Authors reply on comment of editor

Comment: *Two reviewers have read the manuscript. They provided quite contradicting assessments of the work, but there are two main problematic points that are essential. The first is the innovation of the study. Reviewer 2 claims for no novelty at all while reviewer 1 states that the method is not novel but its application is. Indeed the authors state in the manuscript that the WG was never tested in the Alps that have some unique features. However, these features are not considered later and thus the point made for the novelty in application is not convincing. A second, but related point, is that the study is focused in complex topography according to its title; however the topography data (and its derivative such as slopes) were not actually used in the analyses, except for station selection, and so the unique application of this WG does not seem to be justified. In addition to the two above important issues, the manuscript seems to miss some publications of high space-time resolution WG, including in the Swiss environment. More important comments are listed in the two review reports.*

Reply: We would like to thank the editor for her efforts and in putting together this summary on the reviewers' main points. In the following, we respond to these in a general comment. Please also consider our replies to referee #1 and #2.

Indeed, the two reviewers assess the novelty of the study rather differently. It is true, that the precipitation generator itself is methodologically not new, but rather re-built after Wilks and others. The paper clearly states that. However, the main challenge (and the main work) of such a model is to calibrate and implement it. In a practical application a multitude of concrete decisions and assumptions need to be made, such as the selection of time-period, the choice of non-zero precipitation distribution, the setup of calibration (each month or season separately or something else) to name a few. These practical steps, i.e. the "lessons-learnt" and the resulting consequences for the quality of the weather generator, need to be documented. It was not the purpose of this study to invent a novel stochastic modeling approach. The aim is rather to have a sufficiently simple tool ready for current climate that can be re-adjusted for future climate conditions in a subsequent study (a follow-up article is planed).

Following the reviewer comments, we realized though that the latter aspect was not emphasized enough in the abstract and the introduction. In addition, in retrospective, the title was misleading and may have caused wrong expectations. We have changed the title to "Implementation and validation of a multi-site daily precipitation generator over a Swiss river catchment".

We agree that we have not analysed topographical aspects in detail but the topographical analysis was implicit by using the Thur catchment with its different precipitation characteristics among the stations at different altitudes (see Figure 1 of the manuscript). These characteristics, ranging from e.g. a summer wet day frequency of 0.55 at Saentis (2502 m.a.s.l.) to a wet day frequency of 0.3 at Andelfingen (382 m.a.s.l, only about 60 km away), poses a challenge to a weather generator. To test whether a statistical tool is able to capture these different climata is one aim of our study. In the revised manuscript we spell out more clearly, which aspects of this project are novel, and what the implications of those are.

We agree with the editor, that we have missed a few important publications of space-time WGs, in particular from the hydrology-community. In the revised manuscript, we will provide citations to the following studies:

Huser, R., & Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 439–461. doi:10.1111/rssb.12035

Mezghani, A., & Hingray, B. (2009). 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. *Journal of Hydrology*, *377*, 245–260. doi:10.1016/j.jhydrol.2009.08.033

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

We further have to add, that it was difficult to reply to some of the allegations raised by referee #2 due to their unspecific manner. However, based on the more specific comments and the issues raised by referee #1, we hope that we could address the main critical points.

Authors reply on comments of referee #1

The authors would like to thank reviewer #1 for his/her valuable and detailed comments. We will first give a general reply, then answer the specific comments, and subsequently address the technical comments. For clarification, the referee's comments are repeated first and displayed in italic letters, followed by the authors' replies.

General reply

We are happy that referee #1 likes the presentation of our article in terms of readability, clarity, structure and presentation of visual graphics. His/Her specific comments are excellent suggestions for improving the current generator. Some of the suggestions require rather fundamental changes beyond the scope of this paper. Therefore, these particular issues must be tackled more thoroughly in future work. However, as suggested by the reviewer, we will discuss these aspects in the text of the revised manuscript version and put the existing approach in this context.

Following the two reviewers' comments, we will sharpen the motivation and goal of the presented study substantially in order to meet the expectations. The overarching goal is to document the specific implementation of a multi-site weather generator suggested in the literature, validate this implementation for the current climate and apply the generator as a statistical downscaling tool for future local climate. The latter will be presented in a separate, additional study currently in preparation. We therefore deliberately chose a simple tool that can be easily adapted in a climate change context, thereby complementing and improving existing climate change scenarios for Switzerland. Perturbing the generator for a future climate is part of the subsequent article. The aim of the article under revision here is to test and evaluate the implementation of the multi-site precipitation generator under current climate conditions in order to understand its capabilities and caveats. In the revised manuscript we will better motivate our study in the abstract, the introduction and through the text.

We agree with the reviewer, that the novelty of the presented article lies not in the generator-tool itself, but rather in its concrete implementation and application to a Swiss catchment including the presentation of the complex pathway to calibrate it. Following the comment of referee #2, we realized that our manuscript title does not well reflect this aspect and might even evoke wrong expectations among the readers. We have therefore changed the title to *"Implementation and validation of a multi-site daily precipitation generator over a Swiss river catchment"*.

Specific Comments

Comment 1: *This model captures the mean behavior relatively correctly, but does not capture extreme events properly. The mixture of exponential variables, which describe precipitation amounts, is not only unable to reproduce observed extremes (see Fig. 9), but it is certainly even worse at predicting future (non-observed) extremes, owing to the light tail of the assumed exponential variable…The model advocated by Vrac and Naveau (2007), which uses a generalized Pareto distribution (GPD) for the tail, has an asymptotic justification and thus much better at estimating probabilities for rare events. If the goal of the proposed WG is to be used in impact and risk assessment studies (as claimed several times in the paper), possibly in a climate change context, I think that extreme events should be better captured.*

Reply 1: We agree that the proposed mixture model of Vrac and Naveau (2007) better represents daily extreme precipitation amounts. However, fitting this model to daily precipitation amounts at single months introduces a large uncertainty, since the proposed model relies on 5 parameters (gamma shape, gamma scale, threshold, pareto scale and pareto shape). Additionally, although the GPD distribution alleviates the underestimation of daily extreme events, it likely does not improve the reported underestimation of multi-day precipitation sums (Fig. 9). This is because the precipitation amount model currently does not include any autocorrelation. In our view, this aspect is of even greater concern than the underestimation of daily extremes. However, improving our precipitation generator with regard to multi-day extremes goes beyond the scope of this study and will be part of future work. We added a sentence to Section 6 (Summary and outlook), giving a perspective on this aspect.

Concerning further model development, we plan to refine the presented WG in a future study by using multi-state Markov chains (e.g. dry, wet and very wet) in combination with different probability density functions for wet and very wet days in order to improve its main deficiencies. The use of this approach has the advantage to better capture one-day and multi-day extremes as well as to include a temporal memory in the precipitation amount process. We have added a sentence in Section 6.

The reviewer is right that care should be taken when using our simulated time-series as a data basis for any risk and impact assessment studies. We reformulated those sentences concerning this aspect.

Comment 2: *The proposed WG simulates stationary time series (month by month). Does this make sense over a period of 51 years (and for the near future)? A non-stationary model could perhaps explain part of the inter-annual variability, that is not captured in Fig 4. And also, again, if the goal is to use this WG for risk and impact assessment, does it make sense to assume that climate is stationary? Can the WG be used to extrapolate in the future?*

Reply 2: Our goal is to develop a simple tool that generates time-series consistent with mean conditions. In addition, it should be easily adjustable for future climate mean conditions. In a climate change context (to be published in a follow-up article) we will use standard climatological periods of 30 years. In this study, however, we chose a relatively long time-period of 51 years, in order to accurately assess the added value of a multi-site generator against multiple single-site generators. Reducing the time-window of calibration increases the sampling uncertainty (as shown in Figure 3 with artificial data) and hence the determination of added value becomes more uncertain, too. In the revised manuscript (in Section 3.4.1) we will include this reasoning for the particular choice of time-period.

Nevertheless, introducing a non-stationary model would certainly be a valuable extension to the current generator. Although precipitation over the Thur catchment does not feature a trend over the 51 years, one could sample from the observed interannual variability in the WG-parameters and add henceforth an additional stochastic component. This direction of future development is briefly mentioned in the revised manuscript (Section 6).

Comment 3: *The model captures spatial coherency between monitoring stations, but does not describe spatial dependence at ungauged locations. Therefore, it is impossible to simulate precipitation data over the whole catchment, which may be essential for risk assessment (e.g., if the simulated data are needed as input of a hydrological rainfall-runoff model). All pairwise correlations between the different stations are computed empirically, although there is a large geostatistics literature about Gaussian processes, (stationary or non-stationary) correlation functions, etc. Why not fit a correlation function to the data, which would: - allow simulation over the whole catchment, - decrease the number of parameters drastically (therefore also the uncertainty) by exploiting the inherent spatial structure, - automatically yield positive definite correlation matrices (without any adjustment), - easily generalize to many more time series?*

Reply 3: We agree that the suggested geostatistical model is an interesting alternative approach but unfortunately it is not compatible with our proposed model for the following reasons:
(a) The number of parameters in our model is certainly large and a correlation function would reduce this number drastically. However, from a theoretical point of view, geostatistical models (be it isotropic or anisotropic models) rely on the assumption of stationarity, which is not sufficiently fulfilled in regions with a complex topography (e.g. Schiemann et al., 2010), such as the Thur catchment in this study. In addition, from a practical point of view we see some problems in estimating a correlation function. Since we only use a small number of stations (8) the uncertainty of fitting a correlation function would

be substantial. As it can be seen from the Correlogram (see Fig. 1 below) the correlation (*p*) shows large variations for all inter-station distances (*h*).

