchment are reasonably simulated. To answer this user-relevant question, we first analyse seasonal distributions of single-day precipitation area sums over the time-period 1961-2011 (Figure 8). Area sums are defined as the precipitation sum over the eight stations. Note, that days with an area sum of zero were excluded from this analysis and are not shown. The observations (grey boxplots) show in the median only a weak inter-seasonal variability with somewhat higher sums during summer. The spread in daily precipitation is smallest for winter and spring and largest for summer owing to the higher extreme precipitation values observed. Common to all seasons is a distribution that is heavily right-skewed ranging from nearly dry conditions up to about 220 mm day^{-1}. Note, that the spread shown here includes variability from year-to-year but also within the season of the same year.

Compared to observations, the multi-site generator reproduces well the median of the observed daily areal sums. The relative deviations remain rather small, ranging from -8.5% in summer to +1.6% in autumn. Moreover, the multi-site model is able to capture about 95% of the observed variability in the daily sums, while the single-site WG only explains about 13%. Even for extreme areal precipitation, the deficiencies are rather small. Contrary to a multi-site model, the areal sum derived from several single-site WGs over the catchment (red) systematically underestimates median, variability and consequently the magnitude of extreme precipitation amounts (Figure 8). The relative deviations from observations in the median range from -28% in autumn to -18% in spring. The underestimation may be explained by the fact that the single-site model rarely simulates days where all stations are wet (Sect. 4.2.1). Also, the spatial structure of the precipitation amount is not accounted for.

4.2.3 Annual maximum precipitation sums of consecutive days over the catchment

The previous analysis has revealed a pronounced added value when incorporating spatial dependencies in the stochastic simulation of daily areal precipitation sums over the *Thur*.

Similarly to Sect. 4.2.1, we want to go a step beyond and additionally include the temporal structure. Note that by investigating spatial precipitation sums over multi-days, we explore the limits of our WG. We analyse in Figure 9 annual maxima of observed (grey), and modelled (blue and red for multi-site and single-site, respectively) precipitation sums over several consecutive days (2, 5, and 10 days). This means that out of the aggregated catchment-time-series we compute temporal sums over consecutive days and take the maximum in each year.

Regarding the performance of the calibrated WG in multi-site and single-site mode, Figure 9 shows that both are clearly underestimating the observed sums. Yet, the multi-site model exhibits much smaller deviations from the observed distribution than the single-site model, and hence the added value of the multi-site WG is clearly evident. In fact, the sums simulated with the multi-site WG are larger by a factor of around 1.8 than those generated with the single-site WG. Overall, deviations from observations are reduced from about -53% (single-site WG) to about -17% (multi-site WG). The added value of the multi-site model is not constant for different consecutive sums. Differences are larger at shorter multi-day sums and decrease toward longer time-windows. This is related to the fact that the spatio-temporal correlation structure at longer lags is not prescribed in the model as already seen in Sect. 4.2.1 and Table 1. The benefit of a multi-site WG in terms of maximum daily areal precipitation sums is therefore restricted to consecutive sums over a few days only. And as a consequence for time-windows of 30 days (or monthly sums), a single-site WG performs equally good as a multi-site WG (not shown), as both models are calibrated for monthly sums at the eight stations and consequently at the catchment.

5 Discussion

The incorporation of inter-station dependencies in the stochastic model brings substantial added value over multiple single-site models regarding daily and multi-day areal precipitation sums over the *Thur* catchment. Similar benefits from the multi-site WG would be expected for other Alpine catchments and regions with complex topography, where correlations between sites are significant but well below unity. For very homogeneous regimes (inter-station correlation near unity) one single-site WG would be sufficient for the catchment-area, whereas for low spatial correlations several independent single-site WGs can be used.

A stochastic simulation with multi-site correlation structure comes with additional uncertainty from parameter estimations, additional implementation complexity and additional computational costs. The decision for incorporating spatial dependencies must therefore be balanced with the benefit. A careful inspection of the observed precipitation regime and its spatial structure over the catchment prior to the simulation is necessary to decide in favour or against multi-site simulation. This is also important in terms of validation: for a large catchment area that is frequently affected by frontal passages, the validation of the precipitation generator should include more complex space-time dependency analyses. An example is the probability of a certain precipitation amount at a particular station given precipitation at a neighboring station some days earlier.

For many impact applications gridded precipitation data instead of multiple scattered stations would be beneficial. This demand could be achieved by interpolating the spatially consistent synthetic station data over the area of interest. A more sophisticated and elegant method, however, is to build a field generator, for instance by high-dimensional random Gaussian fields (e.g. Pegram and Clothier, 2001), random cascade models (e.g. Over and Gupta, 1996) or Poisson cluster models (e.g. Burton et al., 2008). An alternative would be to rely on geostatistical methods, for instance by prescribing a spatial correlation function at gauged and ungauged locations, that additionally requires specifying also parameters of the WG between the sites (e.g. Wilks, 2009). In regions with complex topography this additional interpolation is not straightforward. It could be alleviated by explicitly including information of topographic aspects (e.g. altitude, aspect and slope) in a GLM- (McCullagh and Nelder, 1989) or Bayesian Hierarchical modelling-approach (Gelman and Hill, 2006). These are appealing frameworks that allow the modelling of physiographic dependencies in the precipitation amount and occurrence model. However, this alone is not sufficient for a space-time weather generator as the spatial dependence of daily precipitation is also determined by spatial autocorrelation and not just the physiographic conditioning of parameters. Clearly, the development of a gridded space-time weather generator dealing with spatial autocorrelation, physiographic conditioning, intermittence and temporal autocorrelation is highly challenging and needs fundamental methodological development. This is beyond the scope in the present study, where our main focus was to develop an easy-to-use statistical downscaling tool for current and future climate.

6 Summary and Outlook

The multi-site precipitation generator of Wilks (1998) has been successfully developed, implemented and tested over the Swiss alpine river catchment *Thur*. The precipitation generator treats precipitation occurrence as a Markov chain and simulates non-zero daily precipitation amounts from a mixture model of two exponential distributions. The spatial dependency is ensured by running the WG with spatially correlated random numbers. The model was calibrated on a monthly basis by using daily station data over a 51-year long time-period from 1961-2011, and extensively compared to the observed record and to simulations based on multiple independent single-site WGs.

Our main findings of this study are:

- The multi-site precipitation generator realistically reproduces key precipitation statistics at single stations, including the annual cycle, quantiles of non-zero precipitation amounts, multi-day spells and multi-day amount statistics.
- The precipitation generator is able to generate relatively large stochastic variability. Nevertheless, it is rather low compared to observed inter-annual variability where it underestimates inter-annual variability by a factor of 3.
- The incorporation of inter-station dependencies in the stochastic process brings substantial added value over multiple single-site WGs. The median of daily area sums are higher by about a factor of 1.3 than those from independent single-site models. In addition, the multi-site WG is able to capture about 95% of the observed variability, while the single-site WG only explains about 13%. Annual maxima of multi-day sums over the catchment increase by about a factor of 1.8 by incorporating the inter-site dependence in the stochastic simulations.
- The added value is largest when the precipitation regime is subject to a large spatial and temporal heterogeneity as it is the case over the *Thur* catchment.

