



Wagener, T., Reinecke, R., & Pianosi, F. (2022). On the evaluation of climate change impact models. *WIREs Climate Change*, 13(3), [e772]. <https://doi.org/10.1002/wcc.772>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1002/wcc.772](https://doi.org/10.1002/wcc.772)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <https://doi.org/10.1002/wcc.772>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

## PERSPECTIVE

# On the evaluation of climate change impact models

Thorsten Wagener<sup>1,2,3</sup>  | Robert Reinecke<sup>1</sup>  | Francesca Pianosi<sup>2,3</sup> 

<sup>1</sup>Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

<sup>2</sup>Department of Civil Engineering, University of Bristol, Bristol, UK

<sup>3</sup>Cabot Institute, University of Bristol, Bristol, UK

## Correspondence

Thorsten Wagener, Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany.  
Email: thorsten.wagener@uni-potsdam.de

## Funding information

Alexander von Humboldt Foundation, German Federal Ministry of Education and Research; Living with Environmental Uncertainty fellowship from the UK Engineering and Physical Sciences Research Council (EPSRC), Grant/Award: EP/R007330/1

**Edited by:** Eduardo Zorita, Domain Editor and Mike Hulme, Editor-in-Chief

## Abstract

In-depth understanding of the potential implications of climate change is required to guide decision- and policy-makers when developing adaptation strategies and designing infrastructure suitable for future conditions. Impact models that translate potential future climate conditions into variables of interest are needed to create the causal connection between a changing climate and its impact for different sectors. Recent surveys suggest that the primary strategy for validating such models (and hence for justifying their use) heavily relies on assessing the accuracy of model simulations by comparing them against historical observations. We argue that such a comparison is necessary and valuable, but not sufficient to achieve a comprehensive evaluation of climate change impact models. We believe that a complementary, largely observation-independent, step of model evaluation is needed to ensure more transparency of model behavior and greater robustness of scenario-based analyses. This step should address the following four questions: (1) Do modeled dominant process controls match our system perception? (2) Is my model's sensitivity to changing forcing as expected? (3) Do modeled decision levers show adequate influence? (4) Can we attribute uncertainty sources throughout the projection horizon? We believe that global sensitivity analysis, with its ability to investigate a model's response to joint variations of multiple inputs in a structured way, offers a coherent approach to address all four questions comprehensively. Such additional model evaluation would strengthen stakeholder confidence in model projections and, therefore, into the adaptation strategies derived with the help of impact models.

This article is categorized under:

- Climate Models and Modeling > Knowledge Generation with Models
- Assessing Impacts of Climate Change > Evaluating Future Impacts of Climate Change

## KEYWORDS

adaptation, sensitivity analysis, uncertainty, validation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *WIREs Climate Change* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

Human activity has become a geologic-scale force, changing landscapes and climate at increasing rates in our effort to supply societies' growing demand for water, energy and food. A fundamental scientific and societal question of our time is: How will this activity alter water, energy, and biogeochemical cycles, and when and where will vulnerability thresholds critical for society and our environment be reached (Gleeson et al., 2020; Rockström et al., 2021)? To address these questions, we need impact models that can reliably translate climate signals into decision-relevant environmental variables (while also considering other possible changes, e.g., to land use) defining future boundary conditions for society and ecosystems. Information provided by impact models across appropriate space–time scales guides decision- and policy-makers in developing adaptation strategies, long-term planning, and infrastructure design suitable for future conditions (Barron, 2009).

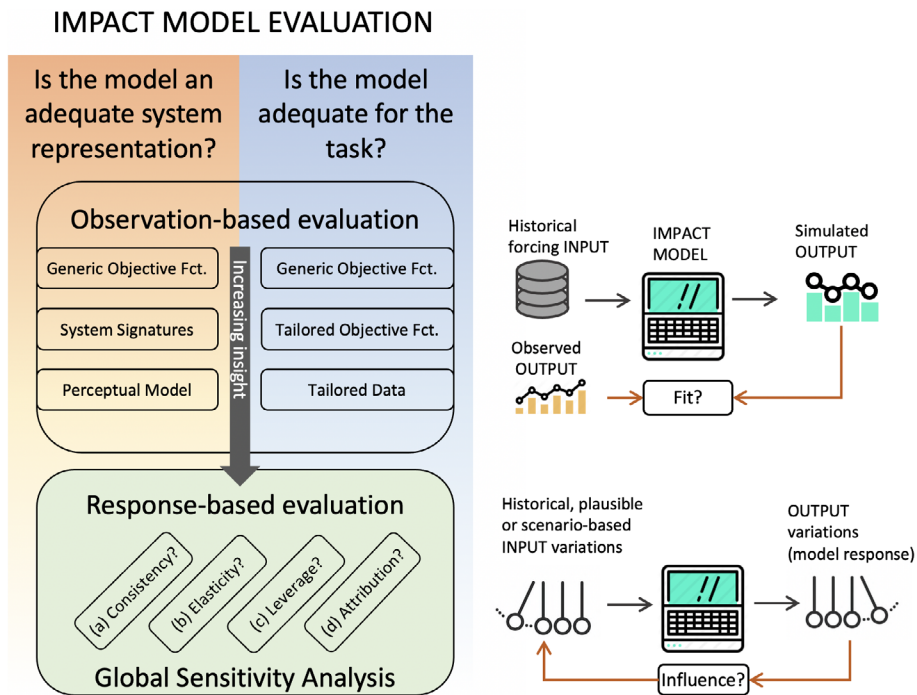
Any impact model is an imperfect representation of the underlying system and will unavoidably be affected by uncertainties introduced along the modeling chain, including those from the climate projections or impact model parameters (Clark et al., 2016; Hakala et al., 2018; Hattermann et al., 2018; Singh et al., 2013; Smith et al., 2018; Wilby & Dessai, 2010). A critical question in the context of using such models is how they have been evaluated regarding both their ability to perform their task adequately, and their suitability as system representation in the first place (Gupta et al., 2012; Klemeš, 1986; Seibert et al., 2000; Wagener et al., 2010). The process of addressing these questions is often referred to as model validation, although the use of this term has been criticized because it suggests that a model can be established as being true, which is not possible when dealing with open systems (Oreskes et al., 1994) or with highly complex systems that can be aggregated into much simpler simulation models in many different ways. Therefore, we prefer here to use the term evaluation, instead of validation, to suggest that we can ever only achieve an incomplete and conditional assessment of a model's suitability (as suggested by Oreskes et al., 1994). In a recent review of validation methods and practices for resource management models, Eker et al. (2018) found that observation-based strategies dominate, even in the context of climate change impact assessment (or “scenario modeling”): that is, a model's ability to reproduce observed system responses over a historical period is taken as evidence that the model is a valid representation of reality and hence suitable for subsequent use. When observations of the predicted variables of interest are not available, for example, because they cannot be measured at relevant scales (e.g., groundwater recharge; Reinecke et al., 2021), or because the region is ungauged (Blöschl et al., 2013), or because the model output is a highly conceptualized variable with no real-world equivalent (e.g., some global integrated assessment variables; Butler et al., 2014), impact models are often not evaluated at all.

