https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

The impact of hydrological model structure on the simulation of extreme runoff events

Gijs van Kempen[1], Karin van der Wiel[2], and Lieke Anna Melsen[1]

[1]Hydrology and Quantitative Water Management, Wageningen University, Wageningen, the Netherlands
[2]Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

Correspondence: lieke.melsen@wur.nl

Abstract. Hydrological extremes affect societies and ecosystems around the world in many ways, stressing the need to make reliable predictions using hydrological models. However, several hydrological models can be selected to simulate extreme events. A difference in hydrological model structure results in a spread in the simulation of extreme runoff events. We investigated the impact of different model structures on the magnitude and timing of simulated extreme high- and low-flow events, by combining two state-of-the-art approaches; a modular modelling framework (FUSE) and large ensemble meteorological simulations. This combination of methods created the opportunity to isolate the impact of specific hydrological process formulations at long return periods without relying on statistical models. We showed that the impact of hydrological model structure was larger for the simulation of low-flow compared to high-flow events and varied between the four evaluated climate zones. In cold and temperate climate zones, the magnitude and timing of extreme runoff events were significantly affected by different parameter sets and hydrological process formulations, such as evaporation. The impact of hydrological model structures on extreme runoff events was smaller in the arid and tropical climate zones. This novel combination of approaches provided insights into the importance of specific hydrological processes formulations in different climate zones, which can support adequate model selection for the simulation of extreme runoff events.

1 Introduction

Extreme high and low flow events, often referred to as floods and droughts, respectively, have high natural, societal and economic impacts. On the global scale, fatalities and economic losses related to flood events have increased dramatically over the past decades (Di Baldassarre et al., 2010; Winsemius et al., 2016), among others due to an increase of settlements in flood prone regions. The impacts of drought events can be recognised in amongst others the water supply, crop production, and the hydro-power sectors (Van Loon, 2015). To mitigate the societal impact of hydrological extremes, knowledge of the processes leading to these extreme events is vital. Hydrological modelling is one of the main tools in this quest for knowledge, but comes with uncertainties. Here we aim to investigate the impact of hydrological model structure on the magnitude and timing of simulated extreme runoff events.

Hydrological mitigation efforts often relate to the return period of the extreme event, a measure that describes the 'extremeness' of the events. It is a traditional method to relate the magnitude of an event to the probability of occurrence of the event (Gumbel, 1941; Salas and Obeysekera, 2014), based on which decision makers can define their policy. Frequency analysis of extremes aims at estimating runoff levels corresponding to certain return periods (Laio et al., 2009). However, the limited length of available observational hydrological records means we frequently rely on a statistical models to estimate return periods (Meigh et al., 1997; Michele and Rosso, 2001; Smith et al., 2015; Sousa et al., 2011), mostly by fitting a Generalised Extreme Value (GEV) distribution.

Despite the wide application of GEV analysis to relate runoff to return periods, there are some important caveats to this method. The statistical models are particularly used for extrapolation - to estimate the probability of yet unobserved extremes. As such, the projected hydrological extremes are highly sensitive to small changes in the parameters of statistical model (Engeland et al., 2004; Smith et al., 2015), leading to distributions that might substantially change when a single data point is added (see e.g. Brauer et al., 2011). Furthermore, the physical processes leading to extrapolated extreme events can not be investigated. A recent alternative to extreme value statistics models, proposed for example by Van der Wiel et al. (2019), is to use large ensemble model simulations: a climate model simulates long time series of meteorological conditions, and with a hydrological model this is translated to runoff, resulting in a long time series that does not require extrapolation for the investigation of events of longer return periods.

In this approach, hydrological models are employed to translate meteorological time series into hydrological time series, from which relevant events can be selected and investigated. Uncertainty is, however, also inevitable in model simulations (Oreskes et al., 1994). In hydrological modelling, different sources of uncertainty can be distinguished, for instance data uncertainty, parameter uncertainty and model structural uncertainty (Ajami et al., 2007). Data uncertainty can be related to random or systematic errors in the model forcing. Parameter uncertainty can be caused by sub-optimal identification of parameter values or equifinality, and model structural uncertainty relates to incomplete or biased model structures (Butts et al., 2004). It is important to gain insight in the uncertainty of environmental models and to communicate these insights to decision makers (Liu and Gupta, 2007), especially in the perspective of extreme events that give rise to policy making.

Data uncertainty and parameter uncertainty can be quantified by a combination of error propagation and sampling (Li et al., 2010; McMillan et al., 2011a). The quantification of model structural uncertainty is more challenging, since it takes a considerable amount of time and effort to set-up and run several model structures. Furthermore, it is difficult to link intermodel differences to alterations in certain hydrological process formulations (Clark et al., 2008), because models often differ in several process formulations. This is where the use of modular modelling frameworks (MMF), a tool which facilitates switching between model structures (Addor and Melsen, 2019), might provide ways forward in the evaluation of model structural uncertainty. In an MMF it is possible to alter a minor part of the model structure, which allows the researcher to isolate choices in the model development process (Knoben et al., 2019). The Framework of Understanding Structural Errors (FUSE, Clark et al.,

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

2008) is an example of a modular modelling framework, which can be used to diagnose differences in hydrological model structures.

This study is designed to evaluate the impact of hydrological model structure on the magnitude and timing of simulated extreme runoff events with different return periods. We combine two state-of-the-art approaches: the hydrological modular modelling framework FUSE, and large ensemble meteorological simulations. The forcing data-set consists of 2,000 years of daily meteorological data, representing the present-day climate conditions. This data set will be used to force several hydrological models within the FUSE framework. The different model structures will be used to evaluate various hydrological process formulations, to determine which process formulations have the largest impact on the simulated magnitude and timing of extreme high- and low-flow events in different climate zones. Due to the length of the forcing time series, the extreme runoff events in the tail of the distribution can be evaluated using simulated values. Hence, we do not rely on statistical models to extrapolate extreme events. As such, this study contributes to the understanding of the impact of model structural uncertainty in hydrological models on simulated extreme runoff events.

2 Methods

We assessed the impact of hydrological model structural uncertainty on extreme runoff events by using large ensemble meteorological simulations in combination with the hydrological modular modelling framework FUSE. We examined four different climate zones, because hydrological processes vary considerably between climate zones (Pilgrim, 1983), which leads to different processes being of importance in controlling the extreme events (Di Baldassarre et al., 2017; Eagleson, 1986). In the R-version of FUSE (Vitolo et al., 2015), ten different model structures were employed, and to capture the complete parameter space, 100 parameter sets were used in every model structure. The simulated extreme runoff events were compared based on their magnitude and timing.

2.1 Meteorological forcing data

We employed a 2,000 year time series of meteorological data, generated by the EC-Earth global coupled climate model (v2.3, Hazeleger et al., 2012). This 2,000 year time series originally consisted of a large ensemble of 400 sets of 5 year runs. In this study, these 400 sets were assumed to be one long time series, which enables extensive return period analysis. This time series represents a period with a simulated absolute Global Mean Surface Temperature (GMST) equal to the observed GMST in the years 2011-2015 based on HadCRUT4 data (Morice et al., 2012). The time series thus represents present-day climatic conditions. In Van der Wiel et al. (2019), this data-set was used to evaluate the benefits of the large ensemble technique for hydrology. Further details on the design of the meteorological forcing data are provided in that paper.

For this study we restricted ourselves to four climate zones (Figure 1). The evaluated climate zones and their corresponding Köppen-Geiger classifications are: arid (BWh), cold (Dfc), temperate (Cfb) and tropical (Aw). This set of climate zones

