Hydrol. Earth Syst. Sci. Discuss., 3, 3099–3146, 2006
www.hydrol-earth-syst-sci-discuss.net/3/3099/2006/

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology

M. Ratto[1], P. C. Young[2,3], R. Romanowicz[2], F. Pappenberge[2], A. Saltelli[1], and A. Pagano[1]

[1]Institute for the Protection and Security of the Citizens, Joint Research Center European Community, Ispra, Italy
[2]Centre for Research on Environmental Systems and Statistic, Lancaster University, Lancaster, UK
[3]Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia

Received: 7 April 2006 – Accepted: 15 May 2006 – Published: 25 September 2006

Correspondence to: Marco Ratto (marco.ratto@jrc.it)

Abstract

In this paper, we discuss the problem of calibration and uncertainty estimation for hydrologic systems from two points of view: a bottom-up, reductionist approach; and a top-down, data-based mechanistic (DBM) approach. The two approaches are applied to the modelling of the River Hodder catchment in North-West England. The bottom-up approach is developed using the TOPMODEL, whose structure is evaluated by global sensitivity analysis (GSA) in order to specify the most sensitive and important parameters; and the subsequent exercises in calibration and validation are carried out in the light of this sensitivity analysis. GSA helps to improve the calibration of hydrological models, making their properties more transparent and highlighting mis-specification problems. The DBM model provides a quick and efficient analysis of the rainfall-flow data, revealing important characteristics of the catchment-scale response, such as the nature of the effective rainfall nonlinearity and the partitioning of the effective rainfall into different flow pathways. TOPMODEL calibration takes more time and it explains the flow data a little less well than the DBM model. The main differences in the modelling results are in the nature of the models and the flow decomposition they suggest. The "quick" (63%) and "slow" (37%) components of the decomposed flow identified in the DBM model show a clear partitioning of the flow, with the quick component apparently accounting for the effects of surface and near surface processes; and the slow component arising from the displacement of groundwater into the river channel (base flow). On the other hand, the two output flow components in TOPMODEL have a different physical interpretation, with a single flow component (95%) accounting for both slow (subsurface) and fast (surface) dynamics, while the other, very small component (5%) is interpreted as an instantaneous surface runoff generated by rainfall falling on areas of saturated soil. The results of the exercise show that the two modelling methodologies have good synergy; combining well to produce a complete modelling approach that has the kinds of checks-and-balances required in practical data-based modelling of rainfall-flow systems. Such a combined approach also produces models that are

suitable for different kinds of application. As such, the DBM model can provides an immediate vehicle for flow and flood forecasting; while TOPMODEL, suitably calibrated (and perhaps modified) in the light of the DBM and GSA results, immediately provides a simulation model with a variety of potential applications, in areas such as catchment management and planning.

1 Introduction

Uncertainty estimation is a fundamental topic in hydrological and hydraulic modelling. Mathematical models are always an approximation to reality and the evaluation of the uncertainties is one of the key research priorities in every modelling process. The most widely used approach to modelling is based on the description of physical and natural systems by deterministic mathematical equations based on well known scientific laws: the so-called reductionist (bottom-up) approach. Uncertainty estimation is dealt with via calibration and estimation procedures that insert the deterministic approach into a stochastic framework (see a recent review in Romanowicz and Macdonald, 2005, and the references cited therein).

A widely used approach to calibration in hydrology, and environmental sciences in general, is the Generalised Likelihood Uncertainty Estimation (GLUE, Beven and Binley, 1992). This is a very flexible, Bayesian-like approach to calibration, which is based on the acceptance that different parameterisations, as well as different model structures, can be equally likely simulators of the observed systems (referred to variously as "ambiguity", "unidentifiability" or "model equifinality"). In GLUE, model realisations are weighted and ranked on a likelihood scale, via conditioning on observations, and the weights are used to formulate a cumulative distribution of predictions.

In the calibration (model identification and estimation) framework, the understanding, rather than just the evaluation of the influence of different uncertainties on the modelling outcome, becomes a fundamental question. This is especially the case if the modelling goal is to reduce the model uncertainties and resources for this task are

limited. Sensitivity analysis (SA) can play an important role in this framework: it can help in better understanding the model structure, the main sources of model output uncertainty and the identification issues (Ratto et al., 2001). For instance, Pappenberger et al. (2006a, b) and Hall et al. (2005) have recently presented cases that achieve such an understanding for flood inundation models, using sensitivity analysis to support their analysis.

The GLUE approach provides a methodology for coping with the problems of lack of identification and over-parameterisation inherent in the reductionist approach, while a detailed sensitivity analysis can make these problems more transparent. Neither of the two approaches, however, can provide a full solution of these problems.

