Climate change impacts model parameter sensitivity - implications for calibration strategy and model diagnostic evaluation

Lieke Anna Melsen[1] and Björn Guse[2,3]

[1]Hydrology and Quantitative Water Managemement, Wageningen University, Wageningen, the Netherlands
[2]GFZ German Research Centre for Geosciences, Section Hydrology, Potsdam, Germany
[3]Christian-Albrechts-University of Kiel, Institute of Natural Resource Conservation, Department of Hydrology and Water Resources Management, Kiel, Germany

Correspondence: Lieke Melsen (lieke.melsen@wur.nl)

Abstract. Hydrological models are useful tools to explore the impact of climate change. To prioritize parameters for calibration and to evaluate hydrological model functioning, sensitivity analysis can be conducted. Parameter sensitivity, however, varies over climate, and therefore climate change could influence parameter sensitivity. In this study we explore the change in parameter sensitivity within a plausible climate change rate. We investigate if changes in sensitivity propagate into the calibration strategy, and compare three hydrological models diagnostically based on the sensitivity results. We employed three frequently used hydrological models (SAC, VIC, and HBV), and explored parameter sensitivity changes across 605 catchments in the United States by comparing a GCM-forced historical and future period. Consistent among all models is that the sensitivity of snow parameters decreases in the future, and that evapotranspiration parameters have a tendency to increase. Which other parameters increase in sensitivity is less consistent among the models. In 43% to 49% of the catchments, dependent on the model, at least one parameter changes in the future in the top-5 most sensitive parameters. The maximum number of changes in the parameter top-5 is two, in 2-4% of the investigated catchments. This requires an adapted calibration strategy for long-term projections, for which we provide several suggestions. The disagreement among the models on the processes that become more relevant in future projections also calls for a strict evaluation of the adequacy of the model structure for long-term simulations.

1 Introduction

Earth and environmental computer models are indispensable tools to explore an uncertain future. Whereas observational studies report on historical changes in streamflow patterns across the contiguous United States (CONUS) that might be attributed to climate change (Stewart et al., 2005; Sagarika et al., 2014), hydrological models are applied in the same region to gain insights into long term changes in the future (e.g. Mizukami et al., 2016; Melsen et al., 2018; Chegwidden et al., 2019; Brunner et al., 2020). These model projections can support water resource managers to prepare for future changes. The models are thus related to costly and impactful decisions (McMillan et al., 2017; Metin et al., 2018).

Given the relevant role of models to support decision making, model functioning should be thoroughly scrutinized. A frequently used tool to evaluate hydrological model functioning is sensitivity analysis (Pianosi et al., 2016; Devak and Dhanya, 2017). Sensitivity analysis is aimed at identifying the relative impact of model parameters on model response. The results of a sensitivity analysis differs over models, target variable of the model response, and applied sensitivity analysis methods (Shin et al., 2013; Razavi and Gupta, 2015; Guse et al., 2016a; Haghnegahdar et al., 2017; Mai and Tolson, 2019).

However, parameter sensitivity also differs across climate, as for instance showed by Demaria et al. (2007), Van Werkhoven et al. (2008) and Melsen and Guse (2019): In a cold catchment with a large fraction of the precipitation falling as snow, snow parameters are supposed to be sensitive, while in a tropical catchment without snowfall, snow parameters are not supposed to show any sensitivity. As such, it is common understanding that parameter sensitivity depends on climate. But reconsidering this fact, this would also imply that parameter sensitivity could change in a changing climate. Therefore, the question is whether, within a plausible rate of climate change, hydrological parameter sensitivity changes. This could have consequences for the way hydrological models should be calibrated for long term projections. Besides, it offers the opportunity to compare different models to evaluate if the same mechanisms are simulated as being relevant for future changes.

Many hydrological models that are used for long term projections have parameters that require calibration to identify their values for the catchment under study. Hydrological model parameters are generally calibrated on discharge time series. But discharge is a lumped catchment response, and therefore only provides limited catchment information (Jakeman and Hornberger, 1993; Guse et al., 2016a). A rule of thumb suggested by Beven (1989) and employed by many modelers is that, given the limited information available in a discharge time series, three to five parameters can be calibrated based on these data (and this might already be a high number, following the famous quote of von Neumann[1]). Global Sensitivity Analysis (GSA) can be employed to identify the parameters that have most influence on the model output (Demaria et al., 2007; van Werkhoven et al., 2009; Pianosi et al., 2016; Borgonovo et al., 2017; Zadeh et al., 2017). Subsequently, the three to five parameters that show most sensitivity are selected for calibration. However, if parameter sensitivity changes with climate change, this could interfere with the parameter prioritization procedure for models used for long term projections.

Besides the potential consequence for calibration, evaluating the relation between change in parameter sensitivity and climate change also provides the opportunity to diagnostically evaluate model functioning during long-term projections. Several studies already investigated the change in parameter sensitivity over time, focusing on specific events or relatively short time scales (Reusser et al., 2011; Herman et al., 2013b; Guse et al., 2014; Massmann and Holzmann, 2015; Pfannerstill et al., 2015). For instance, it was demonstrated that certain parameters are triggered during specific discharge conditions such as high or low flow events (Pfannerstill et al., 2015; Guse et al., 2016b). If certain events will prevail or become less frequent in a future climate, this might change the average parameter sensitivity over the long run. As such, evaluating long-term changes in parameter sensitivity can provide insights into systemic changes. By comparing changes among several model structures, the robustness of simulated systemic changes can be evaluated.

[1] *"With four parameters I can fit an elephant, with five I can make him wiggle his trunk"*, John von Neumann (1903-1957)

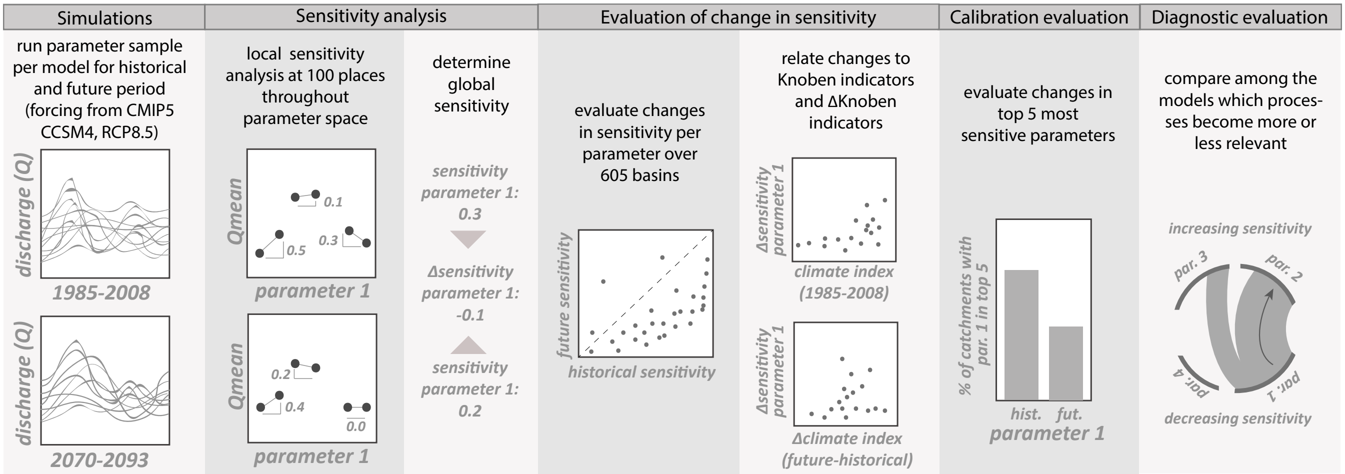

Figure 1. Summary of the methodological approach, to be read from left to right. The first three panels are the calculations, the other four panels are the actual analyses.

In this study we investigate how parameter sensitivity changes as a consequence of climate change. We evaluate if and how this has consequences for parameter prioritization for calibration, and if systemic changes are robust across different model structures. To this end, we apply a hybrid local-global sensitivity analysis method to three frequently used hydrological models in 605 basins across the US, and link changes in sensitivity to changes in climate. Finally, we evaluate the impact on the top-5 most sensitive parameters in each basin, and investigate the transmission of sensitivity from one parameter to the other. We end with a recommendation on how to account for changes in sensitivity in the calibration strategy of models used for long-term projections, and an evaluation of the robustness of systemic changes across different models.

2 Methods

To investigate changes in parameter sensitivity, we employed three frequently used hydrological models. The models were run for a historical and a future period over 605 catchments, forced with a bias corrected and statistically downscaled global circulation model. A hybrid local-global sensitivity analysis method was applied to the simulations of both periods. Then, the differences in parameter sensitivity between the historical and future period were explored in several ways. First, per parameter to investigate which parameters change, and over different climate indicators to investigate how climate and climate change explain changes in sensitivity. Then, we assessed how the top-5 most sensitive parameters would change in the future period, thereby impacting the calibration strategy. Finally, we conduct a diagnostic model evaluation, amongst others by investigating the transmission of sensitivity from one parameter to the other. An overview of the procedure is shown in Figure 1.

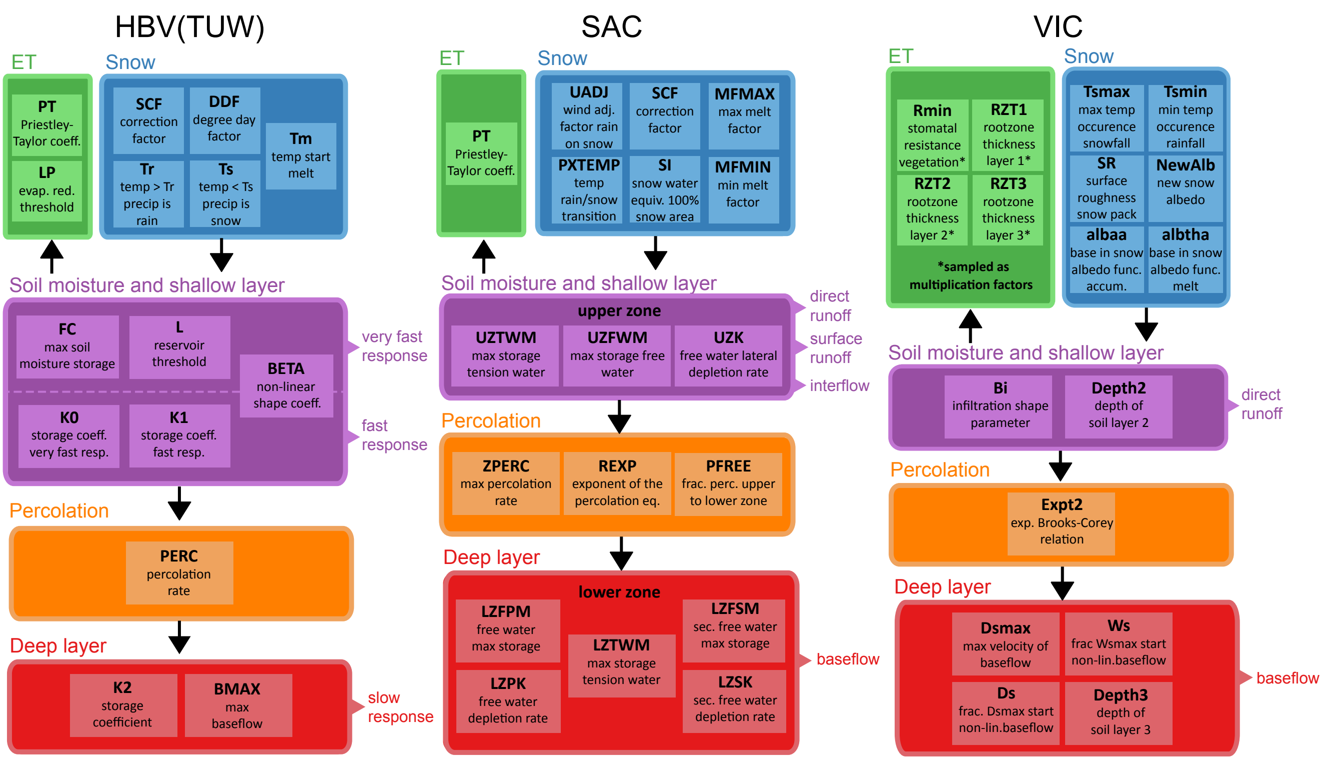

Figure 2. Simplified representation of the model structure of the three models employed in this study. All the parameters that are displayed are included in the sensitivity analysis. Parameters are colored according to the flux or state they influence (evapotranspiration (ET), snow, soil moisture and shallow layer, percolation, deep layer). The colors are used consistently throughout all the figures in this study. Parameter boundaries can be found in Appendix Tables A1, A2, and A3.