We claim that observation-based strategies as defined above are useful but not sufficient for comprehensive model evaluation and that an additional (essentially observation-independent) evaluation stage is necessary to gain confidence in future impact model projections. For reasons that will become apparent in the remainder of the paper, we call this additional stage a “response-based” evaluation. In developing our argument, we find it helpful to clearly distinguish two aims of model evaluation in the context of scenario modeling (Figure 1): model evaluation as an attempt to establish that the model is an adequate representation of the real-world system; and/or, to establish that the model is adequate for a specific task (such as prediction of a single decision-relevant variable)—two aims which can sometimes be in conflict as we discuss below.

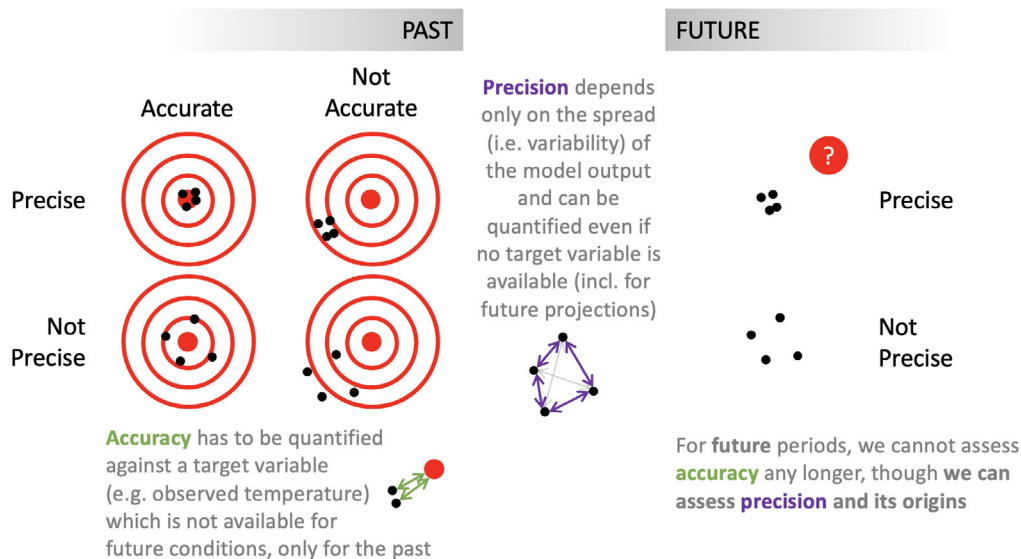
## 2 | OBSERVATION-BASED MODEL EVALUATION STRATEGIES

### 2.1 | Is the model an adequate representation of the real-world system?

First, the modeler might want to demonstrate a model as an adequate representation of the real-world system (Figure 2). The dominant strategy to achieve this is an assessment of a model's accuracy, that is, a quantification of the fit between simulated and observed time-series or spatial patterns of a target variable, such as streamflow or the frequency of hot days. This comparison is most often made using one (or more) statistical objective functions in which the differences between observed and simulated variables are aggregated over time and/or space using summary metrics such as the root mean squared error or Nash–Sutcliffe efficiency (Bennett et al., 2013). Some previous studies have suggested performance thresholds related to such metrics to decide whether a model is adequate or not. For example, Moriasi et al. (2007) suggest fixed quality thresholds for metrics such as the Nash–Sutcliffe efficiency. However, other studies have shown that the use of a single-performance threshold for an objective function across different systems



**FIGURE 1** Conceptualization of impact model evaluation strategies based on both observation-based and response-based strategies. Fct., function



**FIGURE 2** Schematic representation of accuracy (How close are predictions to observations?) and precision (What is the variability of model predictions?) in the context of simulating past and future time periods

can be highly misleading (Knoben et al., 2019), because it is much easier to achieve high performance metrics for some catchments than for others (Van Werkhoven et al., 2008). Flexible performance benchmarks have therefore been suggested as more meaningful alternatives (Schaeffli and Gupta, 2007). A more physically meaningful evaluation strategy is the use of the so-called system signatures (Gupta et al., 2008; Hrachowitz et al., 2014). System signatures can be defined as indices that provide insight into the functional behavior of the underlying system and thus how well the model matches this behavior. Examples of signatures in hydrology are the runoff ratio (defined as the long-term fraction of precipitation that leaves a catchment as streamflow) or the slope of the flow duration curve (which describes how damped a catchment response is) (McMillan, 2021). Such signature-based evaluation is a step toward attempting

to understand the consistency between the simulation model and our current perception of the underlying real-world systems (Euser et al., 2013; Wagener et al., 2021).

However, a simple comparison to historical observation is problematic in the context of climate change impact modeling because the real-world system might evolve with changing climatic boundary conditions, thus requiring different model parameters to reproduce system behavior under different (future) climates (Merz et al., 2011, for hydrologic models or Rosero et al., 2010, for land surface models). How and with what data an impact model is calibrated might therefore strongly define future projections (Müller Schmied et al., 2014). Different studies have tried to create some type of resampling of the past (Fowler et al., 2018) or trading space for time strategies (Singh et al., 2013) to better reflect future conditions during model calibration/evaluation. Other modelers prefer to run their models (especially across large domains) in uncalibrated mode, thus relying on their physical realism. This, however, regularly leaves significant performance gaps between models and observed behavior for climate change impact projections, and eliminating ensemble members that perform poorly on historical observations should increase stakeholder confidence in the ensemble projections (Krysanova et al., 2018), although using all ensemble members—regardless of historical performance—is rather common in large scale impact studies (Gosling et al., 2017). In either approach, there is the additional problem that any discrepancy in historical model performance might not be similar for future model projections. For example, Milly and Dunne (2017) demonstrated how projected future hydrologic drying and wetting trends might be systematically overestimated or underestimated through the choice of evapotranspiration routine.