These results provide confidence that the developed precipitation generator is a helpful tool to realistically simulate mean aspects of the current climate. We therefore conclude that this generator can subsequently be used as a statistical downscaling tool to generate synthetic time-series consistent with mean aspects of the future climate. Although there is substantial improvement compared to a simple delta-change approach, from an end-user perspective some relevant limitations need to kept in mind: The synthetically generated time-series (for current or future climate) do not fully capture the day-to-day and multi-day variability of precipitation. Extreme values and longer spell lengths are hence underestimated. The generator further underestimates the year-to-year variability in monthly precipitation sums.

Therefore, care should be taken when using the precipitation generator as a tool for a broad risk assessment, in particular with respect to extreme events.

These inherent limitations point to potential future refinements of the presented model: (a) To better reproduce extreme precipitation, we intend to implement a three-state Markov chain model with the states dry, wet, and very wet and with state-dependent PDFs. From this, we expect a substantial improvement of one-day and multi-day extremes as well as a better reproduction of multi-day precipitation sums. (b) To alleviate the underestimation of inter-annual variability, we will introduce a non-stationary model. This could be accomplished by sampling from a distribution of observed WG parameters (instead of taking the mean) or by formulating a regression model using large-scale atmospheric variables as predictors (see e.g. Furrer and Katz, 2007).

Beside these methodological improvements the precipitation generator will be subject to two extensions: (a) the coupling of daily minimum and maximum temperature as additional atmospheric variables and (b) the adjustment of the WG parameters to represent a future mean climate. Finally, the time-series over the Thur catchment will serve as input for a hydrological model to assess the added value of multi- versus single-site WGs in terms of runoff and to assess the implications of the systematic biases of the WG for hydrological quantities.

Acknowledgements

This work is supported by the ETH Research Grant CH2-01 11-1. We would like to thank the Center for Climate Systems Modeling (C2SM) at ETH Zurich for providing technical and scientific support.

References

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Automat. Contr., 19(6), 716–723, doi:10.1109/TAC.1974.1100705, 1974.

Allen, M. R. and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle., Nature, 419(6903), 224–32, doi:10.1038/nature01092, 2002.

BAFU: Hydrologischer Atlas der Schweiz HADES, Hydrol. Atlas der Schweiz HADES [online] Available from: http://www.hydrologie.unibe.ch/hades/index.html (Accessed 10 April 2014), 2007.

BAFU: Auswirkungen der Klimaänderung auf Wasserressourcen und Gewässer. Synthesebericht zum Projekt «Klimaänderung und Hydrologie in der Schweiz» (CCHydro), Bern, Switzerland., 2012.

Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation, Hydrol. Earth Syst. Sci. Discuss., 6(3), 4485–4534, doi:10.5194/hessd-6-4485-2009, 2009.

Begert, M., Seiz, G., Schlegel, T., Musa, M., Baudraz, G. and Moesch, M.: Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961-1990, Zurich., 2003.

Bosshard, T., Kotlarski, S., Ewen, T. and Schär, C.: Spectral representation of the annual cycle in the climate change signal, Hydrol. Earth Syst. Sci., 15(9), 2777–2788, doi:10.5194/hess-15-2777-2011, 2011.

Buishand, T. A.: Some remarks on the use of daily rainfall models, J. Hydrol., 36(3), 295–308, 1978.

Buishand, T. A. and Brandsma, T.: Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling distribution, Water Resour. Res., 37(11), 2761–2776, doi:10.1029/2001WR000291, 2001.

Burden, R. and Faires, J. D.: Numerical Analysis, 9th Ed., edited by Michelle Julet., 2010.

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. Model. Softw., 23(12), 1356–1369, doi:10.1016/j.envsoft.2008.04.003, 2008.

Calanca, P.: Climate change and drought occurrence in the Alpine region: How severe are becoming the extremes?, Glob. Planet. Change, 57(1-2), 151–160, doi:10.1016/j.gloplacha.2006.11.001, 2007.

CH2014-Impacts: Toward quantitative scenario of climate change impacts in Switzerland, published by: OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland., 2014.

Chandler, R. E.: Rglimclim, [online] Available from: http://www.homepages.ucl.ac.uk/~ucakarc/work/glimclim.html (Accessed 10 April 2014), 2014.

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(10), 1192, doi:10.1029/2001WR000906, 2002.

Cowpertwait, P. S. P.: A Generalized Spatial-Temporal Model of Rainfall Based on a Clustered Point Process, Proc. R. Soc. A Math. Phys. Eng. Sci., 450(1938), 163–175, doi:10.1098/rspa.1995.0077, 1995.

Fatichi, S., Ivanov, V. Y. and Caporali, E.: Simulation of future climate scenarios with a weather generator, Adv. Water Resour., 34(4), 448–467, doi:10.1016/j.advwatres.2010.12.013, 2011.

Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rainn-gauge observations, Int. J. Climatol., 18, 873–900, doi:10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9, 1998.

Furrer, E. M. and Katz, R. W.: Generalized linear modeling approach to stochastic weather generators, Clim. Res., 34, 129–144, doi:10.3354/cr034129, 2007.

Gabriel, K. R. R. and Neumann, Y.: A Markov chain model for daily rainfall occurrence at Tel Aviv, Quart. J. Roy. Meteorol. Soc., 88, 90–95, 1962.

Gelman, A. and Hill, J.: Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press., 2006.

Gregory, J. ., Wigley, T. M. . and Jones, P. .: Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals., Clim. Dyn., 8, 299–310, doi:10.1007/BF00209669, 1993.

Held, I. M. and Soden, B. J.: Robust Responses of the Hydrological Cycle to Global Warming, J. Clim., 19(21), 5686–5699, doi:10.1175/JCLI3990.1, 2006.

Hertig, E. and Jacobeit, J.: A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area, J. Geophys. Res. Atmos., 118(2), 520–533, doi:10.1002/jgrd.50112, 2013.

Higham, N. J.: Matrix nearness problems and applications, in Applciations of Matrix Theory, edited by M. Gover and S. Barnett, pp. 1–27, Oxford University Press., 1989.

Higham, N. J.: Cholesky factorization, Wiley Interdiscip. Rev. Comput. Stat., 1(2), 251–254, doi:10.1002/wics.18, 2009.

Hughes, J. P., Guttorp, P. and Charles, S. P.: A non-homogeneous hidden Markov model for precipitation occurrence, J. R. Stat. Soc. Ser. C (Applied Stat., 48(1), 15–30, doi:10.1111/1467-9876.00136, 1999.

Hurrell, J. W., Kushnir, Y., Ottersen, G. and Visbeck, M.: The North Atlantic Oscillation: Climatic Significance and Environmental Impact, edited by J. W. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck, American Geophysical Union, Washington, D. C., 2003.