2.1 Models

We investigated for three models whether parameter sensitivity changes within a plausible climate change range: the TUW-model following the structure of HBV (Parajka et al., 2007, hereafter referred to as HBV), SAC-SMA combined with SNOW-17 (Newman et al., 2015), and VIC 4.1.2h (Liang et al., 1994). All three models have previously been used for long-term climate impact projections: e.g. Teutschbein et al. (2011); Wetterhall et al. (2011) for HBV; Koutroulis et al. (2013); Peleg et al. (2015) for SAC-SMA; and Christensen et al. (2004); Wu et al. (2012) for VIC, and are therefore relevant models to consider. The same suit of models was explored in another context in Melsen et al. (2018) and Melsen and Guse (2019).

A simplified representation of the model structures, including a description of the parameters that were accounted for in the sensitivity analysis, are displayed in Figure 2. A more elaborate description of each model can be found in Melsen et al. (2018), and in the respective references of each models: Bergström (1976, 1992) for HBV, Burnash et al. (1973); National Weather Service (2002) for SAC-SMA, and Liang et al. (1994, 1996) for VIC.

2.2 Catchments and Forcing

All three models were run for a historical and future period of 28 years, of which the first five years were omitted from both periods for spin up. As such, the historical period that is analyzed covers 1985-2008, and the future period 2070-2093, 23 years each. The forcing for both periods was obtained from statistically downscaled and bias corrected output from the Community Climate System Model 4.0 (CCSM4, Gent et al., 2011), from Climate Model Intercomparison CMIP5, using Representative Concentration Pathway 8.5 (RCP8.5). Bias correction was done according to the Bias Correction and Spatial Disaggregation (BCSD) method of Wood et al. (2004), based on the Maurer et al. (2002) forcing data.

We consider our study an investigation of the potential that a plausible climate change rate might impact hydrological model parameter sensitivity. As such, we decided to use one climate model only, and conduct the analysis for several hydrological models. CCSM4 is among the better performing climate models when evaluated against observed precipitation and temperature (Knutti et al., 2013; Sheffield et al., 2013) and is therefore selected as providing a 'plausible' projection, but other well performing climate models might still have quite different dynamics (Sheffield et al., 2013).

Finally, the GCM forcing was lumped over the CAMELS basins. The CAMELS data set contains forcing, discharge observations, and catchment characteristics for 671 catchments throughout the contiguous United States with limited human impact (Newman et al., 2014, 2015; Addor et al., 2017). We employed a subset of 605 catchment-averaged forcings, because at the time of calculation, there were still issues with determining the exact catchment area for the remaining 66 catchments.

The hydrological models were not calibrated, since we employ global sensitivity analysis across the full parameter range. Therefore, the 605 catchments should be perceived as 605 different climate instances with an individual level of climate change, rather than as catchment representative models. The models are, however, able to achieve acceptable model performance in these basins when forced with observations and confronted with discharge observations (Melsen et al., 2018), providing credibility to the models to be used in this context.

The goal of this study is to investigate how climate might impact parameter sensitivity within a plausible climate change range. As such, it is of second order importance whether the climate model gives highly accurate predictions or whether the hydrological model can exactly capture catchment behavior. It is, however, important to note that we employed the highest emission scenario (RCP8.5), thereby investigating the effect of the higher ranges of plausible climate change. It can be expected that the impact of climate change on parameter sensitivity will be lower for lower emission scenarios. However, RCP 8.5 is often used to provide an upper boundary for long-term projections, thereby demonstrating the relevance of choosing this scenario.

2.3 Sensitivity analysis methodology

In the selection of the sensitivity analysis method, a few points were considered. First, it had to be a global method, because global sensitivity analysis methods are used to identify the most sensitive parameters for calibration (whereas local methods are generally applied after calibration). Secondly, we had to account for a high number of runs (605 basins, two periods, three models). Therefore, we selected the hybrid local-global method DELSA (Rakovec et al., 2014), which is computationally cheaper than traditional (variance-based) global sensitivity analysis methods.

DELSA evaluates local sensitivity at several places throughout parameter space, as such mimicking global sensitivity analysis. First, 100 parameter samples called base-runs were created based on a space filling sampling strategy. The models were run for all 100 samples. Secondly, the parameters are one-at-a-time perturbed with 1% compared to their base-run value. The effect of this perturbation on the model output, compared to the corresponding base-run, represents the parameter sensitivity:

$$\left.\frac{\partial\psi}{\partial\theta_k}\right|_{\theta_{base-run_i}} = \frac{\psi_{base-run_i} - \psi_{perturbed_i}}{\theta_{base-run_i,k} - \theta_{base-run_i,k}\cdot 1.01}, \quad (1)$$

where ψ denotes the model output that is evaluated, θ refers to the parameter value, k the number of parameters that is evaluated in the sensitivity analysis (in our case 15, 18, and 17 for HBV, SAC, and VIC, respectively), and i the number of base runs (in our case 100). We used the average sensitivity from the 100 samples per parameter per basin as a measure of parameter sensitivity. Each parameter that is displayed in Figure 2 was accounted for in the sensitivity analysis. The applied parameter boundaries for sampling are provided in Appendix Tables A1, A2 and A3 (see also Melsen and Guse, 2019).

Besides the selection of a sensitivity analysis method (which will influence the final results, Razavi and Gupta, 2015; Pianosi et al., 2016), we also had to identify a target variable - the variable that is compared between the base-run and the perturbed run. Whereas performance metrics are quite popular as target variable (Van Werkhoven et al., 2008; Herman et al., 2013a), they are not well suited for global sensitivity analysis (Razavi and Gupta, 2015; Guse et al., 2016a), and besides, it is not possible to obtain model performance for the future. Therefore, this study focuses on mean simulated discharge as target variable. Many other streamflow signatures could have been of interest to evaluate, for instance related to high and low flows, but given the goal of this study, an exploration on the effect of climate change on parameter sensitivity, mean discharge seems the most neutral choice.

The sensitivity analysis was conducted for both the historical and future period, for all 605 basins. The first analysis of the calculations was a simple exploration of which parameters increase and which parameters decrease in sensitivity in the future over all 605 basins to achieve a first insights into potential changes in parameter sensitivity in future and to see which parameters are mainly affected.

2.4 Climate indicators to relate changes in sensitivity

The 605 climate instances from the 605 basins are not a representative sample since certain climates might be over- or under-represented. Therefore, the difference in sensitivity was also related to climate indicators. Given their relevance for discharge, we used the Knoben climate indicators (Knoben et al., 2018) to classify the changes in parameter sensitivity. The Knoben indicators consist of three indicators: aridity index, seasonality, and fraction precipitation falling as snow.

To determine the aridity index, first Thorntwaite's Moisture Index (MI, Knoben et al., 2018; Willmott and Fedema, 1992) is obtained based on mean monthly observations of precipitation, $P(t)$, and evapotranspiration $E_p(t)$. Subsequently, average aridity I_m can be obtained.

$$MI(t) = \begin{cases} 1 - \frac{E_p(t)}{P(t)}, & \text{if} P(t) > E_p(t) \\ 0, & \text{if } P(t) = E_p(t) \\ \frac{P(t)}{E_p(t)} - 1, & \text{if } P(t) < E_p(t) \end{cases} \tag{2}$$

$$I_m = \frac{1}{12} \sum_{t=1}^{t=12} MI(t) \tag{3}$$

Aridity index I_m varies between -1, representing highly arid conditions, and 1, representing humid conditions. The seasonality in the aridity index, $I_{m,r}$, is determined based on the maximum difference in MI over the year:

$$I_{m,r} = \max(MI(1,2,...12)) - \min(MI(1,2,...12)) \tag{4}$$

The seasonality varies between 0 and 2, with 0 indicating no intra-annual variation, and 2 indicating that climate varies from fully arid to fully humid over the year. The last Knoben index is the fraction precipitation falling as snow, f_s.

$$f_s = \frac{\sum P(T(t) \leq T_0)}{\sum_{t=1}^{t=12} P(t)} \tag{5}$$

In this equation, $T(t)$ is the mean monthly temperature, and T_0 the threshold temperature below which precipitation is assumed to occur as snow. The threshold temperature was set to 0°C, in line with Knoben et al. (2018). f_s can have a value between 0, no snow, and 1, all precipitation falling as snow. All three indicators were evaluated based on the climate in the historical period, and based on the change in climate between the future and the historical period (future - historical), this latter one being indicated as Δindicator.

2.5 Evaluation of impact of sensitivity changes on calibration strategy

To evaluate the impact of change in parameter sensitivity on calibration strategy, we determined the top-5 most sensitive parameters for each basin, both for the historical and future period. We analyzed which parameters left and entered the top-5 in the future compared to the historical period, as a consequence of a change in sensitivity. This was again related to the climate indicators of Sect. 2.4, to investigate if in certain climates or climate change rates more changes can be expected in the calibration parameters.

2.6 Diagnostic model evaluation based on changes in sensitivity

To diagnose how the results from the different models have come about, we relate the direct model output (several states and fluxes) to changes in sensitivity. Furthermore, we introduce the concept of 'parameter sensitivity transmission': We evaluate whether any negative correlations exist between parameters with increasing and decreasing sensitivity. Strong negative correlations can be an indication that sensitivity is transmitted from one parameter to the other, so we define transmission as a clear negative correlation in change in sensitivity between two parameters. However, since we evaluate correlation, transmission does not refer to absolute sensitivity values.