## 2.2 | Is the model adequate for the task it is intended for?

Second, the modeler might want to justify a model as being adequate for the task at hand (Figure 1). Modelers might see the second aim in connection with the first—assuming that if the model is an adequate representation of the system, then it should be adequate for any practical task—or they might see it as a more practical alternative to aim one, when this is difficult to achieve. The latter approach translates the basic sentiment of “all models are wrong, but some are useful” (Box, 1976), and thus focuses on establishing that a model is mainly suitable for a chosen purpose. In fact, as any model is an imperfect representation of reality, and thus unavoidably includes model structural issues that often do not allow it to match multiple system observations or different objective functions simultaneously (Kollet et al., 2012), the modeler may have to make choices regarding how far to sacrifice one purpose for the other. For example, Wagener and McIntyre (2005) showed that a rainfall-runoff model can become less and less believable as a hydrologic catchment representation the more it is optimized for a specific water resources management task.

A general comparison of the model to historical data—for example, using standard statistical objective functions mentioned above—might ignore the intended use of the model, even though one might expect it to be a primary driver of the evaluation strategy. For example, Brunner et al. (2021) showed that model evaluation with a widely used statistic of overall fit-to-data does not ensure that a hydrologic model is suitable for flood hazard assessment. Klemeš (1986) introduced multiple ideas for validation strategies in relation to intended model use, for example, related to modeling land-use change. These ideas are still rarely fully implemented. Still, focusing on fitting historical data is emphasizing model performance, rather than model robustness under unavoidable uncertainties. Lastly, comparing the model to historical data might ignore the way stakeholders gain trust in model predictions, especially when modeling change (Eker et al., 2018). Stakeholders might, for example, care strongly whether the modeler's perception of the real-world system reflects an understanding that is consistent with their own, that is, they might ask whether the perceptual model underlying the simulation model is consistent with their experiences (Mahmoud et al., 2009).

## 3 | BEYOND MODEL EVALUATION BASED ON FIT-TO-OBSERVATIONS

Here we claim that the above strategies based on fitting historical observations are necessary and useful (when appropriate observations are available) but by no means sufficient to demonstrate that a model is either a suitable system representation, nor that it is adequate for the task. We suggest that they should be complemented with strategies that demonstrate the “internal consistency” (Oreskes et al., 1994) of the model, that is, the model's input–output response is sufficiently consistent with our current perception of the underlying system (Wagener and Gupta, 2005) and the intended use of the model (Klemeš, 1986).



Our discussion connects closely to the concepts of accuracy and precision, though applied to model predictions instead of measurements (Figure 2). A model is accurate if its predictions are close to the observed system output. They are precise if the variability of predictions from ensemble members is narrow. Key to our discussion here is whether we assess accuracy and precision for past periods (for which we can have historical observations), or whether we are looking at projections for future periods for which observations do not yet exist. Therefore, accuracy can be assessed for the past and not for the future. Although precision, and importantly its causes, can be assessed for both past and future. In this short piece, we argue that the evaluation of precision and its causes should be a key element in climate change impact model evaluation.

We argue that one strategy to perform this observations-free evaluation of precision is through the use of global sensitivity analysis (GSA). GSA describes a set of mathematical tools that enable a modeler to systematically investigate how the variations in the inputs of a model (including parameters, forcing, or initial conditions) propagate into variations in the model's output(s) (Saltelli et al., 2000). A typical result of GSA is a set of sensitivity indices each measuring the relative contribution of every varied input to the variability (uncertainty) of a model's output(s). Some GSA methods also provide information about the input–output mapping, such as threshold values for the inputs or sub-regions of the input variability space that maps into output ranges—for example, extreme values. In the remainder of this paper, we will discuss why it is advantageous to use GSA for impact model evaluation, including why it inherently involves understanding the influence of multiple interacting factors. We will structure our discussion around four specific evaluation questions in the context of impact assessment (named consistency, elasticity, leverage, and attribution), discussing them first in general terms and then through selected, previously published, examples.

In brief, GSA essentially comprises four key steps (Figure 3):

1. *Characterizing variability/uncertainty in the model inputs*, including both numerical inputs such as forcing data and parameters, and “conceptual” ones such as modeling choices and assumptions. Depending on the nature of the input and the information available, input uncertainty may be represented through a probability distribution, range of variability, an ensemble, or a list of plausible values.
2. *Generating input combinations*, typically through statistical sampling from the input distributions defined in Step 1.
3. *Executing the model* for all input combinations and calculating one (or more) summary output metric(s) for each input combinations. Importantly, for future periods, the metric can be a defined using only the model output itself

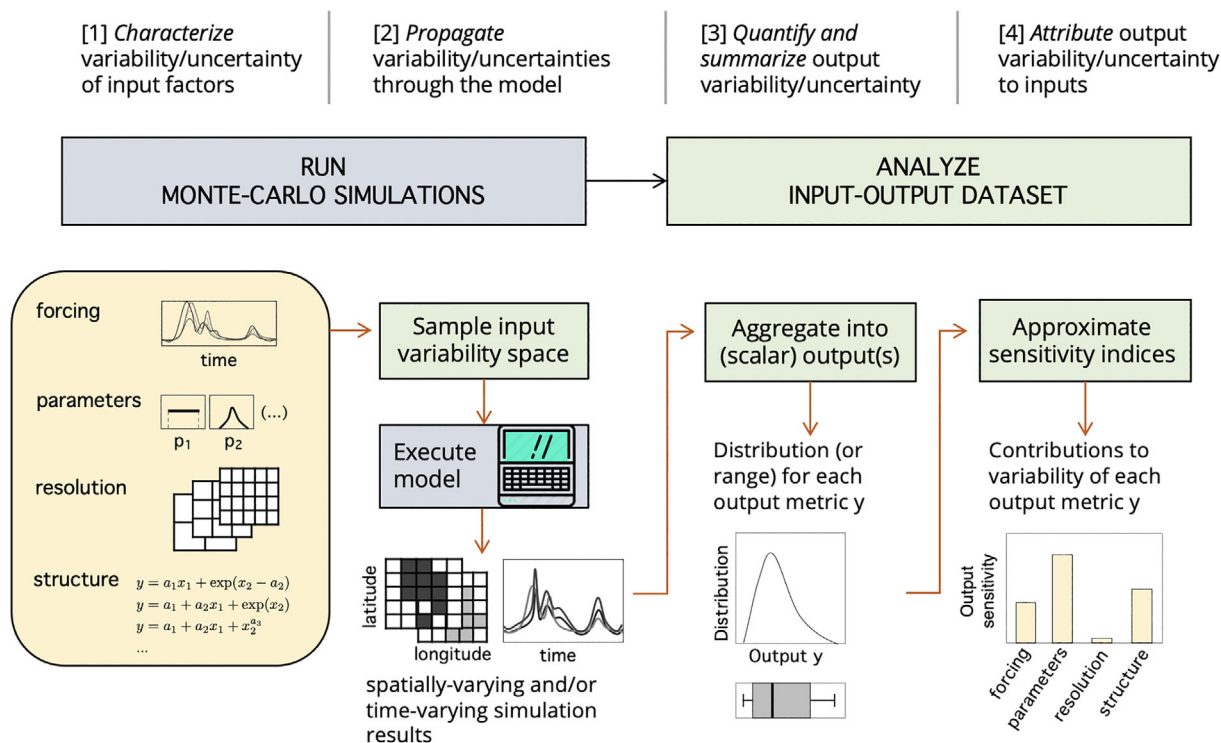


FIGURE 3 The four key steps of global sensitivity analysis (GSA)

