

Evaluation of parameter sensitivity of a rainfall-runoff model over a global catchment set

Léonard Santos, Jafet C. M. Andersson & Berit Arheimer

To cite this article: Léonard Santos, Jafet C. M. Andersson & Berit Arheimer (2022) Evaluation of parameter sensitivity of a rainfall-runoff model over a global catchment set, Hydrological Sciences Journal, 67:3, 342-357, DOI: [10.1080/02626667.2022.2035388](https://doi.org/10.1080/02626667.2022.2035388)

To link to this article: <https://doi.org/10.1080/02626667.2022.2035388>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 01 Mar 2022.



[Submit your article to this journal](#)



Article views: 622



[View related articles](#)



[View Crossmark data](#)

Evaluation of parameter sensitivity of a rainfall-runoff model over a global catchment set

Léonard Santos , Jafet C. M. Andersson  and Berit Arheimer 

Hydrology Research, Swedish Meteorological and Hydrological Institute (Smhi), Norrköping, Sweden

ABSTRACT

This paper presents an evaluation of the parameter sensitivity of a process-based model at the global scale using large-sample data. The analysis was carried out using the Hydrological Prediction of the Environment (HYPE) model, for which soil and snow parameters were evaluated using 187 river flow gauges spread worldwide. As a result, 6 out of 12 soil parameters and 7 out of 10 snow parameters were found to be sensitive. Taking advantage of the global dataset, an additional analysis was used to investigate links between catchment characteristics and parameter sensitivity. Different patterns of sensitivity were observed for different Köppen climate classes, which indicates that parameter regionalization would benefit from calibration based on climate zones. This numerical sensitivity method was compared with the judgement of a set of expert HYPE modellers to understand how numerical results compare with modellers' experience.

ARTICLE HISTORY

Received 14 December 2020
Accepted 18 November 2021

EDITOR

A. Castellarin

ASSOCIATE EDITOR

G. Ayzel

KEYWORDS

large-sample data; Monte Carlo; Saltelli indices; parameter sensitivity; calibration procedure; artificial intelligence; hydrological intuition; global catchment hydrology

1 Introduction

In the last few decades, several global applications of hydrological models have been launched (Bierkens *et al.* 2015, Sood and Smakhtin 2015). The first global hydrological models were physically based land surface models coupled with meteorological or climate models (Wood *et al.* 2011), but recently several continental applications have been made with more empirical and conceptual rainfall-runoff models distributed over catchments or grids (Archfield *et al.* 2015). When applying hydrological models at a large scale, one important issue to address is how to parameterize the model in a geographical domain where the majority of catchments are ungauged (Blöschl *et al.* 2013, Bierkens *et al.* 2015). Numerical models normally suffer from equifinality (Beven 2006) as they use many interacting parameters. It is then important to select the most sensitive parameters for calibration to avoid uncertainty and time-consuming calculations. Since the models are often initially designed for and applied to specific locations or river basins, the parameter selection and identification are generally not adapted to the spatial variability seen in global modelling. A large model domain induces some specificity in the parameter selection, and hence the method used to make this selection needs to be adapted to the global scale in order to balance depth with breadth (Gupta *et al.* 2014).

Different methods exist to select parameters to calibrate. Some modellers select parameters based on expert knowledge of the model (e.g. Hirpa *et al.* 2018). Other modellers use numerical methods, which are often based on sensitivity analysis (e.g. Beven and Binley 1992, Muleta and Nicklow 2005,

Roux *et al.* 2011). These methods are designed to estimate which parameters influence model outputs the most, by running the model with different sets of parameter values. Sensitivity analysis is regularly used in hydrology (e.g. Song *et al.* 2015): Razavi and Gupta (2015) showed that 3% of works on water resources published between 1980 and 2014 match the term “sensitivity” in the Thomson Reuters Web of Science data. Since the numerical methods and the expert judgements have the same goal, it is interesting to compare the results obtained with these two types of methods.

One critical issue for applying sensitivity analysis in large-scale model applications is that it demands heavy computational processing; hence, even if studies were performed at a large scale (Hou *et al.* 2012, Markstrom *et al.* 2016) or the global scale (Chaney *et al.* 2015, Zajac *et al.* 2017, Reinecke *et al.* 2019), most studies so far are only based on one or two catchments. Pfannerstill *et al.* (2015) and Guse *et al.* (2016) showed that a link can be found between parameter sensitivity and process dominance, which is not spatially homogeneous but may vary significantly between catchments. Several recent studies have highlighted the spatial variation of flood-generating processes due to physiographical characteristics (e.g. Hundecha *et al.* 2016, Kuentz *et al.* 2017, Blöschl *et al.* 2019, Stein *et al.* 2020). Accordingly, parameter sensitivity is likely to vary between catchments (as shown by e.g. Van Werkhoven *et al.* 2008, Hou *et al.* 2012, Zajac *et al.* 2017), and a unique sensitivity analysis based on only a few catchments may not be adequate to select parameters for a large-scale model application covering hundreds or thousands of

CONTACT Léonard Santos  leonard.santos@inrae.fr  Swedish Meteorological and Hydrological Institute (Smhi), Norrköping, Sweden

 Supplemental data for this article can be accessed [here](#).

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

catchments, which is the case in global catchment models. For instance, Van Werkhoven *et al.* (2008), Melsen and Guse (2019) and Basijokaite and Kelleher (2021) showed a link between climate and parameter sensitivity for hydrological simulation. If climate has an impact on sensitivity, it is then necessary to analyse it on a set of catchments representative of as many climate classes as possible.

This paper proposes, using an appropriate method, to explore the change in parameter sensitivity over catchments with very different characteristics in order to better select parameters for calibration of a hydrological model at the global scale. We exemplify the method using the process-oriented HYdrological Prediction of the Environment (HYPE) code (Lindström *et al.* 2010) in the World Wide HYPE model setup (Arheimer *et al.* 2020). A variance-based sensitivity analysis was carried out for 187 catchments distributed around the globe that are affected by different climates and dominated by different hydrological processes. The aim is to better understand how process-based model parameter sensitivity is linked to the dominance of hydrological processes that govern the runoff generation mechanism at the global scale. The results were also compared to the experience of a set of expert HYPE modellers to evaluate whether the numerical sensitivity analysis is reflected in their intuitive approaches.

2 Methods and data

2.1 HYPE model description

The HYPE model is a semi-distributed process-based hydrological model. It was initiated in 2002 for water quality purposes (Arheimer and Lindström 2013) but was soon proved to be efficient also for flood and drought forecasting in Sweden (Pechlivanidis *et al.* 2014) and Europe (Pechlivanidis *et al.* 2020). The model has been applied at increasing large scales: country scale (e.g. Strömquist *et al.* 2012), international river basins (e.g. Niger HYPE, Andersson *et al.* 2017), continental scale (e.g. E-HYPE, Donnelly *et al.* 2016) and global scale (World Wide HYPE Arheimer *et al.* 2020).

In terms of process representation, HYPE can be described as process-based even if the hydrological processes are simplified. In terms of spatial representation, the model is based on hydrologically connected catchments, which are further split into hydrological response units (HRUs). The HRUs can be delimited by different combinations of soil, land cover or other elements such as elevation. The process descriptions are spatialized either by catchments (e.g. for lakes or river routing) or by HRUs (e.g. the soil processes). To summarize the river discharge modelling in HYPE, runoff is produced at the HRU scale, routed within the catchment, and then routed between the different catchments and water bodies to the river outlet.

The model can take different hydrological features into account. For example, lakes, floodplains or glaciers can be modelled to calculate the routing of water. As this paper focus on runoff generation, only runoff generation processes will be described here. For soil, this description is also limited to the part of the soil model that is used for the sensitivity analysis. All parameters and processes that can be used in HYPE are described fully in the HYPE model documentation (<http://www.smhi.net/hype/wiki/doku.php>).

The main processes of runoff generation from soil in HYPE are described in Fig. 1. The soil can be represented by three layers. In these three layers, the capacity of water storage is given by two dimensionless soil parameters: the field capacity (w_{cfc} , for water that is only available for evapotranspiration) and the effective porosity (w_{cep} , for moving water that is available for evapotranspiration, percolation and subsurface runoff). The effective porosity is filled once the field capacity is entirely filled. Another part of soil water is more stable and not impacted by any of these three processes. The soil storage capacity at the wilting point (parameter w_{cwp} , also with no dimension) drives the amount of water in this case. Since it only influences water quality, this parameter will not be taken into account in the sensitivity analysis. These three parameters are non-dimensional fractions of the soil layer thickness, and their values are between 0 and 1. They have to be multiplied by the layer thickness to obtain a capacity in millimetres. The values of w_{cfc} , w_{cep} and w_{cwp} can vary for the different soil layers, but in this paper they were kept identical across layers

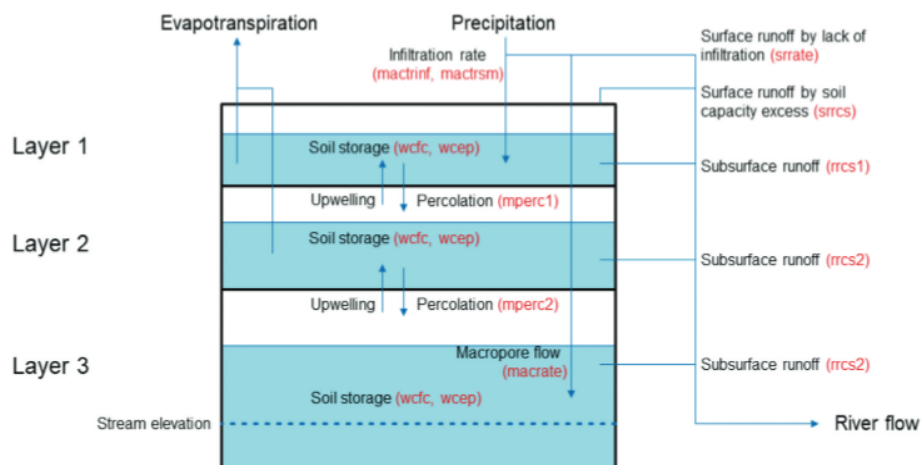


Figure 1. Simplified diagram of HYPE model soil processes, with the corresponding model parameters analysed in this study in brackets (red).

to avoid an increase in the number of parameters. Evapotranspiration water is removed from water in the field capacity and effective porosity stores of the two first soil layers by a linear relationship between water content in the soil and potential evapotranspiration. For the two layers, the potential evapotranspiration is reduced exponentially by a coefficient that depends on the depth of each layer.