(b) The potential to simulate precipitation also at ungauged location is obviously appealing. However, the specification of a correlation function would not be enough. Additionally, further parameters of the precipitation generator (i.e. transition probabilities and parameters of the probability density function) would have to be interpolated consistently in space. This is not straightforward given the complex topography.

(c) Concerning the problem with the positive definite correlation matrices, a correlation function could not alleviate this problem. This is because it is the pair-wise optimization process (see chapter 3.3.2) that yields non-positive correlation matrices and not the estimation of the observed correlation matrices. We clarified this aspect in the revised manuscript in chapter 3.3.2.

Nevertheless, we think it is certainly worth mentioning this alternative modelling approach in the discussion part (Section 5) of our revised manuscript together with an explanation why we have decided against it.

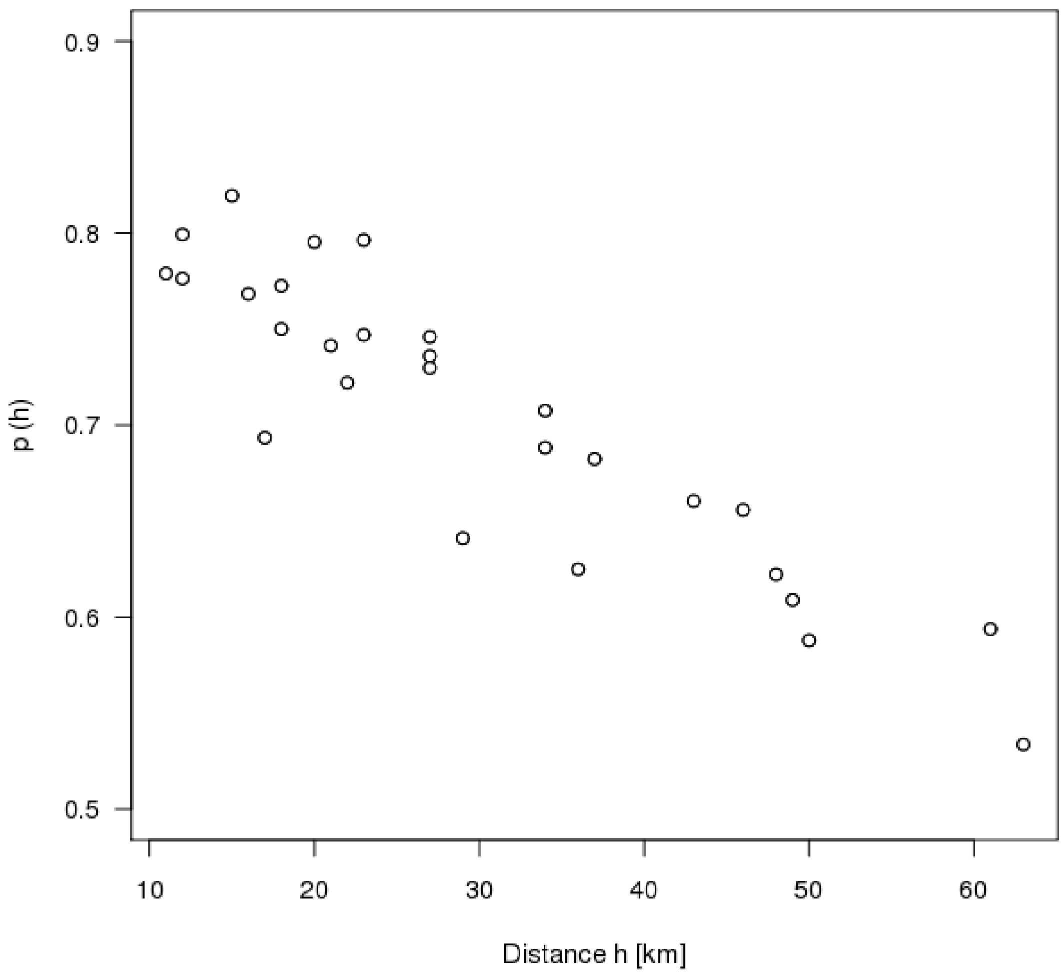

Figure.1: Correlogram for the investigated eight stations of the "Thur" catchment. It describes the spatial correlation p as a function of the inter-station distance h. Each point refers to a station pair. The correlations between the stations was calculated over a time-period of 1961-2011.

Comment 4: *The number of parameters in the model is very large, so simplifications should be considered, e.g., by - fitting a correlation function (see point (3)). This would decrease the number of spatial dependence parameters, though spatial heterogeneity might be difficult to take into account. - fitting a global yearly model, for example using splines or sines/cosines (instead of fitting separately one model per month). This would decrease the number of temporal dependence parameters, and yield a coherent model throughout the whole year.*

Reply 4: We agree with the reviewer that fitting a global yearly model would decrease the number of parameters, for instance by specifying a GLM with cos/sin harmonic functions as co-variates as e.g. described in Furrer and Katz (2007). However, since our generator is later subject to be perturbed for future climate using a delta change approach, we would rather stick to empirical estimates of these parameters for each month separately. In a GLM-context, the prediction of future WG-parameters would have to be done using a number of co-variates (such as e.g. in Beuchat et al. 2012), assuming stationary link functions between the predictors and the predictand. Furthermore, the selection of the (large-scale) predictors is far from trivial. Especially in regions where local effects dominate the precipitation process (e.g. Alps) there is a risk of model over-fitting. Also, to accurately capture the seasonal cycle in the climatic change of certain precipitation characteristics (e.g. wet-day frequency) would be a huge challenge. For these reasons, we opt here for the simpler solution.

Furthermore, the proposed model development would not substantially improve the model in terms of its main deficiencies: i.e. representation of extremes, too low inter-annual variability, and underestimation of multi-day spell lengths. For future research, we would rather spend resources in improving these main deficiencies.

Regarding the spatial correlation function, we refer to reply 3.

Comment 5: *The Markov chain of order 1 does not capture well multi-day spells (see Fig. 7, for example). Maybe, a 2nd-order Markov chain (AR(2)) would do a better job. . . Or maybe a model of type 'ARMA(1,1)' would fit better? Of course, the number of parameters would increase if a more complex model is considered, but this would also better capture long-range dependencies. . .*

Reply 5: To select the order of the Markov chain model we consulted the Bayesian information criterion (BIC). Based on this analysis we did not find a substantial improvement of 2^{nd}-order against 1^{st}-order MC (see Figure 2 below). We therefore opted for the simpler model to also limit the number of parameters. Given a rather short sample size, the risk of introducing large uncertainty would be much higher for a 2^{nd}-order model compared to a 1^{st}-order model. For further details we refer to our reply to your technical comment # 2.
Note, that selecting a higher order Markov model does not alleviate the underestimation of long multi-day spells to a full extent. To improve the duration of multi-day spells, the model would have to be conditioned on other atmospheric variables (in particular circulation-related) (Chandler and Wheater, 2002). This however is out of scope of the present study. Finally, if a correct reproduction of long-lasting spells is the main focus, spell lengths generators (e.g. Racsko et al., 1991) might be the more appropriate way forward.

Regarding the suggestion of specifying an ARMA (1,1) model, in our view this would be inappropriate, since we are dealing here with discrete data (dry/wet) rather than with continuous time-series.

Comment 6: *In Section 4, the authors validate their WG by looking at different temporal or spatial statistics, such as long-term mean, inter-annual variance, PDF of non-zero precipitation, dry and wet spells, annual maximum sums of consecutive days, etc. However, I guess that there is no validation of space-time interactions. For example, if Z(s,t) denotes the precipitation amount at station s and time t, a possibility would be to see if the model and the data agree on statistics of the type 'X =*

P(Z(s2,t+k)>z | Z(s1,t)>z)' for increasing values of z? Here, the statistic X represents the probability that it rains at least z mm at station s2, given that it rained similarly at station s1, k days earlier. Have the authors checked this kind of space-time dependencies?

Reply 6: We have not validated this aspect. Since the investigated catchment is of rather small size, it is very likely that rain occurrence or non-occurrence will be recorded at all stations simultaneously on the same day (see also Table 1 in the manuscript). Certainly, if we have a larger catchment, e.g. those of the Rhine river, it would be interesting and worthwhile to investigate these time-space statistics. A catchment that is frequently affected by frontal passage of rain, this would be a very relevant aspect of analysis.

We added a sentence to the discussion (Section 5) of the revised manuscript.

Comment 7: *The title of the paper is 'Stochastic modeling of [. . .] over complex topography', but the topography information does not appear anywhere in the model. Hence, how could the model be modified in order to incorporate information about altitudes, slopes, etc., and therefore hopefully better predict precipitation at unobserved locations?*

Reply 7: The consideration of topographic effects is a prerequisite for establishing a weather generator that may be applicable comprehensively over a complex topography area (not just multiple sites).Generalized linear models (McCullagh and Nelder, 1989) or Bayesian Hierarchical models (Gelman et al. 2004) are theoretically appealing frameworks that allow modelling of physiographic dependencies into the amount part of our weather generator. This alone is however not sufficient for a space-time weather generator, as the spatial dependence of (daily) precipitation is also determined by spatial autocorrelation not just the physiographic conditioning of parameters. Clearly, the development of a space-time weather generator that deals with spatial autocorrelation, physiographic conditioning, intermittence and temporal autocorrelation is far from trivial and will require fundamental methodological development, before applications can be attempted.

