



Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*. DOI: [10.1002/hyp.10968](https://doi.org/10.1002/hyp.10968)

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1002/hyp.10968](https://doi.org/10.1002/hyp.10968)

[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 <http://onlinelibrary.wiley.com/doi/10.1002/hyp.10968/abstract>. 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/pure/about/ebr-terms.html>

Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis

Francesca Pianosi* and Thorsten Wagener

Department of Civil Engineering, University of Bristol, Bristol, UK

Abstract:

Simulations from hydrological models are affected by potentially large uncertainties stemming from various sources, including model parameters and observational uncertainty in the input/output data. Understanding the relative importance of such sources of uncertainty is essential to support model calibration, validation and diagnostic evaluation and to prioritize efforts for uncertainty reduction. It can also support the identification of ‘disinformative data’ whose values are the consequence of measurement errors or inadequate observations. Sensitivity analysis (SA) provides the theoretical framework and the numerical tools to quantify the relative contribution of different sources of uncertainty to the variability of the model outputs. In traditional applications of global SA (GSA), model outputs are aggregations of the full set of a simulated variable. For example, many GSA applications use a performance metric (e.g. the root mean squared error) as model output that aggregates the distances of a simulated time series to available observations. This aggregation of propagated uncertainties prior to GSA may lead to a significant loss of information and may cover up local behaviour that could be of great interest. Time-varying sensitivity analysis (TVSA), where the aggregation and SA are repeated at different time steps, is a viable option to reduce this loss of information. In this work, we use TVSA to address two questions: (1) Can we distinguish between the relative importance of parameter uncertainty *versus* data uncertainty in time? (2) Do these influences change in catchments with different characteristics? To our knowledge, the results present one of the first quantitative investigations on the relative importance of parameter and data uncertainty across time. We find that the approach is capable of separating influential periods across data and parameter uncertainties, while also highlighting significant differences between the catchments analysed. Copyright © 2016 The Authors. *Hydrological Processes*. Published by John Wiley & Sons Ltd.

KEY WORDS conceptual models; parameter uncertainty; data uncertainty; sensitivity analysis

Received 31 October 2015; Accepted 17 July 2016

PREFACE

One of Keith Beven’s major contributions lies in the field of uncertainty analysis. Following pioneering works by Freeze (1980), he was one of the first to introduce Monte Carlo strategies for uncertainty assessment in hydrological models (Binley *et al.*, 1989; Beven and Binley, 1992), and he has pushed the topic of uncertainty analysis for almost a quarter century by now. His efforts are a main reason why uncertainty analysis in hydrology is more advanced than in most other environmental or natural hazard fields (Rougier *et al.*, 2013). We bring two of Keith’s major contributions together in the research discussed here. First, we use a time-varying implementation of sensitivity analysis that can be traced back to Keith’s work on generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992; Freer

et al., 1996). GLUE brought an easy to implement and effective approach to analyse parameter uncertainty of hydrological models. It also brought us a discussion of the validity of statistical assumptions given the specific nature of hydrological models (including strong nonlinearity and potentially large model structural errors). This discussion also opened the path for an investigation into what metrics are hydrologically relevant, rather than just statistically convenient. These questions, e.g. regarding the appropriate likelihood function based on these statistical assumptions in the context of such models, have still not been answered (e.g. Stedinger *et al.*, 2008). These discussions are unlikely to go away unless we understand how to build likelihood functions that realistically account for all sources of error. Second, we consider Keith’s more recent efforts focused on the implications of ‘disinformative data’, i.e. data points that are erroneous and negatively influence the model calibration or evaluation process. A key question is how we identify data points whose values are the consequence of measurement errors or inadequate observations (such

*Correspondence to: Francesca Pianosi, Department of Civil Engineering, University of Bristol, Bristol, UK.
E-mail: francesca.pianosi@bristol.ac.uk

as a convective rainstorm largely falling between rain gauges). Keith and colleagues pointed out that formal methods to identify such data periods are surprisingly poorly established given the importance of data in our field of study (e.g. Beven and Westerberg, 2011). Here, we try to combine both of these efforts into a single formalized approach to understand the relative importance of data and model parameter uncertainty.

INTRODUCTION

Sensitivity and uncertainty analysis have become common practise in hydrological modelling. One strand of methods is based on Monte Carlo sampling of the parameter space and on conditioning these parameter sets by using one or more objective functions. This approach originates in the regional sensitivity analysis (RSA) method introduced by Young *et al.* (1978) and Spear and Hornberger (1980) in which these authors separate parameter sets into behavioural (well performing) and non-behavioural groups. Parameters matter for the model output if their marginal distributions across these two groups are different. Spear and Hornberger (1980) tested their approach on an algae bloom problem in a lake for which the sampled parameter sets either produced algae bloom (bad) or not (good). This binary separation made the separation into behavioural and non-behavioural groups simple and unambiguous. Beven and Binley (1992) generalized RSA by showing how the parameter sets can be conditioned on any performance metric if appropriately transformed (so that it has some, but not all, of the characteristics of a likelihood function). A parameter set is deemed behavioural if the associated value of the performance metric is above a modeller prescribed threshold (or below, if the performance metric is to be minimized). They found that in most cases, a wide range of behavioural parameter sets can be found for hydrological models. Keith Beven and colleagues referred to this finding as the problem of equifinality. Since then, RSA based on conditioning of performance metrics has been widely applied to investigate parameter uncertainty and their relative influence in hydrological and environmental modelling (e.g. Freer *et al.*, 1996; Wagener *et al.*, 2001; Sieber and Uhlenbrook, 2005).

One issue with using conditioning on performance metrics in RSA is that performance metrics are based on aggregation of model residuals across the whole time series used for calibration. This temporal aggregation process unavoidably leads to a loss of information (e.g. Freer *et al.*, 2003). Sometimes, few but very large residuals will dominate the value of the performance metric, especially if the residuals are squared before aggregation (as carried out for instance when using the root mean squared error or the Nash–Sutcliffe efficiency). Wagener *et al.* (2003)

suggested to apply a variation of the parameter conditioning used in GLUE as a time-varying algorithm in their dynamic identifiability approach (DYNIA). They estimate the performance metric value as a running mean using different window sizes. This approach reduces the loss of information and allows for an assessment of which periods are most informative for parameter calibration or regarding which data points might be erroneous. They visualize the conditional marginal cumulative distribution function (CDF) for each parameter so that they can both separate periods where conditioning takes place or not (i.e. where data is informative or not) and so that they can see which part of the parameter space is better performing (e.g. to test whether different parameter values are required for different system response modes as an indicator of model structural problems such as missing model dynamics). The chosen window size allows for tailoring across influence scales of parameters (Massmann *et al.*, 2014), i.e. parameters controlling the quick recession process require shorter window sizes than those controlling baseflow or water balance processes. There has been a flurry of studies by using different analysis methods for time-varying SA (e.g. Wagner and Harvey, 1997; Wagener *et al.*, 2003; Cloke *et al.*, 2008; Reusser and Zehe, 2011; Kelleher *et al.*, 2013; Herman *et al.*, 2013a, b; Guse *et al.*, 2014).

