14.04.2020

manuscript nhess-2019-340

Cover Letter and Reply to the Referees

Dear Prof. Keiler,

we would like to thank you and the two reviewers (Martina Kauzlaric and anonymous) for taking the time to critically read our manuscript and to provide valuable and constructive feedback. We have addressed the comments and suggestions of referees in this document. Among others, we provide an additional subchapter about the performance of the continuous modelling approach (WeGen), extended the discussion in various points following the suggestions of the reviewers and revised the figures with larger font sizes to improve the readability. We hope the revised manuscript has further improved by the comments and suggestions.

Please, find below the reply to the reviewers and your additional questions (answers in italics). In addition, a separate track change manuscript file is attached to this document.

As discussed, we decided to add Kristian Förster as a co-author to the manuscript. Although his contribution was important, in particular in the early stage of our research, we unfortunately did not include him as a co-author. This was my personal mistake as leading author.

Kristian contributed to the development of the continuous modelling chain and the rainfall runoff modelling. He was responsible for the development and programming of the spatial interpolation scheme of the meteorological data. Finally, he took part in the revision process. To improve transparency, the contribution section of the manuscript was revised with more details of each individual role.

Yours sincerely,
Benjamin Winter

Also on behalf of my co-authors Klaus Schneeberger, Kristian Förster, and Sergiy Vorogushyn.

Revised Authors contribution:

Based on the initial ideas of KS and SV, the study was designed in collaboration of all authors. BW prepared the initial data, implemented and applied the continuous modelling approach and analysed the results. KF programmed the spatial interpolation scheme for the meteorological data and supported the rainfall-runoff modelling. The risk model and the HT-application were mainly developed by KS. The manuscript was drafted by BW with support of SV. All authors contributed to the review and final version of the manuscript.

Reply to Margreth Keiler, Editor

a) Damage has no plural and thus the term 'damages' has a different meaning. please check and adapt accordingly

The manuscript was changed accordingly.

b) The second sentences in your abstracts is not clear if you consider the risk definition your study is based on.

Thank you for this comment. We reformulated the second sentence in the abstract to: "To estimate the risk, i.e. the probability of damage, flood damage needs to be either systematically recorded over long period or it needs to be modelled for a series of synthetically generated flood events" In addition, we extended the paragraph in the introduction.

c) Please provide more information on section 3.3 and 3.4 because without reading Schneeberger et al. (2019) this part is not clear. Furthermore, why did you chose the approach of Borter (1999) which overestimates the risk, and did not consider models like FLEMOps or others?

Some further information was added to section 3.3 and 3.4. The damage model of Borter (1999) was chosen as it originates from Switzerland which is a direct neighbor to the Austrian province Vorarlberg with a similar topographic situation and building structure. Nonetheless, other damage models such as FLEMO could have been used as well. The study rather focuses on the comparison of the results between the two different approaches for the generation of flood event series. Nonetheless, in the best case local functions could be derived, however, the data basis is not available. The notion about the choice of the damage model was also added to the manuscript.

d) Please provide more information why you think a comparison on risk level allows more insights than on hazard level. In this context, I miss a discussion on the effect of the spatial distributions of elements at risk to your results.

Thank you for this good comment. In the study area, the settlements are concentrated alongside the larger valley areas (especially the Rhine valley; c.f. map Figure 1). Thus, the damage corresponding to an event is largely influenced by the region affected. If the comparison is conducted on a hazard only, the impact of wide spread flood events may be overestimated, while the impact of spatially limited events in densely populated areas are underestimated. This discussion was added to the Manuscript (P.13 L5-9)

e) For any currency values, both in text, tables, and figures, tell the reader what year these have been normalized to.

Thank you for this comment. The building values are indexed to the year 2013 (Huttenlau et al. 2015) based on the austrian construction price index (Statistik Austria 2019). This information was added to the section 3.3 and to the caption Figure 6 in the manuscript.

f) Page 10, line 4: please adapt the equation format according to the guidelines

The format of the equation was adapted according to the guidelines on "simple expressions in the body of the text"

Interactive comment, Dirk Diederen

In your paper, the river discharge approach is referred to as the HT-model approachand the weather generator as the WeGen. As you are aware, we also used the HTmodel in our river discharge generator (https://www.natural-hazards-and-earth-system-sciences.net/19/1041/2019/). However, our recent weather (precipitation) generator(https://link.springer.com/article/10.1007/s00477-019-01724-9) also makes use of the (purely statistical) HT-model, so this referencing may cause some confusion in the future. It might be better to refer to the river discharge approach as the RDGen approachC1or something similar.

We thank Dirk Diederen for his notion. We can in principle follow his argument as the HT model is a general statistical model to describe tail dependence. It can be applied to different variables besides discharge peaks. In your cited paper (Diederen et al., 2019), the HT model is however one of the pieces of a larger model to generate synthetic events, which includes e.g. spatial event identification, derivation of event characteristics and construction of new spatially consistent events. In their case, it would indeed be not suitable to name the entire model as HT model. In our case, HT model is used to describe dependence of peak discharges and construct synthetic scenarios and basically represents the hazard module. We indicate that the HT model is used as a hazard module within the PRAMO – probabilistic risk model. WeGen is an alternative hazard module based on a weather generator approach. We therefore would like to keep this notion of 'HTm' referring solely to the hazard module, also to be consistent with the previous works of Schneeberger et al. (2018, 2019), who described in details the setup of the HT model as a part of the more comprehensive PRAMo modelling framework using the very same notation.

Reply to Anonymous Referee #1

General Comments

Winter et al. presented in their paper ("Event generation for probabilistic flood risk modelling: multi-site peak flow dependence model vs weather generator based approach") two approaches to simulate distributed flood risk throughout rural catchments. The manuscript is well written and structured. The methods are sounds and the models used are well established. The results are clearly presented and the conclusions are supported by the results and discussion. The novelty of the paper is not with the development of new methods, but the use of available methods (that are common in hydrological sciences) in the context of risk assessments. I believe that the application presented here will be of interest to the natural hazard community and fall within the scope of NHESS. Below please find some suggestions for the authors to consider. I recommend minor revisions.

First of all, we want to thank the anonymous reviewer for taking the time to critically read our work and to provide additional suggestions for further improvement of the manuscript. We addressed the comments in the revised manuscript.

Specific comments

---

1. Introduction – You do compare the two distributed approaches to the "traditional" approach, but this is not clear from the introduction. I suggest adding a sentence mentioning this.

Many thanks for this comment. We revised the sentence in the introduction section to clarify flood loss is also compared to a homogeneous flood scenario is named "traditional" approach.

Please see P3. L.3-4: "Additionally, the flood risk corresponding to homogeneous flood scenarios of certain return periods ("traditional" approach) is derived and compared to the other two approaches."

---

2. Discussion – Many other models, besides the HT-model and the WeGen model, can be used to estimate distributed risk. For example, one can use a different WG model (say the AWE-GEN model) and a different hydrological model (say the HBV model) with a different outcome – e.g. that the WG-hydrological model approach will systematically underestimate the risk computed by the HT-model. I suggest adding another paragraph in the discussion section, discussing how general are the results of this study.

We agree, that the same framework with different components (weather generator or RR-model) likely lead to alternative results. The outcome of higher systematic risk estimates for the WeGen approach might not be true for other model components. We addressed this important issue in an additional paragraph in the discussion section.

---

3. The WeGen model simulates temperature, but do you use it as input into the hydrological model? It is not clear from the text. If not, I would remove all text mentioning the temperature simulation to avoid confusion.

The conceptual RR-Model HQsim is forced by temperature and precipitation. The information was added to the "Hazard Module II: WeGen" section.

---

4. Some justification is needed for the choice of the HQsim model. Is it able to capture well extreme runoff events? Please discuss the advantages and limitations of using a conceptual semi-distributed rainfall-runoff model to simulate floods.

Thank you for the valuable suggestions. The model was used in different studies regarding extreme runoff in alpine study areas and is inter alia applied for the prognosis system of the Inn River (e.g. Senfter et al. 2009; Achleitner et al. 2012; Bellinger et al. 2012; Dobler and Pappenberger 2013;

Winter et al. 2019). An additional subchapter was added to summarizing information regarding the performance of the weather generator, the disaggregation procedure and the hydrological modelling (see also Referee#2).

Fully distributed, physical based models (e.g. WaSim) will probably perform better in describing certain hydrological process such as for example the evapotranspiration or snowmelt process by using energy balance approaches. In contrast, a conceptual description of the hydrological processes (for example in HQsim) does not need all meteorological variables to solve a full energy balance (temperature, precipitation, radiation, humidity and wind speed). A further increase in model complexity will likely compromise the model parameter identifiability, increase calibration effort and computation burden. The computational efficiency is of major concern for the long-term continuous hourly discharge modelling. The advantages and limitations of choosing a conceptual model are now addressed in the discussion section. (see P.15)

---

5. [page 7, line 25] Terminology: an ensemble of 100 realizations, each consists of 42 members (years). Also later in the text, replace "repetitions" with "realizations".

The term "realizations" was used as suggested.

---

6. Figures 3 and 7. Please use a larger font size for the axes labels.

All Figures have been revised with larger font sizes to increase readability.

---

7. [13, 27-31]. I suggest adding in Figure 5 the known losses from the records (e.g. the August 2005 event) and discussing the models' performance in comparison to the "known" risk. It will give another dimension (from an "expert" knowledge) of the abilities of the different models in assessing the risk.

We agree that a "traditional" validation against known losses would be of great value. Unfortunately, we do not have reliable numbers of the loss event which are directly comparable to the model output. We tried to validate the model based on an insurance portfolio. The portfolio is however only a subset of the overall elements at risk and due to rather low sublimits (maximum payouts) for most objects, the full losses remain unknown. Finally, without a larger set of loss events it is not possible to assign a meaningful return period to the 2005 event to "validate" the risk outcome in a traditional way. This consideration was added to the "Discussion" section (P.17 L8-12).

Reply to Martina Kauzlaric (Referee #2)

General comments and recommendation

The manuscript by Winter et al. presents an interesting comparison of two quite different approaches for estimating flood risk in a probabilistic framework, both valid and currently established in the research community. The application of these models in a complex environment such as the alpine area and coupling it to a risk assessment is indeed new to my knowledge, and I congratulate the authors for their work! The manuscript is well structured, the methods are described in a comprehensible way or supported by relevant sources, and the discussion provides good points; however, generally there are quite a lot of relevant numbers/numeric information as well as background information missing. While the authors neatly site sources, they force the reader to go to look for important and relevant information too often, what is laborious and time-consuming first, and an important drawback for both the evaluation and appraisal of the results, resulting in lacking some important considerations in both the results and the discussion parts. This is a pity, because with not too much effort, you might considerably improve the manuscript and better convey relevant take-home messages. The manuscript generally features high-quality and interesting figures, which however would be more easily readable by using larger fonts. In general, I found quite a few typing errors. I am reporting all those I found in the technical corrections, but I would generally suggest the authors to read the manuscript thoroughly again. Because of these considerations, I think the manuscript requires further work before it can be recommended for publication. Please find my specific and technical comments here following. Please, don't get scared, some are only suggestions, and some questions are out of curiosity or eventual misunderstanding.

First of all, we want to thank Martina Kauzlaric for her comprehensive review with many questions and valuable suggestions to further improve the manuscript. Many thanks also for your positive and motivating words. In this reply, we address all questions and revised the manuscript accordingly. The Figures have also been revised with larger font sizes to improve the readability.

Specific comments

Introduction:

Please state more clearly the limits and the frames of your application (up to which return period and up to which spatial extent do you think these approaches are applicable and transferable -with this set up? In particular: what do you aim at?)

Many thanks for this comment. In our opinion there are no definite limits in terms of return period for both approaches. As mentioned in the introduction, risk estimates for large study areas are beside public authorities especially relevant for the (re-)insurance industry. The so called EU Solvency II regulation defines the loss associated with a 0.5% occurrence probability over a one-year period (RP200) as requirement for internal risk management (European Union (EU) 2009). This specific information was added to the Introduction.

The applicability, however, is strongly depending on available data, whereas in general complexity rises with spatial scale (see Discussion P12-L14/15) and uncertainty rises alongside the extrapolation of the data (this issue was added to the subchapter "Comparison of risk curves"). The weather generator approach is currently limited to about 500-600 climate stations and the spatial scale of 1000x1000 km. A high number of climate stations makes the approach computationally intractable. Moreover, the correlation of daily precipitation over distances beyond about 800-1000 km in Central Europe tend to zero on average that results in random spatial precipitation patterns. However, for specific events, the spatial structure still may be retained over large distances.

P2-L17/18: you might want to also cite more recent literature such as Brunner et al. 2019: Modeling the spatial dependence of floods using the Fisher copula, https://doi.org/10.5194/hess-23-107-2019

We added the citation to the manuscript.

P2-L22 Here also there is some more recent literature, such as Evin et al 2018 (Stochastic generation of multi-site daily precipitation focusing on extreme events, https://doi.org/10.5194/hess-22-655-2018) or appeared very recently Raynaud et al.2019 (Assessment of meteorological extremes using a synoptic weather generator and a downscaling model based on analogs, still under discussion, https://doi.org/10.5194/hess-2019-557)

Many thanks for the good literature suggestions. The citations were added accordingly.

Other two options for generating spatially distributed meteorological fields, more physically based–but also more computationally intensive-, would be to use the output of either global circulation models (e.g. Felder et al.2018, From global circulation to local flood loss: Coupling models across the scales, https://doi.org/10.1016/j.scitotenv.2018.04.170) or of hindcast archives (e.g. National Flood Resilience Review. Tech. rep. HM Government, September 2016. url: https://www.gov.uk/government /publications/nationalflood-resilience-review) and downscale these to the required spatial resolution.