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

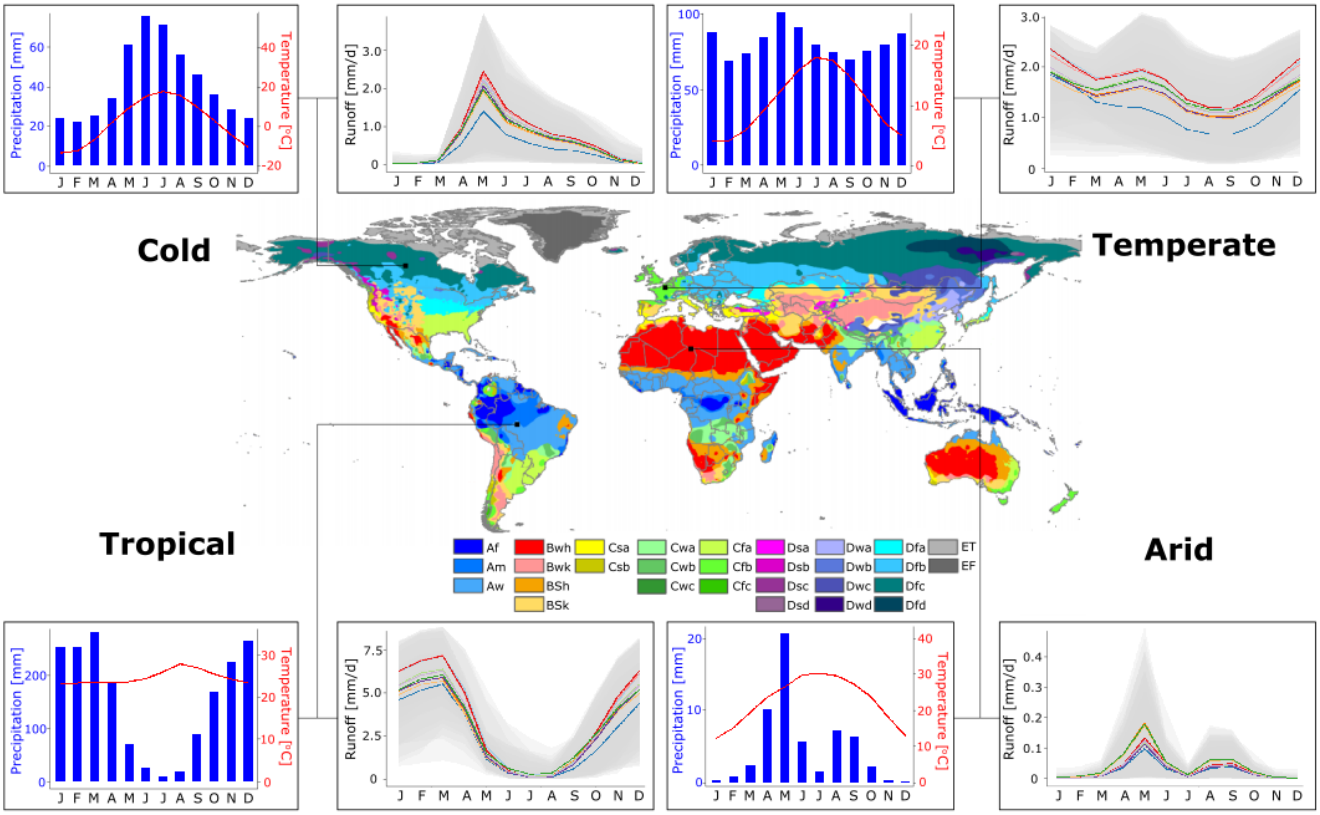

Figure 1. Köppen-Geiger Climate type map indicating the locations of the selected grid cells for the four different climate zones and their corresponding climatology and hydrology (central map taken from Peel et al. (2007), their Figure 10). The climate graphs show simulated climatological monthly precipitation sums (blue bars) and monthly average temperatures (red lines). The hydrological conditions are visualised using simulated monthly average runoff levels. The different line colours represent the ten evaluated model structures and the spread induced by the different parameter sets is shown using grey bands.

offers a comprehensive representation of the global climate zones (Kottek et al., 2006; Peel et al., 2007). Simulated monthly averaged 2 m temperatures and precipitation sums were obtained from the EC-Earth model to classify grid cells based on the Köppen-Geiger criteria, and allow the selection of appropriate grid cells for this study.

Daily 2 m temperature, precipitation and potential evapotranspiration data for the full 2,000 years were then acquired for the four selected grid cells. The 2 m temperature and daily precipitation fluxes were directly available from the EC-Earth model. Potential evapotranspiration fluxes were calculated following the Penman-Monteith procedure (Zotarelli et al., 2015). The precipitation and potential evapotranspiration fluxes were used as input in the FUSE models, the 2 m temperature was used to force the snow module (see Section 2.2).

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

2.2 Framework of Understanding Structural Errors (FUSE)

FUSE is a modular modelling framework, which can be used to diagnose differences in hydrological model structures (Clark et al., 2008). FUSE is developed based on four parent models; the U.S. Geological Survey's Precipitation-Runoff Modelling System (PRMS, Leavesley, 1984), the NWS Sacramento model (Burnash et al., 1973), TOPMODEL (Beven and Freer, 2001) and different versions of the Variable Infiltration Capacity (ARNO/VIC) model (Liang et al., 1994). This framework enables the assessment of intermodel differences in another way compared to other model intercomparison studies (Henderson-Sellers et al., 1993; Reed et al., 2004). In FUSE, each model component can be adapted in isolation and therefore the effect of specific hydrological process formulations can be investigated. In the next subsection we further discuss which model structures we selected and which process formulations were tested.

All model structures used in this study were lumped hydrological models, which were run at a daily time step. We employed a spin-up period of five years, before forcing the hydrological models with the 2,000 year meteorological time series. The simulated monthly average runoff varied among the evaluated model structures and parameter sets (Figure 1). Therefore, it is essential to select an adequate hydrological model for the simulation of runoff levels, and it will likely be of larger importance when simulating extreme runoff events.

FUSE as implemented in R (Vitolo et al., 2015) does not include a snow module. However, snow storage and snow melt might be important components in the hydrological cycle of the colder climate zones. Therefore, a snow module was implemented. First, a threshold temperature was defined at 0°C, below which precipitation is assumed to fall as snow. Secondly, snow melt is simulated by using a simple degree-day method (Kustas et al., 1994):

$$M = a(T_a - T_b), \tag{1}$$

in which M represents snow melt (mm), a the degree-day factor (mm/°C/day), T_a the average daily temperature and T_b the base temperature. The degree-day factor was fixed at a value of 0.475 mm/°C/day based on Kustas et al. (1994), and T_b was set to 0°C. The degree-day method employed daily 2 m temperature data to subdivide the precipitation data into rain and snow and to determine the melt rate. The different FUSE model structures were subsequently forced by these subdivided precipitation fluxes.

2.2.1 Selected structures

In total, 1248 different model structures can be constructed in FUSE as implemented in R (Vitolo et al., 2015) by combining different hydrological process formulations from the parent models. The architecture of the upper and lower layer can be altered, and the process formulations for simulating base flow, evaporation, percolation, surface runoff, interflow and routing can be changed. The lower layer architecture is intimately tied to the process formulation of base flow. Therefore, they need to be changed simultaneously and only a few combinations are possible. In our synthetic experiment, there was no routing

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

module included and the process formulation of interflow was left unchanged throughout this study, as it was not explicitly parameterised in TOPMODEL and ARNO/VIC (Clark et al., 2008).

Table 1. The model structures that were employed in this study. Each letter refers to a specific hydrological process formulation as in Clark et al. (2008), the model IDs are described by Vitolo et al. (2015). The model abbreviations are related to the alteration in the model structure and are used throughout this paper.

Model Component	**Model Number**									
	1	2	3	4	5	6	7	8	9	10
Upper Layer	A	B	C	C	C	C	A	A	A	A
Lower Layer	A	A	A	C	B	B	B	B	C	C
Base Flow	A	A	A	B	C	C	C	C	B	B
Evaporation	A	A	B	B	A	B	A	A	A	A
Percolation	C	C	C	C	C	C	C	B	B	B
Interflow	A	A	A	A	A	A	A	A	A	A
Surface Runoff	A	A	A	A	B	B	A	A	A	B
Routing	A	A	A	A	A	A	A	A	A	A
Model ID	802	800	642	626	808	652	790	880	874	896
Abbreviation	UL1	UL2	LL1	LL2	EV1	EV2	PC1	PC2	SR1	SR2
Alteration	Upper Layer		Lower Layer		Evaporation		Percolation		Surface Runoff	

Ten different model structures were evaluated in this study. Table 1 provides an overview of the selected hydrological model structures. In the odd model numbers, new model structures were constructed and in the even model numbers, a single hydrological process was altered in the model structure relative to the preceding odd model number. By comparing the extreme runoff events simulated between consecutive odd and even numbered model structures, we analysed the impact of a specific hydrological process on extreme event simulation, indicated by the alteration in Table 1.

In contrast to other studies that evaluate different model structures (Atkinson et al., 2002), this study evaluated differences among model structures that are deemed to be equally plausible. Hence, there were no prior expectations of specific models to outperform other models. This means that the emphasis in FUSE is not on the lacking parts of hydrological models, but on the intermodel differences that are caused by different representations of the real world (Clark et al., 2008).

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

2.2.2 Parameters

In this synthetic experiment, the parameters of the hydrological models were not calibrated to real catchment observations. Instead, the parameters of the models were sampled over their full range. Since in calibrated experiments it is always difficult to differentiate the effect of parameter values from the effect of model structure, the parameter sampling approach also created the opportunity to assign the effect on extreme events either to parameter values or to model structure.