While the assessment and consideration of parameter uncertainty have become common practise, there is also the increasing recognition that data uncertainty can have a significant influence on model calibration and validation. Keith Beven and colleagues referred to this problem as that of disinformative data (Beven *et al.*, 2008; Beven and Westerberg, 2011). Specific data points, when erroneous, can have a disproportionate impact on model calibration or evaluation, and approaches to identify them are rather poorly developed in hydrology. Initial work for example demonstrates that event-based water balance estimates can show much larger runoff volumes than feasible given rainfall magnitudes in the streamflow record (Beven and Smith, 2014). Such unrealistic data points can lead to large residuals, which, in turn, strongly influence performance values. It is therefore important to be able to identify these data points and appropriately decide on their validity.

In this paper, we address two questions: (1) Can we distinguish between the relative importance of parameter uncertainty *versus* data uncertainty in time? (2) How do relative influences change in places with different catchment characteristics? To answer these questions, we introduce a time-varying implementation of a recently proposed density-based SA approach called PAWN (Pianosi and Wagener, 2015). As a hydrological model, we use the widely applied lumped Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergstrom, 1995). We repeat our analysis for three catchments in the USA that span different hydroclimatic regimes and geographic

locations. We assess the relative importance of data and parameter error/uncertainty in time and investigate how far SA allows for a formalized approach to identify periods where data uncertainty could have a disproportionally large influence. Because we consider parameters and data as sources of uncertainty, but not model structure, our approach to identify potential disinformation is conditional on the hydrological model being adopted (the HBV model in our case). As such, the approach is complimentary to other methods (e.g. Beven and Smith, 2014) where a more general model is used, which however only captures part of the runoff generation process (for example, the event runoff coefficient).

Our results present one of the first quantitative investigations of the relative importance of parameter and data uncertainty in time. Understanding this variability is relevant for investigations into additional data collection needs and model calibration/evaluation.

MODELS AND METHODS

Hydrological model and study sites

The hydrological model investigated in this study is the lumped HBV conceptual model. It includes three components: a snow accumulation/melting module, a soil moisture accounting module and a flow routing module. The forcing input data are time series of temperature, precipitation and potential evapotranspiration. The model is described in various articles, e.g. Bergstrom (1995), Seibert (1997) and Kollat *et al.* (2012). A schematic is given in Figure 1.

At each time step, the model classifies precipitation as either rainfall or snowfall depending on whether temperature is above or below a given threshold (TS). Snowfall and rainfall contribute to the water balance of the solid and liquid component of the snowpack respectively. Exchanges between the two components are allowed through either snowmelt or refreezing, depending on the temperature being above or below the threshold TS . The amount of snowmelt or refreezing is linearly proportional to temperature via two proportionality coefficients $CFMAX$ and CFR (see again schematic in Figure 1). When the liquid component exceeds the snow pack holding capacity (CWH), the excess water leaves the snowpack and inputs the soil moisture accounting module. The implementation of the soil moisture accounting and flow routing modules are the same as in Kollat *et al.* (2012), which includes three parameters (β , LP and FC) for the soil moisture accounting and six parameters ($PERC$, $K0$, $K1$, $K2$, UZL , $MAXBAS$) for the flow routing. The meaning, units of measurements and range of variation of these parameters are summarized in Table I.

The model is applied to simulate streamflow in three catchments in the USA with very different climatic characteristics: the English River in Iowa, a relatively humid, snow-affected catchment; the French Broad River in North Carolina, a very wet catchment; and the Guadalupe River in Texas, a very dry catchment. Time series of daily streamflows and meteorological inputs (precipitation, temperature and potential evapotranspiration) for these catchments were developed as part of the Model Parameter Estimation Experiment (Duan *et al.*, 2006). The characteristics of the three catchments are summarized in Table II.

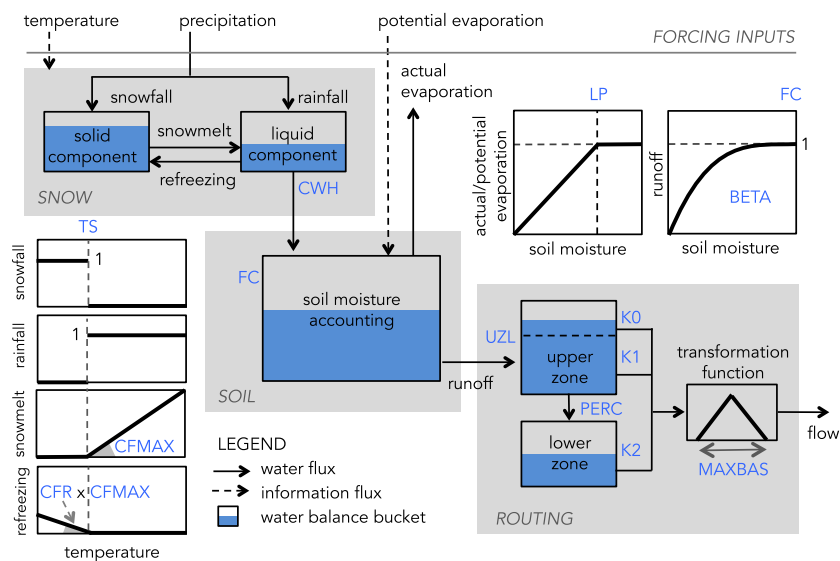


Figure 1. Schematic of the conceptual hydrological model used in this study. Model parameters are highlighted in blue. Their meaning is further explained in Table I

Table I. Parameters of the hydrological model and associated uncertainty ranges for the English, French and Guadalupe catchments.

Name	Description (unit)	Range					
		English		French		Guadalupe	
Snow parameters							
<i>TS</i>	Threshold temperature (C)	-3	3	—	—	—	—
<i>CFMAX</i>	Degree day factor (mm/C)	1	20	—	—	—	—
<i>CFR</i>	Refreezing factor (-)	0	1	—	—	—	—
<i>CWH</i>	Water holding capacity of snow (-)	0	0.8	—	—	—	—
Soil parameters							
<i>BETA</i>	Exponential parameter (-)	1	7	0.4	2.4	1.5	6
<i>LP</i>	Evapotranspiration limit (-)	0.3	1	0.3	1	0.3	1
<i>FC</i>	Field capacity (mm)	50	600	125	500	100	1000
Routing parameters							
<i>PERC</i>	Maximum flux from UZ to LZ (mm/day)	0	100	10	100	0	100
<i>K0</i>	Near surface flow coefficient (day ⁻¹)	0.05	1	0.15	1	0.05	1
<i>K1</i>	UZ outflow coefficient (day ⁻¹)	0.1	1	0.1	0.5	0.05	1
<i>K2</i>	LZ outflow coefficient (day ⁻¹)	0.05	0.1	0.05	0.1	0.05	0.1
<i>UZL</i>	Near surface flow threshold (mm)	0	100	28	90	0	100
<i>MAXBAS</i>	Flow routing coefficient (day)	1	4	1	4	1	4

UZ, upper zone; LZ, lower zone.

Table II. Catchment characteristics.

River	Outlet location	Area (km ²)	Length (km)	P (mm/year)	ROC (-)	PE (mm/year)	WI (-)
English	Kalona, IA	1484	120	868	0.28	994	0.87
French Broad	Ashville, NC	2448	350	1509	0.50	819	1.84
Guadalupe	Spring Branch, TX	3406	370	775	0.14	1528	0.50

P, mean annual precipitation; Q, mean annual streamflow; ROC, runoff coefficient (=Q/P); PE, mean annual potential evaporation; WI, wetness index (=P/PE).

Catchment areas are taken from van Werkhoven *et al.* (2008). P, ROC, PE and WI are computed using daily data for the period from 1 October 1948 to 30 September 2003 from the MOPEX dataset.