Huser, R. and Davison, A. C.: Space-time modelling of extreme events, J. R. Stat. Soc., 76(2), 439–461, doi:10.1111/rssb.12035, 2014.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G. and Vertačnik, G.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, Int. J. Climatol., 34, 1657–1675, doi:10.1002/joc.3794, 2013.

Katz, R. W.: Precipitation as a Chain-Dependent Process, J. Appl. Meteorol., 16(7), 671–676, 1977.

Köplin, N., Viviroli, D., Schädler, B., Weingartner, R. and Bormann, H.: How does climate change affect mesoscale catchments in Switzerland? a framework for a comprehensive assessment., Adv. Geosci., 27, 111–119, doi:10.5194/adgeo-27-111-2010, 2010.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M. J., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C. J., 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, 1–34, doi:10.1029/2009RG000314, 2010.

McCullagh, P. and Nelder, J. A.: Generalized linear models., Chapman and Hall, London England., 1989.

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, doi:10.1016/j.jhydrol.2009.08.033, 2009.

Monahan, J. F.: Numerical Methods of Statistics, Cambridge University Press, Cambridge., 2011.

Over, T. M. and Gupta, V. K.: A space-time theory of mesoscale rainfall using random cascades, J. Geophys. Res., 101(D21), 26319, doi:10.1029/96JD02033, 1996.

Paschalis, A., Molnar, P., Fatichi, S. and Burlando, P.: A stochastic model for high-resolution space-time precipitation simulation, Water Resour. Res., 49(12), 8400–8417, doi:10.1002/2013WR014437, 2013.

Pegram, G. G. S. and Clothier, A. N.: High resolution space–time modelling of rainfall: the "String of Beads" model, J. Hydrol., 241(1-2), 26–41, doi:10.1016/S0022-1694(00)00373-5, 2001.

Racsko, P., Szeidl, L. and Semenov, M. A.: A serial approach to local stochastic weather models, Ecol. Modell., 57, 27–41, doi:10.1016/0304-3800(91)90053-4, 1991.

Richardson, C. W.: Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation, Water Resour. Res., 17(1), 182–190, doi:10.1029/WR017i001p00182, 1981.

Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., 6(2), 461–464, doi:10.1214/aos/1176344136, 1978.

Tallis, G. M. and Light, R.: The Use of Fractional Moments for Estimating the Parameters of a Mixed Exponential Distribution, Technometrics, 10(1), 161–175, doi:10.1080/00401706.1968.10490543, 1968.

Themeßl, J. M., Gobiet, A. and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31(10), 1530–1544, doi:10.1002/joc.2168, 2011.

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, Water Resour. Res., 43(7), W07402, doi:10.1029/2006WR005308, 2007.

Wheater, H. S., Chandler, R. E., Onof, C. J., Isham, V. S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G. and Segond, M.-L.: Spatial-temporal rainfall modelling for flood risk estimation, Stoch. Environ. Res. Risk Assess., 19(6), 403–416, doi:10.1007/s00477-005-0011-8, 2005.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, J. Hydrol., 210(1-4), 178–191, doi:10.1016/S0022-1694(98)00186-3, 1998.

Wilks, D. S.: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models, Agric. For. Meteorol., 93(3), 153–169, doi:10.1016/S0168-1923(98)00125-7, 1999a.

Wilks, D. S.: Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain, Agric. For. Meteorol., 96(1-3), 85–101, doi:10.1016/S0168-1923(99)00037-4, 1999b.

Wilks, D. S.: A gridded multisite weather generator and synchronization to observed weather data, Water Resour. Res., 45(10), n/a–n/a, doi:10.1029/2009WR007902, 2009.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press., 2011.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Prog. Phys. Geogr., 23(3), 329–357, doi:10.1177/030913339902300302, 1999.

Yang, C., Chandler, R. E., Isham, V. S. and Wheater, H. S.: Spatial-temporal rainfall simulation using generalized linear models, Water Resour. Res., 41(11), W11415, doi:10.1029/2004WR003739, 2005.

Table 1. Frequencies (given in percent) of a completely wet or dry catchment together with the frequencies of its spell lengths. The observed (OBS) frequencies are calculated over 1961-2011. The multi-site simulated frequencies are given by the mean of 100 runs over 51 years (1961-2011).

		Wet catchment			**Dry catchment**		
		OBS	***multi-site***	***single-site***	***OBS***	***multi-site***	***single-site***
Overall frequency		25	25	0	45	44	2
Frequencies of spell lengths	**1**	34.8	34.4	0.0	14.1	17.3	2
	2	27.3	29.4	0.0	16.2	20.7	0.0
	3	16.7	18.2	0.0	13.0	18.2	0.0
	4	11.5	9.7	0.0	10.8	14.1	0.0
	5	4.1	4.7	0.0	9.1	10.3	0.0
	6	2.7	2.1	0.0	5.9	7.0	0.0
	7	0.9	0.9	0.0	7.2	4.7	0.0
	8	0.7	0.4	0.0	5.1	3.0	0.0
	9	0.6	0.2	0.0	3.5	1.9	0.0
	10	0.2	0.0	0.0	3.5	1.2	0.0

