Authors reply on comment of editor

We thank the editor for his response. Following Reviewers' comments, major changes have been made to the manuscript. This document details these changes (with the line numbers corresponding to the marked-up manuscript version, with track changes) and provides complete replies to referee #1 and #2. We hope that the revised manuscript responds to their concerns.

Authors reply on comments of referee #1

We thank the referee for this thorough review and for the numerous constructive suggestions. The general presentation of the manuscript has been modified following these suggestions.

1. General comments

Comment R1 #1.1: The title of the paper is a bit misleading. The three models may be used for the spatial assessment of floods and hydrological modelling is mentioned not only in the title, but also throughout the manuscript. However, the precipitation models are not applied in an impact assessment in this study and for this reason in my eyes the title should solely contain the comparison of three precipitation models. It is a bit irritating that the authors refer to the importance of several aspects of the precipitation model performance whose importance is not really demonstrated.

We agree that the title was misleading. It has been replaced by 'Stochastic generation of multi-site daily precipitation focusing on extreme events'. We think that it is important to indicate the emphasis on the reproduction of very large precipitation events, in terms of intensity, duration, and spatial extent.

1

Comment R1 #1.2: The names of the new precipitation models are a bit misleading. First, "1D" and "3D" give the impression of any type of one- and three-dimensional simulation methodology. However, they represent days ("D"). I would rename the models into something more suitable.

This is an excellent suggestion and the names for these versions have been replaced by:

1. Wilks: the model proposed by Wilks (1998),

2. Wilks_EGPD: A first direct extension of 'Wilks', with the E-GPD and a Markov chain of order 4, as suggested by the referee (see comment R1 #1.6),

3. GWEX: the current GWEX-1D model,

4. GWEX_Disag: the current GWEX-3D model. It clearly indicates the disaggregation step which follows up the simulation at a 3-day scale.

Comment R1 #1.3: As far as I understand from the paper, the new GWEX models are actually "Wilks models" but with a new method to simulate the precipitation amounts (using temporally and spatially correlated random numbers from an autoregressive process and using a Student copula for the spatial component). I think this should be stated as such in the paper as the manuscript presents the new models more as a revolution rather than an evolution. So one of the first sentences could be that the paper deals with two modifications of the Wilks approach.

We agree. The fact that GWEX are evolutions of the Wilks model must be clearly stated. In fact, it was already indicated at p.2/l.25 and p.4/l.10 (first version of the manuscript) and throughout the presentation of the models. As suggested by the referee, this point is now indicated directly in the abstract and in the introduction (see p.1/l.3 and p.3/l.15). However, it must be underlined that GWEX is a significant evolution of the model introduced by Wilks (1998). First, as indicated by the referee, the methodology applied to simulate the precipitation amounts is considerably modified. We consider different temporal and spatial dependences, and we also discuss the choice of the marginal distribution in details, which is currently overlooked in the literature of precipitation stochastic models. Second, GWEX-3D (which will be named GWEX_Disag) combines simulations at a 3-day scale and a disaggregation approach, which represents a further step in the complexity of

the model. In our opinion, GWEX cannot only be considered as a slight modification/evolution of the Wilks model.

Comment R1 #1.4: The motivation behind the study and for the new model developments is the impact assessment. However, without the same the reader will not be able to really understand the sensitivity of certain statistics in regard to the assessment of extreme floods. I think the importance of some of the statistical metrics should be explained in more detail referring to the area of their application, and proof must be given of their relevance. Other literature in such a study (complex mountain region) is not very convincing to me.

We thank the referee for this suggestion and additional details regarding the importance of the statistical metrics have been provided in the revised version (section 2, p.5, lines 2-4; section 4, p.16, lines 4-9). In particular, Froidevaux (2014) analyze meteorological events triggering floods in Switzerland. These studies were very briefly mentioned at the beginning of p.12 in the original manuscript and these results are now discussed in more details. However providing a proof of the relevance of these metrics seems complicated without the hydrological application (which is clearly beyond the scope of this paper, as discussed in comment R1 #2.14.). If the referee has more specific metrics that could be presented, we would be glad to include them in our study.

Comment R1 #1.5: The abstract is incomplete and must be much more detailed and specific. What is an "event"? What is "large"? What are "recent advances"? The abstract should mention the Wilks model, the two new models (maybe also a short sentence how they work) and the basic outcomes of the study.

We thank the referee for this constructive suggestion. The abstract has been substantially modified and extended.

Comment R1 #1.6: The Wilks model could likewise be applied with E-GPD distributions for precipitation intensities and Markov chains of the order 4. That is, revealed weaknesses of the Wilks model can easily be addressed. I recommend adapting the Wilks approach for a more objective comparison. The original Wilks approach is not a given and was just one application for a specific dataset in the US and in my eyes it should always be revised for other study areas and climates.

3

We thank the referee for this suggestion. An additional version of the Wilks model, Wilks_EGPD, has been added to the comparison of the three previous models. Wilks_EGPD considers a E-GPD instead of a mixture of exponential distributions for the marginal distributions, and a Markov chain of order 4 instead of order 1.

Comment R1 #1.7: For flood modelling, the lagged cross correlations (see Wilks 1998, page 183) can be very important as they represent the progression of weather systems across the study area. Especially at larger scales the progression of weather events may be important. I strongly recommend plotting these statistics for all three models.
We thank the referee for this recommendation, which is also proposed by the referee #2. Lagged and unlagged inter-site correlations of precipitation amounts are now presented in Figure 10 (p.26) in Section 5.2.

Comment R1 #1.8: The autocorrelation of precipitation is addressed by MAR(1) models in the GWEX models. I would recommend plots for the autocorrelation of the precipitation intensities for some sites to see potential differences in their performance.
We thank the referee for this suggestion. Figure 10 (p.26) also presents an assessment of the autocorrelation of the precipitation amounts at the stations (black points), together with the cross-correlations (gray points).

2. Specific comments

Comment R1 #2.1. Line 8. I think there is a language issue.
This sentence has been reformulated in the revised version (see p.1, lines 14-16).

Comment R1 #2.2. Line 10. Not only conceptual models. There are more recent studies for coupling WGs with impact models.
Thanks for this remark. The authors are not aware of such impact models. If the reviewer has specific references, they can be included in the final version of the manuscript. The word 'conceptual' has been removed in order to include other types of hydrological models (e.g. distributed models).

Comment R1 #2.3. Page 1 bottom/ Page 2 top: In my eyes the classification is not fully correct. All these models are multi-site models. Also resampling methods are multi-site models. I recommend a more suitable classification even though I admit that the variety of the existing multi-site models makes a clear classification more and more difficult (also the authors combined parametric and non-parametric techniques).

We agree that the terminology 'multi-site models' is too vague and does not describe precisely the references given afterwards. This class of models has been renamed by 'Statistical multi-site models'. A detailed summary of the literature for this class of models is now provided, including specific extensions of Wilks model and how the proposed developments differ from them (p.2, lines 21-34).

Comment R1 #2.4. Page 4, Line 8. Are Thiessen polygons suitable for such a complex mountainous study region?

The computation of areal precipitation values is a difficult task considering the spatial and temporal variability of precipitation events, the complex topography of the study area, and the limited number of pluviographs. In Switzerland, Schäppi (2013) shows that the topography impacts rainfall amounts differently according to the type of meteorological event. In a preliminary study, the impact of different interpolation methods (inverse distance, ordinary kriging, kriging with external drift, Thiessen polygons) and different sets of stations (399, 211, 129, 47 and 22 stations) on extreme areal precipitation amounts has been analyzed. The main conclusion was that the number of stations was a much more important factor than the interpolation method. This was the main motivation for the application of the stochastic models to a high number (105) of stations. Furthermore, it is important to notice that applying more complex interpolation methods (e.g. kriging methods) increase significantly the computational cost, which can be prohibitive for the production of long meteorological scenarios.

Comment R1 #2.5. Page 7 "Marginal distributions". Can any proof be given that the more complex fitting of a combined distribution is really significantly better for the simulation of the extremes in this region? Also here, the most prominent argument is other literature.

QQ-plots are provided in the revised version in order to assess the quality of the fitting of these marginal distributions (p.22, Fig.6). However, it is very important to note that local applications give limited proof regarding the

performance of a distribution for the fitting of extreme values. As indicated in "Papalexiou, S. M. and Koutsoyiannis, D. (2013) Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resources Research, 49, 187201", most studies of extreme rainfall are inconclusive because they are too specific to particular areas or stations. The main explanation for these failures is that fitting and inferring the distribution tails is subject to high uncertainties in the estimation of the parameters, even for long time series (this point is also discussed and illustrated in Evin et al., 2016). The references given in the paper (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014) are conclusive precisely because they are the result of a very large number of applications, and give strong arguments in favor of the application of heavy-tailed distributions. Figures 2 and 3 tend to show that low tail-distributions (like a mixture of exponentials) could lead to an under-estimation of extreme precipitations in some regions (regions where ξ is different from 0, in green, yellow and red). In our study area, we acknowledge that the E-GPD does not bring a significant improvement of the performance compared to the mixture of exponential distributions (see Figure 12). However, as stated in the conclusion (p.33, lines 18-22), with only three parameters, the E-GPD provides a parsimonious and flexible representation of the whole of precipitation amounts. Its GPD tail is in agreement with recent results showing that extreme precipitation amounts must be modeled by heavy-tailed distributions (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014). The general framework proposed in this paper can be applied to very distinct precipitation regimes and the possible heavy tail of the E-GPD might be valuable in other areas.

Comment R1 #2.6. Page 9, top of the page. If the Gaussian copula is not suitable for simulating spatially dependent extremes but the Student copula is, this could be demonstrated. I am thinking of readers who want to build the code but are not experts in copulas and want to understand the significance.

The revised manuscript includes an additional version of the Wilks model, Wilks_EGPD, with applies a E-GPD instead of a mixture of exponential distributions for the marginal distributions (see comment R1 #1.6.). The difference between Wilks_EGPD and GWEX models provides a comparison of Gaussian and Student copulas concerning the reproduction of daily precipitation extremes (see Section 5.4., p.28).

Comment R1 #2.7. Page 9 bottom. Why are Markov chains of the order 4 used? Have there been statistical tests or sensitivity studies to underline this decision? Later on, some remarks are given on the simulation of short dry spells, but I think this should be addressed in a more structured way.

At p.5/l.10, it was indicated that Srikanthan and Pegram (2009) apply a 4-order Markov chain and show that it improves the reproduction of dry/wet period lengths. This point is now reminded at p.14, l.8-10.

Comment R1 #2.8. Page 11, Table 11 (and figures). Red and green are not suitable for figures, please change the colours as some people cannot read them otherwise (https://www.nature.com/nature/journal/v510/n7505/full/510340e.html).

We thank the referee for this comment. These colors have been modified and should be suitable for most color-blind people (following the recommendations given in `https://www.nature.com/nmeth/journal/v8/n6/full/nmeth.1618.html`). As we understand this issue, it seems that types of green (bluish green) and red (vermilion) are more adapted to color-blind individuals.

Comment R1 #2.9. Page 12, Line 28. I guess it is very difficult to say if an extreme precipitation amount is unrealistic or not as long as they are physically possible?

It is true that an extreme precipitation amount cannot be considered as unrealistic if the amount is physically possible. However, it is difficult to define what amount can be considered as impossible. Since this constraint was not used in our applications (we always obtain $\xi < 0.25$ in our study area, see dark red areas in Fig. 2 and 3), this remark was removed from the manuscript.

Comment R1 #2.10. Page 16, Line 18-20. If the order of the Markov chain is the issue for short dry spells, this can be easily adapted by using the same order in the original Wilks approach. What was the argument for using the first order Markov chains in the Wilks model? (see comment above)

As indicated in p.14/l.8-10, the direct extension of the Wilks model, Wilks_EGPD, is used to illustrate the impact of using a Markov chain of order 4 compared to order 1.

Comment R1 #2.11. Page 21, Line 8-9. Please explain the seasonal differences with explicit reference to the study area and its climatology for

7

Considering the number of additional figures provided in the revised manuscript, and considering that the assessment of the inter-annual variability is not central in this study (see Comment R1 #2.17.), this plot has been removed from the revised version of the manuscript.

Comment R1 #2.12. Page 22, section 4.4. and figure 10. To me, the performance looks fair for all three models. The main difference is the simulation of higher extremes with the GWEX models. The authors mention the difference but it needs further discussion. Also, how can we know that the extremes of one method are more realistic than from another? While we know little about the validity of the simulated extremes, they may have a big impact on simulated floods, especially in small catchments (but as mentioned before, this is not examined in the paper).

We agree with the referee, the performance looks fair for all three models if we look at figure 10 of the original manuscript. However, this figure only points out differences of behavior between the three models. As mentioned above (see Comment R1 #2.5.), these illustrative examples cannot be used to test the performance of the different models in regard to extreme daily precipitation amounts. The only way to perform such a validation is to apply some metrics on a large set of applications (here, for example, at all the stations), which is done at Figures 12 and 13 of the revised manuscript. Figures 10-13 of the original version, which were showing the fitting of the annual maxima at some stations/basins, have been removed from the manuscript.

Comment R1 #2.13. Page 26 Line 10-13. It is not surprising that the non-parametric disaggregation leads to a better performance. I understand its strengths but it may likewise be a limiting factor in generating extremes. In our opinion, GWEX_Disag represents a compromise between a purely statistical approach and a nonparametric approach. In terms of possible simulated amounts, it is not limiting factor at a 3-day scale and only is a constraint concerning the repartition of 3-day amounts across the daily steps.

Comment R1 #2.14. Page 29, first line 2-9. As already mentioned, I see the motivation behind the study (and it is generally a good one). But without any proof that the differences in the performance of the three precipitation models really have a significant impact on the simulation results of hydrological extremes (also considering all the uncertainties in hydrological

8

models), the significance of the research outcomes remain questionable.

We appreciate this criticism. The two following paragraphs motivate the assessment of these extreme precipitation amounts at different temporal and spatial scales and explain why the hydrological evaluation is not carried out in this study.

First, we would like to remind the key motivation of this study. The proposed stochastic models intend to preserve the most critical properties of precipitation at different spatial and temporal scales, and especially extreme precipitation amounts. We believe that a precipitation model which has these properties has a better chance to reproduce adequately flood properties for small sub-catchments as well as for large basins. Furthermore, empirical evidences have been provided by Froidevaux (2014) and Froidevaux et al. (2015) in our study area (i.e. Switzerland). Using 60 years of gridded precipitation data, Froidevaux et al. (2015) show that, in Switzerland, the generation of floods is mainly influenced by areal precipitation amounts accumulated on short periods (e.g. 1 to 3 days). Typically, the 2-day precipitation sum before floods is the most correlated to the flood frequency and the flood magnitude. These results are obtained by analyzing a wide variety of catchments, their areas ranging from 10 km^2 to 12,000 km^2. This study clearly motivates the multi-scale evaluation in space and time and the relevance of the precipitation metrics shown in our manuscript. These studies were very briefly mentioned at the beginning of p.12 and these results are now discussed in more details in the revised manuscript (see section 2, p.4, lines 2-4; section 4, p.16, lines 1-9).