The infiltration rate of the rainfall in the first layer depends on rainfall intensity (parameter *mactrinf* in mm days⁻¹) and soil moisture storage (parameter *mactrsm*, with no dimension). The surface runoff can also occur when layer 1 is fully saturated: the additional water will reach the river as overland flow and is regulated by the parameter *srrcs* (in days⁻¹). The water that did not infiltrate will either join the surface runoff (this proportion is given by the parameter *srrate*, with no dimension), or join the water in layer 3 by macropore flow in the soil (regulated by the parameter *macrate*, with no dimension). Within the soil, the water that is in the effective porosity can reach the river as subsurface flow (the recession of this flow is given by the parameters *rrcs1* for layer 1 and *rrcs2* for layers 2 and 3, in days⁻¹). The subsurface flow from layer 3 only occurs if the water level is higher than the stream elevation. The water in the effective porosity can also percolate to the layer below if this layer is not entirely filled (parameters *mperc1* for layer 1 and *mperc2* for layer 2, in mm). Last, if the layer is entirely filled, the water will move to the upper layer by upwelling. Finally, frozen soil can be taken into account, but as it is not used for this paper it will not be described here. Table 1 summarizes the soil parameters that are taken into account for this sensitivity analysis and their role in runoff generation.

In addition to soil processes, snow is also taken into account in this paper, because it influences runoff generation more than soil processes do in some regions. Precipitation is considered to occur as snow if the temperature is below the land cover dependent parameter *ttmp* (in °C). The accumulation of snow is not uniformly distributed over the different HRUs; the snow cover is driven by three snow distribution parameters that are land cover dependent (*fscdistmax* and *fscdist0*, which are non-dimensional; and *fscdist1*, in m⁻¹). When the temperature is over *ttmp*, the snow melts. The melting coefficient depends on temperature (with parameter *cmlt* in mm °C⁻¹ days⁻¹) and on radiation (with parameter *cmrad* in mm m² MJ⁻¹). The melt coefficient due to radiation will also depend on snow albedo, which is a function of the age of this snow (calculated with parameters *snalbmin* and *snalbmax* with no dimension and *snalbexp* in days⁻¹). Last, the evaporation from snow is calculated based on potential evapotranspiration and parameterized by the fraction *fepotsnow* (with no dimension). The snow parameters used for this sensitivity analysis and their functions are summarized in Table 2.

Other parameters of the model, designed to tune precipitation adjustment, evapotranspiration from soil and routing in rivers remain constant for this experiment. The values of these parameters were kept from Arheimer *et al.* (2020). This choice was made because, for the catchments used for the study, the soil and snow processes were major compared to other processes.

Being a process-based model, HYPE is a good candidate to evaluate the relationship between hydrological processes, catchment characteristics and sensitivity of the parameters. For this reason, we applied a sensitivity analysis to these

Table 1. List and description of HYPE soil parameters studied.

Parameter name	Role in runoff generation	Unit
<i>rrcs1</i>	Recession coefficient for soil layer 1	[Δt ⁻¹]
<i>rrcs2</i>	Recession coefficient for soil layers 2 and 3	[Δt ⁻¹]
<i>mperc1</i>	Max percolation capacity from soil layer 1 to 2	[mm]
<i>mperc2</i>	Max percolation capacity from soil layer 2 to 3	[mm]
<i>mactrinf</i>	Infiltration capacity threshold	[mm Δt ⁻¹]
<i>macrate</i>	Ratio of macropore flow (when infiltration capacity is exceeded)	[-]
<i>srrate</i>	Ratio of surface runoff (when infiltration capacity is exceeded)	[-]
<i>mactrsm</i>	Soil moisture ratio threshold for macropore flow	[-]
<i>wcwp</i>	Wilting point for soil layers 1, 2 and 3	[-]
<i>wcfc</i>	Field capacity for soil layers 1, 2 and 3	[-]
<i>wcep</i>	Effective porosity for soil layers 1, 2 and 3	[-]
<i>srrcs</i>	Recession coefficient for saturation excess surface runoff	[Δt ⁻¹]

Table 2. List and description of HYPE snow parameters.

Parameter name	Role in runoff generation	Unit
<i>ttmp</i>	Relative threshold temperature for snow melt	[°C]
<i>cmlt</i>	Melting coefficient for snow	[mm °C ⁻¹ Δt ⁻¹]
<i>cmrad</i>	Radiation-based melting coefficient	[mm m ² MJ ⁻¹]
<i>snalbmin</i>	Minimum albedo for snow	[-]
<i>snalbmax</i>	Maximum albedo for snow	[-]
<i>snalbexp</i>	Coefficient for aged snow albedo decrease	[Δt ⁻¹]
<i>fscdistmax</i>	Maximum snow distribution coefficient	[-]
<i>fscdist0</i>	Minimum snow distribution coefficient	[-]
<i>fscdist1</i>	Snow distribution slope coefficient	[m ⁻¹]
<i>fepotsnow</i>	Snow evaporation fraction	[-]

parameters to evaluate their influence on the model results and, by correspondence, the influence of the different processes on the studied catchments.

2.2 Sensitivity method description

Two types of sensitivity analysis exist (Razavi and Gupta 2015). First, the local methods aim to start from a given point in the parameter space and to evaluate the effect of perturbations. The second type of methods, the global ones, aim to evaluate what happens over the entire parameter space (if only a part of the parameter space is prioritized, these methods can be called regional). Local methods are more computationally efficient; however, they are often not sufficient to capture the overall sensitivity (Saltelli and Annoni 2010). For this study, a regional sensitivity analysis was applied, since HYPE computational time on small catchments is moderate.

The sensitivity of the parameters was analysed with respect to the Kling and Gupta efficiency (E_{KG} , Gupta *et al.* 2009) value of monthly simulated discharge compared with discharge observations at 187 gauged catchments. This score varies from minus infinity to 1 with good values being equal to 1. Large negative values taken by the E_{KG} can strongly influence the variance-based sensitivity results and hide parameter influences for E_{KG} values that are above zero. For this reason, E_{KG} values were here limited to vary between -1 and 1 through normalization using the following equation (Mathevet *et al.* 2006):

$$E_{lim} = \frac{E_{KG}}{2 - E_{KG}} \quad (1)$$

This transformation has relatively low impact for positive E_{KG} values but compresses all the possible negative values between -1 and 0. Since the transformation is strictly an increasing function, higher E_{KG} values mean higher E_{lim} (limited Kling and Gupta efficiency) values.

The sensitivity of E_{lim} to each parameter was evaluated using two steps: a Monte Carlo sampling followed by a Sobol' variance-based analysis. First, 50 000 parameter sets were sampled using random and uniform Monte Carlo sampling (the maximum affordable given computational time constraints). The results of the 50 000 runs are classified by the obtained E_{lim} . The 5000 best simulations are used to estimate the initial distribution of each parameter for the variance-based analysis as it is used, for example, in Generalized

Likelihood Uncertainty Estimation (GLUE, Beven and Binley 1992). Indeed, the parameters' initial distributions have a high impact on the results of these analyses. These initial distributions can be estimated based on a literature review if the parameters have a physical meaning, or from expert knowledge if they do not. In the case of this study, a priori knowledge is limited at the global scale. For this reason, these distributions are chosen from the Monte Carlo analysis, which allows us to place greater emphasis on the area of the parameter space in which E_{lim} values are acceptable. From these distributions, the Saltelli (2002) sampling method is applied to calculate both first-order (sensitivity to the parameter by itself) and total-effect (sensitivity to a parameter taking into account its interactions with others) indices.

For this sampling, two matrixes of 2500 parameter sets each are built by splitting the 5000 parameter sets using their rank in terms of E_{lim} . Parameter sets with an odd rank fill the first matrix while parameter sets with an even rank fill the second one. The matrixes are then combined using Saltelli's (2002) methodology: for each parameter, the first matrix is duplicated but by replacing the column that corresponds to the parameter with the corresponding column of the second matrix. As a result, the initial number of parameter sets (2500) is multiplied by the number of analysed parameters plus two (the two initial matrixes) to obtain the total number of parameter sets to run. For no-snow catchments (11 analysed parameters) this calculation gives a number of runs equal to 32 500, and for snow catchments (17 analysed parameters) it gives 47 500.

The combination of these two steps gives an idea of the parameters that produce the most reaction in terms of E_{lim} values. The results of the Monte Carlo runs are used to give a visual overview of the sensitivity of E_{lim} to each parameter over the entire parameter space. This overview is enhanced by the results of the variance-based analysis that give a numerical estimation of the first-order influence and total influence of each parameter. Figure 2 summarizes the methodology used to carry out this sensitivity analysis.