To investigate the appropriate and feasible number of parameter sets required to sufficiently capture the parameter space, the Kolmogorov-Smirnov test was employed (Massey Jr, 1951). The Kolmogorov-Smirnov test evaluates whether the differences in distribution of hydrological model output between a different number of parameter sets is significant. We applied the Kolmogorov-Smirnov test to assess the annual maximum and minimum daily runoff from 10 up to 200 parameter sets, each time with 10 samples increment. The model runs were executed for 30 years to save computation time, because this is considered sufficient to represent the climate conditions (McMichael et al., 2004). The D-statistic describes the largest distance between the Empirical Cumulative Distribution Functions (ECDF), which indicates that when the D-statistic decreases, the ECDFs are more likely to originate from the same data-set. The D-statistic was evaluated based on the difference between the distribution based on a certain sample size, and a reference distribution (in this case, 500 parameter sets).

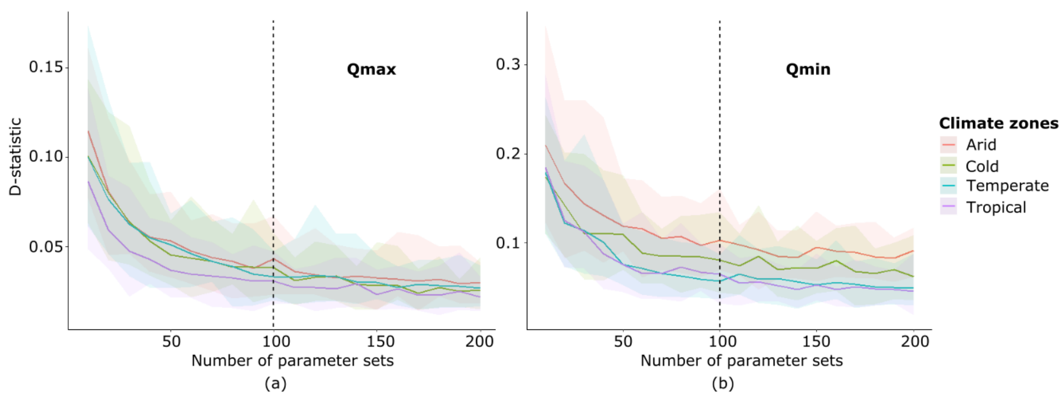

Figure 2. D-statistics for one model structure (UL1) with twelve parameters, which result from the Kolmogorov-Smirnov test. The other model structures show a similar trend (not shown). The bands are a result of the different parameter samples, the different colours represent the four climate zones.

We found that the optimal trade-off between computer time and sufficiently capturing parameter space was at 100 parameter sets, as the D-statistic stabilised at this value (Figure 2). Since there are different process formulations, the number of sampled parameters varied between eleven and fifteen for the different model structures. Nevertheless, for justification we used 100 parameter sets for all model structures, independent of the number of parameters. The parameter sets were generated using Latin Hypercube Sampling, based on the parameter ranges provided in Clark et al. (2008).

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Discussions

2.3 Magnitude of extreme runoff events

The magnitudes of the simulated extreme events were evaluated by comparing the distribution of runoff values based on four return periods: 25, 50, 100 and 500 years. The associated runoff levels were determined by sorting the time series of annual maximum and minimum daily runoff values. This resulted in 2,000 sorted runoff values from which events were selected. For instance, for the 500-year return period, the 4th most extreme value in the sorted time series was taken.

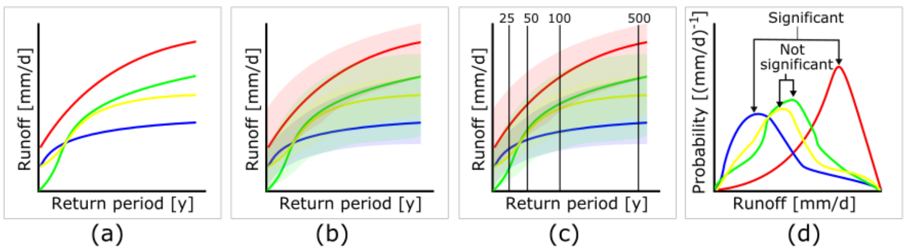

Figure 3. Illustration of the conducted procedure for the comparison of the extreme event magnitude. (a) The different lines represent the model structures. (b) The uncertainty bounds are due to the 100 different parameter samples per model. (c) The parameter samples were compared at different return periods. (d) The projected difference between the distributions at a given return period of the various model structures was tested using a two-sample t-test, an example of a significant and a not-significant difference is shown.

The different model structures yielded different simulated magnitudes for extreme runoff events (Figure 3a). Every model structure was run using 100 different parameter sets, which led to bands around the projected extreme runoff events (Figure 3b). The runoff values and their bands were subsequently evaluated for 25, 50, 100, and 500 year return periods (Figure 3c). The different parameter sets resulted in 100 extreme runoff values at a specific return period for every model structure. In order to test whether the projected difference in the distributions of these runoff values (Figure 3d) was significantly different from the paired model, a two-sample t-test was applied. This test was used to evaluate related model structures based on a change in one single hydrological process formulation (Table 1). By comparing related model structures, the impact of corresponding hydrological process formulations could be isolated for specific climate zones and return periods.

2.4 Timing of extreme runoff events

An asset from the ensemble approach for return period evaluation compared to GEV statistics, is that it also allows us to evaluate the timing of the 500-year events based on the entire 2,000 year time series. Extreme hydrological events do not always result from extreme meteorological conditions, but could also originate from a sequence of moderate weather conditions (Van der Wiel et al., 2020). By assessing the timing of extreme runoff events, we investigated whether the timing of the extreme runoff events is controlled by different model structures and parameter sets or mainly determined by the meteorological forcing

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Discussions

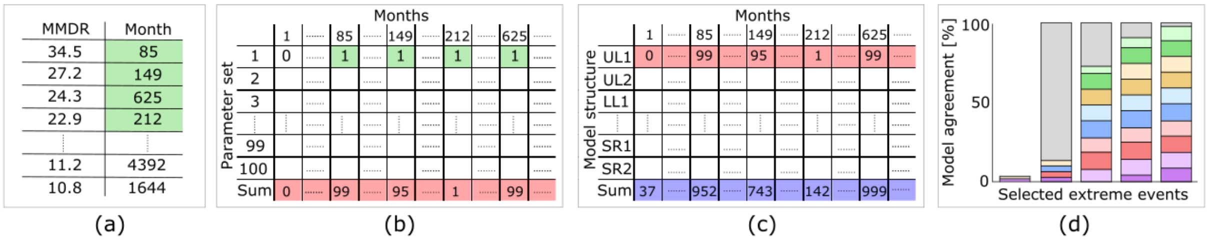

Figure 4. Illustration of the conducted procedure for the comparison of the timing of extreme events, equal or greater than 500-year events. (a) The monthly maximum daily runoff (MMDR) values were sorted and the four most extreme events were selected (green cells), this table shows an example for one parameter set. (b) All parameter sets of one model structure were concatenated and the sum (red cells) indicated the variation in timing in one model structure. A score of 100 means that for all different parameter sets, the same event is selected. (c) All model structures of one climate zone were concatenated and the sum (blue cells) indicated the variation in extreme event timing in all model structures for one climate zone. The values in the blue cells have a maximum score of 1000 (10 models, with 100 parameter samples each). (d) Stacked bar charts are used to visualise the model agreement of specific runoff events. The coloured bars represent the values of the blue cells for different model structures as shown in panel (c), and the grey bars indicate the theoretical maximum for 500-year events: four runoff events with 100 % model agreement.

The timing of extreme runoff events with 500-year return periods were compared. This was done by sorting all the monthly maxima and minima daily runoff values and their corresponding simulation month (Figure 4a). The four most extreme events in this sorted 2,000-year data-set represent the extreme events equal or greater than the 500-year events. When the corresponding months of these four events matched for all parameter sets within one model structure, this would lead to a high value in the red row (Figure 4b), which indicates that the influence of hydrological parameters on the timing of extreme events is negligible. The same procedure was followed for the ten different model structures to evaluate the sensitivity of the timing of the extreme events to the model structures (Figure 4c). Finally, the model agreement of the specific extreme runoff events was evaluated in stacked bar charts (Figure 4d). The colours of the stacked bars represent the different model structures and the height of these bars indicates the model agreement within a specific model structure for different parameter values. The percentage of model agreement was determined by the amount of model simulations that identify a specific extreme runoff event out of a total of 1000 model simulations, where all model simulations employed a unique combination of a model structure and a corresponding parameter set.