Technical Comments

Comment 1: *p. 8742, L.7: A reference about space-time modeling of rainfall extremes in Switzerland is the following: Huser, R. and Davison, A.C. (2014, JRSS B), "Space-time modelling of extreme events".*

Reply 1: We have inserted this reference in the introduction.

Comment 2: *p. 8743, L.25: The BIC is used to select the order of the Markov chain. However, it is known that it typically over-penalizes complicated models (which might explain why a 2nd order model was not retained). What does the AIC say?*

Reply 2: It is true that the BIC criterion tends to be more conservative than the AIC criterion. We chose the BIC as a model selection criterion for two reasons: First, we deliberately aimed for a simple model. Second, at large sample sizes as here with 51 years of daily data, the AIC tends to select over-proportionally higher-order models (e.g. Katz 1981 or Wilks 1998). Both, the AIC and the BIC show a large improvement when we go from a zero-order to a first-order model. However, the difference when going from a first-order to a second-order is almost negligible (see Figure. 2), which is true for both AIC and BIC.

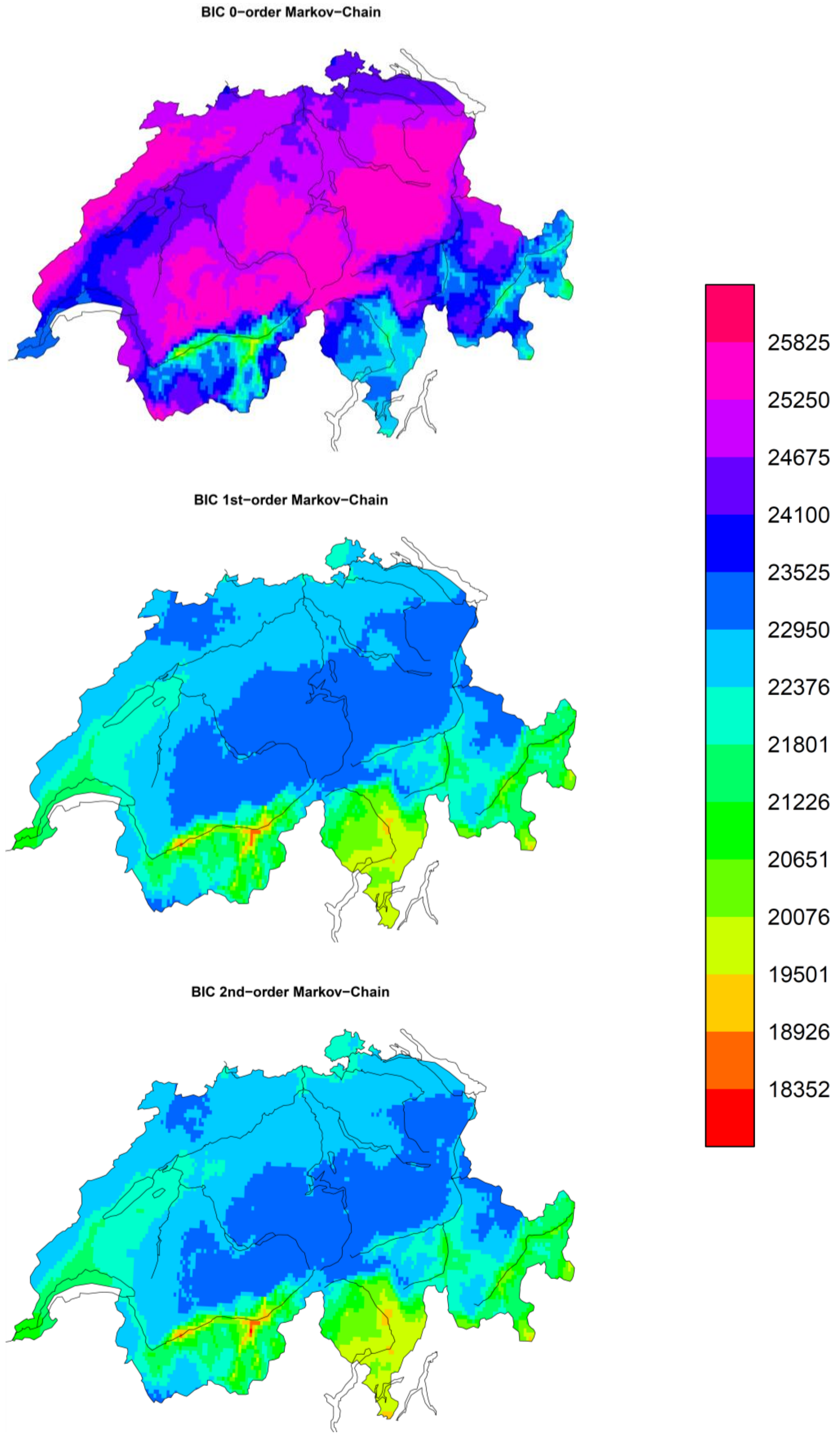

Figure. 2: Bayesian information criterion (BIC) for 0-order (top), first-order (middle) and second-order (bottom) Markov chain over Switzerland. The BIC was calculated over a time-period of 1961-2011 using a daily gridded precipitation data set from MeteoSwiss(Frei and Schär, 1998))

Comment 3: *p. 8744, L.2: A threshold of 1mm day^{-1} was used. How sensitive are the conclusions with respect to this threshold? How many zeros are there?*

Reply 3: The threshold of 1mm/day to distinguish between wet and dry is standard practice for station measurements (see e.g. "Peterson et.al: "http://etccdi.pacificclimate.org/docs/wgccd.2001.pdf", WMO, Rep. WCDMP-47, WMO-TD 1071, Geneve, Switzerland, 143pp"). An analysis of the threshold sensitivity was not performed so far. We expect that our main conclusions about the added value of a multi-site WG also hold for another threshold. This is because the precipitation generator, be it in a single-site or multi-site configuration, is calibrated at individual stations with the same WG parameters (amount and occurrence process) for both configurations. A shift in the threshold would therefore affect both configurations in the same way.

We do not fully understand the question about the number of zeros. In particular, for which season, station, granularity does the reviewer refer to? The number of zeros per given month and station depends on the precipitation regime of the investigated stations. For a high-elevation site such as "Saentis" the wet day frequency ranges from 0.37 in fall up to 0.53 in summer (see Fig. 5 in manuscript). Consequently, the dry day frequency ranges from 0.47 (summer) up to 0.63 (fall), this yields about 15 to 19 dry days (zeros) per month. For a low-elevation site such as "Andelfingen", the wet day frequency is more constant through the year and amounts approximately 0.35 with about 20 dry days per month.

Comment 4: *p. 8745, L.22-23, 'they underestimate [. . .] (gamma distribution)': This sentence is misleading, I think, because the exponential and gamma densities decay at the same rate at infinity, so they are likely to give similar probabilities to extreme events.*

Reply 4: We agree and we will rewrite this sentence.

Comment 5: *p. 8746, L.7-8, 'The parameters [. . .] maximum-likelihood': As already mentioned above, it would be better to have a spatial model linking the parameters, and estimate everything simultaneously (instead of estimating a lot of parameters separately from station to station).*

Reply 5: We refer to our reply to your specific comment # 3. We take up this idea as alternative modelling approach in our discussion part.

Comment 6: *p. 8747, L.7-8: The point 4.2.2 is not very clear to me...*

Reply 6: The point 4.2.2 simply explains how we randomly sample from a mixture of two exponential distributions. In particular, a random number (lying between 0 and 1) is compared to the quantile-function of the mixed distribution to assign the corresponding precipitation amount at a given day. We have rewritten the sentence to clarify.

Comment 7: *p. 8747-8748, S3.3.2: How are the correlation matrices estimated? Empirically? If so, this might induce problems if the number of stations is large, and also it does not ensure that the correlation matrices are positive definite. A better solution is, as explained above, to assume and estimate a correlation function.*

Reply 7: Yes, the correlation matrices were estimated empirically on a monthly basis. Indeed, we encountered problems when including more than 12 stations. The reviewer is also correct that the chances are higher that the correlation matrices are not positive definite. We have reported on both these problems in the original manuscript (page 8750, line 17-18). These limitations are certainly a strong downside of the proposed generator. We aim at improving the generator in this direction in future work.

Comment 8*: p. 8750, 19-20, 'In absence of [. . .] matrix was chosen': How was this fall-back solution implemented? By minimizing a certain norm? If so, which one?*

Reply8: The nearest positive definite matrix was found by applying the R-function "nearPD" from the R-package "matrix". This function uses the algorithm proposed by Higham (2002), which uses a weighted version of the Frobenius norm. In the revised manuscript, we included an additional sentence to clarify this issue.

Comment 9: p*. 8751, S3.4.1: This section shows that the model has a lot of parameters, and that it is crucial to reduce this number to avoid huge uncertainties and optimization issues...*

Reply 9: We partly agree with the reviewer. For the replies we refer to your specific comment # 3 and # 4.

Comment 10: *p. 8752, L.18, 'roughly 19': In fact the theoretical reduction is (asymptotically) sqrt{10000/30} = 18.3. . .*

Reply 10: Yes. Thanks. We will substitute 19 by 18.3 at the indicated position.