2.3 Comparison with expert knowledge on the model

As a complement, the results of the numerical method described above were compared to the choices of 10 experienced HYPE modellers. The aim here was to analyse the difference between the results of the numerical experiment

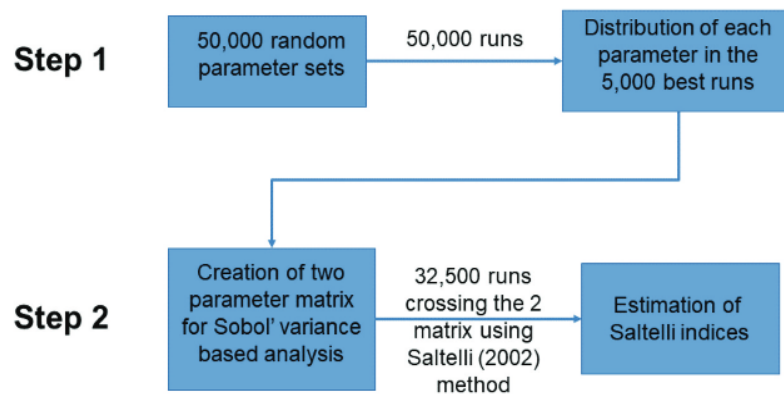


Figure 2. Diagram explaining the sensitivity analysis methodology used in this paper. This methodology was reproduced for each catchment.

and those from expert judgement. A survey was addressed to hydrological modellers at SMHI (the Swedish meteorological and hydrological institute) who were asked to specify the parameters they chose when calibrating HYPE for a large domain with many gauging sites.

To be consistent with this study, several questions were asked. Mainly, the modellers were asked which runoff-generation parameters they used, along with some questions about model set-up, such as geographical domain, modelling purpose, spatial and temporal resolution, etc. In total, information about the parameter choices for calibration from 12 geographical domains, from large scale to small basin scale, was collected. A summary of the results is available in Supplementary material 3. The domains cover mainly Europe (whole continent, whole of Sweden, whole of England and several specific basins), but some applications were located outside of the continent, under different hydrological conditions (Arctic region, Niger basin, India and Cambodia). The model was applied for flood warning systems, evaluation of water quality, water resources and climate change assessment, observation monitoring or process understanding. It was mainly applied at a daily time step, but two modellers used it at an hourly time step. The modellers had up to 13 years' experience with large-scale HYPE model applications and simultaneous parameter estimation from large-sample datasets, using different regionalization methods.

The obtained data were summarized and compared to the numerical experiment results in order to understand their potential differences. Given the variety of model set-ups, the results were also used to check whether modellers were influenced by the model domain or purpose when choosing parameters. The results of this survey were used as a qualitative complement to the results of the sensitivity analysis.

2.4 Data and model runs

As stated in the introduction of this paper, sensitivity analyses are often limited to a small number of catchments. In this study, to be more generic, the sensitivity analysis method described in the previous section was applied to a larger set of gauged catchments extracted from the discharge datasets presented by Arheimer *et al.*

(2020). Open data with basic quality checks by Crochemore *et al.* (2020) was selected to have no influence from lakes or floodplains – the latter to avoid the influence of such hydrological features that can hide the influence of runoff generation parameters. The selection was made continent by continent to optimize the spread over the globe. In total, 20 stations were selected for each continent plus 20 more snowy catchments where possible. The stations were also selected to have at least nine years of data within a 10-year period that depend on the continent on which a given station is located (from 1970 to 1980 in Africa, from 1975 to 1985 for North America and for the snowy catchments, from 1980 to 1990 for Europe, Oceania and South America, and from 1994 to 2004 for Asia). These periods were used to optimize the number of stations by continent because data availability periods differ by geographic zone (Crochemore *et al.* 2020). Some of the data are only available at a monthly time step. Because of this, the E_{lim} score is calculated for all the catchments in the dataset on a monthly time step by aggregating the daily results of the HYPE model.

The catchment delineations for simulating the discharge at each of these stations were extracted from the World-Wide HYPE model version 1.3.3 (Arheimer *et al.* 2020). The input files describing catchment characteristics were modified so that only one uniform HRU covers each catchment (apart from water and glaciers which are treated separately) to achieve uniform catchment runoff response only impacted by one soil class and land cover class for each parameter. These catchments were selected to be small ($\approx 1000 \text{ km}^2$) and independent (having no upstream catchments) so that catchment representation is in some ways lumped for this experiment. This size choice was made because it reduces the computational time to carry out the sensitivity analysis and because the hydrological processes studied here are best represented at this unit in the model.

This selection resulted in a set of 120 catchments that are not affected by snow and another 67 catchments affected by snow (i.e. more than 10% of precipitation occurs when the temperature is below 0°C). They are spread as evenly as possible over the continents (see map Fig. 3).



Figure 3. Map of the 187 gauged catchments used for the experiment.

The selected catchments were evaluated with respect to their climate to check whether climate could be a driver explaining the variability in the sensitivity, as stated by Melsen and Guse (2019). To evaluate the different climates represented among these catchments, Table 3 shows the number of catchments that fall within each Köppen class (from Kottek *et al.* 2006). The table shows that the majority of the catchments are in temperate or continental climate areas, which also represents the most densely gauged regions globally. The Köppen classification uses five main groups of climates (A to E, see Table 3) and includes sub-groups according to the summer temperature, the date of the dry season and specific climatic features (e.g. monsoon in tropical group). To achieve more equal numbers in each class, only the main Köppen classes were considered for tropical, arid and polar climates as there are few catchments in these regions. For the continental and temperate climates an additional sub-division was made using Köppen sub-classes that consider temperature in summer.

For each gauging station, the model was run for the 10-year period of maximum data availability between 1970 and 2004 on the continent where the catchment is located. In addition, a 5-year warm-up period was used before these 10 years to balance the stores in the model. Arheimer *et al.* (2020) used 15 years of initialization, but mainly because of lakes, glaciers and sinks that are not modelled here. Subsequently, the monthly E_{lim} criterion was computed to compare simulated and observed data. The criterion was computed at a monthly time step because the discharge data from some of the stations is at a monthly time step. For this reason, this work is focused on inter- and intra-annual dynamics rather than on short-term hydrological events like floods.

The analysis was split between catchments with and without snow impact since snowmelt has a major influence on soil moisture and runoff, which runs the risk of obscuring all other flow-generating processes. Therefore, the sensitivity of the soil parameter was first evaluated in the no-snow catchments, and then the most influential soil parameters obtained from this analysis were included in the sensitivity analysis of the snow parameters for the 67 snow catchments. By doing this, there is a risk of neglecting a soil parameter that has greater influence on snowy catchments than on no-snow catchments. However, given that having a high number of parameters to evaluate may lead to a need to increase the number of runs, the choice was made to select only the sensitive soil parameters.

Table 3. Number of catchments in each Köppen climate class within the catchment set.

Main Köppen class (abbreviation)	Summer temperature (abbreviation)	Number of catchments
Tropical (A)	-	18
Arid (B)	-	14
Temperate (C)	Hot summer (Ca)	44
	Warm summer (Cb)	36
Continental (D)	Hot summer (Da)	12
	Warm summer (Db)	31
	Cold summer (Dc)	26
Polar (E)	-	6

3 Results

3.1 Parameter sensitivity on the whole catchment set

First, the analysis described in section 2.2 was carried out on the 120 no-snow catchments (red dots in Fig. 3). It gave a general idea of the soil parameter sensitivity. The aim was to reduce the number of parameters on which sensitivity is evaluated, avoiding snow influence.

The evaluation of the Monte Carlo runs was done catchment by catchment (Fig. 4 serves as an example, and the similar plot are available for the other catchments are available in Supplementary material 1). Overall, the median E_{lim} performances by catchments are between -0.25 and 0.25 . This may seem low, but knowing that the calibrated World Wide HYPE model has a median E_{KG} value of 0.4 across the globe (Arheimer *et al.* 2020) it is quite logical to obtain such values. The results show four parameters that seem to influence the value of E_{lim} the most. They are, in order of importance, *wcfc*, *mactrinf*, *rrcs2* and *mperc1*. For all catchments, the most influential parameter seems to be one of these four parameters. However, the parameter influences on E_{lim} vary depending on the catchment. These catchments can be broadly split into four groups. The first group includes catchments in which E_{lim} seems to be sensitive to all four of these parameters. This may happen when the response of the catchment is a balance between surface runoff and subsurface flow. In the second group, the four parameters have an influence but the influence of *wcfc* is very high. This can happen for catchments where there is too much rainfall to reproduce observed flow well. In the third group, E_{lim} is only sensitive to *wcfc*. In the fourth group, only *wcfc* and *rrcs2* show some influence. Catchments in these latter groups (the third and the fourth) are probably dominated by subsurface flow. There is a fifth group of catchments dominated by *mactrinf* and *wcfc*, and a sixth group dominated by *mactrinf* and *mperc1*. Catchments in these latter two groups are probably more dominated by surface runoff. This first analysis is a very simplistic way to understand the sensitivity but gives an overall idea.

In addition, this analysis is carried out over the whole range of each parameter based on a uniform distribution. It is also interesting to have a more precise idea of what the sensitivity is taking into account the distribution of parameters with good E_{lim} values. This is the role of the variance-based analysis. The initial distributions for each parameter obtained from the 5000 best parameter sets values are available in Supplementary material 2. These 5000 parameter sets cover a large part of the parameter space delimited by the initial ranges, but the distribution put greater emphasis on some areas of parameter space. It is also interesting to note that in some catchments, the distribution emphasized separate areas because of local optima created by nonlinearity in the model equations.