Second, we agree that hydrological applications would validate the importance of such properties. Actually, hydrological applications are currently undertaken by the University of Zürich. A conceptual hydrological model (HBV) is applied to 87 sub-basins partitioning the whole study area, using precipitation scenarios produced by GWEX as inputs. Numerous technical issues still need to be resolved. Some basins are ungauged, or with very short streamflow series. The hydrological system of the Aare-Rhine river needs to be treated as a whole since floods at larger spatial scales need also to be investigated. Rating curves have very high uncertainties in some basins and need to be re-evaluated. It is also important to note that this hydrological study (as well as our study) is particularly challenging considering the large spatial extent of the Aare river catchment. These studies stand out from similar studies which are usually limited to few precipitation stations and one "small" catchment (see, e.g., Keller et al., 2015, recently published in

HESS, with an application to 8 precipitation stations located in a catchment with a size of 1700 km^2, to be compared with our study area of 17,000 km^2). The hydrological evaluation of our weather scenarios can thus not be carried out at the present time. It should be presented in future publications, considering the complexity of this work and the amount of results. However, we agree that the hydrological application would emphasize the significance of this study, and this point is discussed in the last section (Section 6, end of page 33, top of page 34).

Comment R1 #2.15. Page 29, Line 21-22. Please explain why, see comments above.
See Comment R1 # 2.11.

Comment R1 #2.16. Page 29, Line 27-28. The issue of larger spatial scales could be addressed by running more analyses at smaller scales. So the key motivation of the study is probably to examine large flood events and their spatial dependences? If so, this should be better explained. But again, without really simulating the floods throughout different scales the arguments for a particular precipitation model choice is questionable.
The key motivation is to develop a stochastic model for precipitation which preserve the most critical properties of precipitation at different spatial and temporal scales, and especially for extreme precipitation amounts. A meteorological model that does not preserve these metrics is unlikely to reproduce adequately flood properties for small sub-catchments as well as for large basins. However, we agree that the assessment of large flood events, in particular their spatial dependency, is very important. This will be done in further studies (see comment R1 #2.14.) by other research teams involved in this project.

Comment R1 #2.17. Page 30. Is the underestimation of the inter-annual variability such a big issue in Switzerland and for flood modelling? I would assume it is more an issue in more arid regions and for example agricultural studies? Some more remarks on the relevance in Switzerland and floods in general would be useful.
Thanks for this remark. We agree that the inter-annual variability is not central in this study, considering that we are interested in flood risk assessment. Indeed, this issue is more critical for other hydrometeorological applications, including agricultural and water resource related ones. Consequently, as in-

dicated in comment R1 # 2.11., this analysis has been removed from the manuscript.

3. Summary of review

Comment R1 #3.1. The abstract needs revision and must be more detailed (see general comments).
See comment R1 # 1.5.

Comment R1 #3.2. The introduction is not very well structured. The arguments for the construction of the new precipitation methodologies are mainly based on other literature and reasoning. The context of the paper should (i) either be revised (comparison of precipitation models) or (ii) proof must be given of the advantages using the new models by really coupling them with a hydrological model and examining the estimated flood events in the study region. I think it is the key weak point of the paper: reference is given to an application, which is not really done. Also, the title and abstract are a bit misleading and the reader may expect a flood modelling study and thus more than what has been presented.
We agree that the introduction was misleading and it has been modified in order to clearly indicate that this study aims at comparing precipitation models, the hydrological context being the motivation for the thorough assessment of areal precipitation extremes.

Comment R1 #3.3. For the three different precipitation models, I would recommend a flow chart with the Wilks model as the central component and then the adaptations that have been done. This makes it easier for the reader to understand all models and what has been changed.
This is an excellent suggestion and a flow chart has been added to the revised version of the manuscript (see Figure 3).

Comment R1 #3.4. Although the level of English is very good, some (minor) mistakes can be found in the manuscript and a native speaker should probably have a final look before resubmission.
A professional native English editor has been hired to proofread the final version of the revised manuscript.

Authors reply on comments of referee #2

Summary

The authors propose extensions of a classical multisite daily rainfall generator initially proposed by Wilks in 1998. The framework of Wilks model is flexible enough to allow many adaptations, and the authors of this paper propose - to add more structure in the dynamics of the model by considering higher order Markov model for the occurrence process and an autoregressive component for the amounts - to use a hybrid distribution for the marginal distribution to deal with heavy tail distributions - to use a Student copula for the spatial structure to catch upper tail dependence. I believe that all these extensions make sense and are interesting to try.

We thank the referee for this review and for these constructive comments. Most of the following suggestions have been incorporated in the modified manuscript.

1. General comments

Comment R1 #1.1. Many extensions of the Wilks model have already been proposed in the literature. I think that a review of this literature must be included in the paper and that the authors should explain why the extension that they propose is original and useful with respect to this literature.

We agree that the differences between GWEX and the existing extensions of Wilks model must be presented in the introduction. A more complete presentation of the literature has been included in the introduction (see p.2, lines 26-34) of the revised version of the manuscript.

Comment R1 #1.2. In my opinion, one weakness of the paper is that the model is formulated as a simulation tool rather than as a proper statistical model. It is also the case for the original Wilks model, but it has then been reformulated by other authors as a statistical model, see e.g. Thompson et al. (2007). I think that the paper would be easier to read for statisticians like me if a similar formalization was done in the paper. In particular, the

various assumptions on the occurrence/amount processes should be written precisely using formulas and the definition of the model should be separated from the discussion on parameter estimation and simulation.

We thank the reviewer for this excellent suggestion. GWEX is now presented using a more formal mathematical formulation and the whole section 3. ("Multi-site precipitation model") has been modified. A specific section is now devoted to parameter estimation (section 3.3, p.11-13).

Comment R1 #1.3. I believe that the validation part must also be improved. First, some usual validation criteria for rainfall generators, such as diagnostics based on the marginal distribution (e.g. qqplot) and the second order structure of the process (autocorrelation and crosscorrelation functions) are not shown and it makes it difficult to see the benefit of using a hybrid distribution and the autoregressive component. Also the chosen validation criteria does not permit to see the interest of using a student Copula (does it really improve the modeling of extremal dependence?).

These remarks have also been made by the referee #1 (comments R1 #1.8, R1 #2.5 and R1 #2.6). QQ-plots are now provided to assess (visually) the quality of the fitting for the marginal distributions (see Figure 6, p.23). An additional figure provides an assessment of the performance concerning the autocorrelations and the reproduction of cross-correlations (see Figure 10, p.26). Finally, an additional model version, a direct extension of Wilks model, with the E-GPD instead of a mixture of exponential distibutions, "Wilks_EGPD", has been added to the three original models. This additional version enables the assessment of the impact of the Student copula (versus a Gaussian spatial structure).

Comment R1 #1.4. Finally, I find the simulation results generally disappointing. If I understand correctly the categorization, we should obtain about 90% of good if the model was able to reproduce the statistics of the observed rainfall? Is it satisfactory to obtain percentage around 50%?.

Yes, we should obtain about 90% of good if the model is able to reproduce the observed statistics, and very few 'poor' cases. As indicated in the paper, our primary criteria to judge the overall performance of a model is the number of metrics for which 'poor' performances are obtained. We agree that these percentages are subjective (why 90%? Is 50% of good cases good enough?) but not more subjective, in our opinion, that the visual inspection of a QQ-plot. Furthermore, the purpose of the CASE framework, as

13

presented in Bennett et al. (2017), is to enable a more systematic comparison of stochastic models. Our study also tries to promote this approach. A more systematic comparison of the models, which includes a consistent way to compute the performance metrics, is important in order to obtain a fair assessment of the strengths/weaknesses of the different models. For this reason, this study applies the classification proposed by Bennett et al. (2017), without modifying the classification.

2. Specific comments

Comment R1 #2.1. Keywords are missing?
In HESS, to the extent of our knowledge, keywords do not appear in the manuscript.

Comment R1 #2.2. End of Page 1/top of page 2. I am not really satisfied by the proposed classification. For example weather type models are often used as multisite rainfall generators (without conditioning to large scale information). Also it would be useful to cite the review papers on rainfall generators here.
We agree that the terminology 'Multi-site models' is too vague here. A similar comment has been done by the referee #1 (see comment R1 #2.3.). We propose to replace 'multi-site models' by 'statistical multi-site models' (see p.2, lines 21-34). Additional references have been incorporated.

Comment R1 #2.3. Section 2.1. The authors go directly from a Markov chain of order p=1 to a Markov chain of order p=4. I would expect that the best value of p is somewhere between these two values. The authors could try to find the optimal value of p, using for example standard model selection criteria.
We thank the reviewer for this suggestion. It is true that an optimal value might be found if there was an easy selection criteria. As this point is not central in our study, a direct comparison of Markov chains of order $p = 1$ and $p = 4$ is deemed sufficient.

Comment R1 #2.4. Equation (5). I am surprised that the authors use a diagonal matrix for A. I would expect that it is useful to add some spatial structure here?
Initial versions of GWEX were applying a full covariance matrix for A. How-

14

ever, it seems that large covariance matrices are often very close to a non positive definite matrix. This is not really problematic during the estimation step, but leads to very unstable results during the simulation step. As applying a diagonal matrix for A does not seem to degrade the performance of GWEX, this solution was retained.

Comment R1 #2.5. Section 2.3 and 3.3 should be merged.
We thank the reviewer for this suggestion. Following comment R2 #1.2., these sections have been re-organized with a specific section devoted to the estimation step. Previous section 3.3 has been removed in the revised manuscript.

Comment R1 #2.6. Section 3. Why is it called "Application"? I do not see any application here.
Following previous comments (comments R2 #1.2. and R2 #2.5.), the whole section 3 has been reorganized. Section 3.1 'Split-sampling procedure' is now presented at the beginning of section 5 "Results" (see p.19). Previous section 3.2 'Regionalization of the ξ parameter' is related to the estimation of the parameters is now presented in section 3.3, p.11-13. Previous section 3.3 'Generation of scenarios' has been removed (see previous comment).

Summary of changes

Overall presentation of the manuscript

Most of the referee's comment are related to the presentation of the methodology and the results. These comments are entirely justified and are appreciated, as they greatly enhance the paper. The following paragraphs summarize what modifications have been made to the manuscript (more details can be found in the response to specific comments):

- **Abstract:** We agree with the referee #1 (comments #1.5. and #3.1.) that the abstract was not specific enough. Additional details have been added (summary of the model developments, key results, etc.)

- **Title and introduction misleading:** As pointed out by the referee #1 (comments R1 #1.1. and R1 #3.2.), the title and the introduction

seemed to indicate that our study shows the results of an hydrological application, which is not the case. The title has been replaced by 'Stochastic generation of multi-site daily precipitation focusing on extreme events'. Vague references to hydrological applications in the introduction have been been removed.

- **Classification of the precipitation models:** Both referees (comment R1 #2.3. and comment R2 #2.2.) rightly indicated that the terminology 'multi-site models' was too vague and did not describe precisely the references given afterwards. This class of models is now named 'Statistical multi-site models'. A detailed summary of the literature for this class of models is provided, including specific extensions of Wilks model and how the proposed developments differ from them.

- **Mathematical formulation:** As suggested by referee #2 (comment R2 #1.2.), we now present a more formal mathematical formulation of GWEX (section 3).

- **Names of the models:** As indicated by referee #1 (comment R1 #1.2.), the current model names are confusing. New names have been given to the different model versions.

- **Flowchart of the models:** As suggested by referee #1 (comment R1 #3.2.), a flow chart has been added in order to clarify the modifications made to the original Wilks model and to illustrate the different model versions (see Figure 6).

- **Specific section devoted to the parameter estimation:** As suggested by referee #2 (comments R2 #1.2., R2 #2.5. and R2 #2.6.), a specific section is now devoted to the estimation step.

Validation and choice of metrics

Both referees (comments R1 #1.7., R1 #1.8., R1 #2.5. and R2 #1.3.) suggested additional validation criteria. Following their suggestions, QQ-plots of the marginal distributions (empirical versus fitted E-GPD or mixture of exponential distributions) is now presented in the revised manuscript (see Figure 6). Additional figures have also been added in order to assess the reproduction of lagged and unlagged cross-correlations (see Figure 10).

Comment R2 #1.4., as well as comments R1 #2.5 and R1 #2.12., to a lesser extent, criticize the evaluation framework and the significance of the results concerning the reproduction of extremes. In this study, validation of extreme values is mostly performed using metrics computed at all the stations and for different spatial scales. In our view, it is difficult to dismiss/validate a particular method using visual inspections of the reproduction of extremes (e.g. using Gumbel plots as in Figures 10-13 of the original manuscript, or QQ-plots). Consequently, previous Figures 10-13 have been removed from the manuscript. These figures were mostly shown to illustrate interesting aspects in terms of extrapolation but seem to be prone to different interpretations in terms of performance. Finally, we now present relative differences in Figures 12-13 (instead f absolute differences), in order to highlight potential under/overestimations at large spatial scales.

In this study, we firmly support the application of the CASE framework (Bennett et al., 2017), which enables a more systematic comparison of stochastic models. A consistent way to compute the performance metrics is important in order to obtain a fair assessment of the strengths/weaknesses of the different models. For this reason, in this study, the classification proposed by Bennett et al. (2017) is not modified. A remark has been added to the revised version of the manuscript (p.16, lines 16-18).

Parameter estimation of the inter-site correlations, and autocorrelations

As indicated in Wilks (1998), direct estimates of the spatial and temporal dependence of precipitation amounts cannot be obtained since non-zero precipitation amounts $Y_t(k)$ is a hidden variable which cannot be observed. In the previous version of the manuscript, these correlations were directly estimated from positive precipitation amounts. However, this method leads to a significant under-estimation of the inter-site correlations of precipitation amounts (zero and non-zero). In the revised version of the manuscript, we follow the methodology proposed by Wilks (1998) and Keller et al. (2015). For each pair of stations, we generate long sequences of precipitation amounts $P_t(k)$ using the estimated parameters of the occurrence process ($\hat{\Pi}$ and $\hat{\omega}_{kl}$), the parameters of the marginal distributions and a correlation coefficient $m_0(k, l)$ indicating the degree of spatial dependence. Similarly to the occurrence process, $\hat{m}_0(k, l)$ is then found iteratively by matching the correlation between

these long random streams with the observed correlation $\text{Corr}(P_t(k), P_t(l))$ (see Wilks, 1998; Keller et al., 2015, for further details). The correlation matrix $\hat{\mathbf{M}}_0$ is then composed of the cross-correlations $\hat{m}_0(k, l)$ obtained for all possible pairs of stations. For each station, the estimates of the lag-1 serial correlation coefficients of the matrix $\mathbf{A}$ are obtained using the same simulation approach (see end of page 12, top of page 13).

These modifications improves greatly the reproduction of extreme precipitation amounts at large spatial scales, in particular for model GWEX concerning 3-day precipitation extremes.

References

Bennett, B., Thyer, M., Leonard, M., Lambert, M., and Bates, B.: A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model, Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2016.12.043, In Press, 2017.