Comment 11: *p. 8754, L.27, '33%': This percentage seems very low! This might be due to nonstationarities that the model is unable to capture, or simply because the model strongly underestimates probabilities of extreme events.*

Reply 11: Indeed, the percentage is low but comparable to other studies applying a similar WG model (e.g. Gregory et al. 1993) as mentioned in the manuscript. We agree, that the underestimation of inter-annual variability in the monthly sums is attributed to the stationarity assumption in the annual cycle of the WG parameters. We consider to highlight the existing discussion of this issue more prominently (see section 6). Also, in the outlook-section we have outlined how this problem could be circumvented (page 8762, line 26).

Regarding the underestimation of extreme events, we suppose there are several issues contributing to this problem. The precipitation generator not only misses the large observed precipitation sums, but also misses to capture the small observed amounts of precipitation sums (compare violet and grey shading in Figure 4). The variability underestimation is therefore more likely caused by forcing our generator solely by mean conditions, whereas in real-world there are large fluctuations from year-to-year.

Authors reply on comment of referee # 2

Comment: *I have several problems with this scientific article. The first one is that this looks more like a class rather than a scientific paper. Authors expend a lot of time describing single site weather generators for nothing. All that is repetition and it can be found everywhere. The second problem is that authors are very superficial in their literature review and over simplified the publications already published. I'm extremely sure that they haven't read the papers they mention in the review, so they have no idea of the detailed method developed in each of these previous approaches. Another problem is that the title made me think that they developed something especial for complex topography areas, something the other developer scientist didn't take into account. In fact, authors only apply the weather generator in some weather stations in the mountains. Some authors have done the same and published their work without advertising the issue. The methodology per se is nothing new. Authors propose no new ideas, no new methods, and no nothing. This is an application of methods already published and they add nothing to the science of stochastic modelling. Unfortunately, from my point of view, this proposed article is not to the level of scientific publication.*

Reply: The authors would like to thank referee # 2 for his/her efforts.

The main critics of Referee #2 is the novelty of the article. We understand this point in the sense that our precipitation generator is an implementation and calibration of a generator proposed in the literature by Wilks and others. We mention this at several locations in the original manuscript. The novelty lies in its application to a catchment over Switzerland with a cascade of methodological challenges to calibrate and implement it. We are convinced that it is of scientific relevance to document this specific implementation, i.e. the "lessons-learnt", in order to make use of this weather generator in further studies and to make it available to other groups. Having said that, we acknowledge that our manuscript-title may have been misleading (as claimed by the reviewer). From the title alone, one would expect a novel approach on stochastic modeling taking into account topographical information, which is not the case here. In the revised version of the manuscript, we have changed the title to: "*Implementation and validation of a multi-site daily precipitation generator over a Swiss river catchment*". In addition, we fully acknowledge that the purpose of the study needs to be better motivated in the abstract, introduction and conclusions. The presented precipitation generator is mainly developed to be used in a subsequent study as a statistical downscaling technique for a future climate (to be published in a follow-up article in preparation). Therefore, we deliberately chose a tool which is sufficiently easy to handle also in a climate change context. The main aim of the article under review is to describe the generator, its implementation and validation for current climate. We have largely rewritten the abstract and introduction with these arguments. Note, that some of the citations in the original manuscript have been omitted in the revised version.

Regarding the point concerning the review of existing literature, we realize that we indeed missed to cite some relevant literature (see also comment by the editor), in particular from the hydrology-community. We apologize for that and have included in the revised version citations to the following stochastic precipitation-modelling approaches:

Huser, R., & Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 439–461. doi:10.1111/rssb.12035

Mezghani, A., & Hingray, B. (2009). 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. *Journal of Hydrology*, *377*, 245–260. doi:10.1016/j.jhydrol.2009.08.033

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

Most of the other comments of referee #2 are difficult to address, since the critical points are not substantiated. Examples are the claim that we have not read the papers and that we have no idea about the published methods in detail. We strongly object to this substantial allegation of non-scientific practice.

It is rather difficult, if not impossible, to respond point-by-point to these review comments without more details on what the referee is specifically criticizing. We note that the journal guidelines for reviewers contain the following paragraph:
“Referees should explain and support their judgments adequately so that editors and authors may understand the basis of their comments. Any statement that an observation, derivation, or argument had been previously reported should be accompanied by the relevant citation.http://www.hydrology-and-earth-system-sciences.net/review/obligations_for_referees.html.

In general, we have to assume that referee #2 criticizes similar points as referee #1 and hence we hope that our reply to referee #1 also addresses the majority of the comments by referee #2.

List of all relevant changes *(page and line numbers refer to the marked-up version of the revised manuscript)*

Page	Line	Changes
1	1-4	Title has been changed to *Implementation and validation of a multi-site daily precipitation generator over a Swiss river catchment*
1 2	16-28 1-24	Abstract has been completely reformulated to better motivate the study
2 3 4 5	26-31 1-32 1-31 1-10	Introduction has been completely reformulated to better motivate the study
4	4-10	Additional citations about time-space WGs included
4	15	Additional citation about extreme precipitation generator included
6	17-22	Explanation about the choice of the Markov model order of occurrence model included
9	10-12	Sentence was re-written to clarify
10	23-25	Sentence was added to clarify issue about non-positive definite correlation matrices
12	5-14	Sentences were re-written to clarify how positive definite matrices were calculated in case of the fall-back solution
12	20-22	Sentence was added to explain why we used a calibration time-period of 51 years
20	13	A discussion section was added
20	26-30	Sentences about time-space validation in large catchment is added
21	1-21	Sentences about alternative time-space WGs was added, including field generators and geostatistical models

21	22-32	Summary and outlook section was reformulated:
22 23	1-30 1-5	Key findings of this study and limitations have been reformulated
23	8-18	Sentences about further model development and improvement has been included
23	19-26	Sentences about further model extensions has been included.

~~Stochastic modelling of spatially~~Implementation and ~~temporally consistent~~validation of a multi-site daily precipitation ~~time-series~~generator over ~~complex topography~~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] {~~Climate Physics,~~ 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 ~~over topographically complex terrain~~ require high-resolution precipitation time-series that have a spatio-temporal correlation structure consistent with observations~~. This consistency is essential~~, for ~~spatially distributed modelling of processes with non-linear responses to precipitation input (e.g. soil water and river runoff modelling).~~simulating either current or future climate conditions. In this ~~regard~~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 ~~that approximate~~ consistent with the ~~observed temporal and spatial dependencies.~~local precipitation characteristics and its expected future changes. In this study, we present ~~a stochastic~~ the implementation and validation of a multi-site ~~precipitation generator and validate it over the hydrological catchment *Thur* in the Swiss Alps. The model~~daily precipitation generator following ideas of Wilks (1998). The generator consists of

several Richardson-type WGs ~~that are~~ run with spatially correlated random number streams ~~reflecting~~. We investigate the applicability of the ~~observed correlation structure among all possible station pairs. A first-order two-state Markov process simulates intermittence of daily precipitation, while precipitation amounts are simulated from a mixture model of two exponential distributions. The model is~~ 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 ~~separately for each month over the time-period 1961-2011.~~

~~The~~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~~: the multi-site WG is capable to capture about 95% of the observed variability in~~. Limitations are seen in reproducing daily ~~area~~and multi-day extreme precipitation sums, ~~while the summed time-series from multiple single-site WGs only explains about 13%. Limitation of the WG have been detected in reproducing~~ observed variability from year to year~~, a component that has not been considered in the WG calibration.~~ and in reproducing long dry spell lengths. Given the ~~obtained~~ performance~~,~~ of the presented ~~stochastic model is expected to be~~ generator, we conclude that it is a useful tool to ~~re-sample the observed record and valuable to be used as a statistical downscaling method in a~~generate precipitation series consistent with the mean aspects of the current and future climate ~~change context~~.

1 Introduction

In Switzerland, precipitation is a key weather variable with high relevance for sectors such as energy production, infrastructure, tourism, ~~security,~~ agriculture and ecosystems. Owing to a complex topography, daily precipitation varies strongly in space and ~~in~~ time (Frei and Schär, 1998; Isotta et al., 2013). The spatial distribution of daily precipitation frequency and intensity ~~clearly~~ 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 ~~until~~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. 0 introduces the hydrological catchment of the river *Thur* together with the used station data. In Sect. 0 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. ~~We conclude the article by~~5 includes a discussion and finally, Sect. 6 provides a summary and an outlook ~~(~~.~~Sect.~~ 6~~).~~

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 ~~our~~this study~~ here~~, we selected eight evenly distributed measurement stations (Figure 1a) of MeteoSwiss that ~~meet several requirements: (a) they~~ all provide homogenized time-series covering a 51-year period from 1961-2011 (Begert et al., 2003), ~~(b) they~~and that sufficiently ~~reflect~~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

~~Core~~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. ~~Based on earlier work by~~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

~~Our procedure to~~To model occurrence at a single station ~~is based~~we rely on ~~the concept of~~ 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 Bayesian information criterion~~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)~~ revealing~~. Both the AIC and the BIC revealed a ~~better match compared~~ substantial improvement when going from a zero-order to a first-order model, but the additional gain at a second- or higher-~~-~~order ~~Markov models~~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 ~~dichotomise~~ subsequently ~~into~~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 ~~to be distinguished~~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*).