The variance-based sensitivity analysis provided two results. The first-order Saltelli index (denoted S) measures the first-order influence of each parameter (Fig. 5 left). The higher the S estimator, the higher the parameter sensitivity (Fig. 5). This shows that the most influential parameters remain the same for the second step of the sensitivity analysis. With this variance-based analysis, the E_{lim} criterion seems to be more sensitive to *mactrinf* than to *wcfc*. This is

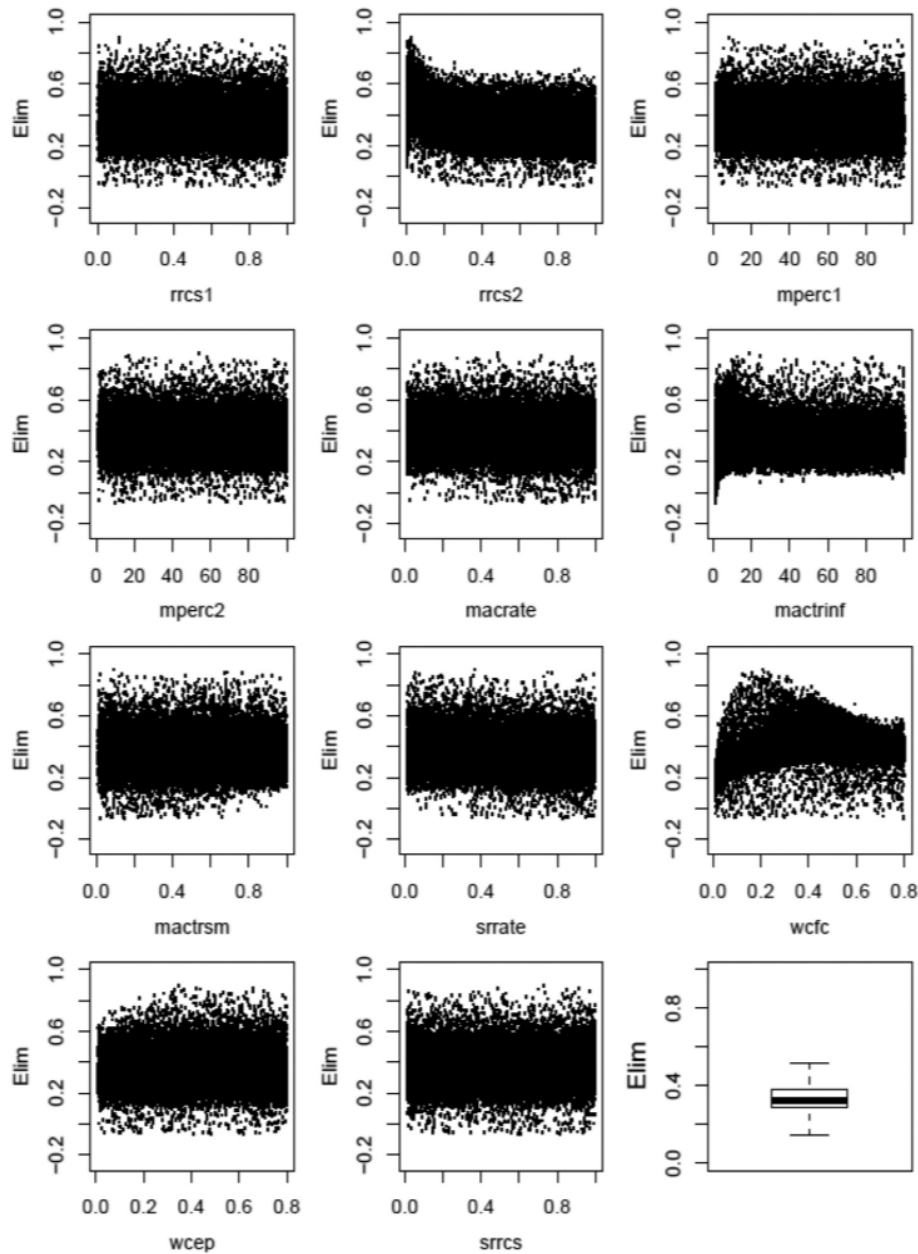


Figure 4. Response in terms of monthly E_{lim} value for each tested parameter value for the Acheron River catchment in Victoria (Australia).

mainly due to the fact that the prior distribution of $wcfc$ is more limited. This means that this parameter has less importance when distribution over the best values of E_{lim} is taken into account. Still, using an appropriate range for $wcfc$ is important, as shown by the Monte Carlo analysis. The Saltelli analysis also provides an opportunity to rank the parameters to which the E_{lim} is less sensitive. It seems that *macrate* and *srrate* have a small influence in some catchments, while for the others (*mactrsm*, *srrcs*, *wcep*, *rrcs1* and *mperc2*), the index values remain low.

The Saltelli total index (S_t) quantifies the sensitivity of a parameter considering its interactions with other parameters; it gives the total influence of each parameter on E_{lim} (Fig. 5, right). Its value then represents the total influence of the parameter. The distributions of S_t indices over the 120 catchments are similar to the distributions of

the S index. This indicates that the influence of a parameter in interaction with others is correlated to its influence by itself. However, it is clear that *macrate* and *srrate* now have some influence in certain catchments when interactions are considered.

As a result of these analyses on the no-snow catchments, it was determined that the parameters *rrcs1*, *mperc2*, *mactrsm*, *wcep* and *srrcs* do not show enough sensitivity to be considered for the snow catchments.

The snow parameters analysis is based on the same protocol. The soil parameters are also included in the analysis because they may impact some catchments' response. However, the parameters mentioned above that were least sensitive were not included. The analysis was carried out on the snow catchment set (shown in blue in Fig. 3).

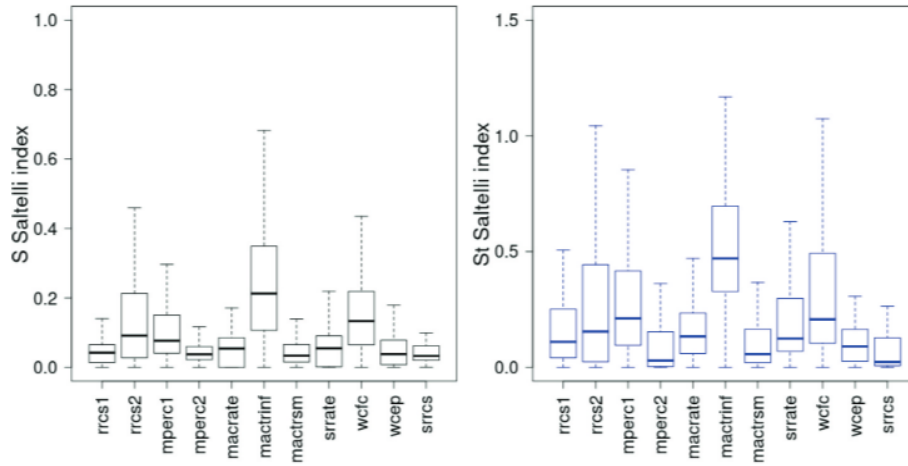


Figure 5. Distribution of the S (left) and St (right) sensitivity estimators over the 120 no-snow catchments for each parameter. A high index value means the parameter is sensitive on its own (left) or in interaction with other parameters (right).

The analysis of the obtained parameter distributions after the Monte Carlo runs on the snow catchments is more difficult, for two reasons. The first reason is that the performances are not very good for this set of catchments. The other reason is that there are more parameters that influence the E_{lim} value. It is thus more difficult to understand the interactions. The Monte Carlo-based sensitivity remains high for soil parameters *wfcf*, *mactrinf* and *rrcs2* for a significant number of catchments, and, at the same time, the sensitivity of snow parameters is high. The most influential seems to be *ttmp*, but *cmlt*, *cmrad* and *fepotsnow* also show clear patterns for the majority of catchments. All these added influences are difficult to interpret, and therefore the variance-based analysis is of critical importance.

The results of the variance-based sensitivity analysis tend to confirm the analysis done on the Monte Carlo runs. Looking at the first-order influence of each parameter (Fig. 6, left), the four dominant snow parameters are the same as stated above. As for no-snow catchments, the soil parameters are dominated by *wfcf* and *mactrinf*. For the less influential snow parameters, it seems that *fscdist0*, *fscdistmax* and *snalbmin* are more important than the others. These patterns remain the same when the interactions

between parameters are considered (Fig. 6, right). The S_t index also shows that the snow parameters have more influence than the soil parameters.

In addition to the sensitivity, the results of the Monte Carlo runs were used to limit the range of values for each parameter. For sensitive parameters, the plots that compare E_{lim} to parameter values (Fig. 4) were used to find values of parameters where there is no sensitivity. For example, in Fig. 4, E_{lim} is not sensitive to *mactrinf* above 30 mm. As a consequence, in this catchment, the value of *mactrinf* can be limited to a range between 1 and 30 mm. The same analysis can be extended to the other catchments in order to obtain ranges that are as general as possible. These ranges are compiled in Table 4 for the most sensitive parameters and can be valuable for a calibration to limit equifinality.

3.2 Link between sensitivity and process dominance

As stated in the introduction, the majority of the studies on sensitivity analysis are done for one or a couple of catchments. Since this work presents results on different catchments around the world, it provides a good opportunity to investigate

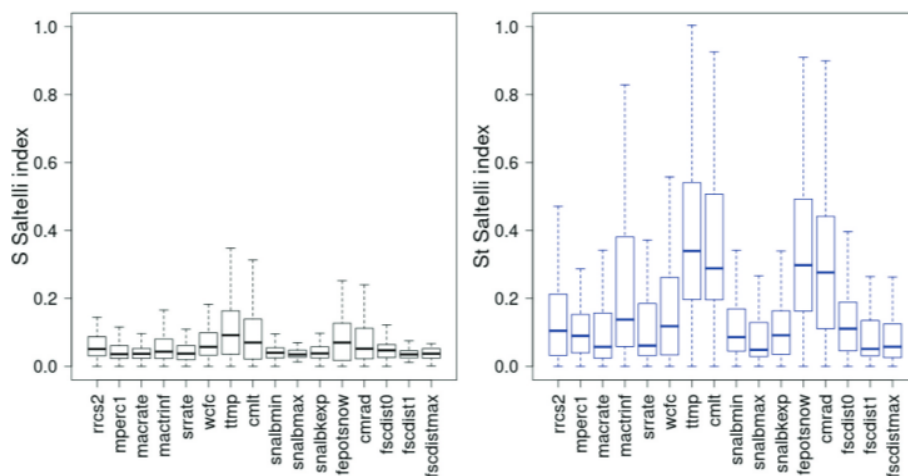


Figure 6. Distribution of the S (left) and St (right) sensitivity estimators over the 67 snow catchments for each parameter. A high index value means the parameter is sensitive on its own (left) or in interaction with other parameters (right).

Table 4. List of the parameters with the most influence on E_{lim} , and the range of values within which they have influence.

