Reviewer #2

This manuscript tries to investigate the uncertainties resulting from different hydrological model components when assessing the impacts of climate change on streamflow. To do so, they design a modeling framework that incorporates three runoff yield schemes, one runoff routing scheme, several GCM and RCP. I think the topic is interesting and the manuscript is overall well-prepared. However, I think there are still several issues have to be addressed before considering for publication in HESS.

The authors choose annual mean discharge, annual peak discharge or 100-yr flood discharge to analyze the uncertainties. I doubt if it's meaningful to investigate annual mean values in a 750 km2 catchment. In figure 5 they even investigate changes and uncertainties in much smaller sub-basin. Because I think, according to their methodology, in such a small catchment the annual mean runoff is simply controlled by precipitation and evaporation. On the other hand, when investigate the annual peak values (here it's not clear how they define 'peak' values, from daily or hourly?), the routing may play a more significant role in the timing and magnitude of simulated streamflow. My concern is if the authors can still reach the same conclusions if they use daily streamflow when perform uncertainties analysis because I believe in such small catchment different runoff yield schemes have more effects on daily streamflow instead annual streamflow.

Response: The annual mean discharge was defined as the average of daily streamflow in a year. To clarify it, we have inserted the following sentence in the manuscript:
"*Here, the annual mean discharge was defined as the average of daily streamflow in a year*."
The annual peak discharge was defined as the maximum daily streamflow in a year. It was described in L322-323: "*annual maximum daily discharge* (*Q_p*)"

I also want to hear opinions from the authors regarding the choose of runoff yield scheme. When perform regional or global simulations using LSM, people usually can only use one runoff yield option, either saturation-excess (e.g. NoahMP, CLM) or infiltration-excess (e.g. VIC). However, when focus on the specific catchment, you can definitely choose a runoff yield scheme that is suitable for the hydrological regime of that catchment. I'm not challenging your work, just want to hear some discussion.

Response: This is a good point. One of the main objectives of this study was to "*evaluate and compare the performance of hydrologic models with different approaches representing runoff generation process…*" (L111-113). The results in this study showed that STP performs better than the other two methods. This finding can be informative for future studies associated with hydrologic model selection. We have inserted the following discussion in the manuscript:

"*This study can also provide useful information for selecting hydrologic models for climate change impact analysis. As discussed in section 3.1, the STP-HRR model is more suitable than the other two models for the study region, mainly due to its ability to represent the non-linear hydrological response to precipitation forcings. This implies hydrologic models adopting the saturation excess runoff generation algorithms may be more suitable for areas with a Mediterranean climate. The uncertainties from hydrologic models are larger than those from the hydrologic model parameters for all hydrologic variables (e.g., discharge, runoff and seasonality), suggesting the inter-model variability is larger than the intra-model variability* *(from model parameters). This implies that model selection is more important than the parameter selection, and that the parameter equifinality (or non-uniqueness) is less of a concern when quantifying climate change impacts on hydrologic fluxes when using an ensemble of GCM forcings. In this study, only the runoff generation algorithm was investigated. Other hydrologic model components, such as ET algorithm and routing method, also have many variants. The choice of these components can also make a difference in the total uncertainties in simulated runoff and streamflow. Therefore, further study integrating different algorithms for these components can be conducted in the future. This complete analysis can be useful to guide stakeholders to select appropriate hydrologic algorithms for climate change impacts analysis and to develop actionable adaptation and mitigation strategies.*"

Line 134-141. The authors use MODIS products to estimate the PET, However, they don't provide any detail regarding how to convert PET into ET for runoff yield simulation. In eq(1)~(7) I don't see any variable related to ET.

Response: The ET was extracted from soil at each time step, and then the soil content was updated which was used for water balance calculation in next time step. We have added the following content to the manuscript:

"The evapotranspiration was estimated using Eq. S15…. After the water fluxes (runoff, ET and water movement between soil layers) were determined, the soil moisture was updated which would be used for the water balance calculation in the next time step."

$$ET = min\ (PET, W - W_{min}) \qquad (S15)$$

where PET is the potential evapotranspiration estimated using the method proposed by Raoufi and Beighley (2017); W is water content in the upper soil layer; W_{min} is the minimum water content in the soil, defined as 0.15× W_s; W_s is soil water content as saturation.