3 Results

3.1 Magnitude of extreme runoff events

We compared the distribution of the magnitudes of the extreme high- and low-flow events for related model structures, based on four different return periods, and for four different climate zones. Alterations in the hydrological process formulations lead to a difference in the magnitude of extreme runoff events, as depicted in Figure 3a. Figure 5 shows the same information,

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

but now based on actual simulations of high-flow (Figure 5a) and low-flow (Figure 5b) in the arid climate zone for four selected models. The model structures that are related by an alteration in the process formulation of percolation, simulate an increasing difference in extreme high-flow magnitude for longer return periods (Figure 5a). Based on the t-test conducted on the distributions of the 500-year return period, this results in a significant impact of alterations in the process formulation of percolation for this return period (as displayed in Figure 6). In contrast, the model structures related by an alteration in the process formulation of evaporation, simulate comparable runoff values across all return periods (Figure 5a). Therefore, there is no significant impact on the magnitude of extreme high-flow events caused by this hydrological process formulation (Figure 6).

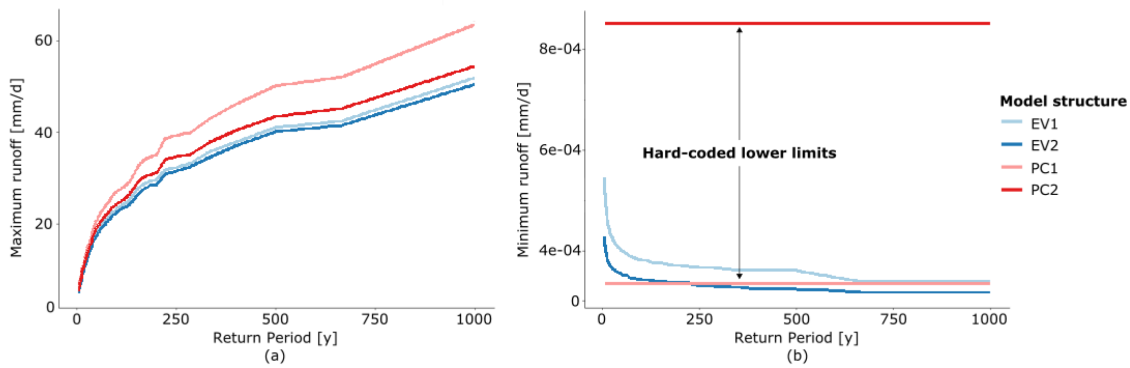

Figure 5. The ensemble mean of the annual maximum (a) and minimum (b) daily runoff levels at different return periods in the arid climate zone. The ensemble mean is obtained based on 100 parameter sets. Four model structures are visualised, which are related by alterations in the evaporation (EV1, EV2) and percolation (PC1, PC2) process formulations (Table 1).

This section will describe the impact of model structures on extreme event magnitude for different climate zones, hydrological process formulations and return periods. A two-sample t-test was employed to calculate the p-values (Figure 6), which were used to distinguish the statistically significant ($p < 0.05$) and non-significant ($p > 0.05$) differences in the distribution of extreme event magnitudes as in Figure 3d.

An alteration in the model structure has significant impact in about a quarter of the model output comparisons during high-flow events (Figure 6a). The difference between the magnitude distributions of the high-flow events is non-significant for alterations in the architecture of the upper and lower layer and in the process formulation of surface runoff. This means that the magnitude of high-flow events for all climate zones and return periods are not significantly sensitive to changes in the formulation of these hydrological processes. In the arid climate zone, the impact of alterations in model structures on high-flow events has least impact. This indicates that the magnitudes of the high-flow events are mainly controlled by the meteorological forcing. In the cold and temperate climate zones, the high-flow events are sensitive to alterations in the process formulation of two hydrological processes; evaporation and percolation. This indicates that the magnitudes of the high-flow events are not only determined by the meteorological forcing, but there is also a notable impact of the hydrological model structure, specif-

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Discussions

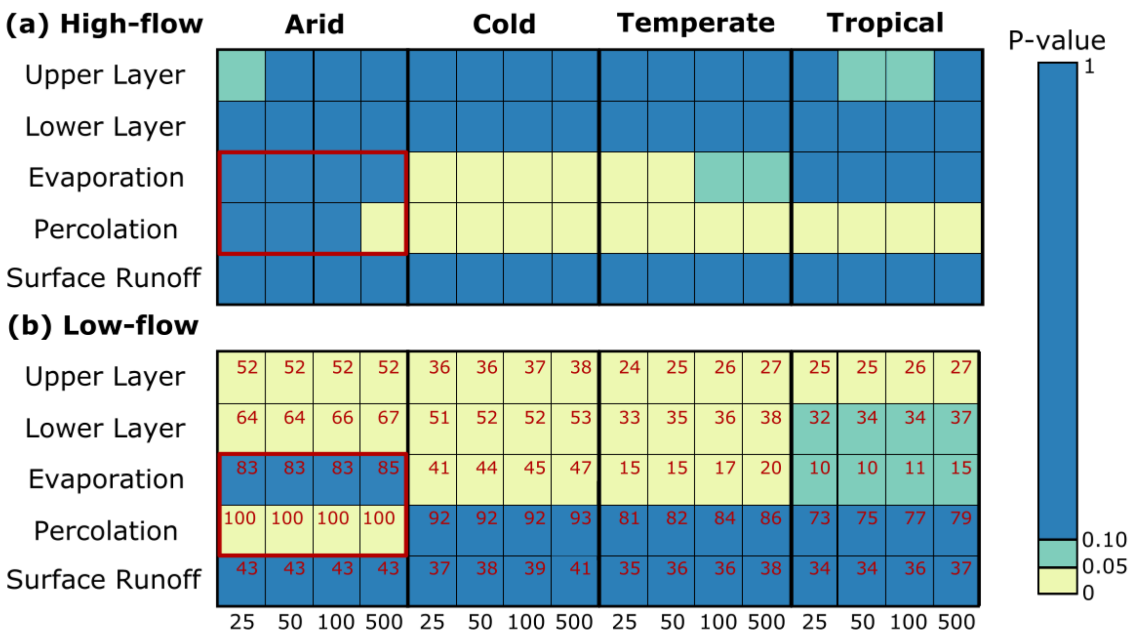

Figure 6. Statistically significant (p < 0.05) and non-significant (p > 0.05) differences between the distribution of magnitudes for extreme runoff events, assessed by a two-sample t-test. The colours indicate whether an alteration in the model structure has a statistically significant impact on the magnitude of extreme high- (a) and low-flow (b) events. This is shown for the four climate zones (arid, cold, temperate and tropical, indicated at the top) and the four different return periods (25, 50, 100 and 500 years, indicated at the bottom). The red values in (b) indicate the percentage of simulations which reached a hard-coded lower limit. The red boxes indicate the magnitude distributions that were shown in Figure 5.

ically for the formulation of these two processes. Finally, in the tropical climate zone, the high-flow events are only sensitive to alterations in the process formulation of percolation. The other hydrological process formulations do not significantly affect the magnitude of high-flow events in this climate zone.

For low-flow events, the model structure has a greater impact on the simulation of extreme runoff events. An alteration in the model structure has significant impact in half of the model output comparisons during low-flow events (Figure 6b). In the arid climate zone, the low-flow events are sensitive to alterations in the architecture of the upper and lower layer and in the process formulation of percolation. In the cold and temperate climate zones, the low-flow events are also sensitive to alterations in the architecture of the upper and lower layer. Furthermore, changes in the process formulation of evaporation lead to significant differences. In the tropical climate zone, the low-flow events are less sensitive to alterations in the architecture of the lower layer and the process formulation of evaporation. However, these low-flow events are sensitive to alterations in the upper layer architecture. In most climate zones, formulations of multiple hydrological processes significantly impact the simulation of the magnitude of low-flow events, which implies that the model structure is an important source of uncertainty. The meteorological

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

forcing is clearly not the only factor controlling the magnitude of simulated low-flow events.

However, certain conditions in the model formulation create an additional source of uncertainty in the projection of low-flow events. Some specific combinations of model structures and parameter sets lead to a hard-coded lower limit in the runoff output, which implies that the annual minimum runoff is equal at different return periods as demonstrated in Figure 5b. These hard-coded lower limits are not the result of a single model structure, but appear from a combination of model structure and specific parameter values. When two related models simulate a certain hard-coded lower limit, this will result in the same p-value for all return periods. The red numbers in Figure 6b indicates that these hard-coded lower limits are frequently reached in the simulations. The limits can result in significant differences between models with altered process formulations. This is however not a direct result of a different process formulation, but rather the result of a numerical artefact. As such, the effect of the process formulations themselves on low-flow simulations can be overestimated. In the physical world, these lower limits might indicate a zero flow situation where the model formulation is not relevant anymore.

