Climatic Change
https://doi.org/10.1007/s10584-020-02858-4

Grappling with uncertainties in physical climate impact projections of water resources

Rutger Dankers[1] · Zbigniew W. Kundzewicz[2]

Received: 6 December 2019 / Accepted: 30 August 2020/Published online: 07 October 2020

Abstract
This paper reviews the sources of uncertainty in physical climate impact assessments. It draws on examples from related fields such as climate modelling and numerical weather prediction in discussing how to interpret the results of multi-model ensembles and the role of model evaluation. Using large-scale, multi-model simulations of hydrological extremes as an example, we demonstrate how large uncertainty at the local scale does not preclude more robust conclusions at the global scale. Finally, some recommendations are made: climate impact studies should be clear about the questions they want to address, transparent about the uncertainties involved, and honest about the assumptions being made.

Keywords Physical climate impact projections · Uncertainty · Multi-model ensembles

1 Introduction

Ongoing climate change will affect virtually all sectors of our society in multiple, complex, ways. Assessment of these expected impacts has been an integral part of the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC) since the beginning (IPCC 1990). Increasingly, these impact projections have been based on impact model (IM) simulations. Ranging from simple statistical models to sophisticated process-based models that can be used across multiple sectors, IMs are typically driven by the meteorological output of global or regional climate models and effectively translate the projection of a future climate into a projection of an impact.

This article is part of a Special Issue on "How evaluation of hydrological models influences results of climate impact assessment," edited by Valentina Krysanova, Fred Hattermann, and Zbigniew Kundzewicz

✉ Rutger Dankers
rutger.dankers@wur.nl

[1] Wageningen Environmental Research, Wageningen University & Research, Postbus 47, 6700 AA Wageningen, The Netherlands

[2] Institute for Agricultural and Forest Environment, Polish Academy of Sciences, Bukowska 19, 60-809 Poznan, Poland

Of course, running an additional impact model will often, although not necessarily always, add to the overall uncertainty that is inherent to climate change projections (Zscheischler et al. 2018; Kundzewicz et al. 2018). Over the past decade or so, several impact model intercomparison projects (MIPs) have been initiated in an effort to explore some of this impact model uncertainty, and enable coordinated assessments of climate impacts within and across sectors. Examples of these include the Integrated Project Water and Global Change, WATCH[1] (Harding et al. 2011), the Agricultural Model Intercomparison and Improvement Project (AGMIP)[2], and the Inter-Sectoral Impact Model Intercomparison Project ISIMIP[3] (Warszawski et al. 2014). The analysis of the outcomes of these MIPs has raised new questions about how to interpret the results of these multi-model experiments. As we will see, the focus has increasingly been on model evaluation, with the implicit or explicit assumption that better IM performance over a past period, as measured by the comparison of simulation and observation, yields greater trust in its projections for the future.

Naturally, model evaluation is a key step in the development and application of any model of the environment. However the assumption that a model can somehow be "validated" by comparing its predictions with observations has also been challenged (see e.g., Oreskes et al. 1994; Parker 2013). The reasons for this include the notion that natural systems are complex, our understanding of them is incomplete, and the observation data we use are poorly defined and uncertain. Within the context of climate change, a specific problem is that future climate conditions can be very different from the historical climate. For example, a simple statistical model might accurately reproduce historical streamflow, but it would be unwise to apply the same model to a very different future climate while claiming it is trustworthy because it compares well with past observations.

In this paper we will review how multi-model ensembles are used to produce physical climate impact projections, focusing on global-scale hydrological models in particular. How water resources and streamflow characteristics will change into the future is arguably one of the most pertinent questions to the problem of the impact of climate change.

2 Types and sources of uncertainty

In this section we will briefly review the sources of uncertainty in climate impact assessments. A prime source of uncertainty—literally a lack of certainty, or precise knowledge—stems from the use of models. Since we cannot examine the behaviour of a catchment under future conditions in a laboratory, climate impact projections are by necessity model-based, and models are, inevitably, subject to considerable uncertainty (Oreskes et al. 1994). It is common to subdivide uncertainty within the modelling process into (1) uncertainty about the model structure or, in other words, about how to represent the physics of the system; (2) uncertainty about the input data and model parameter values, which extends to the data used for model calibration and evaluation; and (3) the residual unpredictability of events for given models and parameters.

The first two sources can be taken together as "epistemic" uncertainty (Beven 2016) (after the Greek word for "knowledge") and arises from the fact that models are an abstraction of

[1] See www.eu-watch.org

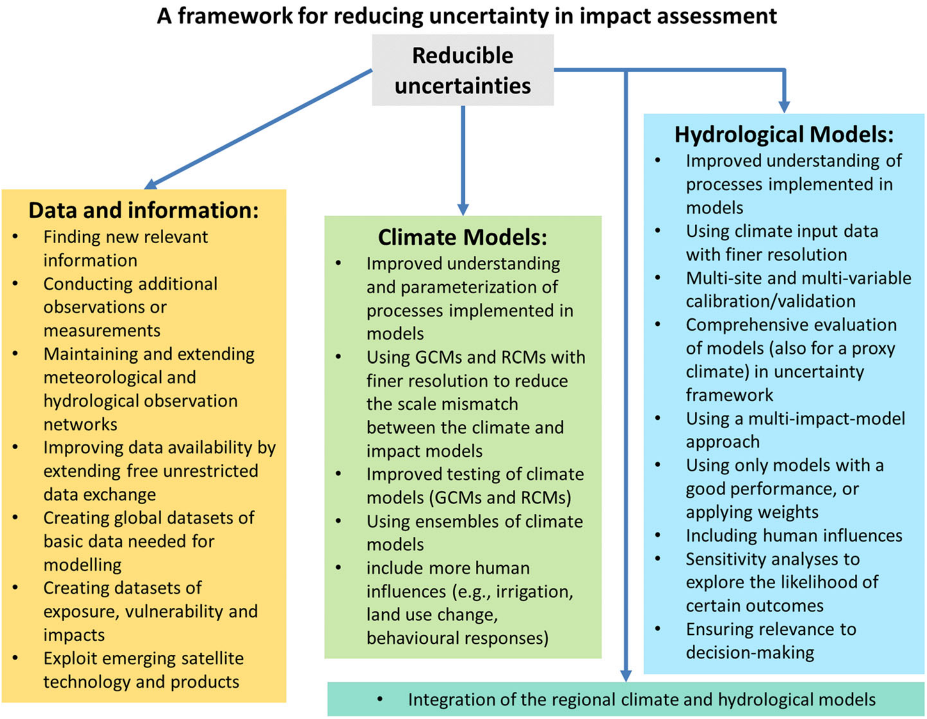
[2] See https://agmip.org/

[3] See https://www.isimip.org

reality. Some of the key processes are still not well understood or represented. Limits to computing power and data availability (e.g., on soil properties) mean that sometimes processes need to be represented in a simplified way, or are missing altogether; for example, groundwater is missing in many models or is parameterized very simplistically. Often it is not clear which representation, and which parameter values, would be most suitable for the application at hand.

The latter source of uncertainty is sometimes called “aleatory” uncertainty (from the Latin word for dice) and arises from natural variability or randomness that essentially cannot be reduced. Although the distinction between epistemic and aleatory uncertainty is not always easy to make in practice, it is still useful framework to keep in mind. In particular, the existence of epistemic uncertainties implies that model errors may not always follow a simple statistical distribution (Beven 2013).

For climate impact assessments, we also need to consider the entire modelling chain, from socio-economic emission scenarios to climate models, including climate model downscaling and bias correction, to impact models, impact assessments, and adaptation decisions. Each component in this “cascade” will have its own associated uncertainties (Beven et al. 2018). Kundzewicz et al. (2018) comprehensively discuss the sources of uncertainty in the emission – climate change – impact chain. Some of these uncertainties might be reducible, that is, by adding new information to the process the range of possible outcomes could be constrained (see Fig. 1).