Many thanks for this comment. Of course, global and regional climate models represent another way to generate synthetic meteorological fields. However, due to their much longer computational times, only a few realizations of typically about 100 years lengths are feasible. Stochastic weather generators and the HT-model have an advantage to generate hundreds and thousands of possible realizations needed to robustly estimate flood risks. We think, the stochastic methods are still in advantage compared to the climatic models for the purpose of risk assessment.

Study area:

From the map in Figure 1 it seems you are also simulating the Rhine at Lustenau, is this true? If yes, I assume you are using observations at some gauging station upstream of the inflow of the Ill river into the Rhine? If this is the case (and in any case?), I think it would be quite important to note this later on, as the nodes/communities simulated downstream in the Rhine valley should be considered a bit' differently.

Thank you for this important question. Based on the available risk maps for Vorarlberg, there are no inundation areas designated for the river Rhine due to its high protection level (up to HQ300; see e.g. http://vogis.cnv.at/atlas/init.aspx?karte=wasserbuch&ks=gewaesser, Vorarlberg Map Service in German). The gauge Lustenau (Höchster Brücke) at the river Rhine is actually included in the HT-modelling procedure however is not connected to any community node points and also not included in the hydrological modelling framework of the WeGen approach. Even if they are not considered in this study, there is a risk of dam failures, which could have a devastating effect in Vorarlberg. We added an additional statement to the Discussion section of the manuscript (P.17 L3-7).

Methods and Data:

Please introduce how many meteorological stations and gauging stations are available for this study, for how many years, instead of first mentioning it in the results part.

In total 17 gauging stations (1971-2013) are applied for the HT-approach and data of 45 meteorological stations with daily time series from 1971-2013 are included in the WeGen approach. Stations with shorter time series were not considered in the study. This information was added to the Methods and Data section accordingly. (P.4 L1-4)

Hazard Module II: please give more details about how good is the modeling chain working (refer also to comments further below under Results).

An additional subchapter was added summarizing the results of the Hazard Module II (WeGen). In the subchapter information regarding the performance of the weather generator, the disaggregation procedure and the hydrological modelling are included. More detailed information is given in Winter et al. (2019).

P6-L21/22: What do you mean concretely by saying “whereas the underlying hydrological boundary conditions are based on the considerations of the Austrian flood risk zoning project HORA”?

As stated, the model chain is not coupled (yet) with a hydraulic model to simulate inundation maps seamlessly. Instead the inundated areas and corresponding water depths for the loss calculation are taken from the “official” inundation maps provided by public authorities. The hydrological loads for the 1D and 2D hydrodynamic simulations are thereby based on the Austrian flood risk zoning project HORA (Merz et al. 2008). The paragraph was rephrased accordingly.

Please add some information about the experimental set up: 100 x 42 years for the analysis of the spatial coherence and 30 x 1000 years for the rest (and also explain why 30 x 1000 years).

For a fair comparison of quantile values between the simulation result and observed data, the time series need to be of identical length. With 42 years of data available at the gauging stations, the simulation was set up accordingly for the analysis of the spatial coherence (See P.9 L6-7).

As the simulations are theoretically not limited in regard of overall length, all further analyses are based on a total length of 1000 years. The length was chosen to be far above the highest return period of available homogeneous inundation data (HQ300) and the number of 30 realizations were calculated due to the computational limitations of the continuous simulation on an hourly time step. This information was added to the corresponding result section (See P.11 L 7-11).

Results:

On the analysis of spatial dependence: Even though it is visible from the maps you show on the diagonal, it would be more fair to mention (and consider at all?) in the text that the four stations you are showing in Fig. 3 are not completely “independent”, in the sense that Kennelbach is the downstream station of Thal and in turn Gisingen is the downstream station of Schruns. If you intentionally chose this set up –and I could see good reasons for making this choice-, please state it, and explain why.

Thank you for this comment. We choose the gauges Kennelbach and Gisingen as they are the gauges closest to the outlets of the two largest catchments in the study area. This two catchments comprising about 80% of the total study area. We intentionally choose Schruns and Thal which are subcatchments of Gisingen and Kennelbach, respectively. By choosing this examples we show two strongly related gauges in nested catchments (but never completely dependent) as well as two relatively independent gauging stations (e.g. Schruns and Thal). The explanation for the choice of the gauges has been added to the manuscript. (P.9 L8- P.10 L2)

Furthermore, by reading Winter et al. 2019 it occurred to me that first, Thal and Gisingen are actually the two stations with the worst performance both, in the calibration and in the validation periods (if the hydrological model is not able to reproduce well the hydrological features of some subcatchments, depending on the reasons for the low performance, I wouldn’t place too much confidence in the results for any other application of this, and rather ask myself if I am ev. not propagating some structural problem in the modeling chain), and second, that apparently both Kennelbach and Gisingen are influenced by hydropower operations (you also state in Winter et al. 2019 “...the influence of hydropower reservoirs cutting peak discharges, especially for the upper Ill catchment. This effect is not considered in the hydrological model set-up, but is contained in the discharge records.” => even though I would generally expect this kind of weather generator to overestimate spatial dependence in a complex mountainous environment, this information is relevant when judging the strength of the dependence shown by the WeGen approach). I think you should provide more “background” and or critical information, and accordingly discuss more critically the results. You might be actually attaching too much guilt to the weather generator and/or the WeGen approach.

It is correct that the models at Thal and Gisingen do not perform well. Both the spatial dependence and the hydrological model deficiency play a role in poor performance at the Thal and Gisingen. However, the tendency to overestimate the spatial dependence in comparison to the observed data *is present for many station pairs as for example the gauges "Hoher Steg" (NSE 0.85/0.80) and "Kennelbach" (NSE 0.73/0.69) shown in Figure 1.*

We agree that further information regarding the performance of the modelling chain will improve the manuscript. An additional subchapter was added to the "results" containing information about the performance of the weather generator, disaggregation procedure and the rainfall runoff model based on the Winter et al. (2019).

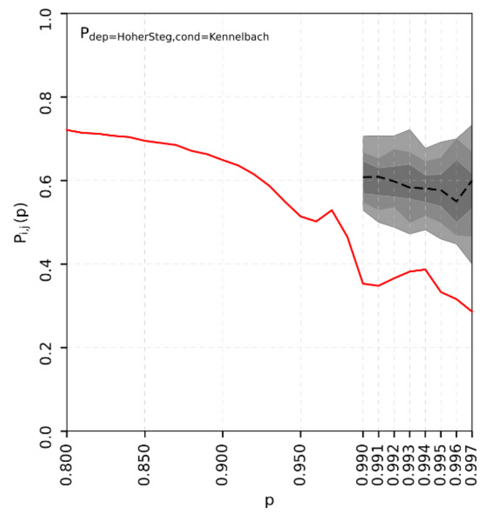

Figure 1 Spatial dependence measure observed vs. WeGen approach; conditioning site Kennelbach, dependent site Hoher Steg.

Please do correct me if I am wrong. Another information I am missing here is to what correspond the exceedance probabilities 0.99 and 0.997 (what is the return period we are talking about here? I am not sure I fully understood how you derived the quantiles, sorry if this might be a stupid question).

This is actually a very good question. The thresholds are based on the quantile values of the time series. So, in the example shown in figure 3, the 0.99 to 0.997 quantile values are based on 42 years of input data. As the data analysis is based on 3-day block maxima values to guarantee the independence of events (Schneeberger and Steinberger 2018). Accordingly, based on the empirical cdf, a p-value of 0.99 refers to a return period of approximately 1 year and a p-value of 0.997 refers to a return period of roughly 3 years. The text was revised accordingly, for a better understanding of the quantile values. (P.10)

A follow-up consideration: To my knowledge, flood protection measures in Austria are designed whenever possible against a 100 years flood event. For example, Felder et al.2017 (The effect of coupling hydrological and hydrodynamic models on maximum flood estimation, http://dx.doi.org/10.1016/j.jhydrol.2017.04.052) have shown that there might be considerable potential in re-shaping the hydrograph by coupling a simple 1D hydrodynamic model, in particular in terms of the timing of the peak. While I assume that the effect of retention in the floodplains in your study area is negligible, I would assume that this might become more important downstream for floods with return periods larger than 100 years, let's say for example in the main Rhine valley. As you also look at return periods up to 300 years in the vulnerability module, and you actually make use of inundation maps generated by hydrodynamic models, when and where do you think the coupling to a hydrodynamic model becomes relevant and what are in this sense the limits of the applicability of the WeGen approach?

Yes, we do agree with this comment that the so-called hydrodynamic interactions in the river network may affect the risk estimates, i.e. dike overtopping and failure upstream with associated inundation and water storage would reduce the risk downstream. The higher the return period of evet is, the *stronger the effect of hydrodynamic interactions is expected to be. Similarly, the larger the potential storage area in the hinterland is, which is the case for the lowland parts of the river network, the stronger the effect is. By including a 1D-2D hydrodynamic modelling, the hydrodynamic interactions can be explicitly considered in the WeGen approach. This could be a potential future extension of the modelling approach, particularly for the lowland parts of the network. On the contrary, the HT-approach is not suitable for coupling with the unsteady continuous hydrodynamic models, since it is not mass-conservative and delivers only dependent discharge peaks and not the full continuous flood hydrographs as boundary conditions. This discussion is now provided in the revised manuscript (Discussion section, P.15 L9-12).*

Please comment on the larger spread produced by the WeGen approach.

The weather generator tends to overestimate the spatial correlation of extreme precipitation. It is parameterized by using an isotropic correlation function by mixing low intensity large scale and high intensity local rainfalls. Hence, the generated fields of extreme precipitation tend to have larger spatial extent than observed. Please see the corresponding paragraph in the manuscript (P.12 L6 - 17)

Discussion:

Weather generators in general: any weather generator makes a quite strong assumption about the tail behaviour, so that the higher the return period resp. the extremeness of the simulated precipitation, the larger should be the structural model uncertainty, which in turn is expected to quite influence the corresponding estimated hydrological load. While a 100 years event might be just at the boundary of what we might be able to extrapolate from about 40 years observations – with still some degree of confidence- anything beyond will very likely be strongly related to the tail models. Could you please elaborate on this, and state what do you think might be the impact of the use of another weather generator on your results?

We agree that the extrapolation of weather generators make strong assumptions about the tail distribution and such the uncertainty raises alongside the exceedance probability which is directly propagated to the hydrological loads. The extrapolation beyond RPs of 100 years based on typically available data series of a few decades is associated with large uncertainties. Nonetheless, information about higher return periods are often required in practice (e.g. In Austria HQ300 is applied to define "residual risk areas" or the RP of 200 years is defined in the Solvency II definition in the European (re)insurance context (European Union (EU) 2009)). To our knowledge, there are no studies comparing risk assessments driven by different weather generators. Hence, it is difficult to make a reliable statement how decisive the tail dependence is with regards to the final risk estimates. On the one side, the effect of tail dependence is expected to increase with the return period. On the other side, the events with high return periods have low probability and might have little impact on the average risk (expected annual damage) (area under the risk curve). So, this is a question whether we look at the loss estimate of a e.g. 1000-year flood or we are interested in the annual expected damage. For the first, the tail dependence might be more important, for the second rather less important. We believe, more studies are needed to compare different weather generators and their impact on risk assessments.

The same framework with different components (e.g. weather generator or RR-model) likely lead to alternative results. The outcome of higher systematic risk estimates for the WeGen approach might not be true for other model components and thus should not be generalized. We added an additional paragraph in the Discussion section regarding this issue. (P14 L17-22)

P13-L26: Actually in Figure 5 you are showing the "overall" uncertainties of the two modelling chains, what do you mean with and why do you write single uncertainty sources here?

As stated on P.13-L20-21, the uncertainty presented in Figure 5 only shows the uncertainty which corresponds to the multiple realizations and does not account for other sources of uncertainties (e.g.

*parameter uncertainties). Also, no comprehensive uncertainty assessment by propagating the uncertainties of all sub-models throughout the model chain is included in the current work, it is still possible to have a look at each individual modelling step. We reformulated the statement and the sentence about '*single uncertainty sources*' was deled.*

This is just a consideration /suggestion: Of course volumes cannot be considered by applying the HT model, however besides flood peaks, flood volumes can play an important role in flood risk analysis. You correctly mention that one of the advantages of applying WeGen is the ability to produce continuous hydrographs (and accordingly event volumes), however you might want to mention it explicitly? Flood volumes play an important role for hydraulic infrastructure such as reservoirs/lakes/etc. (and thus in hydraulic design engineering), and also in the case of presence of floodplains with retention potential. On the other side, volumes might be another validation measure for the WeGen approach, as –depending also on how good is working the hydrological model-indirectly indicate how well or bad is the weather generator doing by reproducing persistence at longer time scales (a week and beyond), as I would generally expect this kind of weather generator to be underestimating persistence. This is something you might want to check in the future?

Many thanks for this interesting suggestion. We agree that flood volumes are an important characteristic of flood events and especially relevant for risk assessment (e.g. Dung et al. 2015; Lamb et al. 2016). We will elaborate more on this in the discussion. Future research can focus on the ability of the model chain WeGen-RR to reproduce flood volumes. This is however not straightforward. This can be approached by the assessment of the areal precipitation volume for different durations and spatial aggregations (from small sub-catchments to the entire basin) and secondly, by comparing the observed and simulated flood event volume statistics similarly to the flood peak statistics. We added a notion about the relevance of flood volumes beside peak estimates to describe the severity of a flood event in the discussion section.