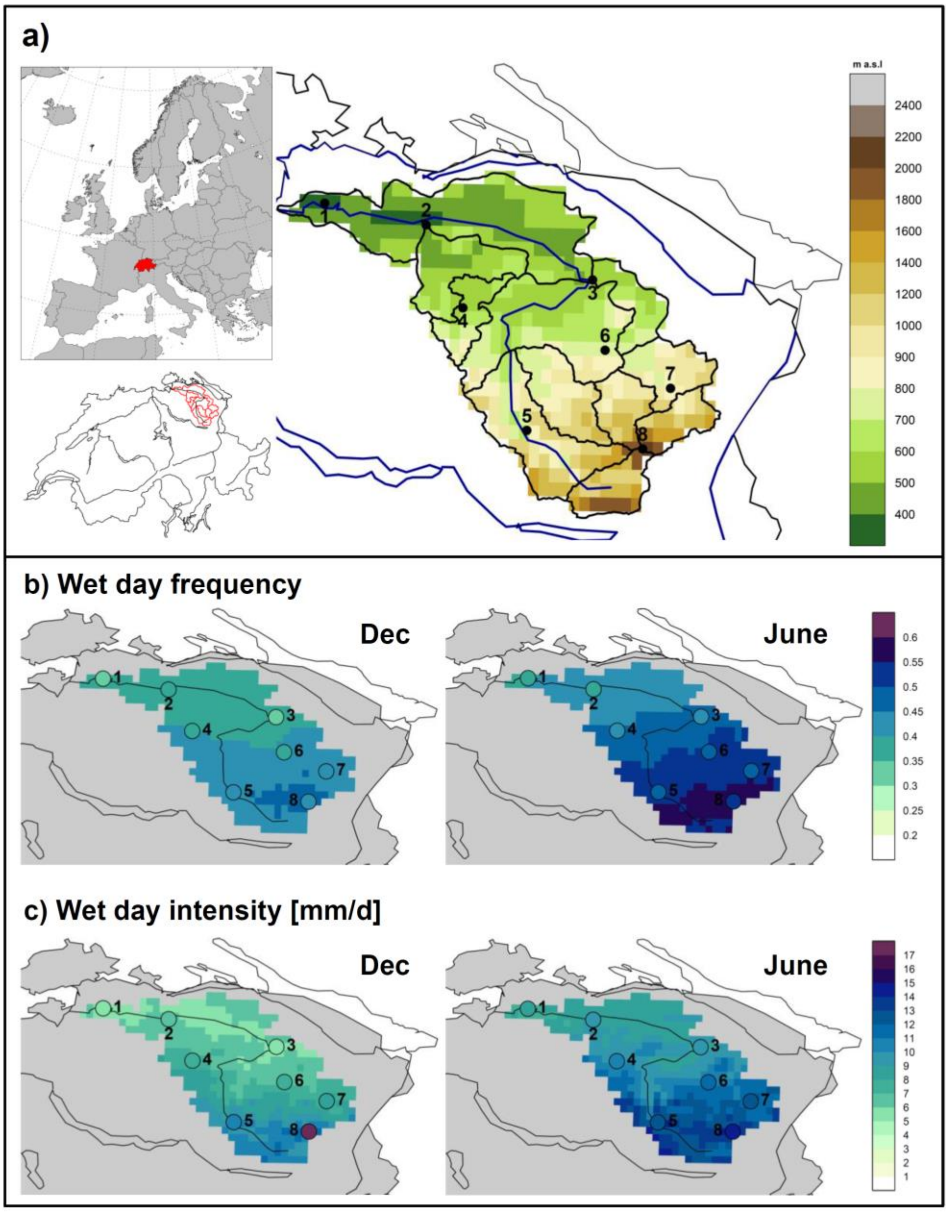

Figure 1. a) The catchment of the river *Thur*, located in north-eastern Switzerland, together with the underlying topography (in m.a.s.l.). The dots indicate the locations of the investigated stations. 1: *Andelfingen* (AFI), 2: *Frauenfeld* (FRF), 3: *Bischofszell* (BIZ), 4: *Eschlikon* (EKO), 5: *Ebnat-Kappel* (EBK), 6: *Herisau* (HES), 7: *Appenzell* (APP), 8: *Saentis* (SAE). b) Observed precipitation climatology of the wet day frequency (1961-2011) derived from a 2km x 2km gridded daily precipitation dataset (Frei and Schär, 1998) for December and June. c) The same as in b), but for wet day intensity (in mm day^{-1}). The filled circle symbols point to the station locations (as in a) together with the observed station measurements.

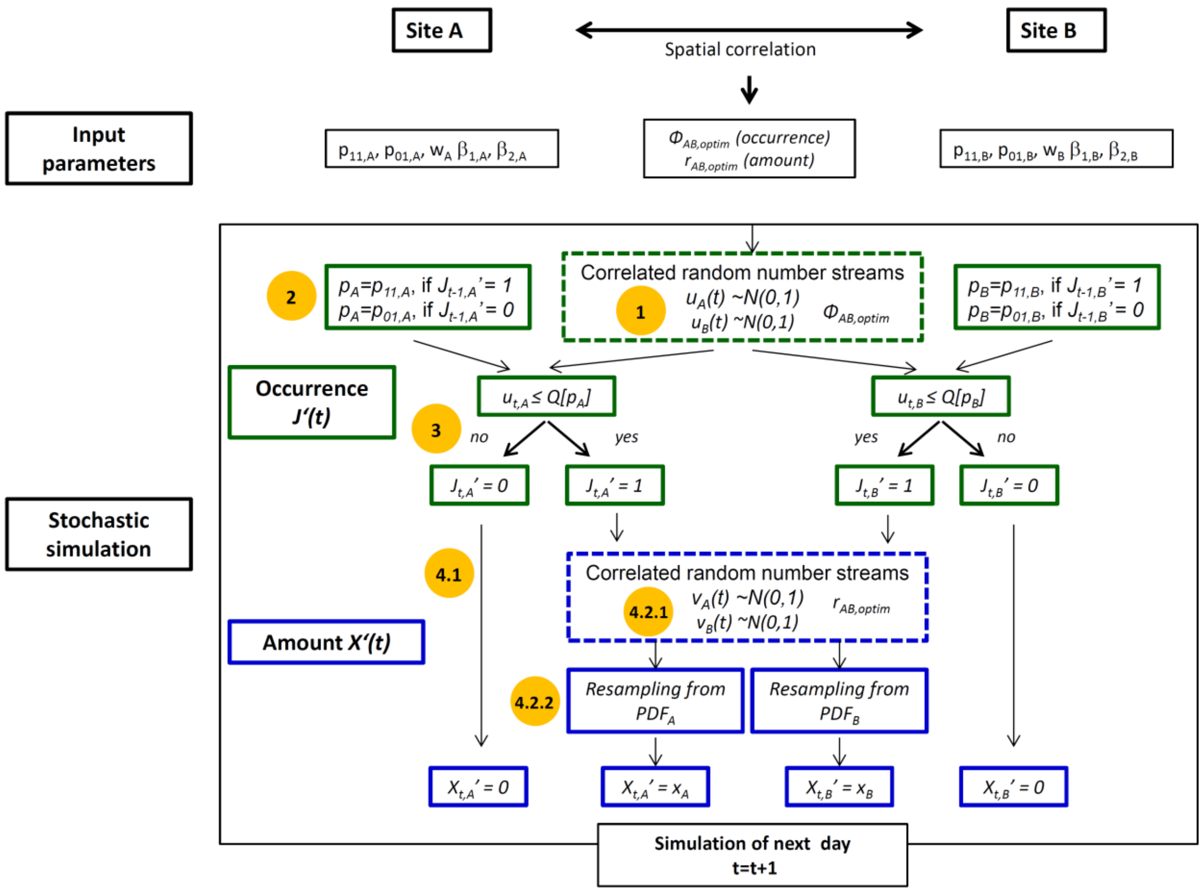

Figure 2. Technical workflow of a multi-site precipitation generator after Wilks (1998) at the example of two fictitious sites *A* and *B*. In general, it is a combination of multiple single-site precipitation generators that are calibrated at each site individually (see input parameters) and run simultaneously with spatially correlated random number streams (dashed boxes). The correlated random number streams (of similar length as the simulation period) are determined beforehand (see Section 3.3.2). The orange-labelled numbers in indicate the steps for single-site precipitation simulation (see Section 3.3.1).