A framework for reducing uncertainty in impact assessment

Reducible uncertainties

Data and information:

- Finding new relevant information
- Conducting additional observations or measurements
- Maintaining and extending meteorological and hydrological observation networks
- Improving data availability by extending free unrestricted data exchange
- Creating global datasets of basic data needed for modelling
- Creating datasets of exposure, vulnerability and impacts
- Exploit emerging satellite technology and products

Climate Models:

- Improved understanding and parameterization of processes implemented in models
- Using GCMs and RCMs with finer resolution to reduce the scale mismatch between the climate and impact models
- Improved testing of climate models (GCMs and RCMs)
- Using ensembles of climate models
- include more human influences (e.g., irrigation, land use change, behavioural responses)

Hydrological Models:

- Improved understanding of processes implemented in models
- Using climate input data with finer resolution
- Multi-site and multi-variable calibration/validation
- Comprehensive evaluation of models (also for a proxy climate) in uncertainty framework
- Using a multi-impact-model approach
- Using only models with a good performance, or applying weights
- Including human influences
- Sensitivity analyses to explore the likelihood of certain outcomes
- Ensuring relevance to decision-making

- Integration of the regional climate and hydrological models

Fig. 1 A general framework for reducing uncertainty in assessment of climate change impact on water resources (modified from Kundzewicz et al. 2018)

Climate projections are uncertain first of all because we are uncertain about future levels of greenhouse gas emissions and concentrations. Projections of long-term climate change are therefore based on a set of assumptions, or scenarios, about how these factors will develop, such as the Representative Concentration Pathways (RCPs) (Moss et al. 2010). Strictly speaking, the RCPs should not be interpreted as forecasts and no likelihood or preference is attached to individual scenarios (van Vuuren et al. 2011). Instead, multiple plausible futures need to be considered (see Maier et al. 2016). Climate projections, and by extension projections of impacts, are therefore not predictions in the classical sense, but rather scenario studies.

In the more immediate future, climate modelling uncertainties are more dominant than emission uncertainties, because near-term climate is strongly conditioned by past emissions (committed warming) (see, e.g., Hawkins and Sutton 2009, 2011). At decadal timescales, climate predictions are also subject to initial condition uncertainty, where small errors in the initial state of the model can grow into marked differences in the development of the climate system (Suckling 2018). This source of uncertainty is particularly relevant in weather forecasting and seasonal to decadal climate prediction but also affects longer-term climate projections, especially when looking at climate variability and extremes.

Climate impact studies often select a subset of the available set of global climate models (GCMs) to provide the meteorological input for impact models. Especially at smaller scales GCMs can exhibit significant biases, and the application of downscaling techniques (either empirical-statistical or dynamic) and/or bias correction techniques is often required. Although the aim is often to reduce biases by bringing climate model output closer to observations, these techniques are, in effect, models in their own right based on assumptions that may or may not hold under future climate change. Some have therefore argued that these techniques hide rather than reduce the uncertainty (Ehret et al. 2012).

For impact modellers it may seem natural to assume that uncertainty from climate models dwarfs uncertainty due to impact models, but at least a number of studies have suggested that impact model uncertainty can also be significant (e.g., Haddeland et al. 2011; Vetter et al. 2017; Beck et al. 2017). However, few studies have made a thorough end-to-end assessment of the uncertainties involved, as was done for the entire flood risk chain by Metin et al. (2018).

Several recent papers discuss the issue of “deep uncertainty” in environmental modelling and risk assessment (Spiegelhalter and Riesch 2011; Beven et al. 2014, 2018). This issue resembles the concept of “vague uncertainties” of Budescu and Wallsten (1987) and similar concepts dating back to the 1920s (Knight 1921), and extends the notion of epistemic uncertainty in the model itself to include problems such as unknown inadequacies in the modelling process, and possible disagreements about the framing of the problem. An implication of deep uncertainty is that we may never be able to fully describe or “quantify” the uncertainty inherent to future climate change and its impacts.

3 Evaluation of multi-model ensembles

These days multi-model ensembles have become the primary route to explore the uncertainty in projections of climate change and its impacts. The impact MIPs that have been initiated over the last decade (Harding et al. 2011; Warszawski et al. 2014) followed the example of similar initiatives in the climate modelling community, and many of the questions that surround the interpretation of climate model ensembles (see Parker 2013) also apply to ensembles of impact models. Can we use them to infer probabilities of future climate impacts? Are robust findings especially trustworthy? As decisions on climate-change mitigation and adaptation are potentially influenced by the outcomes of these ensemble studies, answering questions such as these is not purely an academic exercise (Parker 2013).

A common finding is that different models, whether hydrological models or land surface schemes, lead to different results, in other words, impact model uncertainty is a significant component in the overall uncertainty (e.g., Haddeland et al. 2011). Although this result mirrors what has been found in the climate modelling community, it may still come as a surprise to some, considering that hydrological models are less complex than climate models.

3.1 Sources of bias

Disentangling what is causing these differences between models has proved a greater challenge. Key uncertain processes appear to be evapotranspiration (ET) and snow accumulation and melt. Haddeland et al. (2011) noted that Land Surface Models (LSMs) generally simulated lower snow accumulation and melt than Global Hydrological Models (GHMs). While GHMs typically use a conceptual degree-day approach to simulate snow accumulation and melt, LSMs normally include a more complex energy balance scheme that also simulates snow sublimation, which explains some of the differences. Similarly, Beck et al. (2017) found that in regions dominated by snow GHMs performed better than LSMs, which they ascribed to more data-demanding snow routines or misrepresentation of frozen soil and snowmelt processes by the LSMs.

Haddeland et al. (2011) noted that models calculating evapotranspiration based only on temperature yielded different results than those that included radiation and humidity, and differences tended to be smaller in wet climates than in dry climates. As a consequence, while the largest absolute differences in simulated runoff were found in the tropics, the largest relative differences occurred in arid areas, with models generally overpredicting runoff in arid and semiarid basins. Guo et al. (2017) found that also the scheme used in converting potential to actual evapotranspiration can have a major impact on the results, yielding a more than sevenfold difference in estimated runoff sensitivity.

An overestimation of runoff in dry basins was also found by Zaherpour et al. (2018). The majority of the models mostly overestimated the mean annual runoff and all indicators of upper and lower extreme runoff, and in particular low flow indicators. Capturing the seasonal dynamics of streamflow proved also difficult, with models struggling to get the timing right particularly in northern basins, while in southern areas the magnitude of the seasonal cycle was often more problematic (Zaherpour et al. 2018).

Few studies have mentioned the representation of soil moisture and groundwater dynamics as a source of model biases. Both are highly uncertain components of the overall water balance even from an observational perspective, and even the more physically based models usually simulate groundwater in a very simplified manner unlikely to resemble the actual processes. Improving groundwater processes in models, for example by assimilating satellite data, could improve hydrologic simulations (Lo et al. 2010; Koirala et al. 2014).

A major obstacle to understanding which schemes for evapotranspiration, snow, or soil moisture perform better under what conditions, is a lack of suitable observation datasets to evaluate these processes separately. Most studies evaluate the model performance with observations of river discharge, and infer deficiencies in the model. However, biases in the simulated flow may be caused by a number of factors, including biases in the meteorological input data (see e.g., Haddeland et al. 2012; Müller Schmied et al. 2016) or a lack of understanding of the soil properties, and even in simple models these causes may interact in complex ways. Since discharge is an integrated measure of processes over the entire basin, some of these biases may counter each other, leading to plausible results. Good performance in simulating discharge at the catchment outlet therefore does not guarantee that all processes in the basin have been represented realistically.