Soil parameter	Range	Snow parameter	Range
rrcs2	0.001 to 0.2	ttmp	−3 to 5
mperc1	1 to 25	cmlt	0.1 to 5
mactrinf	1 to 50	cmrad	0.01 to 0.7
macrate	0.01 to 1	snalbmin	0.1 to 0.6
srrate	0.01 to 1	fscdistmax	0.4 to 1
wcfc	0.01 to 0.8	fscdist0	0.01 to 1
		fepotsnow	0 to 1

how the location of the catchment can influence the sensitivity of parameters. To analyse this, the different parameters are grouped according to the process they regulate, in order to evaluate the most influential process in the catchments (Table 5). The sum of the Saltelli (S) index of the involved parameters in each process is then analysed to compare the influence of the different processes. Since the S index represents the influence of the parameter itself, the sum of all the S indices of the parameters that are involved in a given process is an indicator of the influence of the process in the catchment.

The map in Fig. 7 shows the most influential process for each catchment analysed. The map shows a complex spatial pattern of process influence. It is clear that snow processes are the most influential for the catchments where more than 10% of precipitation falls as snow. As shown in Fig. 6, the snow parameter has the greatest influence on the E_{lim} value in general. It is, then, expected that the snow process is dominant in these snow catchments. However, the fact that a lot of parameters are involved in the snow process while the soil processes are split can contribute to exaggerate this observation. For the catchments that are not influenced by snow, the map

Table 5. Parameters associated with the four evaluated runoff-generating processes.

Runoff-generating process	Parameters involved
Soil storage capacity	wcfc, wcep
Saturation excess runoff	rrcs1, rrcs2, srrcs, mperc1, mperc2
Infiltration excess runoff	mactrinf, macrate, srrate, mactrsm
Snow processes	ttmp, cmlt, fepotsnow, cmrad, snalbmin, snalbmax, snalbexp, fscdist0, fscdist1, fscdistmax

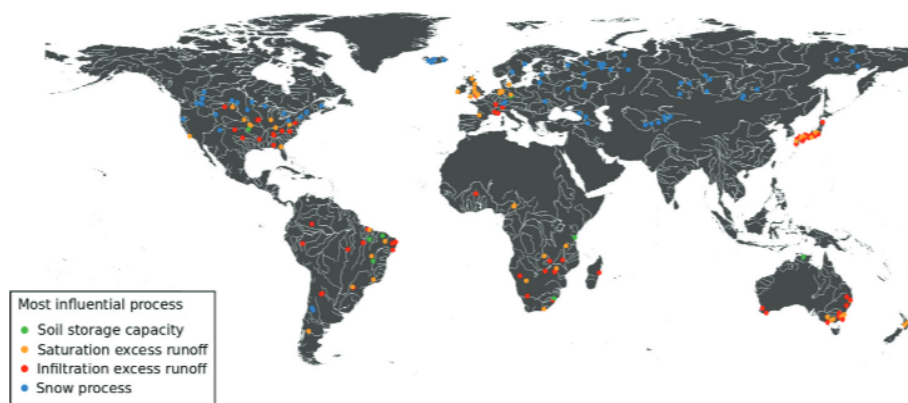
shows clearly that the process for which the E_{lim} criteria is the most sensitive differs for each catchment. Thus, with this map, it is difficult to draw conclusions regarding a geographical pattern of the process influences in each catchment. For example, the catchments where the runoff by saturation has more impact are not clearly distinct from the ones where the runoff by lack of infiltration is the major process (as illustrated e.g. in Japan). However, some patterns can still be observed. For instance, in Europe, the runoff by saturation looks more influential in the northern part and the runoff by lack of infiltration is the most influential in Mediterranean area. This corresponds with the fact that Mediterranean catchments are more likely affected by less regular rainfall and high-intensity events.

3.2.1 Effect of climate

Since climate is often the main cause for variability in runoff generation processes (Kuentz *et al.* 2017, Knoben *et al.* 2018), it is interesting to evaluate the impact of climate characteristics on the sensitivity of the parameters. For this, catchments were grouped by their main climate class (first column in Table 3). These main climate classes are tropical, arid, temperate, continental and polar. The box plots in Fig. 8 show the sum of all the S index values of the parameters involved in each process (denoted S_{sum}), and its distribution across catchments in each climate class. The box plots show some interesting tendencies. First, the spread of each distribution shows that the diversity within each class is high. However, some patterns still can be observed.

For the catchments in tropical areas there is no dominant process, but it is interesting to note that the soil capacity parameters seem to have more relative influence here than in any other climate class. The high amount of both rainfall and evapotranspiration in these areas can explain why the soil capacity has this high influence on the hydrological response. The model is built so that a large part of evapotranspiration is handled within the soil layers. By consequence, in these areas, the soil storage needs to be tuned correctly to be able to correctly store the large amount water that goes to the routing part of the model and that evaporates.

The catchments located in the arid areas are most influenced by the parameters that are linked to the runoff by lack of infiltration. These areas tend to be affected by occasional rainfall with high intensity that can create runoff even if the soil is

**Figure 7.** Map of the most influential process for each gauged station, obtained from the sensitivity analysis.

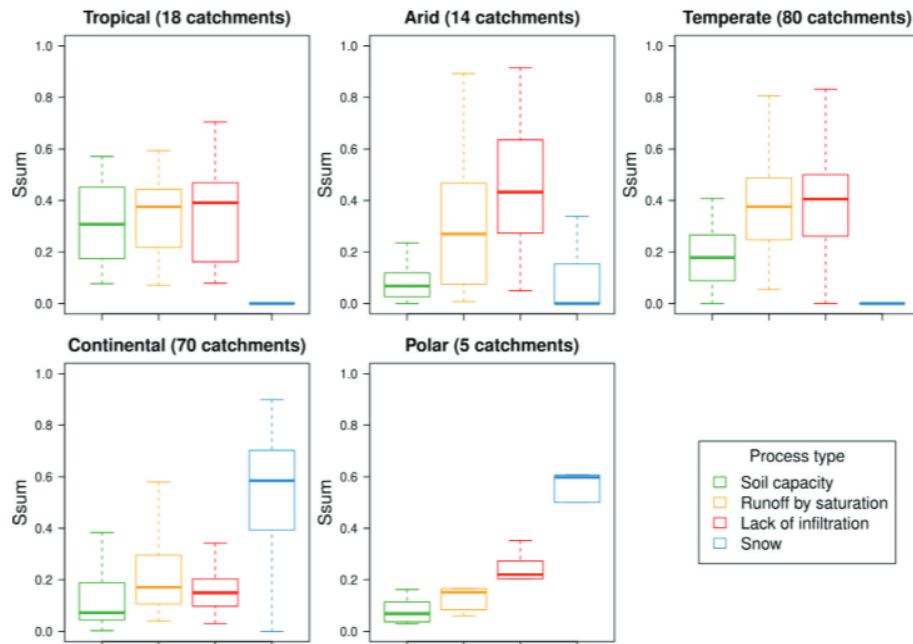


Figure 8. Distribution of process influence (in term of sum of sensitivity indices) for the catchments that belong to the Köppen main classes.

dry. To represent this process in the model, the water tends to bypass the soil layers to flow directly to the river, which explains the relatively low sensitivity of soil capacity parameters to E_{lim} here. The catchments in temperate areas seem to be dominated by the two types of runoff generation. In this climate zone, some catchments are mainly affected by stratiform rainfall while others are affected by less frequent high-intensity events, which thus results in different runoff generation processes. The flowpaths taken by water in the model are then more balanced, which explains the equivalent range of sensitivity index values taken by saturation and infiltration excess parameters.

As expected, the continental and polar regions are dominated by snow (see Fig. 7). It is, however, interesting to note that the response of the catchment in the polar region is more affected by parameters that drive runoff by infiltration, while the parameters that drive runoff by saturation seems more influential in continental areas. For polar regions, this can be linked to a low permeability of soil, for instance due to permafrost. Under polar climate, then, as for arid regions but not for the same reasons, the water also tends to bypass the soil layers in the model, which has an influence on the E_{lim} sensitivity to model parameters.

To summarize, Fig. 8 shows that the main climate classes seem to have an influence on the sensitivity to the different HYPE runoff parameters. It shows that water tends to use different paths in the model to correctly represent the hydrological specificity of each catchment class. These sensitivity patterns were easily linked to the known hydrological specificity in these climate regions, which shows that the process representation within the model has enough flexibility to represent hydrological processes under various climate conditions. However, considering the high spread within the different classes, these broad climate classes do not appear to be the only driver.

This analysis can be complemented by evaluating some more features of the Köppen classification. For instance, it also considers the summer climate for the continental and temperate classes. The influence of the summer climate on sensitivity is analysed by splitting the temperate and continental catchments depending on whether the summer is hot, warm or cold (see the second column of Table 3). A sub-group with very cold winters also exists for the continental class but is not observable in the catchment set. For each of these sub-groups of catchments, the distribution of the sensitivity of each process was evaluated as for Fig. 8. The result is plotted in Fig. 9. For temperate regions, the catchment set contains only catchments with hot and warm summers. The results show a clear difference of sensitivity between the two sub-groups. For catchments with hot summers, the runoff by lack of infiltration is the most influential process. By contrast, it is the runoff by saturation that is most influential for catchments with warm summers. For continental regions, the difference in terms of sensitivity among the three classes is not as high as for the temperate region (second row in Fig. 9) but the influence of snow is higher as the summer is colder, which is logical.