Characterization of the uncertainty sources

The goal of this study is to assess the relative importance of parameter and data uncertainty for model accuracy. We group sources of uncertainty into six groups: (1) the observational uncertainty in the precipitation time series, (2) the uncertainty in the potential evapotranspiration time series, (3) the uncertainty in the four parameters of the snow accumulation/melting model component, (4) the uncertainty in the three parameters of the soil moisture accounting component, (5) the uncertainty in the six parameters of the flow routing component, and (6) the uncertainty in streamflow observations used to evaluate model performance. These sources of uncertainty are characterized as follows.

Parameter uncertainty is described by assuming independent uniform distributions with the ranges reported in Table I. These ranges were defined by combining a priori knowledge about the physical meaning of each parameter and a preliminary evaluation of the model's behaviour in each study area. Kollat *et al.* (2012) provide a set of wide parameter ranges that should cover

catchments with any hydroclimatic characteristics across the USA. Sampling from those ranges, we ran Monte Carlo simulations and identified behavioural parameterizations by defining a set of thresholds on model performances (we considered root mean squared error, absolute mean error and bias). The range of variation of the behavioural parameterizations was then taken as uncertainty range for each study site. These ranges are reported in Table I. As it can be noted from the table, some of these ranges are still quite large, and for some parameters (for example, *LP* and *K2*), they are the same as in Kollat *et al.* (2012), which means that the conditioning on performances did not constraint them. This is a consequence of our choice of quite loose performance thresholds, which, in turn, reflects the fact that our aim is to define ranges that reasonably reflect our uncertainty in model parameters, rather than to identify a small set of highly performing parameterizations.

As for data uncertainty, a range of studies has assessed how much uncertainty can be expected in certain measurements of hydrological and meteorological vari-

ables. They vary with location, with instrument, with time period of measurement, etc. A nice summary is presented by McMillan *et al.* (2012). Here, we assume what we believe are typical ranges for the circumstances present in our case studies.

Precipitation uncertainty was described using storm-dependent rainfall depth multipliers as proposed by Kavetski *et al.* (2002, 2006), which corresponds to the assumption that precipitation errors are multiplicative and that the magnitude of the multiplicative error varies from storm to storm. The results discussed in the succeeding texts are obtained using storm-dependent multipliers drawn from a uniform distribution over the interval [0.6, 1.4], which corresponds to assuming a maximum error in precipitation data of $\pm 40\%$.

Multiplicative errors are also used for potential evapotranspiration; however, here a constant multiplier is used for the entire time series. We assumed a uniform distribution over [0.8, 1.2] for this multiplier, thus allowing for a maximum error of $\pm 20\%$.

Finally, for flow data, we used an additive error model where errors are described by an autocorrelated heteroschedastic Gaussian process, with zero mean and variance linearly proportional to the flow (Schoups and Vrugt, 2010). The two parameters of this model are set to ensure that 99% of the errors on flow fall within the interval $\pm 0.2q_t^{\text{obs}}$, i.e. a maximum error in flow observations of $\pm 20\%$. More details about this model and the procedure to set its parameters are given in the Appendix.

Definition of the performance metric

In our study, the performance metric used to synthetically measure the model accuracy is the root mean squared error (RMSE). Because our goal is to investigate how relative influences vary in time, we compute one value of the RMSE for each time step along the simulation period, by using a moving window centred around that time step, i.e.

$$y = \sqrt{\frac{1}{2w+1} \sum_{k=t-w}^{t+w} (q_k^{\text{sim}} - q_k^{\text{obs}})^2} \quad (1)$$

where q_k^{sim} is the simulated flow on day k , q_k^{obs} is the observed flow, t is the time step under analysis, and w is the semi-length of the moving window.

Global sensitivity analysis method: PAWN

In global SA (GSA), each source of uncertainty (or *input factor* in the GSA terminology) is associated with a sensitivity index that measures the relative influence of that factor on the model performance. Here, sensitivity indices will be computed according to a density-based

method called PAWN (Pianosi and Wagener, 2015). For each factor, say the i th, the PAWN sensitivity index is defined as

$$S_i = \max_{x_i} \max_y |F_y(y) - F_{y|x_i}(y|x_i)| \quad (2)$$

where F_y is the unconditional distribution of the performance metric y , i.e. the one induced by variations of all the factors, and $F_{y|x_i}$ is the conditional distribution of y , i.e. the one induced by variations of all factors but the i th, which is fixed to a nominal value x_i .

The rationale of Equation 2 is the following. If the unconditional and conditional distributions are very similar, it means that variations in the i th factor do not significantly affect the variability of y , and therefore, that factor has little influence. Conversely, the larger the difference between the two distributions, the more influential the input factor. This is captured by the inner maximum in Equation 2, which provides a measure of the distance between the two CDFs. The outer maximum in Equation 2 instead is used to remove the variability in the results that might arise from different choices of the nominal value x_i . By taking the maximum with respect to x_i , we ensure that the sensitivity index of Equation 2 is zero only if the i th factor has no influence at any point in its space of variability.

In the operational implementation of the method, the outer maximum in Equation 2, i.e. the one with respect to the conditioning value of x_i , is approximated by the sample mean over a prescribed number of conditioning values (e.g. 10). For each of these, the inner maximum, i.e. the maximum absolute difference between CDFs, is approximated by using empirical distribution functions. These are obtained by evaluating the model against input samples where all input factors vary (unconditional distribution) and against samples where the i th input is fixed to the conditioning value and the others vary (conditional distribution). The PAWN method is implemented in the Sensitivity Analysis for Everybody (SAFE) Toolbox (Pianosi *et al.*, 2015), which is freely available for academic use.

Density-based methods have a number of advantages. In the first place, they can be applied to any type of input factor, including time series of model forcing inputs or output observations, as in our study. This is not possible for other GSA methods. For instance, RSA compares probability distributions of the input factors in the behavioural and non-behavioural group. Thus, it presumes that each input factor x_i is a scalar variable that can be sensibly appointed a CDF, which is not the case when an input factor is a group of parameters or a time series, as in our study. In PAWN instead, sensitivity indices are defined based on the values of y only, as shown in Equation 2. Therefore, they can be computed regardless

of the mathematical properties and meaning of the input factors. Variance-based SA methods (Saltelli *et al.*, 2007) also possess this property. However, variance-based methods are not suitable when the output distribution is highly skewed or multimodal and variance would be a poor measure of uncertainty (Borgonovo, 2007, 2014). Density-based methods, instead, are applicable also in those situations because they assess changes in the entire distribution of the output y , rather than in one of these moments only. Furthermore, because the PAWN index is defined on CDFs that are efficiently approximated by empirical distribution functions, its application requires a relatively limited number of model evaluations (Pianosi and Wagener, 2015).

Finally, another advantage of PAWN is that sensitivity indices can be easily tailored to focus on a subregion of output values of particular interest, for instance, below a prescribed threshold \bar{y} . This is achieved by simply adjusting Equation 3 as follows:

$$S_i = \max_{x_i} \max_{y \leq \bar{y}} |F_y(y) - F_{y|x_i}(y|x_i)| \quad (3)$$

In our context where the output y is a performance metric (to be minimized), the threshold \bar{y} would represent a minimum level of performance, and by using Equation 4 means that only model evaluations that achieve that minimum performance contribute to the sensitivity indices. Using the RSA/GLUE terminology, we might call these model evaluations behavioural. However, there is a subtle difference with respect to RSA. In RSA, both behavioural and non-behavioural samples contribute to

determine sensitivity. Indeed, it is the very separation between the two groups that is used to measure sensitivity. In Equation 4, instead, the separation is only used to filter out non-behavioural samples. The rationale is that if we set the threshold to a reasonably loose value, Equation 4 will ensure that any model evaluation with unreasonably large deviations from observations do not bias the SA results.