Another approach to coping with uncertainty in environmental processes is the "top-down" Data Based Mechanistic (DBM) method of modelling, introduced by Young over many years (see Young, 1998, and the prior references therein). Here, the models are derived directly from the data and the approach is based on identifying and estimating a stochastic dynamic relationship between the input and output variables using advanced time series analysis tools, with possible non-linearities introduced by means of non-linear State-Dependent Parameter (SDP) transformation of the model variables. Unlike "black-box" modelling, however, DBM models are only accepted if the estimated model form can be interpreted in a physically meaningful manner. This DBM methodology can also be applied to the analysis and simplification of large deterministic models (e.g., Young et al., 1996).

In contrast to the GLUE technique, the DBM approach chooses, from amongst all possible model structures, only those that are clearly identifiable: i.e., those that have an inverse solution in which the model parameters are well defined in statistical terms. Moreover, this technique estimates the uncertainty associated with the model parameters and variables, normally based on Gaussian assumptions. If the Gaussian assumptions are strongly violated, however, the DBM analysis exploits Monte Carlo Simulation (MCS) to evaluate the uncertainty. For instance, it is common in DBM analysis to estimate the uncertainty in derived physically meaningful model parameters, such as residence times and flow partition parameters, using MCS analysis. Also, when the identified model is nonlinear, then the uncertainty in the state and output variables (e.g. internal, unobserved flow variables and their combination in the output flow) is normally evaluated using MCS analysis.

In this paper, we will discuss the problems of calibration and uncertainty estimation for hydrologic systems from both points of view: the bottom-up, reductionist approach; and the top-down, data-based mechanistic approach. We will also highlight the role of sensitivity analysis in better calibrating the mechanistic (reductionist) models and in highlighting both their weaknesses and possible mis-specifications. Moreover, we will show the key role of data-based mechanistic modelling in providing a succinct and well identifiable explanation of the data, which can also be used to guide the analyst in improving the features of the bottom-up, mechanistic model.

2 Data-based mechanistic modelling

The Data-based Mechanistic (DBM) approach to modelling rainfall-flow processes goes back a long way (Young, 1974) but the more powerful technical tools for its application have been developed much more recently (see e.g. Young, 1998, 2003; Young and Beven, 1994). DBM modelling involves four, relatively objective, stages that exploit advanced methods of time series analysis: identification of the model structure; estimation of the parameters that characterize this identified model structure; interpretation of the estimated model in physically meaningful terms; and validation of the estimated model on rainfall-flow data that is different from the calibration data used in the identification and estimation analysis.

Normally, the identification stage of the modeling is based on the non-parametric estimation of a nonlinear State-Dependent Parameter (SDP) transfer function model, where the SDP relationships that define the nonlinear dynamic behaviour of the model are estimated in a graphical form, with the SDPs plotted against the states on which they are identified to be dependent. In the case of rainfall flow models, there is nor-

mally only one such SDP relationship: this is associated with the rainfall input and the state dependency is identified, paradoxically at first, in terms of the measured flow variable (see later). In physical terms, this nonlinear function defines the connection between the measured rainfall and the "effective rainfall", i.e. the rainfall that is effective in causing flow variations. The relationship between this effective rainfall and the flow is then normally identified as a linear process and is estimated in the form of a 2nd (or on rare occasions 3rd order Stochastic Transfer Function (STF) model, whose impulse response represents the underlying unit hydrograph behaviour.

In the estimation stage of the DBM modelling, the non-parametric SDP nonlinearity is parameterized in a parsimonious (parametrically efficient) manner and the complete model, composed of this nonlinearity in series with the linear STF model, is estimated using appropriate statistical methods. Normally, the parameterization of the nonlinearity is in terms of a power law or an exponential function. The approach used in the later example has been developed in the previous DBM modeling of rainfall flow processes. Here, the STF model and the power law parameters, are estimated simultaneously using a special, nonlinear least squares optimization procedure that exploits the Refined Instrumental Variable (RIV) transfer function estimation algorithm in the CAPTAIN Toolbox for Matlab (See http://www.es.lancs.ac.uk/cres/captain/). Depending on the application of the DBM model, this simple optimization procedure can be extended to handle a more sophisticated model of the additive noise process that includes allowance for both autocorrelation and heteroscedasticity (changing variance). In the case of the heteroscedasticity, the variance is normally considered as a SDP function of the flow, with higher variance operative at higher flows.