(e.g., frequency of occurrence of an event or exceedance of a threshold). The output sample can be used to quantify the output variability/uncertainty.

4. *Analyzing the input–output dataset* to derive a set of sensitivity indices (or input–output mapping information). The definition of the sensitivity indices, and the calculation procedure to approximate their values from the input–output dataset, vary depending on the GSA method. For example, some methods use correlations between inputs and output samples to measure sensitivity, while other methods measure sensitivity to an input through the reduction in output variance (or some other statistic) when fixing that input. Starting points to learn about the different methods and their practical implementation can be found in Saltelli et al. (2000) and Pianosi et al. (2016). We refer to Norton (2015) for analytical approaches to sensitivity analysis and for the use of emulators in this context.

### 3.1 | Consistency: Do modeled process controls match our system perception?

We argue that the “internal consistency” (between model behavior and perception of the real-world system), which Oreskes et al. (1994) demanded, cannot be demonstrated without looking at the input–output relationship (or “response surface”) established by the model. GSA enables us to analyze this input–output relationship and to understand which model parameters (and therefore modeled processes) dominate the variability in the model output, even (if needed) disaggregated in space and time (Wagener & Pianosi, 2019). We can then ask whether the dominant processes in the model are consistent with those we expect to dominate in the real-world system. And, with respect to climate change studies, which of the model parameters (and therefore modeled processes) show sensitivity to changing climatic boundary conditions (Confalonieri et al., 2010). Clearly, we have no way to measure the correct dominant processes under potential future conditions, though, we might have empirical or theoretical arguments to support our expectations.

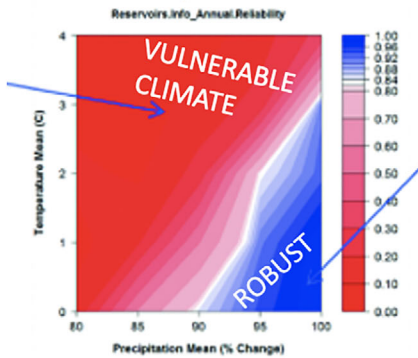
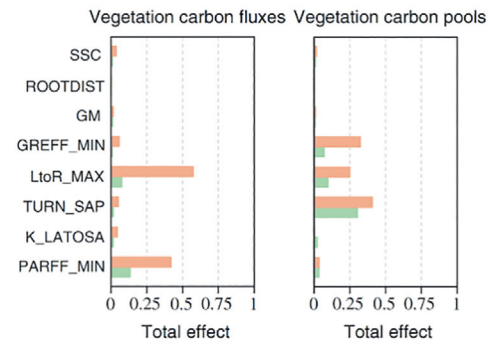
For example, Pappas et al. (2013) utilized GSA to analyze the consistency in the behavior of an ecosystem model, LPJ-GUESS, with both expectations and empirical evidence (Figure 4a). Their analysis revealed that some processes exerted an unexpectedly high or low control on the model output, thus suggesting a need for model structural improvement. Other examples used GSA to reveal an unexpected climate dependence of environmental model parameters (Rosero et al., 2010; Van Werkhoven et al., 2008), implying that this dependence must be considered for scenario analyses. The integration of signatures—discussed in Section 2.1—with GSA can be particularly insightful (Van Werkhoven et al., 2008). There are interesting methodological similarities with pattern-oriented modeling or the use of stylized facts from other areas of modeling such as ecology and integrated assessment, that have yet to be explored (Grimm et al., 2005; Schwanitz, 2013).

### 3.2 | Elasticity: Is my model's sensitivity to changing forcing as expected?

How natural or coupled natural-human systems respond to changing climatic forcing has been challenging impact modelers for a long time (Nemec & Schaake, 1982). For water resources systems, Schaake (1990) introduced the concept of elasticity—originally from economics—to quantify the sensitivity of streamflow to changes in precipitation. Others subsequently defined and studied elasticities with respect to other climatic forcing variables (see review in Sankarsubramanian et al., 2001) using historical data (Vogel et al., 1999) or simulation models (Nijssen et al., 2001; Vano et al., 2012). We use the term elasticity here to describe the analysis of the extent to which a model's response surface is controlled by (potential) climatic forcing. For example, Saltelli et al. (2020) disentangled how different factors influence streamflow in the Colorado River, highlighting how the warming-driven loss of reflective snow increases evapotranspiration. Factor interactions would suggest that GSA should be a helpful tool to disentangle the diverse interaction of uncertain forcing and modeled system responses.

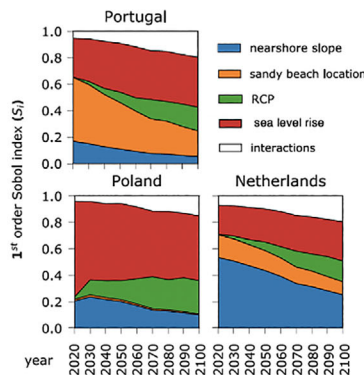
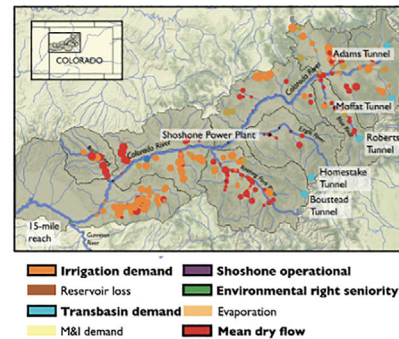
Approaching climate impact assessment as a bottom-up (or “scenario-neutral”; Prudhomme et al., 2010) rather than a top-down (scenario-based) problem has added some energy to the assessment of plausible input (forcing) spaces for impact models (Figure 4b). Bottom-up strategies start by defining stakeholder relevant thresholds to subsequently determine which combinations of climatic forcing produce model outputs above or below these thresholds (Brown et al., 2012; Poff et al., 2016; Wilby & Dessai, 2010). Such bottom-up approaches have been proposed as more appropriate than scenario-modeling given the “deeply” uncertain future we face in a non-stationary world. Deep uncertainty refers to a lack of knowledge or a lack of agreement regarding the probability distribution of parameters describing a

(a) **CONSISTENCY** - Do modelled process controls match my system perception (expectations)? Pappas et al. (2013) used GSA to evaluate the physical realism of an ecosystem model, LPJ-GUESS model, for long-term model applications. They found that the model output's sensitivity to parameters (especially SSC) controlling the water balance was much lower than what would be expected from empirical evidence. The authors' concluded that future model improvement *should concentrate on a more mechanistic representation of plant-water relations*, among other things.