Experimental set-up

In our application, we use 3000 random samples to build the unconditional CDFs, and 1000 random samples to build each conditional CDF. For each of the six input factors, conditional CDFs are computed at 20 conditioning values. The total number of model evaluations is therefore equal to $3000 + 1000 \times 20 \times 6 = 123\ 000$. For each model evaluation, the procedure to generate and propagate the six sources of uncertainty throughout the model is as follows (Figure 2). (1) Generate a time series of perturbed precipitation by multiplying the original time series by a randomly sampled multiplier (one per each storm). (2) Generate a time series of perturbed potential evaporation by multiplying the original time series by a randomly sampled multiplier (same value for the entire time series). (3) Generate three vectors of parameters for the snow, soil and routing component by uniform random sampling from the ranges in Table I. (4) Run the hydrologic model and obtain a time series of simulated flows. (5) Generate a time series of perturbed flows by adding a randomly generated sequence of autocorrelated heteroscedastic Gaussian errors to the original time series (details about the flow error model in Appendix). (6)

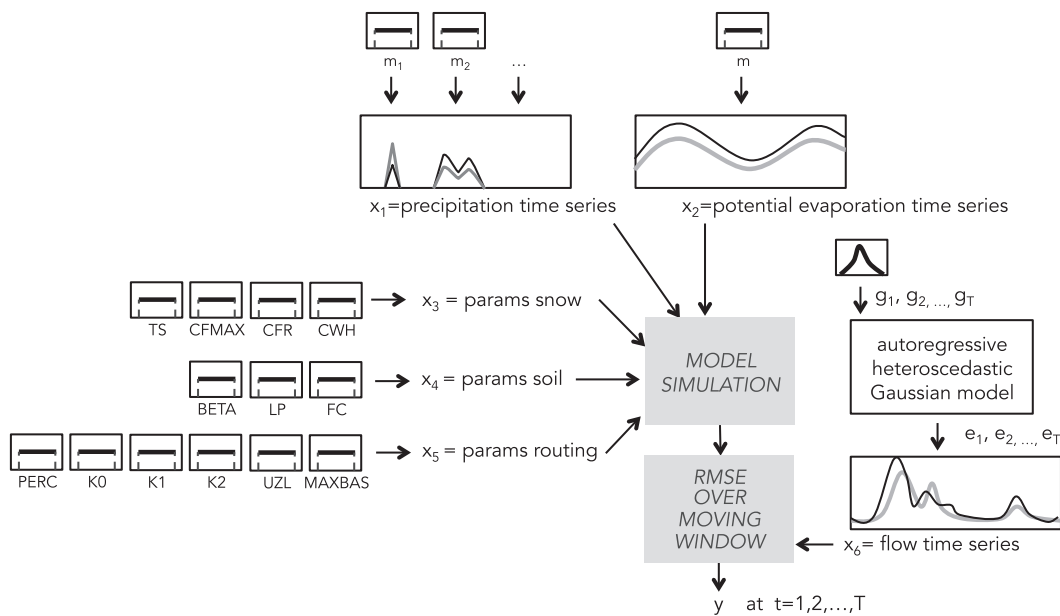


Figure 2. Schematic of the sampling and model evaluation procedure

Apply Equation 1 to the simulated and (perturbed) flow observations and obtain a time series of model performances y at $t=1, \dots, T$. The simulation period is from 1 October 1948 to 30 September 1951.

In our application, we set $w=15$ (days) so that the window size is 31 days. As a threshold value \bar{y} , we use twice the mean of observed flows over the moving window, i.e. $\bar{y} = (q_{t-w}^{obs} + \dots + q_{t+w}^{obs}) / (2w + 1)$. In other words, at each time step, we discard those model evaluations where deviations from flow observations are on average higher than twice the mean flow over that window. As anticipated, this threshold value is quite loose, and it is only meant to avoid that the analysis be biased from few samples corresponding to very bad model performance. All computations were performed using the SAFE Toolbox (Pianosi *et al.*, 2015).

RESULTS

Figure 3 shows the sensitivity indices of the six sources of uncertainty for the three catchment sites. Panels on the

left show the time-varying sensitivity of the RMSE computed over a time window of $2w + 1 = 31$ days, while panels on the right show the aggregate sensitivity of the RMSE computed over the entire simulation period.

English River

We first analyse the top panels in Figure 3, which refer to the English River. By comparing the two panels, we notice in the first place that aggregation indeed induces a loss of information. For example, by looking at aggregate results (top right), we would conclude that uncertainty in flow observations is the least important for this catchment; however, when looking at time-varying results (top left), we see that it can have a strong influence although very localized on some specific events. Our results thus confirm that GSA of aggregate performance metrics might not convey the same information as GSA of disaggregate (time varying) metrics.

In general, parameter uncertainty is more influential than data uncertainty in this catchment. Parameters of the soil moisture accounting module are the most influential

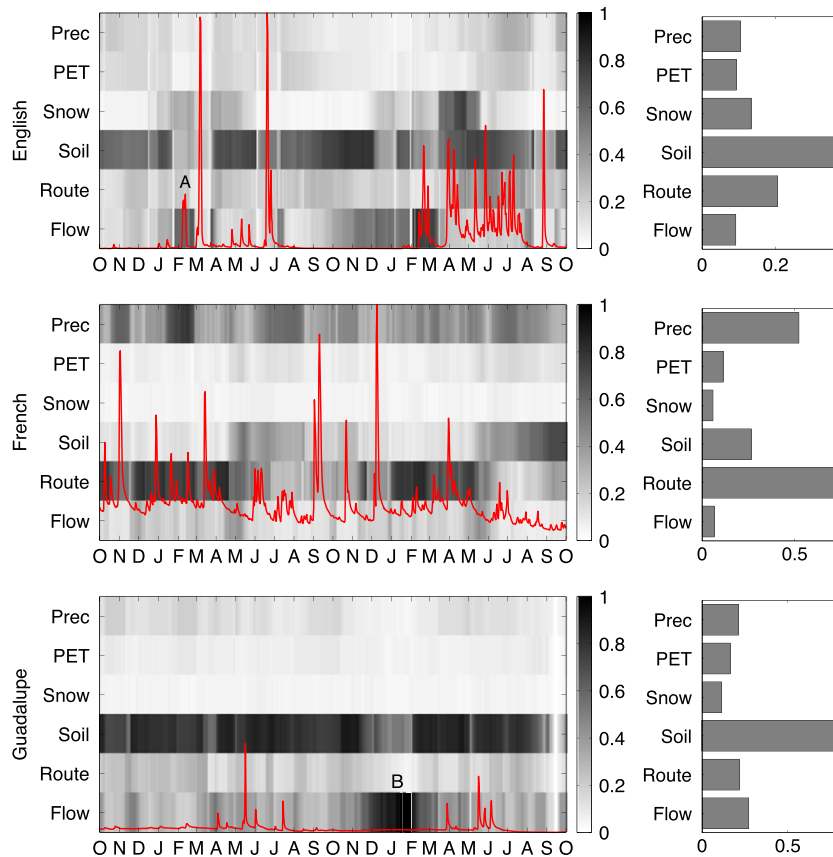


Figure 3. PAWN sensitivity indices (ranging from 0 to 1) of RMSE for the English River catchment (top panel), the French Broad River catchment (middle) and the Guadalupe River (bottom). Uncertainty sources considered by sensitivity analysis are precipitation data (*rain*), potential evapotranspiration data (*evap*), parameters of the snowmelt/accumulation component (*snow*), parameters of the soil moisture accounting component (*soil*), parameters of the flow routing component (*route*) and flow data (*flow*). Left panels: sensitivity indices of RMSE computed over a moving window of 31 days (for the sake of clarity, only the last 2 years of the simulation period are shown). Red line is the time series of observed flow. Right panels: sensitivity indices of RMSE computed over the entire simulation period

among the three groups. As expected, the snow parameters are particularly influential in those times of the year where snowmelt occurs while they have no influence in summer time, which confirms that the model's behaviour is consistent with the system's behaviour.