The physical interpretation of this estimated DBM model is normally straightforward. As pointed out above, the input SDP function can be considered as an effective rainfall nonlinearity, in which the SDP is dependent on the flow. Although unusual at first sight, this state dependency makes physical sense because the flow is acting simply as a surrogate for the soil moisture content in the catchment (catchment storage: see the previously cited references). Thus, when the flow is low, implying low soil moisture, the rainfall is multiplied by a small gain and the effective rainfall is attenuated; whilst at high flow and soil moisture, this gain and the associated effective rainfall are large, so that the rainfall has a much larger effect on the flow.

The effective rainfall is the input to the STF part of the model, which characterizes the dominant catchment dynamics and defines the underlying catchment scale hydrograph. Typically, this 2nd order STF model can be decomposed into a parallel form that reveals the major identified flow pathways in the effective rainfall-flow system. These consist of a "quick" flow component, with a short residence time that seems to account for the surface and near surface processes; and a "slow" component with a long residence time (time constant), that can be associated with the replenishment of the groundwater storage and the consequent displacement of groundwater into the river channel. Of course, these components are unobserved (not measured directly) and so their identification is one of statistical inference based on the specific model form: in this case, a generic nonlinear differential equation, the specific structure of which is identified from the data with the minimum of a priori assumptions. Consequently, the assumption of another model form will probably result in different flow decomposition. Indeed, as we shall see later, this is the situation in the present context, where the flow decompositions of the DBM model and TOPMODEL are quite different. Moreover, the interpretation of the decomposed flow components is subjective, depending on not only the model form but also on how the modeller views the unobserved flow components within the hydrological context.

We see from the above analysis of the DBM model that, although it is identified and estimated in a relatively objective, inductive manner from the rainfall-flow data, without the imposition of any strong assumptions about the physical nature of the hydrological processes active in the catchment, it can be interpreted quite straightforwardly in a manner which has sensible (if not universally accepted) hydrological meaning. However, any model of a real system is only acceptable in technical terms if it can be well validated. The final validation stage in DBM modelling is similar to that used in most hydrological modelling exercises. The model estimated on the calibration data set is

applied, without re-estimation, to a new set (or sets) of rainfall-flow data from the same catchment. If this validation is successful, in the sense that the simulated output of the model matches the measured flow to within the uncertainty levels quantified during the model estimation (calibration), then the model is deemed "conditionally valid" and can be used in whatever capacity it is intended.

The DBM model for the River Hodder, as considered in the later example, is well suited for use in data assimilation and flow forecasting (see Young, 1998; 2003). However the main purpose of the modelling in the present paper is to investigate the dynamic characteristics of the catchment, so that these can be compared with those inferred by the alternative TOPMODEL (Beven and Kirkby, 1979). In contrast to the DBM model, TOPMODEL is obtained from the rainfall-flow data by a process of hypothesis and deduction, with the model evaluation assisted by the application of Global Sensitivity Analysis (GSA). However, to aid comparison between the two alternative models, calibration and validation are carried out using the same data sets.

In comparing the models and their associated modelling methodologies (see e.g. Young, 2002), it is important to note that the DBM model is obtained by a relatively objective inductive analysis of the data, with prior assumptions kept to the minimum but with the aim of producing a model that can be interpreted in a reasonable, physically meaningful manner. The sensitivity analysis in this case is an implicit part of the statistical identification and estimation analysis: indeed, sensitivity functions computed within the estimation process are influential in identifying the model structure and order. On the other hand, the TOPMODEL structure represents one particular hypothesis about the nature of the hydrological processes involved in the transfer of rainfall into river flow and the model synthesis follows the hypothetico-deductive approach. In this case, the pre-defined model structure is evaluated critically by overt global sensitivity analysis, which is used to specify the most sensitive and important parameters; and the subsequent exercises in calibration and validation are carried out in the light of this sensitivity analysis.

3 Global sensitivity analysis and model calibration

Since some aspects of sensitivity analysis are not all that well known, this section will provide an outline of the main topics in sensitivity analysis that are important within the context of the present paper. We can formalise any mathematical or computational model as a mapping $Y = f(X_1, \ldots, X_k)$, where X_i are the uncertain input factors. We interpret the term "factor" in a very broad sense: a factor is anything that can be subject to some degree of uncertainty in the model. All X_i's are treated as random variables characterised by specified distributions, implying that the output Y is also a random variable, with a probability distribution whose characterisation is the object of uncertainty analysis. Sensitivity analysis is "The study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" (Saltelli et al., 2000). This definition is general enough to cover a variety of strategies for sensitivity analysis, while committing the strategy to some sort of quantitative partition of the output uncertainty (however this uncertainty is defined) into factors-related components.