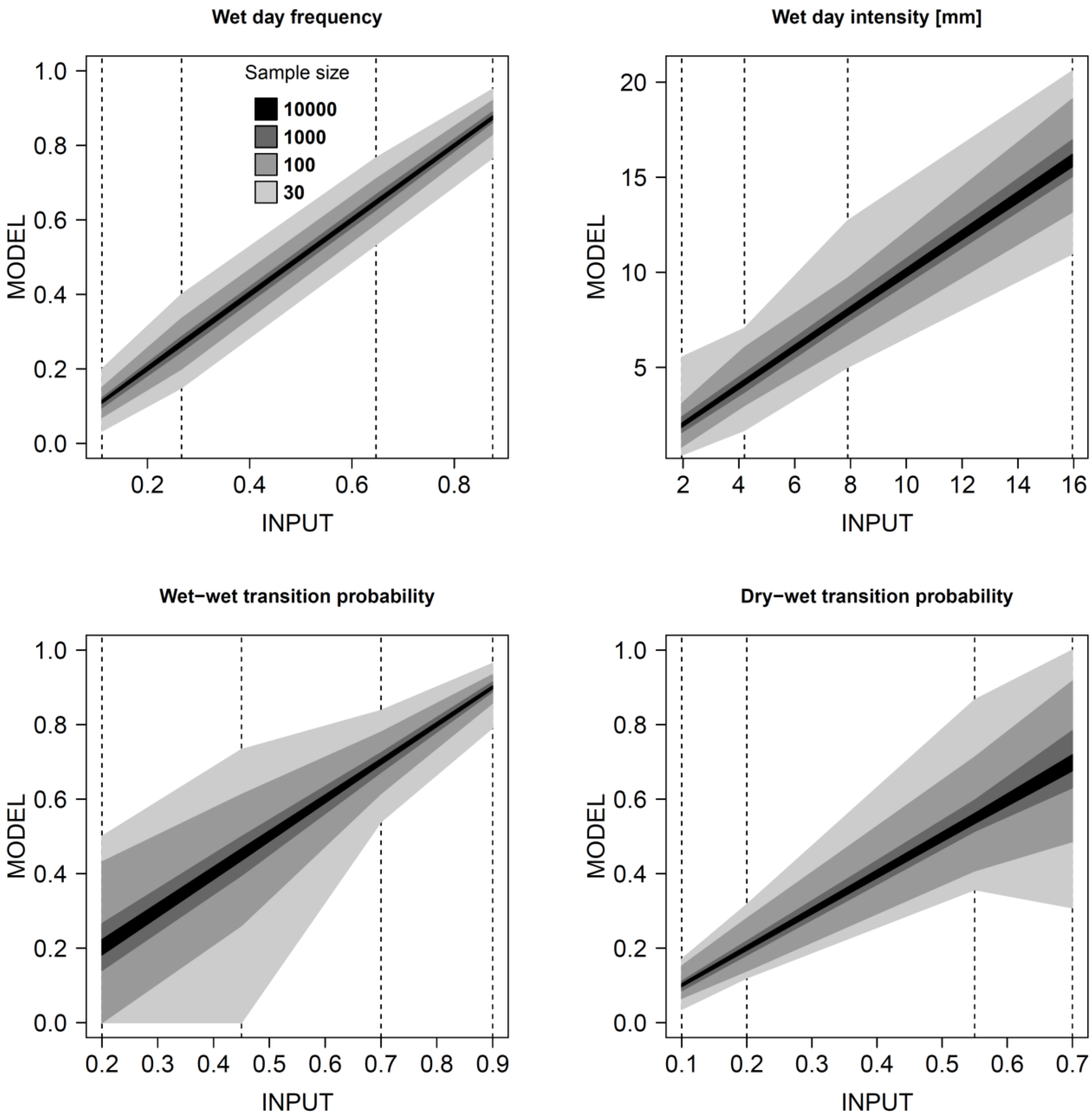

Figure 3. Reproduction of average wet day frequency (wdf), mean wet day intensity (wdi), wet-wet transition probability (p_{11}) and dry-wet transition probability (p_{01}) for the four idealized climate regime ranging from very dry (left) to very wet (right) as indicated by dashed lines. The shaded areas correspond to the range between the 2.5% and the 97.5% empirical quantiles of 100 realizations. Results are shown for sample sizes of 10000, 1000, 100 and 30 (grey shading).