The goal of this analysis is to investigate to what extent sensitivity is transmitted directly from the decreasing parameter to the increasing parameter. When there is no direct relation, it can indicate that sensitivity changes at several places within the model structure. The transmission of sensitivity can give insights into which processes become more relevant in the future, at the expense of processes that become less relevant - a systemic change as a result of climate change. A comparison among the different model structures will indicate their (dis)agreement on the change in relevant processes.

3 Results

To place the results into context, we first briefly discuss the change in climate between the historical and future period, and the changes in several simulated water balance terms for both periods for the three employed hydrological models. Subsequently, we discuss the change in sensitivity between the historical and future period, the relation between changes in sensitivity and climate, the impact of sensitivity changes on calibration strategy, and finally the model diagnostic evaluation.

3.1 Changes in climate and simulated water balance terms between historical and future period

Fig. 3 provides an overview on the change in climate (expressed in temperature and precipitation) between the historical and future period using CCSM4 and RCP8.5. In all the investigated basins the mean temperature will increase in the future with at least 2.6°C. The median temperature increase across the 605 basins is 4.1°C, aligning with the most extreme emission scenario that was employed to run the climate model. Most of the investigated basins will receive more precipitation in the future.

Fig. 3 also depicts several simulated water balance terms for the three employed hydrological models. Note that the models were not calibrated and that for each of the 605 basins, the mean across the parameter ensemble was used for this figure. It is therefore not an indication of what exactly might happen in the future in the investigated basins, but an indication of the flexibility of the models in responding to changes in climate. There is consistency among the models that in most basins,

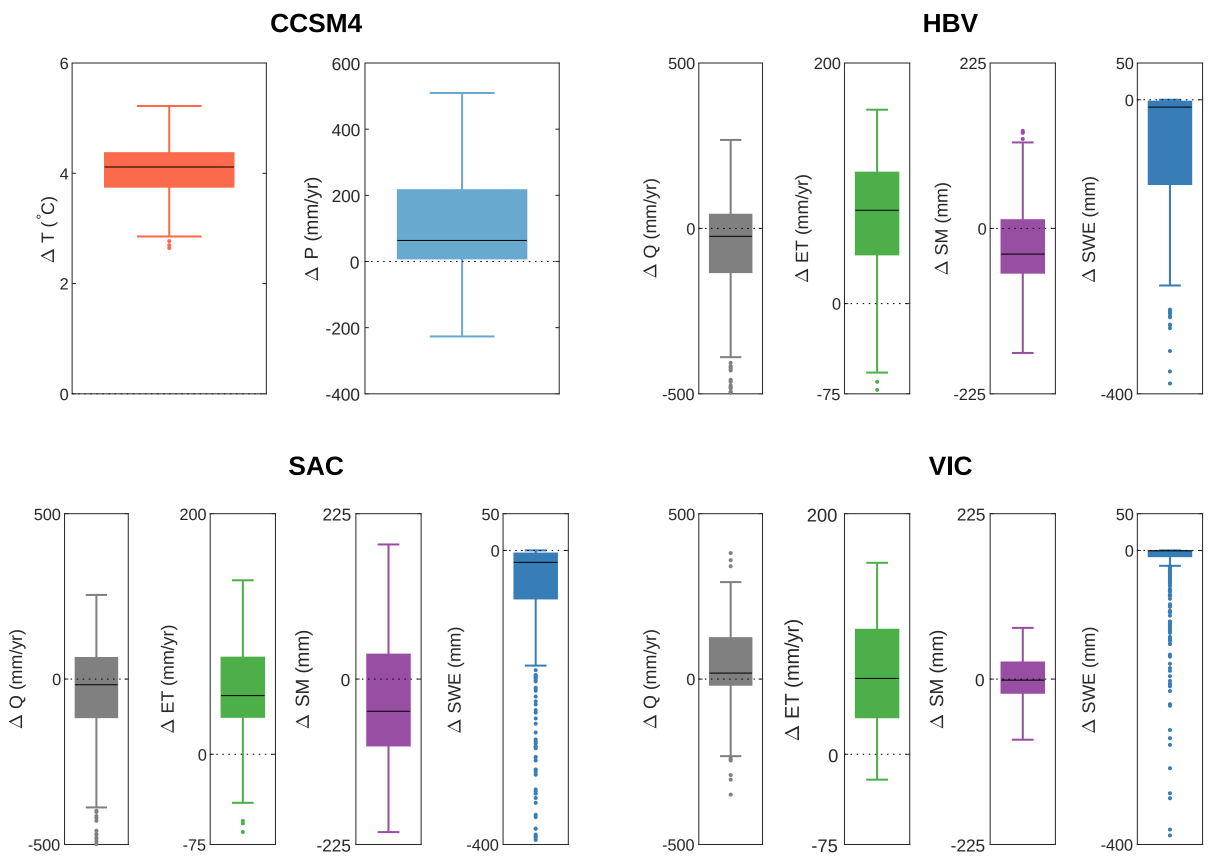

Figure 3. Changes between historical (1985-2008) and future (2070-2093) period. The top left panels depict the change in temperature (ΔT) and precipitation (ΔP) across the 605 basins, as obtained with CCSM4 (RCP8.5). The other boxplots depict the change in mean discharge (ΔQ), mean evapotranspiration (ΔET), mean soil moisture (ΔSM) and mean snow water equivalent (ΔSWE) between both periods for the 605 basins, as simulated by the three different hydrological models. Each model was run for a full parameter sample per basin, the average change across the parameter sample per basin was used to create the boxplots.

evapotranspiration will increase in the future. There is also agreement among the models that snow water equivalent will decrease in all basins, although the magnitude of change differs among the models. The signal in discharge has a less clear direction, which is also consistent among the models, although VIC seems to hinge on a general increase in discharge while HBV has slightly more basins where discharge would decrease. Both HBV and SAC simulate a decrease in soil moisture in most basins, whereas the median change in soil moisture across the 605 basins with VIC simulations is only -1.4 mm. The models seem to broadly agree on the general direction of change in several of the simulated water balance terms, but differences among models can already be observed and might be more pronounced for individual basins.

3.2 Changes in sensitivity between historical and future period

Fig. 4 shows the distribution of change in sensitivity between the historical (1985-2008) and future period (2070-2093) over all 605 basins for the three employed models. Consistent over all three models is a decrease in the sensitivity of snow parameters in the future. The parameters that show increasing sensitivity cannot consistently be associated to one specific process. Whereas a strong decrease in sensitivity requires a high sensitivity in the historical period, this is not required for a strong increase in sensitivity. It can be observed, however, that model parameters that display an increase in sensitivity were also already sensitive in the historical period.

In the HBV model especially the snow correction factor (SCF) displays a large decrease. This is also the parameter with the highest sensitivity in the historical period, therefore having the highest potential to decrease. The other three snow parameters in HBV displayed lower sensitivity in the historical period, and also show a less consistent decrease in the future. Also in the SAC and VIC models, the snow parameter that displayed the highest sensitivity in the historical period (SCF in SAC and Snowrough in VIC, respectively) show the strongest decrease, although less consistent than SCF in HBV.

Among the three models, different parameters related to different processes display an increase in sensitivity in the future. In HBV, evapotranspiration and soil parameters increase in sensitivity in the future with the largest increase in the evapotranspiration parameter PT, while there is hardly any observable change in sensitivity in percolation and deep layer parameters. In the SAC and VIC model, there are parameters associated to all processes except snow, that tend to mainly increase in sensitivity in the future. Like for HBV, also in SAC the evapotranspiration parameter PT has the highest increase. In the VIC model, the depth of the second soil layer (Depth2) shows the largest positive change in sensitivity.

3.3 Relationship between climatic variables and sensitivity changes

Since the 605 basins employed in the previous section are not a representative, balanced sample over climates and climate changes, the results are split out over climate indicators. Fig. 5 depicts how parameter sensitivity changes between historical and future period, related to the three Knoben climate indicators. From the figure, it can be seen that the patterns relating parameter sensitivity to climate and climate change indicators are weak. The aridity index seems to have relatively most explanatory value, followed by seasonality and the fraction of precipitation falling as snow, respectively. The change in sensitivity of snow and evapotranspiration parameters can be related to current mean temperature and precipitation and projected changes in mean temperature (Fig. A1), but the patterns vary per model.

In most cases, the patterns that can be identified relate to the projected change in climate. For instance in both SAC and HBV, the sensitivity of snow parameter SCF decreases especially in regions with a strong decrease in aridity index and in regions that were humid (positive aridity index in our definition) in the historical period. Soil moisture/shallow layer parameter Depth2 (VIC) and percolation parameter Expt2 (VIC) demonstrate a more pronounced increase in regions with decreasing aridity in-

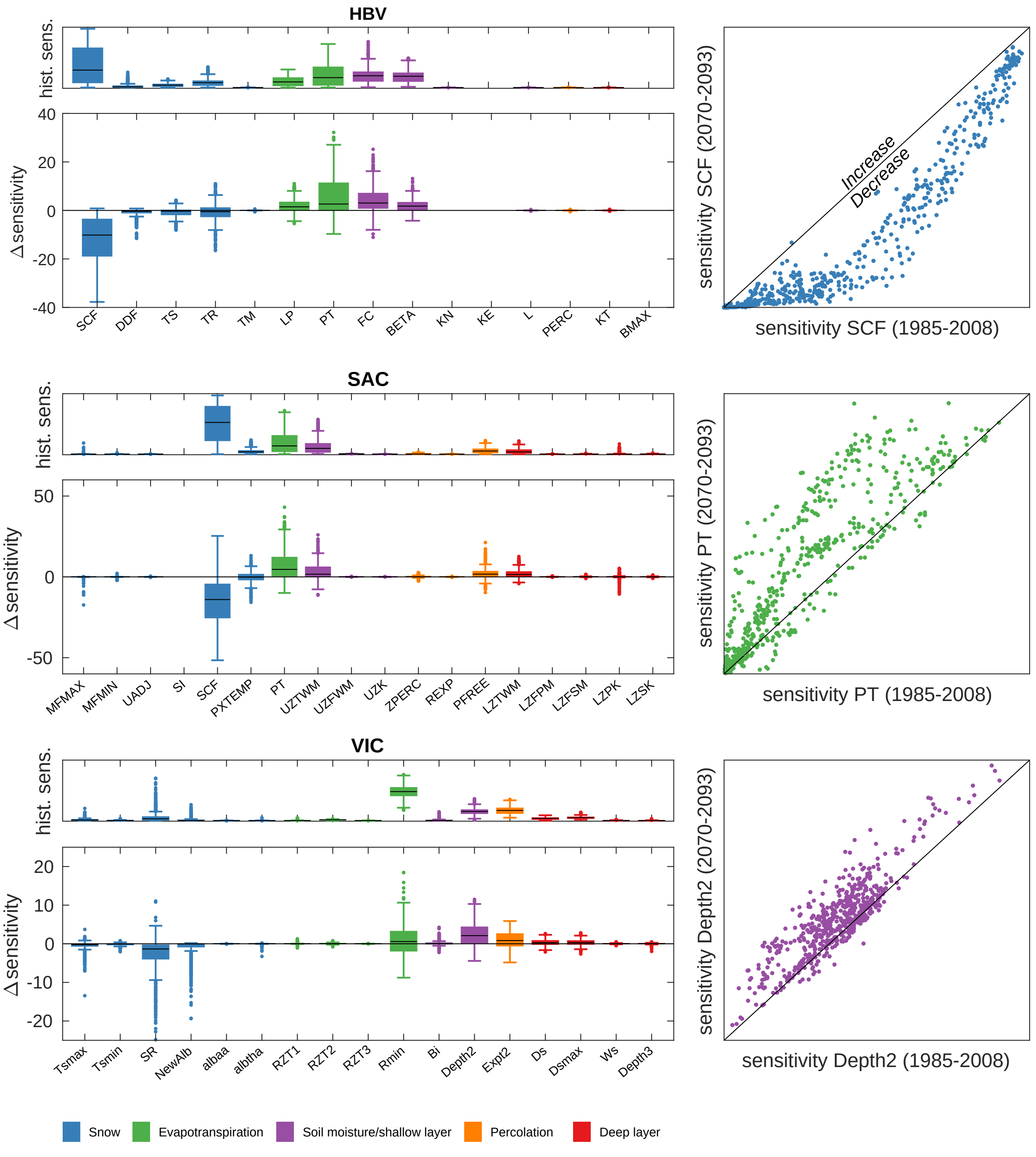

Figure 4. The distribution of change in parameter sensitivity (Δsensitivity) over 605 basins for the period 2070-2093 compared to 1985-2008, displayed per parameter per model. Above each Δsensitivity panel, historical sensitivity is displayed. The panels on the right show the data for a selected case per model.