Technical corrections

Figure 2: it is full of typing errors (refer to Obseravtions, topographie, Geometrie)

The Figure was revised accordingly.

Please use the word realizations instead of repetitions

The term "realizations" is now used throughout the manuscript.

Please use more consistently the word severity (e.g. in the of Figure 4 use return period instead of level of severity => what might be confusing, as you define and quantify severity by the UoFH later on)

As suggested the caption was changed to return period to avoid confusion.

P2-L3/4: what do you mean with floods hazard characteristics?

For example, inundation depth or flow velocity. This information was added for clarification.

P6-L24: "...a linear interpolated interpolation..." please reformulate better

The term "interpolated" was deleted.

P7-L8: ...can be statistically.

corrected

P10: please reformulate the last sentence (90% of exceeding sites sounds weird)

The sentence will be revised to: "The most widespread event (UoFH=77) corresponds to about 90% of the sites exceeding the threshold."

P11-L11: just a suggestion=> capability instead of feature?

Thank you for the suggestion. We will use the term "capability" instead of feature.

P11-L15/16: might effect ...?=> please reformulate

'effect' was replaced by 'affect'

P12-L5: …estimate of (=> better with? Or by?)) WeGen approach

The manuscript was changed to "...estimate by "

P12-L14: just a suggestion: instead of On the contrary => At the same time? On the other hand?

*Thank you for the suggestion. The phrase "*At the same time" was used instead.

P7-L21: dependence matrices instead of dependence metrices

P7-L26: Each simulation instead of Each simulations

P8-L1: the data are "too" few instead of the data are "to" few

P10-L6/7: either remove a , in "a significat lower damages.."or change damages to singular

P11-L1: …simulate (remove a) complex spatially heterogeneous patterns.

P11-L5: On the contrary. . .only indirectly …

P11-L13/14: One possible reason could ...be?

P12-L2: instead of overall estimation => overestimation?

Thank you for the grammar and language corrections. The manuscript was revised accordingly.

References

Achleitner, S., *et al.,* 2012. Analyzing the operational performance of the hydrological models in an alpine flood forecasting system. *Journal of Hydrology,* 412-413, 90–100. doi: 10.1016/j.jhydrol.2011.07.047.

Bellinger, J., *et al.,* 2012. The impact of different elevation steps on simulation of snow covered area and the resulting runoff variance. *Advances in Geosciences,* 32, 69–76. doi: 10.5194/adgeo-32-69-2012.

Dobler, C., and Pappenberger, F., 2013. Global sensitivity analyses for a complex hydrological model applied in an Alpine watershed. *Hydrological Processes,* 27 (26), 3922–3940. doi: 10.1002/hyp.9520.

Dung, N.V., *et al.,* 2015. Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta. *Journal of Hydrology,* 527, 704–717. doi: 10.1016/j.jhydrol.2015.05.033.

European Union (EU), 2009. on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II): Directive 2009/138/EC of the European Parliament and the Council. *Official Journal of the European Union (*L335).

Huttenlau, M., *et al.,* 2015. Analysis of loss probability relation on community level: a contribution to a comprehensive flood risk assessment. *In:* S.M. Sener, C.A. Brebbia, and O. Ozcevik, eds. Disaster Management 2015, 20-22 May 2015 2015, 171–182.

Lamb, R., *et al.,* 2016. Have applications of continuous rainfall-runoff simulation realized the vision for process-based flood frequency analysis? *Hydrological Processes,* 30 (14), 2463–2481. doi: 10.1002/hyp.10882.

Merz, R., Blöschl, G., and Humer, G., 2008. National flood discharge mapping in Austria. *Natural Hazards,* 46 (1), 53–72. doi: 10.1007/s11069-007-9181-7.

Schneeberger, K., and Steinberger, T., 2018. Generation of Spatially Heterogeneous Flood Events in an Alpine Region—Adaptation and Application of a Multivariate Modelling Procedure. *Hydrology,* 5 (1), 5. doi: 10.3390/hydrology5010005.

Senfter, S., *et al.,* 2009. Flood Forecasting for the River Inn. *In:* E. Veulliet, S. Johann, and H. Weck-Hannemann, eds. *Sustainable Natural Hazard Management in Alpine Environments.* Berlin, Heidelberg: Springer Berlin Heidelberg, 35–67. doi: 10.1007/978-3-642-03229-5_2.

Statistik Austria, 2019. *Baupreisindex*.

Winter, B., *et al.,* 2019. A continuous modelling approach for design flood estimation on sub-daily time scale. *Hydrological Sciences Journal,* 88 (11), 1–16. doi: 10.1080/02626667.2019.1593419.

Event generation for probabilistic flood risk modelling: multi-site peak flow dependence model vs. weather generator based approach

Benjamin Winter[1,2], Klaus Schneeberger[1,2], Kristian Förster[3], and Sergiy Vorogushyn[4]

[1]Institute of Geography, University of Innsbruck, Innrain 52f, A-6020 Innsbruck, Austria
[2]alpS, Grabenweg 68, A-6020 Innsbruck, Austria
[3]Institute of Hydrology and Water Resources Management, Leibniz University Hannover, Appelstr. 9A, D-30167 Hannover, Germany
[4]GFZ German Research Centre for Geosciences, Hydrology Section, Telegrafenberg, D-14473 Potsdam, Germany

Correspondence: B. Winter (Benjamin.Winter@uibk.ac.at)

Abstract. Flood risk assessment is an important prerequisite for risk management decisions. To estimate the risk, ~~flood damages need~~ i.e. the probability of damage, flood damage needs to be either systematically recorded over long period or ~~they need~~ it needs to be modelled for a series of synthetically generated flood events. Since damage records are typically rare, time series of plausible, spatially coherent event precipitation or peak discharges need to be generated to drive the chain of process models. In the present study, synthetic flood events are generated by two different approaches to model flood risk in a meso-scale alpine study area (Vorarlberg, Austria). The first approach is based on the semi-conditional multi-variate dependence model applied to discharge series. The second approach ~~is based~~ relies on the continuous hydrological modelling of synthetic meteorological fields generated by a multi-site weather generator and using an hourly disaggregation scheme. The results of the two approaches are compared in terms of simulated spatial patterns of peak discharges and overall flood risk estimates. It could be demonstrated that both methods are valid approaches for risk assessment with specific advantages and disadvantages. Both methods are superior to the traditional assumption of a uniform return period, where risk is computed by assuming a homogeneous return period (e.g. 100-year flood) across the entire study area.

1 Introduction

In recent decades several large flood events occured across Europe resulting in direct damage exceeding one billion Euro (Kundzewicz et al., 2013). Growing flood damage due to socio-economic and land-use changes as well as a possible increase of flood hazards in a warmer climate (IPCC, 2018) calls for robust flood risk assessment. A reliable estimation of flood ~~damages~~ damage is an essential prerequisite for profound decision making (de Moel et al., 2015). The most straightforward estimation of possible flood risk would be a statistical evaluation of documented flood ~~damages~~ damage across the area of interest. In practice, systematic damage records are rare and mostly not available for longer periods (Downton and Pilke, 2005), whereas the major interest e.g. in the re-insurance industry is on losses due to extreme events ~~(European Union, 2009) that are rare by definition. Since risk assessment is currently not feasible based on empirical data, modelling approaches based on synthetic~~

~~flood scenarios are often deployed (Lamb, et al., 2010, Falter et al. 2015).~~ such as the 200 years return period to fulfill the Solvency II regulations. (European Union, 2009).

Following the European flood directive flood risk is defined as "the combination of the probability of a flood event and of the potential adverse consequences [...]" (European Union, 2007). In other words, flood risk is defined by the probability of damage. Hence, for risk estimation, a flood event including its probability of occurrence (hazard) on the one hand and the vulnerability of exposed values on the other hand need to be considered (Klijn et al., 2015). Since risk assessment is currently not feasible based on empirical data, modelling approaches based on synthetic flood scenarios are often deployed (e.g. Lamb et al., 2010; Falter et al., 2015; Schneeberger et al., 2019).

In a traditional approach, the hydrological load is estimated by means of extreme value statistics using river gauge data and transformed into corresponding inundated areas by hydrodynamic models (Teng et al., 2017). The monetary damage can then be assessed in combination with susceptibility functions which describe the relationship between one or more flood hazard characteristics (e.g. inundation depth, flow velocity) and damage for the elements at risk (Merz et al., 2010). This approach implies two strong assumptions. First, the return period of flood discharge is assumed to be equal to the return period of the resulting damage. Second, a uniform return period across the entire study area is considered and resulting damage estimates are accumulated. The first assumption can be relaxed by modelling a continuous series of synthetic flood events. As a result, a long series of damage values can be generated and used for analysing damage frequency distribution (Achleitner et al., 2016). The second assumption of homogeneous flood return periods may be valid for small areas (de Moel et al., 2015). With increasing scale, the assumption of a homogeneous return period becomes unlikely, as precipitation and flood footprints are inhomogeneous in space. This assumption can lead to an overestimation of risk for specific return periods in large river basins ~~(Thieken et al., 2015; Vorogushyn et al., 2018)~~(Thieken et al., 2015; Vorogushyn et al., 2018; Metin et al., 2020). To overcome the second limitation, realistic spatially heterogeneous events need to be generated across the area of interest which fully represent the spatial variability of flooding (Schneeberger et al., 2019).

Generation of spatially heterogeneous flood events in terms of precipitation fields or discharges is of current scientific interest ~~(Keef et al., 2013; Falter et al., 2015; Falter, 2016; de Moel et al., 2015; Speight et al., 2017; Diederen et al., 2019; Schneeberger et al., 20~~ (Keef et al., 2013; Falter et al., 2015; Falter, 2016; de Moel et al., 2015; Speight et al., 2017; Diederen et al., 2019; Diederen and Liu, 201 . There are different approaches to generate large event series of heterogeneous flood events. One possibility is the application of multivariate statistical methods to discharge series, such as copula models ~~(Jongman et al., 2014; Serinaldi and Kilsby, 2017)~~ (Jongman et al., 2014; Serinaldi and Kilsby, 2017; Brunner et al., 2019) or the semi-parametric conditional model proposed by Heffernan and Tawn (2004) (hereinafter referred to as 'HT-model'). These models consider the pairwise dependence of peak discharges at multiple locations and generate synthetic series of multiple dependent flow peaks. The second possibility is based on the generation of spatially distributed meteorological fields by a weather generator, either station-based with subsequent interpolation ~~(Falter et al., 2016; Falter, 2016; Breinl et al., 2017)~~ (Falter et al., 2016; Falter, 2016; Breinl et al., 2017; Evin et al., 2018; Ray or raster-based (Buishand and Brandsma, 2001; Peleg et al., 2017). Synthetic meteorological fields are subsequently used to drive hydrological simulations to generate streamflow values across the study area.

The two presented approaches estimate the hydrological load in the river network at multiple locations, but are different in their nature. This leads to the key question of the present study: Does it matter which approach is chosen in the context of flood risk modelling, and what are advantages and disadvantages of the two? We answer this question by comparing the set of heterogeneous flood events from the HT-model with the one resulting from a weather generator and subsequent rainfall-runoff modelling. Both methods are embedded in a probabilistic flood risk model used to estimate the effect of chosen methods on flood losses. To the authors' best knowledge, there is no study to date in which the two approaches are directly compared. Additionally, the flood risk corresponding to homogeneous flood scenarios of certain return periods ("traditional" approach) is derived and compared to the other two approaches.

This paper is organised as follows: First, the study area is shortly described. In the second section the flood risk model is introduced and the two different approaches for heterogeneous event generation are ~~described~~ presented in details. Section three presents the results of the comparison, which are discussed in the following section. Finally, conclusions summarise the major findings.

2 Study Area

The flood risk model is applied in the westernmost province of Austria, Vorarlberg. The region is characterised by a strong altitudinal gradient between the Rhine Valley ($\approx$ 400 m a.s.l.) and the high mountain ranges of the Alps ($>$ 3000 m a.s.l.). As a result of the high relief energy, the rivers are characterised by a fast hydrological response with short concentration times. The mountainous landscape of in total 2600 km^2 is dominated by forest, meadows and pastures with only small percentage of settlement area (Sauter et al., 2019). Due to steep topography, asset values are concentrated in the lowlands of larger valley floors, especially alongside the Rhine and Ill rivers. Vorarlberg is characterised by one of the highest precipitation amounts in Austria conditioned by predominantly westerly flows and strong orographic effects (BMLFUW, 2007). During the last decades, the province was affected by several severe flood events in 1999, 2002, 2005 and 2013. The most devastating recent flood event in August 2005 caused about €180 million direct tangible losses for the private and public sector, including infrastructures (Habersack and Krapesch, 2006). Figure 1 provides an overview of the study area, including the river network, settlement areas and the locations of river gauging stations as well as meteorological stations.