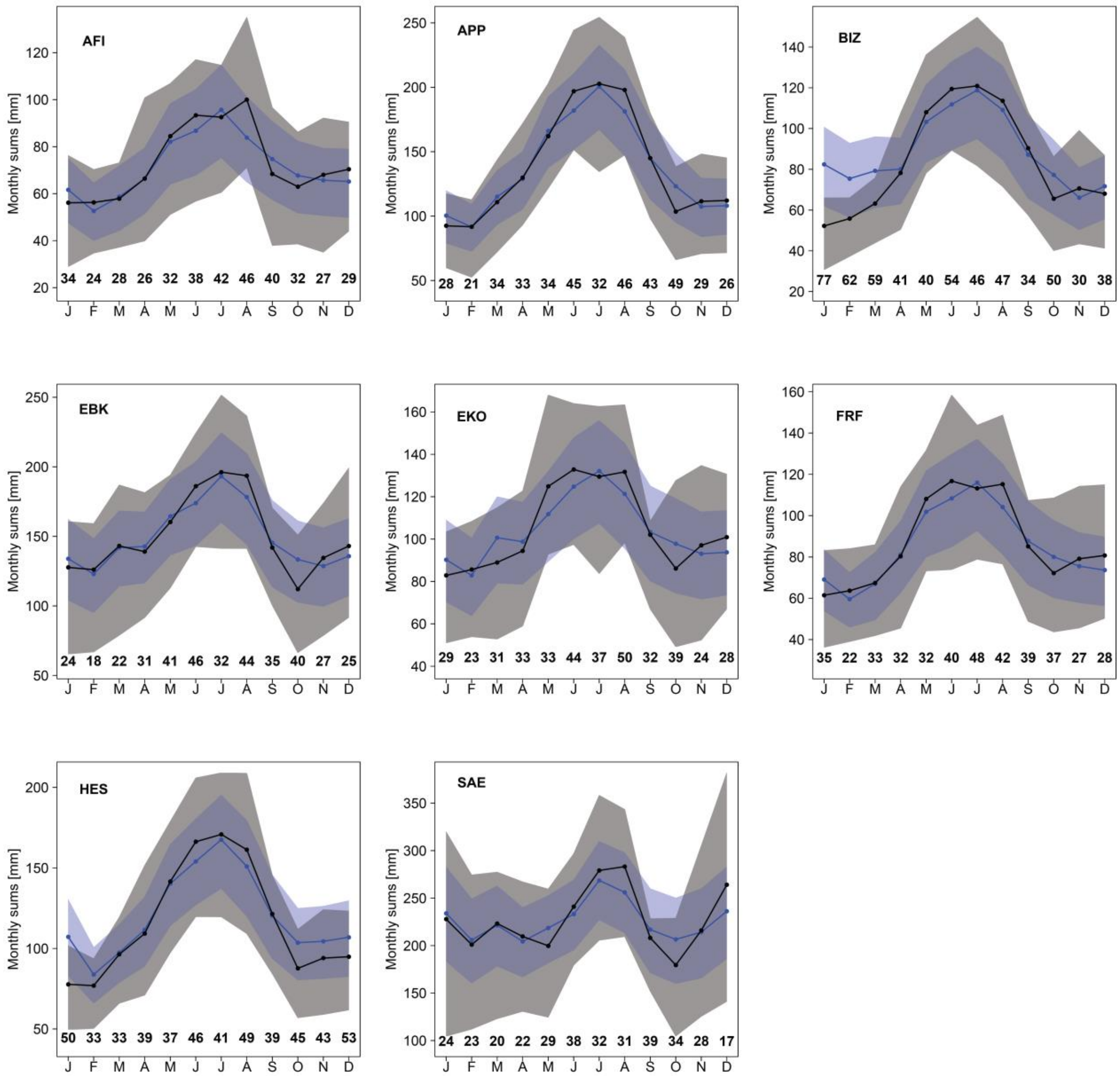

Figure 4. Long-term mean and variability of monthly precipitation sums during the period 1961-2011 for eight stations in the *Thur* catchment. The black (blue) lines refer to the mean annual cycle of observed (modelled) precipitation sums. The grey (blue) shaded areas represent the inter-quartile ranges of observed (simulated) monthly precipitation sums. The simulation comprises 100 realizations covering each 51 years. The numbers at the bottom indicate for each month the percentage of variance explained by the precipitation generator. Note that the scale of the y-axis differ between different stations.

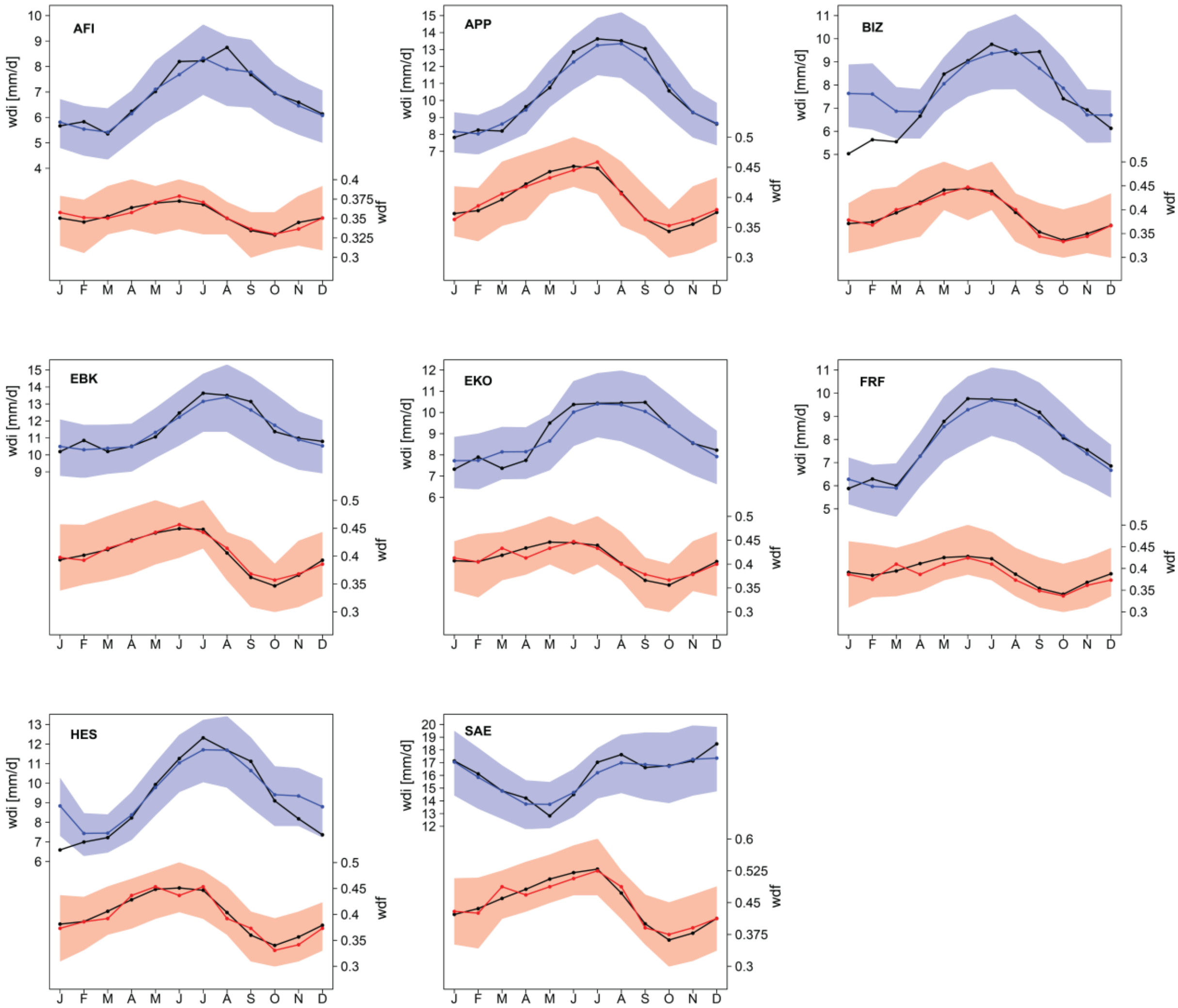

Figure 5. Observed and modelled monthly mean wet day intensity (blue) and frequency (red) at eight stations during 1961-2011. The black (coloured) lines indicate the observed (modelled) values. The blue (red) shaded areas correspond to the inter-quartile range across the set of synthetic daily time-series. They comprise 100 runs covering each 51 years.