Both differences and similarities can be identified between the distributions of runoff values for the high- and low-flow events. Alterations in hydrological model structures more often result in significant differences in low-flows (50 %) compared to high-flows (24 %), which implies a larger model structural uncertainty in the magnitude of low-flow events (although this might be the result of numerical artefacts). High-flow events mainly depend on precipitation, while the influence of other runoff generating processes such as soil moisture and base flow is marginal (Zhang et al., 2011). The situation during high-flow events is often characterised by a precipitation surplus. Therefore, there will be more or less continuous groundwater recharge by percolation in the unsaturated zone (Knutsson, 1988), which explains why the formulation of percolation appears as a relevant hydrological process to estimate the magnitude of high-flow events.

Hydrological models are traditionally designed to simulate the runoff response to rainfall and therefore, it seems to be more challenging to simulate low-flow events (Staudinger et al., 2011). The low-flow events are mainly sensitive to alterations in the architecture of the upper and lower layer. Earlier research indicates the importance of the lower layer architecture and the process formulation of base flow in simulating low-flow events (Staudinger et al., 2011). The architecture of the upper and lower layer defines the water content in these layers (Clark et al., 2008). This water content is controlling the runoff-generating processes during low-flow events due to a precipitation deficit and reduces the importance of the percolation process (Andersen et al., 1992). Therefore, alterations in the process formulation of percolation mainly affect high-flow events in the wet climate zones (Figure 6). An exception is that low-flow events in the arid climate zone seem to be affected by the process formulation of percolation as well. However, in this situation the hard-coded lower limit is reached in both model structures for all parameter sets (Figure 6b).

Besides these differences, there are also similarities in the simulation of high- and low-flow events. The magnitudes of high- and low-flow events in the cold and temperate climate zone have a similar response to alterations in all hydrological process

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

formulations. Furthermore, alterations in the process formulation of surface runoff have no significant impact on the magnitude of both types of extreme runoff events. This might be due to the lacking implementation of infiltration excess overland flow in FUSE (Clark et al., 2008). This could be an important factor for surface runoff, especially in arid climate zones (Reaney et al., 2014). Another factor might be the temporal resolution of the model runs: the models are run at a daily time step, while surface runoff is especially relevant at shorter time steps (Morin et al., 2001; Melsen et al., 2016).

3.2 Timing of extreme runoff events

The timing of extreme high-flow events is evaluated using stacked bar charts. Figure 7 shows the percentage of model agreement on the timing of extreme high-flow events with a return period equal or greater than 500-years, as earlier depicted in Figure 4d. For the low-flow events, the timing evaluation appeared to be impossible due to the hard-coded lower limits, as will be discussed further in this section.

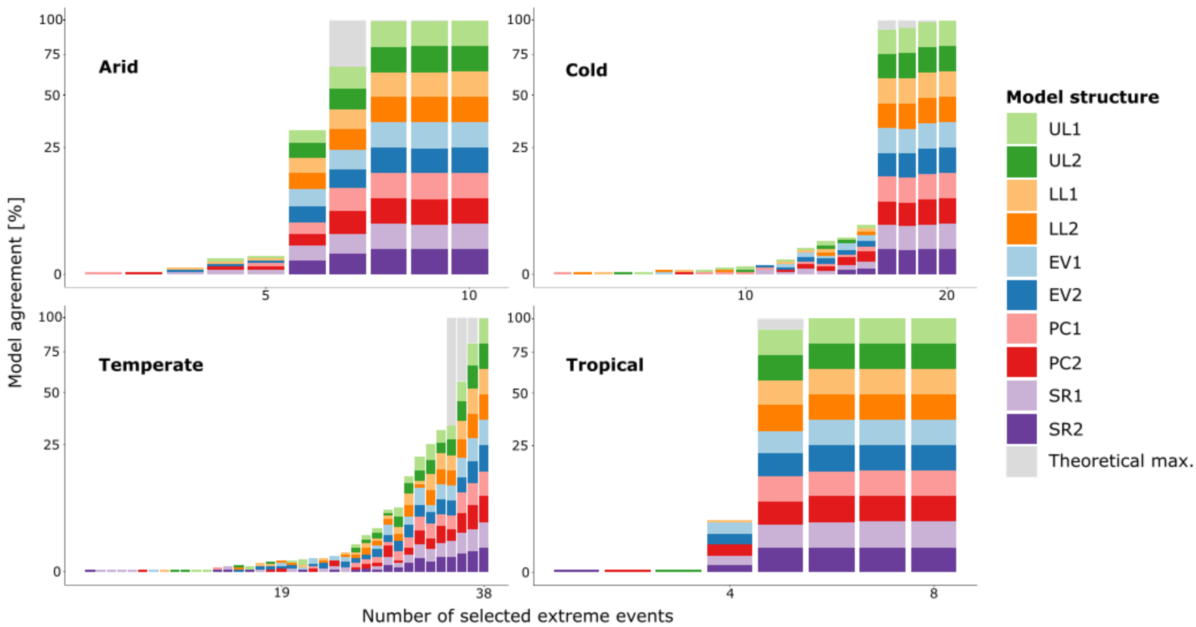

Figure 7. Stacked bar charts that visualise the percentage of model agreement for extreme high-flow events. Four different climate zones are evaluated in the different subplots. Extreme runoff events are identified when they are equal or greater than the 500-year return level. The linked model structures (Table 1) are related to each other by comparable colours. The grey bars indicate the theoretical maximum of four events with 100 % model agreement, which would imply a negligible impact of model structure and parameters on high-flow event timing.

The impact of different hydrological process formulations and parameter sets on the timing of extreme high-flow events varies between the selected climate zones. In the arid and tropical climate zones, there are multiple events with a model agreement exceeding 99 %. In these cases, almost all model simulations agree on the timing of these extreme events. Just ten and

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

eight runoff events were selected (out of a total of 24,000 events) as extreme high-flow events in the arid and tropical climate zones, respectively (Figure 7). This means that there are only a few model simulations that show deviant behaviour by simulating the most extreme runoff events at a different point in the time series. For these climate zones, this implies that the timing is mainly prescribed by the meteorological forcing. This might be explained by the precipitation climatology in these climate zones. On average, in the arid climate zone the daily precipitation sum exceeds 1 mm only during eleven days a year. Precipitation is therefore scarce and characterised by short events of high-intensity (Goodrich et al., 1995), which propagate into extreme runoff events. In the tropical climate zone, there is a high precipitation rate throughout the complete time series. However, there is a pronounced wet season from October until April (Figure 1). There are multiple extreme precipitation events larger than 150 mm/d. The 500-year extreme runoff events are initiated by these extreme precipitation events.

In both the cold and temperate climate zone, there is only one event with a model agreement exceeding 99 % (Figure 7). In the cold and temperate climate zones, there are 20 and 38 different events selected as extreme events, respectively. The selected runoff events with the highest model agreement are initiated by the most extreme precipitation events, whereas the selected extreme runoff events with a low model agreement are most likely initiated by compound events (Van der Wiel et al., 2020; Zscheischler et al., 2018). Hence, the timing of extreme high-flow events may depend more on hydrological processes, and consequently vary over hydrological model structure and parameter values in these climate zones. The stacked bar charts indicate which model structures lead to the selection of events with low agreement. Some model structures seem to show deviant behaviour, but there is no convincing pattern visible; most model structures seem to be represented in low-agreement events. Therefore, there is no clear relationship between the extreme runoff events with a low model agreement and specific model structures. We hypothesise that this uncertainty can be assigned to the difference in parameter sets.

To evaluate the timing of extreme low-flow events, a similar approach was applied compared to the high-flow events. However, for several combinations of model structures and parameter sets, the simulations touched upon the hard-coded lower limit in the runoff levels as discussed in the previous section (Figure 5b). For these simulations, it is not possible to select the four most extreme events, which invalidates our method to investigate the impact of different model structures on the timing of low-flow events. The simulated hard-coded lower limit is mostly occurring in the drier climate zones. In general, dry climate conditions lead to lower runoff levels, which more frequently results in models simulating hard-coded lower limits. In the arid climate zone, the runoff levels drop to a hard-coded lower limit in 69 % of all the model simulations. In the cold climate zone, 53 % of the model combinations simulate a hard-coded lower limit. In this climate zone, the temperature regularly drops to below zero (Figure 1), which indicates that precipitation falls as snow instead of rain. This transition affects the runoff-generating processes (Immerzeel et al., 2009), which results in lower runoff levels during colder periods (Figure 1). In the temperate and tropical climate zones, 39 % and 36 %, respectively, of all combinations of model structures and parameter sets simulate a lower limit in the runoff levels. The hard-coded lower limits hamper the analysis of the timing of extreme low flows, but this does show that these hard-coded lower limits, which are probably implemented for numerical stability, are problematic for the

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Discussions

investigation of extreme low-flow events. However, if the hard-coded lower limits indicate a no-flow situation, it would justify that the timing analysis as presented here is not applicable.