3 Methods and Data

The probabilistic flood risk model (PRAMo) used in the presented work consists of three different modules: The Hazard module comprising the generation of long time series of flood events, the Vulnerability module used to evaluate possible adverse consequences of flood events with a certain exceedance probability, and the Risk Assessment module which combines the results of the Hazard and Vulnerability modules to estimate the loss per event and resulting risk (Schneeberger et al., 2019). The output of the flood risk model are expected annual ~~damages~~ damage and exceedance probability curves of ~~damages~~damage. PRAMo was previously driven by the synthetic flood event series of coherent peak discharges generated by the HT-model

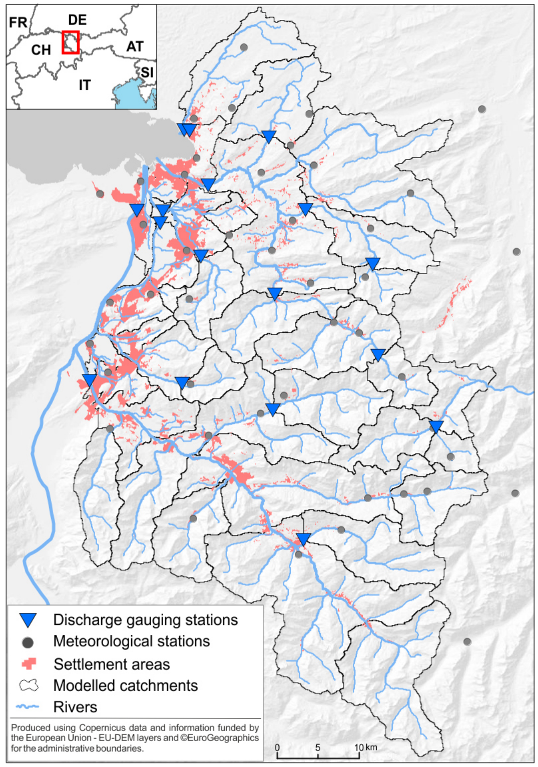

Figure 1. Study area and the location of meteorological and river gauging stations.~~(Map sources: Produced using Copernicus data and information funded by the European Union - EU-DEM layers and ©EuroGeographics for the administrative boundaries.)~~

(Schneeberger and Steinberger, 2018). A second event generation approach based on a multi-site, multi-variate weather generator and continuous rainfall-runoff modelling was recently introduced by Winter et al. (2019) and is used for comparison with the HT-model based approach and the assumption of homogeneous return periods. Figure 2 provides an overview of the modules and the simulation steps, which are described in more details in the following.

In this study, data of 17 gauging stations (1971-2013) are applied for the HT-approach. The continuous simulation of the WeGen approach is based on daily time series from 1971-2013 for 45 meteorological stations (c.f. Figure 1). At hourly time steps data for only 23 sites starting from 2001 are available. Stations without hourly information were interpolated by an inverse distance-weighting scheme (for details see Winter et al. 2019).

3.1 Hazard Module I: HT-model

The Hazard module generates time series of spatially distributed synthetic flood events. In the first approach, we apply the conditional extreme value model (HT-model) proposed by Heffernan and Tawn (2004) to peak flows. In this approach, flood events are understood as a set of spatially consistent peak discharges at multiple locations of stream gauges. Spatial consistency is ensured by considering the correlation structure of peak flows from the past observation period. Discharge time series at 17

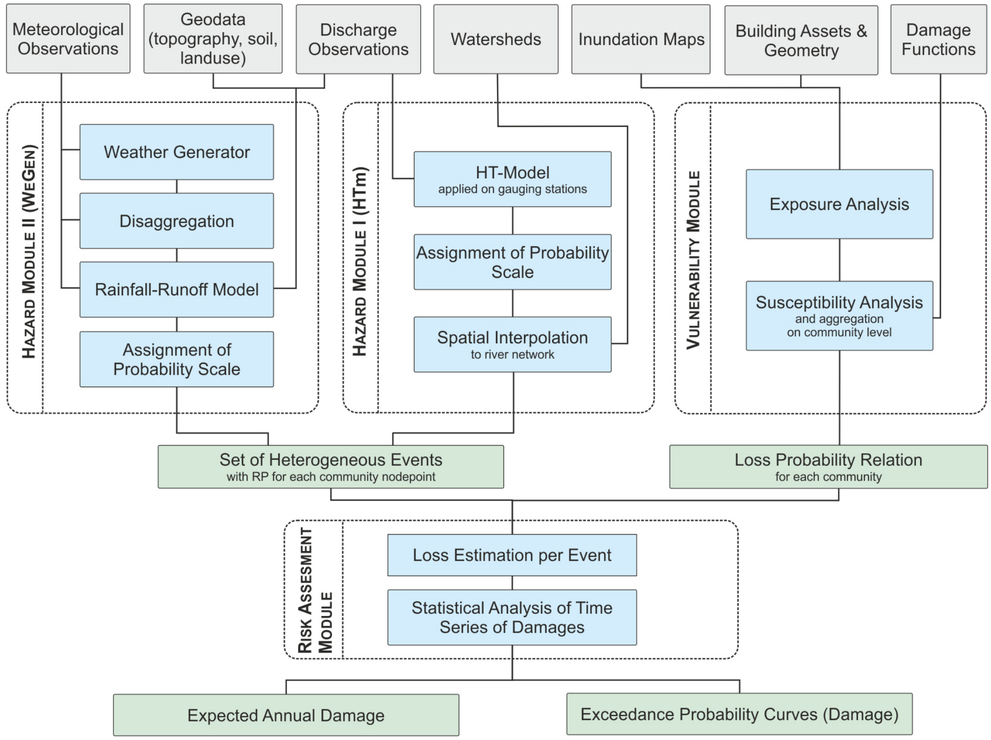

Figure 2. Flowchart of the PRAMo flood risk model including two different approaches for flood event generation.

gauges across the study area are used to parameterise the HT-model. In the first step, the observed data are standardised by a marginal model to a Laplace distribution. In the second step, the dependency between the stations is modelled for the case that peak flow at one station is above a certain threshold. According to Lamb et al. (2010), the HT-model can be interpreted as a multi-site peak-over-threshold approach. Due to strong seasonality of streamflow in Vorarlberg, the HT-model is separately parameterised for winter and summer periods (Schneeberger and Steinberger, 2018).

For the set of synthetic flood peaks at each of the 17 gauge locations we estimate the return period based on the Generalized Extreme Value (GEV). A flood event is characterised by exceedance of a certain streamflow at a single or multiple location with a defined time period. As threshold for defining a widespread flood event, a return period of 30 years was selected in the present study. The output of the HT-model in terms of synthetic flood peaks is available at the locations of gauging stations. Hence, for the river segments without observations, the flows and their respective return periods need to be estimated. We apply the top-kriging approach (Skøien et al., 2006) for the spatial interpolation of model results to the entire river network. This method takes into account the nested structure of river catchments which makes the results more robust compared to traditional regional regression based approaches (Laaha et al., 2014; Archfield et al., 2013). A more detailed description of the HT-model is provided in Schneeberger and Steinberger (2018) and Schneeberger et al. (2019).

3.2 Hazard Module II: WeGen

The second approach is based on a stochastic weather generator used to drive a hydrological model. Long-term daily precipitation and temperature series are generated, with a multi-site, multi-variate weather generator based on the auto-regressive model (Hundecha et al., 2009). Daily precipitation amounts are generated from mixed gamma and generalised Pareto distributions fitted to individual weather stations. The mixed distribution is shown to better capture extreme precipitation ~~by~~ while robustly modelling the bulk of precipitation amounts (Vrac and Naveau, 2007). In respect to seasonal patterns, the fitting is applied on a monthly base. Occurrence and amount of precipitation are modelled considering the autocorrelation and inter-site correlation structure. The mean temperature is then modelled conditioned to the simulated precipitation (Hundecha and Merz, 2012).

As the study area is characterised by mostly alpine topography with short catchment response times, the hydrological model needs to be driven by meteorological input at sub-daily resolution to estimate realistic peak flows (e.g. Dastorani et al., 2013). A non-parametric k-nearest neighbour algorithm based on the method of fragments is applied to disaggregate the generated daily values to hourly time steps (Winter et al., 2019). For a day to disaggregate, the generated daily values of temperature and precipitation from the weather generator are compared against observed daily data at all stations. Subsequently, k-nearest neighbours in terms of lowest euclidean distances between generated and observed daily values are selected. Next, one matching day is randomly sampled from the selected neighbours and the corresponding relative temporal patterns from the match day are transferred to the input day (method of fragments). In contrast to the previous study (Winter et al., 2019), a centred moving window of 30 days is applied instead of the identical months in order to restrict the search of possible matching days. The modification increases the variability between the disaggregated days and reduces the maximum search distance on a temporal scale, especially for days at the beginning and end of a month.

Following the generation of meteorological data at the locations of the weather stations, a spatial interpolation to continuous meteorological fields is necessary for the application of the rainfall-runoff model. Complex methods for spatial interpolation can be applied (e.g. Goovaerts, 2000; Plouffe et al., 2015), however, for the long term simulation a computationally efficient approach is needed. The interpolation was carried out by a inverse distance-weighting scheme including a step wise lapse rate to account for the complex topography (Bavay and Egger, 2014).

Finally, the semi-distributed conceptual rainfall-runoff model HQsim is applied to simulate streamflow across all catchments of the study area ~~(Kleindienst, 1996; Senfter et al., 2009; Achleitner et al., 2012).~~ (Kleindienst, 1996). HQsim is forced by precipitation and temperature data and was previously used in various studies in alpine catchment areas (e.g. Senfter et al., 2009; Achleitner . A simulated annealing algorithm is used for the model calibration against observed discharge data at the gauging stations (Andrieu et al., 2003). From a long synthetic discharge series, relevant flood events are identified and extracted. For this, a flood frequency analysis at all ~~point~~ points of interest based on fitting the GEV-distribution using the L-moments is carried out. Analogously to the HT-model approach, a threshold of ~~≥ 30~~ 30 year return period, at least at one site across the study area is applied to define relevant flood events. A more detailed description of the modelling chain, including the disaggregation procedure, is given in Winter et al. (2019).

3.3 Vulnerability Module

While the Hazard module computes the hydrological load, the Vulnerability module assesses the possible negative consequences in terms of exposed objects and monetary damage. The module is based on the widely used approach of combining the exposure and susceptibility of elements at risk in the inundated areas (Koivumäki et al., 2010; Merz et al., 2010; Huttenlau and Stötter, 2011; Meyer et al., 2013; Cammerer et al., 2013; de Moel et al., 2015; Falter, 2016; Wagenaar et al., 2016). The module calculates losses for each community in the study area for a number of predefined return periods (or probabilities) (i.e. RP = 30, 50, 100, 200 and 300 years). The results of the Vulnerability module are loss-probability relations for each community, describing the expected damage for the corresponding return periods. To derive a continuous relation, a linear interpolation between available data points (RP-damage) is applied. The loss-probability relations are used as input in the Risk Assessment Module and combined with the simulated return periods (Hazard Module) at each community to derive risk curves.

At the scale of a community (in average 28 km^2), a homogeneous return period of hydrological load is assumed and associated with the total community loss. For the loss calculation we use "official" inundation maps. The inundation maps are based on 1D hydrodynamic modelling in rural areas and 2D modelling in urban areas (IAWG, 2010)~~, whereas the underlying hydrological boundary conditions are based on the considerations of the~~. The boundary conditions for the hydrodynamic simulation are taken from the Austrian flood risk zoning project HORA (Merz et al., 2008). ~~To continuously describe the loss-probability relation for each community, a linear interpolated interpolation between available data points (damages) is applied.~~

The estimation of monetary damage for the elements at risk is based on the relative damage functions combined with the total asset values. A damage function describes the relative loss of value as a function of water depth (Merz et al., 2010). If available, additional damage influencing parameters, such as flow velocity or contamination can be considered for damage assessment (Merz et al., 2013). In ~~the present study, the~~ accordance to Schneeberger et al. (2019), the one parametric damage model ~~according to~~ of Borter (1999) is applied in the present study. The damage model was derived for Switzerland, which is a direct neighbour to the Austrian province Vorarlberg with a similar topography and building structure. More precise, site-specific damage functions are not available for the study region.

The damage estimation is conducted on a single object basis for residential buildings only. To derive the flood losses, the available inundation maps are combined with the asset datasets and damage function. Subsequently, the object based loss data are aggregated for each community. The absolute building values indexed to 2013 according to the construction price index (Statistik Austria, 2019) are derived by calculating mean cubature values from local insurance data, and transferred to the entire building stock of the study area (Huttenlau et al., 2015). Since derived values are based on insurance data, they are consequently defined as replacement values.

3.4 Risk Assessment Module

The risk assessment module brings together the results of the hazard and vulnerability modules to generate a time series of losses and calculates the resulting risk curve for the area of interest (Schneeberger et al., 2019). In order to combine the results, each spatial unit (community) is represented by a defined model node point at the river network. For each generated heterogeneous flood scenario, the recurrence intervals are derived for all model node points (hazard module) and combined with the respective loss-probability relation to compute losses (vulnerability module). By integrating the losses at all model node points, i.e. for each community, the total loss for every generated event can be ~~evaluated and~~ calculated. By evaluating the overall modelled time series of events a continuous time series of ~~losses~~ damage is generated. Finally, the time series of damage can be ~~statistical~~ statistically analysed to derive the expected annual damage (EAD) and to construct risk curves (Schneeberger et al., 2019). More detailed information about the Vulnerability and Risk Assessment Module, including a schematic overview of the module interaction is provided in Schneeberger et al. (2019).