(b) **ELASTICITY** - Is a model's sensitivity to changing forcing adequate? Brown et al. (2019) outlined the so-called 'decision-scaling' strategy to identify suitable decision support options for climate change adaptation. Weather generator tools are used to create an unbiased sample of system responses to plausible climate change. The approach includes a 'climate stress test' based on a multidimensional sensitivity analysis to reveal the sector's sensitivity to climate change and other uncertain factors.

(c) **LEVERAGE** - Do modelled decision levers show adequate sensitivity? Hadjimichael et al. (2020) used GSA to analyze the factors controlling water shortage frequency, including both environmental and management factors (color shows dominant factor). Here, we can see where tributaries shown are likely more sensitive to natural streamflow availability than to human-controlled water demand for an extreme dry year. Quinn et al. (2020) then assessed the value of this approach for exploratory modelling.



(d) **ATTRIBUTION** - What is the attribution of uncertainty sources throughout the projection? Athanasiou et al. (2020) used GSA to study how influential different uncertainties are on coastal land loss projections across the EU under sea level rise (sample shown). Their temporal attribution of uncertainties showed that the choice of geophysical input datasets (i.e. nearshore slope and sandy beach location) can be a significant source of uncertainty.

FIGURE 4 Example studies representative for the four different response-based questions we might ask and address using global sensitivity analysis (GSA). Figures are modified from Pappas et al. (2013), Brown et al. (2019), Hadjimichael et al., (2020), and Athanasiou Saltelli et al. (2020)

system, its boundary, or the system itself (Lempert & Collins, 2007). Such bottom-up approaches have been applied widely, including in hydrology (Singh et al., 2014), ecology (Poff et al., 2016), water resources systems analysis (Ghile et al., 2014; Quinn Saltelli et al., 2020), and natural hazards studies (Almeida et al., 2017).



### 3.3 | Leverage: Do modeled decision levers show adequate influence?

A key purpose of impact models is to clarify the influence of human decisions on the modeled system output under climate change, especially in the context of developing adaptation or intervention strategies (Beltrame et al., 2021). We might, for example, want to understand how much land-use choices like deforestation/reforestation impact the level of downstream flooding under future climate conditions, or we might want to know the value of increased air-conditioning on reducing human losses from excessive heat. We, therefore, must demonstrate that the parameters reflecting these intervention levers (such as those describing human actions exert an adequate control on the model output consistent with our current understanding (Butler et al., 2014). In other words, when assessing leverage, we investigate the impact of those model factors that are under our control—and included as such in our model—in contrast to elasticity which focuses on the system forcing outside our control.

For example, Hadjimichael et al. (2020) set up a diagnostic evaluation framework with GSA at its core to study water scarcity vulnerabilities in institutionally complex river basins (Figure 4c). The authors assessed how hydrological/climatic, institutional, and demand factors impact frequency and duration of water shortages in a subbasin of the Colorado River, USA. Figure 4c depicts a result from their study in which the authors demonstrated that the tributaries shown are likely more sensitive to natural streamflow availability than to human-controlled water demand for an extremely dry year, 2002. Saltelli et al. (2020) subsequently used this approach for exploratory modeling of water scarcity vulnerability under plausible future conditions. An early application of this concept is the study of Pastres et al. (1999), who applied GSA to a water quality model for a shallow water system, and found that the main decision-relevant factor, nitrogen load, was less influential on the occurrence of anoxic crises than the history of the system defined by the initial density of benthic algae. This uncertainty in the initial condition thus prevented the model from being able to identify potential management options.

### 3.4 | Attribution: Can we attribute uncertainty sources throughout the projection horizon?

Uncertainty is an unavoidable aspect of any impact projection, starting from the emission scenarios themselves, their translation into climate projections, the properties of all sub-systems involved, etc. (Clark et al., 2016). These uncertainty sources are nicely visualized in the uncertainty cascade of Wilby and Dessai (2010), who showed the compounding presence of uncertainty through the scenario modeling chain. However, quantification of these uncertainties is rather difficult (Rougier & Crucifix, 2018; Stephenson et al., 2012), especially since they will change along the projection horizon, for example, some will grow the further out we project (Collins et al., 2012). It is therefore valuable to understand which uncertainties dominate during which specific period within the projection horizon. Again, GSA can help by attributing output uncertainty to its sources, so that we can understand their relative importance over time. Although GSA does not solve the problem of having to decide the magnitude of the individual input uncertainties in the first place (see the issue of deep uncertainty discussed above), it nonetheless can guide the modeler regarding the importance of these choices.

For example, Saltelli et al. (2020) used time-varying GSA to reveal the relative contribution of different uncertainties on coastal land loss projections given sea level rise scenarios (Figure 4d). They demonstrated that both the contributions of different uncertain input factors and their temporal evolution varies significantly across EU countries for projections covering the next 80 years. Failing to consider these uncertainties may invalidate conclusions and thus subsequent decisions. Another example of uncertainty attribution along a projection horizon is the study by Le Cozannet et al. (2015) who showed that uncertainty in coastal flood defense vulnerability projections was dominated by local factors such as bathymetry for near-term projections before climate change scenario uncertainty started to dominate.

## 4 | CONCLUSION

Evaluation of climate change impact models—an important task for developing robust adaptation strategies and for gaining stakeholder confidence, and in line with the evaluation of climate models themselves (Eyring et al., 2019)—cannot be based on assessing a model's fit to historical observations alone, even though this approach is still common practice (Eker et al., 2018). Such observation-based assessment can contribute toward establishing confidence in a simulation model (Fowler et al., 2018) and in its ability to perform its intended task (Klemeš, 1986; Refsgaard & Knudsen, 1996).