3.3 Synthesis

This study evaluates the spread introduced by different hydrological model structures on the magnitude and timing of extreme runoff events. Our results reveal that the spread in magnitude and timing is very much related to each other. If the variation in the magnitude of extreme runoff events is large, there is often also a spread in the timing of these events.

The magnitude and timing of the extreme high-flow events in the arid climate zone are mainly controlled by the meteorological forcing. This is contrary to previous studies in which the runoff in dry catchments was more sensitive to different hydrological models (Jones et al., 2006; Lidén and Harlin, 2000), but here we specifically refer to high-flow events in arid climates. In this climate zone, precipitation is scarce and often characterised by extremely variable, high-intensity and short-duration events (Goodrich et al., 1995). Consequently, runoff in arid climate zones is characterised by a dominance of Hortonian overland flow (Segond et al., 2007). This runoff-generating process is not included in the implementation of FUSE, which might reduce the impact of different model structures (Clark et al., 2008). There is more spread in the magnitudes of low-flow events. Alterations in multiple hydrological processes result in significant differences. Besides that, the hard-coded lower limits also affect the spread in the simulated low-flow events.

In the cold and temperate climate zones, there is more spread in the simulations regarding the magnitude and timing of extreme runoff events. The magnitudes of extreme high- and low-flow events are sensitive to alterations in multiple hydrological process formulations, which implies that several hydrological processes are important in the runoff-generating processes in these climate zones, as also discussed by Scherrer and Naef (2003). In different model simulations different high-flow events are identified as most extreme runoff events, which leads to a spread in the timing of these events. This spread is partly assigned to the difference in parameter sets.

In the tropical climate zone, the spread in the magnitude and timing of extreme runoff events is small, which indicates that the extreme events are mainly controlled by the meteorological forcing. There is only one process formulation that simulates a significant impact on the magnitude; percolation and the upper layer architecture for the high- and low-flow events, respectively. The formulation of the percolation process controls the high-flow events in the tropical climate zone, as there are months with large amounts of precipitation (Figure 1). Due to these large amounts of precipitation, water is subjected to percolation through the succeeding layer (Bethune et al., 2008; Savabi and Williams, 1989). The role of the upper layer architecture in the simulation of low-flow events might be related to evaporation dynamics - although the evaporation formulation has less significant impact ($0.05<p<0.1$).

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

We found no distinct relationship between the length of return periods and the degree of uncertainty in the magnitude of extreme runoff events. There are situations in which the difference between related distributions of high-flow events become significant when the length of the return period increases, e.g. the percolation process formulation in the arid climate zone. On the other hand, there are distributions of related model structures that are significantly different at shorter return periods, e.g. the evaporation process formulation in the temperate climate zone. This contrast might be explained by the difference in importance of specific hydrological processes or parameters for events at different return periods.

4 Discussion

Calibration. We designed a synthetic experiment to conduct controlled experiments on the role of model structure on the simulation of extreme runoff events. There are, however, a few implications when using a synthetic approach. In this study, the models were not calibrated in order to isolate the impact of different model structures. However, it is common practice to use a pre-defined model structure, which is fitted to the local circumstances via parameter calibration (McMillan et al., 2011b). In this study the complete parameter range was sampled: all combinations of parameter values were considered equally plausible and interdependence of parameters was not considered since we used the Latin Hypercube Sampling approach (Clarke, 1973; Helton and Davis, 2003). By implementing more detailed information based on the local circumstances of the different climate zones, the width of these parameter ranges could have been reduced. Smaller parameter ranges would probably lead to more realistic runoff values (Cooper et al., 2007), and might have revealed a relatively higher impact of model process formulation on model results. However, when calibrating hydrological models to simulate extreme runoff events, other challenges remain. Especially the limited availability of historical observations can create a problem for the reliable calibration of extreme events (Wagener et al., 2010); since many observation records do not exceed a length of 50 years, models are forced to simulate outside of their calibration range. This will negatively influences model performance, as for instance demonstrated by Imrie et al. (2000).

Meteorological forcing data. The 2,000-year meteorological time series used in this study originally consists of a simulated large ensemble of 400 sets of 5-year runs. These 400 sets were concatenated artificially. This concatenation might lead to strange transitions of meteorological conditions once every 5 years, as the December month is followed by the next January month of a new 5-year set. Nevertheless, we decided to treat this large ensemble as a single time series, in order to allow for extensive return period analysis. We consider the effect of the concatenation limited since we only evaluate the annual and monthly maximum and minimum daily runoff levels. The employed time series does not allow for the evaluation of multi-year droughts, despite these events being extremely relevant considering their societal impact. However, the hard-coded lower limits that were found in this study and hampered some of the low-flow analyses, would also pose a serious problem for analysing multi-year droughts, even if realistic forcing time series were used.

Model selection. Model selection is a crucial step in hydrological modelling. Different hydrological models might lead to substantially different outcomes (Melsen et al., 2018). When hydrologists are familiar with a certain model, they tend to stick to this model, even though other models might be more adequate for a specific objective (Addor and Melsen, 2019). Model intercomparison studies can provide guidance for model selection and improve model adequacy in the future. This study evaluates the impact of alterations in model structures on extreme runoff events. Some alterations in the model structure lead to significant impacts in the simulation. For example, in the tropical climate zone, the formulation of the percolation process is important. This information can be regarded in model selection of future studies, which will result in more adequate model selection. On the other hand, it should be noted that the framework employed in this study, FUSE, is only representative for a particular suit of bucket-based models. Whereas these models are suitable for long term simulations due to their low data demand and high computational efficiency, results might look different when a more process-based framework, such as SUMMA (Clark et al., 2015a, b), would have been employed.

Societal impact. This study evaluated the translation of meteorology to hydrological extreme impact events. Return periods were used to sort runoff events based on their extremeness, as return periods are frequently used in policy design (Marco, 1994; Read and Vogel, 2015). However, this study does not translate hydrological impact events to the societal impact, which implies that fatalities and economic losses are not examined. This relationship might be affected by non-linear effects, similar to the meteorology-hydrology relationship (Van der Wiel et al., 2020). Therefore, a direction for future research is to link societal impact to return periods of extreme runoff events. The accurate assessment of vulnerability and societal impact requires information related to exposure and sensitivity (Cardona et al., 2012).

5 Conclusions

Hydrological extremes are natural hazards that affect a large number of people on a global scale. Several hydrological models were employed to simulate these extremes, with the aim to investigate the impact of hydrological model structure on the simulation of extreme runoff events. The combination of two state-of-the-art approaches, the hydrological modular modelling framework FUSE and large ensemble meteorological simulations to study extreme events, provided insights into uncertainties of the simulations. Parameters of the hydrological models were sampled in a synthetic experiment, which enabled the examination of the impact of different hydrological process formulations on the magnitude and timing of extreme high- and low-flow events, independent of calibration.

The impact of hydrological process formulations on magnitude and timing of extreme runoff events varies among different climate zones (Figure 8). In the arid climate zone, the magnitude and timing of the extreme high-flow events are not affected by changing process formulations or parameter sets. The magnitudes of the low-flow events are significantly affected by alterations in the architecture of the upper and lower layer and the process formulation of percolation. In the cold and temperate climate zones, we found a larger spread in the simulations of the extreme runoff events. Multiple hydrological processes sig-

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

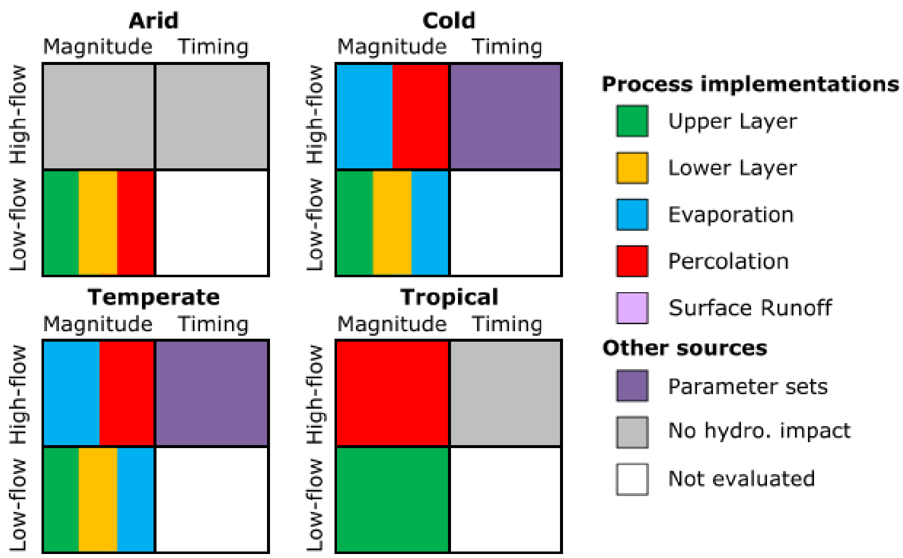

Figure 8. Summary of the results: indicated are the process formulations that significantly affect the distributions of the extreme runoff events in the different climate zones. The process implementations refer to the formulation of hydrological processes in different model structures. "No hydro. impact" indicates that the effect of the hydrological model was limited, which implies that neither alterations in the model structure, nor in the parameter sets significantly affected the simulated extreme runoff events.