dex. Sometimes also the historical climate, combined with the projected change, can show some organization. For example, the sensitivity of the evapotranspiration parameter PT in both SAC and HBV is particularly increasing in regions with high historical aridity index, and changes are more pronounced with larger projected changes, either an increase or a decrease, in aridity index.

Given that no clear patterns were revealed based on the Knoben indicators, we also explored patterns related directly to climate: the mean temperature and mean precipitation and their projected changes. These results can be found in Figure A1 in the Appendix. The snow parameters mainly decrease in sensitivity in basins with a historically mean temperature between 5 and 15°C, dependent on the model. In these basins, the fraction of snow will decrease in a warmer climate, whereas in basins with a lower mean temperature, snow will remain a relevant process in the future (Fontrodona Bach et al., 2018).

In HBV, the decrease in sensitivity in SCF is highest in catchments with a mean historical temperature between 5 and 10°C. An increase in the sensitivity of evapotranspiration parameter PT occurs in basins with projected changes in precipitation (both positive and negative). Also here, the largest increase is found in basins with a mean historical temperature between 5 and 10°C. An increase in the sensitivity of shallow layer parameter FC is related to no change or a decrease in precipitation in the future.

In SAC, snow parameter SCF decreases in sensitivity in basins with a mean historical temperature of about 10°C. In these basins, the sensitivity of evapotranspiration parameter PT and lower zone parameter UZTWM increases. Similar to HBV, evapotranspiration parameter PT changes in sensitivity in basins with both positive and negative changes in precipitation in the future.

In VIC, the patterns are weakest. Here, we see a decrease of snow parameters in basins with mean temperatures lower than 5°C, and even around zero combined with no change in precipitation. Evapotranspiration parameter Rmin increases in sensitivity in basins with increase in precipitation and vice-versa. Shallow layer parameter Depth2 and to a lower extent percolation parameter Expt2 decreases in sensitivity in basins with mean temperatures between 0 and 10°C and with decreasing precipitation.

3.4 Impact of sensitivity changes on calibration strategy

In this section we explore to what extent the changes in parameter sensitivity that were observed in the previous sections propagate into the calibration procedure. Fig. 6 depicts the percentage of catchments in which parameters appeared in the top-5, both historically and in the future. Snow parameters drop out of the top-5 in some cases, while the relevance of parameters with already many top-5 notations further increases. This indicates that the variation among catchments in top-5 parameters decreases in the future. Although changes in top-5 parameters are observed, the overall top-5 of the parameters is in most cases maintained (51 to 57% of the catchments, dependent on the model). In 41 to 45% of the catchments, one parameter changes in the top-5. The maximum number of changes in the parameter top-5 per catchment is two, which occurs only in 2 to 4% of the

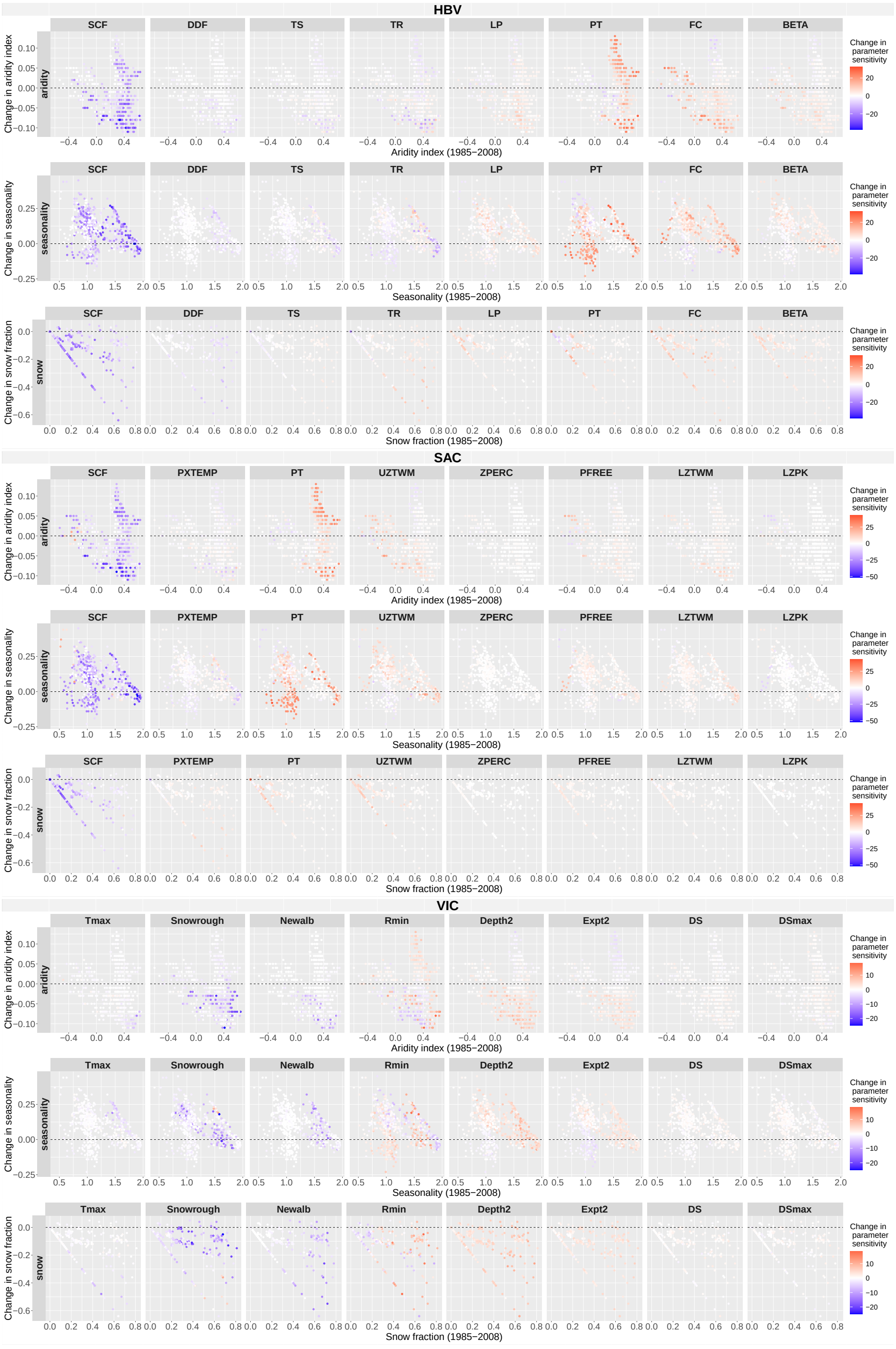

Figure 5. Change in parameter sensitivity versus historical climate indicators and change in climate indicators for 605 basins. The climate indicators are aridity index (-1 highly arid, +1 highly humid), seasonality, and fraction of precipitation falling as snow, as defined by Knoben et al. (2018). Parameter sensitivity for the historical period is expressed in dot size, change in parameter sensitivity in colour: red indicates an increase in sensitivity, blue a decrease.

investigated basins.

For HBV, snow parameter SCF historically has a top-5 notation in 76% of the basins, in the future this drops to 57% - a relative drop of 24%. The largest increase in top-5 notations for HBV is found for evapotranspiration parameter LP (a relative increase of 22%). In SAC, snow parameter SCF loses its top-5 notation in 11% of the basins where it used to be relevant. Lower zone parameter LZTWM shows the strongest increase in top-5 notations (a relative increase of 24%). In VIC, mainly the snow parameter Snowrough loses top-5 notations (a relative decrease of 53%). Deep layer parameters gain most notations, especially DS (a relative increase of 49%).

The results of the left three panels in Fig. 6 cannot directly be generalized because the 605 explored basins are not a well-balanced sample in terms of climate and climate change. Therefore, the change in top-5 parameters is also again displayed against the Knoben indicators (right panels in Fig. 6). It can be observed that one change in parameter top-5 can occur over all climates and climate changes. Only VIC is showing fewer changes in basins in between constant and seasonal, and with decreasing seasonality. Also two changes in parameter top-5 seems to occur across all climates and climate changes. Only for HBV, this seems to be constrained to wet catchments that become drier (lower aridity index) in the future.

In conclusion, the impact of changes in parameter sensitivity on calibration strategy remains limited to a maximum of two parameter changes in the parameter top-5, at least for the explored climates and climate changes. The changes in top-5 positions are model dependent and do not demonstrate a clear relation to climate or climate change.

3.5 Diagnostic model evaluation based on changes in sensitivity

The evaluated changes in parameter sensitivity in response to climate change can be perceived as a way to evaluate models diagnostically, especially since we can compare the results for three different hydrological models. The parameter sensitivity in the historical period (the top panels in Fig. 4) already shows that the models activate different processes to simulate historical discharge. Our analysis of change in sensitivity demonstrates that the models also respond differently to changes in forcing.

There are a few points where all three models agree: all models simulate a decrease in snow in the future across all basins, and an increase in ET across most basins (Fig. 3). This is also visible in the change in sensitivity of the parameters related to these processes. In all models, the snow parameters tend to decrease most in sensitivity (dependent on their historical sensitivity), and a median increase in the sensitivity of ET parameters was found (although clearly weakest for VIC, Fig. 4). These results are robust across different formulations for snow and ET processes: SAC and HBV share the same ET formulation and employ a comparable snow formulation, but VIC employs a completely different formulation for both ET and snow. Yet, all three models agree on these signals.