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

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

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O.: Flood triggering in Switzerland: the role of daily to monthly preceding precipitation, Hydrology and Earth System Sciences, 19, 3903–3924, 2015.

Keller, D. E., Fischer, A. M., Frei, C., Liniger, M. A., Appenzeller, C., and Knutti, R.: Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment, Hydrology and Earth System Sciences, 19, 2163–2177, 2015.

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

Schäppi, B.: Measurement and analysis of rainfall gradients along a hillslope transect in the Swiss Alps, Ph.D. thesis, ETH Zürich, Zürich, Switzerland, 2013.

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

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

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

Stochastic generation of multi-site daily precipitation ~~for the assessment of~~ focusing on extreme ~~floods in Switzerland~~events

Guillaume Evin[1], Anne-Catherine Favre[1], and Benoit Hingray[1]

[1]Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP[*], IGE, F-38000 Grenoble, France

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

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

1 Introduction

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

In the last two decades, a ~~fair~~ number of precipitation models have been proposed to deal with the temporal and spatial properties of daily precipitation, for both intermittency and amount, ~~which~~ and all have different strengths and ~~limitations. An important proportion~~ weaknesses. Many of these models use exogenous variables to predict the statistical properties of precipitation ~~,~~ using generalized linear models (Chandler and Wheater, 2002; Mezghani and Hingray, 2009; Serinaldi and Kilsby, 2014b), atmospheric analogs (Lafaysse et al., 2014), or modified Markov models (Mehrotra and Sharma, 2010). Introducing a link between exogenous atmospheric variables can be ~~interesting~~ used to reconstruct past events, ~~for~~ make

---

[*]Institute of Engineering Univ. Grenoble Alpes

predictions, or ~~to~~ downscale GCM-based simulations of future climate. Such models are classically referred to as statistical downscaling models (see Maraun et al., 2010, for a review). Closely related to this approach, weather 'types' or 'regimes' ~~(see Ailliot et al., 2015, for a review)~~ (Ailliot et al., 2015) can be used to specifically account for different atmospheric circulation patterns. Using Hidden Markov Model (HMM) with transitions between these weather states, stochastic weather generators

5　can then simulate various aspects of the precipitation process (Rayner et al., 2016).

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

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

10　– **Random fields:** Spatio-temporal precipitation models can simulate precipitation fields over a regular grid. ~~These developments are particularly interesting~~ This approach is particularly useful for hydrological applications, since areal precipitation values over a ~~catchment are directly obtained~~ basin are obtained directly. Poisson cluster-based models ~~(Burton et al., 2008, 2010; McRe~~ simulate rain disk cells, with random centers, radius and intensity, over the study area. Meta-Gaussian models (Vischel et al., 2009; Kleiber et al., 2012; Allard and Bourotte, 2015; Baxevani and Lennartsson, 2015; Bennett et al., 2017) are

15　based on truncated and transformed random Gaussian fields. Closely related, the turning band method can be used to simulate intermittent precipitation fields with different type of ~~advections~~ advection (Leblois and Creutin, 2013). These model structures are appealing since they are able to simulate realistic precipitation fields at fine spatial scales. However, their complexity leads to numerous technical issues during parameter estimation and simulation, notably in terms of computational cost. Moreover, they are usually unable to represent large regions ~~, with~~ comprising very distinct precipitation

20　regimes.

– ~~**Multi-site**~~ **Statistical multi-site models:** ~~Since the pioneering work of Wilks (1998), numerous weather generators have been developed to fit directly the statistical~~ In this last type of weather generator, the properties of precipitation are directly fitted at a limited number of stations ~~(Bárdossy and Pegram, 2009; Srikanthan and Pegram, 2009; Baigorria and Jones, 2010;~~ ~~For both precipitation occurrence and amount, multi-site generators are able to preserve~~ using different statistical structures.

25　This type of generator preserves the inter-dependency between all pairs of stations, even when the area under study exhibits ~~very~~ different precipitation regimes~~(e. g. in mountainous areas).~~. Bárdossy and Pegram (2009) and Rasmussen (2013) combine a multivariate autoregressive process and transformations (V-transform, power transformation) to simultaneously model precipitation occurrence and amount. More precisely, with these models, transformed precipitation amounts follow truncated distributions. Alternatively, Wilks (1998) proposes a multi-site model in which precipitation occurrence and

30　amount are handled separately. Several extensions to this popular structure have been proposed in the literature. Thompson et al. (2007) the Wilks model as a hidden Markov model, inferring three precipitation states ('dry', 'light' and 'heavy'). Mehrotra and Sharma (200 semi-parametric techniques to add more flexibility to the spatial structure of precipitation occurrence and amount. Srikanthan and Pegram (2009) propose a modified version in which daily, monthly and annual amounts are nested such that precipitation statistics are preserved for all these levels of aggregation.

~~The context of this work is the risk assessment of extreme flooding events using the "continuous simulation" method. Very long series of daily precipitation (e.g. 10, 000 years)are generated for the present climate and used as inputs of a conceptual hydrological model. As the hydrological process and extreme floods are influenced by the hydrological configurations (for example, different levels of soil saturation), these precipitation scenarios must reproduce low to very extreme daily precipitation~~
5 ~~events at different temporal and spatial scales.~~

~~In this paper, we develop a precipitation model, called GWEX, which will be used to generate these long scenarios over a large area with various precipitation regimes. GWEX is applied to 105 stations of the Aare river catchment in Switzerland. As multi-site models have~~ Mehrotra et al. (2006) compare three different precipitation models, the Wilks model, a HMM and a resampling approach, and provide strong arguments in favor of the Wilks model in terms of performance, computation time,
10 model, and level of complexity of the model structure. Furthermore, as indicated above, this model offers a flexible structure which can be applied to a large number of stations with very different precipitation regimes ~~, GWEX relies on the structure proposed by Wilks (1998). The underlying idea is to separate the process representing the precipitation occurrences at the different stations from the process generating the amountsof the precipitation events.~~

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

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

30 – While non-parametric densities with Gaussian kernels (Mehrotra and Sharma, 2007a, 2010) offer ~~a great~~the flexibility to fit the observed range of precipitation amounts, their tail also ~~belong~~belongs to the domain of attraction of the Gumbel distribution and ~~suffer~~suffers from the same drawbacks.

Alternatively, current statistical procedures consisting in fitting a flexible distribution to the bulk of the observations and ~~use~~using it for extrapolation ~~is~~are highly questionable, as major assumptions are usually violated ~~, as it has been extensively~~

discussed by Klemeš (2000a, b)(Klemeš, 2000a, b). Since the tail of the distribution on precipitation amounts at each station will dictate the generation of the most extreme precipitation events, important features of GWEX are:

- to apply an application of a heavy-tailed distribution to precipitation amounts at each station (Naveau et al., 2016), following the conclusions drawn by Papalexiou et al. (2013); Serinaldi and Kilsby (2014a); Cavanaugh et al. (2015),

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

Furthermore, following Bárdossy and Pegram (2009), GWEX also employs the copula theory to introduce a tail dependence between the precipitation amounts simulated at the different stations. The second version of the GWEX model includes a disaggregation method, the observed precipitation amounts being fitted at a 3-day scale in a first step. This paper compares the performance of the different model versions and assesses the impact of the different statistical components (e.g. heavy-tailed distribution, tail dependence, etc.).

The global methodology is first described in Section ??, with a presentation of the study area, the We first describe the study area in Section 2. The features of different multi-site precipitation models , and the are then described in Section 3. The evaluation framework, which presented in Section 4, aims at assessing the performances of GWEX performance of these models at different spatial and temporal scales. Section ?? details the applications 5 presents an application of these daily precipitation models to 105 stations located in the Aare river catchment. Section 5 synthesizes the results , Switzerland, with a summary of the results focusing on the reproduction of extreme events. Section 6 concludes Finally Section 6 presents our conclusions.

2 Material and methods

1.1 Data and study area

2 Data and study area

The Aare River basin covers the northern part of the Swiss Alps and has an area of 17,700 km^2. Basin elevations approximately range from 310 m.a.s.l. in Koblenz (entrance to Germany in the north) to 4270 m.a.s.l. at the Finsteraarhorn summit (in the south of the area basin). The mean annual precipitation for the basin as a whole is 1300 mm. The basin can be divided into five main sub-basins with different hydrometeorological regimes highly governed by regional terrain features (Jura mountains in the North-West north-west; Northern Alps in its southern part, the south of the basin and lowlands in the middle).

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

4

The ~~weather scenarios are used~~ proposed precipitation models are designed to simulate, via a conceptual hydrological model, flood scenarios for the whole Aare River ~~Basin~~ basin and for its different sub-basins. For Switzerland, Froidevaux et al. (2015) show that the generation of floods is mainly influenced by areal precipitation amounts accumulated over short periods (e.g. 1 to 3 days). These results are obtained by analyzing a wide variety of basins, their areas ranging from 10 km^2 to 12,000 km^2.

5 Therefore, the properties of the weather scenarios must be evaluated at different spatial and temporal scales, from the high resolutions required ~~for simulating~~ to simulate the hydrological behavior of the system (e.g. sub-daily, 100 km^2) to lower resolutions relevant at the scale of the entire basin (e.g. ~~n-days~~n-days, 17,700 km^2).

~~Following Mezghani and Hingray (2009), a multi-scale evaluation in space and time is thus carried out. For instance,~~ In this study, the performance of ~~GWEX are~~ the different precipitation models is evaluated at the station scale, at the scale of ~~5 and~~ 15

10 and 5 sub-basins partitioning the Aare ~~river catchment (see Figure 1)~~River basin, and at the scale of the entire study area (see Section 5). Note that for those evaluations, areal estimates of precipitation are obtained from the precipitation amounts at the stations using the Thiessen polygon method.
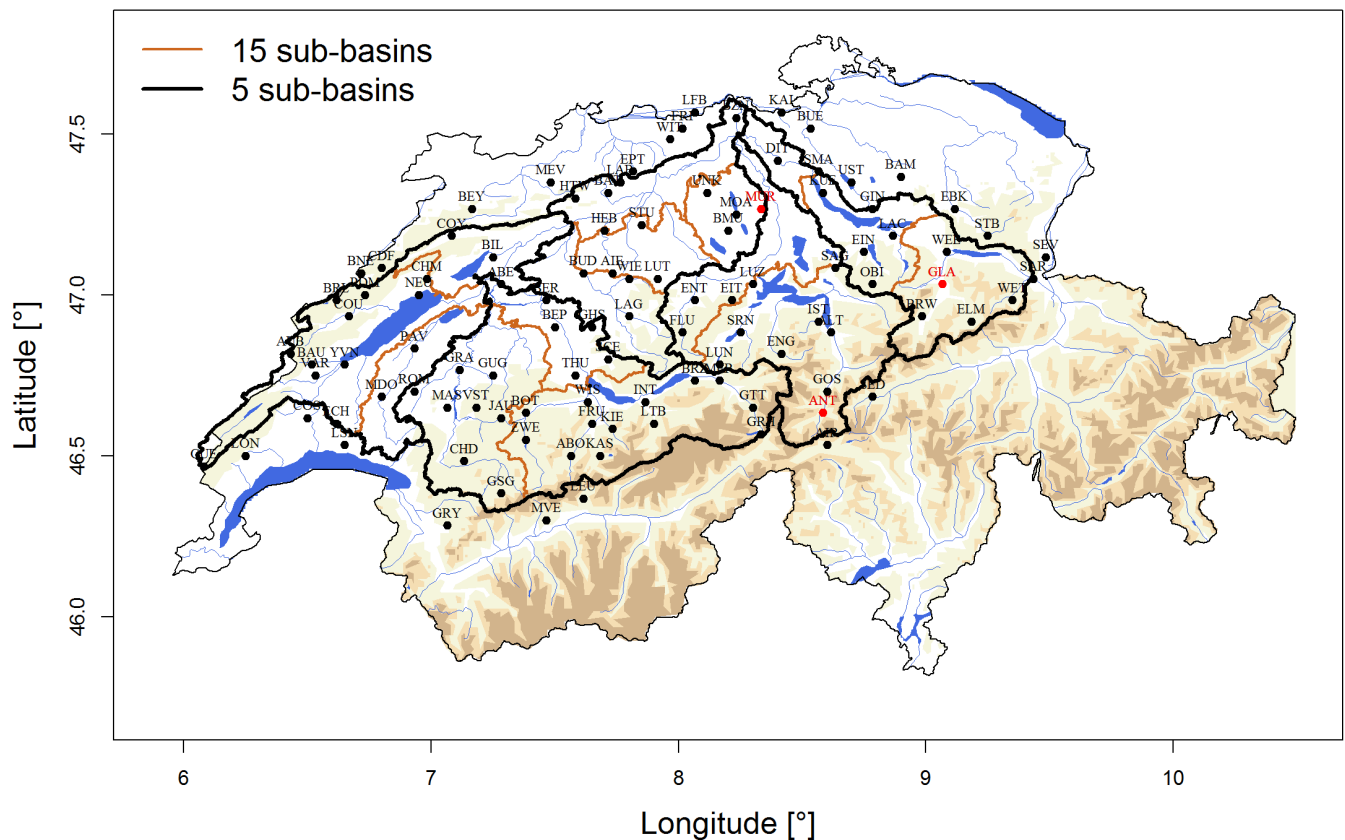

Figure 1. Location of the 105 precipitation stations in Switzerland. Different partitions of the Aare ~~river catchment are considered~~ River basin into 5 and ~~the names of the five main~~ 15 sub-basins are ~~indicated~~shown.

3 Multi-site precipitation model

As indicated above, GWEX ~~is a~~ refers to multi-site precipitation ~~model which relies~~ models that rely strongly on the structure proposed by Wilks (1998). ~~Precipitation amounts are modeled independently of precipitation occurrences, which act as a mask.~~

~~As proposed by Wilks (1998), the occurrence process~~ At each location k, let $P_t(k)$ be a random variable representing the accumulated precipitation over day t. The structure proposed by Wilks considers a hidden occurrence process $X_t(k)$ that can be represented by a two-state Markov chain ~~, representing 'dry' and 'wet' days:~~ as follows:

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

~~In practice, these states are obtained using a low precipitation threshold (0.2 mm). In the present case, the seasonality of the occurrence process is taken into account by estimating model parameters on a monthly basis~~ Precipitation amount $P_t(k)$ is then defined as:

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

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

~~At-site occurrence process~~

3.1 Precipitation occurrence process

3.1.1 At-site occurrence process

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

$$\Pr\{X_t(k) = 1 | X_{t-1}(k), \ldots, X_1(k)\} = \Pr\{X_t(k) = 1 | X_{t-1}(k), \ldots, X_{t-p}(k)\}, \tag{3}$$

i.e. the probability of having a wet day at time t depends only on the p previous states, for days $t-1, \ldots, t-p$. While many authors suppose that a first-order Markov is sufficient (e.g. Wilks, 1998; Keller et al., 2015), Srikanthan and Pegram (2009)

apply a 4-order Markov chain and show that it improves the reproduction of dry/wet period lengths. In this study, different orders for this Markov chain are considered.