Interestingly, uncertainty in precipitation data does not seem to have a strong influence in this catchment. This might be due to a limitation in the assumed error model for precipitation, which only allows for variations in precipitation intensity but not in the temporal distribution of precipitation days. Allowed variations in the parameters instead are such that they might amplify or reduce timing errors, which are likely more influential than amount errors.

As anticipated before, uncertainty in flow data has a relatively high influence in some specific events. To understand this better, we analysed some of those events in more detail. Figure 4 shows this analysis for the event labelled as A in the top left panel of Figure 3. The top panel in Figure 4 shows the time series of observed precipitation, temperature and flow for this event. It can be noticed that the flow peak observed around day 500 has no clear explanation from the input forcing data: There is no precipitation prior to the event, neither a significant temperature increase that could produce a large amount of snowmelt (notice that a similar increase in temperature occurring some days before did not produce any increase in flow). This event might thus be an example of disinformative data. Time-varying SA (TVSA) attributes a key role to uncertainty in flow observations because if those observations were lower, then the model performance could be significantly higher, and vice versa. This is exemplified in the bottom panels in Figure 4. The left one reports, as an example, two sampled time series of flow perturbations used in our TVSA. The dashed line generates a perturbed time series where flow observations are increased; the continuous line generates a perturbed time series where flow observations are reduced. The bottom right panel in Figure 4 shows the conditional CDFs of the RMSE for day 500 when these two flow time series are used (while varying all other sources of uncertainty). The red line in this figure is the unconditional CDF, which is obtained by varying all sources of uncertainty including flow observations. The figure shows that when reducing flow observations (continuous line), the CDF is shifted towards the left, i.e. lower values of RMSE become more frequent. This means that, regardless of the variations in the other input factors (parameters and forcing inputs), the model would be likely to perform better if flow observations were lower. The opposite would happen if flow observations were higher. Uncertainty in flow observations plays a key role with

respect to other sources of uncertainty in that event, and this is why its sensitivity index is high. This is an example of how TVSA could be used to determine disinformative data.

French Broad River

TVSA results for the French Broad River are given in the middle panel of Figure 3. They show that in this catchment, uncertainty in precipitation data has much larger influence than in the English River and is as influential as parameter uncertainty during some high flow events. Among different groups of parameters, routing parameters are more influential than soil ones. A possible reason for this is that this catchment is very wet, and therefore, soil parameters matter less. Furthermore, there are many small events where timing errors might relate more to the routing than to the runoff production (soil) parameters – especially compared with drier catchments.

Results for the snow parameters deserve a further comment. Because this catchment is not affected by snow accumulation and melt, the relevant module is actually switched off here. This is obtained by fixing the snow parameter TS to $-\infty$ and all other snow parameters ($CFMAX$, CFR and CWH) to zero, so that all precipitation is turned into rainfall and immediately diverted to the soil

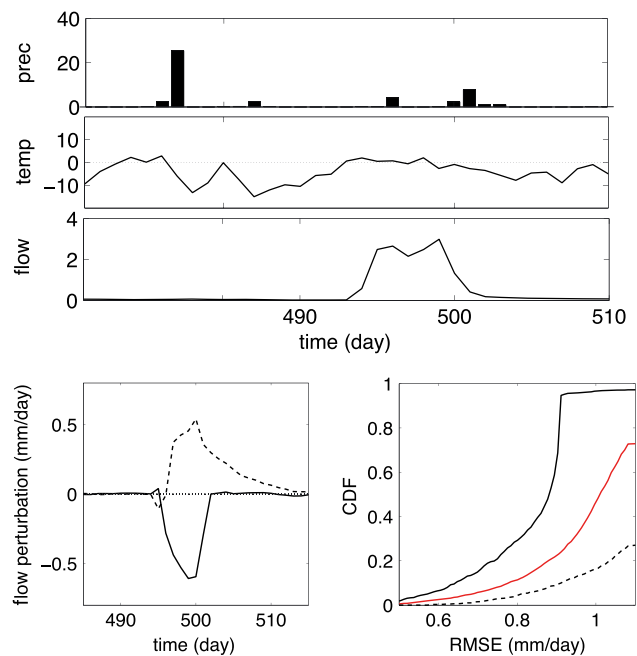


Figure 4. Top: observed precipitation (mm/day), temperature (C) and flow (mm/day) for the English River catchment around day 500. Bottom left: example of two time series of the perturbation e_t applied to the flow observations around day 500. Bottom right: unconditional CDF of the RMSE for day 500 (red) and two conditional CDFs (black) obtained from the flow perturbation time series in the left panel. These are the CDFs used to compute the PAWN sensitivity index for flow uncertainty source (for the sake of clarity, only 2 out of the 20 conditional CDFs are displayed)

moisture accounting module. Hence, when performing TVSA, the model is evaluated against the same combination of snow parameters $([-\infty, 0, 0, 0])$. In principle, we should therefore obtain a zero-valued sensitivity index for the *snow* parameter group. In practice, this does not happen because the sensitivity index of Equation 4 is approximated using empirical CDFs, and the empirical CDFs of two different samples can differ (by a small amount) even if the underlying probability distribution is the same. Indeed, in both the time-varying and aggregate cases, we obtain very low but non-zero sensitivity values. Although physically meaningless, these values are interesting because they give us a reference to evaluate the accuracy of the sensitivity indices of the other input factors. For instance, in the middle right panel of Figure 3, we see that the sensitivity index of flow data uncertainty is of the same order as that of the *snow* parameters, which means that the measured sensitivity to flow data uncertainty is within the range of numerical approximation errors, and might thus be regarded as negligible.

Guadalupe River

Finally, the bottom panels of Figure 3 show the TVSA results for the Guadalupe River near Spring Branch, Texas. Similarly to the French Broad River, this catchment is not affected by snowmelt, and the sensitivity estimates for the snow parameters are reported only as a reference to infer the approximation accuracy of the other indices. Here, uncertainty in soil parameters is by far the most influential source of uncertainty, which is consistent with the fact that the catchment is very dry, and therefore, the soil dynamics, which control the separation of precipitation into runoff and evaporation, dominates over routing for the lumped model we are analysing here.