It is also interesting to note that runoff by lack of infiltration still has more influence in catchments with hot summers than in warm-summer continental catchments, and that its relative influence compared to runoff by saturation processes increases again for cold-summer catchments. These results highlight an interesting feature of how the sensitivity is changing within each Köppen class. For example, the sensitivity pattern of temperate catchments with hot summers is similar to the sensitivity pattern of arid catchments. The sensitivity pattern of continental catchments with cold summers is also close to the pattern observed for polar regions. This observation shows that the Köppen classification is probably not sufficient to describe the relationship between sensitivity and climate. As

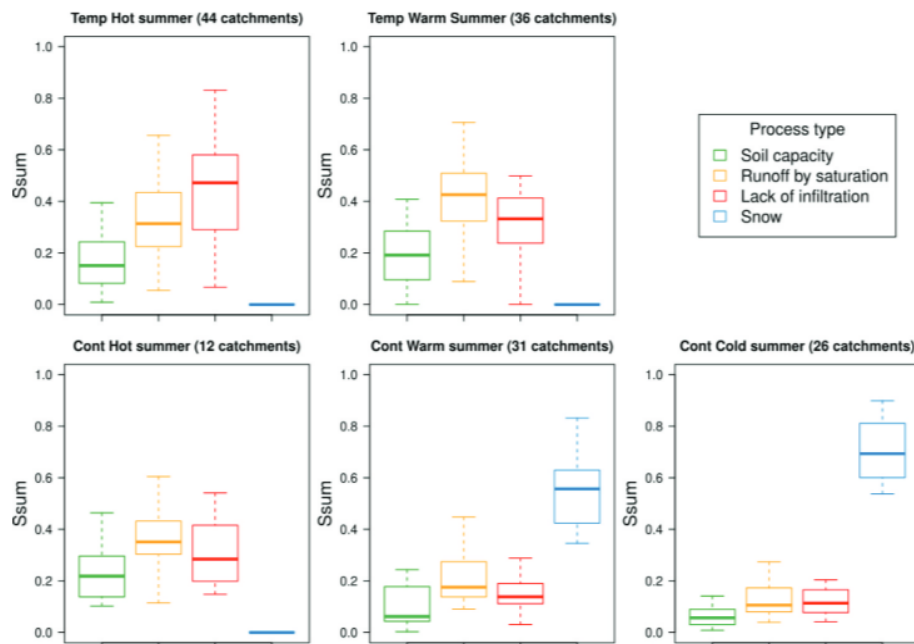


Figure 9. Distribution of process influence (in term of sensitivity indices) for the catchments with, hot, warm and cold (left to right) summers in temperate (top) and continental (bottom) climates.

stated by Knoben *et al.* (2018), climate is perhaps better described by continuous variables than by discrete classes, which is reflected in these results. However, the classification still allowed us to identify some interesting behaviour and a clear link between parameter sensitivity and climate.

In continental and temperate climate classes, the Köppen classification also gives an idea of whether and when the dry season occurs (never, in winter or in summer). It was also evaluated as for the summer climate, but the results did not show major differences in terms of parameter sensitivity between catchments with different dry periods.

To conclude, the analysis of the sensitivity results vs. Köppen classification showed that the climate condition of the modelled catchment is important to consider when using sensitivity analysis in the literature to select the parameter to calibrate. When transposing the result of a sensitivity analysis made on other catchment(s), it is important to check that the climate conditions of the modelled catchment are similar to the catchment(s) used to make the sensitivity analysis.

3.2.2 Impact of land cover

The relationship between sensitivity and land cover was evaluated because it is used in HYPE, and especially in World-Wide HYPE, to create the HRUs (Arheimer *et al.* 2020) in the catchments. Some HYPE parameters are, by definition, spatialized by land cover (*srrcs* and *snow* parameters). It is then interesting to check if the sensitivity is impacted by the land cover of catchments, to understand whether different parameters need to be calibrated for different land cover types. Figure 10 shows the distribution of dominant processes for catchments with four given dominant land cover types. A catchment is seen to be dominated by a land cover if it covers more than 50% of its area. No clear trend can be observed, but

some interesting patterns emerge. First, runoff by saturation seems to have more influence (compared to other processes) in crop- and broad-leaved-tree-dominated catchments. The catchments dominated by grass and needle trees seem, in comparison, more influenced by runoff due to lack of infiltration. However, it is difficult to give an interpretation to these results. Possibly, the fact that broad-leaved trees intercept more water than grass or needle trees can explain that the runoff by lack of infiltration has less influence in catchments dominated by these broad-leaved trees. However, interception by crops is on average not as high as for broad-leaved trees, and the soil can even be bare in some periods. This difficulty in interpreting the results may be due to the fact that the analysed land-cover classes are very generic and the internal variability within these classes may be high.

3.3 Comparison to expert judgement

In order to compare expert judgement to the sensitivity analysis, Fig. 11 shows how many of the surveyed experts used each soil parameter. This number is seen as a proxy for the influence of a parameter and is compared to the median Saltelli indices for each parameter (see Fig. 5) over the no-snow catchment set.

Figure 11 shows that expert judgement and the numerical sensitivity analysis have some results in common. For example, parameters *rrcs2*, *mperc1* and *wcfc*, which were shown by the sensitivity analysis to have high influence, are often used by the experts, while *mactrsm*, which is rarely used by the experts, has a low sensitivity value. However, the figure also shows discrepancies between the results of the study and the expert judgement. For example, the *mactrinfr* parameter is used by only half of the experts although it has the highest Saltelli indices values.

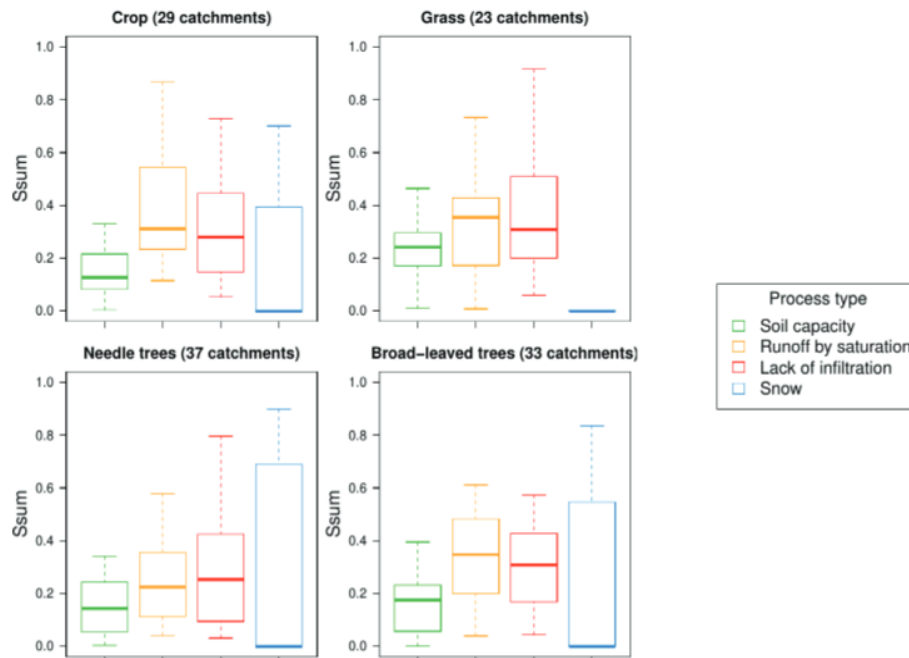


Figure 10. Distribution of process influence (in term of sensitivity indices) for the catchments with four different dominant land covers.

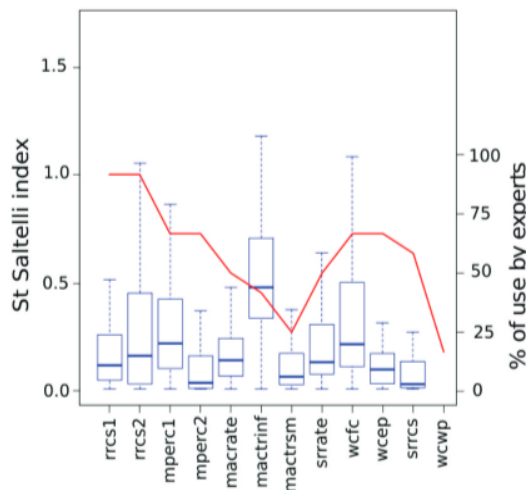


Figure 11. Comparison between St sensitivity estimators over the 120 no-snow catchments and the percentage of use of each parameter by the surveyed experts.

Conversely, parameters like *mperc2* and *rrcs1* are used by a majority of the experts although their Saltelli index values are relatively low.

These differences may occur for various reasons. The first reason is that the model domains in which the expert modellers applied HYPE are mainly located in Europe (especially northern Europe). Figure 7 shows that this area is more influenced by runoff by saturation-excess rather than infiltration-excess. It is then logical that the *mactrinf* parameter is less used than saturation excess parameters. Another reason may be the difference in set-up between the experiment presented in this paper and the model applications by the experts. Time step in particular may have an impact on the parameter choice. As the sensitivity analysis was performed at a monthly time step, parameter influence may differ

compared to a daily time step (which was mainly used by the surveyed experts). For example, the parameter *rrcs1*, which governs runoff from layer 1, represents a short-term process, which is probably more sensitive at high temporal resolution. It is then logical that this parameter was used by the experts, even though it has low Saltelli index values at a monthly scale. Last, the fact that some of the surveyed modellers are doing manual calibration may impact the result. A modeller with several years of experience is potentially able to deal with parameter equifinality while calibrating the model. Also, the data available to use in order to constrain some parameters may lead the modeller to choose parameters without taking sensitivity into account.

The results of the survey also confirm the fact that climate has an impact on parameter choice. The parameters linked to runoff by lack of infiltration are used in tropical and semi-arid locations (India, Niger basin) but not necessarily in northern Europe where this process has less importance. It is also interesting to note that time step has an impact (even if small) on parameter choice. The parameter *srrate* in particular was preferably chosen by experts who worked at an hourly time step. Conversely, the model purpose does not seem to influence the choice, as HYPE was designed to be an integrated tool for different applications. In fact, the majority of HYPE modellers who were part of this study applied the model for multiple purposes and in several contexts. For example, the model S-HYPE is used for both flood forecasting and water quality evaluation. Finally, it is important to note that the surveyed experts also used parameters that were not part of this study. For instance, all the modellers used parameters linked to evapotranspiration, which are not included in the sensitivity analysis.