At each site, the probability of having a wet day at day t is given by the transition probability $\Pr\{X_t(k)=1|X_{t-1}(k),\ldots,X_{t-p}(k)\}$. $\Pr\{X_t(k)=1|X_{t-1}(k)=i_1,\ldots,X_{t-p}(k)=i_p\}$, where $i_1,\ldots,i_p$ are equal to 0 or 1. This Markov chain is thus fully charac-
5   terized by a transition matrix Π with ~~dimensions~~ dimension 2^p.

These transition probabilities are estimated directly by the proportion of wet days $X_t(k)=1$ following observed sequences $\{X_{t-1}(k),\ldots,X_{t-p}(k)\}$.

3.1.2   Spatial occurrence process

The spatial dependence of the precipitation states $X_t(k)$ is modeled using an unobserved Gaussian stochastic process $\mathbf{U}_t=\{U_t(1),\ldots,U_t(K)$
10   where K is the number of stations. Here, Gaussian random variables $U_t(k),k=1,\ldots,K,$ are temporally independent and $\mathbf{U}_t$ follows a multivariate normal distribution:

$$\mathbf{U}_t \sim N(0,\Omega_X), \tag{4}$$

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

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

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

~~Spatial occurrence process~~

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

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

where $\pi_0(s)=\Pr\{X_t(s)=0\}$ and $\pi_1(s)=\Pr\{X_t(s)=1\}$ denote the probabilities of having dry and wet states at location s, respectively, and $\pi_{00}(k,l)=\Pr\{X_t(k)=0,X_t(l)=0\}$ denotes the joint probability of having dry states at both locations k and l. ~~$\hat{\rho}$ estimates can thus be obtained using the empirical (i.e. observed) counterparts of these probabilities.~~
25   ~~Following Wilks (1998), for two locations k and l, a bivariate normal distribution with mean 0, variance 1 and a correlation parameter ω can be employed to reproduce $\hat{\rho}$. More precisely, Gaussian variates can be converted to uniform variates using~~

~~the probability integral transform. The correlated uniform variates are then compared to the transition probabilities in order to generate new states $X_t(k)$ and $X_t(l)$ (see Wilks, 1998, for further details).~~

~~The relationship between ω and $\hat{\rho}$~~ The relationship between ω_{kl} and ρ_{kl} is not direct since the ~~at-site occurrence process also influences $\hat{\rho}$~~ temporal persistence of dry and wet events introduced at each station with a Markov chain also influences ρ_{kl}

5   (Wilks, 1998). Figure 2 illustrates this relationship~~for~~, obtained for the month of July via Monte-Carlo simulations, for two close stations, GOS and ANT. ~~For these two stations~~In a first step, transition probabilities with a Markov chain of order 4 are ~~computed for the month of January~~estimated for these two stations. Given these transition probabilities, stochastic simulations of occurrence are then generated for different values of ~~ω~~ω_{kl}, leading to different values of ~~ρ~~ρ_{kl}. Since this relationship is monotonic (see Fig. 2), ~~a value ω corresponding to an empirical estimate $\hat{\rho}$ can be found iteratively~~ it can be used to identify

10   the value ω_{kl} leading to a specific $\hat{\rho}_{kl}$, namely the empirical value obtained from the observed time series of occurrence. The estimate of ω_{kl} is found by iterating until the evaluation of the correlation between the simulated precipitation states, ~~ρ, matches $\hat{\rho}$. It must be noticed~~ ρ_{kl}, matches $\hat{\rho}_{kl}$. Note that a very high value for ~~$\hat{\rho}$~~$\hat{\rho}_{kl}$ cannot always be reached, even if ~~$\omega = 1$~~$\omega_{kl} = 1$. This is however a situation which rarely occurs in practice.
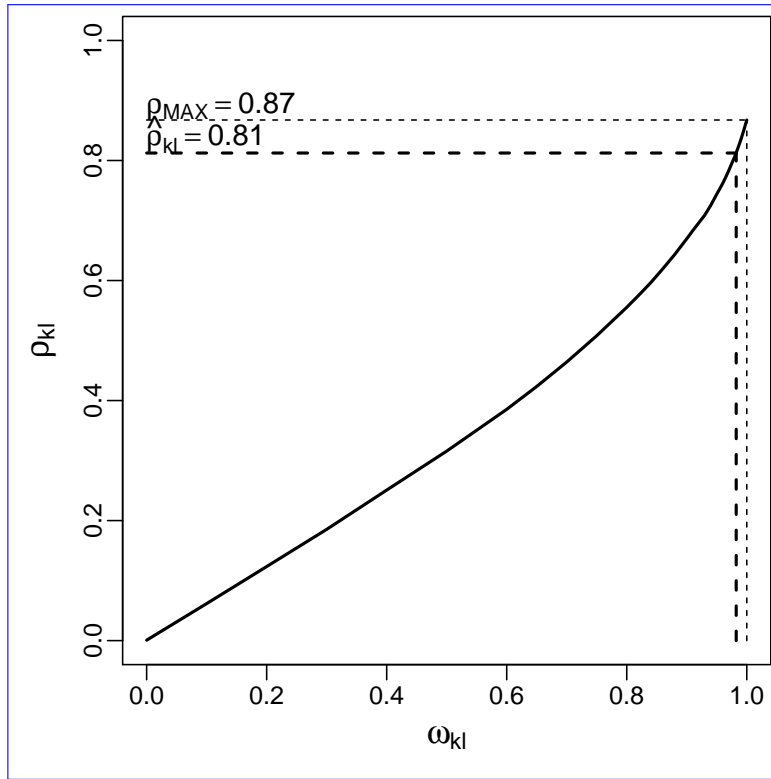

Figure 2. Illustration of the relationship between ~~ω~~ ω_{kl} and ~~ρ~~ ρ_{kl} for the month of ~~January,~~ July and for stations GOS and ANT. A Markov chain of order 4 is considered in this example. The correlation between the observed states is ~~$\hat{\rho} = 0.81$~~ $\hat{\rho}_{kl} = 0.81$ and can be reproduced using a bivariate Gaussian distribution with a correlation parameter of ~~$\omega = 0.98$~~$\omega_{kl} = 0.98$. The maximum correlation ρ which can be obtained if ~~$\omega = 1$~~$\omega_{kl} = 1$ is ~~$\rho_{MAX} = 0.88$~~$\rho_{MAX} = 0.87$.

8

3.2   Precipitation intensity process

3.2.1   ~~Precipitation intensity process~~

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

- marginal ~~heavy-tail~~ heavy-tailed distributions,

- a tail-dependent spatial distribution,

- an autocorrelated temporal process.

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

~~**Marginal distributions**~~

3.2.1   Marginal distributions

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

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

This distribution can be described by a smooth transition between a gamma-like distribution and ~~an~~ a heavy-tailed Generalized Pareto distribution (GPD). This transition is obtained via a transformation function, $G(\nu)$, such that the whole range of

precipitation intensities is modeled without a threshold selection (Naveau et al., 2016):

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

where

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

5   with $a_+ = \max(a,0)$, is the standard cumulative distribution function of the GPD, $\sigma > 0$ is a scale parameter and $G(\nu) = \nu^\kappa, \kappa > 0$. Thus, a 3-parameter set $\{\sigma, \kappa, \xi\}$ needs to be estimated at each station.

~~This distribution has been proposed by Papastathopoulos and Tawn (2013) under the name of extended GP-Type III distribution and has been shown to model adequately precipitation intensities (Naveau et al., 2016). Here, we refer to this distribution under the name of E-GPD.~~

10   ~~Local estimations of the GPD tail exhibiting a lack of robustness, the ξ parameter of the E-GPD is estimated using a regionalization method similar to Evin et al. (2016). Neighborhoods around each station are first obtained using homogeneity tests, following the concept of regions-of-influence (RoI) proposed by Burn (1990). The ξ parameters are then estimated using the precipitation data gathered in this region (see section ?? for details). The two remaining parameters, the scale parameter σ and the parameter of the transformation κ, are estimated at each station using a method of moments based on probability~~

15   ~~weighted moments (see Naveau et al., 2016, for further details).~~

~~**Spatial and temporal dependence of precipitation amounts**~~

3.2.2   Spatial and temporal dependence of precipitation amounts

Spatial and temporal dependence of precipitation amounts is represented using a Multivariate Autoregressive model of order 1 ~~,~~ (MAR(1)). A MAR(1) process has been used by different authors (Bárdossy and Pegram, 2009; Rasmussen, 2013) to ~~represent~~

20   ~~simultaneously~~ simultaneously represent spatial and temporal dependences. Let $\mathbf{Z}_t$ denote a vector of K Gaussian ~~variates with mean 0. The~~ random variables with mean 0 defined as:

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

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

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

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

can be expressed as:

$$\Omega_Z = \mathbf{M}_0 - \mathbf{A}\mathbf{M}_0'\mathbf{A}, \tag{11}$$

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

5   Elements of $\mathbf{Z}_t$ are first obtained using the following transformation:

$$Z_t(k) = \Phi^{-1}\big[\hat{F}\{Y_t(k)\}\big],$$

where $\hat{F}$ is the empirical distributionfunction and $\Phi[.]$ indicates the standard normal cumulative distribution function. Innovations ϵ_t are often assumed to follow a standard multivariate normal distribution. However, the upper tail dependence of the multivariate normal distribution is 0, which means that extreme precipitation amounts simulated at the different sites are not spatially

10   dependent. To introduce a tail dependence between at-site extremes, a possibility is to use a Student copula to represent the dependence structure of ϵ_t, providing an additional parameter, ν, related to the tail dependence. Both dependence structures will be considered in the following.

Using the

3.3   Parameter estimation

15 #### 3.3.1   Occurrence process

Following Wilks (1998), parameters related to the occurrence process $X_t(k)$ are estimated using the method of moments, $\mathbf{M}_0$ is i.e. using the empirical counterparts of the parameters. Observed states are first obtained using a low precipitation threshold (e.g. 0.2 mm). The matrix Π of transition probabilities are then estimated directly by pairwise covariances between the elements of $\mathbf{Z}_t$ using the Kendall's rank correlation τ, which can be directly related to the Pearson correlation coefficient ρ_P for elliptical

20 distributions (McNeil et al., 2005, p.97):

$$\rho_P = \sin\left(\frac{\pi}{2} \times \tau\right),$$

including Gaussian and Student multivariate distributions. The Kendall's τ does not depend on the marginal distributions, unlike the linear Pearson correlation ρ_P, and has the advantage to be a robust estimator of the degree of dependence, since it is calculated from the ranks of the data alone. Since Ω_Z is not necessarily the proportion of wet days $X_t(k) = 1$ following

25 observed sequences $\{X_{t-1}(k), \ldots, X_{t-p}(k)\}$. Concerning the spatial occurrence process, $\hat{\rho}_{kl}$ estimates are obtained using the empirical counterparts of π_{00}, π_0 and π_1 (see Eq. 6), which correspond respectively to the proportion of days for which dry states are observed simultaneously at two locations ($\hat{\pi}_{00}$) and to the proportions of dry days $\hat{\pi}_0$ and wet days $\hat{\pi}_1$. The correlation matrix $\hat{\Omega}_X$ is then composed of the cross-correlations $\hat{\omega}_{kl}$ obtained for all possible pairs of stations. If $\hat{\Omega}_X$ is not positive-definite, the closest positive-definite matrix is considered (Rousseeuw and Molenberghs, 1993; Rebonato and Jaeckel, 2011).

Furthermore, the seasonality of the occurrence process is taken into account by estimating these parameters on a monthly basis.

3.3.2 Intensity process

E-GPD distributions are first fitted to precipitation amounts available at each location k. Local estimations of the GPD tail
5   exhibiting a lack of robustness, we propose to estimate the ξ parameter of the E-GPD (see Eq. ~~11) , the closest positive-definite~~ ~~matrix is taken as the covariance matrix of ϵ_t if necessary.~~ 8) using a regionalization method similar to that of Evin et al. (2016), which can be summarized as follows:

~~Innovations ϵ_t are often assumed to follow a standard multivariate normal distribution, which means that their dependence~~ ~~structure is modeled by a Gaussian copula. However, the upper tail dependence of the Gaussian copula is 0, which means that~~
10   ~~extreme precipitation events simulated from~~

1. Following Burn (1990), for each station, a region-of-influence (RoI) is delimited by a circle around the site, the radius being determined using homogeneity tests. All the stations inside this RoI are then considered homogeneous up to a ~~Gaussian copula are not spatially dependent. This motivates the use of a Student copula to represent the dependence~~ ~~structure of ϵ_t,~~ scale factor.

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

This regionalization method is applied to the precipitation data available from 666 stations in Switzerland, for ~~which an~~ ~~additional parameter, ν, is related to the tail dependence. Given Ω_Z~~ 4 different seasons:

  – **Winter:** December, January and February,

20   – **Spring:** March, April and May,

  – **Summer:** June, July and August,

  – **Autumn:** September, October and November.

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

Concerning the spatial and temporal dependence of precipitation amounts, direct estimates of $\mathbf{M}_0$ and $\mathbf{A}$ cannot be obtained since non-zero precipitation amounts $Y_t(k)$ are not observed. Here, we follow the methodology proposed by Wilks (1998) and
30 Keller et al. (2015). For each pair of stations, we generate long sequences of precipitation amounts $P_t(k)$ using the estimated

parameters of the occurrence process ($\hat{\Pi}$ and $\hat{\omega}_{kl}$), the parameters of the marginal distributions and a correlation coefficient $m_0(k,l)$ indicating the degree of spatial dependence. Similarly to the occurrence process, $\hat{m}_0(k,l)$ is then found iteratively by matching the correlation between these long random streams with the observed correlation $\text{Corr}(P_t(k), P_t(l))$ (see Wilks, 1998; Keller et al. The correlation matrix $\hat{\mathbf{M}}_0$ is then composed of the cross-correlations $\hat{m}_0(k,l)$ obtained for all possible pairs of stations. For

5   each station, the estimates of the lag-1 serial correlation coefficients of the matrix $\mathbf{A}$ are obtained using the same simulation approach.

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

10   5.5.3.).

3.3.3 ~~Model versions~~

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

3.4 Model versions

15   Different versions of the proposed multi-site precipitation model are considered in this paper, each corresponding to different extensions of the Wilks model. A flowchart summarizing the increasing complexity of these models is presented in Figure 3. ~~The performances of these different versions will then be presented in Section 5.~~

~~Wilks~~

3.4.1 Wilks

20   A first benchmark version of the multi-site model, referred to here as 'Wilks', is considered~~, which~~. It closely matches the multi-site model proposed by Wilks (1998). In particular:

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

– ~~A threshold of 0.2 mm separates dry and wet states.~~

– The marginal distribution on precipitation amounts is a ~~mixed-exponential~~ mixture of exponential distribution, for which

25   the pdf is defined as:

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

The parameters w, β_1 and β_2 are estimated using the ~~Expectation-Maximisation~~ Expectation-Maximization (EM) method (Dempster et al., 1977).

13

– Precipitation amounts are not considered to be temporally correlated, i.e. the matrix **A** in ~~equation 10~~Eq. (10) is a zero matrix. Furthermore, innovations ϵ_t follow a standard multivariate normal distribution and represent the spatial correlations.