In the history of sensitivity analysis, hydrologists and environmental scientists have contributed with one of the strategies today considered as a good practice and named by its proponents "regionalised sensitivity analysis", RSA (Young et al., 1978, 1996; Hornberger and Spear, 1981; Spear et al., 1994; Young, 1999). Beside RSA, we would also like to outline, in this section, other strategies that have received acceptance amongst practitioners, together with a discussion of the "settings" for sensitivity analysis. This is because, as Saltelli et al. (2004) have argued, the effectiveness of a sensitivity analysis is greater if the purpose of the analysis is specified unambiguously beforehand. Over time, practitioners have identified cogent questions for sensitivity analysis. These questions define the setting, which in turn allows for the selection of the strategy. These settings and the associated methods are described succinctly in the next sub-sections.

3.1 Variance based methods

Variance-based sensitivity indices are the most popular measures of importance used in global SA. The two key measures are the main effect

$$V_i = V[E(Y|X_i)] \tag{1}$$

and the total effect

$$V_{Ti} = E[V(Y|\mathbf{X}_{-i})] \tag{2}$$

where $\mathbf{X}_{-i}$ indicates the array of all input factors except X_i, V and E denote variance and expectation operators. All measures are usually normalised by the unconditional variance of Y, to obtain the sensitivity indices, scaled between 0 and 1:

$$\begin{aligned} S_i &= V_i/V(Y) \\ S_{Ti} &= V_{Ti}/V(Y) \end{aligned} \tag{3}$$

The S_i spectrum is sufficient to characterise the entire sensitivity pattern of Y only for additive models, for which

$$\sum_{i=1}^{k} S_i = 1 \tag{4}$$

Equation (4) tells us that all the variance of the model Y can be explained in terms of first order effects. Models for which (4) does not hold are termed non-additive. For non-additive models $\sum_{j=1}^{r}\left(S_j\right) \le 1$ and these models are characterised by the existence of so-called interaction effects, leading to the most general variance decomposition scheme (Sobol, 1990),

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \sum_i \sum_{j>i} \sum_{l>j} S_{ijl} + \ldots S_{12\ldots k} = 1 \tag{5}$$

The complete decomposition (5) comprises an array of $2^k - 1$ sensitivity terms, giving rise to the so-called "curse of dimensionality", since the expression and its decomposition is neither cheap to compute, nor does it provide a succinct and easily readable portrait of the model characteristics. In this context, total indices provide the major part of the information needed to complement the main effects, with only k additional indices to be estimated. Total indices measure the overall effect of input factors, including both main effects and all possible interactions with any other input factor. Returning to the additivity of models, when a model is additive, we will have $S_i = S_{Ti}$ for all X_i's; while, in general, $S_i \le S_{Ti}$ and the difference between main and total effects is due to all interaction terms involving X_i. The main effects and total effects are also strictly linked to two important sensitivity settings that are extremely relevant in the calibration context: Factors prioritization and Factors fixing.

3.1.1 Factors prioritization (FP)

Assume that, in principle, the uncertain input factors X_i can be "discovered", i.e. determined or measured, so as to find their true value. One legitimate question is then "which factor should one try to determine first in order to have the largest expected reduction in the variance of Y"? This defines the "factors fixing" setting. Saltelli and Tarantola (2002) have shown that the main effect provides the answer to the FP setting, so that ranking input factors according to the S_i values allows the analyst to guide research efforts that reduce the uncertainty in Y, by investing resources on the factor having the largest S_i.

3.1.2 Factors fixing (FF)

Another aim of sensitivity analysis is to simplify models. If a model is used systematically in a Monte Carlo framework, so that input uncertainties are systematically propagated into the output, it might be useful to ascertain which input factors can be fixed, anywhere in their range of variation, without sensibly affecting a specific output of interest, Y. This may be useful for simplifying a model in a larger sense, because we may be able then to condense entire sections of our models if all factors entering in a section are non-influential. Saltelli and Tarantola (2002) also showed that the total

effect provides the answer to the FF setting, and the ranking of input factors according to the S_{Ti} values allows us to restrict the research efforts by "fixing" the factors having null S_{Ti}'s. It is useful to note here that the condition $S_i = 0$ alone is not sufficient for fixing factor X_i. This factor might be involved in interactions with other factors, so that although its first order term is zero, there might be non-zero higher order terms.

Both the FP and the FF settings are extremely important in the calibration context: FP matches the need of highlighting the key input factors driving the uncertainty of the model predictions and possibly reducing them; while FF matches the need to identify irrelevant compartments of the model that, subsequently, can be simplified. For example, when applied to the likelihood weights in a GLUE procedure, as in Ratto et al. (2001), input factors can be classified as:

- factors with high main effect: the analyst has to concentrate on these to reduce the prediction uncertainty (FP);
- factors with small total effect: such factors have a negligible effect on the model performance and can be fixed at a nominal value (FF).
- factors with small main effect but high total effect: here, such a situation flags an influence mainly through interaction, implying lack of identification.