Similarly to the English River, uncertainty in precipitation data has a limited influence in relation to parameter uncertainty. As for uncertainty in flow observations, TVSA reveals a very high sensitivity in one specific time period, i.e. the beginning of the third year of the simulation period (see letter B in Figure 3). We therefore analysed the third year of the simulation period in more detail. Figure 5 depicts the simulated and observed flows in that period and exhibits two very different behaviours in dry conditions. At the beginning of the year (days 800–850), the observed hydrograph (red line) flattens at a value of about 0.05 mm/day, while at the end of the year (days 1050 onwards), it goes to zero. In both periods, precipitation events (shown at the top of the panel) are equally infrequent and low. This evidence suggests an inconsistency in the data, with days 800–850 suggesting that the catchment can sustain a flow of about 0.05 mm/day even after a prolonged dry period and days 1050–1096 suggesting that the flow goes to zero. This is why

TVSA shows a higher sensitivity to flow observations around days 800–850: Just as for the English River, the most effective way to improve RMSE in this period is by perturbing (i.e. decreasing) observed flows.

The second influential factor in the dry period is soil parameters. To further investigate the time-varying relationship between soil parameters and model performance, we applied the DYNIA (Wagener *et al.* (2003) to the available sample of 3000 independent Monte Carlo simulations (which we generated to build the unconditional CDFs for PAWN). As an example, the left panel in Figure 6 shows the DYNIA results for the exponential parameter *BETA*. Here, the colour scale represents the frequency distribution of *BETA* over the subsample of the top 5% performing simulations. It can be noticed that in the dry period around days 800–850, higher performances are more frequently obtained with low values of *BETA* (around 1.96), while in the dry period just after day 1050, they are obtained with high values of *BETA* (around 5.55). The reason is that increasing *BETA* reduces the runoff from the soil moisture accounting component (see right panel in Figure 6), thus allowing for increased evaporation and reduced flow, while reducing *BETA* increases the runoff and therefore the flow. In other words, DYNIA reveals how parameter *BETA* can be varied to (almost) close the water balance and compensate for inconsistencies in data.

This example shows how we can use the combination of TVSA, output visualization and DYNIA to understand model shortcomings and potentially disinformative data periods. TVSA highlights an unexpected period of parameter sensitivity, while the streamflow plot shows that the model is not capable of encapsulating the observed flow. Time-varying parameter analysis then further suggests how the model is trying to reach the observed flow by decreasing parameter *BETA*. Given that the model cannot reach the observations, it is likely that data error is to blame, rather than model structural shortcomings.

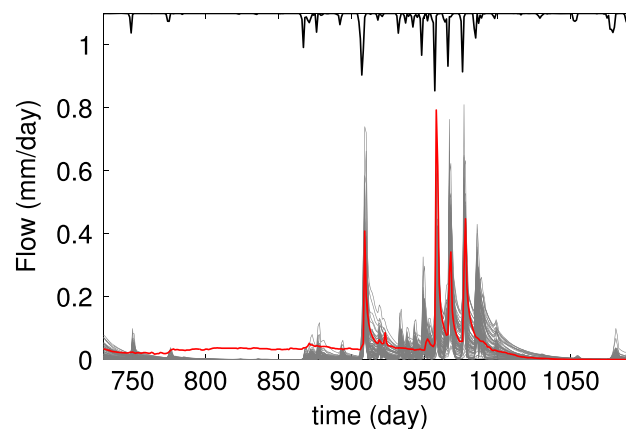


Figure 5. Observed (red) and simulated (grey) flow for the Guadalupe River in the third year of the simulation period. Black line shows the (reversed) precipitation data

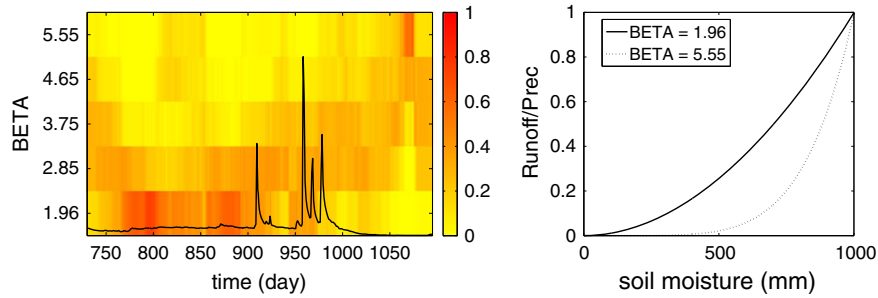


Figure 6. Left: dynamic identifiability analysis (DYNIA) of the soil parameter $BETA$ for the Guadalupe River in the third year of the simulation period. The colour scale here represents the frequency distribution of $BETA$ in the subsample of the top 5% performing simulations. Black line is the time series of observed flow. Right: the effect of two different choices of parameter $BETA$ on the runoff from the soil moisture accounting routine (underlying equation is $Runoff = Prec (SM/FC)^{BETA}$ where SM is the soil moisture content and FC is field capacity)

Impact of the moving window size

The last analysis we performed was aimed at evaluating the impact of the chosen window size. In fact, the limited influence of precipitation data uncertainty in the English and Guadalupe catchments might be attributed to the fact that performances are averaged over a relatively large moving window (31 days), while precipitation data uncertainty might influence the model accuracy on shorter timescales. To verify this explanation, we repeated our TVSA using a moving window of smaller size. We tried different sizes down to a minimum of 3 days ($w=1$). This analysis showed that changing the window size does not significantly impact sensitivity to precipitation errors. As an example, Figure 7 reports the sensitivity results for the extreme case $w=1$ (intermediate results with $1 \leq w \leq 15$ are qualitatively similar). Shortening the moving window increases the sensitivity to precipitation uncertainty, but the increase is rather small. We can thus conclude that precipitation uncertainty has limited impact on the model performance in the English and Guadalupe catchments regardless of the considered timescale. However, as discussed earlier, the uncertainty here investigated only deals with the intensity of precipitation data and not their temporal distribution.

Figure 7 also shows a main difference with respect to the previous sensitivity results, that is, the loss in sensitivity to the snow component parameters in the English catchment (compare the row labelled as *snow* in the top panels of Figures 3 and 7). The reason is that the snow accumulation and melt process is relatively slow, and therefore, its impact can be more clearly detected over a 31-day simulation period than a 3-day period. These results confirm that the choice of the window size can significantly impact sensitivity estimates, as also demonstrated in previous studies (e.g. Massmann *et al.*, 2014). While we cannot suggest a formal, objective way to define the window size a priori, we advise to check the impact of this choice by repeating the analysis for different window sizes. Here, the particular hydrological

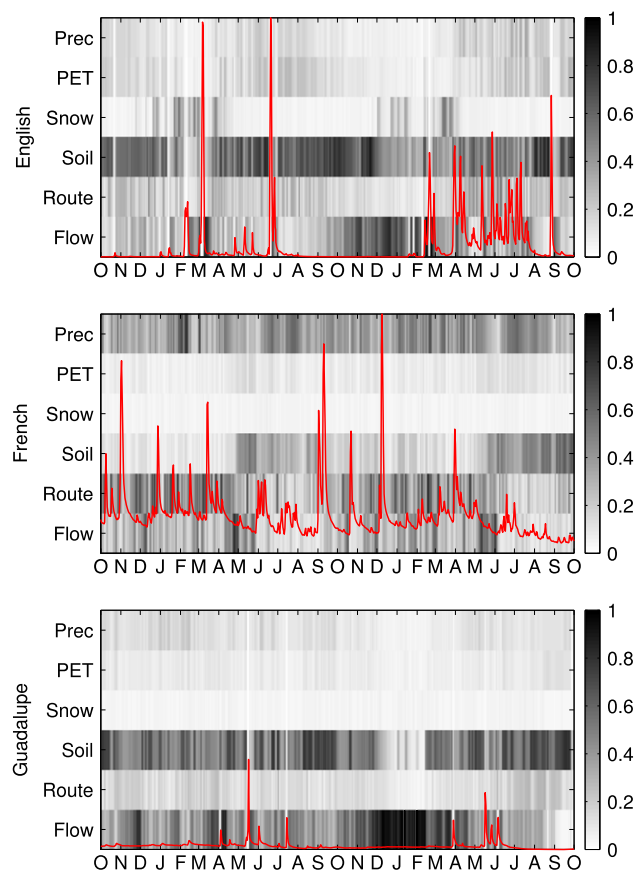


Figure 7. Same as left panels of Figure 3 but using a moving window of 3 days to compute the RMSE

meaning of each parameter is important to consider, e.g. parameters defining storage sizes require larger window sizes than quick residence times. If large differences are detected, either they can be given a physically meaningful interpretation or they should be further investigated as they could help in identifying conceptual weaknesses in the model and/or in the GSA set-up. Notice that such posterior analysis is not computationally expensive because calculating the time-varying

performance metric for a different window size and the associated sensitivity indices does not require re-running the simulation model.