Line 255. The authors calibrate several parameters related to runoff. But they don't document how they fix the value of soil depth, from dataset or by calibration. In Line 233 they state that the soil depth is based on a previous study but I don't see any description in (Feng et al., 2019). In their modeling framework, they use quite simple water balance scheme to account for the soil water movement, in this case the soil depth is an important variable determining the soil water holding capacity.

Response: The soil depth data was originally from the Soil Survey Geographic (SSURGO) Data Base for Santa Barbara County. This reference has been added to the manuscript.

Line 256. sim-topmodel uses groundwater depth to calculate runoff yield. Do you spin up the model to reach the equilibrium state?

Response: Yes. We did spin up the model for one year. The following text has been inserted to the manuscript L212:

"The models spun up for one year to ensure the equilibrium status."

Line 290. If I understand correctly, here should be "parameter", which is different from "parameterization

Response: Thanks for pointing it out. We have corrected it.

This paper presents some limited results of evaluating the impact of different formulations in runoff generation schemes when simulating streamflow. My major objection with the paper is that it really is not assessing the uncertainty but rather the variability of the simulated streamflow and how each of the forcings, model parameters or formulations contribute to it. Although that is valuable in itself, the authors claim that the objective is to identify the uncertainties in the context of climate change simulations. However, that is not what was done here. The calibration of the parameters was done using an observation-based forcing dataset and although I can understand the rationale, I believe that any calibration of parameters should have been done in a way that would emulate the intended application (i.e. using GCM output to drive the hydrology model). I believe historical simulation are available from CMIP5 and if so they should be used to evaluate the actual uncertainty of simulated streamflow within the framework that the authors have developed. The end of 21st century simulations should be a final experiment (if included at all) given the objective of the paper. Consequently, I recommend major revisions before publication that will include new simulations that test the different model parameter sets, runoff generation schemes and downscaled GCM output for the period when streamflow measurements are available, so that the actual uncertainty can be quantified. In addition, I believe the study area is rather limited and an opportunity is being missed by not including additional basins with different physiography and climate.

Response: In this study, we did use GCM simulations as the forcings of hydrologic models for the historical period. For each simulation scenario (i.e., the combination of hydrologic model, parameter set, GCM and RCP), we simulated runoff and discharge using GCM outputs for both historical and future periods, and then the relative changes (%) between future and historical simulations were quantified. The total uncertainty in these projected changes from all model combinations (3 hydrologic models, 3 parameter sets, 10 GCMs and 2 RCPs; 3*3*10*2=180) was quantified. The uncertainty contributions to the total uncertainty were then quantified using the ANOVA method.

To clarify it, we have inserted the following text to the manuscript:

"Here, we used GCM outputs as the forcings of hydrologic models for both historical (1986-2005) and future (2081-2100) periods. For each simulation scenario (i.e., the combination of hydrologic model, parameter set, GCM and RCP), the historical and future daily streamflow and monthly runoff were simulated, and the relative changes (%) were quantified."

To deepen this study, we have also expanded the analysis by including more metrics about the volume and composite of runoff (i.e., monthly surface, subsurface and total runoff), as well as the hydrologic seasonality (wet season length and timing of wet season onset), considering these quantities are of great importance for the study region (Myers et al., 2019;Feng et al., 2019).

We have added the following figures and texts in the manuscript:

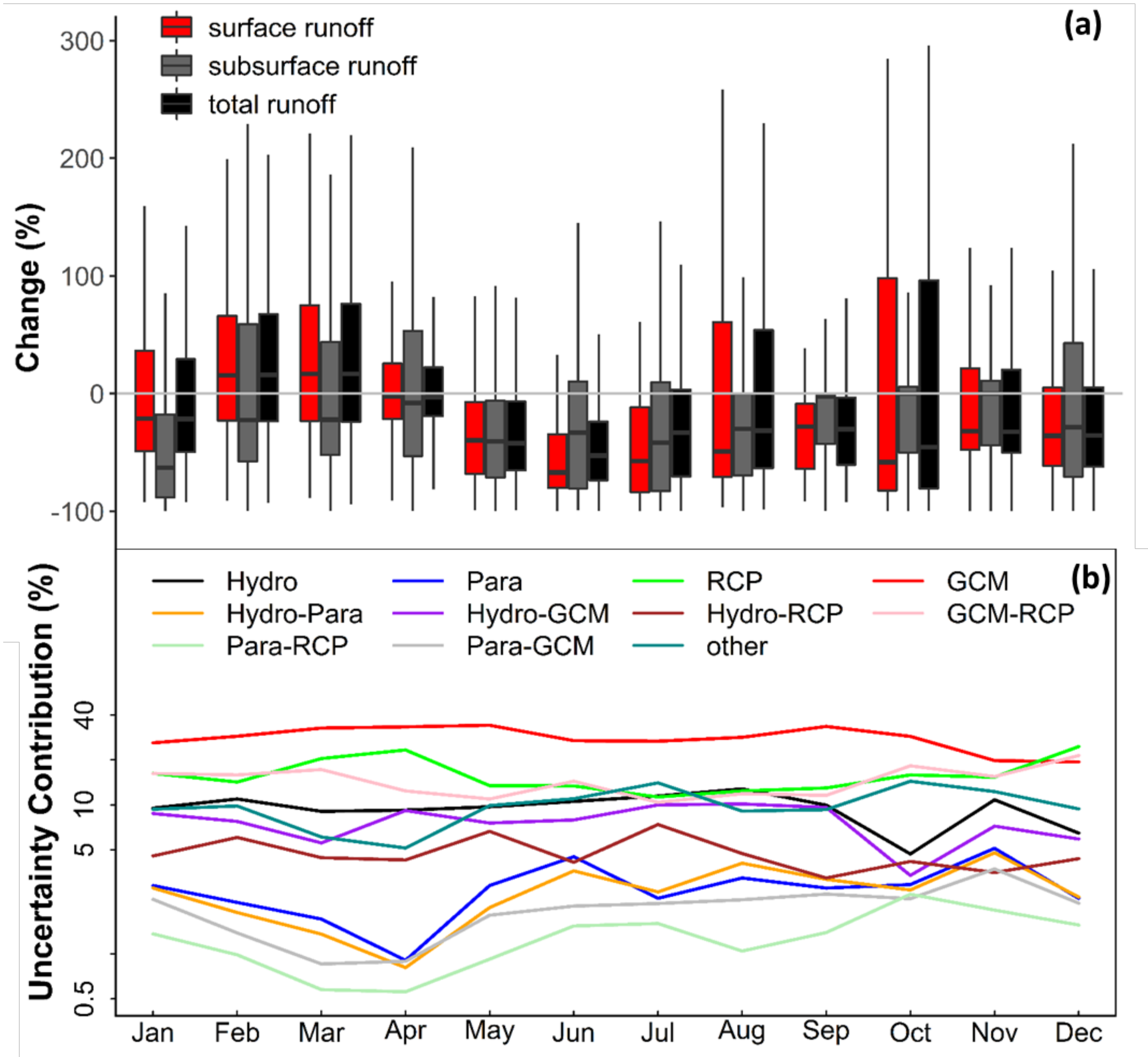

Figure 6: (a) Projected relative changes (%) in monthly surface runoff, subsurface runoff and total runoff in the whole study region during 2081-2100 as compared to historical period (1986-2005); (b) Relative contributions (%) of the uncertainties for the projected changes in the monthly total runoff; Hydro = Hydrologic models; Para = hydrologic model parameters; GCM = General Circulation Models; RCP = Representative concentration pathways (emission scenarios); "other" is the uncertainty from the 3rd and 4th orders of interactions between the 4 major sources (i.e., GCMs, RCPs, Hydrologic models and parameters).

"The projected changes in monthly runoff (surface, subsurface and total) during 2081-2100 compared to 1986-2005 range between -100% and 300% (Figure 6a). Surface runoff will probably increase in February and March, and decrease in other months (Figure 6a). This is because in the future, the onset of wet season will be delayed and more severe storm events will

occur during the shorter wet season (Feng et al., 2019). The decrease in subsurface runoff in all months is probably because the decrease in the frequency (or total number) of storm events (Feng et al., 2019). The changes of monthly total runoff show similar pattern with the surface runoff, suggesting the more pronounced changes in surface runoff as compared to subsurface runoff. The major uncertainty sources are GCM and RCP, which account for ~45% of total uncertainty (Figure 6b). Hydrologic models contribute to ~10% of total uncertainty (Figure 6b). This suggests that the climate patterns (e.g., storm event frequency and intensity) are more important factors controlling the runoff generation than the hydrologic model algorithms."