Few studies have attempted to evaluate large-scale hydrological models more broadly (see overviews in Zaherpour et al. (2018) and in Krysanova et al., in this SI). Zhang et al. (2016) used observations on evapotranspiration from around the world, as well as observations of streamflow in evaluating two different models. They found that the ET simulated by the models compared better with the observations than runoff, with runoff biases typically, but not always, being the opposite of biases in ET. An important caveat, though, is the limited number of years of ET data that was available, often not located in the same catchment as the streamflow observations.

So while there is potential for different types of observations to be used in model evaluation, limitations in the spatial and/or temporal coverage, measurement uncertainties, and potentially even conceptual differences between the variables in the model and the processes and properties that can be observed in the real world remain an issue. To overcome some of these problems, Beck et al. (2015) produced a global dataset of observation-based estimates of hydrological streamflow characteristics. Although still based on empirical models, such estimates may provide a useful benchmark to evaluate the performance of GHMs. Nevertheless, we need to keep in mind that all observational data are uncertain (Beven et al. 2019) and it is therefore essential that model evaluation is undertaken within an uncertainty analysis framework (Lane et al. 2019).

Different types of observations, and in particular satellite observations of quantities like snow cover, land surface temperature, and leaf area index, could also be used in multi-objective model calibration and parameterization (Zhang et al. 2016). Studies have shown that multi-objective calibration against multiple data sources can improve the model performance in simulating processes such as snow accumulation and melt, as well as streamflow (Crow et al. 2003; Udnæs et al. 2007; Parajka and Blöschl 2008; Zhang et al. 2009). However, such an approach is not usually adopted in GHM development and application, and in fact, GHMs are not usually calibrated even to discharge at the catchment outlet (see Gaedeke et al. and Krysanova et al. in this SI).

3.2 Influence of model type and calibration

When compared with observed river discharge, hydrological models sometimes show smaller biases than LSMs (e.g., Beck et al. 2017). To some extent this is expected, as hydrological models broadly solve only the water balance, while LSMs aim to close both the water and the energy budget of the land surface. Haddeland et al. (2011) noted that both the mean and median runoff fractions for the LSMs were lower than those of the GHMs, although the range was wider.

To understand these differences, it is worth keeping the aims in mind with which these models have been developed. Many hydrological models were developed with a focus on predictive skill, and thus tend to be very parsimonious and conceptual. Typically, they are highly abstracted and contain a small number of parameters that can easily be calibrated with observations, and therefore tend to outperform more physically based models when using traditional evaluation methods.

In contrast, LSMs have been built based on an understanding of the main processes and to explore the interaction between processes. Although LSMs represent these processes on a physical basis, they often cannot outperform GHMs as including more and more complex processes also implies larger uncertainties, especially when data are limited to constrain those processes in the models. For instance, LSMs use sophisticated energy balance approaches to model ET but only limited observations exist to evaluate these approaches, let alone calibrate the parameters involved.

Nevertheless, GHMs do not always outperform LSMs: Zaherpour et al. (2018) and Krysanova et al. (this SI) include at least one LSM with smaller biases than several GHMs. Beck et al. (2017) found that the LSMs performed similarly to (uncalibrated) GHMs in rainfall-dominated regions, while in snow-dominated regions the GHMs performed consistently better. Similarly Zhang et al. (2016) found that both LSMs and GHMs can simulate monthly and interannual variability and trends in streamflow reasonably well, even if they cannot adequately reproduce the long-term volumes. They concluded that both types of model can be used for comparative regional and global water balance assessments and projections of future trajectories.

Many studies have found that models that have been calibrated (usually with streamflow data) perform better when compared with river discharge observations (e.g., Beck et al. 2017). Krysanova et al. (2018) noted several problems related to the use of uncalibrated GHMs, including poor performance in many basins and a high spread in climate impact projections, sometimes leading to conflicting results.

Although observations of runoff are not available everywhere around the globe, some GHMs have successfully been calibrated. For example, the WaterGAP model was tuned to long-term average discharge at over a thousand gauging stations (Müller Schmied et al. 2014). Methods for regionalizing model parameters exist but may need to be improved or applied more consistently (Beck et al. 2017). Calibration is easier for catchment-scale models, although even these are typically calibrated at the outlet point only, and good performance there does not guarantee unbiased simulations throughout the catchment. Since calibration may correct for biases in the input data as well as in the model, the better performance may also be restricted to a particular meteorological dataset. However, the sensitivity of a particular model parameterisation to changing input datasets is not commonly assessed.

Hattermann et al. (2017) compared hydrological projections from nine global and nine regional hydrological models with an emphasis on model validation, looking at sensitivity of annual discharge to climate variability and of seasonal dynamics to climate change. The mostly uncalibrated GHMs showed a considerable bias in the long-term average monthly discharge, although they did in many cases reproduce the intra-annual variability well. In contrast, the regional models, tuned to the specific catchments, were better able to reproduce streamflow conditions in the reference period.

Perhaps surprisingly, Hattermann et al. (2017) found that the sensitivity of both types of models (evaluated for their respective ensembles) was quite similar in most basins. They concluded that the GHMs can be useful tools when looking at large-scale impacts of climate variability and change. For local applications, the regional-scale models should be preferred.

3.3 Stationarity of model parameters

A key concern in the application of calibrated models should be the stability or stationarity of model parameters. For example, Merz et al. (2011) found that the optimal values of calibrated parameters changed considerably with time. Assuming time invariant parameters led to significant biases in their simulations, with errors increasing with the time lag between the simulation and calibration periods. A similar result was found by Li et al. (2012), who also noted that some model parameters were significantly more sensitive to the choice of calibration period than others. The use of calibrated models may therefore result in better model performance when compared with historical observations, and therefore higher trust in the model, but it does not necessarily imply reduced uncertainty in the future projections. The assumption of parameter stationarity may introduce additional uncertainties in the simulated response to climate change that is not normally explored (for example through sensitivity analyses).

Several approaches have been proposed to address the issue of non-stationary parameters and the related problems of miscalibration and overcalibration (see also Andréassian et al. 2012). Li et al. (2012) used a Monte-Carlo approach to explore the uncertainty and possible equifinality in hydrological model parameters. They also recommend calibrating a model on wetter periods of the historical record if it is being used to simulate wet climate scenarios, and vice versa for drier scenarios. Similarly, Coron et al. (2012) recommend testing the model robustness and propose a generalized split-sample test to provide insights into the model's transposability over time under various conditions. Krysanova et al. (2018) suggested evaluating models using a proxy climate test.

Westra et al. (2014) proposed a strategy for diagnosing and interpreting hydrological nonstationarity, consisting of investigating potential systematic errors in the calibration data, exploring time-varying model parameters, and trialling alternative model structures. They suggested that time-varying parameters could be a diagnostic for model misspecification: in other words, deficiencies in model structure are likely to express themselves as differences in the estimated parameters when calibrated to climatologically different periods. Wallner and Haberlandt (2015) also investigated the impact of nonstationarity on model performance for different flow indices and time scales and showed that non-stationary parameters can improve the performance with an acceptable growth in parameter uncertainty. Like Li et al. (2012), they also found that some model parameters are highly correlated to some climate indices.

In other cases, for example parameters that relate to the groundwater stores, the assumption of time invariance may hold better, but without further exploration this is still an assumption—and should be recorded as such.

Singh et al. (2011) proposed a trading-space-for-time framework that utilizes the similarity between the predictions under change and in ungauged basins. They noted that the trading-space-for-time approach resulted in a stronger watershed response to climate change for both high and low flow conditions, compared with simulations based on historically calibrated parameters.

However, Stephens et al. (2020) warn against the use of historical periods as proxies for future climate conditions, as levels of carbon dioxide were lower than what is expected in the future. Long-term changes in the ecohydrological functioning of a catchment need to be considered, as relatively brief periods in the past that were wetter or drier than average are unlikely to provide good guidance to what will happen under persistent changes in the future. They conclude that many studies likely underestimate the potential for nonstationarity in hydrologic assessments, especially in case of drier future conditions.