~~GWEX-1D~~

5 ~~A first version of the GWEX model presented in this section, labeled GWEX-1D,~~

3.4.2   Wilks_EGPD

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

3.4.3   GWEX

The initial GWEX model has the following ~~specifications~~characteristics:

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

15 – ~~A threshold of 0.2 mm separates dry and wet states.~~

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

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

~~GWEX-3D~~

~~As will be shown in Section 5, GWEX-1D model tends to underestimate extreme amounts for different temporal scales (e.g. 3~~
20 ~~days). It motivated the investigation of~~

3.4.4   GWEX_Disag

In this paper, an alternative version, ~~GWEX-3D. GWEX-3D~~ referred to as GWEX_Disag, is also proposed. GWEX_Disag is applied to 3-day precipitation amounts ~~, with the same specifications than GWEX-1D~~and has the same characteristics as GWEX, except that:

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

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

With ~~GWEX-3D~~GWEX_Disag, daily scenarios are first generated at a 3-day scale ~~, and~~ and then disaggregated at a daily scale using a method of fragments ~~(see, e.g., Buishand, 1991)~~(e.g. Wójcik and Buishand, 2003). Simulated 3-day amounts are disaggregated using the temporal structures of the closest observed 3-day amounts, in terms of similarity of the spatial fields. The ~~details of~~ same observed 3-day sequence is thus used to disaggregate the 3-day amounts simulated at the 105 stations, which ensures the spatial coherence of these disaggregated amounts. Details of the disaggregation method are provided in Appendix A. Compared to ~~GWEX-1D, GWEX-3D model presents~~ GWEX, GWEX_Disag offers the following advantages:

– 3-day precipitation amounts are directly modeled and have a better chance to be adequately reproduced,

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

3.5 ~~Multi-scale evaluation~~
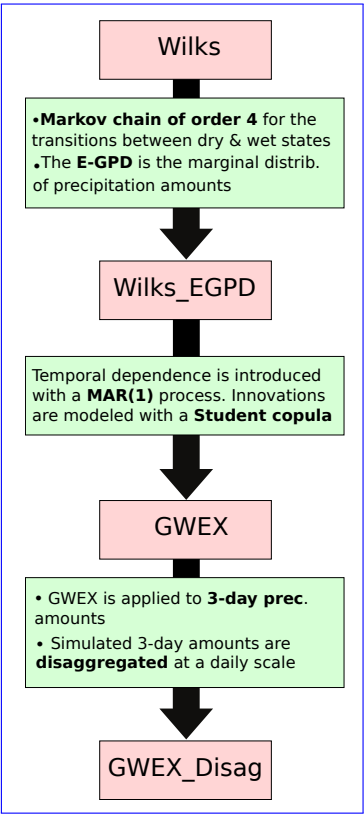
Figure 3. Flowchart of the different model versions. The differences between the models are summarized inside green boxes.

4 Multi-scale evaluation

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

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

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

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

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

Otherwise, we consider that ~~performances have been obtained, which indicates~~ performance is 'poor', indicating that the model fails to reproduce this particular statistical property.

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

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

Performance Classification	Key	Test
'good'	▪	Observed metric inside 90% limits (case 1)
'fair'	▪	Observed metric outside 90% limits but within the 99.7% limits (case 2) OR ~~Absolute~~ absolute relative difference between the observed metric and the average simulated metrics is 5% or less (case 3)
'poor'	▪	Otherwise (case 4)

For illustration purposes, we also present the results of the evaluation for three precipitation stations ~~and sub-catchments~~ corresponding to different hydrological regimes (see ~~table~~ Table 2). Figure 1 shows the 3 (~~over~~ out of 105) selected precipitation stations~~and the 3 (over 5) representative catchments~~. Station ANT (at Andermatt) is located in a glacial ~~catchment~~basin, station GLA (at Glarus) in a nival ~~catchment~~ basin and station MUR (at Muri) in a pluvial ~~catchment~~basin.

~~Two selected sub-catchments (Reuss and Limmat) include these stations and a third sub-catchment (Neuchâtel) covers the west part of the study area.~~

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

	Mean elevation [m]	Season	Triggering events
Glacial	> 1900	summer	showers + snow melt
Nival	1200 − 1900	summer, spring	showers, long rain
Pluvial	< 1200	summer	long rain

~~In this work, we focus on daily and 3-day precipitation maxima, high discharge events being usually triggered by meteorological events with a duration of several days, in late summer and autumn (Froidevaux, 2014).~~

5 ~~Application~~

5 Results

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

5.1 ~~Split-sampling procedure~~

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

~~The~~

5.1 Parameter estimation and generation of scenarios

The different model parameters are estimated with the 45-year estimation set of observations, following the methodology described in section 3.3, except for the ξ parameter of the E-GPD which is estimated using all available precipitation data in Switzerland~~, following the regionalization method described below~~. This approach ensures that robust estimates are obtained for this parameter, which is crucial in our context since extreme simulated precipitation amounts are highly sensitive to the ξ parameter. ~~All the other parameters are estimated with the estimation set of 45 years, following the methodology described in section 3.~~

5.2 ~~Regionalization of the ξ parameter~~

~~For the different stations, the ξ parameter of the E-GPD (see Eq. 8) is estimated using a regionalization method. This methodology is similar to what is proposed by Evin et al. (2016) and can be summarized as follows:~~

1. ~~For each station, a neighborhood is obtained using homogeneity tests. All the stations inside this region-of influence (RoI) are then considered homogeneous up to a scale factor.~~

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

~~This regionalization method has been applied to the precipitation data from 666 stations available in Switzerland, for 4 different seasons:~~

   - ~~**Winter:** December, January and February,~~

   - ~~**Spring:** March, April and May,~~

19

- **Summer:** June, July and August,

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

In this work, the For GWEX, the estimation of the ξ parameter is bounded between 0 and 0.25. When $\xi < 0$, the E-GPD distribution has an upper bound. When $\xi > 0.25$, extremely fat tails are obtained, which usually lead to unreasonable simulated precipitations. As shown by many recent studies (see, e.g. Serinaldi and Kilsby, 2014a), negative and high estimates of ξ are usually due to the parameter uncertainty and are not realistic.

For GWEX-1D, the estimation of the ξ parameter is performed at a daily scale. In order to highlight spatial patterns of ξ over Switzerland, we show the maps of the interpolated parameter estimates in Figure 4. Fat tails are obtained in the South and East southern and eastern parts of the Aare river catchmentRiver basin, particularly during spring and summer seasons. In the south of Switzerland, a region with high estimates ($\xi \sim 0.2$), highlighted in red, is obtained for the summer and automn autumn seasons. These high ξ estimates are coherent consistent with the presence of strong convective storms in this mountainous region during this period of the year (Rudolph and Friedrich, 2012).

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

5.2 Generation of scenarios

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

For each multi-site precipitation model investigated in this paper (Wilks, GWEX-1D and GWEX-3DWilks_EGPD, GWEX and GWEX_Disag), we generate 100 daily precipitation scenarios with these parameter estimates, each scenario having a length of 100 years. These scenarios are compared to the precipitation observed during the validation period of 40 yearsfor the 40-year validation period.
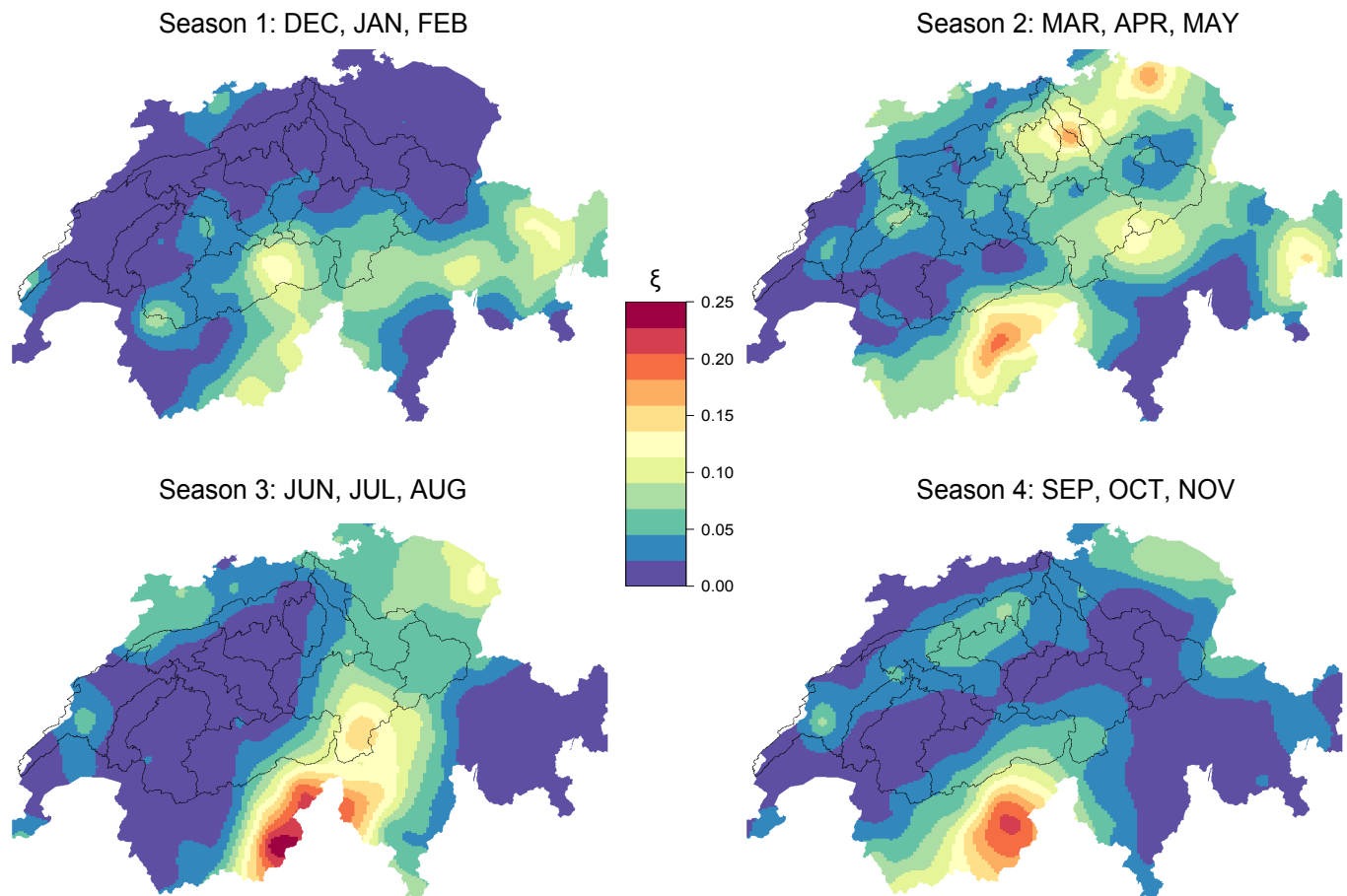
6 Results

Figure 4. ~~Regionalization of the~~ <u>Regionalized</u> ξ ~~parameter~~ <u>parameters</u> at a daily scale, for the different seasons. <u>Here, we present the spatial interpolation of at-site estimates for a better readability of their variability.</u>

~~This section presents the results of the multi-scale evaluation framework (see Section **??**) for several metrics related to the occurrence process of the precipitation events, daily amounts, monthly totals and precipitation extremes. As much as possible, synthetic assessments are provided, with several statistics being provided for all the spatial scales of interest. Illustrative examples are shown in order to support the conclusions drawn from these synthetic results.~~

5.1 Occurrence process

The ~~comparison of the~~ monthly number of wet days obtained from observed and simulated precipitation data are ~~shown~~ <u>compared</u> in Figure 7. The average number of wet days is adequately reproduced by all models, with approximately 30% of cases with 'poor' ~~performances~~<u>performance</u>. These 'poor' ~~performances~~ <u>performance cases</u> seem to occur mainly during the winter and spring seasons. The standard deviation of the monthly number of wet days indicates the inter-annual variability of this metric. While the magnitudes of the standard deviations from the simulated ~~precipitations~~ <u>precipitation</u> roughly match the
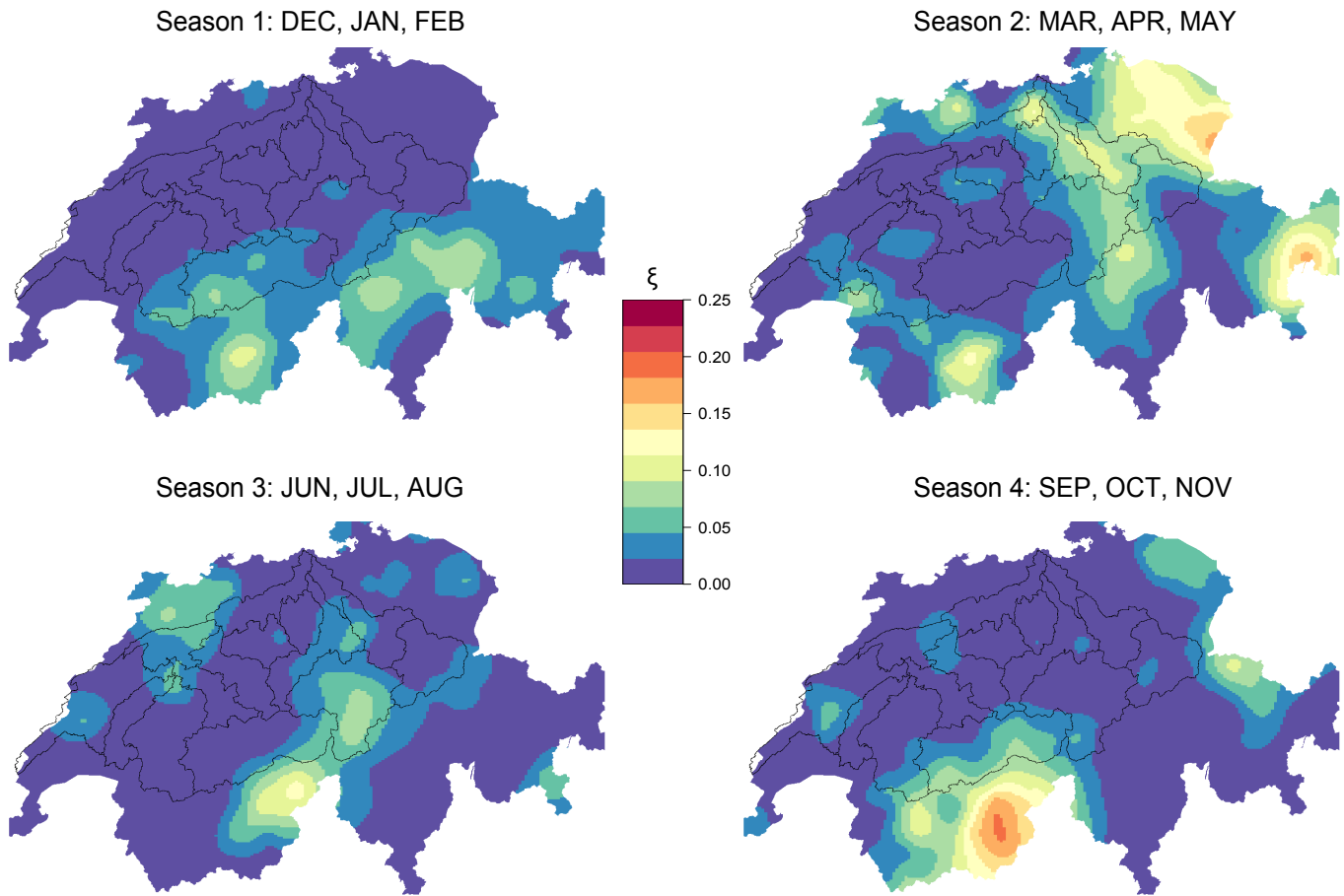
Figure 5. ~~Regionalization of the~~ Regionalized ξ ~~parameter~~ parameters at a 3-day scale, for the different seasons. <u>Here, we present the spatial interpolation of at-site estimates for a better readability of their variability.</u>

corresponding observed standard deviations, it seems that the highest observed variabilities are ~~under-estimated~~ underestimated by all the models, ~~this defect being more apparent for~~ most markedly by the Wilks model.