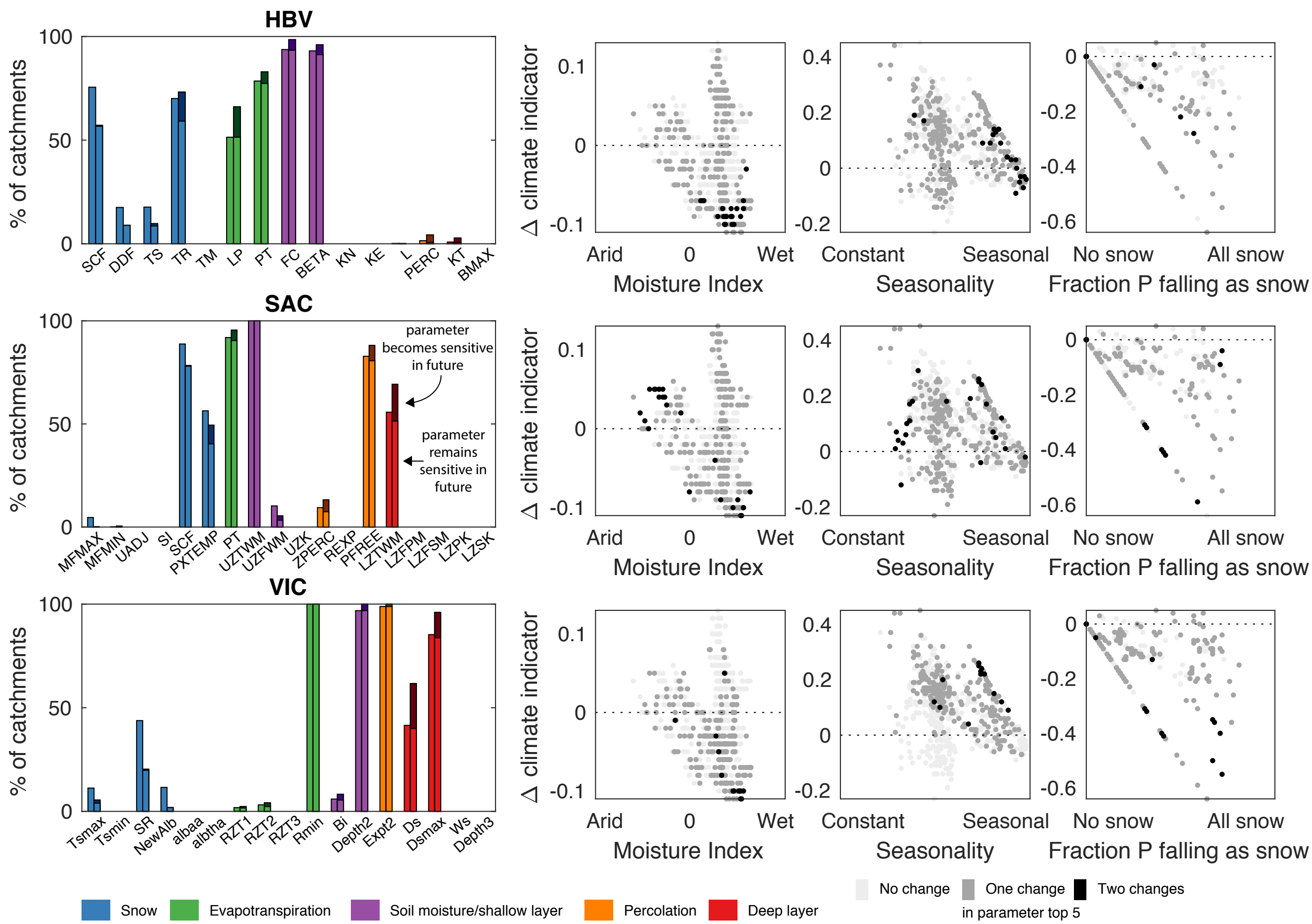

Figure 6. Impact of change in parameter sensitivity on top-5 position, where top-5 refers to the five most sensitive parameters per basin - generally the parameters that are calibrated. The left panels show how often a parameter appears in the top-5 both historically and in the future. The right panels relates the number of changes in the parameter top-5 to climate and climate change indicators.

However, many other changes in sensitivity can be observed where the models disagree, for instance the role of percolation and soil moisture/the shallow layer. To further explore how the models respond to climate change in terms of parameter sensitivity, the transmission of sensitivity is explored by means of the negative correlation between change in sensitivity among two parameters. An example is the left panel of Fig. 7, depicting a negative correlation between the change in sensitivity of snow parameter SCF and the change in sensitivity of evapotranspiration parameter LP for HBV, which can indicate a transmission of sensitivity from SCF to LP. The chord diagrams in Fig. 7 show the correlations between the parameters with decreasing and increasing sensitivities. All three models display a decrease in sensitivity of the snow parameters, but this sensitivity is transmitted to different process parameters in the three models. In HBV, mainly to evapotranspiration and shallow layer parameters, in SAC evapotranspiration, percolation, and deep layer parameters, and in VIC to shallow layer and deep layer parameters. Weak transmissions can indicate that parameter sensitivity changes at several places in the model structure.

Besides the clear transmission from snow parameter SCF to evapotranspiration parameter LP in HBV, it is also visible in Fig. 7 that snow parameter TS mainly transmits to evapotranspiration parameter PT: different snow parameters transmit to different evapotranspiration parameters. What is also visible in the chord diagram is that snow parameter DDF is mainly transmitting to snow parameter TR, explaining the increase in sensitivity in some regions for this snow parameter. Not displayed in the chord diagram for clarity, is that snow parameter TR then again transmits to evapotranspiration parameters LP, PT, and shallow layer parameter FC.

For SAC, the snow parameter SCF demonstrates a clear negative relation with many parameters that increase in sensitivity. High correlations were found with percolation parameter PFREE (-0.63), deep layer parameter LZTWM (-0.72), and evapotranspiration parameter PT (-0.72). The correlations between snow parameter PXTEMP and the parameters with increasing sensitivity are less pronounced. Two other parameters in SAC experienced a slight decrease in parameter sensitivity; snow parameter MFMAX and shallow layer parameter UZFWM, especially visible in their loss of top-5 positions in Fig. 6. These two parameters, however, did not display any negative correlation with any of the parameters that experience a clear increase in top-5 positions.

For VIC, the negative correlations are generally weaker than what was found for the other two models, but still some insights can be obtained from the chord diagram. For instance that snow parameter Newalb mainly transmits to shallow layer parameter Depth2 (-0.59), while snow parameter Snowrough is the only one that shows a correlation with deep layer parameter DSmax (-0.32). Shallow layer parameter Infilt increases in number of top-5 positions (Fig. 6) but did not display any clear relation with the parameters that decrease in sensitivity.

Whereas the models agree on the decline in snow water equivalent and decreased sensitivity of snow parameters despite employing different snow formulations, the models disagree on changes related to many other processes. Since the three models differ in many aspects in their model structure, the difference in response to changing forcing cannot directly be related

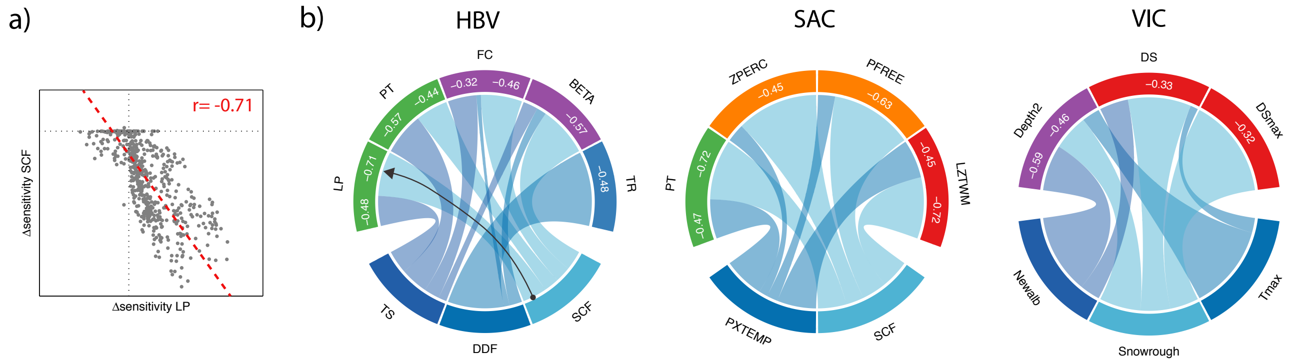

Figure 7. Indication of parameter sensitivity transmission. Panel a) an example for HBV: the decrease in sensitivity of parameter SCF shows a strong **negative?** correlation with an increase in the sensitivity of parameter LP, which can indicate that SCF transmits sensitivity to LP. Since we focus on transmission, we only evaluate negative correlations. Panel b) The chord (circle) diagrams display transmission of sensitivity, indicated with a band from the parameter that decreases in sensitivity to the parameter that increases in sensitivity. The width of the band indicates the strength of the negative correlation. The example from panel a is indicated with an arrow in the chord diagram of HBV. The white number indicates the strength of the correlation; -0.71 between SCF and LP. In all three chord diagrams, the lower part shows the parameters that decrease in sensitivity, and the upper part the parameters that increase in sensitivity, with the white number indicating the strength of the correlation (for clarity, negative correlations lower than 0.32 are not displayed). Colors are according to the process they represent (with different shades of blue used for snow parameters for clarity). The chord diagrams are focused around the most relevant parameters based on Fig. 6.

to specifics of the model structure. The results, however, do show that the internal functioning of the models differ when used for long term simulations, and this might impact the results and subsequently the conclusions of the model study.

4 Discussion

4.1 Changes in states, fluxes, and sensitivity between historical and future period

A first evaluation of the different states and fluxes that are simulated by the models for the future period demonstrates that the three models agree that in general, snow water equivalent will decrease and evapotranspiration will increase under RCP 8.5 as simulated by CCSM4. This same signal is propagated into the sensitivity of the parameters related to these processes: in all three models, the sensitivity of snow parameters tends to decrease, and the sensitivity of evapotranspiration parameters tends to increase (although the models disagree on the magnitude of change). Since we conducted parameter sensitivity evaluated for the mean discharge, these results imply that the impact of snow on discharge will decrease, while the impact of evapotranspiration on discharge will increase.

For other states and fluxes simulated by the models, such as soil moisture and percolation, the models agree less on the change in sensitivity (Fig. 4). This can first and foremost be attributed to a difference in model structure, but the impact of model structure can be further emphasized by the target variable that we used for our sensitivity analysis. We evaluated the sensitivity of parameters to simulate mean discharge. For instance for HBV, percolation parameters historically already did not display a strong sensitivity for mean discharge, and the sensitivity of percolation parameters does not change in the future. This does not automatically imply that HBV does not simulate a change in percolation as a consequence of climate change, but mainly that mean discharge and percolation are decoupled (at least in comparison to other processes) in the HBV model structure. Another example is that all models simulate a substantial increase in evapotranspiration, but for VIC this does not lead to a substantial increase in sensitivity of parameters related to ET. Other signatures as target variable might therefore give different results.