3.4 Parameter uncertainty

In this context, it is noteworthy that relatively few studies have examined the effects of uncertainty in model parameters on climate impact projections, even though techniques to estimate this uncertainty have been around for more than two decades (e.g., Beven and Binley 1992). Similar efforts in climate modelling are now well-established through the application of "perturbed-parameter" ensembles (Murphy et al. 2007; Frame et al. 2009), where a single GCM is run multiple times with different values for some of the key parameters. Due to computational limits, a formal sampling of the full parameter space is out of reach for state-of-the-art, complex earth system models. In practice, the key parameters and parameter values are chosen from ranges considered plausible on the basis of expert judgement. Statistical methods have also been used to estimate the set of projections that would be produced if more comprehensive sampling of parameter uncertainty in the model could be performed (see, e.g., Sexton et al. 2012).

A common finding from these studies is that the uncertainty range in perturbed-parameter ensembles overlaps with those of multi-model ensembles, and Beck et al. (2017) speculate that the same may also be true for hydrological models. When properly designed, such multi-parameterization ensembles may allow a more probabilistic analysis of the results, including the adoption of probabilistic verification techniques that have been widely used in ensemble weather prediction and hydrological forecasting (see, e.g., Franz and Hogue 2011) and could also be used in the evaluation of climate impact models.

3.5 Summarizing multi-model ensemble results

A number of recent studies focused on summarizing the results across the ensemble of multiple models, rather than analysing the differences between them. In their analysis of an ensemble of six global-scale hydrological models, Zaherpour et al. (2018) found that, contrary to expectations, the ensemble mean failed to perform better than any individual model. Similarly, Beck et al. (2017), evaluating 10 state-of-the-art macro-scale hydrological models, found that the multi-model ensemble mean generally did not perform better than the best-performing model or models in the ensemble. These findings are somewhat different from studies of multi-model ensembles in weather and climate modelling, where the ensemble mean is often found to outperform any individual model (e.g., Tebaldi and Knutti 2007; Sanderson and Knutti 2012). More in line with these other fields, Beck et al. (2017) noted that the inclusion of less-accurate models did not severely degrade the overall performance of the ensemble.

The ensemble mean is a straightforward, widely used, method of summarizing the performance of an ensemble of hydrological models. However, the results of Zaherpour et al. (2018) and Beck et al. (2017) suggest that users should not assume a priori that the ensemble mean produces the most trustworthy projections. Zaherpour et al. (2018) recommended the use of weighting individual models based on their performance in the evaluation period. Similar and new techniques using advanced methods for model weighting and process-based observational constraints are already being used in the climate modelling community (see e.g., Giorgi and Mearns 2003; Gillett 2015; Sanderson et al. 2017; Eyring et al. 2019).

However, when analysing the results of multi-model ensembles, even when using simple statistics such as the ensemble mean, we need to ask ourselves whether the usual statistical assumptions actually hold. Noting that many of the studies discussed here are "ensembles of opportunity" and were never designed for such a statistical analysis at the outset, a key concern is to what extent these models can be considered independent. For the Coupled Model Intercomparison Project CMIP5 ensemble of GCMs, Knutti et al. (2013) established that many GCMs were not only strongly tied to their predecessors, but also exchanged ideas and code with other models, implying that the CMIP5 models were neither independent of each other nor independent of the earlier generation. They argued that this interdependence of models complicates the interpretation of multi-model ensembles but largely goes unnoticed. The same may also apply to the ensembles of GHMs and LSMs discussed here, yet the degree of interdependence between these models has never been thoroughly examined.

Keeping in mind that multi-model ensembles are a way to explore structural uncertainty in model formulation, one should ask whether these ensembles of opportunity are indeed sampling the relevant space of possible alternative model structures, if that space could even be specified (Parker 2013). The point of multi-model ensemble studies is not to produce only an ensemble mean that may or may not compare better than individual models with observations in a particular area. Instead we need to look at the full range of responses, better understand why some of the differences occur, and better understand what this tells us about the uncertainty in the projections of climate change impacts.

3.6 Incorporating human factors

River basins around the world are increasingly being modified by human activities, such as building reservoirs and extracting water for irrigation. To enable applications in water resources management, many large-scale hydrological models have now included these anthropogenic factors. This demonstrably enhances model simulation capabilities and enables a more realistic comparison with observations (Zaherpour et al. 2018; Veldkamp et al. 2018).

Veldkamp et al. (2018) compared the results of five state-of-the-art GHMs with observations to examine the role of human impact parameterization (HIP) in the streamflow simulation. Their finding was that inclusion of human activity in GHMs can significantly improve the model performance and this finding is robust across both managed and near-natural catchments and across the GHMs. The inclusion of HIP was found to lead to a significant improvement (decrease in the bias of the long-term mean monthly discharge and an improvement in the modelled hydrological variability ratio). Including HIP in the GHMs also leads to an improvement in the simulation of hydrological extremes. While HIP generally leads to an improvement in the absolute magnitude of simulated high flows, its impact on low flows is mixed.

Liu et al. (2017) noted that parameterizing anthropogenic water uses in GHMs is likely to introduce additional uncertainty in GHMs. Using four GHMs, they conducted the first quantitative investigation of between-model uncertainty resulting from the inclusion of human impact parameterizations. The differences between the two experiments were found to be significantly related to the fraction of irrigation areas of basins. Liu et al. (2017) also discussed differences in the parameterizations of irrigation, reservoir regulation, and water withdrawals, towards potential directions of improvements for future GHM development. Further discussion on including human interventions in hydrological models can also be found in Nazemi and Wheater (2015), Pokhrel et al. (2016), and Wada et al. (2017).

4 Uncertainty and scale

The uncertainty in climate impact projections is intrinsically linked to the scale of the analysis. To illustrate this point, we revisit here the results of Dankers et al. (2014), who provided a first assessment of changes in flood hazard at the global scale based on a relatively large ensemble of climate and impact model simulations from the first (fast-track) phase of ISIMIP

(Warszawski et al. 2014). The ISIMIP fast-track experiments were aimed at providing a rapid assessment of projections of climate impacts, whereas later phases have included a greater focus on model evaluation. In total, nine models provided simulations of daily river discharge at a global 0.5-degree grid to the ISIMIP archive. Each IM was driven by bias-corrected simulations of five GCMs (see Hempel et al. (2013) for details) for up to four scenarios of atmospheric greenhouse gas concentrations (Moss et al. 2010). As an indicator of present-day flood hazard, Dankers et al. (2014) estimated the 30-year return level of river flow (*Q30*) at each grid cell for the 30-year period 1971–2000.

Projections of flood hazard and extreme events in general are typically subject to large uncertainty arising not only from climate and hydrological modelling uncertainties, but also from uncertainties associated with estimating the frequency (or probability) of hydrological extremes from relative short timeseries. The uncertainty related to estimating extremes is in essence a sampling uncertainty and is a function of the length of the timeseries being used. Extreme river flows such as the *Q30* directly relate to the hazard of a flood event happening along a given stretch of a river, but not to the hazard of flooding of a specific area, which would require additional inundation modelling. Note that the *Q30* is not a very extreme discharge level: while the probability of exceedance (P_e) in any given year is 1/30 (0.03), in any given 10-year period it amounts to almost a third (0.29). However, from 30 years of data it can be estimated more robustly than other indicators that are sometimes used, such as the 100-year return level.