It is more difficult to draw conclusions from the analysis about snow parameters, since few modellers used them. However, it is interesting to see that the most sensitive

parameters obtained from the numerical experiment (*ttmp* and *cmlt*) are also the most frequently used by the experts. The two other sensitive parameters (*fepotsnow* and *cmrad*) were only used by one modeller, and the others were used by either one expert or none. Also for snow parameters, some parameters other than the ones included in the numerical analysis were used by the experts.

4 Discussion

The use of the global model World-Wide HYPE provided the possibility to analyse parameter sensitivity on a catchment dataset with global coverage. This type of analysis is useful both to support the development of global-scale hydrological modelling and to evaluate how parameter sensitivity changes among geographical regions.

4.1 Support for global-scale modelling

As stated in the introduction, global applications of hydrological models induce some specificity, especially when it comes to parameter estimation (Bierkens *et al.* 2015). Depending on model characteristics, parameter estimation can vary. Some modellers assign pre-defined parameter values to physical characteristics (such as soil, land cover, etc.), for instance in PCR-GLOBWB (Sutanudjaja *et al.* 2018) or WaterWorld (Mulligan 2013), while other modellers estimate parameter values based on observed hydrological data (e.g. Zhang *et al.* 2016, Beck *et al.* 2017, Arheimer *et al.* 2020). However, since the necessary data are typically not available with high enough accuracy at this geographical scale, both approaches may present some difficulties. For example, the model H08 (Hanasaki *et al.* 2018) was not calibrated globally because of the lack of data; instead, the model must be calibrated when applied to a specific basin. The HYPE model was calibrated in a stepwise procedure when applied globally to limit equifinality, but due to the large biases remaining in certain locations, it is still recommended to re-calibrate it against local observations before using it in any local management applications (Arheimer *et al.* 2020).

All process-based models struggle with equifinality and a shortage of observed data for parameter estimations. For practical reasons, sensitivity analysis or expert judgement is required to reduce the number of parameters to estimate. This also reduces the amount of data needed for model calibration. Our results show that parameter sensitivity varies depending on geographical domain, which confirms the results obtained by e.g. Van Werkhoven *et al.* (2008), Hou *et al.* (2012) and Zajac *et al.* (2017). Hence, we recommend a careful selection of parameters by taking climate classes and catchment characteristics into account. To do this selection, a parameter sensitivity analysis on a large number of catchments well spread over the model domain can be performed. The sensitivity analysis methodology presented in this paper may not, however, be adapted to every global-scale model since computational time can be limiting. The Morris method, which is less demanding in terms of model runs, may be a good alternative for such models with high computational time. Also, if parameter selection is based on

expert judgement, the work of this paper shows that the specificity of the model may have an impact on the parameter choice.

The specificity of the experiment is a key factor to consider in the process of selecting parameters for calibration. The choice of the sensitivity estimation methodology may highly influence the results if it is not done with sufficient attention. For example, in the case of a variance-based analysis, the results may be influenced by the choice of the initial parameter distribution. To tackle this influence of experimental design, Sarrazin *et al.* (2016) proposed guidance to estimate the confidence of the obtained results. For this work we did not develop such a systematic approach, but we evaluated the impact of the size of the initial distribution. The results in this paper were compared to results obtained if the best 2500 or the best 10 000 simulations were used from the Monte Carlo sampling instead of 5000. This evaluation showed similar results over the whole catchment set (with local differences) regardless of the number of parameter sets extracted from Monte Carlo runs. The choice of this number of simulations may also influence the comparability between catchments since the distributions extracted from the 5000 best simulations differ among catchments.

The quality of the results can also be analysed by indicators of inconsistencies of the Saltelli indices. The indices cannot be higher than 1 or lower than 0, and the sum of first-order indices must be lower than zero. In this study, a significant number of catchments have shown some of these inconsistencies in the Saltelli indices (around 50 catchments show any of these types of issue), probably due to a too-low number of runs. For this reason, the meaningfulness of the results was checked by comparing them to the results obtained using Sobol' *et al.*'s (2007) calculation method (which requires a lower number of runs). The results obtained from the latter method were similar to the results presented in this paper, although the quality indicators were better. This confirms that even if some issues may appear for some catchments, the results of this paper remain relevant.

Also, in the case of this work, the fact that the analysis was made on a monthly time step is likely to influence the sensitivity. It is possible that the pattern of sensitivity would have been different at a daily time step as the most influential processes are different. For example, the parameter *rrcs1*, that drives fast subsurface flow, may have more influence at a daily time step as shown by the expert judgement results. It is also important to note that the criteria, based on the E_{KG} , have an impact on sensitivity (e.g. Van Werkhoven *et al.* 2008, Markstrom *et al.* 2016, Wang and Solomatine 2019). Thus, the results could have been different if the evaluation criterion was different, and the model was tuned to reduce only the model bias, for example. Accordingly, the evaluation criteria used to perform the sensitivity analysis should be selected with care in relation to the model goals. Moreover, the choice of sensitivity analysis method may influence the results.

Overall, it is important to note that the results are linked to the choice of experiment. The aim of this paper is not to recommend a particular method but rather to show that a sensitivity analysis adapted to the global scale is necessary

in order to efficiently select parameters for calibration. The link between expert judgements and sensitivity teaches us that, to some extent, the sensitivity analysis captures the parameters that are the most used in HYPE modelling. However, the discrepancies confirm that some parameters may be more adapted to specific goals and set-ups.

4.2 On the link between sensitivity and catchment characteristics

This work shows that parameter sensitivity is likely to change over different regions. It also shows that some patterns of sensitivity can be linked to climate, which confirms that climate should be taken into account for parameter regionalization (as shown by e.g. Merz *et al.* 2011). The study confirms, for HYPE and at the global scale, the outcomes found by Van Werkhoven *et al.* (2008), Melsen and Guse (2019) and Basijokaite and Kelleher (2021). The flowpaths taken by the water within the model vary for each climate class, even though some of the climates studied were only represented by a few catchments. The discharge data available at the global scale are more densely available for a small number of classes. For example, only six and 14 catchments were situated in polar and arid areas, respectively. Even if the sensitivity pattern is clear and logical for these classes, the conclusions from these results should not be exaggerated. Still, these low numbers in some classes do not erase the influence of climate zone on parameter sensitivity, but further evaluation may be needed to validate the results in the regions currently lacking data. It is also important to note that the choice of simulation period may influence the results, given that the climate has changed over the last decade. Indeed, because data availability periods differ by continent, different periods were used for each continent. However, these simulation periods are not so different (they all fall between 1970 and 1990 except for Asia), thus the impact of these different periods may be limited.

Probably, climate is not the only driver of parameter sensitivity. The box plot ranges show that not all the catchments within each Köppen class seem to have the same sensitivity. As HYPE is a process-based model, a link between sensitivity and process can be hypothesized, and if past studies stated that climate is a main driver for hydrological processes (Donnelly *et al.* 2016, Kuentz *et al.* 2017, Knoben *et al.* 2018), it is not the only one. Drivers such as catchment topography, geology or soil type may influence the sensitivity, as well as land cover (as shown in section 3.2.2). With this in mind, we argue that future studies on sensitivity should investigate the impact on sensitivity of additional catchment characteristics (e.g. slope, soil permeability) or of flow signatures (e.g. baseflow index, coefficient of variation, flashiness) to investigate whether such information is suitable for regionalization of parameter values, as suggested by Kuentz *et al.* (2017). The effect of climate can also be further refined by studying the sensitivity against a continuum-based climate classification (Knoben *et al.* 2018), since the link between sensitivity and Köppen classes appears to have a pattern that changes progressively. Interestingly, this observation was already made by Van Werkhoven *et al.* (2008), who identified a Spearman

correlation between Sobol' indices and aridity indices (one of the hydrological indices used by Knoben *et al.* 2018 for their classification).

Finally, the assumption that parameter sensitivity is linked to processes may be model dependent. As stated by Pfannerstill *et al.* (2015), the link between the two can be verified, but it will of course also largely depend on the model representation of each process. For instance, Melsen and Guse (2019) found a discrepancy between the parameter sensitivity of three models in drought representation. This shows the importance of reproducing such a large-scale sensitivity analysis while trying to apply a model at the global scale.

5 Conclusion

When setting up hydrological models, parameter estimation is of critical importance. As available data is not sufficient to estimate every parameter at the model resolution, it is necessary to set up a methodology to be able to select the most relevant parameters and to decrease data need. This paper uses such a methodology by applying a parameter sensitivity analysis on a large sample of catchments across the globe. The work shows that parameter sensitivity varies for different catchments in different regions. The driver that has the greatest influence on the sensitivity at the global scale appears to be climate, which confirms the need to take climate variation into account when using a sensitivity analysis (as advised by e.g. Wagener and Pianosi 2019, Razavi *et al.* 2021). The proposed analysis shows that parameter sensitivity and process dominance vary with Köppen class. As parameter sensitivity is linked to climate, it is important to take into account catchments with as varied climate as possible for sensitivity analysis. In addition, land cover also seems to have an impact on sensitivity, and we believe that more catchment characteristics (such as slope or dominant soil type) may have an impact. We hope that these observations will encourage the global modelling community to perform relevant sensitivity analysis, prior estimating model parameters.

The survey of HYPE expert modellers confirms that climate has an impact on parameter sensitivity but also emphasizes that sensitivity may depend on the model application. For instance, the selected time step seems to influence experts' choice of parameters to calibrate. Investigating this may require further work, using numerical sensitivity analysis at different time steps. The sensitivity is expected to change with the time step, as process influence increases with temporal resolution for certain processes (Massmann *et al.* 2014, Melsen *et al.* 2016).