3.5 Assessment of spatial coherence of generated events

A core element of the probabilistic flood risk model is the generation of plausible, spatially heterogeneous flood events. To investigate the spatial coherence of synthetic events generated by two different approaches, two spatial dependence measures proposed by Keef et al. (2009) are applied. The first measure $P_{i,j}(p)$ describes the probability that a dependent site i exceeds a certain threshold~~given,~~ , given that a conditional site j is exceeding a threshold $q_p(Q_j)$ as well:

$$P_{i,j}(p) = Pr(Q_i > q_p(Q_i)|Q_j > q_p(Q_j)), \tag{1}$$

where (p) is the level of extremeness (quantile) and Q_i and Q_j are the dependent and conditioned runoff series, respectively. The calculation of the thresholds are based on a three day block maxima, which was found to be appropriate in this region (Schneeberger and Steinberger, 2018). The second spatial dependence measure $N_j(p)$, is an overall summary metric and describes the average probability of all dependent sites i to be high, given that the conditional site j is high, defined as:

$$N_j(p) = \frac{\sum_{i \neq j} Pr(Q_i > q_p(Q_i)|Q_j > q_p(Q_j))}{n-1} \tag{2}$$

In case of the WeGen approach the dependence ~~metrices~~ matrices were computed for the peak discharges at the gauging station locations resulting from the combined simulations of the weather generator and rainfall-runoff model.

4 Results

4.1 Simulation Results of the Continuous Modelling Approach (WeGen)

To assess the performance of the continuous modelling approach extreme precipitation of simulated data are compared to observed station data (daily: 1971–2013; hourly 2001–2013) for the weather generator and disaggregation procedure. The median and the uncertainty range represented by the 5% and 95% quantiles of 100 model realizations are compared to the observed data. Figure 3a shows the results for 99% quantile of daily precipitation (wet days) for all 45 station and spring (MAR-APR-MAY), summer (JUN-JUL-AUG) and autumn (SEP-OCT-NOV). In general, the characteristics of the observed daily precipitation is well reproduced by the weather generator. A few stations, however, show a slight underestimation in summer (mainly June and August). The validation results for all months separately including maximum and minimum simulated daily temperatures are provided by Winter et al. (2019). To validate the disaggregation procedure, the hourly data are first aggregated to daily data and subsequently disaggregated back to hourly time steps. For the comparison of disaggregated precipitation, 99%, 99.9% and 99.95% quantiles are calculated and compared to the observed values. The results for the 99.9% quantile show a good agreement between observed and simulated precipitation intensities for the three analysed rainfall durations 1, 3 and 6 hours (Figure 3b). Results for the 99% and 99.95% quantile are shown in Winter et al. (2019).

The rainfall-runoff model is calibrated (2001-2007) and validated (2008-2013) in classical split-sample approach (Klemeš, 1986) for all catchments of the study area against observed river gauging data. On average, a Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1 of 0.68 and 0.67 and a Kling-Gupta efficiency (KGE; Kling et al., 2012) of 0.75 and 0.74 are achieved for the calibration and validation periods, respectively. Detailed results for the individual catchments, including a comparison of design flood estimates with a flood frequency analysis and a design storm approach are given in Winter et al. (2019).

4.2 Spatial Patterns of Generated Flood Events

For the analysis of spatial coherence, 100 simulations using each of the two event generation approaches (HT-model and WeGen) were carried out. Each ~~simulations~~ simulation comprised 42 years of data corresponding to the length of the observed discharge series. Figure 4 illustrates exemplary results for four gauging stations and both methods. Each plot shows the dependence measure between the two stations depicted on the maps in the principal diagonal. The gauge Kennelbach and Gisingen are the two largest catchments of the study area (about 80% of the total area). The examples Schruns and Thal are subcatchments of Gisingen and Kennelbach, respectively and thus represent two strongly related gauge pairs. The measure is calculated for discharge values with exceedance probability between $p = 0.99$ and $p = 0.997$, above which the data are ~~to~~ too few $(n < 15)$ to calculate a meaningful $P_{i,j}$ value. Based on the empirical distribution function of the 3-day block maxima series, a p-value of 0.99 refers to a return period of approximately 1 year and a p-value of 0.997 refers to a return period of roughly 3 years.

In general, the spatial dependence declines with the level of extremeness. For more extreme runoff situations, the dependence structure is less stable and ~~represented by~~ prone to a large variability. The HT-model results in the lower triangle ~~,~~ reproduce the observed spatial patterns between the stations well. The observed measure is in $\approx 90\%$ of the cases inside the simulated data range (~~2.5 — 97.5~~ $2.5-97.5\%$ quantile). The results of the WeGen approach ~~follows~~ follow the general observed patterns of lower dependence (e.g. $P_{i,j}(p) \approx 0.2$ for Thal (2) vs. Schruns (4)) and higher dependence (e.g. $P_{i,j}(p) \approx 0.5$ Kennelbach (1) vs. Thal (2))~~, however, ~~. However, the results are biased towards a higher dependence, such that only half of the results correspond well to the observed data.

To analyse the dependence structure of high flows across the study area, the ~~summary~~ measure $N_j(p)$ ~~was~~ is calculated for all node points corresponding to different communities. The measure ~~was~~ is calculated for p values corresponding to the 1-year, 10-year and 100-year return period. ~~The~~ As the simulation of the two approaches are not limited to the length of the observed data, the results are based on the median of 30 ~~repetitions~~ realizations of 1000 years of HT-model and WeGen simulations

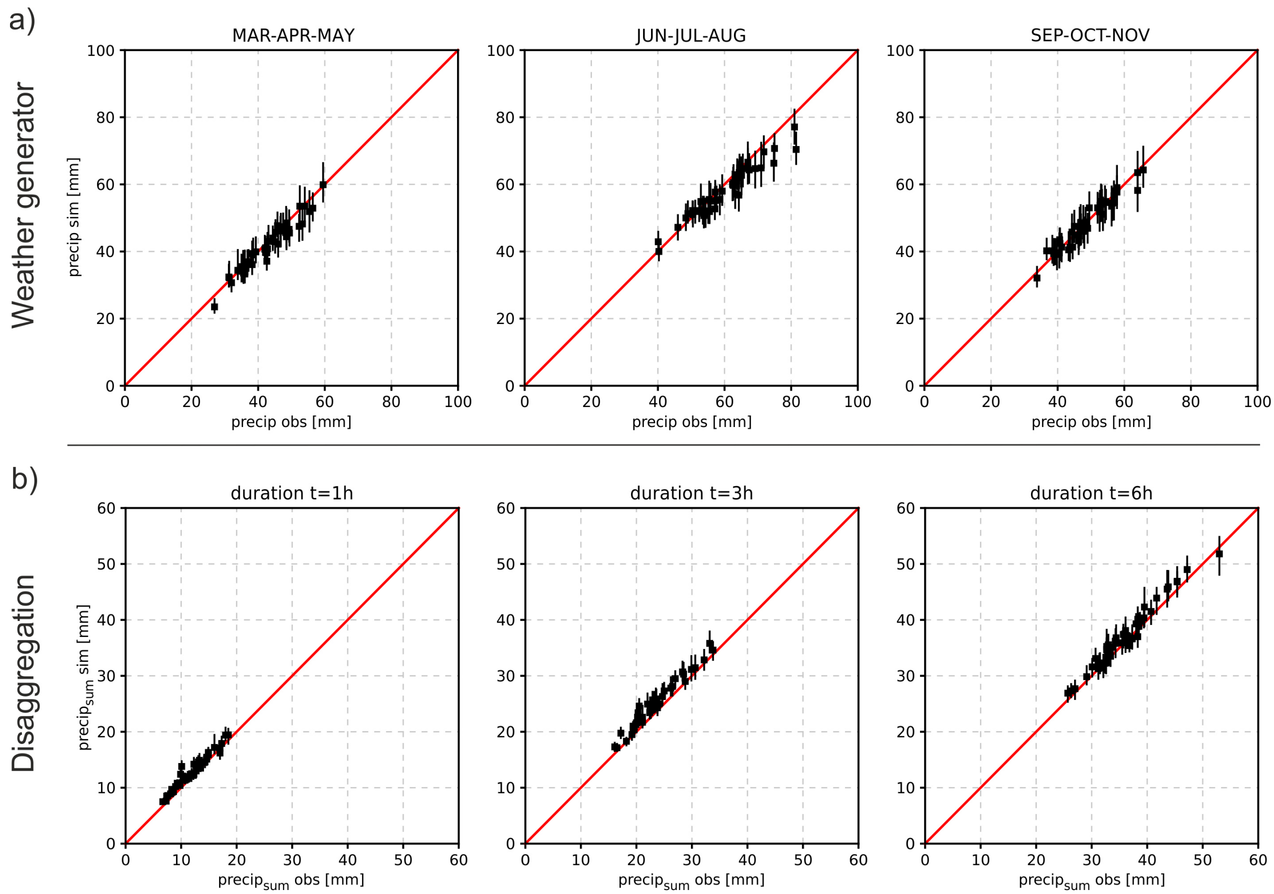

Figure 3. Validation results of the weather generator and the revised disaggregation procedure for all stations ($n = 45$). The bars represent the median and the 5 to 95% quantile range of 100 realizations for the weather generator and disaggregation. a) Weather generator: 99% quantile of daily precipitation for generated data compared with observed data for spring, summer and autumn. b) Disaggregation: 99.9% quantile of 13 years of disaggregated data is compared to observed data, for the precipitation sum of 1, 3 and 6 h duration.

(Figure 5). The ~~decline of spatial dependence towards higher return periods is shown for both~~ length is chosen to be far above the highest return period of available homogeneous inundation data (RP300) and the number of 30 realizations is dictated by the computational limitations of the continuous simulation on an hourly time step. Both approaches (see Figure 5 a-c) show a decline of spatial dependence towards higher return periods. The general patterns of lower spatial dependence in the southern part of the study area and of the individual northern catchments is visible. The node points downstream are characterised by a higher dependence. For high return period of 100 years (Figure 5 c), the simulated spatial dependence is higher for the HT-model than for the WeGen results in contrast to the findings for the lower return periods. The results are regionally different. Whereas the dependence measure is higher for the HT-model in the western part of the study area, the north-eastern catchments show higher degree of dependence for the WeGen approach.

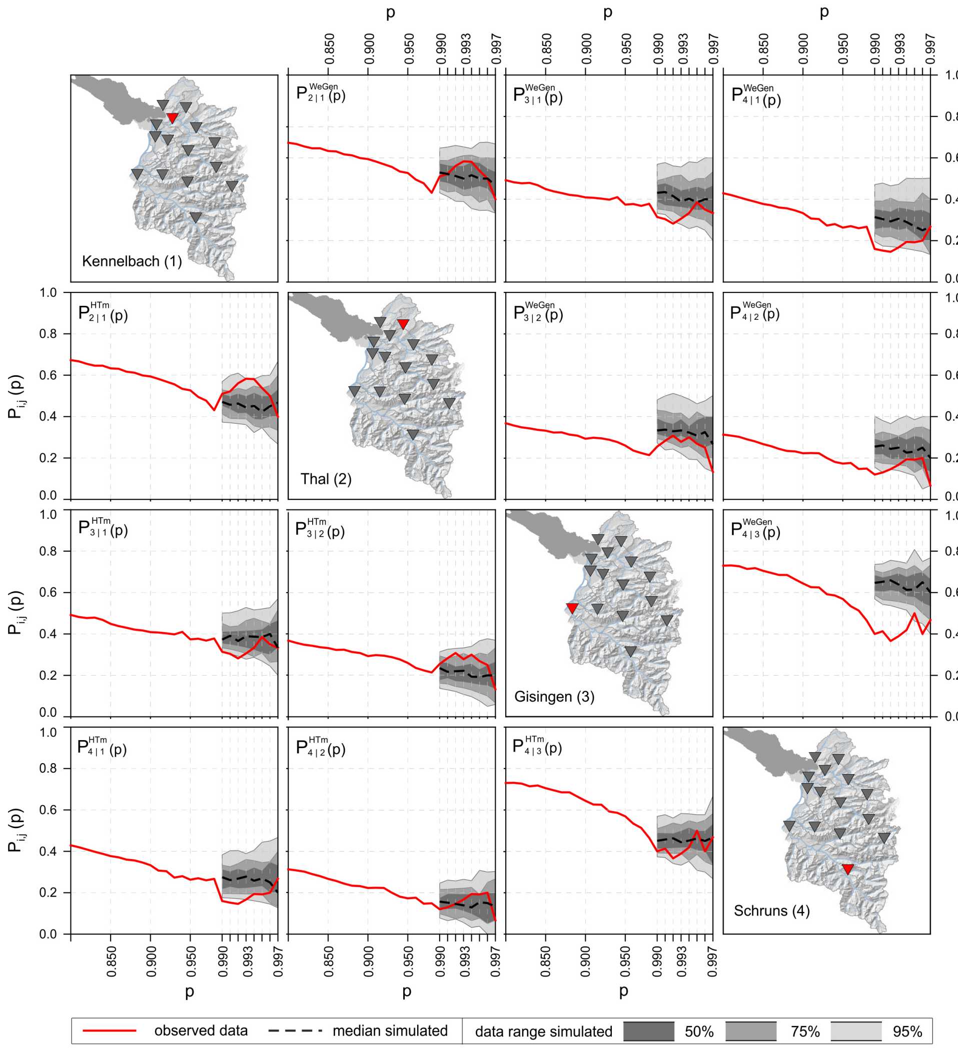

Figure 4. Comparison of observed (42 years) and simulated conditioned exceedance probability $P_{i,j}(p)$. The range of the simulated results is based on 42 years of simulation with 100 ~~repetitions~~realizations. The plots in the lower triangle correspond to the HT model, whereas those in the upper triangle show the WeGen results.