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

$$p_{11} = P\{J_t = 1 \mid J_{t-1} = 1\}$$
$$p_{01} = P\{J_t = 1 \mid J_{t-1} = 0\} \qquad (1)$$

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

$$p_{11} + p_{10} = 1$$
$$p_{00} + p_{01} = 1 \qquad (2)$$

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):

$$\hat{p}_{01} = \frac{n_{01}}{n_{0\bullet}}$$
$$\hat{p}_{11} = \frac{n_{11}}{n_{1\bullet}} \qquad (3)$$

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):

$$\pi = \frac{p_{01}}{1 + p_{01} - p_{11}} \qquad (4)$$

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

$$r_1 = p_{11} - p_{01} \qquad (5)$$

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:

$$r_L = (p_{11} - p_{01})^L \quad (6)$$

3.2 Precipitation amount process

As will be detailed in Sect. 3.3, ~~at wet days,~~ 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:

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

$f(x)$ 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 ~~This random number is then substituted by the~~The corresponding quantile ~~(x_A)~~ of the ~~cumulative distribution~~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 ~~later introduced~~ multi-site extension ~~(Sec.~~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 ~~used the concept of the~~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 ~~inter-station correlations~~the applied pairwise optimization process (see section (2) above) this is not always fulfilled~~ and depends on the number of stations with incomplete records~~. 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 ~~the~~for smaller ~~the~~ sample ~~size is~~sizes; in fact, the uncertainty range ~~enlarges~~increases by a factor of ~~roughly 19~~18.3 when the sample size is reduced from ~~a sample size of~~ 10000 ~~down~~ 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):

$$E[S(T)] = T \cdot wdf \cdot wdi \tag{8}$$

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).

~~Let us now~~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 ($Var[X \geq 1\ mm\ day^{-1}]$) and the variance of the number of wet days (*Var [N(T)]*) as proposed by Wilks and Wilby (Wilks and Wilby, 1999):

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

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. ~~(5)).~~(6)). Similar to the underestimation of variability in precipitation sums, higher-order Markov chains ~~(Wilks, 1999a)~~(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. ~~The~~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 ~~our~~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 ~~our~~the calibrated WG in multi-site and single-site mode, ~~Fig. 8~~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. ~~An~~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.

~~4.2.4~~5 Discussion

Formatted: Heading 1

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.

~~5~~6 Summary and Outlook

~~A multi-site daily precipitation generator~~The multi-site precipitation generator of Wilks (1998) has been successfully developed, implemented and tested over the Swiss alpine river catchment *Thur*. ~~The generator is built after suggestions by Wilks (1998). Core of our multi-site precipitation generator is a Richardson-type WG with simulation of daily precipitation occurrence as a chain dependent process and simulation of~~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 ~~dependencies between the stations are imposed~~dependency is ensured by running the ~~precipitation models~~WG with spatially correlated ~~but serially independent~~ 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 ~~inter-~~compared to the observed record and to simulations based on multiple ~~(~~independent~~)~~ single-site WGs.

Our main findings of this study are:

- ~~Our developed~~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.
- ~~Based on its good performance in a range of spatio-temporal precipitation aspects, our weather generator is expected to serve as a helpful data provision tool for multiple applications including climate change assessments.~~
- 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 ~~over heterogeneous catchment areas such as the *Thur* catchment:~~

  ~~(a)~~. The median of daily area sums are higher by about a factor of 1.3 ~~higher~~ 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%.
- ~~(b)~~ 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 ~~expected to become most distinct~~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 ~~give us~~provide confidence that the developed precipitation generator is a ~~very~~ helpful tool to realistically simulate mean aspects of the current climate. ~~Nevertheless~~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 ~~remain: the~~ need to kept in mind: The

Formatted: Bulleted + Level: 1 + Aligned at: 0.63 cm + Indent at: 1.27 cm synthetically generated time-series (for current or future climate) do not fully capture the day-to-day and multi-day variability of precipitation~~ to a full extent.~~. Extreme values and longer spell lengths are hence underestimated~~ and should not be the focus of any such analysis with the data at hand. Furthermore, our generator ~~. 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 ~~problem ~~could be ~~alleviated~~accomplished by sampling ~~the input ~~from a distribution of observed WG parameters ~~from the observed distribution ~~(instead of ~~solely ~~taking the ~~best estimate. A more sophisticated way would be to use a model that incorporates ~~mean) or by formulating a regression model using large-scale atmospheric variables as predictors ~~to estimate~~(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~~, such as for instance demonstrated by Furrer and Katz (2007) using Generalized linear models (GLMs).~~ 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.

~~In light of these inherent limitations, care should be taken when using the generated time-series as basis for a comprehensive risk assessment of different climatic impacts. To increase robustness in our results here, the generator should be ideally applied to further catchments of different sizes and in different time-periods. This would entail a better quantification of the benefits and limitations. In any case, the presented generator is subject to further developments, including the extension to a multi-variate weather generator and its adaptation~~

~~for climate change studies. If proven skilful, it is planned to use the weather generator as a downscaling technique to simulate spatially and temporally consistent daily precipitation time-series at the local scale consistent with large-scale climate model projections of a future climate.~~

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.

~~Baigorria, G. A. and Jones, J. W.: GiST: A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data, J. Clim., 23(22), 5990–6008, doi:10.1175/2010JCLI3537.1, 2010.~~

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.

~~Brissette, F. P., Khalili, M. and Leconte, R.: Efficient stochastic generation of multi-site synthetic precipitation data, J. Hydrol., 345(3-4), 121–133, doi:10.1016/j.jhydrol.2007.06.035, 2007.~~

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~~.: 2.1 Bisection Method, in~~. D.: Numerical Analysis, ~~pp. 48–55~~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.

~~Dubrovsky, M.: CREATING DAILY WEATHER SERIES WITH USE OF THE WEATHER GENERATOR, Environmetrics, 8, 409–424, 1997.~~

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.

~~Hirschi, M., Stoeckli, S., Dubrovsky, M., Spirig, C., Calanca, P., Rotach, M. W., Fischer, A. M., Duffy, B. and Samietz, J.: Downscaling climate change scenarios for apple pest and~~

~~disease modeling in Switzerland, Earth Syst. Dyn., 3(1), 33–47, doi:10.5194/esd-3-33-2012, 2012.~~

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.

~~Lall, U. and Sharma, A.: A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series, Water Resour. Res., 32(3), 679–693, doi:10.1029/95WR02966, 1996.~~

~~Leander, R. and Buishand, T. A.: A daily weather generator based on a two-stage resampling algorithm, J. Hydrol., 374(3-4), 185–195, doi:10.1016/j.jhydrol.2009.06.010, 2009.~~

~~Maraun, D., Osborn, T. J. and Rust, H. W.: The influence of synoptic airflow on UK daily precipitation extremes. Part I: Observed spatio-temporal relationships, Clim. Dyn., 36(1-2), 287–301, doi:10.1007/s00382-011-1176-0, 2009.~~

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~~.1.INTRODUCTION~~, 2010.

~~Mehrotra, R~~McCullagh, P. and ~~Sharma, A.: A nonparametric stochastic~~ Nelder, J. A.: Generalized linear models., Chapman and Hall, London England., 1989.

Mezghani, A. and Hingray, B.: A combined downscaling~~ framework~~-disaggregation weather generator for ~~daily rainfall at multiple locations, J. Geophys. Res., 111, D15101~~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.~~1029/2005JD006637, 2006~~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.

~~Rajagopalan, B. and Lall, U.: A k -nearest-neighbor simulator for daily precipitation and other weather variables, Water Resour. Res., 35(10), 3089–3101, doi:10.1029/1999WR900028, 1999.~~

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

~~Richardson, C. W. and Wright, D. A.: WGEN : A Model for Generating Daily Weather Variables, USDA-ARS., 1984.~~

~~Rodríguez-Iturbe, I., de Power, B. F. and Valdés, J. B.: Rectangular pulses point process models for rainfall: Analysis of empirical data, J. Geophys. Res., 92(D8), 9645, doi:10.1029/JD092iD08p09645, 1987.~~

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

~~Semenov, M. A. and Barrow, E. M.: Use of a stochastic weather generator in the development of climate change scenarios, Clim. Change, 35, 397–414, doi:10.1023/A:1005342632279, 1997.~~

~~Srikanthan, R. and Pegram, G. G. S.: A nested multisite daily rainfall stochastic generation model, J. Hydrol., 371(1-4), 142–153, doi:10.1016/j.jhydrol.2009.03.025, 2009.~~

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.

~~Zucchini, W. and Guttorp, P.: A Hidden Markov Model for Space-Time Precipitation, Water Resour. Res., 27(8), 1917–1923, doi:10.1029/91WR01403, 1991.~~

Formatted: English (U.K.)