Dankers et al. (2014) noted that in individual river basins the uncertainty in the projections of changes in flood hazard can be large, and often even the direction of change (i.e., an increase or decrease) is not clear. Figure 2 summarizes the changes in *Q30* at the outlet of 12 major river basins across the world by the end of this century under two RCPs. Here, changes in *Q30* were calculated by estimating the 30-year return level separately for the period 2070–

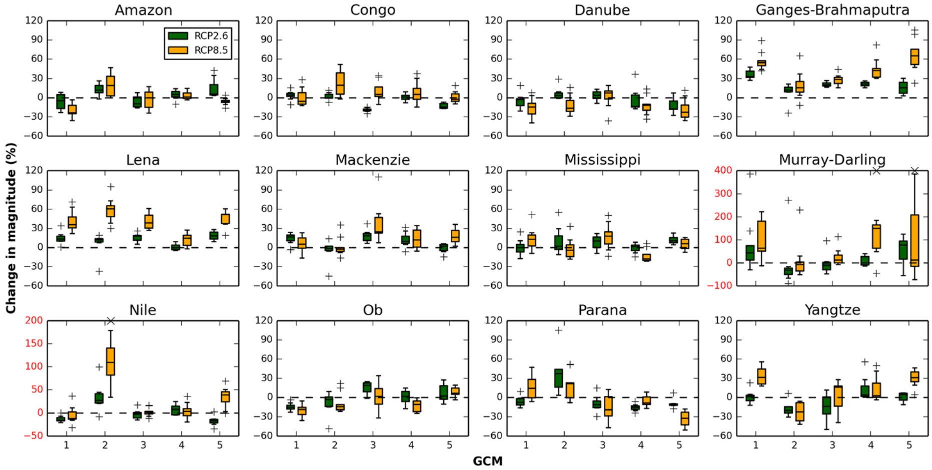


Fig. 2 Relative change (%) in the 30-year return level of river flow (Q30) at the outlet of 12 major river basins as simulated by nine impact models (IMs), each driven by five general circulation models (GCMs) for two different RCPs. Changes in Q30 were calculated by estimating the 30-year return level separately for the historical (1971–2000) and scenario (2070–2099) periods (see Dankers et al. 2014 for details). The distribution of changes in Q30 across the nine IMs is shown by boxplots for each driving GCM, indicated by numbers on the horizontal axis: 1 = HadGEM2-ES; 2 = IPSL-CM5A-LR; 3 = MIROC-ESM-CHEM; 4 = NorESM1-M; 5 = GFDL-ESM2M. Outliers that fall outside the range of the vertical axis are indicated with x. Note the deviating scale for the Murray–Darling and Nile Rivers

2099, and the uncertainty associated with estimating an extreme return level after fitting an extreme value distribution (Coles 2001) may have influenced the results to some extent.

In some river basins (e.g., the Lena) there is a robust signal for an increase in extreme river flow levels across all model combinations, particularly under RCP8.5. But often the signal is much less clear, and in many cases the IMs do not agree on the direction of change even though they have been driven by the same climate forcing (e.g., the Mississippi). Similarly, different driving GCMs sometimes yield conflicting results on the sign of change in *Q30* (e.g., the Yangtze). This highlights once again that, in addition to GCM uncertainty, IM uncertainty arising from differences between the impact models can be a significant component of the overall uncertainty.

These results are complementary to those obtained by, for example, Hirabayashi et al. (2013) who similarly found a low consistency in the direction of change in 100-year discharge in many rivers across a larger ensemble of (uncorrected) GCM simulations driving a single flood inundation model. Likewise, Rojas et al. (2012) found large discrepancies in the magnitude of change in flood hazard at the scale of individual river basins in Europe in an ensemble of 12 climate simulations driving a single hydrological model. More recently, Do et al. (2020) studied, at the global scale, historical and future changes of annual maxima of 7-day streamflow, using a comprehensive streamflow archive and six GHMs. Models show a low to moderate capacity to simulate spatial patterns of historical trends, highlighting the role of model structural uncertainty.

In many cases, these global IMs were not tuned to local-scale conditions, and we should ask ourselves if we should use their results to understand climate impacts in a single basin. It may be better to represent their results at global or perhaps regional (sub-continental) scale. To obtain a global aggregate picture we can calculate a global exceedance rate (*E*), summarizing how often in a given year the historical *Q30* is exceeded globally:

$$E_s = \frac{\sum_{i=1}^{N}\sum_{d=1}^{D}\left[Q_{i,d} > Q30_i\right]}{N}$$

where E_s is the global exceedance rate for a given model simulation, N is the number of land grid cells, D the number of days in a year, and $Q_{i,d}$ is the simulated river discharge.

In essence, E is a measure of the frequency of occurrence of high-flow events (not necessarily flood events) worldwide. It has the advantage that changes in this frequency can be calculated without the need for fitting a new extreme value distribution to a future time period as was done in, for example, Fig. 2.

Since the P_e of *Q30* in any grid cell is 0.03 in any given year, we can expect the *Q30* to be exceeded at roughly 3% of the land grid points in any year ($E = 0.03$). In a stationary climate, E would remain at its expected baseline level, and indeed, in the historical part of the simulations (1971–2000) the average E across the ensemble of 45 GCM-IM combinations is 0.032 ± 0.010 (Fig. 3).

After the first decade of the twenty-first century, however, the simulations suggest a rapid increase in the global exceedance frequency. This increase is robust across all GCM/IM combinations, albeit stronger in some GCM simulations than others (Fig. 3). Under the high-end greenhouse gas scenario RCP8.5, E is on average 0.152 ± 0.045 in the last two decades of the century, suggesting that globally *Q30* levels will be exceeded almost five times more often than in the historical period. The aggressive mitigation scenario RCP2.6 avoids most of the strong increase in E after mid-century, but E is still 0.075 ± 0.024 or more than double the historical rate by the end of the present century.

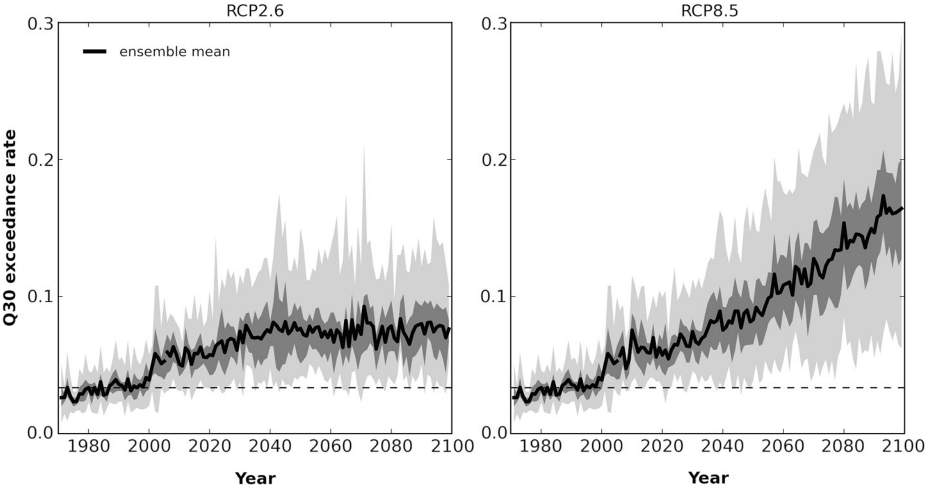


Fig. 3 Change in the Q30 global exceedance rate *E* across the ensemble of 45 GCM-IM combinations under the scenarios RCP2.6 (left panel) and RCP8.5 (right panel). The dark shaded area shows the interquartile range in *E* across the ensemble, the light shaded area the total range. The dashed horizontal line shows the expected baseline *E* in the historical part of the simulations (1971–2000)

In both RCPs, the simulations driven by the GCM NorESM1-M generally show the smallest increases in *E* and MIROC-ESM-CHEM the largest. But there are differences between the IMs, too, with the MATSIRO simulations resulting in the smallest increases in global exceedance frequency on average, and PCR-GLOBWB and (in RCP8.5) JULES the largest.