4.2 Climate indicators to relate changes in sensitivity

We evaluated change in sensitivity against three climate indicators; aridity index, seasonality, and fraction precipitation falling as snow. We were not able to identify a clear relation between climate indicator, change in climate indicator, and change in parameter sensitivity. In our approach, we investigated if any temporal relations exist. Another way to evaluate change in sensitivity would be to evaluate spatial relations. van Werkhoven et al. (2008) for instance, demonstrate that spatial gradients in model sensitivity exist that relate to climate. If we can establish a temporal relation in the same way van Werkhoven et al. (2008) could demonstrate spatial relations, space-for-time trading would be an option to determine which parameters become sensitive in the future. The lack of a clear relation between climate, climate change, and parameter sensitivity however, also demonstrates that we should critically evaluate the adequacy of the model structures for long-term projections.

4.3 Impact of sensitivity changes on ranking of sensitive parameters

We investigated how parameter sensitivity changes as a consequence of climate change. We also explored the use of sensitivity analysis to provide the most relevant parameters (factor prioritisation) for an effective model calibration (Saltelli et al., 2006; Reusser et al., 2011). Within this context, we have shown how changes in parameter sensitivity propagate into the selection of relevant parameters for model calibration. We assumed a general calibration strategy where the modeller selects the five most sensitive parameters for calibration. Certainly many other calibration strategies exist. For example, one could select all the parameters that exceed a certain sensitivity-threshold as suggested by van Werkhoven et al. (2009) or when compared to a dummy parameter as suggested by Zadeh et al. (2017), resulting in a higher or lower number of parameters for the calibration, or simply include all the parameters in the model if the model is highly parsimonious (Melsen et al., 2014). Our results are, however, still relevant in the context of other calibration strategies, as the changes in sensitivity will still influence the calibration results. That is, it is difficult to calibrate a parameter if the model is hardly or not sensitive to changes in its values in current-day climate.

The implication of our result is that, the more the parameter sensitivity changes, the more parameter identifiability decreases for long-term projections. Accordingly, we can expect that in particular the parameters that will enter the top-5 in the future are probably not well identified in the historical period. Therefore, we provide suggestions to account for changing sensitivities in the calibration strategy of hydrological models for long-term projections. A first strategy, related to methods that have been suggested for changing parameters over time (Merz et al., 2011; Vaze et al., 2010), is to conduct sensitivity analysis over different parts of the observation-period, and calibrate the model on the period that best resembles the parameter sensitivity of the future scenario. A risk, however, is that the calibration period becomes too short to determine stable parameter values (Yapo et al., 1996). A second strategy is to sample the parameters that will become sensitive in the future. Provided that, in the current climate, the model is not sensitive to changes in parameter θ_i, the value of θ_i cannot be inverted through calibration. However, in the future the value of θ_i does become relevant. In order to correctly capture spread in long-term projections that results from uncertainty in this parameter-value, the value of θ_i should be sampled. Hereby, we have to emphasise that in this context, parameter uncertainty is specifically related to expected changes in the relevance of the associated processes in future. A third option could be to increase the effort in finding data to be able to calibrate a parameter directly to the associated process. Sensitivity analyses on different processes demonstrate that the sensitivity signal increases using the associated process as target variable (Guse et al., 2016a). In this way, it can be expected that parameters are better identifiable and more robust for future simulations.

4.4 Diagnostic interpretation

In the previous sub-section we provide suggestions to further validate the calibration procedure of models employed for long-term projections. It seems a valid question, however, whether our models are fit for this purpose at all. The results of the sensitivity analysis indicate a change in relevant processes in the future which is captured differently among the three investigated models. This emphasizes the need to improve model structure for long term projections, as suggested by Fowler et al. (2018), Grigg and Hughes (2018) and Westra et al. (2014).

Assuming that a sensitivity analysis conducted over 23 years of daily data is robust and thus that the observed changes in sensitivity can be attributed to a changing climate rather than to noise, our results demonstrate that parameter sensitivity is nonstationary (Koutsoyiannis and Montanari, 2015). Nonstationarity of parameter sensitivity fits in the growing body of literature identifying nonstationarity when simulating the hydrological system on the long term (e.g. Milly et al., 2008; Thirel et al., 2015; Fowler et al., 2016, 2018). Nonstationarity is not only disclosed through a change in sensitivity, but also through a change in parameter values over time (Vaze et al., 2010; Merz et al., 2011). The identification of nonstationarity in parameter values is the result of the simplified model representations, not capturing dynamics and/or processes that are relevant in the real world. Fowler et al. (2018) provides a framework to evaluate model improvement under nonstationary conditions; Grigg and Hughes (2018); Westra et al. (2014) and Duethmann et al. (2020) adapted model structure to account for nonstationarity, leading to improved model results. This study reinforces this direction of research; even though the decrease in sensitivity among all three models can consistently be found for the snow parameters, the increase in sensitivity can be attributed to different processes in the three models, which might indicate that a relevant process is missing in any of the models, stressing the need to carefully assess whether these models are appropriate for long-term projections. The differences in which processes and associated parameters becomes more relevant among the models shows that there is no consensus how the hydrological system will change in future.

A decrease of sensitivity of snow parameters and an increase in the sensitivity of evapotranspiration parameters in a warming climate could be expected based on expert judgement, and at least the three models agree on those signals despite employing different formulations to compute these processes. However, the models disagree on the other processes that will become more or less relevant in the future, while changes in these processes are not straight forward to estimate based on expert judgement. It is, for instance, not easy to judge whether the relatively higher amount of rain in the future (due to a decrease in snow) goes on average more to higher evaporation or to higher infiltration. As such, we have to acknowledge that the models differ in the processes they use to simulate future changes, and that we cannot easily differentiate the right from the wrong models.

5 Conclusions

The sensitivity of the parameters in three investigated hydrological models changes within a plausible changing climate. The three models agree that especially the snow parameters decline in sensitivity, while evapotranspiration parameters show a tendency to increase. Which other parameters increase in sensitivity is, however, less consistent among the models; sometimes mainly ET and soil moisture/shallow layer parameters, sometimes mainly percolation and/or deep layer parameters. These differences occur due to differences in three model structures. We were not able to identify a clear pattern in which kind of climates and expected climate changes most changes in parameter sensitivity take place.

The change in parameter sensitivity propagates into the calibration strategy. Typically, a global sensitivity analysis is conducted to determine the most sensitive parameters, and based on that, the top-5 most sensitive parameters are selected for calibration. Dependent on the model, 43% to 49% of the 605 investigated catchments has at least one parameter changing in the top-5 in the future. The maximum number of changes in the top-5 parameters is two, in 2-4% of the catchments. Since these results were obtained for the highest emission scenario (RCP8.5), fewer changes might be expected for lower emission scenarios.

Some parameters become sensitive in the future, but are currently not sensitive. Therefore, their value cannot be obtained through calibration based on current data. One way to account for changes in sensitivity is to identify a historical period that mimicks the future projected sensitivity. Another approach is to sample the parameter that becomes sensitive in the future, to account for predictive uncertainty as a consequence of the uncertainty in this parameter value. A third approach is to invert the value of this parameter based on observations specifically related to the process that the parameter is related to.

Besides implications for the calibration strategy when using models for long-term projections, our results also have implications for model selection for this purpose. The results demonstrate that the three employed models consider different processes as becoming more or less relevant in the future; they simulate different systemic changes. Whereas the models agree on systemic changes that can be excepted based on expert judgement (decreased relevance of snow and increased relevance of evapotranspiration in a warming climate), the models disagree on other processes that are more difficult to judge. These results not only stress the need, but also the challenge in carefully assessing model structure adequacy when applying models for long-term projections.

Acknowledgments

BG thanks for financial support from the German Research Foundation ("Deutsche Forschungsgemeinschaft", DFG) via the FOR 2416 "Space-Time Dynamics of Extreme Floods (SPATE)" research group. The authors would like to thank three anonymous reviewers for their constructive feedback, and Thorsten Wagener for his suggestions.

Data availability. The model output will be made available on the 4TU website

Author contributions. LM and BG designed the study together. LM conducted the calculations and wrote the first draft of the paper. LM and BG processed the data together and finalized the manuscript together.

Competing interests. There are no competing interests

475 **Appendix A: Parameter ranges**

Table A1. Selected parameters, their classification, and their boundaries for the HBV model. The parameters and their boundaries are based on Parajka et al. (2007); Uhlenbrook et al. (1999); Abebe et al. (2010). The Priestley-Taylor parameter is based on Lhomme (1997).

	Name	Unit	Lower boundary	Upper boundary	Description
1	Tm	$^{\circ}$C	-3.0	3.0	Temperature where snow melt starts
2	Ts	$^{\circ}$C	Tr-0.01	Tr-3	Temp. below which precipitation is snow
3	Tr	$^{\circ}$C	0.0	3.0	Temp. above which precipitation is rain
4	DDF	mm $^{\circ}C^{-1}$ d^{-1}	0.04	12	Degree day factor
5	SCF	-	0.1	5.0	Snow correction factor
6	LP	-	0.0	1.0	Evaporation reduction threshold
7	PT	-	1.0	1.74	Priestley-Taylor coefficient
8	FC	mm	0.0	2000	Max soil moisture storage
9	BETA	-	0.0	20	Non-linear shape coefficient
10	K0	day	0.0	2.0	Storage coefficient of very fast response
11	K1	day	2.0	30	Storage coefficient of fast response
12	L	mm	0.0	100	Reservoir threshold
13	PERC	mm d^{-1}	0.0	100	Percolation rate
14	K2	day	30	250	Storage coefficient of slow response
15	BMAX	day	0.0	30	Max baseflow of low flows

Table A2. Selected parameters and their boundaries for the SAC model. The parameter boundaries are based on Newman et al. (2015), the Priestley-Taylor parameter has been adapted based on Lhomme (1997).

	Name	Unit	Lower boundary	Upper boundary	Description
1	MFMAX	mm $^\circ C^{-1}$ $6h^{-1}$	0.8	3.0	Max snow melt factor
2	MFMIN	mm $^\circ C^{-1}$ $6h^{-1}$	0.01	0.79	Min snow melt factor
3	UADJ	km $6h^{-1}$	0.01	0.40	Wind adjustment factor for rain on snow
4	SI	mm	1.0	3500	snow water equivalent for 100% snow area
5	SCF	-	0.1	5.0	Snow undercatch correction factor
6	PXTEMP	$^\circ C$	-3.0	3.0	Temperature for rain/snow transition
7	PT	-	1.0	1.74	Priestley-Taylor coefficient
8	UZTWM	mm	1.0	800	Upper zone max storage of tension water
9	UZFWM	mm	1.0	800	Upper zone max storage of free water
10	UZK	day^{-1}	0.1	0.7	Upper zone free water lateral depletion rate
11	ZPERC	-	1.0	250	Max percolation rate
12	REXP	-	0.0	6.0	Exponent of the percolation equation
13	PFREE	-	0.0	1.0	Frac. percolating from upper to lower zone
14	LZTWM	mm	1.0	800	Lower zone max storage of tension water
15	LZFPM	mm	1.0	800	Lower zone max storage of free water
16	LZFSM	mm	1.0	1000	Lower zone max storage of sec. free water
17	LZPK	day^{-1}	1^{-5}	0.025	Lower zone prim. free water depletion rate
18	LZSK	day^{-1}	1^{-3}	0.25	Lower zone sec. free water depletion rate

Table A3. Selected parameters and their boundaries for the VIC model based on Demaria et al. (2007); Chaney et al. (2015); Melsen et al. (2016); Mendoza et al. (2015).