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

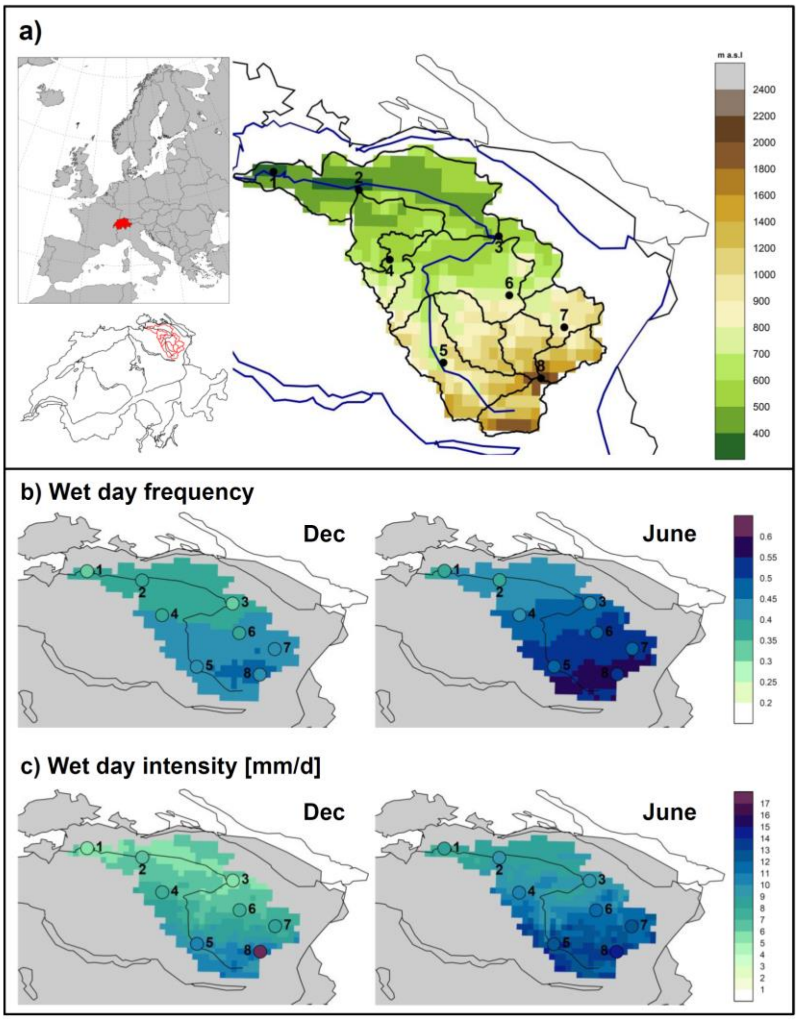

Figure 1. a) The catchment of the river *Thur*, located in north-eastern Switzerland, together with the underlying topography (in m.a.s.l.). The dots indicate the locations of the investigated stations. 1: *Andelfingen* (AFI), 2: *Frauenfeld* (FRF), 3: *Bischofszell* (BIZ), 4: *Eschlikon* (EKO), 5: *Ebnat-Kappel* (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.

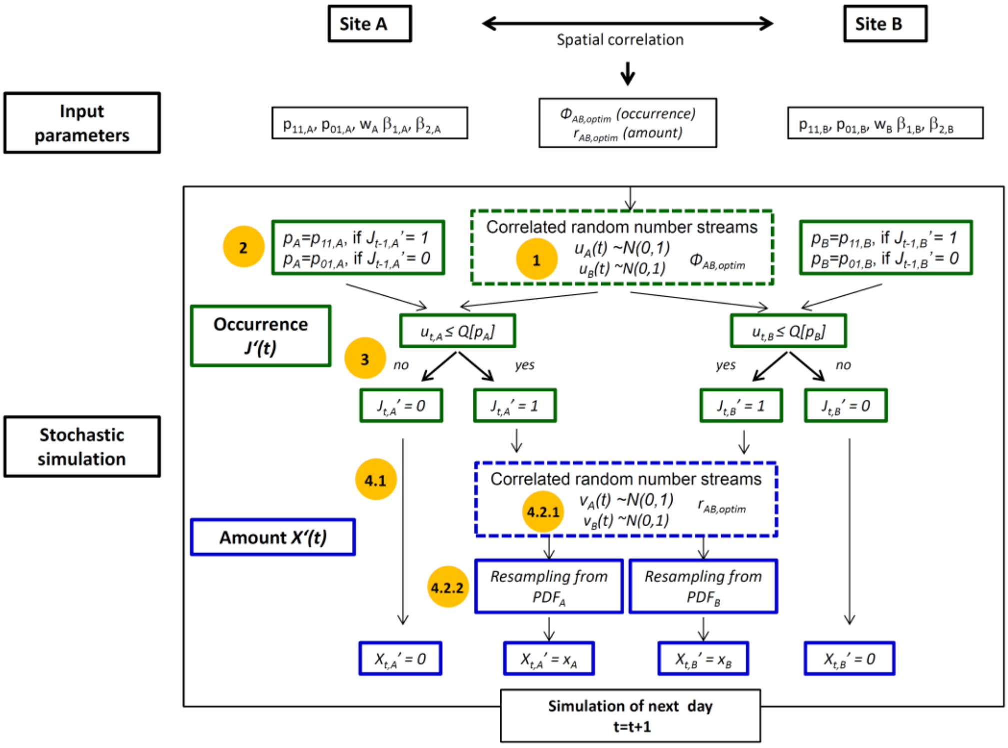

Figure 2. Technical workflow of a multi-site precipitation generator after Wilks (1998) at the example of two fictitious sites *A* and *B*. In general, it is a combination of multiple single-site precipitation generators that are calibrated at each site individually (see input parameters) and run simultaneously 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).

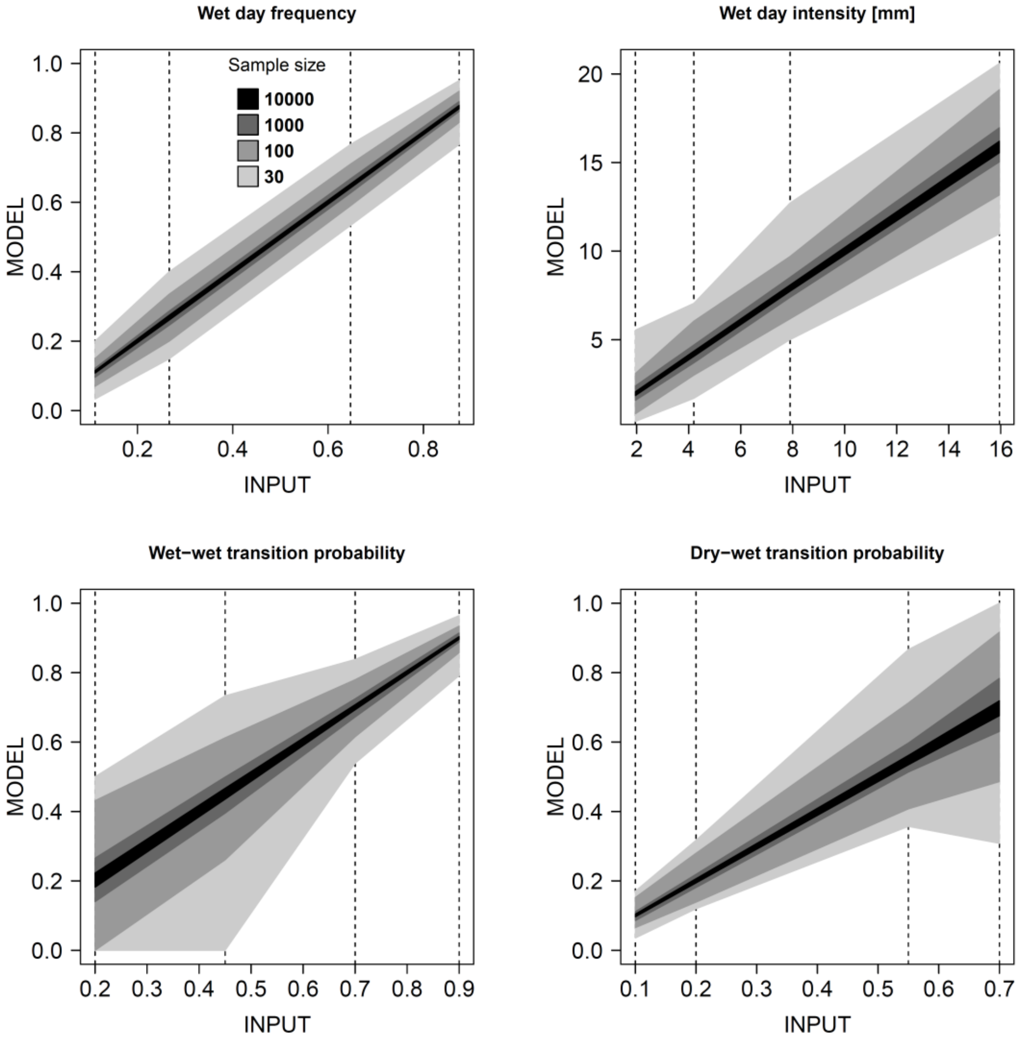

Figure 3. Reproduction of average wet day frequency (wdf), mean wet day intensity (wdi), wet-wet transition probability (p_{11}) and dry-wet transition probability (p_{01}) for the four idealized climate regime ranging from very dry (left) to very wet (right) as indicated by dashed lines. The shaded 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).