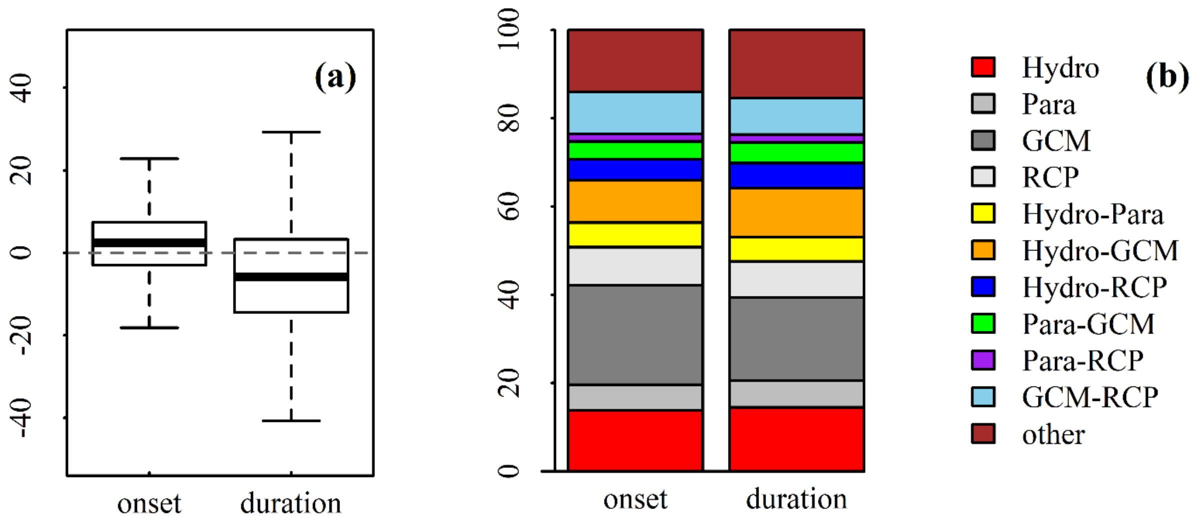

Figure 9: **(a)** Projected change (days) in the onset and duration of wet season in SBC; positive (negative) values indicate later (earlier) onset or longer (shorter) duration of the wet season; **(b)** relative contributions (%) of the uncertainties of the projected changes in seasonality. Hydro = Hydrologic models; Para = hydrologic model parameters; GCM = General Circulation Models; RCP = Representative concentration pathways (emission scenarios); "other" is the uncertainty from the 3rd and 4th orders of interactions between the 4 major sources (i.e., GCMs, RCPs, Hydrologic models and parameters).

"Consistent with the work of Feng et al. (2019), this study suggests a delayed onset and shorter duration of wet season (Figure 9a). The median changes show that the wet season will start later by 3 days, and become shorter by ~6 days. The major uncertainty sources for both onset and duration of wet season are GCM (~20%) and hydrologic models (~15%). Different from discharge and runoff, the seasonality shows more uncertainty from hydrological models (15% vs 12%) and model parameters (~6% vs 2%) (Figure 9b). This is because the seasonality integrates the runoff generation, paths and transport processes for both surface and subsurface runoff, which are important for the timing and quantity of simulated discharge."

Some additional comments are outlined below:

* How does the uncertainties in the prescribed ET affect the results? Why weren't they accounted for?

Response: We agree that the uncertainty from ET models will affect the results. However, the main focus of this study is to investigate the uncertainty contributions of runoff generation schemes and associated parameters for the climate change impact assessment. Therefore, we used the same ET method and routing algorithm for all three hydrologic models. To clarify it, we have added the following text in the discussion section:

"Compared to previous studies (e.g., Vetter et al. (2015), Schewe et al. (2014), Hagemann et al. (2013);(Troin et al., 2018), and Asadieh and Krakauer (2017)), this work identifies relatively low uncertainty contributions from hydrologic models. The main reason for this is that the hydrologic model uncertainty in this study was only from runoff generation algorithms and associated parameters. As is, the three hydrologic models share common algorithms for ET and plane/channel routing, and the same model configuration (e.g., soil matrix and model unit definition). These similarities among models likely reduced the differences in simulated runoff and discharge. In addition, the uniform calibration approach and parameter selection criteria were also likely to eliminate user/method bias which is common in studies that consider more than one hydrologic model. In contrast, the hydrologic models used in previous studies have their own model configurations, and ET and routing algorithms. For example, the VIC model (here VIC refers to the original VIC model, and is different from the model used in this study; to clarify, in following text, VIC refers to the original VIC model while VIC-HRR refers to the model used in this study) applies an ET algorithm different from the one used in this study (Raoufi and Beighley, 2017), uses the grid-based model units ignoring the spatial arrangement, and has its own routing scheme which adopts the synthetic unit hydrograph concept. These differences between models likely resulted in the larger uncertainties in the simulation from hydrologic models in previous studies.

This study can also provide useful information for selecting hydrologic models for climate change impact analysis. As discussed in section 3.1, the STP-HRR model is more suitable than the other two models for the study region, mainly due to its ability to represent the non-linear hydrological response to precipitation forcings. This implies hydrologic models adopting the saturation excess runoff generation algorithms may be more suitable for areas with a Mediterranean climate. The uncertainties from hydrologic models are larger than those from the hydrologic model parameters for all hydrologic variables (e.g., discharge, runoff and seasonality), suggesting the inter-model variability is larger than the intra-model variability (from model parameters). This implies that model selection is more important than the parameter selection, and that the parameter equifinality (or non-uniqueness) is less of a concern when quantifying climate change impacts on hydrologic fluxes when using an ensemble of GCM forcings. In this study, only the runoff generation algorithm was investigated. Other hydrologic model components, such as ET algorithm and routing method, also have many variants. The choice of these components can also make a difference in the total uncertainties in simulated runoff and streamflow. Therefore, further study integrating different algorithms for these components can be conducted in the future. This complete analysis can be useful to guide *stakeholders to select appropriate hydrologic algorithms for climate change impacts analysis and to develop actionable adaptation and mitigation strategies."*

* Abstract needs some attention, especially after l. 21 in terms of cohesiveness. Right now, it reads as bullet points stitched together. Some proofreading needed for redundant articles and grammatical errors.

Response: We have modified the abstract as follows. We have also taken a careful proofreading for the whole manuscript.

"Assessing the impacts of climate change on hydrologic systems is critical for developing adaptation and mitigation strategies for water resource management, risk control and ecosystem conservation practices. Such assessments are commonly accomplished using outputs from a hydrologic model forced with future precipitation and temperature projections. The algorithms used for the hydrologic model components (e.g., runoff generation) can introduce significant uncertainties in the simulated hydrologic variables. Here, a modeling framework was developed that integrates multiple runoff generation algorithms with a routing model and associated parameter optimizations. This framework is able to identify uncertainties from both hydrologic model components and climate forcings as well as associated parameterization. Three fundamentally different runoff generation approaches: runoff coefficient method (RCM, conceptual), variable infiltration capacity (VIC, physically-based, infiltration excess) and simple-TOPMODEL (STP, physically-based, saturation excess), were coupled with the Hillslope River Routing model to simulate surface/subsurface runoff and streamflow. A case study conducted in Santa Barbara County, California, reveals increased surface runoff in February and March while decrease in other months, a delayed (3 days, median) and shortened (6 days, median) wet season, and increased daily discharge especially for the extremes (e.g., 100-yr flood discharge, Q100). The uncertainties of the projected changes in these hydrologic variables are large (e.g., 400% for monthly runoff and 340% for Q100). For runoff and discharge, general circulation models (GCMs) and emission scenarios are two major uncertainty sources, accounting for about half of the total uncertainty. For the changes in seasonality, GCMs and hydrologic models are two major uncertainty contributors (~35%). In contrast, the contribution of hydrologic model parameterization to the total uncertainty of changes in these hydrologic variables is relatively small (<6%), limiting the impacts of hydrologic model parameter equifinality in climate change impact analysis. This study also provides insights on how to optimize hydrologic model selection for projecting future hydrologic conditions."

* l. 53: what is the need for naming the "land-atmosphere interactions" as "runoff generation process" when the latter is clearly one of the processes that manifest from those interactions?