However, we suggest here that additional evaluation is necessary to achieve greater confidence in model behavior for climate change impact studies, both related to a model being an acceptable system representation and to it being appropriate for the task at hand. In this context, GSA is a valuable tool to complement any observation-based strategy because it allows us to make the model and its simulations significantly more transparent (Razavi et al., 2021; Saltelli et al., 2020; Wagener & Pianosi, 2019).

Appropriate GSA strategies can be defined that enable the modeler to simultaneously address the four evaluation questions we set out in this paper, thus reducing the computational burden that would occur through separate analyses. The four evaluations differ with respect to the input factors included or the way the results need to be interpreted. In summary, (1) Consistency evaluates whether the controlling processes in the model are consistent with expectations or empirical evidence, thus focusing on the model parameters and often relying heavily on the modeler's depth of understanding of the real-world system. (2) Elasticity focuses on evaluating how the modeled system responds to changes in climatic forcing, and whether this response is as expected. (3) Leverage evaluates whether the parameters that represent decision-making levers have sufficient influence on the modeled system response in the presence of other uncertainties. (4) Attribution evaluates which uncertainties dominate the modeled system response at what time, thus guiding potential uncertainty reduction to enhance the model's value.

Recent studies have further demonstrated that GSA can be performed even on highly complex models (Saltelli et al., 2020), on those covering a global domain (Reinecke et al., 2019); and that it can include wider model assumptions, such as model resolution (Savage et al., 2016). However, the complexity of models and therefore the computational burden of such analyses remains the main bottleneck for the widespread application of GSA. Advancing these methods for the purpose of climate change impact study evaluation therefore remains an exciting area of research.

## ACKNOWLEDGMENT

The authors thank the two reviewers for their constructive criticism that helped to improve the paper.

Open Access funding enabled and organized by Projekt DEAL.

## AUTHOR CONTRIBUTIONS

**Robert Reinecke:** Conceptualization (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Francesca Pianosi:** Conceptualization (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Thorsten Wagener:** Conceptualization (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Thorsten Wagener  <https://orcid.org/0000-0003-3881-5849>

Robert Reinecke  <https://orcid.org/0000-0001-5699-8584>

Francesca Pianosi  <https://orcid.org/0000-0002-1516-2163>

## RELATED WIREs ARTICLES

[Changing climate and glacio-hydrology in Indian Himalayan Region: A review](#)

[Climate, weather, and water in history](#)

[Is snow in the Alps receding or disappearing?](#)

[The future of the Nile: Climate change, land use, infrastructure management, and treaty negotiations in a transboundary river basin](#)

[On doing hydrology with dragons: Realizing the value of perceptual models and knowledge accumulation](#)

## FURTHER READING

Athanasiou, P., van Dongeren, A., Giardino, A., Voudoukas, M. I., Ranasinghe, R., & Kwadijk, J. (2020). Uncertainties in projections of sandy beach erosion due to sea level rise: An analysis at the European scale. *Scientific Reports*, 10, 11895. <https://doi.org/10.1038/s41598-020-68576-0>

- Brown, C., Steinschneider, S., Ray, P., Wi, S., Basdekas, L. and Yates, D. (2019). Decision Scaling (DS): decision support for climate change. In Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M. and Popper, S.W. (eds.) *Decision Making under Deep Uncertainty*. Springer, 255–287. <https://doi.org/10.1007/978-3-030-05252-2>
- Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J. K., & McMillan, H. K. (2021). TOSSH: A toolbox for streamflow signatures in hydrology. *Environmental Modelling & Software*, 138, 104983.
- Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research*, 54(11), 8692–8717.
- Hadjimichael, A., Quinn, J., and Reed, P. (2020). Advancing diagnostic model evaluation to better understand water shortage mechanisms in institutionally complex river basins. *Water Resources Research*, 56, e2020WR028079. <https://doi.org/10.1029/2020WR028079>
- Hornberger, G. M., & Spear, R. C. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, 12, 7–18.
- Maples, S., Foglia, L., Fogg, G. E., & Maxwell, R. M. (2020). Sensitivity of hydrologic and geologic parameters on recharge processes in a highly-heterogeneous, semi-confined aquifer system. *Hydrology and Earth System Sciences*, 24, 2437–2456. <https://doi.org/10.5194/hess-24-2437-2020>
- Milly, P. C., & Dunne, K. A. (2020). Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. *Science*, 367(6483), 1252–1255. <https://doi.org/10.1126/science.aay9187>
- Singh, R., Wagener, T., van Werkhoven, K., Mann, M. E., & Crane, R. (2011). A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate—Accounting for changing watershed behavior. *Hydrology and Earth System Sciences*, 15(11), 3591–3603.

## REFERENCES

- Almeida, S., Holcombe, E. A., Pianosi, F., & Wagener, T. (2017). Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change. *Natural Hazards and Earth System Sciences*, 17(2), 225–241.
- Barron, E. J. (2009). Beyond climate science. *Science*, 326, 643.
- Beltrame, L., Rose Vineer H., Walker J.G., Morgan E.R., Vickerman P., Wagener T. (2021). Discovering environmental management opportunities for infectious disease control. *Scientific Reports*, 11(1), 1–10. [doi.org/https://doi.org/10.1038/s41598-021-85250-1](https://doi.org/10.1038/s41598-021-85250-1).
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, 40, 1–20.
- Blöschl, G., Sivapalan, M., Wagener, T., Savenije, H., & Viglione, A. (2013). *Runoff predictions in ungauged basins: Synthesis across processes, places and scales*. Cambridge University Press.
- Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791–799. <https://doi.org/10.1080/01621459.1976.10480949>
- Brown, C., Ghile, Y., Lavery, M., & Li, K. (2012). Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. *Water Resources Research*, 48(9), 1–12.
- Brunner, M. I., Melsen, L. A., Wood, A. W., Rakovec, O., Mizukami, N., Knoben, W. J. M., & Clark, M. P. (2021). Flood spatial coherence, triggers, and performance in hydrological simulations: Large-sample evaluation of four streamflow-calibrated models. *Hydrology and Earth System Sciences*, 25, 105–119. <https://doi.org/10.5194/hess-25-105-2021>
- Butler, M. P., Reed, P. M., Fisher-Vanden, K., Keller, K., & Wagener, T. (2014). Inaction and climate stabilization uncertainties lead to severe economic risks. *Climatic Change*, 127, 463–474. <https://doi.org/10.1007/s10584-014-1283-0>
- Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R., & Brekke, L. D. (2016). Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, 2(2), 55–64.
- Collins, M., Chandler, R. E., Cox, P. M., Huthnance, J. M., Rougier, J. C., & Stephenson, D. B. (2012). Quantifying future climate change. *Nature Climate Change*, 2, 403–409. <https://doi.org/10.1038/nclimate1414>
- Confalonieri, R., Belloccchi, G., Tarantola, S., Acutis, M., Donatelli, M., & Genovese, G. (2010). Sensitivity analysis of the rice model WARM in Europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environmental Modelling & Software*, 25(4), 479–488.
- Eker, S., Rovenskaya, E., Obersteiner, M., & Langan, S. (2018). Practice and perspectives in the validation of resource management models. *Nature Communications*, 9, 5359.
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, 17(5), 1893–1912.
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., ... Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9, 102–110. <https://doi.org/10.1038/s41558-018-0355-y>
- Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., & Zhang, L. (2018). Simulating runoff under changing climatic conditions: A framework for model improvement. *Water Resources Research*, 54(12), 9812–9983.