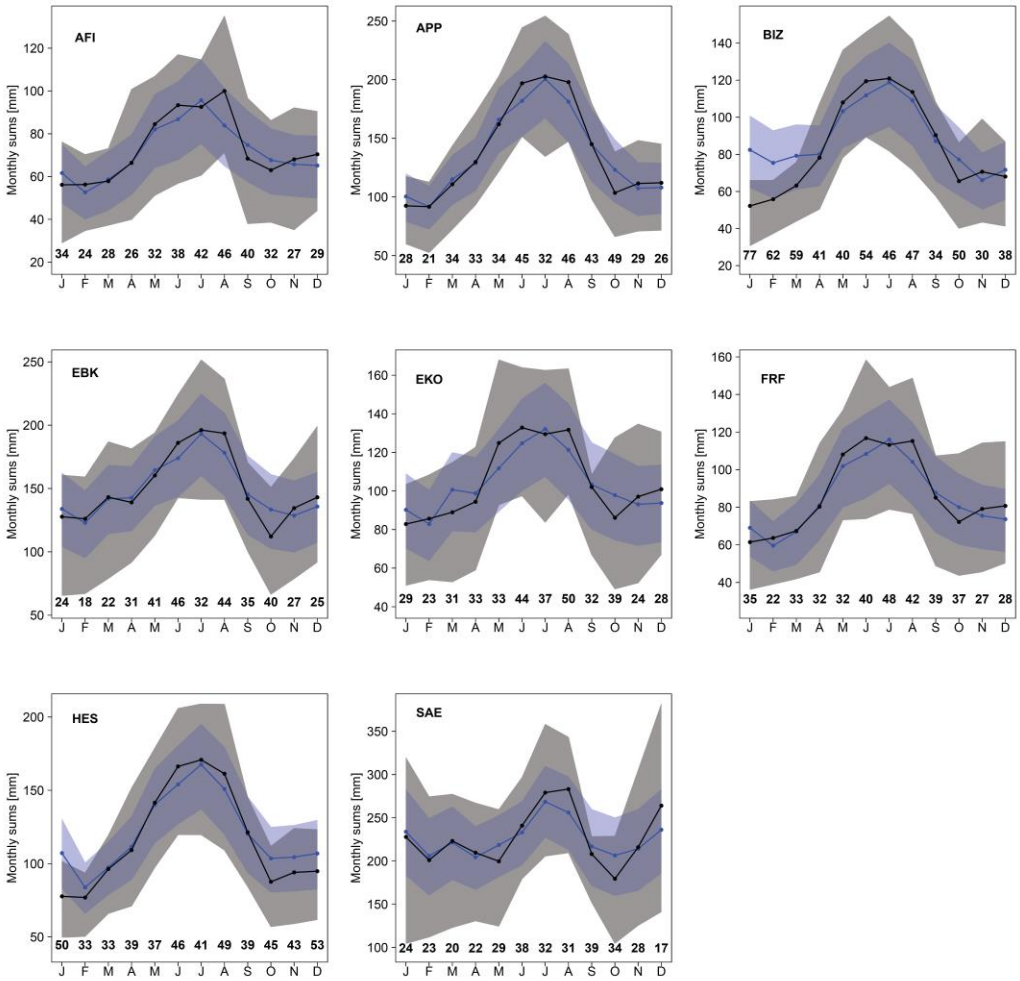

Figure 4. Long-term mean and variability of monthly precipitation sums during the period 1961-2011 for eight stations in the *Thur* catchment. The black (blue) lines refer to the mean annual cycle of observed (modelled) precipitation sums. The grey (blue) shaded areas represent the inter-quartile ranges 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.

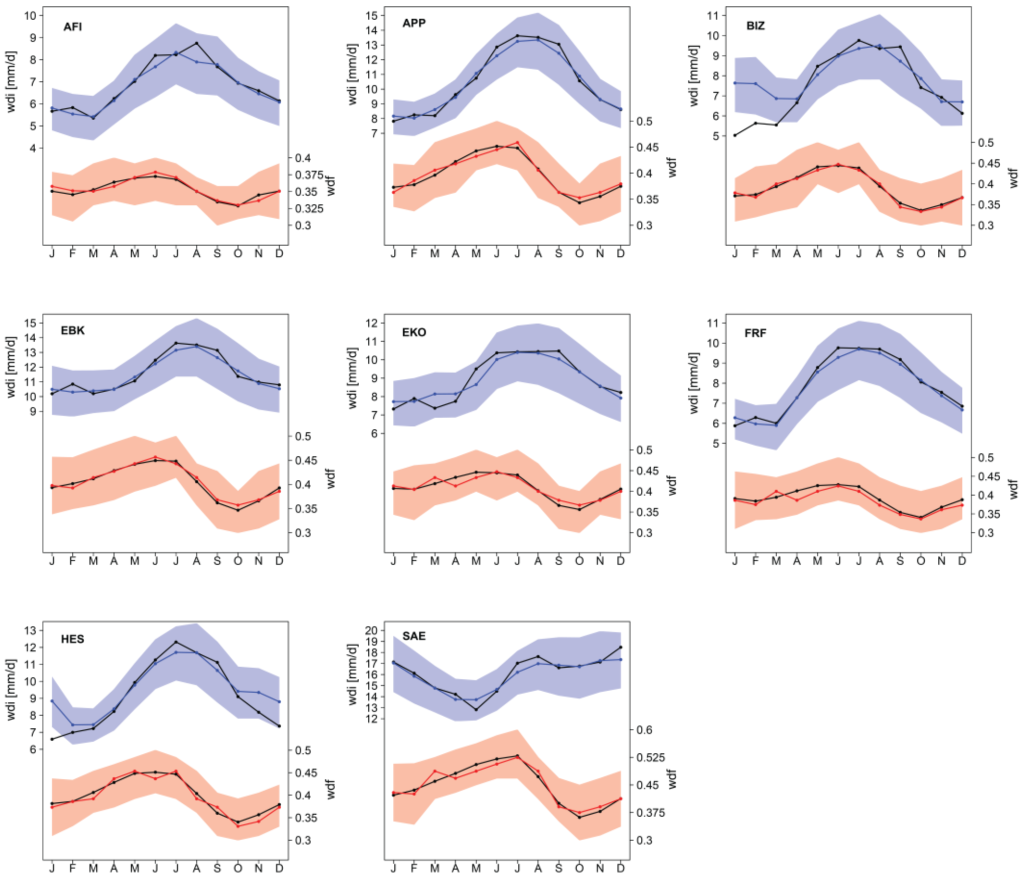

Figure 5. Observed and modelled monthly mean wet day intensity (blue) and frequency (red) at eight stations during 1961-2011. The black (coloured) lines indicate the observed (modelled) values. The blue (red) shaded areas correspond to the inter-quartile range across the set of synthetic daily time-series.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

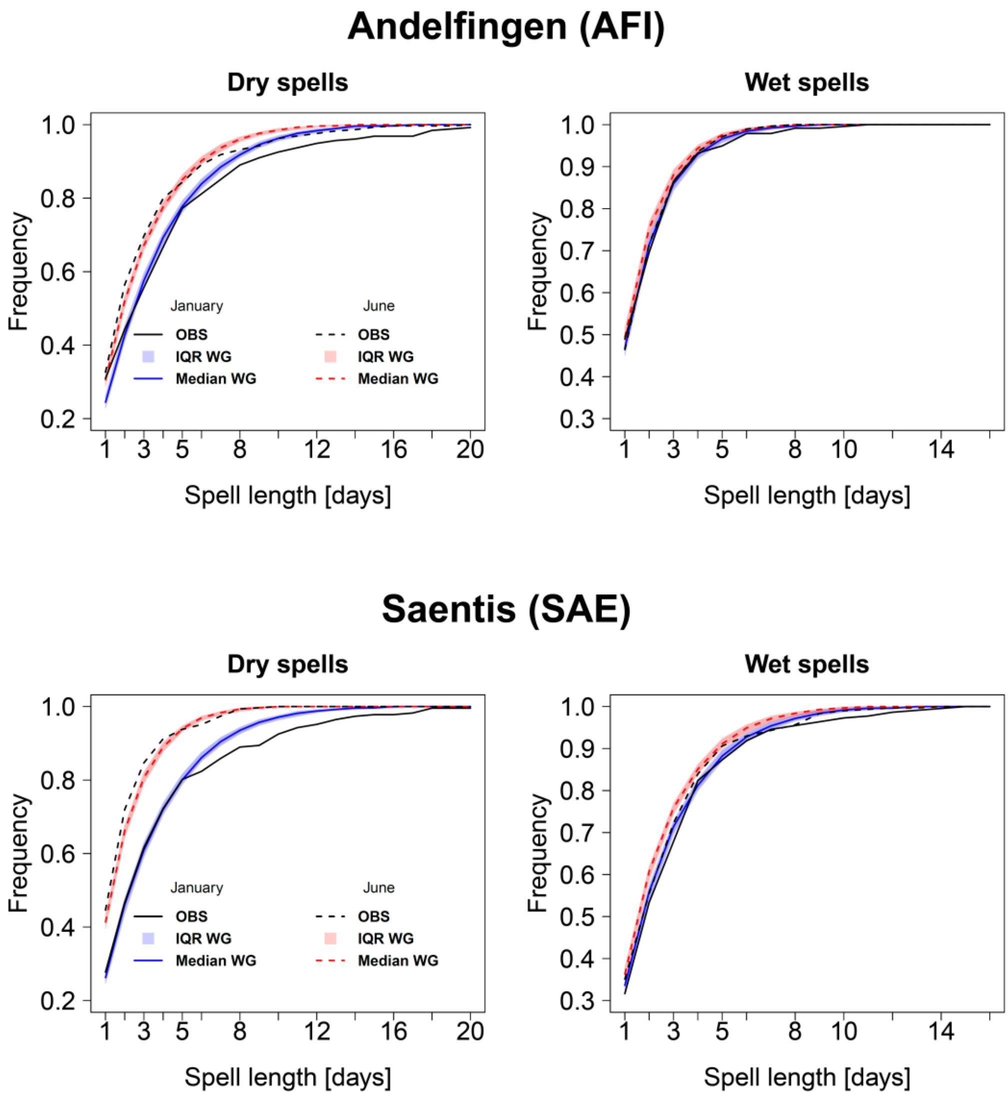

Figure 6. Cumulative distribution of the observed and simulated dry (left) and wet (right) spell length frequencies for the lowland station *Andelfingen* (top) and the mountain station *Saentis* (bottom). Results are for January and June during the time period of 1961-2011. The coloured area (line) represents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