Last, this paper also gives an overall idea of the parameter sensitivity of the HYPE model, which, to our knowledge, has never been evaluated in the literature. Even if the results of the analysis are linked to the choice of methodology (sensitivity method, objective function, time step etc.), it also shows more generally how parameter sensitivity links to processes in HYPE.

Acknowledgements

The authors thank the HYPE-expert modellers from SMHI for their advice and for taking part in the survey. This research benefited from several R&D projects at the Hydrological Research Unit, especially FANFAR (funded by

the European Union's Horizon 2020 research and innovation programme under grants agreement # 780118). The authors also thank Dr Georgy Ayzel who agreed to edit this paper and two anonymous reviewers for their valuable comments that helped to improve the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Union's Horizon 2020 research and innovation programme [grant agreement 780118].

ORCID

Léonard Santos  <http://orcid.org/0000-0003-2615-7279>

Jafet C. M. Andersson  <http://orcid.org/0000-0001-5269-7549>

Berit Arheimer  <http://orcid.org/0000-0001-8314-0735>

References

- Andersson, J.C.M., *et al.*, 2017. Process refinements improve a hydrological model concept applied to the Niger River basin. *Hydrol. Process*, 31, 4540–4554. doi:10.1002/hyp.11376
- Archfield, S.A., *et al.*, 2015. Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, 51 (12), 10078–10091. doi:10.1002/2015WR017498
- Arheimer, B., and Lindström, G., 2013. Implementing the EU Water Framework Directive in Sweden. In: A. Viglione, *et al.*, eds. *Runoff Prediction in Ungauged Basins: synthesis across Processes, Places and Scales*, 353–359. Cambridge: Cambridge University Press.
- Arheimer, B., *et al.*, 2020. Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, 24 (2), 535–559. doi:10.5194/hess-24-535-2020
- Basijokaite, R. and Kelleher, C., 2021. Time-varying sensitivity analysis reveals relationships between watershed climate and variations in annual parameter importance in regions with strong interannual variability. *Water Resources Research*, 57 (1). doi:10.1029/2020WR028544
- Beck, H.E., *et al.*, 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21 (12), 6201–6217. doi:10.5194/hess-21-6201-2017
- Beven, K., 2006. A manifesto for the equifinality thesis. *Model Parameter Estim. Exp.* 320, 18–36. doi:10.1016/j.jhydrol.2005.07.007
- Beven, K. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6 (3), 279–298. doi:10.1002/hyp.3360060305
- Bierkens, M.F.P., *et al.*, 2015. Hyper-resolution global hydrological modelling: what is next? *Hydrol. Process*, 29, 310–320. doi:10.1002/hyp.10391
- Blöschl, G., *et al.*, 2019. Changing climate both increases and decreases European river floods. *Nature*, 573 (7772), 108–111. doi:10.1038/s41586-019-1495-6
- Blöschl, G., *et al.*, Eds., 2013. *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*. Cambridge: Cambridge University Press.
- Chaney, N.W., *et al.*, 2015. Flood and drought hydrologic monitoring: the role of model parameter uncertainty. *Hydrology and Earth System Sciences*, 19 (7), 3239–3251. doi:10.5194/hess-19-3239-2015
- Crochemore, L., *et al.*, 2020. Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, 65 (5), 699–711. doi:10.1080/02626667.2019.1659509
- Donnelly, C., Andersson, J.C.M., and Arheimer, B., 2016. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal*, 61 (2), 255–273. doi:10.1080/02626667.2015.1027710
- Gupta, H.V., *et al.*, 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377 (1–2), 80–91. doi:10.1016/j.jhydrol.2009.08.003
- Gupta, H.V., *et al.*, 2014. Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18 (2), 463–477. doi:10.5194/hess-18-463-2014
- Guse, B., *et al.*, 2016. On characterizing the temporal dominance patterns of model parameters and processes. *Hydrological Processes*, 30 (13), 2255–2270. doi:10.1002/hyp.10764
- Hanasaki, N., *et al.*, 2018. A global hydrological simulation to specify the sources of water used by humans. *Hydrology and Earth System Sciences*, 22 (1), 789–817. doi:10.5194/hess-22-789-2018
- Hirpa, F.A., *et al.*, 2018. Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology*, 566, 595–606. doi:10.1016/j.jhydrol.2018.09.052
- Hou, Z., *et al.*, 2012. Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model. *Journal of Geophysical Research: Atmospheres*, 117 (D15). doi:10.1029/2012JD017521
- Hundecha, Y., *et al.*, 2016. A regional parameter estimation scheme for a pan-European multi-basin model. *Journal of Hydrology: Regional Studies*, 6, 90–111. doi:10.1016/j.ejrh.2016.04.002
- Knoben, W.J.M., Woods, R.A., and Freer, J.E., 2018. A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, 54 (7), 5088–5109. doi:10.1029/2018WR022913
- Kottek, M., *et al.*, 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15 (3), 259–263. doi:10.1127/0941-2948/2006/0130
- Kuentz, A., *et al.*, 2017. Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21 (6), 2863–2879. doi:10.5194/hess-21-2863-2017
- Lindström, G., *et al.*, 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, 41 (3–4), 295–319. doi:10.2166/nh.2010.007
- Markstrom, S.L., Hay, L.E., and Clark, M.P., 2016. Towards simplification of hydrologic modeling: identification of dominant processes. *Hydrology and Earth System Sciences*, 20 (11), 4655–4671. doi:10.5194/hess-20-4655-2016
- Massmann, C., Wagener, T., and Holzmann, H., 2014. A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales. *Environmental Modelling & Software*, 51, 190–194. doi:10.1016/j.envsoft.2013.09.033
- Mathevet, T., *et al.*, 2006. A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. *IAHS-AISH Publ.* 307, 211–219.
- Melsen, L.A. and Guse, B., 2019. Hydrological drought simulations: How climate and model structure control parameter sensitivity. *Water Resources Research*, 55 (12), 10527–10547. doi:10.1029/2019WR025230
- Melsen, L.A., *et al.*, 2016. HESS opinions: The need for process-based evaluation of large-domain hyper-resolution models. *Hydrology and Earth System Sciences*, 20 (3), 1069–1079. doi:10.5194/hess-20-1069-2016
- Merz, R., Parajka, J., and Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, 47, W02531. doi:10.1029/2010WR009505
- Muleta, M.K. and Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, 306 (1–4), 127–145. doi:10.1016/j.jhydrol.2004.09.005
- Mulligan, M., 2013. WaterWorld: a self-parameterising, physically based model for application in data-poor but problem-rich environments globally. *Hydrology Research*, 44 (5), 748–769. doi:10.2166/nh.2012.217

- Pechlivanidis, I.G., *et al.*, 2014. Uncertainty in the Swedish operational hydrological forecasting systems. In *Vulnerability, uncertainty, and risk*, 253–262.
- Pechlivanidis, I.G., *et al.*, 2020. What are the key drivers controlling the quality of seasonal streamflow forecasts?. *Water Resour Res*, 56 (e2019WR026987). doi:10.1029/2019WR026987
- Pfannerstill, M., *et al.*, 2015. Process verification of a hydrological model using a temporal parameter sensitivity analysis. *Hydrology and Earth System Sciences*, 19 (10), 4365–4376. doi:10.5194/hess-19-4365-2015
- Razavi, S. and Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in Earth and Environmental systems models. *Water Resour Res*, 51, 3070–3092. doi:10.1002/2014WR016527
- Razavi, S., *et al.*, 2021. The future of sensitivity analysis: an essential discipline for systems modeling and policy support. *Environmental Modelling and Software*, 137, 104954. doi:10.1016/j.envsoft.2020.104954
- Reinecke, R., *et al.*, 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. doi:10.5194/hess-23-4561-2019
- Roux, H., *et al.*, 2011. A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments. *Natural Hazards and Earth System Sciences*, 11 (9), 2567–2582. doi:10.5194/nhess-11-2567-2011
- Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145 (2), 280–297. doi:10.1016/S0010-4655(02)00280-1
- Saltelli, A. and Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25 (12), 1508–1517. doi:10.1016/j.envsoft.2010.04.012
- Sarrazin, F., Pianosi, F., and Wagener, T., 2016. Global Sensitivity Analysis of environmental models: convergence and validation. *Environmental Modelling & Software*, 79, 135–152. doi:10.1016/j.envsoft.2016.02.005
- Sobol', I.M., *et al.*, 2007. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety*, 92 (7), 957–960. doi:10.1016/j.ress.2006.07.001
- Song, X., *et al.*, 2015. Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, 523, 739–757. doi:10.1016/j.jhydrol.2015.02.013
- Sood, A. and Smakhtin, V., 2015. Global hydrological models: a review. *Hydrological Sciences Journal*, 60 (4), 549–565. doi:10.1080/02626667.2014.950580
- Stein, L., Pianosi, F., and Woods, R., 2020. Event-based classification for global study of river flood generating processes. *Hydrological Processes*, 34 (7), 1514–1529. doi:10.1002/hyp.13678
- Strömqvist, J., *et al.*, 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. *Hydrological Sciences Journal*, 57 (2), 229–247. doi:10.1080/02626667.2011.637497
- Sutanudjaja, E.H., *et al.*, 2018. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geosci Model Dev*, 11, 2429–2453. doi:10.5194/gmd-11-2429-2018
- Van Werkhoven, K., *et al.*, 2008. Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, 44 (1). doi:10.1029/2007WR006271
- Wagener, T. and 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. doi:10.1016/j.earscirev.2019.04.006
- Wang, A. and Solomatine, D.P., 2019. Practical experience of sensitivity analysis: comparing six methods, on three hydrological models, with three performance criteria. *Water*, 11. doi:10.3390/w11051062
- Wood, E.F., *et al.*, 2011. Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.* 47. 10.1029/2010WR010090
- Zajac, Z., *et al.*, 2017. The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, 548, 552–568. doi:10.1016/j.jhydrol.2017.03.022
- Zhang, Y., *et al.*, 2016. Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. *Journal of Hydrometeorology*, 17 (3), 995–1010. doi:10.1175/JHM-D-15-0107.1