Analysis of variance (ANOVA) of the simulated *E* in 2080–2099 shows that both GCM and IM have a significant effect ($p < 0.001$) on the overall variability in the results, with a significant ($p < 0.001$) interaction between the two factors. In RCP8.5, the partial effect size (η_p^2) for the factor GCM is of similar magnitude to that for the IMs (0.61 vs 0.58, respectively), while in RCP2.6 the GCMs contribute more to the total variation ($\eta_p^2 = 0.41$) than the IMs ($\eta_p^2 = 0.31$). This highlights that even at the global scale IM uncertainty has a significant impact on projections of changes in *Q30* exceedance frequency.

So while in individual basins the GCM-IM combination give very different results and sometimes will not even agree on the direction of flood hazard change, at the global scale the signal for an increase in the global exceedance rate *E* is remarkably robust across model combinations. This finding is analogous to Fischer et al. (2013) who found that spatially aggregated projections of precipitation extremes can be highly robust even if they are very uncertain at the local scale. In other words, we can state with more confidence that the RCP8.5 scenario will lead to more frequent flood events globally than where exactly these events will occur.

5 Conclusions

A number of concluding remarks can be drawn from the previous discussion, as well as the example of changes in flood hazard at global scale. First of all, it is clear that in addition to

GCM uncertainty, IM (structural) uncertainty can be a significant component of the overall uncertainty in the projections of climate-change impact on water resources. This implies that studies that are based on a single hydrological model may well be overconfident and do not adequately sample the uncertainty range even if they use multiple driving climate models and/or realizations to account for the uncertainty in the projected climate change.

Following Beven (2013), a multi-model approach can be an effective way to explore some of the epistemic uncertainty in impact modelling. However, the mere fact of using a multi-model ensemble does not mean that all sources of uncertainty in the projections have been fully represented or quantified.

We also need to question ourselves whether we can analyse multi-model ensembles statistically as if they are similar to aleatory uncertainty, with simple, known, error distributions. This is especially true for "ensembles of opportunities" that were not designed for such a statistical analysis on the outset. We need to understand better to what extent the models included in some of the MIPs share the same modelling approaches and process descriptions, and where they are different.

Several studies (e.g., Haddeland et al. 2011; Beck et al. 2017) have found that hydrological models, especially when calibrated at basin scale, tend to outperform the more complex LSMs in reproducing hydrological variability, when compared with observations (usually limited to records of river discharge only). However, in the context of climate change a key process—and a good example of some of the "deeper" uncertainties involved—is the response of the vegetation to the changing climate (Davie et al. 2013; Stephens et al. 2020). At least in a qualitative sense it is well-established that higher concentrations of atmospheric CO_2 will affect the water-use efficiency of plants, yet unlike the LSMs most hydrological models ignore this process altogether. If the actual sensitivity to elevated CO_2 concentrations is high, studies that use only hydrological model simulations on the basis of their seemingly better performance in the past risk underestimating the true uncertainty in the projected impacts.

We have seen that uncertainty is larger at the smaller scale of individual river basins. This provides a challenge to local adaptation decisions, as greater uncertainty may, for example, require greater protective measures in order to keep the flood hazard at the same level (cf. Hunter 2012). The implication is that global-scale modelling projections will not necessarily provide the best guidance for local-scale decisions. A different approach, more tuned to local conditions, may well be required in order to reduce uncertainty at local scale. For example, it may be possible to reduce the spread in multi-model ensemble results by down-weighting or eliminating models that are clearly unable to reproduce important aspects of the water cycle in a particular catchment, while in global scale applications it is unlikely that any one model will be "good" or "bad" everywhere around the globe. However, this desire to narrow the model spread needs to be carefully balanced against the need to sample the full uncertainty range, including the extremes, in order to avoid overconfident projections that may result in wrong adaptation decisions (Knutti 2010).

In the face of large uncertainty that is unlikely to be fully sampled by a limited set of hydrological or impact models, a more productive approach could be to focus on the information that a model or set of models can provide to enable (quasi-) 'optimal' decisions (Gupta et al. 2012; Nearing and Gupta 2015). One way to deal with large uncertainty is to evaluate the sensitivity of the decision against a range of possible climate outcomes, thus highlighting critical vulnerabilities that may warrant further attention (cf. Prudhomme et al. 2010). Kundzewicz et al. (2017) noted that it is rather naïve to expect that reliable (in a statistical sense) quantitative projections of future flood hazard may become available. Hence, in order to reduce flood risk, one should focus attention on identification of current and future risks and vulnerability hotspots and improve the situation in areas where such hotspots occur (Kundzewicz et al. 2017).

Perhaps a comprehensive evaluation of all the uncertainties involved in the cascade of climate and impact models, and an honest appraisal of the "deep" uncertainties associated with the modelling process, may feel overwhelming. Yet, that is no reason for ignoring these uncertainties (Pappenberger and Beven 2006). Since every analysis is conditional on the assumptions about the sources of epistemic uncertainty, Beven et al. (2018) recommend to record these assumptions and evaluate their impact on the uncertainty estimate (see also Beven et al. 2018).

When operational meteorologists produce their weather forecasts, they tend to use the output of numerical weather prediction models as a guide and interpret the model simulations in the light of known and unknown limitations of the models, and their own expert insight into the evolving weather situation. On occasion, they will deviate from the model guidance and produce their own assessment of the expected weather. In a similar way, impact modellers (being disciplinary specialists) need to interpret, or help their users interpreting, the output of their simulations. This may extend to being able to understand the driving climate models, and the reason for some of the differences observed in these climate models. Impact models can be very useful tools to test our hypotheses on expected future climate impacts, but ultimately they are based on our limited knowledge and judgement and subject to the assumptions that were made during their development, and hence, they need to be used with caution (Spiegelhalter and Riesch 2011).

6 Recommendations

A number of recommendations can be derived from the previous discussion, both for model development and for model use. First of all, we feel that the community will benefit from developing a common language and clear framework for the treatment of uncertainties in climate impact assessments. For model development, there is a need to explore parameter and structural uncertainty in a more consistent manner, akin to the perturbed parameter ensembles used in climate modelling and the approach used by Lane et al. (2019) for river flow and flood prediction in Great Britain, as opposed to the current "ensembles of opportunity." At the same time, MIPs could be exploited more fully to better understand the mechanisms and processes that lead to different responses in the models and could explain part of the uncertainty in climate impact projections.

Given the scale of human interference in the hydrological cycle, it is imperative that human impacts are included in large-scale models to improve the realism of the models. However, this requires an ongoing effort in data collection and further development of these schemes. Finally, there are processes that could be highly uncertain in the current generation of models, in particular the water-use efficiency of the vegetation under higher CO_2 concentrations, and the representation of groundwater dynamics. More effort is needed to investigate the sensitivity of climate impact projections to these processes, and to improve the realism in the models.

With regard to the use of impact models, a strong message that comes through is that climate impact projections should not be based on a single model, or indeed the ensemble mean. Users need to be aware that the true uncertainty is likely to be larger than what has been sampled by current multi-model ensembles. In ensemble weather forecasting, where—unlike in climate impact studies—model predictions can routinely be compared with the actual outcomes, a common finding is that the ensembles are often "underspread," in other words fail to capture the full range of outcomes especially at longer lead times. In this area of application, the aim is often to increase the model spread rather than reducing it. In a similar way, climate impact studies should aim to capture the full uncertainty range, enabling users to seek robust decisions that perform well across a wide range of possible future climate impact scenarios (Kalra et al. 2014).

Acknowledgements We thank two reviewers and the guest editor of this special issue for their very useful comments that helped us to improve our paper.