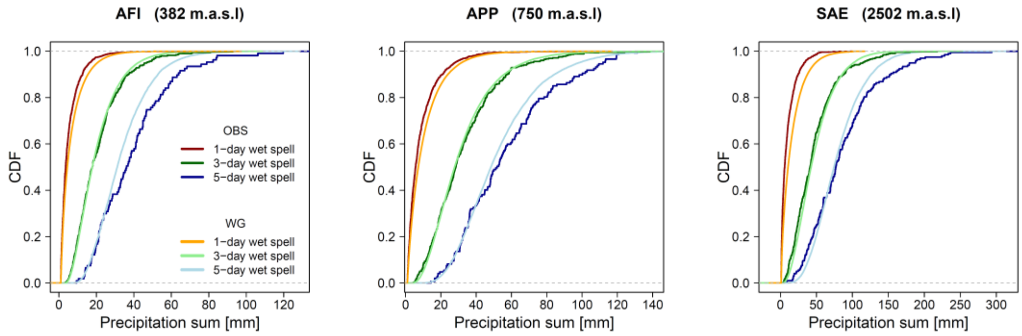

Figure 7. Cumulative distribution functions (CDFs) of multi-day precipitation sums for the three stations *Andelfingen* (AFI), *Appenzell* (APP) and *Saentis* (SAE). The lines represent the CDFs of non-zero precipitation amounts over one day (red), over three consecutive wet days (green) and over five 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.

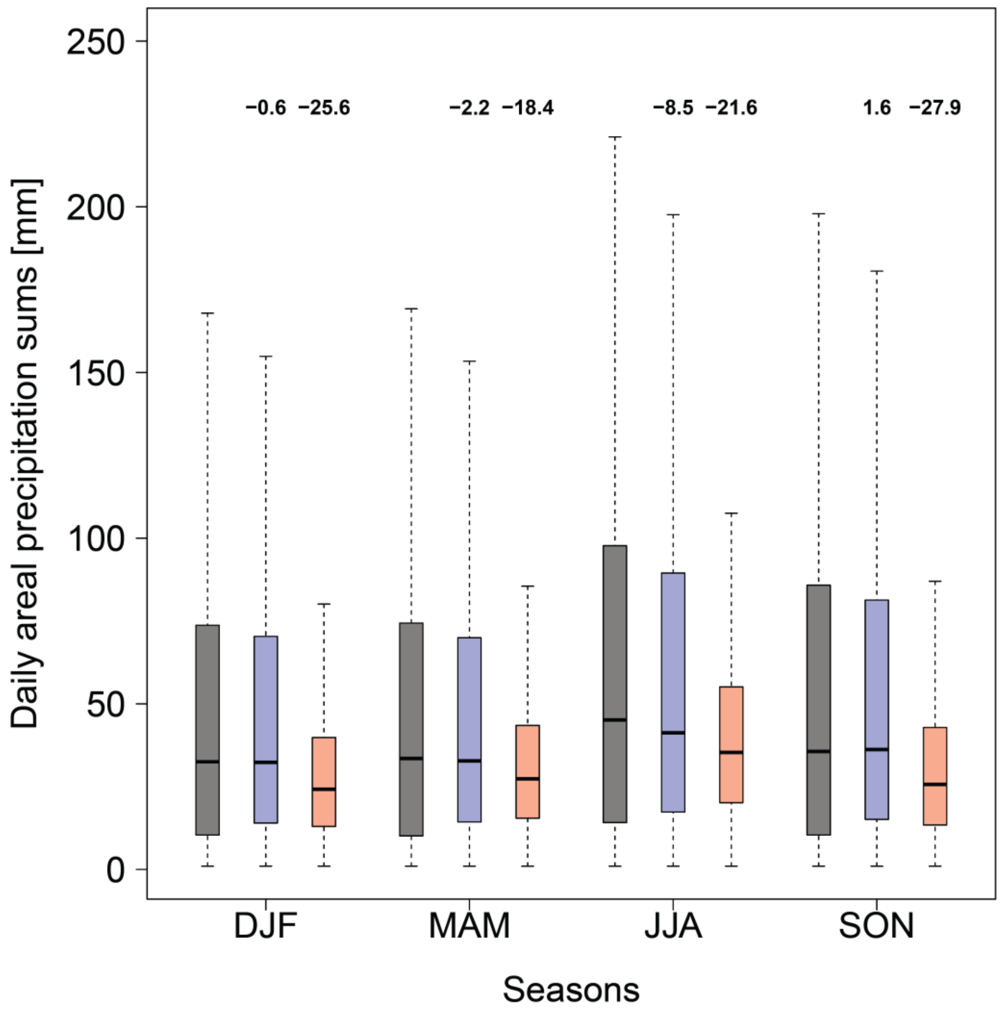

Figure 8. Daily non-zero precipitation sums over the catchment for the four seasons during 1961-2011. Daily Precipitation intensity of the eight stations are summed and days with an area sum of zero are excluded. Boxplots of observed daily sums (grey), of multi-site simulated time-series (blue) and of 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.

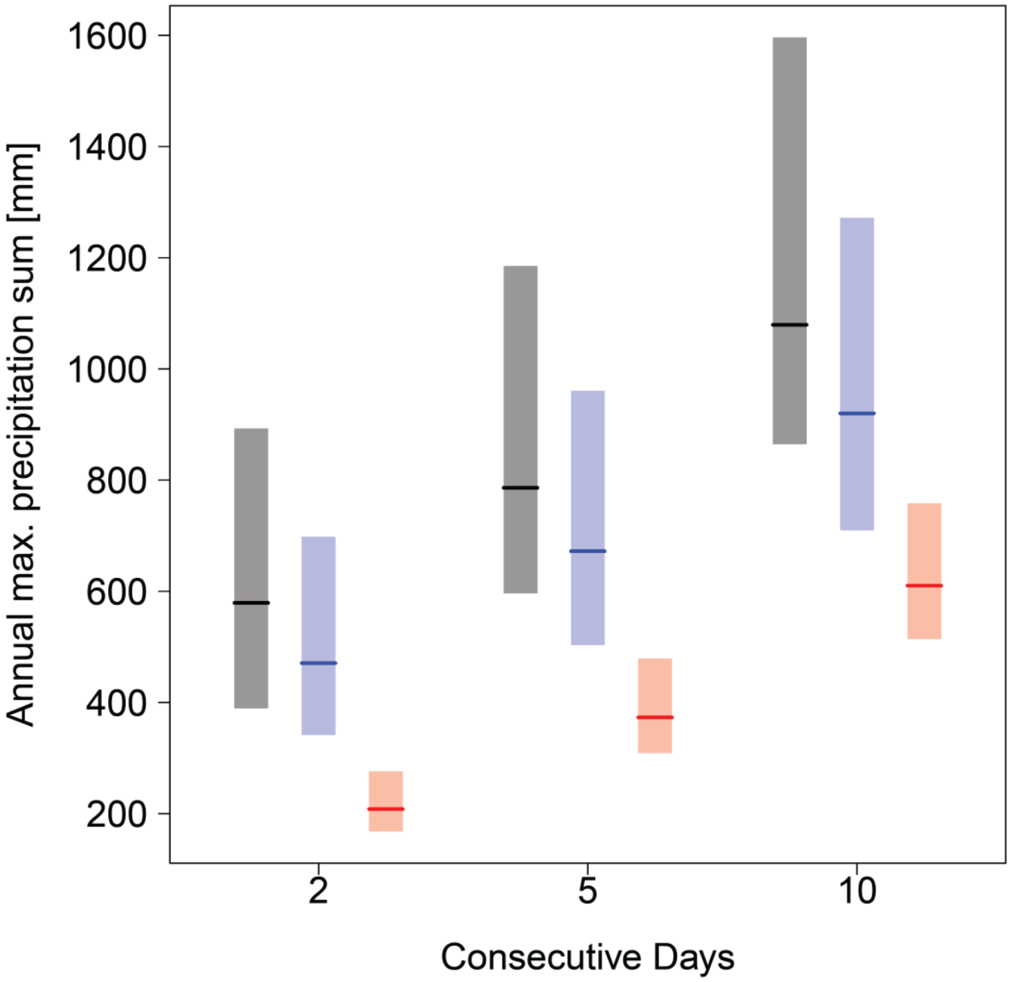

Figure 9. Annual maximum precipitation summed over all eight stations and over consecutive days. The analysis is done for all days of year. The bars (horizontal line) indicate the range between the 2.5% and the 97.5% empirical quantiles of the yearly maximum area sums during 1961-2011. The observations 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.