Response: We have modified the text to following:

"Generally, hydrologic models have modules simulating water partitioning at land surface (named as runoff generation process in this study), evapotranspiration, and water transportation along terrestrial hillslopes and channels (named as routing process here)."

* l. 175- : Not sure whether this much detail is needed for the description of the runoff generation models, since they are well established.

Response: We have moved the text associated with runoff generation models to the supporting information.

* l. 354: does that mean that there is bias in the validation data (i.e. streamflow)?

Response: Yes, we think so. The typical uncertainty for streamflow gauge data is 6%-19% in small watershed based on previous studies (e.g., Harmel et al., 2006). The work of Beighley et al. (2003) also identified the bias in the 1995 January event.

We discussed this in L363-368:

"The uncertainties in gauge measurements can also be a bias source. For example, in typical conditions the uncertainty in streamflow measurements ranges between 6%-19% in small watersheds, but it can be higher during large storm events when accurate stage measurements are more difficult (Harmel et al., 2006). Beighley et al. (2003) also identified the overestimation of gauge records for the 1995 January event at Gauge 11119940."

* l. 362-363: this highlights another problem that has not been addressed in this study: the downscaling of GCM outputs to drive the hydrology model.

Response: We agree that the uncertainty from the downscaling of GCM outputs can impact the results. The focus of this study is the uncertainty contribution from the runoff generation models and associated parameters. Therefore, we didn't include different downscaling methods and quantify their uncertainties. However, we do think it is a great idea, and we may investigate it in our future studies.

References:

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., and Seibert, J.: Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments, Water Resources Research, 50, 7541-7562, 10.1002/2014wr015549, 2014.
Asadieh, B., and Krakauer, N. Y.: Global change in streamflow extremes under climate change over the 21st century, Hydrology and Earth System Sciences, 21, 5863, 2017.
Beighley, R. E., Melack, J. M., and Dunne, T.: Impacts of California's climatic regimes and coastal land use change on streamflow characteristics, JAWRA Journal of the American Water Resources Association, 39, 1419-1433, 10.1111/j.1752-1688.2003.tb04428.x, 2003.
Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., Kao, S.-C., 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, 10.1029/2018ef001047, 2019.
Feng, D., Beighley, E., Raoufi, R., Melack, J., Zhao, Y., Iacobellis, S., and Cayan, D.: Propagation of future climate conditions into hydrologic response from coastal southern California watersheds, Climatic Change, 153, 199-218, 10.1007/s10584-019-02371-3, 2019.
Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., and Wiltshire, A. J.: Climate change impact on available water resources obtained using multiple global climate and hydrology models, Earth Syst. Dynam., 4, 10.5194/esd-4-129-2013, 2013.

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Flörke, F., Gosling, S. N., Hoffmann, P., Liersch, S., Masaki, Y., Motovilov, Y., Müller, C., Samaniego, L., Stacke, T., Wada, Y., Yang, T., and Krysnaova, V.: Sources of uncertainty in hydrological climate impact assessment: a cross-scale study, Environmental Research Letters, 13, 015006, 10.1088/1748-9326/aa9938, 2018.

Myers, M. R., Barnard, P. L., Beighley, E., Cayan, D. R., Dugan, J. E., Feng, D., Hubbard, D. M., Iacobellis, S. F., Melack, J. M., and Page, H. M.: A multidisciplinary coastal vulnerability assessment for local government focused on ecosystems, Santa Barbara area, California, Ocean & Coastal Management, 104921, https://doi.org/10.1016/j.ocecoaman.2019.104921, 2019.

Raoufi, R., and Beighley, E.: Estimating daily global evapotranspiration using penman–monteith equation and remotely sensed land surface temperature, Remote Sensing, 9, 1138, 2017.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, Proceedings of the National Academy of Sciences, 111, 3245-3250, 10.1073/pnas.1222460110, 2014.

Troin, M., Arsenault, R., Martel, J.-L., and Brissette, F.: Uncertainty of Hydrological Model Components in Climate Change Studies over Two Nordic Quebec Catchments, Journal of Hydrometeorology, 19, 27-46, 10.1175/jhm-d-17-0002.1, 2018.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, Earth System Dynamics, 6, 17, 2015.