Figures 8 and 9 show the distributions of observed and simulated dry and wet spells, respectively, for the three illustrative stations. Concerning the distributions of dry spell lengths, ~~GWEX-1D and GWEX-3D models both~~ the Wilks_EGPD, GWEX

5 and GWEX_Disag models lead to adequate ~~performances, the performances~~ performance, the performance being classified as 'good' in 48%~~and 51~~, 48% and 49% of the cases, respectively. The performance of ~~Wilks model are~~ the Wilks model is slightly lower because of an imprecise reproduction of the frequency of the shortest dry spells. This difference ~~of performances between Wilks and GWEX-1D models~~ in performance is explained by the order of the Markov chain used to simulate the transitions between dry and wet states, which is the only difference between the occurrence processes of ~~these models~~ Wilks

10 and Wilks_EGPD or GWEX. The 4-order Markov ~~of the GWEX-1D model seems to be~~ chain of the Wilks_EGPD and GWEX
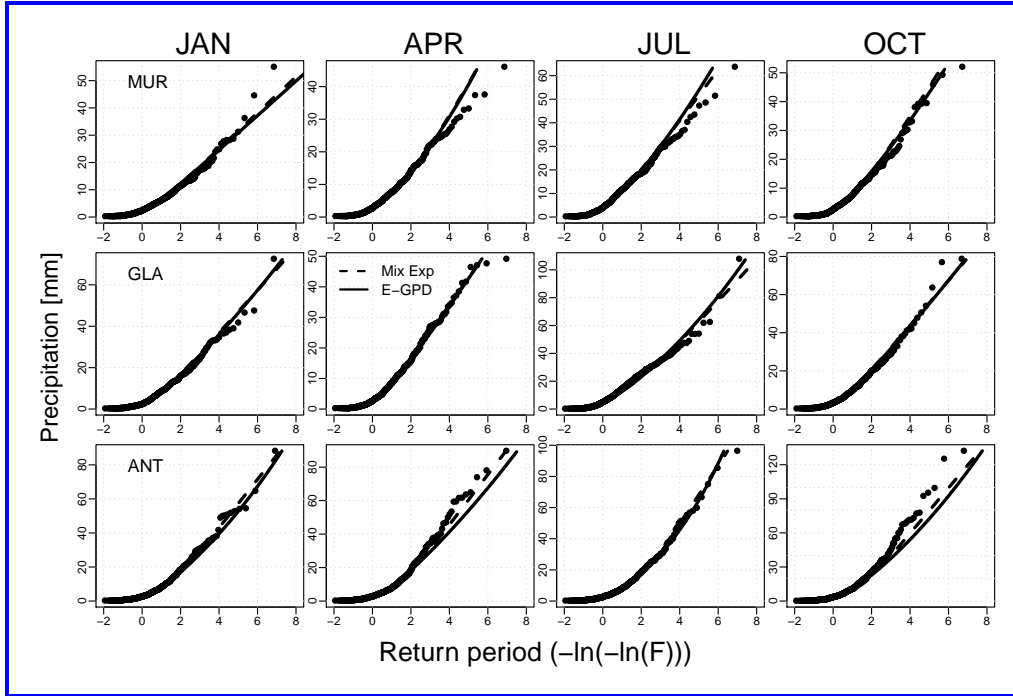
22

Figure 6. Empirical and fitted distributions (dashed curves for mixture of exponentials and solid curves for E-GPD) at a daily scale, for the three illustrative stations and for the months of January, April, July and October.

models seems to provide a more adequate representation of these transitions than the first-order Markov chain of the Wilks model, confirming previous findings (Srikanthan and Pegram, 2009).

The frequencies of wet spell lengths are adequately reproduced by ~~Wilks and GWEX-1D~~ the Wilks, Wilks_EGPD and GWEX models, with more than 50% of 'good' ~~performances~~performance. The lower overall performance of ~~GWEX-3D~~
5 GWEX_Disag for this metric is due to a slight underestimation of the longest wet spells for some stations (which is however not the case for the stations shown in Fig. 9).

5.2 Inter-site correlations of precipitation amounts

Figure 10 compares observed and simulated inter-site correlations for the different model versions. Unlagged cross-correlations, which represent the spatial dependence, are close to the 1:1 diagonal line, as expected given that these correlations are explicitly
10 taken into account by all model versions. However, a slight underestimation can be observed, especially concerning correlations above 0.8. This underestimation is a side-effect of the transformation applied to obtain a positive-definite matrix (see section 3.3).

An adequate reproduction of lag-1 inter-site correlations is important for the reproduction of persistent precipitation events. Simulated lag-1 cross-correlations are close to 0 for the Wilks and Wilks_EGPD models, as expected given that these versions
15 ignore the temporal dependence. Consequently, these two model versions significantly underestimate observed lag-1 cross-correlations,
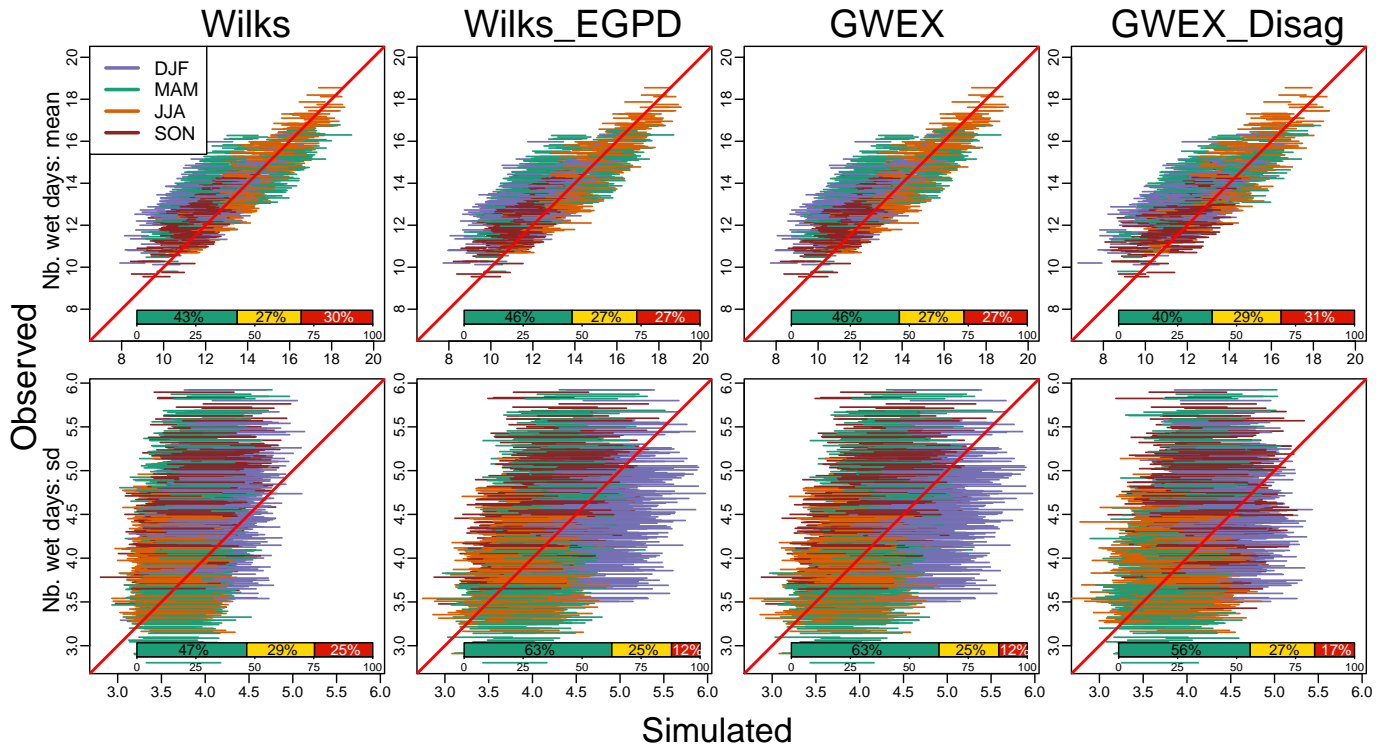
23

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

which range between 0 and 0.4. Concerning GWEX, lag-1 serial autocorrelations at the stations (black points in the bottom plots) are perfectly aligned along the 1:1 line, as expected given that they are explicitly fitted by the MAR(1) process. Simulated and observed lag-1 cross-correlations are roughly in agreement, though the largest observed cross-correlations are underestimated. This is also the case to a lesser extent for GWEX_Disag. However, the agreement between observed and simulated cross-correlations is much stronger.
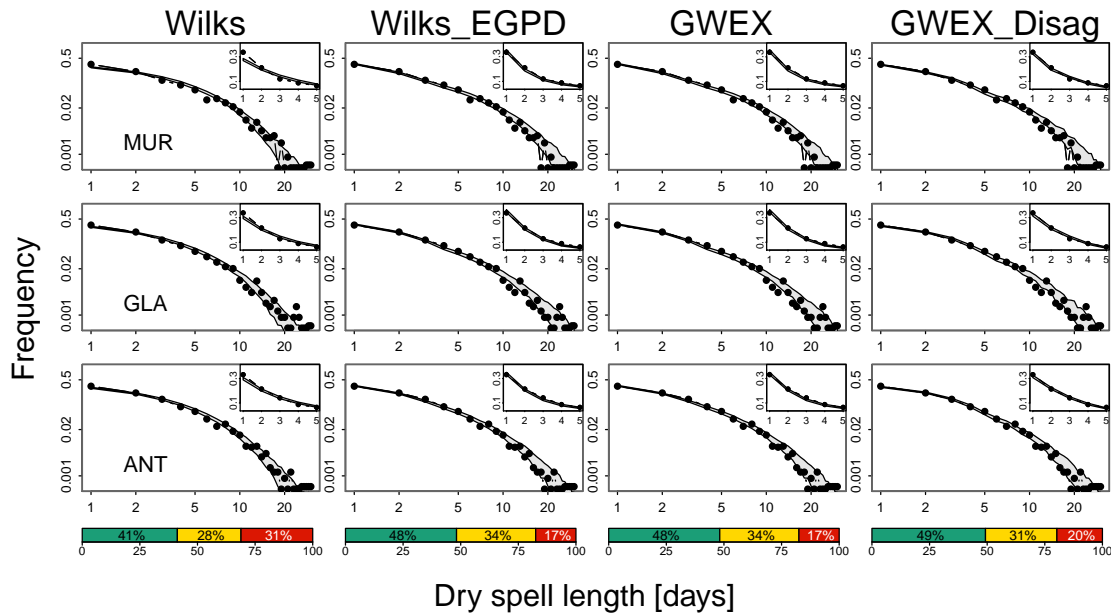
Figure 8. Distribution of dry spell lengths at the stations: ~~The~~ The 90% probability limits are shown. ~~The overall~~ Overall performance ~~represents a percentage~~ is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.
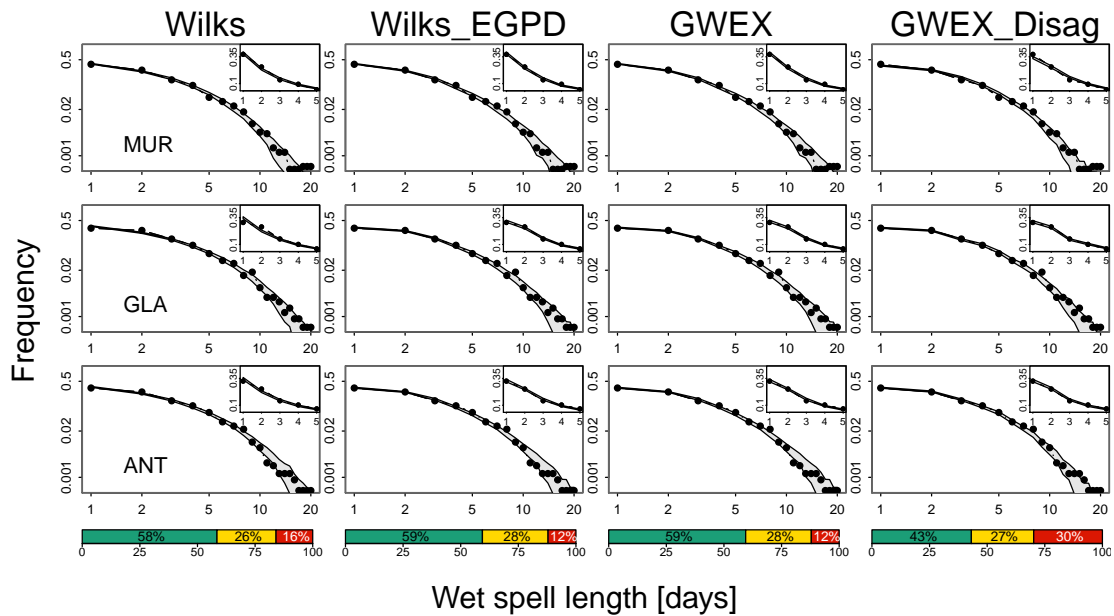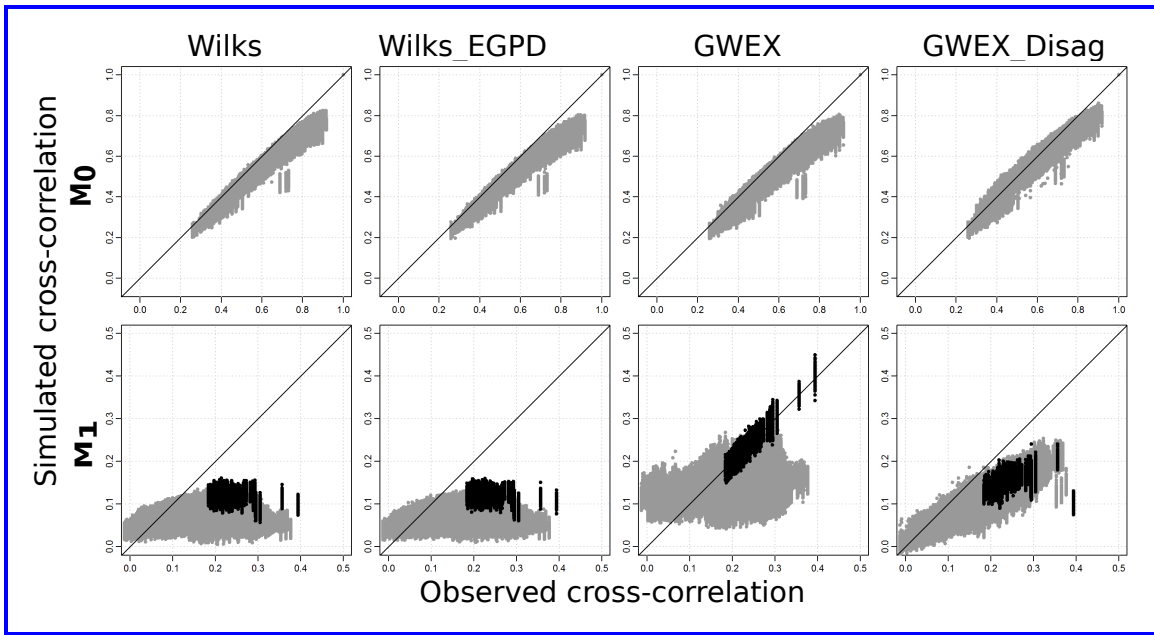

Figure 9. Distribution of wet spell lengths at the stations: ~~The~~ The 90% probability limits are shown. ~~The overall~~ Overall performance ~~represents a percentage~~ is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all sites. Inset plots provide a zoom for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

25

Figure 10. Comparison of unlagged inter-site correlations (M_0) and lag-1 inter-site correlations (M_1) in observed and simulated precipitation series, for the winter (DJF) and summer (JJA) seasons and for the different model versions considered. Black points indicate lag-1 serial autocorrelations at the stations.