	Name	Unit	LB	UB	Description
1	Tsmax	$^\circ$C	0.0	3.0	Max temp. where snowfall can occur
2	Tsmin	$^\circ$C	Tsmax-0.01	Tsmax-3.0	Min temp. where rainfall can occur
3	SR	-	$5 \cdot 10^{-5}$	0.5	Surface roughness of the snow pack
4	NewAlb	-	0.7	0.99	New snow albedo
5	albaa	-	0.88	0.99	Base in snow albedo function for accum.
6	albtha	-	0.66	0.98	Base in snow albedo function for melt
7	RZT1	-	0.5	2	Multipl. factor rootzone thickness layer 1
8	RZT2	-	0.5	2	Multipl. factor rootzone thickness layer 2
9	RZT3	-	0.5	2	Multipl. factor rootzone thickness layer 3
10	Rmin	-	0.1	10	Multipl. factor min. stom. res. vegetation
11	Bi	-	10^{-5}	0.4	Infiltration shape parameter
12	Depth2	m	0.1	3.0	Depth of soil layer 2
13	Expt2	-	4.0	30	Exponent of the Brooks-Corey relation
14	Ds	-	10^{-4}	1.0	Frac. Dsmax non-linear baseflow starts
15	Dsmax	mm d^{-1}	0.1	50	Max velocity of the baseflow
16	Ws	-	0.2	1.0	Frac. Wsmax non-linear baseflow starts
17	Depth3	m	0.1	3.0	Depth of soil layer 3

Appendix B: Change in sensitivity versus temperature and precipitation

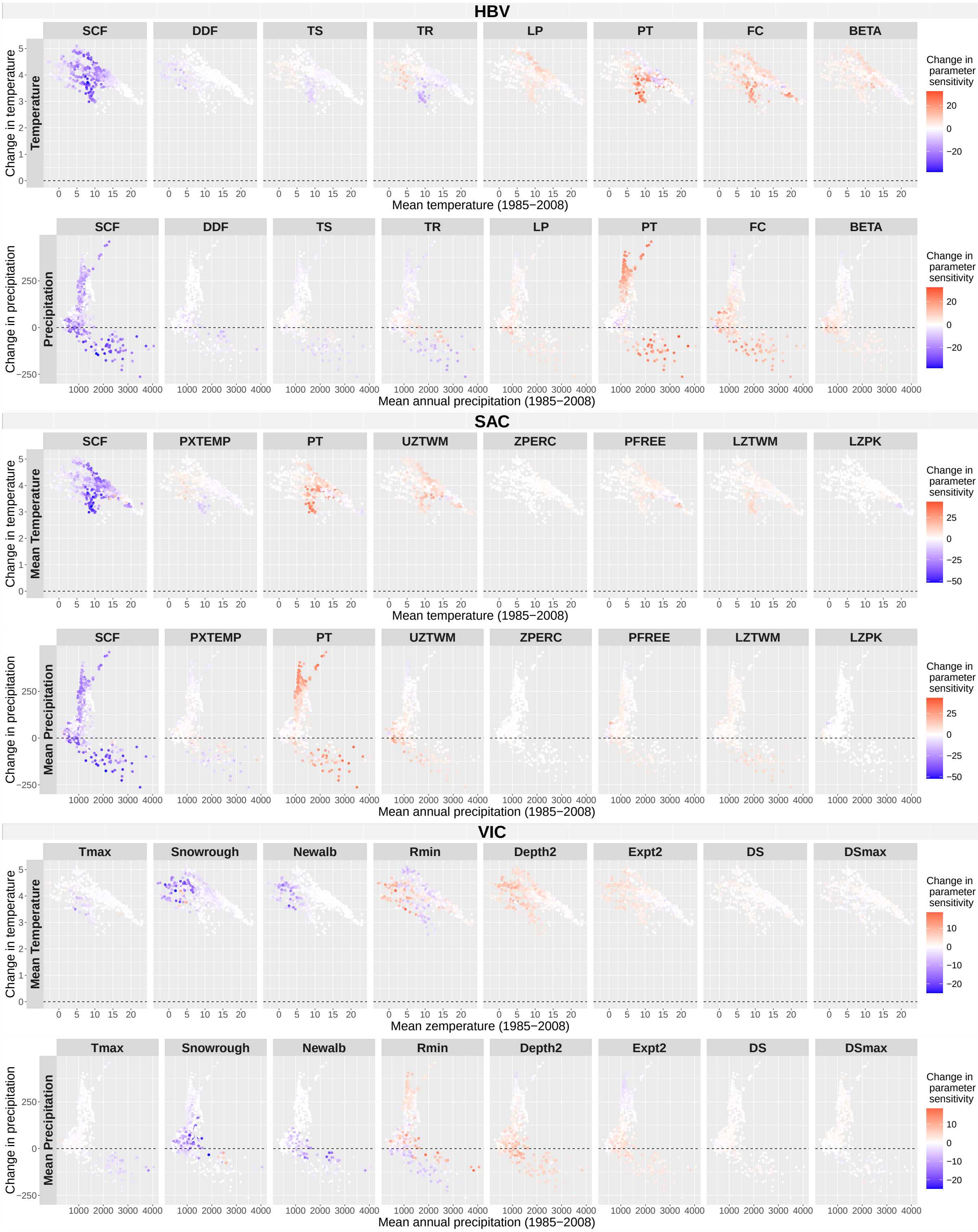

Figure A1. Change versus historical values in mean temperature and mean precipitation over 605 basins, with change in parameter sensitivity indicated. Parameter sensitivity for the historical period is expressed as dot size. Change in parameter sensitivity in colour. Red colours indicate an increase in sensitivity, blue a decrease.

References

Abebe, N., Ogden, F., and Pradhan, N.: Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation, J. Hydrol., 389, 301–310, https://doi.org/10.1016/j.jhydrol.2010.06.007, 2010.

Addor, N., Newman, A., Mizukami, N., and Clark, M.: The CAMELS data set: catchment attributes and meteorology for large-sample studies. version 1.0., Boulder, CO: UCAR/NCAR., https://doi.org/10.5065/D6G73C3Q, 2017.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, Tech. rep., SMHI Report RHO 7, Norrköping, 1976.

Bergström, S.: The HBV model - its structure and applications, Tech. Rep. 4, SMHI reports hydrology, 1992.

Beven, K.: Changing ideas in hydrology - The case of physically-based models, J.Hydrol., 105, 157–172, https://doi.org/10.1016/0022-1694(89)90101-7, 1989.

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., and Hill, M.: Making the most out of a hydrological model data set: Sensitivity analysis to open the model black box, Water Resour. Res., 53, 7933–7950, https://doi.org/10.1002/2017WR020767, 2017.

Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., and Clark, M. P.: Future streamflow regime changes in the United States: assessment using functional classification, Hydrol. Earth Syst. Sci., 24, 3951–3966, https://doi.org/10.5194/hess-24-3951-2020, 2020.

Burnash, R., Ferral, R., and McGuire, R.: A generalized streamflow simulation system - conceptual modeling for digital computers, Tech. rep., U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources., 1973.

Chaney, N., Herman, J., Reed, P., and Wood, E.: Flood and drought hydrologic monitoring: the role of model parameter uncertainty, Hydrol. Earth Syst. Sc., 19, 3239–3251, https://doi.org/10.5194/hess-19-3239-2015, 2015.

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., Kao, S., Mao, Y., Mizukami, N., Mote, P. W., Pan, M., Pytlak, E., and Xiao, M.: How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates, Earth's Future, 7, 623–637, https://doi.org/10.1029/2018EF001047, 2019.

Christensen, N., Wood, A., Voisin, N., Lettenmaier, D., and Palmer, R.: The effects of climate change on the hydrology and water resources of the Colorado River basin, Climatic Change, 62, 337–363, 2004.

Demaria, E. M., Nijssen, B., and Wagener, T.: Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model, J. Geophys. Res., 112, D11113, https://doi.org/10.1029/2006JD007534, 2007.

Devak, M. and Dhanya, C. T.: Sensitivity analysis of hydrological models: review and way forward, J. Water Clim. Change, 8, 557–575, https://doi.org/10.2166/wcc.2017.149, 2017.

Duethmann, D., Blöschl, G., and Parajka, J.: Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change?, Hydrol. Earth Syst. Sci., 24, 3493–3511, https://doi.org/10.5194/hess-24-3493-2020, 2020.

Fontrodona Bach, A., van der Schrier, G., Melsen, L., Klein Tank, A., and Teuling, A.: Widespread and Accelerated Decrease of Observed Mean and Extreme Snow Depth Over Europe, Geophys. Res. Let., 45, 12,312–12,319, https://doi.org/10.1029/2018GL079799, 2018.

Fowler, K., Peel, M., Western, A., Zhang, L., and Peterson, T.: Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models, Water Resour. Res., 52, 1820–1846, https://doi.org/10.1002/2015WR018068, 2016.

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., and Zhang, L.: Simulating runoff under changing climatic conditions: A framework for model improvement, Water Resour. Res., 54, 9812–9832, https://doi.org/10.1029/2018WR023989, 2018.

Gent, P., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z., and Zhang, M.: The Community Climate System Model Version 4, J. Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

Grigg, A. and Hughes, J.: Nonstationarity driven by multidecadal change in catchment groundwater storage: A test of modifications to a common rainfall-run-off model, Hydrol. Process., 32, 3675–3688, https://doi.org/10.1002/hyp.13282, 2018.

Guse, B., Reusser, D. E., and Fohrer, N.: How to improve the representation of hydrological processes in SWAT for a lowland catchment - Temporal analysis of parameter sensitivity and model performance, Hydrol. Process., 28, 2651–2670. doi: 10.1002/hyp.977, 2014.

Guse, B., Pfannerstill, M., Gafurov, A., Fohrer, N., and Gupta, H.: Demasking the integrated information of discharge: Advancing sensitivity analyses to consider different hydrological components and their rates of change, Water Resour. Res., 52, 8724–8743, https://doi.org/10.1002/2016WR018894, 2016a.

Guse, B., Pfannerstill, M., Strauch, M., Reusser, D. E., Volk, M., Gupta, H. V., and Fohrer, N.: On characterizing the temporal dominance patterns of model parameters and processes, Hydrol. Process., https://doi.org/10.1002/hyp.10764, 2016b.

Haghnegahdar, A., Razavi, S., Yassin, F., and Wheater, H.: Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty, Hydr. Proc., 31, 4462–4476, https://doi.org/10.1002/hyp.11358, 2017.

Herman, J., Kollat, J., Reed, P., and Wagener, T.: Technical note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models, Hydrol. Earth Syst. Sc., 17, 2893–2903, https://doi.org/10.5194/hess-17-2893-2013, 2013a.

Herman, J., Reed, P., and Wagener, T.: Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior, Water Resour. Res., 49, 1400–1414, https://doi.org/10.1002/wrcr.20124, 2013b.