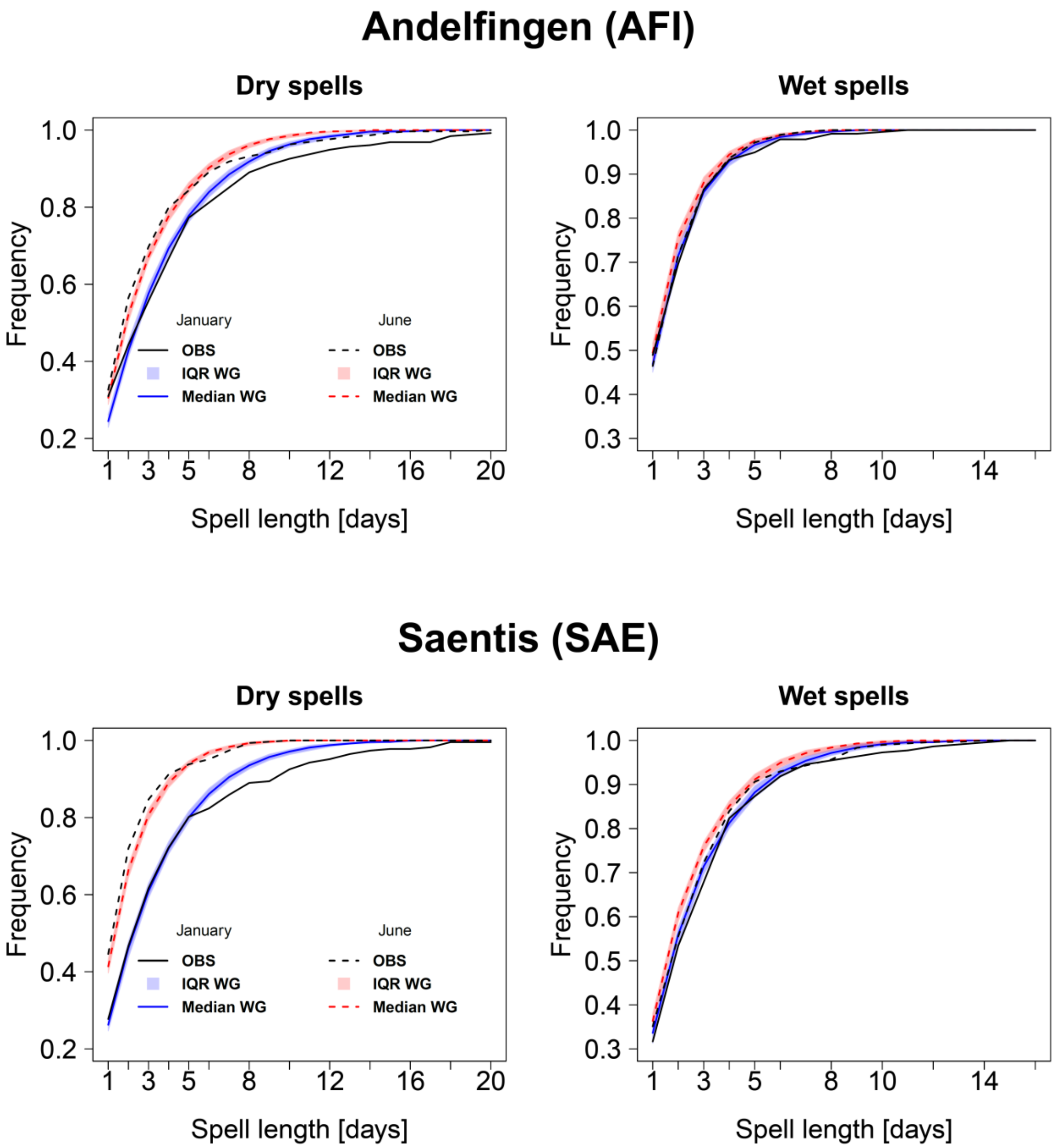

Figure 6. Cumulative distribution of the observed and simulated dry (left) and wet (right) spell length frequencies for the lowland station *Andelfingen* (top) and the mountain station *Saentis* (bottom). Results are for January and June during the time period of 1961-2011. The coloured area (line) represents the inter-quartile range (median) of the 100 realizations covering each 51 year-long daily time-series.

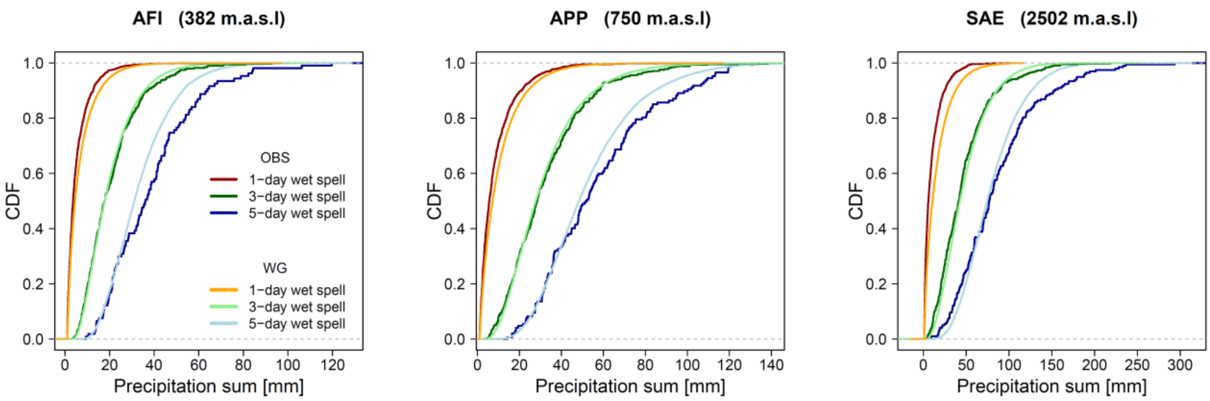

Figure 7. Cumulative distribution functions (CDFs) of multi-day precipitation sums for the three stations *Andelfingen* (AFI), *Appenzell* (APP) and *Saentis* (SAE). The lines represent the CDFs of non-zero precipitation amounts over one day (red), over three consecutive wet days (green) and over five consecutive wet days (blue). Darker and lighter colours refer to observations and simulations, respectively. The observed CDFs have been derived from a 51-year long daily time-series between 1961 and 2011, those of the weather generator from 100 realizations of 51-year long daily simulations. Note that the scaling of the horizontal axis differs between different stations.