4.3 Comparison of risk curves

To compare the effect of the two approaches of synthetic event generation on the overall estimated loss, flood risk curves are calculated. Confidence intervals are derived based on 30 ~~repetitions~~ realizations of 1000-year simulations. Furthermore, the risk curve based on the assumption of homogeneous return period floods across all catchments is derived based on 5 inundation maps corresponding to the return periods between 30 and 300 years. The two synthetic event generators result in a comparable range of overall estimated flood risk (see Figure 6). The WeGen approach systematically overestimates the risk computed by

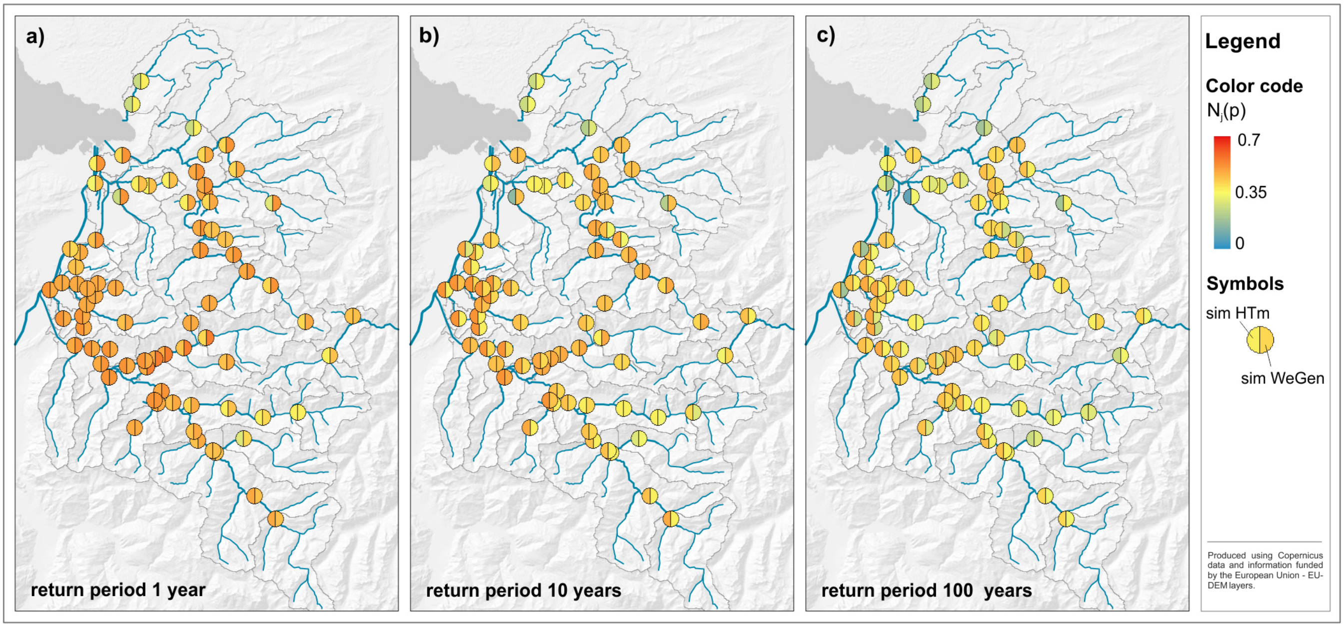

Figure 5. Spatial dependence measure $N_j(p)$ for the community node points at the river network and three different ~~levels of severity~~return periods. The results show the median for the HT-model and WeGen approach based on 30 ~~repetitions~~ realizations of 1000 years simulation.~~(Map sources: Produced using Copernicus data and information funded by the European Union - EU-DEM layers.)~~

the HT-model. The relative difference between the estimated median values ($\cancel{\frac{WeGen-HT-model}{WeGen}}(WeGen-HTm)/WeGen$) is approximately 17.5%. The uncertainty increases with increasing return period of damage alongside the extrapolation of the input time series. On average 172 damage events are generated per 1000 years of simulation in the WeGen approach compared to about 167 for the HT-model. Both approaches show a significant lower ~~damages~~ damage in comparison to the assumption of ~~the homogeneous return~~homogeneous scenarios for specific return periods. The estimated damage of a homogeneous 100-year flood scenario is ≈50% above the HT-model results and still 40% above the WeGen approach.

The sets of generated heterogeneous flood events reflect a large variability of plausible spatial patterns. Hence the estimated flood risk is the result of a combination of these patterns. Figure 7 shows multiple examples of generated flood events corresponding to an estimated damage of 100±1 million Euro for both model approaches. The general severity in terms of flood hazard (without consideration of flood risk) is given by the Unit of Flood Hazard (UoFH). The measure UoFH is a simple proxy of hazard severity defined as the total number of sites at which the threshold of 30 years return period is exceeded (Schneeberger et al., 2019). Even though, the selected severity of displayed flood event is rather high, some of the generated events are still spatially limited. The event with the lowest UoFH of 46 corresponds to ≈ 50% of all sites exceeding the 30-year threshold. The most widespread event ~~corresponds to ≈ 90% of exceeding sites (UoFH~~(UoFH=~~77).~~ 77) corresponds to about 90% of the sites exceeding the threshold. This result reflects the spatial distributions of elements at risk with a settlement concentration alongside the larger valley areas in the study area (c.f. Figure 1). Thus, the damage corresponding to an event is

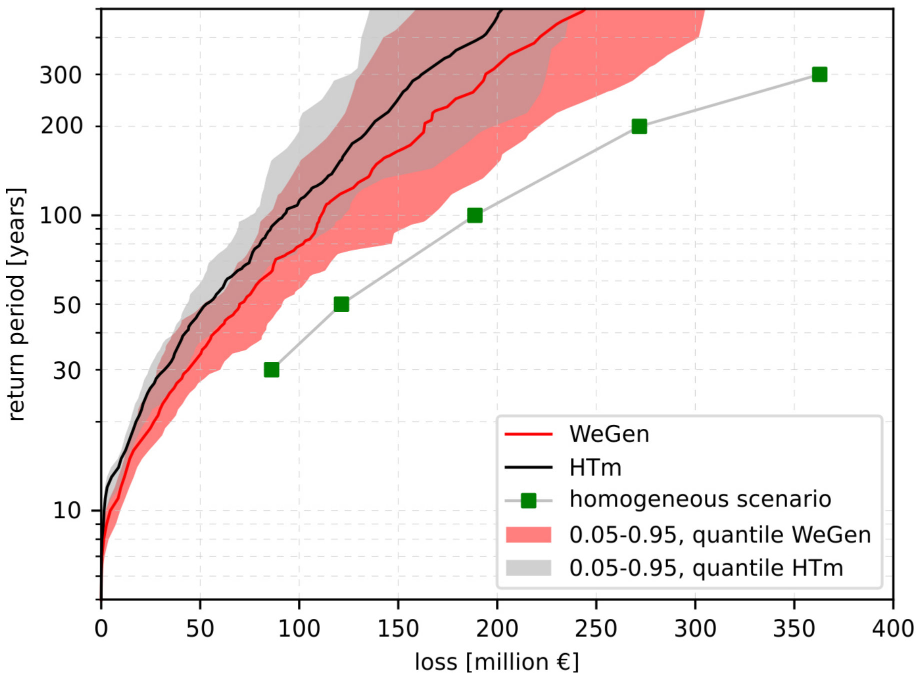

Figure 6. Risk curves for WeGen and HT-model approach in comparison to the results of a homogeneous scenario. The median and quantile confidence intervals are based on 30 ~~repetitions~~ realizations of 1000 years of simulation. Monetary values are normalized to the year 2013.

largely influenced by the region affected. If the overall comparison is conducted on hazard level only, the impact of widespread flood events may be overestimated, while the impact of spatially limited events in densely populated areas are underestimated.

5 Discussion

Both approaches, the HT-model and the WeGen approach, simulate ~~a~~ complex spatially heterogeneous patterns of synthetic flood events. In the present study, the HT-model outperforms the WeGen approach in terms of reproducing the observed dependence patterns of peak flows at the gauging stations. The HT-model makes use of the observed river gauging data and models their dependence structure directly. On the contrary, the WeGen approach models the dependence structure only ~~indirect~~ indirectly based on the meteorological input data.

The overall river network and especially small ~~tributaries with no explicitly river gauging stations~~ ungauged tributaries do however rely on the top-kriging interpolation in case of the HT-model approach and are not able to react independently to the larger river system. This explains the higher dependence structure on the community node points, while at the river gauges the results do correspond well to the observed values. Nevertheless, in both cases the ~~feature~~ capability to capture spatial effects of a certain spatial scale ~~is~~ in the end ~~depending~~ depends on the density of the measuring network and its data quality.

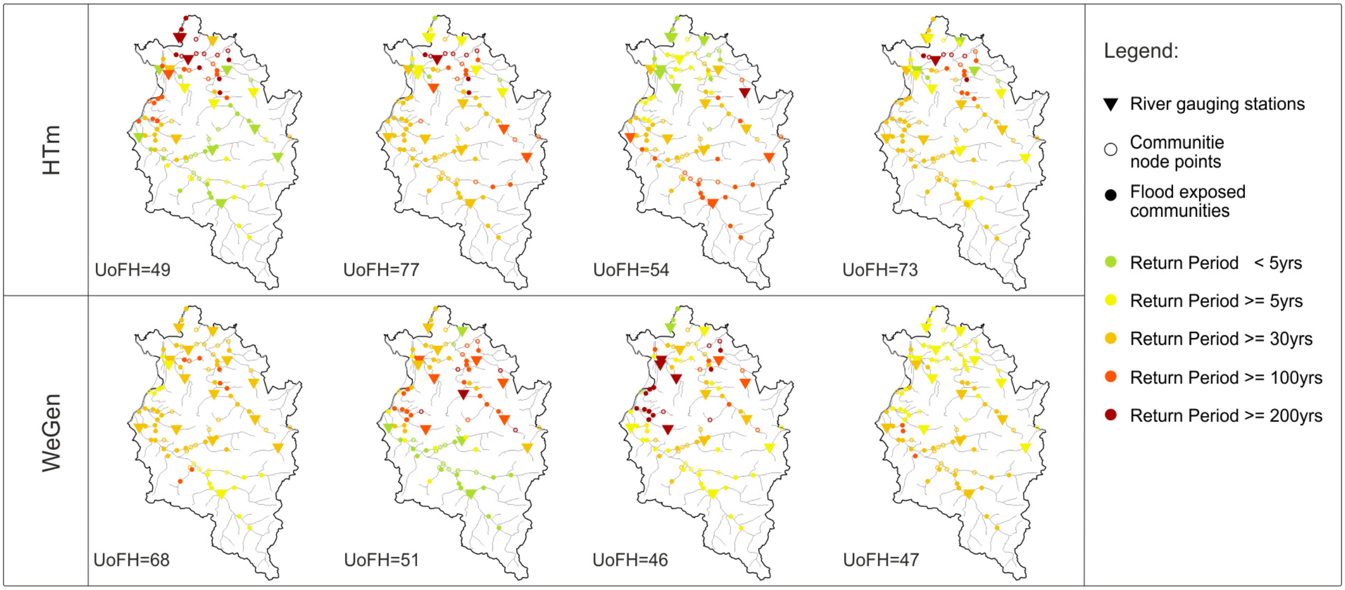

Figure 7. Examples of flood events with an estimated flood damage of 100±1 million Euro flood damage for HT-model and WeGen approach. The general severity of flood events is characterised by the Unit of Flood Hazard (UoFH).

The WeGen approach seems to overestimate the overall spatial dependence in the study area in comparison to the observed values. This was also found in a previous study, comparing a different set of gauging stations (Winter et al., 2019). One possible reason could be, the spatial interpolation of the meteorological data by the rather simple IDW-approach, without consideration of shading effects etc. and the rather short length of hourly input data for the disaggregation procedure might ~~effect~~ affect the spatial patterns towards a stronger dependence. More importantly, the WeGen model itself seems to overestimate the dependence between stations particularly for higher return period thresholds. This is in line with the results of the recent evaluation of the weather generator (Ullrich et al., 2019), which suggest an overestimation of correlation of extreme precipitation between individual stations leading to an ~~overall estimation~~ overestimation of areal rainfall. The correlation structure of the weather generator is fitted on a monthly base, independently of the rainfall intensities and thus does mix low intensity large scale rainfalls and ~~rather regional~~ small scale convective events. The simulated stronger spatial dependence in certain ~~parts~~ areas with high damage potential also contributes to the higher flood risk estimate ~~of~~ by the WeGen approach.

Only one possible combination of weather generator, disaggregation procedure and rainfall runoff model was applied for the WeGen approach. Thus, by the application of an alternative weather generator with different assumptions about the spatial dependence or tail distribution, the resulting risk estimates may change. This counts as well for the application of an different rainfall runoff model or alternative disaggregation procedure (e.g. Müller-Thomy et al., 2018). Thus, the result of a higher risk estimate for the WeGen approach in comparison to the HT approach can not be generalized to other model combinations.

Both approaches for synthetic event generation differ substantially in terms of estimated damage from the one assuming a uniform return period across the whole study area (Figure 6). The flood losses for individual return periods above the 30-year threshold under the homogeneous assumption are largely overestimated. This result confirms the necessity to take heterogeneous spatial patterns into account. An event where every community in the study area is affected by discharges exceeding the 30-year return period during a single event is rare. Based on total of 30000 years of simulation less then 10% of the communities experience losses simultaneously in more than 50% of events (Figure 8). It can be expected that with increasing spatial scale, the likelihood that a large number of communities will experience high return period discharges and losses in a single event will decrease (Metin et al., 2020). Therefore, generation of spatially consistent heterogeneous flood events is particularly important with increasing spatial scale. ~~On the contrary~~At the same time, considering dependence of meteorological and hydrological variables at multiple locations with increasing scale and increasing number of dependent locations becomes more challenging.