- Ghile, Y. B., Taner, M. U., Brown, C. M., & Grijzen, J. G. (2014). Bottom-up climate risk assessment of infrastructure investment in The Niger River Basin. *Climatic Change*, *122*(1–2), 97–111.
- Gleeson, T., Wang-Erlandsson, L., Porkka, M., Zipper, S. C., Jaramillo, F., Gerten, D., Fetzer, I., Cornell, S. E., Piemontese, L., Gordon, L. J., Rockström, J., Oki, T., Sivapalan, M., Wada, Y., Brauman, K. A., Flörke, M., Bierkens, M. F. P., Lehner, B., Keys, P., ... Famiglietti, J. S. (2020). Illuminating water cycle modifications and Earth System resilience in the Anthropocene. *Water Resources Research*, *56*(4), e2019WR024957.
- Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., Breuer, L., Ding, J., Haddeland, I., Kumar, R., Kundu, D., Liu, J., van Griensven, A., Veldkamp, T. I. E., Vetter, T., Wang, X., & Zhang, X. (2017). A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1°C, 2°C and 3°C. *Climatic Change*, *141*(3), 577–595. <https://doi.org/10.1007/s10584-016-1773-3>
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H. H., Weiner, J., Wiegand, T., & Deangelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, *310*(5750), 987–991. <https://doi.org/10.1126/science.1116681>
- Gupta, H.V., Wagener, T. and Liu, Y. (2008). Reconciling theory with observations: Towards a diagnostic approach to model evaluation. *Hydrological Processes*, *22*. <https://doi.org/10.1002/hyp.6989>.
- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*(8), 1–16.
- Hakala, K., Addor, N., & Seibert, J. (2018). Hydrological modeling to evaluate climate model simulations and their bias correction. *Journal of Hydrometeorology*, *19*(8), 1321–1337. <https://doi.org/10.1175/jhm-d-17-0189.1>
- Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Florke, F., Gosling, S. N., Hoffmann, P., Liersch, S., Masaki, Y., Motovilov, Y., Muller, C., Samaniego, L., Stacke, T., Wada, Y., Yang, T., & Krysanova, V. (2018). Sources of uncertainty in hydrological climate impact assessment: A cross-scale study. *Environmental Research Letters*, *13*(1), 1–14.
- Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., & Gascuel-Oudou, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, *50*, 7445–7469.
- Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13–24.
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331.
- Kollet, J. B., Reed, P. M., & Wagener, T. (2012). When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resources Research*, *48*(3), 1–19.
- Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., & Kundzewicz, Z. (2018). How the performance of hydrological models relates to credibility of projections under climate change. *Hydrological Sciences Journal*, *63*(5), 696–720. <https://doi.org/10.1080/02626667.2018.1446214>
- Le Cozannet, G., Rohmer, J., Cazenave, A., Idier, D., van de Wal, R., de Winter, R., Pedreros, R., Balouin, Y., Vinchon, C., & Oliveros, C. (2015). Evaluating uncertainties of future marine flooding occurrence as sealevel rises. *Environmental Modeling & Software*, *73*, 44–56.
- Lempert, R. J., & Collins, M. (2007). Managing the risk of an uncertain threshold response: Comparison of robust, optimum, and precautionary approaches. *Risk Analysis*, *27*(4), 1009–1026.
- Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., ... Winter, L. (2009). A formal framework for scenario development to support environmental decision-making. *Environmental Modeling and Software*, *24*, 798–808.
- McMillan, H. K. (2021). A review of hydrologic signatures and their applications. *Wiley Interdisciplinary Reviews: Water*, *8*(1), e1499.
- Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2), 1–17.
- Milly, P. C., & Dunne, K. A. (2017). A hydrologic drying bias in water-resource impact analyses of anthropogenic climate change. *Journal of the American Water Resources Association*, *53*(4), 822–838.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. & Veith, T.L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*, 885–900. <https://doi.org/10.13031/2013.23153>
- Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, *18*, 3511–3538. <https://doi.org/10.5194/hess-18-3511-2014>
- Nemec, J., & Schaake, J. C. (1982). Sensitivity of water resources systems to climate variation. *Hydrological Sciences Journal*, *27*, 327–343.
- Nijssen, B., O'Donnell, G. M., Hamlet, A. F., & Lettenmaier, D. P. (2001). Hydrologic sensitivity of global rivers to climate change. *Climatic Change*, *50*(1), 143–175.
- Norton, J. (2015). An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software*, *69*, 166–174. <https://doi.org/10.1016/j.envsoft.2015.03.020>
- Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641–646.
- Pappas, C., Fatichi, S., Leuzinger, S., Wolf, A., & Burlando, P. (2013). Sensitivity analysis of a process-based ecosystem model: Pinpointing parameterization and structural issues. *Journal of Geophysical Research – Biogeosciences*, *118*, 505–528. <https://doi.org/10.1002/jgrg.20035>