Funding The analysis of flood projections in Section 4 was partly funded by the Joint UK Department of Energy and Climate Change/Department for Environment, Food and Rural Affairs Met Office Hadley Centre Climate Programme (GA01101) when the first author (RD) was working at the Met Office in Exeter, United Kingdom.

References

Andréassian V, Le Moine N, Perrin C et al (2012) All that glitters is not gold: the case of calibrating hydrological models. Hydrol Process 26:2206–2210. https://doi.org/10.1002/hyp.9264

Beck HE, de Roo A, van Dijk AIJM et al (2015) Global maps of Streamflow characteristics based on observations from several thousand catchments. J Hydrometeorol 16:1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Beck HE, van Dijk AIJM, de Roo A et al (2017) Global evaluation of runoff from 10 stateof- the-art hydrological models. Hydrol Earth Syst Sci 21:2881–2903. https://doi.org/10.5194/hess-21-2881-2017

Beven K (2013) So how much of your error is epistemic? Lessons from Japan and Italy. Hydrol Process 27: 1677–1680. https://doi.org/10.1002/hyp.9648

Beven K (2016) Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol Sci J 61:1652–1665. https://doi.org/10.1080/02626667.2015.1031761

Beven KJ, Almeida S, Aspinall WP et al (2018) Epistemic uncertainties and natural hazard risk assessment—part 1: a review of different natural hazard areas. Nat Hazards Earth Syst Sci 18:2741–2768. https://doi.org/10.5194/nhess-18-2741-2018

Beven K, Asadullah A, Bates P, et al (2019) Developing observational methods to drive future hydrological science: can we make a start as a community? Hydrol Process Hyp 13622. https://doi.org/10.1002/hyp.13622

Beven KJ, Aspinall WP, Bates PD et al (2018) Epistemic uncertainties and natural hazard risk assessment—part~2: what should constitute good practice? Nat Hazards Earth Syst Sci 18:2769–2783. https://doi.org/10.5194/nhess-18-2769-2018

Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6:279–298. https://doi.org/10.1002/hyp.3360060305

Beven K, Younger P, Freer J (2014) Struggling with epistemic uncertainties in environmental Modelling of natural hazards. In: Vulnerability, uncertainty, and risk. American Society of Civil Engineers, pp. 13–22

Budescu DV, Wallsten TS (1987) Subjective estimation of precise and vague uncertainties. In: Wright G, Ayton P (eds) Judgmental forecasting. John Wiley & Sons Ltd, Chichester, pp 63–82

Coles S (2001) An introduction to statistical modeling of extreme values, 1st Edition. Springer

Coron L, Andréassian V, Perrin C, et al (2012) Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. Water Resour Res 48:W05552+. https://doi.org/10.1029/2011wr011721

Crow WT, Wood EF, Pan M (2003) Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals. J Geophys Res 108:4725. https://doi.org/10.1029/2002JD003292

Dankers R, Arnell NW, Clark DB et al (2014) First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. Proc Natl Acad Sci U S A 111. https://doi.org/10.1073/pnas.1302078110

Davie JCS, Falloon PD, Kahana R et al (2013) Comparing projections of future changes in runoff from hydrological and biome models in ISI-MIP. Earth Syst Dyn 4:359–374. https://doi.org/10.5194/esd-4-359-2013

Do HX, Zhao F, Westra S et al (2020) Historical and future changes in global flood magnitude - evidence from a model-observation investigation. Hydrol Earth Syst Sci 24:1543–1564. https://doi.org/10.5194/hess-24-1543-2020

Ehret U, Zehe E, Wulfmeyer V et al (2012) HESS opinions “Should we apply bias correction to global and regional climate model data?”. Hydrol Earth Syst Sci 16:3391–3404. https://doi.org/10.5194/hess-16-3391-2012

Eyring V, Cox PM, Flato GM et al (2019) Taking climate model evaluation to the next level. Nat Clim Chang 9:102–110. https://doi.org/10.1038/s41558-018-0355-y

Fischer EM, Beyerle U, Knutti R (2013) Robust spatially aggregated projections of climate extremes. Nat Clim Chang 3:1033–1038. https://doi.org/10.1038/nclimate2051

Frame D., Aina T, Christensen C., et al (2009) The climateprediction.net BBC climate change experiment: design of the coupled model ensemble. Philos Trans R Soc A Math Phys Eng Sci 367:855–870. https://doi.org/10.1098/rsta.2008.0240

Franz KJ, Hogue TS (2011) Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community. Hydrol Earth Syst Sci 15:3367–3382. https://doi.org/10.5194/hess-15-3367-2011

Gillett NP (2015) Weighting climate model projections using observational constraints. Philos Trans R Soc A Math Phys Eng Sci 373:20140425. https://doi.org/10.1098/rsta.2014.0425

Giorgi F, Mearns LO (2003) Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. Geophys Res Lett 30:1629–n/a. https://doi.org/10.1029/2003gl017130

Guo D, Westra S, Maier HR (2017) Impact of evapotranspiration process representation on runoff projections from conceptual rainfall-runoff models. Water Resour Res 53:435–454. https://doi.org/10.1002/2016WR019627

Gupta H V, Clark MP, Vrugt JA et al (2012) Towards a comprehensive assessment of model structural adequacy. Water Resour Res 48:W08301+. https://doi.org/10.1029/2011wr011044

Haddeland I, Clark DB, Franssen W et al (2011) Multimodel estimate of the global terrestrial water balance: setup and first results. J Hydrometeorol 12:869–884. https://doi.org/10.1175/2011JHM1324.1

Haddeland I, Heinke J, Voß F et al (2012) Effects of climate model radiation, humidity and wind estimates on hydrological simulations. Hydrol Earth Syst Sci 16:305–318. https://doi.org/10.5194/hess-16-305-2012

Harding R, Best M, Blyth E et al (2011) WATCH: current knowledge of the terrestrial global water cycle. J Hydrometeorol 12:1149–1156. https://doi.org/10.1175/JHMD-11-024.1

Hattermann FF, Krysanova V, Gosling SN, et al (2017) Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. Clim Change 1–16. https://doi.org/10.1007/s10584-016-1829-4

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bull Am Meteorol Soc 90:1095–1108. https://doi.org/10.1175/2009BAMS2607.1

Hawkins E, Sutton R (2011) The potential to narrow uncertainty in projections of regional precipitation change. Clim Dyn 37:407–418. https://doi.org/10.1007/s00382-010-0810-6

Hempel S, Frieler K, Warszawski L et al (2013) A trend-preserving bias correction: the ISI-MIP approach. Earth Syst Dyn 4:219–236. https://doi.org/10.5194/esd-4-219-2013

Hirabayashi Y, Mahendran R, Koirala S et al (2013) Global flood risk under climate change. Nat Clim Chang 3:816–821. https://doi.org/10.1038/nclimate1911

Hunter J (2012) A simple technique for estimating an allowance for uncertain sea-level rise. Clim Chang 113:239–252. https://doi.org/10.1007/s10584-011-0332-1

IPCC (1990) The IPCC impacts assessment. Australian Government Publishing Service

Kalra N, Hallegatte S, Lempert R, et al (2014) Agreeing on robust decisions: new processes for decision making under deep uncertainty. The World Bank

Knight FH (1921) Risk, uncertainty and profit. Houghton Mifflin, New York

Knutti R (2010) The end of model democracy? Clim Chang 102:395–404. https://doi.org/10.1007/s10584-010-9800-2

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

Koirala S, Yeh PJ-F, Hirabayashi Y et al (2014) Global-scale land surface hydrologic modeling with the representation of water table dynamics. J Geophys Res Atmos 119:75–89. https://doi.org/10.1002/2013JD020398

Krysanova V, Donnelly C, Gelfan A et al (2018) How the performance of hydrological models relates to credibility of projections under climate change. Hydrol Sci J 63:696–720. https://doi.org/10.1080/02626667.2018.1446214