5.3 Daily amounts

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

5   is ~~are~~ also adequately reproduced (Fig. 11, bottom plots). ~~However, we can notice that these standard deviations are slightly underestimated by models Wilks and GWEX-1D at the scale of the basins, which is not the case of the GWEX-3D model.~~
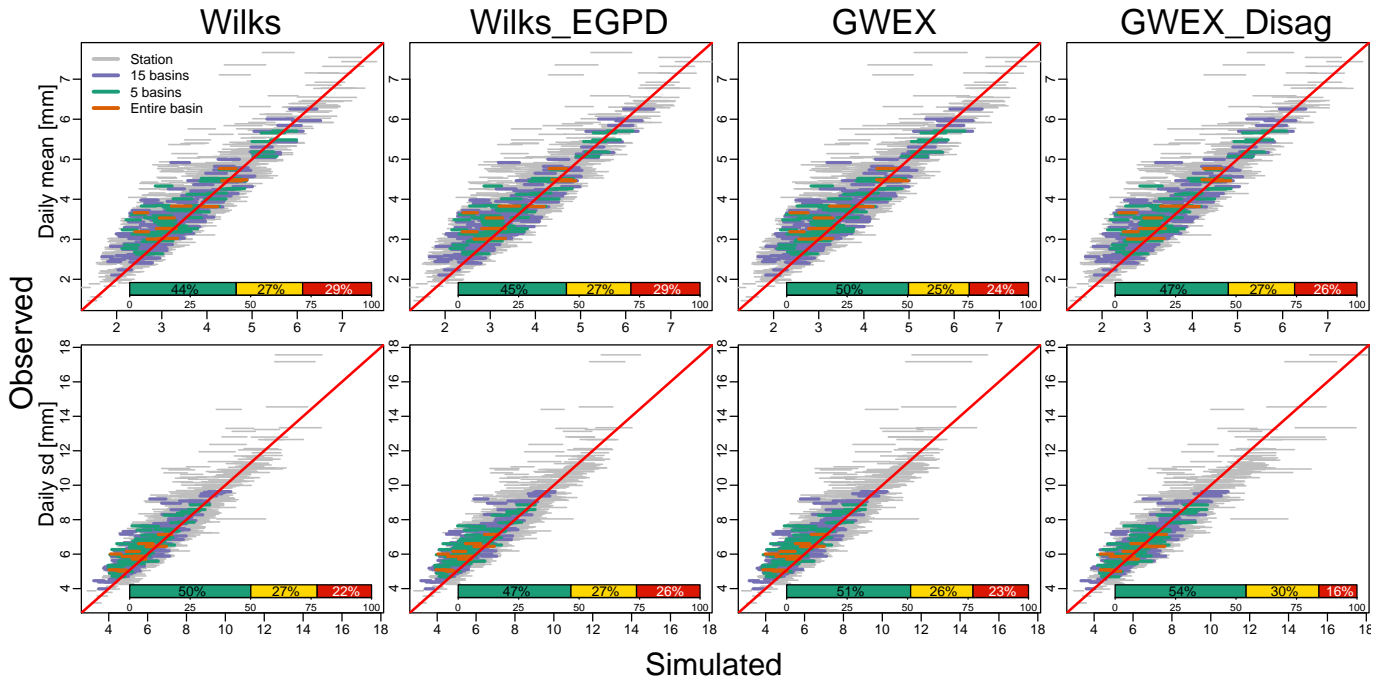

Figure 11. Daily amounts for all spatial scales and months: inter-annual mean (top) and standard deviation (sd, bottom). The 90% probability limits are shown. ~~The overall~~ Overall performance ~~represents a percentage~~ is represented by the indicated percentages of 'good', 'fair' and 'poor' performance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

5.4 ~~Inter-annual variability~~Extreme precipitation amounts

~~The reproduction of the standard deviations of aggregated precipitation amounts at a monthly scale is used to assess the inter-annual variability (Figure ??),~~ Figures 12 and 13 show the relative differences, expressed as a percentage, between observed and simulated 10-year and 50-year return periods, at daily and 3-day scales, respectively, for all spatial scales~~and months. For the winter months, the standard deviations of these monthly totals are underestimated at all spatial scales (Fig. ??, top plots). We can clearly interpret this deficiency as an under-estimation of the inter-annual variability of these aggregated amounts. This deficiency has been identified in many stochastic precipitation models (see, e.g. Wilks and Wilby, 1999; Bennett et al., 2017) and different remedies have been proposed in the literature (Mehrotra and Sharma, 2007a; Mehrotra et al., 2012). However, we can notice that this underestimation is moderate for GWEX-3D (32%~~. The percentiles corresponding to these return periods are estimated empirically using the Gringorten formula (Gringorten, 1963). These figures provide an overview of model performance regarding extreme precipitation amounts.

At the daily scale (Figure 12), there is no major difference in performance between the four models. For the 10 years and 50-year return periods, the number of 'poor' ~~cases against 79% for Wilks). Furthermore, this under-estimation of the inter-annual variability is not present for the summer months (Fig. ??, bottom plots) .~~

~~Monthly totals for all spatial scales and months: inter-annual standard deviation (sd) for the winter (DJF) and summer (JJA) seasons. 90% probability limits are shown. The overall performance represents a percentage of all spatial scales and the months corresponding to the season.~~

5.5 ~~Extreme precipitation amounts~~

~~Figures ?? and ?? show a comparison of the observed and simulated annual maximum precipitation for the three illustrative stations, at a daily and at a 3-day scale, respectively. At a daily scale,~~ performance cases is below 20% for all models. The relative differences are globally centered around zero, which means that the mixture of exponentials (Wilks model) and the ~~three precipitation models exhibit different behaviors for the simulated maxima. Maxima from Wilks are linear on a Gumbel scale, which is expected as daily intensities are generated from a mixture of exponential distributions. GWEX-1D, with the E-GDP distribution, generates larger extreme precipitation amounts than Wilks , but also than GWEX-3D. For example, at station ANT, the 95% quantile (upper limit of the 90% intervals)~~E-GPD (Wilks_EGPD, GWEX and GWEX_Disag models) all produce reasonable performance at this temporal scale. However, if we compare the 50-year return periods simulated by the Wilks and Wilks_EGPD models, we note an increase of 10% of 'good' performance cases (from 65% to 75%), which can be explained by a slight underestimation of the largest ~~daily annual maxima obtained from the 100-year scenarios exceed 350 mm for GWEX-1D and is below 300 mm for GWEX-3D~~maxima with Wilks, for some stations.

~~At a 3-day scale, larger discrepancies can be observed between the three models (Fig. ??). In particular, observed maxima are strongly underestimated by Wilks~~at stations GLA and ANT, which is not the case (or, at least, not as clearly), for the two other models. We can thus assume that the temporal dependency~~Comparing Wilks_EGPD and GWEX, the scores are almost

identical, which suggests that the tail dependence ~~introduced by the MAR(1) process (Eq. 10) leads to a better reproduction of the largest precipitation amounts cumulated on several days. As expected, GWEX-3D performs well at a 3-day scale, which justifies the strategy consisting in fitting directly 3-day amounts.~~

~~Simulated and observed daily annual maxima at the stations: 50% and 90% probability limits are shown.~~

5 ~~Simulated and observed 3-day annual maxima at the stations: 50% and 90% probability limits are shown.~~

~~Observed and simulated annual maximum precipitation at~~ Student copula in GWEX does not produce a significant improvement for the reproduction of extremes. However, if we focus on the largest spatial scales ~~are shown in Figures ?? and ??, at the daily and 3-day scales, respectively. At a daily scale, a slight under-estimation~~ (at the basins), and in particular on the entire Aare River basin (orange lines), it seems that the slight underestimation of the 50-year return periods obtained with Wilks_EGPD

10 is reduced thanks to this tail dependence. GWEX_Disag also reproduces adequately the largest precipitation amounts at all spatial scales, even if a slight overestimation of the maxima ~~by the Wilks and GWEX-1D models~~ at the largest spatial scales can be suspected~~,~~. Nevertheless, this performance shows that the disaggregation process leads to an adequate reproduction of the daily maxima.

At the ~~simulated maxima being larger with GWEX-3D, especially at the scale of the entire Aare river catchment (bottom~~

15 ~~plots). At a~~ 3-day scale ~~, a dramatic~~ (Figure 13), the underestimation of the maxima ~~can be observed with Wilks and GWEX-1D. The slight underestimation observed at the scale of the stations, especially for the Wilksmodel, is far more severe at larger~~ by Wilks and Wilks_EGPD is clear at all spatial scales. ~~GWEX-3D~~ GWEX does not suffer from ~~such~~ the same shortcomings, which ~~can probably be explained by its direct representation of the spatial dependence at the~~ means that the MAR(1) process (Eq. 10) improves the temporal structure of the largest 3-day ~~scale.~~

20 ~~Simulated and observed daily annual maxima at the scale of the basins: 50% and 90% probability limits are shown.~~

~~Simulated and observed 3-day annual maxima at the scale of the basins: 50% and 90% probability limits are shown.~~

~~Figures 12 and 13 show the observed and simulated 10-year and 50-year return periods, at a daily and a 3-day scales, respectively, for all spatial scales. These return periods are estimated empirically using the Gringorten formula (Gringorten, 1963). These figures summarize the previous illustrations and provide a more synthetic view of the model performances regarding~~

25 ~~extreme~~ precipitation amounts. ~~At a daily scale, there is no major difference of performances between the three models. For the 50-year return periods, the number of 'poor' performance cases is below 20% for all models. However, at the~~ GWEX_Disag being fitted at a 3-day scale, ~~the under-estimation of the maxima by Wilks and GWEX-1D, as previously discussed, is clearly highlighted.~~

~~For GWEX-3D, the~~ this model logically leads to an adequate reproduction of extreme 3-day precipitation amounts. The

30 strategy consisting in simulating 3-day precipitation amounts, which are then disaggregated at a daily scale, presents several advantages:

– The model being fitted at a 3-day scale, 3-day maxima are adequately reproduced.

– As the method of fragments uses observed 3-day ~~distributions~~ temporal structures to disaggregate 3-day amounts, the daily amounts resulting from a generated 3-day maxima are physically plausible. In particular, the temporal and spatial

structures of large and persistent observed precipitation events are ~~employed, which brings a coherence~~ used, which ensures consistency between the generated extreme events at the daily and 3-day scales.

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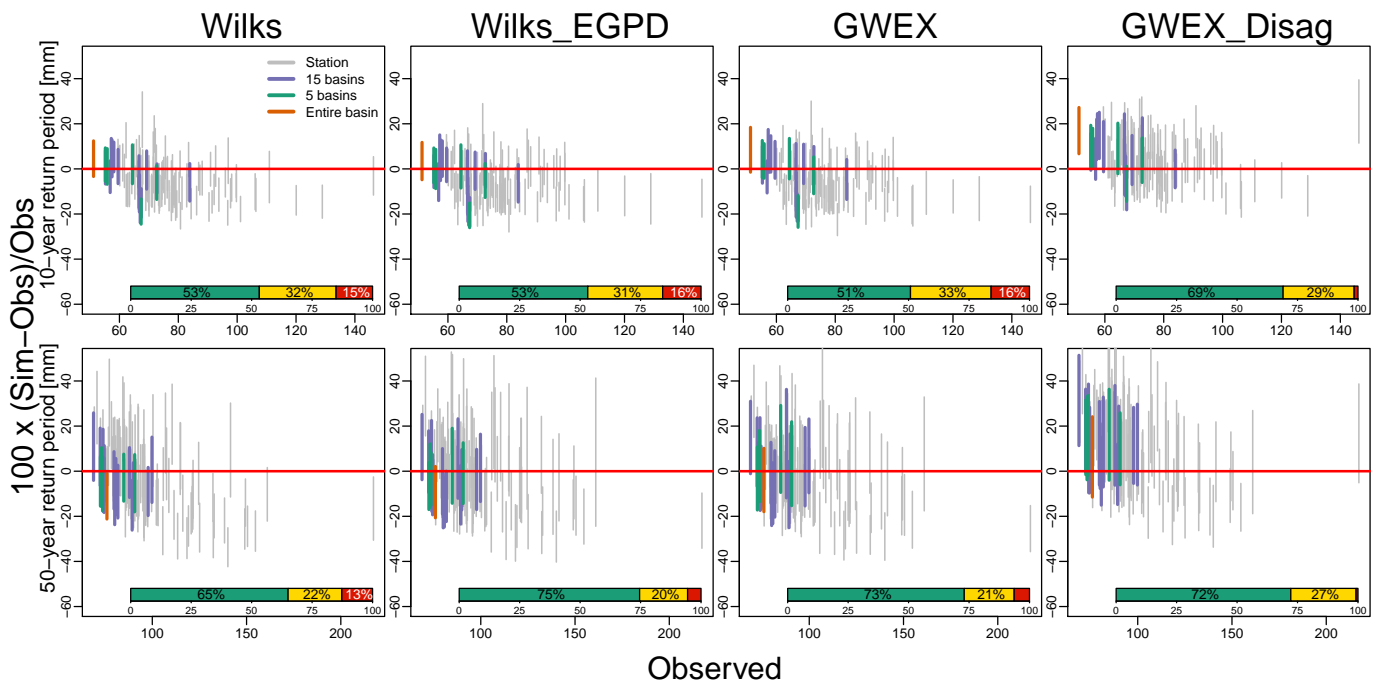