- Pastres, R., Chan, K., Solidoro, C., & Dejak, C. (1999). Global sensitivity analysis of a shallow-water 3D eutrophication model. *Computer Physics Communications*, *117*, 62–74.
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232.
- Poff, N. L. R., Brown, C. M., Grantham, T. E., Matthews, J. H., Palmer, M. A., Spence, C. M., Wilby, R. L., Haasnoot, M., Mendoza, G. F., Dominique, K. C., & Baeza, A. (2016). Sustainable water management under future uncertainty with eco-engineering decision scaling. *Nature Climate Change*, *6*(1), 25–34.
- Prudhomme, C., Wilby, R.L., Crooks, S., Kay, A.L. & Reynard, N.S. (2010). Scenario-neutral approach to climate change impact studies: Application to flood risk. *Journal of Hydrology*, *390*(3–4), 198–209. doi.org/https://doi.org/10.1016/j.jhydrol.2010.06.043
- Quinn, J. D., Hadjimichael, A., Reed, P. M., & Steinschneider, S. (2020). Can exploratory modeling of water scarcity vulnerabilities and robustness be scenario neutral? *Earth's Futures*, *8*, e2020EF001650. https://doi.org/10.1029/2020EF001650
- Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., ... Maier, H. R. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, *137*, 104954.
- Refsgaard, J. C., & Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, *32*, 2189–2202. https://doi.org/10.1029/96WR00896
- Reinecke, R., Foglia, L., Mehl, S., Herman, J. D., Wachholz, A., Trautmann, T., & Döll, P. (2019). Spatially distributed sensitivity of simulated global groundwater heads and flows to hydraulic conductivity, groundwater recharge, and surface water body parameterization. *Hydrology and Earth System Sciences*, *23*(11), 4561–4582.
- Reinecke, R., Hannes, M. S., Trautmann, T., Andersen, L. S., Burek, P., Flörke, M., Gosling, S. N., Grillakis, M., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Thiery, W., Wada, Y., Yusuke, S., & Döll, P. (2021). Uncertainty of simulated groundwater recharge at different global warming levels: A global-scale multi-model ensemble study. *Hydrology and Earth System Sciences*, *25*(2), 787–810. https://doi.org/10.5194/hess-25-787-2021
- Rockström, J., Gupta, J., Lenton, T. M., Qin, D., Lade, S. J., Abrams, J. F., Jacobson, L., Rocha, J. C., Zimm, C., Bai, X., Bala, G., Bringezu, S., Broadgate, W., Bunn, S. E., DeClerck, F., Ebi, K. L., Gong, P., Gordon, C., Kanie, N., ... Winkelmann, R. (2021). Identifying a safe and just corridor for people and the planet. *Earth's Futures*, *9*, 1–7. https://doi.org/10.1029/2020EF001866
- Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L. E., Yatheendradas, S., & Niu, G.-Y. (2010). Quantifying parameter sensitivity, interaction and transferability in hydrologically enhanced versions of Noah-LSM over transition zones. *Journal of Geophysical Research*, *115*, D03106. https://doi.org/10.1029/2009JD012035
- Rougier, J. C., & Crucifix, M. (2018). Uncertainty in climate science and climate policy. In E.A. Lloyd & E. Winsberg (eds.) *Climate Modelling*, 361–380.
- Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Espeland, W. N., Kay, J., Piano, S. L., Mayo, D., Pielke, R., Jr., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., ... Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, *582*, 482–484.
- Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modelling. *Statistical Science*, *15*(4), 377–395.
- Sankarsubramanian, A., Vogel, R. M., & Limbrunner, J. F. (2001). Climate elasticity of streamflow in the United States. *Water Resources Research*, *37*(6), 1771–1781.
- Savage, J.T.S., Pianosi, F., Bates, P., Freer, J. and Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model, *Water Resour. Res.*, *52*, 9146–9163. https://doi.org/10.1002/2015WR018198.
- Schaake, J. C. (1990). From climate to flow. In P. E. Waggoner (Ed.), *Climate change and U.S. water resources*, Chapter 8 (pp. 177–206). John Wiley.
- Schwanitz, V. J. (2013). Evaluating integrated assessment models of global climate change. *Environmental Modelling & Software*, *50*, 120–131. https://doi.org/10.1016/j.envsoft.2013.09.005
- Schaefli, B. and Gupta, H.V. (2007). Do Nash values have value? *Hydrological Sciences*, *21*, 2075–2080. https://doi.org/10.1002/hyp.6825
- Seibert, J., Uhlenbrook, S., Leibundgut, C. & Halldin, S. (2000). Multiscale calibration and validation of a conceptual rainfall-runoff model. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, *25*(1), 59–64. doi.org/https://doi.org/10.1016/S1464-1909(99)00121-5
- Singh, R., van Werkhoven, K., & Wagener, T. (2013). Climate change impacts in gauged and ungauged basins of the Olifants River, South Africa: A trading space for time approach. *Hydrological Sciences Journal*, *59*, 29–55. https://doi.org/10.1080/02626667.2013.819431
- Singh, R., Wagener, T., Crane, R., Mann, M. E., & Ning, L. (2014). A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds. *Water Resources Research*, *50*(4), 3409–3427.
- Smith, K. A., Wilby, R. L., Broderick, C., Prudhomme, C., Matthews, T., Harrigan, S., & Murphy, C. (2018). Navigating cascades of uncertainty—As easy as ABC? Not quite. *Journal of Extreme Events*, *5*(01), 1850007.
- Stephenson, D. B., Collins, M., Rougier, J. C., & Chandler, R. E. (2012). Statistical problems in the probabilistic prediction of climate change. *Environmetrics*, *23*(5), 364–372.
- van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, *44*, W01429. https://doi.org/10.1029/2007WR006271

- Vano, J. A., Das, T., & Lettenmaier, D. P. (2012). Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *Journal of Hydrometeorology*, 13(3), 932–949. <https://doi.org/10.1175/JHM-D-11-069.1>
- Vogel, R. M., Wilson, I., & Daly, C. (1999). Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, 125(3), 148–157.
- Wagener, T., Gleeson, T., Coxon, G., Hartmann, A., Howden, N., Pianosi, F., Rahman, M., Rosolem, R., Stein, L., & Woods, R. (2021). On doing hydrology with dragons: Realizing the value of perceptual models and knowledge accumulation. *WIREs Water*, 8, 1–7. <https://doi.org/10.1002/wat2.1550>
- Wagener, T., & McIntyre, N. (2005). Identification of rainfall–runoff models for operational applications [Identification de modèles pluie–débit pour des applications opérationnelles]. *Hydrological Sciences Journal*, 50(5), 735–750.
- Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B., & Wilson, J. S. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research*, 46, W05301.
- Wagener, T., & Pianosi, F. (2019). What has global sensitivity analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. *Earth-Science Reviews*, 194, 1–18. <https://doi.org/10.1016/j.earscirev.2019.04.006>
- Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*, 65(7), 180–185. <https://doi.org/10.1002/wea.543>

**How to cite this article:** Wagener, T., Reinecke, R., & Pianosi, F. (2022). On the evaluation of climate change impact models. *Wiley Interdisciplinary Reviews: Climate Change*, e772. <https://doi.org/10.1002/wcc.772>