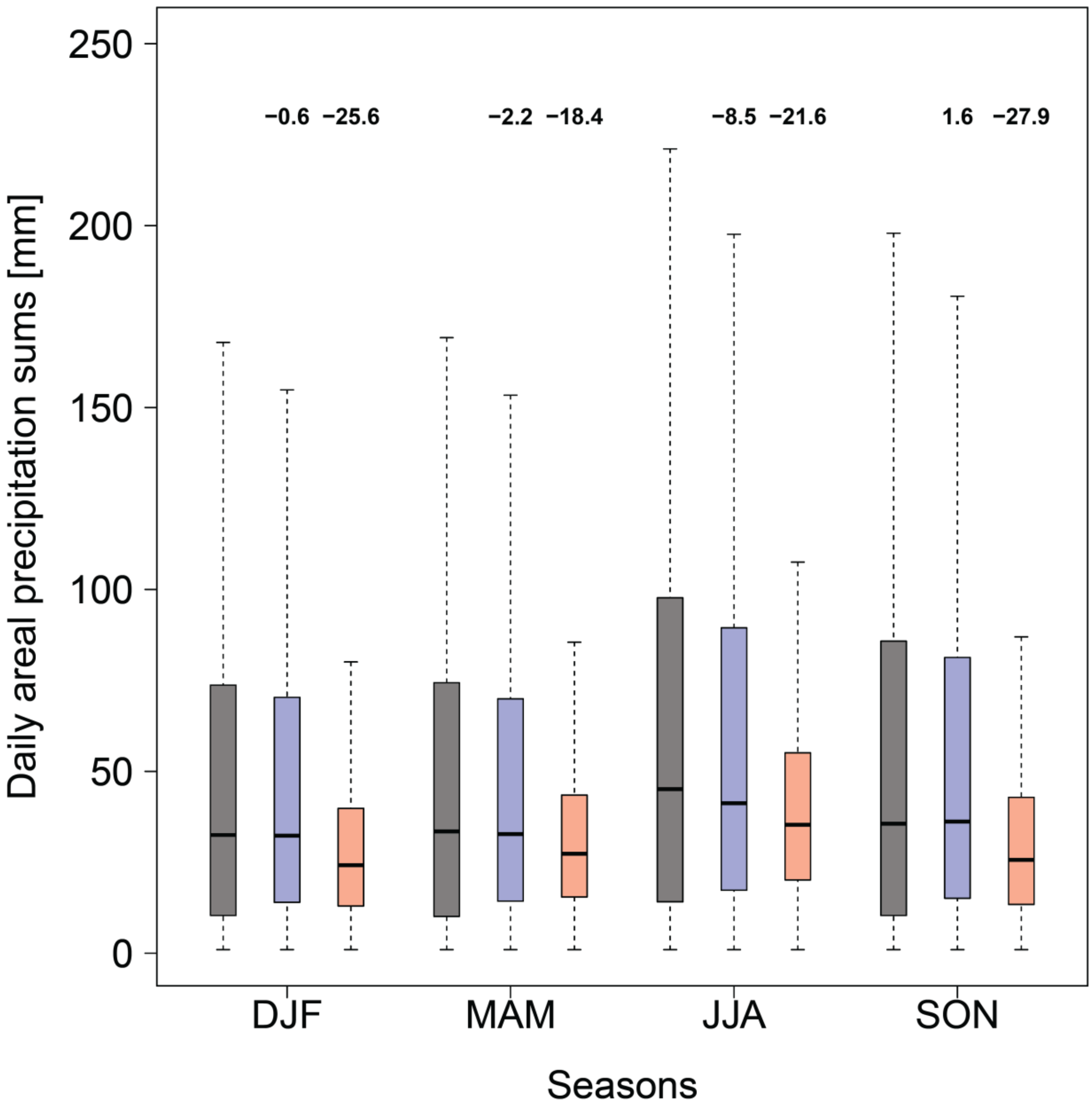

Figure 8. Daily non-zero precipitation sums over the catchment for the four seasons during 1961-2011. Daily Precipitation intensity of the eight stations are summed and days with an area sum of zero are excluded. Boxplots of observed daily sums (grey), of multi-site simulated time-series (blue) and of single-site simulated time-series (red) are shown. The WG models were run 100 times over a 51 year time-period. The numbers (in percentage) indicated above the corresponding model represent the relative deviation of the simulated median from the observed.

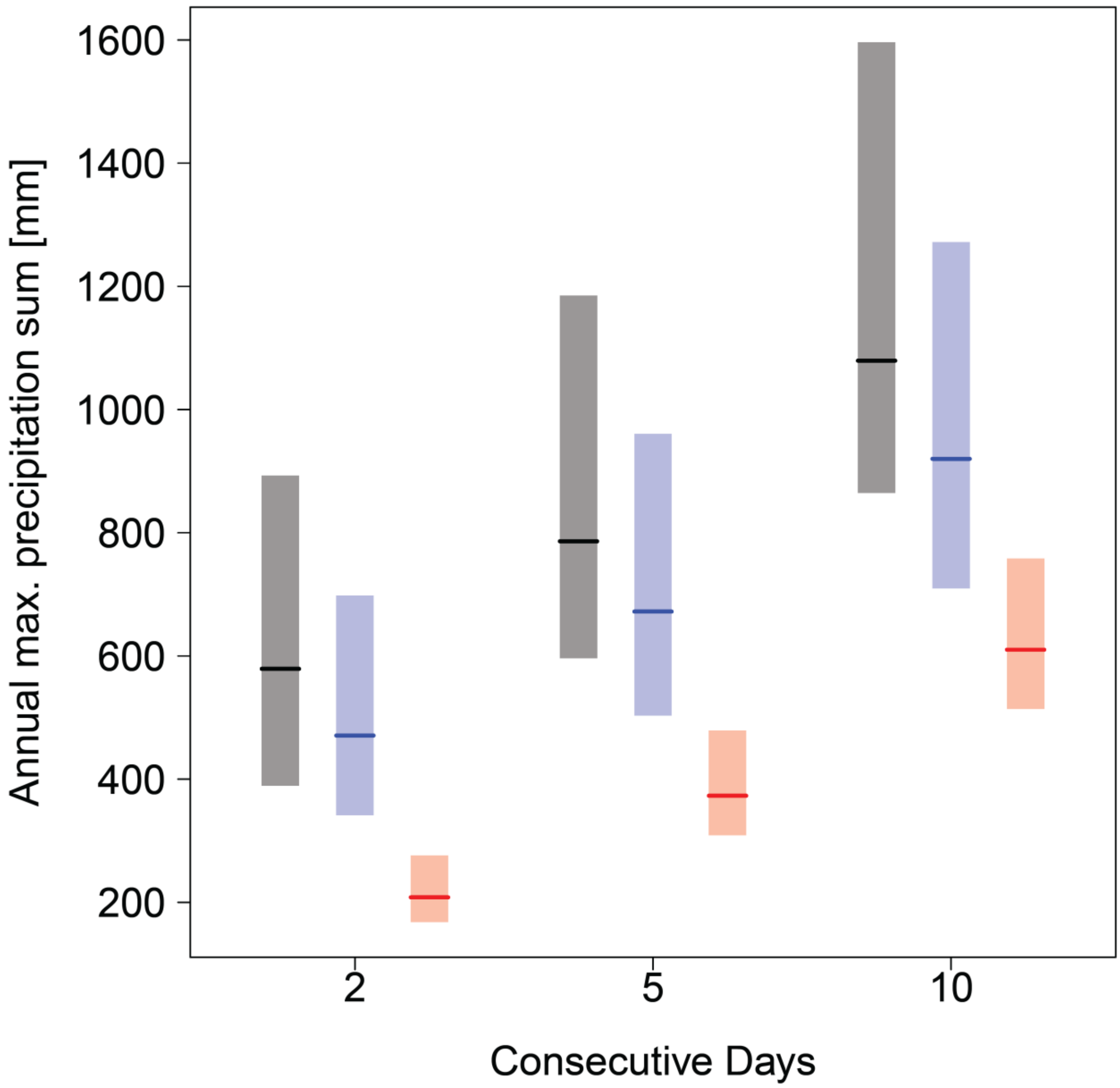

Figure 9. Annual maximum precipitation summed over all eight stations and over consecutive days. The analysis is done for all days of year. The bars (horizontal line) indicate the range between the 2.5% and the 97.5% empirical quantiles of the yearly maximum area sums during 1961-2011. The observations are plotted in grey, the multi-site simulations in blue and the single-site simulations in red. The observations comprise 51 years, the models were run 100 times over a 51 year time-period.