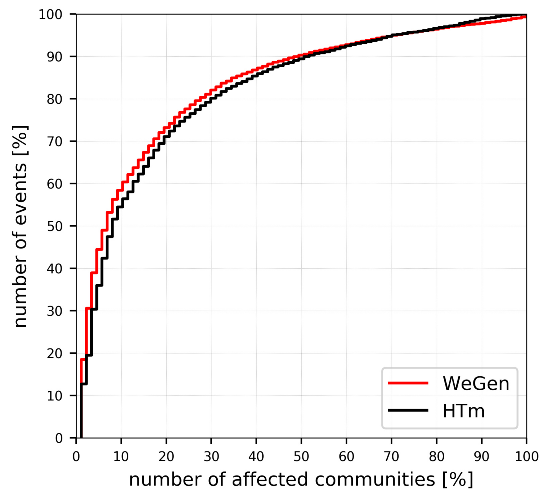

Figure 8. Relative number of flood events exceeding a 30-year flood threshold and corresponding relative number of affected communities. The results are based on 30000 years of simulation.

A fundamental difference between the two approaches resides in the way of considering the hydrological processes. The HT-model takes a purely statistical approach by analysing the dependence of peak discharges above a certain threshold. It does not explicitly consider hydrological processes which generate extremes. For instance, the non-linearity of catchment response is not explicitly taken into account, but only so far it is imprinted in the previously observed peaks used for model parameterisation. The combination of the weather generator and rainfall-runoff modelling describes the hydrological processes in a spatially consistent and time continuous way. Hence, the effect of soil moisture accumulation and pre-event catchment conditions are explicitly modelled. ~~Furthermore, a~~ By the application of a fully distributed, physically based model the hydrological process description could even be improved for example by solving full energy balance equations for snow melt or evapotranspiration (e.g. Förster et al., 2014, 2018). On the downside, a further increase in model complexity might compromise the model parameter identifiability, increase calibration effort and computational burden and increase input data demand (temperature, precipitation, radiation, humidity and wind speed).

In general, continuous hydrological modelling generates full hydrographs at all locations that allows for direct coupling with hydraulic models as for example applied in Falter et al. (2015) and Falter (2016). The direct coupling of the WeGen approach with a 1D-2D hydrodynamic model would also allow to consider hydrodynamic interactions in the river network and their possible affect on the risk estimates. This may for example be the reduction of risk downstream due to dike overtopping and failure upstream. In case of the HT-model, only peak discharge of events is estimated, not the entire hydrograph. Hence, these results cannot be used directly as a boundary condition for unsteady hydraulic simulations. Assumptions on the shape of a hydrograph would be required.

In addition, the continuous modelling approach is capable to explicitly model scenarios of changing hydrological boundary conditions. For instance, changes in the climate system can be taken into account in the generation of meteorological fields by conditioning the rainfall and temperature probability distributions (e.g. Hundecha and Merz, 2012). Also possible changes in land use can be considered by parameterising hydrological models accordingly (Rogger et al., 2017). As the HT-model approach is based on observed streamflow only, change scenarios may be included in terms of trends. However, they cannot be modelled explicitly. A continuous simulation approach requires a vast amount of processed data including multiple data interfaces between the different modelling steps and results is high computational costs. This is especially true if sub-daily simulations are applied that require an additional disaggregation scheme. In contrast, the purely statistical HT-model convinces with its efficient data processing, easily applicable on local computers. A further advantage of the HT-model is the transferability of the approach. While each of the modelling steps of the continuous approach, from weather generator to the hydrological models needs to be implemented, calibrated and validated for every new study area, the HT-model only needs to be fitted to new discharge time series which is less complex. Different advantages and disadvantages of both approaches are finally summarised in Table 1.

Both presented approaches are subject to different uncertainties. The confidence intervals presented in Figure 6 are for example based on the random processes generating heterogeneous flood events of each method (~~repetition ensemble~~multiple realizations). However, there are other uncertainties which are not explicitly addressed, as for example uncertainties related to the topological kriging of the HT-model results or uncertainties related to the hydrological model in the WeGen approach. Some uncertainties pertain to both methods such as the choice and fitting of the extreme value distributions. A comprehensive assessment ~~of these uncertainties~~ by propagating the uncertainties of all sub-models throughout the model chain is currently precluded by computational constraints particularly relevant for the WeGen approach. ~~However, single uncertainty sources can be evaluated in multiple simulations and visualised in the form of uncertainty bounds (Figure 6)~~

A further important point, currently not considered in both approaches are dike failure scenarios. In the study area, for example no inundation is considered for the River Rhine due to its high protection level. Nonetheless, the probability of a dike failure is non-zero and could have a devastating effect. In this sense, the consideration of flood volumes beside peak estimates could be another important extension to describe the severity of flood events (e.g. Dung et al., 2015; Lamb et al., 2016).

Table 1. Summary of advantages and disadvantages of the WeGen and HT-model approach to generate heterogeneous flood events.

Categories		HT-model		WeGen
Computational complexity	(+)	low processing costs (local processing)	(-)	processing intensive (HPC necessary)
			(-)	complex data interfaces between different models
Output	(-)	Return periods at all sites for modelled events only	(+)	continuous hydrographs at all modelled sites
Hydraulic coupling	(-)	event hydrographs need to be deducted to drive a hydraulic model	(+)	continuous description of hydraulic boundary conditions allows unsteady hydraulic modelling
Processes	(-)	No information about individual hydrological processes	(+)	continuous description of hydrological system and modelled processes
Hydrological changes	(-)	no explicit modelling of scenarios (e.g. climate or land use scenarios) possible	(+)	scenarios can be modelled explicitly (e.g. climate or land use scenarios)
	(+)	runoff trends can be integrated		
Transferability	(+)	model is well transferable to other study areas	(-)	model chain is transferable, however all components must be setup and calibrated for new study areas

A traditional validation of the overall risk model in terms of a comparison of observed to simulated data is hardly possible as comprehensive databases of loss events are often not available (Thieken et al., 2015). In the present study, damage data based on a insurance portfolio were available for the 2005 event. The data are, however, only a subset of the overall elements at risk and due to rather low sublimits (maximum insurance payout), the full losses remain unknown. Finally, without a larger set of loss events it is not possible to assign a meaningful return period to the 2005 event to validate the risk outcome in a traditional way. Nonetheless, by applying and comparing different methods, the plausibility of the results can be checked (Molinari et al., 2019). Furthermore, the uncertainties related to the choice of methods to generate heterogeneous flood events seem to be lower in comparison to other aspects of the probabilistic flood risk model, such as the choice of the applied damage functions (Winter et al., 2018).

6 Conclusions

The question whether the choice of method to generate heterogeneous flood events for flood risk modelling matters can be answered in different ways. Both approaches, the HT-model and continuous WeGen approach, were generally capable of modelling spatially plausible flood events across the study area. By direct comparison to observed spatial patterns, the HT-model approach performed better than the WeGen approach in our study area in terms of correctly representing the observed dependence structure. A stronger modelled dependence of extreme precipitation resulted in ~~higher~~ high areal rainfall in the WeGen approach and higher overall risk compared to the HT-model. The median damage from 30000 years of simulation is about 17.5% larger in the WeGen approach than in the HT-model. The representation of the dependence structure for simulation of extremes needs to be further improved for the weather generator. Nevertheless, the choice of method to generate heterogeneous flood events might have smaller impact than, for example, the choice of the applied damage functions (Winter et al., 2018).

To conclude, both methods are valid approaches to overcome the simplified assumption of uniform return period across a study area. Accordingly, when designing a flood risk study, the choice of the approach should consider the specific advantages and disadvantages of the two methods and data availability. If computational efficiency and quick transferability are in focus, the HT-model approach might be a better choice. In contrast, if unsteady hydraulic modelling is required for the ~~underlying~~ targeted application, the continuous modelling of generated meteorological fields is more appropriate.

Code and data availability. For Austria, daily meteorological and river gauging data are available at https://ehyd.gv.at. The applied meteorological data for the DWD stations are freely available at https://opendata.dwd.de. Underlying loss data are not publicly available. The MeteoIO is available at https://models.slf.ch/p/meteoio/. The applied weather generator and RR-model are currently not publicly available.

Author contributions. Based on the initial ideas of KS and SV, the study was designed in collaboration of all authors. BW prepared the initial data, implemented and applied the continuous modelling approach and analysed the results. KF programmed the spatial interpolation scheme for the meteorological data and supported the rainfall-runoff modelling. The risk model and the HT-application were mainly developed by KS. The manuscript was drafted by BW with support of SV. All authors contributed to the review and final version of the manuscript.

Competing interests. The authors declare that there is no conflict of interest.

Acknowledgements. This work results from the research project HiFlow-CMA [KR15AC8K12522] funded by the Austrian Climate and Energy Fund (ACRP 8th call). We would like to thank all the institutions that provided data, the Zentralanstalt für Meteorologie and Geodynamik (ZAMG), the Deutscher Wetterdienst (DWD), and particularly the Hydrographischer Dienst Vorarlberg. The simulations were conducted using the Vienna Scientific Cluster (VSC). Finally, we want to thank the editors and reviewers for taking their time to critically evaluate this article.

References

Achleitner, S., Schöber, J., Rinderer, M., Leonhardt, G., Schöberl, F., Kirnbauer, R., and Schönlaub, H.: Analyzing the operational performance of the hydrological models in an alpine flood forecasting system, Journal of Hydrology, 412-413, 90–100, https://doi.org/10.1016/j.jhydrol.2011.07.047, 2012.

Achleitner, S., Huttenlau, M., Winter, B., Reiss, J., Plörer, M., and Hofer, M.: Temporal development of flood risk considering settlement dynamics and local flood protection measures on catchment scale: An Austrian case study, International Journal of River Basin Management, pp. 1–17, https://doi.org/10.1080/15715124.2016.1167061, 2016.

Andrieu, C., Freitas, N., Doucet, A., and Jordan, M.: An Introduction to MCMC for Machine Learning, Machine Learning, 50, 5–43, https://doi.org/10.1023/A:1020281327116, 2003.

Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., and Kiang, J. E.: Topological and canonical kriging for design flood prediction in ungauged catchments: An improvement over a traditional regional regression approach?, Hydrology and Earth System Sciences, 17, 1575–1588, https://doi.org/10.5194/hess-17-1575-2013, 2013.

Bavay, M. and Egger, T.: MeteoIO 2.4.2: a preprocessing library for meteorological data, Geoscientific Model Development, 7, 3135–3151, https://doi.org/10.5194/gmd-7-3135-2014, 2014.

Bellinger, J.: Uncertainty Analysis of a hydrological model within the Flood Forecasting of the Tyrolean River Inn: Dissertation, 2015.

BMLFUW, ed.: Hydrologischer Atlas Österreichs, Lieferung, 2007.

Borter, P.: Risikoanalyse bei gravitativen Naturgefahren: Fallbeispiele und Daten, Naturgefahren, Bern, 1999.

Breinl, K., Strasser, U., Bates, P., and Kienberger, S.: A joint modelling framework for daily extremes of river discharge and precipitation in urban areas, Journal of Flood Risk Management, 10, 97–114, https://doi.org/10.1111/jfr3.12150, 2017.

Brunner, M. I., Furrer, R., and Favre, A.-C.: Modeling the spatial dependence of floods using the Fisher copula, Hydrology and Earth System Sciences, 23, 107–124, https://doi.org/10.5194/hess-23-107-2019, 2019.

Buishand, T. A. and Brandsma, T.: Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest-neighbor resampling, Water Resources Research, 11, 2761–2776, https://doi.org/10.1029/2001WR000291, 2001.

Cammerer, H., Thieken, A. H., and Lammel, J.: Adaptability and transferability of flood loss functions in residential areas, Natural Hazards and Earth System Science, 13, 3063–3081, https://doi.org/10.5194/nhess-13-3063-2013, 2013.

Dastorani, M., Koochi, J., and Darani, H. S.: River instantaneous peak flow estimation using daily flow data and machine-learning-based models, Journal of Hydroinformatics, pp. 1089–1098, https://doi.org/10.2166/hydro.2013.245, 2013.

de Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E., and Ward, P. J.: Flood risk assessments at different spatial scales, Mitigation and Adaptation Strategies for Global Change, https://doi.org/10.1007/s11027-015-9654-z, 2015.

Diederen, D. and Liu, Y.: Dynamic spatio-temporal generation of large-scale synthetic gridded precipitation: with improved spatial coherence of extremes, Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-019-01724-9, 2019.

Diederen, D., Liu, Y., Gouldby, B., Diermanse, F., and Vorogushyn, S.: Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment, Natural Hazards and Earth System Sciences, 19, 1041–1053, https://doi.org/10.5194/nhess-19-1041-2019, 2019.

Dobler, C. and Pappenberger, F.: Global sensitivity analyses for a complex hydrological model applied in an Alpine watershed, Hydrological Processes, 27, 3922–3940, https://doi.org/10.1002/hyp.9520, 2013.

Downton, M. W. and Pilke, R. A.: How Accurate are Disaster Loss Data? The Case of U.S. Flood Damage, Natural Hazards, 35, 211–228, https://doi.org/10.1007/s11069-004-4808-4, 2005.