Figure 12. Daily annual maxima for all spatial scales: Relative differences, expressed as a percentage, between observed and simulated 10-year (top plots) and 50-year (bottom plots) return periods. The 90% probability limits are shown. ~~The overall~~ Overall performance ~~represents a percentage~~ is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all spatial scales.
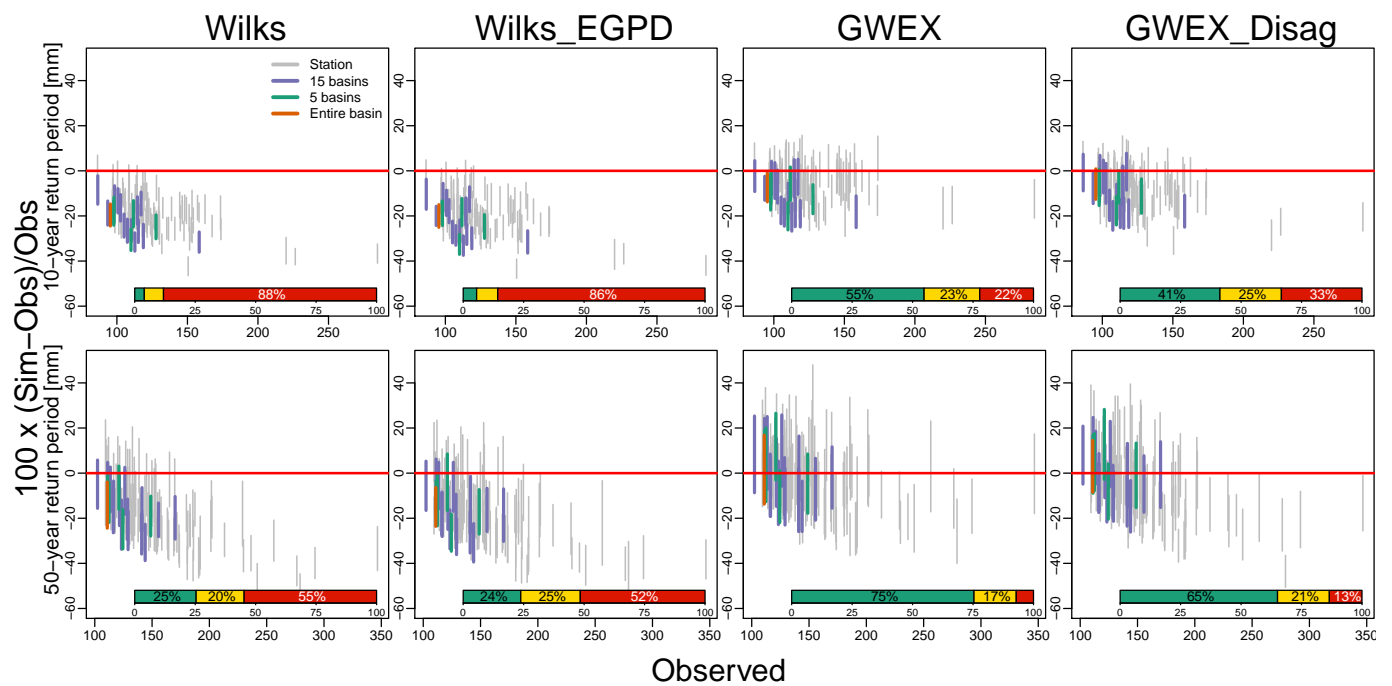
Figure 13. 3-day annual maxima for all spatial scales: Relative differences, expressed as a percentage, between observed and simulated 10-year (top plots) and 50-year (bottom plots) return periods. The 90% probability limits are shown. ~~The overall~~ Overall performance ~~represents a percentage~~ is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all spatial scales.

6 Conclusions and outlook

~~The motivation for the development of precipitation models is usually the risk assessment of natural disasters~~ Precipitation models are usually developed for the purpose of risk assessment in relation to natural hazards (e.g. droughts, floods). ~~The majority of~~ Most existing precipitation models ~~aims~~ aim at reproducing a wide range of statistical properties of precipitation, at different scales, in order to be used as a general tool in different contexts. In this study, our main objective ~~is~~ was to provide a precipitation generator ~~used in combination of an~~ that could be used together with a hydrological model for the evaluation of extreme flooding events ~~,~~ in a region covering approximately half of Switzerland. As a consequence, we ~~are~~ were especially interested in the reproduction of extreme precipitation amounts at medium to large spatial scales. As the daily and 3-day precipitation amounts are a major determinant of ~~the flood magnitude~~ flood magnitudes in large Swiss ~~catchments~~ basins (Froidevaux et al., 2015), an adequate reproduction of precipitation at these time scales ~~is~~ was also required.

In this paper, we ~~thus develop a~~ considered different multi-site precipitation ~~model~~ models targeting the reproduction of extreme amounts at multiple temporal (daily, 3-day) and spatial scales. ~~Two versions are considered, which are both based on the structure proposed by Wilks (1998). The first model version, GWEX-1D, enhance existing multi-site precipitation models using~~ Different extended versions of the model introduced by Wilks (Wilks, 1998) have been proposed. A first direct extension, Wilks_EGPD, considers a Markov chain of order 4 instead of order 1) for the at-site occurrence process. Furthermore, taking advantage of recent advances regarding extreme ~~precipitations. In particular, an~~ precipitation, a heavy-tailed distribution (instead of a mixture of exponential distributions), the E-GPD, is applied to the precipitation intensities at each station. ~~Temporal and~~ Two important extensions of Wilks_EGPD, named GWEX and GWEX_Disag, are then considered. In GWEX model, temporal and spatial dependencies of the occurrence and intensity process are introduced using the copula theory and a multivariate autoregressive process. ~~In the second model version, GWEX-3D, the same structure is applied~~ A second version, GWEX_Disag, applies the same model, but at a 3-day scale. The 3-day simulated amounts are then disaggregated using an adaptation of the method of fragments ~~(Buishand, 1991)~~ (Wójcik and Buishand, 2003).

~~GWEX-1D and GWEX-3D are compared to the multi-site precipitation model proposed by Wilks (1998). The application of a multi-scale evaluation framework leads to the following conclusions:~~

- ~~A 4-order Markov chain outperforms a first-order Markov chain for the transitions between dry and wet states, notably in terms of reproduction of dry spell lengths.~~

- ~~For winter months, the inter-annual variability of monthly aggregated amounts is clearly under-estimated by all the models. This under-estimation is not observed for summer months.~~

- ~~At the scale of the stations, daily amounts (average, standard deviations and extremes) are reasonably well reproduced by all models.~~

- ~~At a 3-day scale, precipitation extremes are severely under-estimated by Wilks and GWEX-1D. This under-estimation is observed at all spatial scales but is more pronounced at larger spatial scales.~~

~~As the GWEX-3D model outperforms the other precipitation models tested in this study, this is our recommended model for the evaluation of extreme flood events.~~

In this study, we support the ~~arguments in favor~~ use of a systematic evaluation framework. The CASE framework proposed by Bennett et al. (2017) provides ~~useful tools~~ a useful tool in this respect, making ~~possible a fair comparison of performances~~ it possible to fairly compare performance between precipitation models. Regarding the reproduction of extreme precipitation, ~~we notice that evaluations are usually~~ evaluations until now have usually been qualitative (e.g. interpretations based on one or two examples ~~are provided and interpreted) ,~~) and limited in terms of spatial scales (often only at the stations ~~only~~). The evaluation of extreme precipitation amounts proposed in this paper is multi-scale in time (daily and 3-day scale) and ~~in~~ space (at the stations, for two different ~~dissections~~ divisions of the study area into sub-basins, and for the entire Aare ~~river catchment).Illustrative examples of the reproduction of annual maxima are supplemented with synthetic representations of these performances.~~ River basin).

~~A possible enhancement of the GWEX-3D is the~~ The different multi-site precipitation models have been applied to 105 stations located in Switzerland. A multi-scale evaluation led to the following conclusions:

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

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

- With only three parameters, the E-GPD provides a parsimonious and flexible representation of the whole of precipitation amounts. Its GPD tail is in agreement with recent results showing that extreme precipitation amounts must be modeled by heavy-tailed distributions (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014a). Furthermore, robust estimates of the parameter controlling the heaviness of the distribution tail are obtained using a regionalization method. In our study area, the E-GPD does not bring a significant improvement of the ~~inter-annual variability produced by the model. The solution proposed by Mehrotra et al. (2012), which consists in using a predictor based only on the aggregated number of precipitation occurrences over the previous 365 days, seems promising. Indeed, it avoids the introduction of atmospheric predictors, and preserve the purely stochastic behavior of the model.~~ performance compared to the mixture of exponential distributions. However, the general framework proposed in this paper can be applied to very distinct precipitation regimes and the possible heavy tail of the E-GPD might be valuable in other areas.

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

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

Future research will investigate if the floods ~~resulting from~~ simulated by a hydrological model using the generated precipitation scenarios ~~through an hydrological model~~ have statistical properties in agreement with observed floods. An extensive

investigation is currently underway with a distributed version of the HBV hydrological model, applied to 87 sub-basins of the whole study area and using precipitation scenarios produced by GWEX as inputs. This hydrological evaluation of our weather scenarios will be presented in future publications.

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

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

5

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

   – **Season:** Periods $\tilde{\mathbf{D}}$ and $\mathbf{D}$ must belong to the same season, as defined in Section ~~??~~3.3.

   – **Mean intensity:** Simulated and observed precipitation fields must have the same order of magnitude. Let $q_{0.5}$, $q_{0.75}$~~and~~, $q_{0.9}$ and $q_{0.99}$ denote the quantiles of the mean observed precipitation intensities over all the stations associated ~~to~~ with probabilities 0.5, 0.75~~and~~, 0.9, and 0.99, respectively. Observed and simulated 3-day periods are

10

   classified in ~~4~~ 5 groups according to their mean intensity $\bar{\mathbf{Y}} = \frac{1}{n}\sum_k Y_{\mathbf{D}}(k)$: dry periods ($\bar{\mathbf{Y}} < q_{0.5}$), moderately wet periods ($q_{0.5} \leq \bar{\mathbf{Y}} < q_{0.75}$), wet periods ($q_{0.75} \leq \bar{\mathbf{Y}} < q_{0.9}$)~~and~~, very wet periods (~~$q_{0.9} \geq \bar{\mathbf{Y}}$~~ $q_{0.9} \leq \bar{\mathbf{Y}} < q_{0.99}$) and extremely wet periods ($q_{0.99} \geq \bar{\mathbf{Y}}$).

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

15

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

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

   This score measures the similarity between the simulated spatial field for the period $\tilde{Y}_{\mathbf{D}}(k)$ and the observed spatial field for the period $\tilde{\mathbf{D}}$ ~~, but also take~~ and also takes into account the similarity between the spatial fields for the previous days

20

   $\tilde{d} - 1$ and $d - 1$.

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

3. For each simulated period $\tilde{\mathbf{D}}$, the observed precipitation fields corresponding to the 10 lowest scores are retained. For

25

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

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

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

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

References

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

Allard, D. and Bourotte, M.: Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process, Stochastic Environmental Research and Risk Assessment, 29, 453–462, 2015.

Baigorria, G. A. and Jones, J. W.: GiST: A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data, Journal of Climate, 23, 5990–6008, 2010.

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

Bennett, B., Thyer, M., Leonard, M., Lambert, M., and Bates, B.: A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model, Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2016.12.043, In Press, 2017.

Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation, Hydrology and Earth System Sciences, 13, 2299–2314, 2009.

Buishand, T. A.: Extreme rainfall estimation by combining data from several sites, Hydrological Sciences Journal, 36, 345–365, 1991.

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

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., and O'Connell, P. E.: RainSim: A spatial–temporal stochastic rainfall modelling system, Environmental Modelling & Software, 23, 1356–1369, 2008.

Burton, A., Fowler, H. J., Kilsby, C. G., and O'Connell, P. E.: A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts, Water Resources Research, 46, W11 501, 2010.

Cavanaugh, N. R., Gershunov, A., Panorska, A. K., and Kozubowski, T. J.: The probability distribution of intense daily precipitation, Geophysical Research Letters, 42, 2015GL063 238, 2015.

Chandler, R. E. and Wheater, H. S.: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, Water Resources Research, 38, 1192, 2002.

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

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society. Series B (Methodological), 39, 1–38, 1977.

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

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

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O.: Flood triggering in Switzerland: the role of daily to monthly preceding precipitation, Hydrology and Earth System Sciences, 19, 3903–3924, 2015.

Gringorten, I. I.: A plotting rule for extreme probability paper, Journal of Geophysical Research, 68, 813–814, 1963.

Hundecha, Y., Pahlow, M., and Schumann, A.: Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes, Water Resources Research, 45, W12 412, 2009.

Keller, D. E., Fischer, A. M., Frei, C., Liniger, M. A., Appenzeller, C., and Knutti, R.: Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment, Hydrology and Earth System Sciences, 19, 2163–2177, 2015.

Kleiber, W., Katz, R. W., and Rajagopalan, B.: Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, Water Resources Research, 48, W01 523, 2012.

5 Klemeš, V.: Tall Tales about Tails of Hydrological Distributions. I., Journal of Hydrologic Engineering, 5, 227–231, 2000a.

Klemeš, V.: Tall Tales about Tails of Hydrological Distributions. II., Journal of Hydrologic Engineering, 5, 232–239, 2000b.

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

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

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

Leonard, M., Lambert, M. F., Metcalfe, A. V., and Cowpertwait, P. S. P.: A space-time Neyman–Scott rainfall model with defined storm extent, Water Resources Research, 44, W09 402, 2008.

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

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

McRobie, F. H., Wang, L.-P., Onof, C., and Kenney, S.: A spatial-temporal rainfall generator for urban drainage design, Water Science and Technology: A Journal of the International Association on Water Pollution Research, 68, 240–249, 2013.

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

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

Mehrotra, R. and Sharma, A.: Development and Application of a Multisite Rainfall Stochastic Downscaling Framework for Climate Change Impact Assessment, Water Resources Research, 46, W07 526, 2010.

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

Mehrotra, R., Westra, S., Sharma, A., and Srikanthan, R.: Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach, Water Resources Research, 48, W01 536, 2012.

Mezghani, A. and Hingray, B.: A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin, Journal of Hydrology,
35   377, 245–260, 2009.

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

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

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, Hydrology and Earth System Sciences, 17, 851–862, 2013.

5 Papastathopoulos, I. and Tawn, J. A.: Extended generalised Pareto models for tail estimation, Journal of Statistical Planning and Inference, 143, 131–143, 2013.

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

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

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

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

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

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

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

20 Sharif, M. and Burn, D. H.: Improved K -Nearest Neighbor Weather Generating Model, Journal of Hydrologic Engineering, 12, 2007.

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

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

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

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, Water Resources Research, 43, W07 402, 2007.

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

30 Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Progress in Physical Geography, 23, 329–357, 1999.

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

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K.: A technique for generating regional climate scenarios using a nearest-
35 neighbor algorithm, Water Resources Research, 39, 1199, 2003.