Jakeman, A. and Hornberger, G.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29, 2637–2649, https://doi.org/10.1029/93WR00877, 1993.

Knoben, W., Woods, R., and Freer, J.: A quantitative hydrological climate classification evaluated with independent streamflow data, Water Resour. Res., 54, 5088–5109, https://doi.org/10.1029/2018WR022913, 2018.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1194–1199, https://doi.org/10.1002/grl.50256, 2013.

Koutroulis, A., Tsanis, I., Daliakopoulos, I., and Jacob, D.: Impact of climate change on water resources status: A case study for Crete Island, Greece, J. Hydrol., 479, 146–158, https://doi.org/10.1016/j.jhydrol.2012.11.055, 2013.

Koutsoyiannis, D. and Montanari, A.: Negligent killing of scientific concepts: the stationarity case, Hydrolog. Sci. J., 60, 1174–1183, https://doi.org/10.1080/02626667.2014.959959, 2015.

Lhomme, J.: An examination of the Priestley-Taylor equation using a convective boundary layer model, Water Resour. Res., 33, 2571–2578, https://doi.org/10.1029/97WR01897, 1997.

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., 99, 14,415–14,458, https://doi.org/10.1029/94JD00483, 1994.

Liang, X., Wood, E. F., and Lettenmaier, D. P.: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, Global Planet. Change, 13, 195–206, https://doi.org/10.1016/0921-8181(95)00046-1, 1996.

Mai, J. and Tolson, A.: Model Variable Augmentation (MVA) for diagnostic assessment of sensitivity analysis results, Water Resour. Res., 55, 2631–2651, https://doi.org/10.1029/2018WR023382, 2019.

Massmann, C. and Holzmann, H.: Analysing the Sub-processes of a Conceptual Rainfall-Runoff Model Using Information about the Parameter Sensitivity and Variance, Environ. Model. Assess., 20, 41–53, https://doi.org/10.1007/s10666-014-9414-6, 2015.

Maurer, E., Wood, A., Adam, J., Lettenmaier, D., and Nijssen, B.: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States, J. Climate, 15, 3237–3251, https://doi.org/http://dx.doi.org/10.1175/JCLI-D-12-00508.1, 2002.

McMillan, H., Seibert, J., Petersen-Overleir, A., Lang, M., White, P., Snelder, T., Rutherford, K., Krueger, T., Mason, R., and Kiang, J.: How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications, Water Resour. Res., 53, 5220–5228, https://doi.org/10.1002/2016WR020328, 2017.

Melsen, L. and Guse, B.: Hydrological drought simulations: How climate and model structure control parameter sensitivity, Water Resour. Res., 55, 10 527–10 547, https://doi.org/10.1029/2019WR025230, 2019.

Melsen, L., Teuling, A., van Berkum, S., Torfs, P., and Uijlenhoet, R.: Catchments as simple dynamical systems: A case study on methods and data requirements for parameter identification, Water Resour. Res., 50, 5577–5596, https://doi.org/10.1002/2013WR014720, 2014.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale prealpine basin, Hydrol. Earth Syst. Sci., 20, 2207–2226, https://doi.org/10.5194/hess-20-2207-2016, 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.

Mendoza, P. A., Clark, M., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resour. Res., 51, 716–728, https://doi.org/10.1002/2014WR015820., 2015.

Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, Water Resour. Res., 47, W02531, https://doi.org/10.1029/2010WR009505, 2011.

Metin, A. D., Dung, N. V., Schröter, K., Guse, B., Apel, H., Kreibich, H., Vorogushyn, S., and Merz, B.: How do changes along the risk chain affect flood risk?, Nat. Hazards Earth Syst. Sci., 18, 3089–3108, https://doi.org/10.5194/nhess-18-3089-2018, 2018.

Milly, P., Betancourt, J., Falkenmark, M., Hirsch, R., Kundzewicz, Z., Lettenmaier, D., and Stouffer, R.: Stationarity is dead: Whither water management?, Science, 319, 573–574, https://doi.org/10.1126/science.1151915, 2008.

Mizukami, N., Clark, M. P., Gutmann, E. D., Mendoza, P. A., Newman, A. J., Nijssen, B., Livneh, B., Hay, L. E., Arnold, J. R., and Brekke, L. D.: Implications of the Methodological Choices for Hydrologic Portrayals of Climate Change over the Contiguous United States: Statistically Downscaled Forcing Data and Hydrologic Model, J. Hydrometeor., 17, 73–98, https://doi.org/10.1175/JHM-D-14-0187.1, 2016.

National Weather Service: II.3-SAC-SMA Conceptualization of the Sacramento soil moisture accounting model, Tech. rep., National Oceanic and Atmospheric Administration (NOAA), 2002.

Newman, A., Sampson, K., Clark, M., Bock, A., Viger, R., and Blodgett, D.: A large sample watershed-scale hydrometeorological dataset for the contiguous USA, Boulder, CO: UCAR/NCAR, https://doi.org/10.5065/D6MW2F4D, 2014.

Newman, A., Clark, M., Sampson, K., Wood, A., Hay, L., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrol. Process., 21, 435–446, https://doi.org/10.1002/hyp.6253, 2007.

Peleg, N., Shamir, E., Georgakakos, K., and Morin, E.: A framework for assessing hydrological regime sensitivity to climate change in a convective rainfall environment: a case study of two medium-sized eastern Mediterranean catchments, Israel, Hydro. Earth Syst. Sci., 19, 567–581, https://doi.org/10.5194/hess-19-567-2015, 2015.

Pfannerstill, M., Guse, B., Reusser, D., and Fohrer, N.: Process verification of a hydrological model using a temporal parameter sensitivity analysis, Hydrol. Earth Syst. Sci., 19, 4365–4376, https://doi.org/10.5194/hess-19-4365-2015, 2015.

Pianosi, F., Beven, K., Freer, J., Hall, J., Rougier, J., Stephanson, D., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, Environ. Model. Softw., 79, 214–232, https://doi.org/10.1016/j.envsoft.2016.02.008, 2016.

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., and Uijlenhoet, R.: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, Water Resour. Res., 50, 409–426, https://doi.org/10.1002/2013WR014063, 2014.

Razavi, S. and Gupta, H.: What do we mean by sensitivity analysis? The need for comprehensive characterization of global sensitivity in earth and environmental systems models, Water Resour. Res., 51, 3070–3092, https://doi.org/10.1002/2014WR016527, 2015.

Reusser, D. E., Buytaert, W., and Zehe, E.: Temporal dynamics of model parameter sensitivity for computationally expensive models with FAST (Fourier Amplitude Sensitivity Test), Water Resour. Res., 47(7), doi:10.1029/2010WR009 947, 2011.

Sagarika, S., Kalra, A., and Ahmad, S.: Evaluating the effect of persistence on long-term trends and analyzing step changes in streamflows of the continental United States, J. Hydrol., 517, 36–53, https://doi.org/j.jhydrol.2014.05.002, 2014.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Sensitivity analysis practices: Strategies for model-based inference, Reliability Engineering & System Safety, 91 (10-11), 1109–1125, 2006.

Sheffield, J., Barrett, A., Colle, B., Fernando, D., Fu, R., Geil, K., Hu, Q., Kinter, J., Kumar, S., Langenbrunner, B., Lombardo, K., Long, L., Maloney, E., Mariotti, A., Meyerson, J., Mo, K., Neelin, J., Nigam, S., Pan, Z., Ren, T., Ruiz-Barradas, A., Serra, Y., Seth, A., Thibeault, J., Stroeve, J., Yang, Z., and Yin, L.: North American Climate in CMIP5 Experiments. Part I: Evaluation of Historical Simulations of Continental and Regional Climatology, J. Climate, 26, 9209–9244, https://doi.org/10.1175/JCLI-D-12-00592.1, 2013.

Shin, M., Guillaume, J. H. A., Croke, B. F. W., and Jakeman, A. J.: Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R, J. Hydrol., 503, 135–152, https://doi.org/10.1016/j.jhydrol.2013.08.047, 2013.

Stewart, I. T., R.Cayan, D., and Dettinger, M. D.: Changes toward earlier streamflow timing across western North America, J. Climate, 18, 1136–1155, https://doi.org/10.1175/JCLI3321.1, 2005.

Teutschbein, C., Wetterhall, F., and Seibert, J.: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale, Climate Dynamics, 37, 2087–2105, https://doi.org/10.1007/s00382-010-0979-8, 2011.

Thirel, G., Andréassian, V., and Perrin, C.: On the need to test hydrological models under changing conditions, Hydr. Sci. J., 60, 1165–1173, https://doi.org/10.1080/02626667.2015.1050027, 2015.

Uhlenbrook, S., Seibert, J., Leibundgut, C., and Rodhe, A.: Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure, Hydrolog. Sci. J., 44, 779–797, 1999.

Van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Characterization of watershed model behavior across a hydroclimatic gradient, Water Resour. Res., 44, https://doi.org/10.1029/2007WR006271, 2008.

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Rainfall characteristics define the value of streamflow observations for distributed watershed model identification, Geophys. Res. Let., 35, L11 403, https://doi.org/10.1029/2008GL034162, 2008.

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Sensitivity-guided reduction of parameteric dimensionality for multi-objective calibration of watershed models, Adv. Water Resour., 32, 1154–1169, https://doi.org/10.1016/j.advwatres.2009.03.002, 2009.

Vaze, J., Post, D., Chiew, F., Perraud, J., Viney, N., and Teng, J.: Climate non-stationarity - Validity of calibrated rainfall-runoff models for use in climate change studies, J. Hydrol., 394, 447–457, https://doi.org/10.1016/j.jhydrol.2010.09.018, 2010.

Westra, S., Thyer, M., Leonard, M., Kavetski, D., and Lambert, M.: A strategy for diagnosing and interpreting hydrological modelnonstationarity, Water Resour. Res., 50, 5090–5113, https://doi.org/10.1002/2013WR014719, 2014.

Wetterhall, F., Graham, L., Andreasson, J., Rosberg, J., and Yang, W.: Using ensemble climate projections to assess probabilistic hydrological change in the Nordic region, Nat. Hazards Earth Syst. Sci., 11, 2295–2306, https://doi.org/10.5194/nhess-11-2295-2011, 2011.

Willmott, C. and Fedema, J.: A more rational climatic moisture index, The Professional Geographer, 44, 84–88, https://doi.org/10.1111/j.0033-0124.1992.00084.x, 1992.

Wood, A., Leung, L., Sridhar, V., and Lettenmaier., D.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model output, Clim. Change, 15, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e, 2004.

Wu, H., Kimball, J. S., Li, H., Huang, M., Ruby Leung, L., and Adler, R. F.: A new global river network database for macroscale hydrologic modeling, Water Resour. Res., 48, W09701, https://doi.org/10.1029/2012WR012313, 2012.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, J. Hydrol., 181, 23–48, https://doi.org/10.1016/0022-1694(95)02918-4, 1996.

Zadeh, F. K., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., and Bauwens, W.: Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model, J. Env. Mod. Softw., 91, 210–222, https://doi.org/10.1016/j.envsoft.2017.02.001, 2017.