LIMITATIONS OF OUR APPROACH AND FUTURE RESEARCH DIRECTIONS

This study demonstrates the use of TVSA to quantify the relative influence of different sources of uncertainty on the accuracy of a lumped hydrological model. While our results provide some interesting insights on how such influences vary in time and across catchments, it should be highlighted that they hold true under a number of assumptions and choices that we made in setting up the GSA. These choices include the hydrological model being used; the case study sites; several definitions of the experimental set-up of GSA, e.g. the simulation period, the definition of the performance metric and the selection and characterization of the uncertainty sources; and finally the choice of GSA method itself. In this section, we discuss some of these choices and their possible implications on our results and give directions for further research.

Our analysis is applied to the conceptual HBV model in the formulation presented in Kollat *et al.* (2012); hence, we cannot exclude that sensitivity estimates would be different if a different rainfall–runoff model was used. Herman *et al.* (2013a) compare TVSA for three different hydrological models and find significant inter-model differences. However, that study considers parameter uncertainty only and not the relative influence of parameter *versus* data uncertainty. Also, because in our approach we analyse groups of parameters related to the three model components (snow, soil, routing) rather than individual parameters, using different equations to represent individual hydrological processes in those components might have a smaller impact on group sensitivities. Another interesting direction for further research would be to regard the variability of model equations as an additional source of uncertainty (so-called model structure uncertainty) and expand our approach to assess the relative influence of structure uncertainty with respect to parameter and data uncertainty.

Another subjective choice is that of the performance metric. In this study, we use the RMSE, a metric that tends to be particularly responsive to how well the model reproduces the timing and shape of the hydrograph (e.g. Gupta *et al.*, 2009 and references therein) and therefore is usually sensitive to the parameters of both the soil moisture accounting and the flow routing component. Other metrics might produce different sensitivities. Such differences are typically significant when using aggregate performance metrics (an example is given by Shin *et al.*

(2013)); however, in the case of TVSA, our own experience is that this choice has a rather small impact unless the window size is very large (e.g. several months, see Wagener *et al.*, 2003).

As for the GSA method, a range of options has been used for TVSA in the past, including the Fourier Amplitude Sensitivity Testing (FAST) (Reusser and Zehe, 2011), regional sensitivity analysis (Wagener *et al.*, 2003; Sieber and Uhlenbrook, 2005), Sobol' (Kelleher *et al.*, 2013) and PAWN in this study. Given that the results across these studies are quite consistent with each other, we believe that also this choice is less crucial (but for very large window sizes), although we have not yet thoroughly tested this perception thus far.

Finally, one aspect we know that does have a large impact on our results is the characterization of the uncertainty sources. For example, we already discussed how the error model used to generate equiprobable time series of precipitation, which does not allow for timing errors, might have reduced the sensitivity to precipitation uncertainty. We might also expect an increased (reduced) sensitivity to data uncertainty if we increased (reduced) the variability of data errors (here set to $\pm 40\%$ for precipitation and $\pm 20\%$ for potential evaporation and flow data). Similarly, the definition of parameter ranges might be very important. For example, Kelleher *et al.* (2013) found that it was possible to separate parameter influences in time only after substantially reducing the uncertainty range of one parameter (cross-sectional area), because, otherwise, the variability of that parameter dominated the model response by producing unreasonable model outputs. In our approach, such effect might be mitigated by the fact that we filter out output samples that do not satisfy a minimum performance target (see the discussion after Equation 3). Yet, understanding the influence of the chosen characterization for the uncertainty sources remains a crucial aspect in the interpretation and transferability of GSA results.

CONCLUSIONS

In this paper, we investigate the relative importance of parameter and data uncertainty on the performance of a spatially lumped conceptual rainfall–runoff model via TVSA. We find that TVSA can reveal information on local sensitivities that would be hidden in SA of aggregate performance and that TVSA could provide a formal method to identify periods where data might be disinformative due to observational errors. We also find that the relative importance of different factors changes across catchments with different characteristics. Routing parameters have higher influences in a wet catchment where the runoff coefficient is higher and quick recession

phase is longer, while soil parameters are the main source of uncertainty in a dry catchment, where estimating the amount of water lost to the atmosphere as evapotranspiration has a large influence. Uncertainty in precipitation data has a significant influence in a wet catchment, while its influence is much more limited in a snow affected and a dry catchment. It is important to stress that the transferability of our conclusions beyond our test catchments has yet to be tested. The results are further limited by the assumptions made in our study set-up as discussed earlier.

While addressing these limitations left for future research, this work demonstrates that (1) the relative importance of data and parameter uncertainty, both in time and across different places, can be formally investigated by TVSA; (2) TVSA is a generic methodology that can be tailored and applied to other case studies in a relatively straightforward way; and (3) TVSA provides interesting insights for model diagnostic, identification of disinformative data and prioritization of efforts for uncertainty reduction.

ACKNOWLEDGEMENTS

This work was supported by the Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1]. The data used in this research are available from National Weather Service (2011) (available at http://www.nws.noaa.gov/oh/mopex/mo_datasets.htm). The SAFE Toolbox for Global Sensitivity Analysis is freely available for research purposes from www.bris.ac.uk/cabot/resources/safe-toolbox/. The specific MATLAB script used to develop the analysis presented in this paper is available upon request to the authors. The authors are grateful to the associate editor, two anonymous reviewers and Jan Seibert for their useful comment.