Kundzewicz ZW, Krysanova V, Benestad RE et al (2018) Uncertainty in climate change impacts on water resources. Environ Sci Pol 79:1–8. https://doi.org/10.1016/J.ENVSCI.2017.10.008

Kundzewicz ZW, Krysanova V, Dankers R, et al (2017) Differences in flood hazard projections in Europe–their causes and consequences for decision making. Hydrol Sci J 62. https://doi.org/10.1080/02626667.2016.1241398

Lane RA, Coxon G, Freer JE et al (2019) Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. Hydrol Earth Syst Sci 23:4011–4032. https://doi.org/10.5194/hess-23-4011-2019

Li CZ, Zhang L, Wang H et al (2012) The transferability of hydrological models under nonstationary climatic conditions. Hydrol Earth Syst Sci 16:1239–1254. https://doi.org/10.5194/hess-16-1239-2012

Liu X, Tang Q, Cui H et al (2017) Multimodel uncertainty changes in simulated river flows induced by human impact parameterizations. Environ Res Lett 12:025009. https://doi.org/10.1088/1748-9326/aa5a3a

Lo M-H, Famiglietti JS, Yeh PJ-F, Syed TH (2010) Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. Water Resour Res 46. https://doi.org/10.1029/2009WR007855

Maier HR, Guillaume JHA, van Delden H et al (2016) An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? Environ Model Softw 81:154–164. https://doi.org/10.1016/J.ENVSOFT.2016.03.014

Merz R, Parajka J, Blöschl G (2011) Time stability of catchment model parameters: implications for climate impact analyses. Water Resour Res 47:W02531+. https://doi.org/10.1029/2010wr009505

Metin AD, Dung NV, Schröter K et al (2018) 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

Moss RH, Edmonds JA, Hibbard KA et al (2010) The next generation of scenarios for climate change research and assessment. Nature 463:747–756. https://doi.org/10.1038/nature08823

Müller Schmied H, Adam L, Eisner S et al (2016) Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. Hydrol Earth Syst Sci 20:2877–2898. https://doi.org/10.5194/hess-20-2877-2016

Müller Schmied H, Eisner S, Franz D et al (2014) Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. Hydrol Earth Syst Sci 18:3511–3538. https://doi.org/10.5194/hess-18-3511-2014

Murphy J, Booth BB, Collins M et al (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philos Trans R Soc A Math Phys Eng Sci 365:1993–2028. https://doi.org/10.1098/rsta.2007.2077

Nazemi A, Wheater HS (2015) On inclusion of water resource management in earth system models—part 1: problem definition and representation of water demand. Hydrol Earth Syst Sci 19:33–61. https://doi.org/10.5194/hess-19-33-2015

Nearing GS, Gupta HV (2015) The quantity and quality of information in hydrologic models. Water Resour Res 51:524–538. https://doi.org/10.1002/2014WR015895

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the Earth sciences. Science 263:641–646. https://doi.org/10.1126/science.263.5147.641

Pappenberger F, Beven KJ (2006) Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resour Res 42:W05302+. https://doi.org/10.1029/2005wr004820

Parajka J, Blöschl G (2008) The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. J Hydrol 358:240–258. https://doi.org/10.1016/J.JHYDROL.2008.06.006

Parker WS (2013) Ensemble modeling, uncertainty and robust predictions. WIREs Clim Chang 4:213–223. https://doi.org/10.1002/wcc.220

Pokhrel YN, Hanasaki N, Wada Y, Kim H (2016) Recent progresses in incorporating human land-water management into global land surface models toward their integration into Earth system models. Wiley Interdiscip Rev Water 3:548–574. https://doi.org/10.1002/wat2.1150

Prudhomme C, Wilby RL, Crooks S et al (2010) Scenario-neutral approach to climate change impact studies: application to flood risk. J Hydrol 390:198–209. https://doi.org/10.1016/j.jhydrol.2010.06.043

Rojas R, Feyen L, Bianchi A, Dosio A (2012) Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations. J Geophys Res 117:D17109+. https://doi.org/10.1029/2012jd017461

Sanderson BM, Knutti R (2012) On the interpretation of constrained climate model ensembles. Geophys Res Lett 39:L16708+. https://doi.org/10.1029/2012gl052665

Sanderson BM, Wehner M, Knutti R (2017) Skill and independence weighting for multimodel assessments. Geosci Model Dev 10:2379–2395. https://doi.org/10.5194/gmd-10-2379-2017

Sexton DMH, Murphy JM, Collins M, Webb MJ (2012) Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. Clim Dyn 38:2513–2542. https://doi.org/10.1007/s00382-011-1208-9

Singh R, Wagener T, van Werkhoven K et al (2011) A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate – accounting for changing watershed behavior. Hydrol Earth Syst Sci 15:3591–3603. https://doi.org/10.5194/hess-15-3591-2011

Spiegelhalter DJ, Riesch H (2011) Don't know, can't know: embracing deeper uncertainties when analysing risks. Philos Trans R Soc London A Math Phys Eng Sci 369:4730–4750. https://doi.org/10.1098/rsta.2011.0163

Stephens CM, Marshall LA, Johnson FM et al (2020) Is past variability a suitable proxy for future change? A Virtual Catchment Experiment. Water Resour Res 56. https://doi.org/10.1029/2019WR026275

Suckling E (2018) Seasonal-to-decadal climate forecasting. In: Weather & climate services for the energy industry. Springer International Publishing, Cham, pp 123–137

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos Trans R Soc A Math Phys Eng Sci 365:2053–2075. https://doi.org/10.1098/rsta.2007.2076

Udnæs H-C, Alfnes E, Andreassen LM (2007) Improving runoff modelling using satellitederived snow covered area? Hydrol Res 38:21–32. https://doi.org/10.2166/nh.2007.032

van Vuuren DP, Edmonds J, Kainuma M et al (2011) The representative concentration pathways: an overview. Clim Chang 109:5–31. https://doi.org/10.1007/s10584-011-0148-z

Veldkamp TIE, Zhao F, Ward PJ et al (2018) Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. Environ Res Lett 13:055008. https://doi.org/10.1088/1748-9326/aab96f

Vetter T, Reinhardt J, Flörke M et al (2017) Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. Clim Chang 141:419–433. https://doi.org/10.1007/s10584-016-1794-y

Wada Y, Bierkens MFP, de Roo A et al (2017) Human–water interface in hydrological modelling: current status and future directions. Hydrol Earth Syst Sci 21:4169–4193. https://doi.org/10.5194/hess-21-4169-2017

Wallner M, Haberlandt U (2015) Non-stationary hydrological model parameters: a framework based on SOM-B. Hydrol Process 29:3145–3161. https://doi.org/10.1002/hyp.10430

Warszawski L, Frieler K, Huber V et al (2014) The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): project framework. Proc Natl Acad Sci 111:3228–3232. https://doi.org/10.1073/pnas.1312330110

Westra S, Thyer M, Leonard M et al (2014) A strategy for diagnosing and interpreting hydrological model nonstationarity. Water Resour Res 50:5090–5113. https://doi.org/10.1002/2013wr014719

Zaherpour J, Gosling SN, Mount N et al (2018) Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environ Res Lett 13:65015

Zhang Y, Chiew FHS, Zhang L, Li H (2009) Use of remotely sensed actual evapotranspiration to improve rainfall–runoff modeling in Southeast Australia. J Hydrometeorol 10:969–980. https://doi.org/10.1175/2009JHM1061.1

Zhang Y, Zheng H, Chiew FHS et al (2016) Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. J Hydrometeorol 17:995–1010. https://doi.org/10.1175/JHM-D-15-0107.1

Zscheischler J, Westra S, van den Hurk BJJM et al (2018) Future climate risk from compound events. Nat Clim Chang 8:469–477. https://doi.org/10.1038/s41558-018-0156-3

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.