nificantly affect the magnitude of the high- and low-flow events, which implies that the model structure is an important source of uncertainty. Therefore, it is essential to select an adequate hydrological model when simulating extreme events in cold and temperate climate zones. Besides that, there is a spread in the timing of high-flow events, caused by different parameter sets in these climate zones. The magnitudes of the high- and low-flow events in the tropical climate zone are affected by the formulation of percolation and upper layer, respectively. The timing of these events is hardly affected by hydrological model structure or parameter sets, which implies that the timing of these events is dictated by the meteorological forcing. The timing of low-flow events is not evaluated in this study, as the analysis was hampered by hard-coded lower limits.

The results revealed a spread in the simulation of extreme runoff events as a consequence of different hydrological model structures. The impact of different model structures is larger for the simulation of low-flow events compared to high-flow events. For the low-flow events hard-coded lower limits were found, implemented for numerical stability. This revealed the numerical challenge that comes with simulating extremely low values. The extreme events were assessed at different return periods. However, no clear relationship was found between the model structural uncertainty in the magnitude of extreme runoff events and the return period length.

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Discussions

Insights provided by this study contribute to a better understanding of the importance of the hydrological model formulation of specific processes in different climate zones. These insights can be used in future studies, which will result in more adequate model selection leading to more reliable predictions of extreme runoff events.

Code and data availability. All codes to process the data (R-code) and the results themselves are available upon request from the corresponding author. The meteorological forcing and all simulated runoff data of the four evaluated climate zones will be published online (4TU), upon acceptance of the manuscript.

Author contributions. KW and LM designed the study in consultation with GK. KW provided the meteorological forcing data, which GK employed to carry out the hydrological simulations and analyses. GK wrote the manuscript with support from KW and LM.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. KW acknowledges the HiWAVES3 project, funded by the Netherlands Organisation of Scientific Research (NWO) under Grant Number ALWCL.2 016.2.

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

References

Addor, N. and Melsen, L. A.: Legacy, rather than adequacy, drives the selection of hydrological models., Water Resour. Res., 55, 378–390, https://doi.org/10.1029/2018WR022958, 2019.

Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resour. Res., 43, https://doi.org/10.1029/2005WR004745, 2007.

Andersen, M. N., Jensen, C. R., and Lösch, R.: The interaction effects of potassium and drought in field-grown barley. I. Yield, water-use efficiency and growth., Acta Agriculturae Scandinavica B., 42, 34–44, https://doi.org/10.1080/09064719209410197, 1992.

Atkinson, S. E., Woods, R. A., and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resour. Res., 38, 50–1–50–17, https://doi.org/10.1029/2002wr001487, 2002.

Bethune, M. G., Selle, B., and Wang, Q. J.: Understanding and predicting deep percolation under surface irrigation., Water Resour. Res., 44, https://doi.org/10.1029/2007WR006380, 2008.

Beven, K. and Freer, J.: A dynamic topmodel., Hydrol. Process., 15, 1993–2011, https://doi.org/10.1002/hyp.252, 2001.

Brauer, C., Teuling, A., Overeem1, A., van der Velde, Y., Hazenberg1, P., Warmerdam, P., and Uijlenhoet, R.: Anatomy of extraordinary rainfall and flash flood in a Dutch lowland catchment, Hydrol. Earth Syst. Sci., 15, 1991–2005, https://doi.org/10.5194/hess-15-1991-2011, 2011.

Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system: Conceptual modeling for digital computers., US Department of Commerce, National Weather Service, and State of California, Department of Water Resources., 1973.

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, J. Hydrol., 298, 242–266, https://doi.org/10.1016/j.jhydrol.2004.03.042, 2004.

Cardona, O. D., Van Aalst, M. K., Birkmann, J., Fordham, M., Mc Gregor, G., Perez, R., Pulwarty, R. S., Schipper, E. L. S., and Sinh, B. T.: Determinants of risk: exposure and vulnerability. In Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change., Cambridge University Press, 2012.

Clark, M., Nijssen, B., Lundquist, J., Kavetski, D., Rupp, D., Woods, R., Freer, J., Gutmann, E., Wood, A., Brekke, L. D., Arnold, J., Gochis, D., and Rasmussen, R.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498–2514, https://doi.org/10.1002/2015WR017198, 2015a.

Clark, M., Nijssen, B., Lundquist, J., Kavetski, D., Rupp, D., Woods, R., Freer, J., Gutmann, E., Wood, A., Gochis, D., Rasmussen, R., Tarboton, D., Mahat, V., Flerchinger, G., and Marks, D.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, Water Resour. Res., 51, 2515–2542, https://doi.org/10.1002/2015WR017200, 2015b.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, https://doi.org/10.1029/2007WR006735, 2008.

Clarke, R. T.: A review of some mathematical models used in hydrology, with observations on their calibration and use, J. Hydrol., pp. 1–20, https://doi.org/10.1016/0022-1694(73)90089-9, 1973.

Cooper, V. A., Nguyen, V. T. V., and Nicell, J. A.: Calibration of conceptual rainfall–runoff models using global optimisation methods with hydrologic process-based parameter constraints., J. Hydrol., 334, 455–466, https://doi.org/10.1016/j.jhydrol.2006.10.036, 2007.

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

Di Baldassarre, G., Montanari, A., Lins, H., Koutsoyiannis, D., Brandimarte, L., and Blöschl, G.: Flood fatalities in Africa: from diagnosis to mitigation., Geophys. Res. Lett., 37, https://doi.org/10.1029/2010GL045467, 2010.

Di Baldassarre, G., Martinez, F., Kalantari, Z., and Viglione, A.: Drought and flood in the Anthropocene: feedback mechanisms in reservoir operation., Earth Syst. Dynam., 8, 1–9, https://doi.org/10.5194/esd-8-225-2017, 2017.

Eagleson, P. S.: The emergence of global-scale hydrology., Water Resour. Res., 22, 6S–14S, https://doi.org/10.1029/WR022i09Sp0006S, 1986.

Engeland, K., Hisdal, H., and Frigessi, A.: Practical extreme value modelling of hydrological floods and droughts: a case study., Extremes, 7, 5–30, https://doi.org/10.1007/s10687-004-4727-5, 2004.

Goodrich, D. C., Faurés, J.-M., Woolhiser, D. A., Lane, L. J., and Sorooshian, S.: Measurement and analysis of small-scale convective storm rainfall variability, J. Hydrol., 173, 283–308, https://doi.org/10.1016/0022-1694(95)02703-R, 1995.

Gumbel, E. J.: The return period of flood flows., The annals of mathematical statistics, 12, 163–190, https://doi.org/10.2307/2235766, 1941.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., Van Noije, T., Van der Linden, E., and Van der Wiel, K.: EC-Earth V2. 2: description and validation of a new seamless earth system prediction model., Clim. Dynam., 39, 2611–2629, https://doi.org/10.1007/s00382-011-1228-5, 2012.

Helton, J. C. and Davis, F. J.: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems., Reliab. Eng. Syst. Safe, 81, 23–69, https://doi.org/10.1016/S0951-8320(03)00058-9, 2003.

Henderson-Sellers, A., Yang, Z. L., and Dickinson, R. E.: The project for intercomparison of land-surface parameterization schemes., B. Am. Meteorol. Soc., 74, 1335–1350, https://doi.org/10.1175/1520-0477(1993)074<1335:TPFIOL>2.0.CO;2, 1993.

Immerzeel, W. W., Droogers, P., De Jong, S. M., and Bierkens, M. F. P.: Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing., Remote Sens. Environ., 113, 40–49, https://doi.org/10.1016/j.rse.2008.08.010, 2009.

Imrie, C. E., Durucan, S., and Korre, A.: River flow prediction using artificial neural networks: generalisation beyond the calibration range., J. Hydrol., 233, 138–153, https://doi.org/10.1016/S0022-1694(00)00228-6, 2000.