REFERENCES

- Bergstrom S. 1995. The HBV model. In *Computer Models of Watershed Hydrology*, Singh VP (ed). Water Resources Publications: Highlands Ranch, Colorado, USA: 443–476.
- Beven K, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**(3): 279–298.
- Beven KJ, Smith PJ, Freer J. 2008. So just why would a modeller choose to be incoherent? *Journal of Hydrology* **354**(1): 15–32.
- Beven K, Westerberg I. 2011. On red herrings and real herrings: disinformation and information in hydrological inference. *Hydrological Processes* **25**: 1676–1680.
- Beven K, Smith P. 2014. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *Journal of Hydrologic Engineering* **A4014010**.
- Binley A, Elgy J, Beven K. 1989. A physically based model of heterogeneous hillslopes: 1. Runoff production. *Water Resources Research* **25**(6): 1219–1226.
- Borgonovo E. 2007. A new uncertainty importance measure. *Reliability Engineering and System Safety* **92**: 771–784.
- Borgonovo E. 2014. Transformation and invariance in the sensitivity analysis of computer experiments. *Journal of the Royal Statistical Society, Series B* **76**(5): 925–947.
- Cloke H, Pappenberger F, Renaud J. 2008. Multi-method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes. *Hydrological Processes* **22**(11): 1660–1674.
- Duan Q, Schaake J, Andrassian V, Franks S, Goteti G, Gupta H, Gusev Y, Habets F, Hall A, Hay L, Hogue T, Huang M, Leavesley G, Liang X, Nasonova O, Noilhan J, Oudin L, Sorooshian S, Wagener T, Wood E. 2006. Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. *Journal of Hydrology* **320**(1–2): 3–17.
- Freer J, Beven K, Ambroise B. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* **32**(7): 2161–2173.
- Freer J, Beven K, Peters N. 2003. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In *Calibration of watershed models*, Duan Q, Gupta H, Sorooshian S, Rousseau A, Tourcotte R (eds), Water Science and Application Series 6. American Geophysical Union: Washington, DC.
- Freeze RA. 1980. A stochastic-conceptual analysis of rainfall-runoff processes on a hillslope. *Water Resources Research* **16**(2): 391–408.
- Gupta HV, Kling H, Yilmaz KK, Martinez GF. 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377**(1–2): 80–91.
- Gusev B, Reusser DE, Fohrer N. 2014. How to improve the representation of hydrological processes in SWAT for a lowland catchment. A temporal analysis of parameter sensitivity and model performance. *Hydrological Processes* **28**(4): 2651–2670.
- Herman JD, Reed PM, Wagener T. 2013a. Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research* **49**(3): 1400–1414.
- Herman JD, Kollat JB, Reed PM, Wagener T. 2013b. From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models. *Hydrology and Earth System Sciences* **17**(12): 5109–5125.
- Kavetski D, Franks S, Kuczera G. 2002. Confronting input uncertainty in environmental modelling. In *Calibration of watershed models*, Duan Q, Gupta H, Sorooshian S, Rousseau A, Tourcotte R (eds), Water Science and Application Series 6. American Geophysical Union: Washington, DC.
- Kavetski D, Kuczera G, Franks S. 2006. Bayesian analysis of input uncertainty in hydrological modelling. II. Application. *Water Resources Research* **42**(3). DOI: 10.1029/2005WR004376
- Kelleher C, Wagener T, McGlynn B, Ward AS, Gooseff MN, Payn RA. 2013. Identifiability of transient storage model parameters along a mountain stream. *Water Resources Research* **49**(9): 5290–5306.
- Kollat JB, Reed PM, Wagener T. 2012. When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resources Research* **48**(3). W03520, DOI: 10.1029/2011WR011534
- Massmann C, Wagener T, 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.
- McMillan H, Krueger T, Freer J. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes* **26**(26): 4078–4111.
- National Weather Service. 2011. Model parameter estimation experiment (MOPEX), available at http://www.nws.noaa.gov/oh/mopex/mo_datasets.htm, Hydrologic Science and Modelling Branch, NWS Office of Hydrologic Development, Silver Springs, MD.
- Pianosi F, Sarrazin F, Wagener T. 2015. A matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software* **70**: 80–85.
- Pianosi F, Wagener T. 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software* **67**: 1–11.

- Reusser DE, Zehe E. 2011. Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. *Water Resources Research* **47**(7). DOI: 10.1029/2010WR009946
- Rougier J, Sparks R, Hill L. 2013. *Risk and Uncertainty Assessment for Natural Hazards*. Cambridge University Press: Cambridge, UK.
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. 2007. Introduction to Sensitivity Analysis. In *Global Sensitivity Analysis, The Primer*. John Wiley & Sons, Ltd.: Chichester, UK.
- Schoups G, Vrugt JA. 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research* **46**, W10531. DOI: 10.1029/2009WR008933
- Seibert J. 1997. Estimation of parameter uncertainty in the HBV model. *Nordic Hydrology* **28**(4–5): 247–262.
- Shin M-J, Guillaume JH, Croke BF, Jakeman AJ. 2013. Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. *Journal of Hydrology* **503**: 135–152.
- Sieber A, Uhlenbrook S. 2005. Sensitivity analyses of a distributed catchment model to verify the model structure. *Journal of Hydrology* **310**(1–4): 216–235.
- Spear R, Hornberger G. 1980. Eutrophication in peel inlet. II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Resource Research* **14**: 43–49.
- Stedinger JR, Vogel RM, Lee SU, Batchelder R. 2008. Appraisal of generalised likelihood uncertainty estimation (glue) methodology. *Water Resources Research* (1): 43–49.
- van Werkhoven K, Wagener T, Reed P, Tang Y. 2008. Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research* **44**(1). W01429, DOI: 10.1029/2007WR00627
- Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. 2001. A framework for development and application of hydrological models. *Hydrology and Earth System Sciences* **5**(1): 13–26.
- Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes* **17**(2): 455–476.
- Wagner BJ, Harvey JW. 1997. Experimental design for estimating parameters of rate-limited mass transfer: analysis of stream tracer studies. *Water Resources Research* **33**(7): 1731–1741.
- Young PC, Spear RC, Hornberger GM. 1978. *Modeling Badly Defined Systems: Some Further Thoughts*. Proceedings SIMSIG Conference, Canberra; 24–32.

APPENDIX.

THE AUTOREGRESSIVE HETEROSCEDASTIC GAUSSIAN MODEL FOR FLOW ERRORS

Time series of perturbed flow observations $q_t^{\text{obs,per}}$ are obtained from the available flow observations q_t^{obs} via the additive model:

$$q_t^{\text{obs,per}} = q_t^{\text{obs}} + e_t \quad (4)$$

In order to preserve the persistence of the original time series, the error term e_t is described as an autoregressive process, i.e.

$$e_t = \alpha e_{t-1} + g_t \quad (5)$$

where g_t is randomly generated from a zero mean Gaussian distribution. In order to account for the fact that observational errors are typically larger at higher flow values, the standard deviation of the Gaussian distribution is not constant in time, but it is assumed proportionally to the flow observation on the current time step, i.e.

$$\text{STD}[g_t] = \beta q_t^{\text{obs}} \quad (6)$$

The model thus requires specifying two parameters, α and β . In our application, we set the parameter α to 0.8 and derived the value for β as

$$\beta = 0.2 \frac{\sqrt{(1-\alpha^2)}}{3} \quad (7)$$

This choice for β ensures that at each time step, the perturbation e_t is within the interval $\pm 0.2 q_t^{\text{obs}}$ with probability 0.99. The proof is as follows.

Given that e_t is a Gaussian variable, at each time step, its value falls with 0.99 probability within the interval $\pm 3\text{STD}[e_t]$. From Equation 4, the standard deviation of e_t is related to that of g_t by the equation

$$\text{STD}[e_t] = \frac{1}{\sqrt{(1-\alpha^2)}} \text{STD}[g_t] \quad (8)$$

By replacing Equation 6 into Equation 8, we obtain

$$\text{STD}[e_t] = \frac{\beta}{\sqrt{(1-\alpha^2)}} q_t^{\text{obs}} \quad (9)$$

Therefore, at each time step, e_t is within the interval $\pm 3\beta/\sqrt{(1-\alpha^2)} q_t^{\text{obs}}$ with probability 0.99. In order to ensure that relative errors do not exceed ± 0.2 , we need to ensure that $3\beta/\sqrt{(1-\alpha^2)} = 0.2$ and therefore Equation 7.