Dung, N. V., Merz, B., Bárdossy, A., and Apel, H.: Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta, Journal of Hydrology, 527, 704–717, https://doi.org/10.1016/j.jhydrol.2015.05.033, 2015.

European Union: European Union on the assessment and management of flood risks: Directive 2007/60/EC of the European Parliament and the Council, Official Journal of the European Community, 2007.

European Union: on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II): Directive 2009/138/EC of the European Parliament and the Council., Official Journal of the European Union, 2009.

Evin, G., Favre, A.-C., and Hingray, B.: Stochastic generation of multi-site daily precipitation focusing on extreme events, Hydrology and Earth System Sciences, 22, 655–672, https://doi.org/10.5194/hess-22-655-2018, 2018.

Falter, D.: A novel approach for large-scale flood risk assessments : continuous and long-term simulation of the full flood risk chain: Dissertation, Potsdam, 2016.

Falter, D., Schröter, K., Dung, N. V., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H., and Merz, B.: Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain, Journal of Hydrology, 524, 182–193, https://doi.org/10.1016/j.jhydrol.2015.02.021, 2015.

Falter, D., Dung, N. V., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F., and Merz, B.: Continuous, large-scale simulation model for flood risk assessments: proof-of-concept, Journal of Flood Risk Management, 9, 3–21, https://doi.org/10.1111/jfr3.12105, 2016.

Förster, K., Meon, G., Marke, T., and Strasser, U.: Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany), Hydrology and Earth System Sciences, 18, 4703–4720, https://doi.org/10.5194/hess-18-4703-2014, 2014.

Förster, K., Garvelmann, J., Meißl, G., and Strasser, U.: Modelling forest snow processes with a new version of WaSiM, Hydrological Sciences Journal, 63, 1540–1557, https://doi.org/10.1080/02626667.2018.1518626, 2018.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of Hydrology, 228, 113–129, https://doi.org/10.1016/S0022-1694(00)00144-X, 2000.

Habersack, H. and Krapesch, G.: Hochwasser 2005 - Ereignisdokumentation: der Bundeswasserbauverwaltung, des Forsttechnischen Dienstes für Wildbach- und Lawinenverbauung und des Hydrographischen Dienstes, 2006.

Heffernan, J. E. and Tawn, J. A.: A conditional approach for multivariate extreme values (with discussion), Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66, 497–546, https://doi.org/10.1111/j.1467-9868.2004.02050.x, 2004.

Hundecha, Y. and Merz, B.: Exploring the relationship between changes in climate and floods using a model-based analysis, Water Resources Research, 48, n/a, https://doi.org/10.1029/2011WR010527, 2012.

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, https://doi.org/10.1029/2008WR007453, 2009.

Huttenlau, M. and Stötter, J.: The structural vulnerability in the framework of natural hazard risk analyses and the exemplary application for storm loss modelling in Tyrol (Austria), Natural Hazards, 58, 705–729, https://doi.org/10.1007/s11069-011-9768-x, 2011.

Huttenlau, M., Schneeberger, K., Winter, B., Reiss, J., and Stötter J.: Analysis of loss probability relation on community level: a contribution to a comprehensive flood risk assessment, in: Disaster Management and Human Health Risk IV, edited by Sener, S. M., Brebbia, C. A., and Ozcevik, O., pp. 171–182, 2015.

IAWG: HORA / Vorarlberg: Hydraulische Neuberechnung für Vorarlberg, unpublished, 2010.

IPCC: Impacts of 1.5C of Global Warming on Natural and Human Systems, 2018.

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C. J. H., Reinhard Mechler, Botzen W.J.W., Bouwer, L. M., Pflug, G., Rojas, R., and Ward, P. J.: Increasing stress on disaster-risk finance due to large floods, Nature Climate Change, 4, 264–268, https://doi.org/10.1038/NCLIMATE2124, 2014.

Keef, C., Tawn, J., and Svensson, C.: Spatial risk assessment for extreme river flows, Journal of the Royal Statistical Society: Series C (Applied Statistics), 58, 601–618, https://doi.org/10.1111/j.1467-9876.2009.00672.x, 2009.

Keef, C., Tawn, J. A., and Lamb, R.: Estimating the probability of widespread flood events, Environmetrics, 24, 13–21, https://doi.org/10.1002/env.2190, 2013.

Kleindienst, H.: Erweiterung und Erprobung eines anwendungsorientierten hydrologischen Modells zur Gangliniensimulation in kleinen Wildbacheinzugsgebie: Diplomarbeit, 1996.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, https://doi.org/10.1080/02626668609491024, 1986.

Klijn, F., Merz, B., Penning-Rowsell, E. C., and Kundzewicz, Z. W.: Preface: climate change proof flood risk management, Mitigation and Adaptation Strategies for Global Change, 20, 837–843, https://doi.org/10.1007/s11027-015-9663-y, 2015.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424-425, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Koivumäki, L., Alho, P., Lotsari, E., Käyhkö, J., Saari, A., and Hyyppä, H.: Uncertainties in flood risk mapping: a case study on estimating building damages for a river flood in Finland, Journal of Flood Risk Management, 3, 166–183, https://doi.org/10.1111/j.1753-318X.2010.01064.x, 2010.

Kundzewicz, Z. W., Pińskwar, I., and Brakenridge, G. R.: Large floods in Europe, 1985–2009, Hydrological Sciences Journal, 58, 1–7, https://doi.org/10.1080/02626667.2012.745082, 2013.

Laaha, G., Skøien, J. O., and Blöschl, G.: Spatial prediction on river networks: Comparison of top-kriging with regional regression, Hydrological Processes, 28, 315–324, https://doi.org/10.1002/hyp.9578, 2014.

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., and Batstone, C.: A new method to assess the risk of local and widespread flooding on rivers and coasts, Journal of Flood Risk Management, 3, 323–336, https://doi.org/10.1111/j.1753-318X.2010.01081.x, 2010.

Lamb, R., Faulkner, D., Wass, P., and Cameron, D.: Have applications of continuous rainfall-runoff simulation realized the vision for process-based flood frequency analysis?, Hydrological Processes, 30, 2463–2481, https://doi.org/10.1002/hyp.10882, 2016.

Merz, B., Kreibich, H., and Apel, H.: Flood risk analysis: uncertainties and validation, Österreichische Wasser- und Abfallwirtschaft, 60, 89–94, https://doi.org/10.1007/s00506-008-0001-4, http://dx.doi.org/10.1007/s00506-008-0001-4, 2008.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A.: Assessment of economic flood damage, Natural Hazards and Earth System Science, 10, 1697–1724, https://doi.org/10.5194/nhess-10-1697-2010, 2010.

Merz, B., Kreibich, H., and Lall, U.: Multi-variate flood damage assessment: a tree-based data-mining approach, Natural Hazards and Earth System Science, 13, 53–64, https://doi.org/10.5194/nhess-13-53-2013, 2013.

Metin, A. D., Dung, N. V., Schröter, K., Vorogushyn, S., Guse, B., Kreibich, H., and Merz, B.: The role of spatial dependence for large-scale flood risk estimation, Natural Hazards and Earth System Science, 20, 967–979, https://doi.org/10.5194/nhess-20-967-2020, 2020.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola, P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfurtscheller, C., Poussin, J., Przyluski, V., Thieken, A. H., and Viavattene, C.: Review article: Assessing the costs of natural hazards – state of the art and knowledge gaps, Natural Hazards and Earth System Science, 13, 1351–1373, https://doi.org/10.5194/nhess-13-1351-2013, 2013.

Molinari, D., de Bruijn, K. M., Castillo-Rodríguez, J. T., Aronica, G. T., and Bouwer, L. M.: Validation of flood risk models: Current practice and possible improvements, International Journal of Disaster Risk Reduction, 33, 441–448, https://doi.org/10.1016/j.ijdrr.2018.10.022, 2019.

Müller-Thomy, H., Wallner, M., and Förster, K.: Rainfall disaggregation for hydrological modeling: is there a need for spatial consistence?, Hydrology and Earth System Sciences, 22, 5259–5280, https://doi.org/10.5194/hess-22-5259-2018, 2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Peleg, N., Fatichi, S., Paschalis, A., Molnar, P., and Burlando, P.: An advanced stochastic weather generator for simulating 2-D high-resolution climate variables, Journal of Advances in Modeling Earth Systems, 9, 1595–1627, https://doi.org/10.1002/2016MS000854, 2017.

Plouffe, C. C., Robertson, C., and Chandrapala, L.: Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka, Environmental Modelling & Software, 67, 57–71, https://doi.org/10.1016/j.envsoft.2015.01.011, 2015.

Raynaud, D., Hingray, B., Evin, G., Favre, A.-C., and Chardon, J.: Assessment of meteorological extremes using a synoptic weather generator and a downscaling model based on analogs, https://doi.org/10.5194/hess-2019-557, 2019.

Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., Chaplot, V., Gallart, F., Glatzel, G., Hall, J., Holden, J., Holko, L., Horn, R., Kiss, A., Kohnová, S., Leitinger, G., Lennartz, B., Parajka, J., Perdigão, R., Peth, S., Plavcová, L., Quinton, J. N., Robinson, M., Salinas, J. L., Santoro, A., Szolgay, J., Tron, S., van den Akker, J. J. H., Viglione, A., and Blöschl, G.: Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research, Water resources research, 53, 5209–5219, https://doi.org/10.1002/2017WR020723, 2017.

Sauter, I., Kienast, F., Bolliger, J., Winter, B., and Pazur R.: Changes in demand and supply of ecosystem services under scenarios of future land use in Vorarlberg, Austria, Journal of Mountain Science, 12, 2793–2809, https://doi.org/10.1007/s11629-018-5124-x, 2019.

Schneeberger, K. and Steinberger, T.: Generation of Spatially Heterogeneous Flood Events in an Alpine Region—Adaptation and Application of a Multivariate Modelling Procedure, Hydrology, 5, 5, https://doi.org/10.3390/hydrology5010005, 2018.

Schneeberger, K., Huttenlau, M., Winter, B., Steinberger, T., Achleitner, S., and Stötter, J.: A Probabilistic Framework for Risk Analysis of Widespread Flood Events: A Proof-of-Concept Study, Risk Analysis, 39, 125–139, https://doi.org/10.1111/risa.12863, 2019.

Senfter, S., Leonhardt, G., Oberparleiter, C., Asztalos, J., Kirnbauer, R., Schöberl, F., and Schönlaub, H.: Flood Forecasting for the River Inn, in: Sustainable Natural Hazard Management in Alpine Environments, edited by Veulliet, E., Johann, S., and Weck-Hannemann, H., pp. 35–67, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-03229-5_2, 2009.

Serinaldi, F. and Kilsby, C. G.: A Blueprint for Full Collective Flood Risk Estimation: Demonstration for European River Flooding, Risk analysis : an official publication of the Society for Risk Analysis, 37, 1958–1976, https://doi.org/10.1111/risa.12747, 2017.

Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydrology and Earth System Sciences, pp. 277–287, 2006.

Speight, L. J., Hall, J. W., and Kilsby, C. G.: A multi-scale framework for flood risk analysis at spatially distributed locations, Journal of Flood Risk Management, pp. 124–137, https://doi.org/10.1111/jfr3.12175, 2017.

Statistik Austria: Baupreisindex, https://www.statistik.at/web_de/statistiken/wirtschaft/produktion_und_bauwesen/konjunkturdaten/baupreisindex/index.html, 2019.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B., Dutta, D., and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental Modelling & Software, 90, 201–216, https://doi.org/10.1016/j.envsoft.2017.01.006, 2017.

Thieken, A. H., Apel, H., and Merz, B.: Assessing the probability of large-scale flood loss events: a case study for the river Rhine, Germany, Journal of Flood Risk Management, pp. 247–262, https://doi.org/10.1111/jfr3.12091, 2015.

Ullrich, S., Hegnauer, M., Dung, N. V., de Bruijn, K., Merz, B., Kwadijk, J., and Vorogushyn, S.: Comparative evaluation of two types of stochastic weather generators coupled to hydrological models for flood estimation: (Oral), 2019.

Vorogushyn, S., Bates, P. D., de Bruijn, K., Castellarin, A., Kreibich, H., Priest, S., Schröter, K., Bagli, S., Blöschl, G., Domeneghetti, A., Gouldby, B., Klijn, F., Lammersen, R., Neal, J. C., Ridder, N., Terink, W., Viavattene, C., Viglione, A., Zanardo, S., and Merz, B.: Evolutionary leap in large-scale flood risk assessment needed, Wiley Interdisciplinary Reviews: Water, 5, 1–7, https://doi.org/10.1002/wat2.1266, 2018.

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, Water Resources Research, 43, https://doi.org/10.1029/2006WR005308, 2007.

Wagenaar, D. J., de Bruijn, K. M., Bouwer, L. M., and de Moel, H.: Uncertainty in flood damage estimates and its potential effect on investment decisions, Natural Hazards and Earth System Sciences, 16, 1–14, https://doi.org/10.5194/nhessd-3-607-2015, 2016.

Winter, B., Schneeberger, K., Huttenlau, M., and Stötter, J.: Sources of uncertainty in a probabilistic flood risk model, Natural Hazards, pp. 431–446, https://doi.org/10.1007/s11069-017-3135-5, 2018.

Winter, B., Schneeberger, K., Dung, N. V., Huttenlau, M., Achleitner, S., Stötter, J., Merz, B., and Vorogushyn, S.: A continuous modelling approach for design flood estimation on sub-daily time scale, Hydrological Sciences Journal, 88, 1–16, https://doi.org/10.1080/02626667.2019.1593419, 2019.