Jones, R. N., Chiew, F. H., Boughton, W. C., and Zhang, L.: Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models, Adv. Wat. Res., 29, 1419–1429, https://doi.org/10.1016/j.advwatres.2005.11.001, 2006.

Knoben, W. J., Freer, J. E., Fowler, K. J., Peel, M. C., and Woods, R. A.: Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1. 2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations., Geosci. Model Dev., 12, 2463–2480, https://doi.org/10.5194/gmd-12-2463-2019, 2019.

Knutsson, G.: Humid and arid zone groundwater recharge—a comparative analysis., Estimation of Natural Groundwater Recharge, 113, 40–49, https://doi.org/10.1007/978-94-015-7780-9_32, 1988.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen-Geiger climate classification updated., Meteorol. Z., 15, 259–263, https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Kustas, W. P., Rango, A., and Uijlenhoet, R.: A simple energy budget algorithm for the snowmelt runoff model, Water Resour. Res., 30, 1515–1527, https://doi.org/10.1029/94wr00152, 1994.

Laio, F., Di Baldassarre, G., and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes., Water Resour. Res., 45, https://doi.org/10.1029/2007WR006666, 2009.

Leavesley, G. H.: Precipitation-runoff modeling system: User's manual., vol. 83, US Department of the Interior, 1984.

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Li, Z., Shao, Q., Xu, Z., and Cai, X.: Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China., J. Hydrol., 385, 76–83, https://doi.org/10.1016/j.jhydrol.2010.01.025, 2010.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models., J. Geophys. Res-Atmos., 99, 14 415–14 428, https://doi.org/10.1029/94JD00483, 1994.

Lidén, R. and Harlin, J.: Analysis of conceptual rainfall–runoff modelling performance in different climates, J. Hydrol., 238, 231–247, https://doi.org/10.1016/s0022-1694(00)00330-9, 2000.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework., Water Resour. Res., 43, https://doi.org/10.1029/2006WR005756, 2007.

Marco, J.: Flood risk mapping, in: Coping with Floods, pp. 353–373, Springer Netherlands, https://doi.org/10.1007/978-94-011-1098-3_20, 1994.

Massey Jr, F. J.: The Kolmogorov-Smirnov test for goodness of fit., J. Am. Stat. Assoc., 46, 68–78, https://doi.org/10.1080/01621459.1951.10500769, 1951.

McMichael, A. J. and, C.-L. D., Kovats, S., Edwards, S., Wilkinson, P., Wilson, T., R., N., Hales, S., Tanser, F., Le Sueur, D., Schlesinger, M., and Andronova, N.: Comparative Quantification of Health Risks: Chapter 20 Global climate change, World Health Organization Geneva, 2004.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Input uncertainty in hydrological models: an evaluation of error models for rainfall., J. Hydrol., 1-2, 83–94, 2011a.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure., Hydrol. Process., 25, 511–522, https://doi.org/10.1002/hyp.7841, 2011b.

Meigh, J. R., Farquharson, F. A. K., and Sutcliffe, J. V.: A worldwide comparison of regional flood estimation methods and climate., Hydrolog. Sci. J., 42, 225–244, https://doi.org/10.1080/02626669709492022, 1997.

Melsen, L., Teuling, A., Torfs, P., Uijlenhoet, R., Mizukami, N., and Clark, M.: HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models, Hydrol. Earth Syst. Sci., 20, 1069–1079, https://doi.org/10.5194/hess-20-1069-2016, 2016.

Melsen, L., Addor, N., Mizukami, N., Newman, A., Torfs, P., Clark, M., Uijlenhoet, R., and Teuling, A.: Mapping (dis)agreement in hydrologic projections, Hydrol. Earth Syst. Sci., 22, 1775–1791, https://doi.org/10.5194/hess-22-1775-2018, 2018.

Michele, C. D. and Rosso, R.: Uncertainty assessment of regionalized flood frequency estimates., J Hydrol. Eng., 6, 453–459, https://doi.org/10.1061/(ASCE)1084-0699(2001)6:6(453), 2001.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set., J. Geophys. Res-Atmos, 117, https://doi.org/10.1029/2011JD017187, 2012.

Morin, E., Enzel, Y., Shamir, U., and Garti, R.: The characteristic time scale for basin hydrological response using radar data., J. Hydrol., 252, 85–99, https://doi.org/10.1016/S0022-1694(01)00451-6, 2001.

Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, Science, 263, 641–646, https://doi.org/10.1126/science.263.5147.641, 1994.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, 2007.

Pilgrim, D. H.: Some problems in transferring hydrological relationships between small and large drainage basins and between regions., J. Hydrol., 65, 49–72, https://doi.org/10.1016/0022-1694(83)90210-X, 1983.

Read, L. K. and Vogel, R. M.: Reliability, return periods, and risk under nonstationarity, Water Resour. Res., 51, 6381–6398, https://doi.org/10.1002/2015wr017089, 2015.

Reaney, S. M., Bracken, L. J., and Kirkby, M. J.: The importance of surface controls on overland flow connectivity in semi-arid environments: Results from a numerical experimental approach., Hydrol. Process, 28, 2116–2128, https://doi.org/10.1002/hyp.9769, 2014.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D. J., and Participants, D.: Overall distributed model intercomparison project results., J. Hydrol., 298, 27–60, https://doi.org/10.1016/j.jhydrol.2004.03.031, 2004.

Salas, J. D. and Obeysekera, J.: Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events., J. Hydrol. Eng., 19, 554–568, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820, 2014.

Savabi, M. and Williams, J.: Chapter 7. Water balance and percolation, USDA–Water Erosion Prediction Project: Hillslope profile model documentation., 1989.

Scherrer, S. and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland., Hydrol. Process., pp. 391–401, https://doi.org/10.1002/hyp.1131, 2003.

Segond, M. L., Wheater, H. S., and Onof, C.: The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, J. Hydrol., 347, 116–131, https://doi.org/10.1016/j.jhydrol.2007.09.040, 2007.

Smith, A., Sampson, C., and Bates, P.: Regional flood frequency analysis at the global scale., Water Resour. Res., 51, 539–553, https://doi.org/10.1002/2014WR015814, 2015.

Sousa, P. M., Trigo, R. M., Aizpurua, P., Nieto, R., Gimeno, L., and García Herrera, R.: Trends and extremes of drought indices throughout the 20th century in the Mediterranean., Nat. Hazard Earth Sys., 11, 33–51, https://doi.org/10.5194/nhess-11-33-2011, 2011.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., and Tallaksen, L. M.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrol. and Earth Syst. Sci., 15, 3447–3459, https://doi.org/10.5194/hess-15-3447-2011, 2011.

Van der Wiel, K., Wanders, N., Selten, F. M., and Bierkens, M. F. P.: Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2° C Warmer World., Geophys. Res. Lett., 46, 2093–2102, https://doi.org/10.1029/2019GL081967, 2019.

Van der Wiel, K., Selten, F. M., Bintanja, R., Blackport, R., and Screen, J.: Ensemble climate-impact modelling: extreme impacts from moderate meteorological conditions, Environ. Res., https://doi.org/10.1088/1748-9326/ab7668, 2020.

Van Loon, A. F.: Hydrological drought explained., Wiley Interdisciplinary Reviews: Water, 2, 359–392, https://doi.org/10.1002/wat2.1085, 2015.

Vitolo, C., Wells, P., and Dobias, M.: R-Fuse: First Official Release, https://doi.org/10.5281/ZENODO.14005, 2015.

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Suresh, C. R., Basu, N. B., and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, Water Resour. Res., 46, https://doi.org/10.1029/2009WR008906, 2010.

Winsemius, H. C., Aerts, J. C., Van Beek, L. P., Bierkens, M. F., Bouwman, A., Jongman, B., Kwadijk, J., Ligtvoet, W., Lucas, P. L., van Vuuren, D. P., and Ward, P. J.: Global drivers of future river flood risk., Nat. Clim. Change, 6, 381–385, https://doi.org/10.1038/nclimate2893, 2016.

Zhang, X., Hörmann, G., Gao, J., and Fohrer, N.: Structural uncertainty assessment in a discharge simulation model., Hydrolog. Sci. J., 56, 854–869, https://doi.org/10.1080/02626667.2011.587426, 2011.

https://doi.org/10.5194/nhess-2020-154
Preprint. Discussion started: 14 July 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

Zotarelli, L., Dukes, M. D., Romero, C. C., Migliaccio, K. W., and Morgan, K. T.: Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method)., Environmental Science, 2015.

Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events., Nat. Clim. Chang., 8, 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.