The latter situation, when it characterises the entire spectrum of input factors, would flag a particularly badly defined calibration problem, whereby the analyst is not allowed to identify any prioritisation to reduce prediction uncertainty; nor introduce any fixing to simplify the model structure. This usually corresponds to a highly over-parameterised, unidentifiable model.

3.2 Monte Carlo filtering and Regionalised Sensitivity Analysis

We now present an altogether different setting for sensitivity analysis, strictly related to calibration. We call this "Factors Mapping" and it relates to the situations where we are especially concerned with particular points or portions of the distribution of the output Y. For example, we are interested in Y being above or below a given threshold: e.g. Y could be a dose of contaminant and we are interested in how much (how often) a threshold level for this dose level is being exceeded; or Y could be a set of constraints, based on the information available on observed systems. The latter situation is typical in calibration. In these settings, we will naturally tend to partition the realisation of Y into "good" and "bad". This leads very naturally to Monte Carlo Filtering (MCF), where one runs a Monte Carlo experiment producing realisations of the output(s) of interest corresponding to different sampled points in the input factors space. Having done this, one "filters" the realizations of Y.

Regionalised Sensitivity Analysis (RSA, see Young et al., 1978, 1996; Hornberger and Spear, 1981; Spear et al., 1994; Young 1999, and the references cited therein) is an MCF procedure that aims to identify which factors are most important in leading to realisations of Y that are either in the "behaviour" or "non-behaviour" regions. Note that, in this context, one is not interested in the variance of Y but in which factors produce realizations of Y in the specified zone. In the simplest cases, RSA can answer this question by examining, for each factor, the subset corresponding to "behaviour" and "non-behaviour" realisations. It is intuitive that, if the two subsets are dissimilar from one another (as well as dissimilar, one would expect, from the initial marginal distribution of that factor), then that factor is influential. Standard statistical tests such as the Smirnov test are usually applied for this purpose.

The GLUE technique can be seen as an extension of the RSA methodology where, instead of the binary classification behaviour/non-behaviour, model realisations are weighted and ranked on a likelihood scale via conditioning on observations

3.3 Other methods

In general, the variance-based measures constitute good practice for tackling settings. The main problem is computational cost. Estimating the sensitivity coefficients takes many model realisations. Accelerating the computation of sensitivity indices of all or-

ders, or even simply of the S_i, S_{Ti} couple, is the most intensely researched topic in sensitivity analysis. Recently, various authors presented efficient techniques for estimating main effects and low order effects up the 3rd level, using a single Monte Carlo sample of size N (Li et al, 2002, 2006; Oakley and O'Hagan, 2004; Ratto et al., 2004, 2006). This makes the estimation of main effects very efficient (N=250 to 1000). However, the estimation of total effects still requires a larger number of runs: $Ntot{=}Nk{+}2$, where k is the number of input factors and N is as large as 500–1000. So, it would be useful to have methods capable of providing approximate sensitivity information at lower sample sizes. One such simple method, the Elementary Effect Test, is to average derivatives over the space of factors. Elementary effects are defined as

$$EE_i^j = \frac{\left| Y\left(x_1^j, x_2^j, \dots x_{i-1}^j, x_i + \Delta_i^j, x_{i+1}^j, \dots x_k^j\right) - Y\left(x_1^j, x_2^j, \dots x_k^j\right) \right|}{\Delta_i^j} \tag{6}$$

where, for each factor X_i and selected grid points j, a modulus incremental ratio is computed. Then, for each factor, the EE_i^j computed at different grid points are averaged and the factors ranked based on:

$$EET_i = \sum\nolimits_{j=1}^{r} EE_i^j, \tag{7}$$

where r is the sample size for each factor, usually of the order of 10, for an overall cost of $r(k{+}1){<<}Ntot$. EET_i is a useful measure as it is efficient numerically and it is very good for factor fixing: indeed, it is a good proxy for S_{Ti}. Moreover, EET is rather resilient against type II errors, i.e. if a factor is seen as non-influential by EET, it is unlikely to be seen as influential by any other measure (see Saltelli et al., 2004).

4 DBM rainfall-flow model for Hodder catchment

The dataset available in this example contains hourly data for rainfall, evaporation and flow observations at Hodder Place, North West England, 1991. The hourly flows are measured at Hodder Place gauging station on the River Hodder in North West England. The River Hodder has a catchment area of 261 km^2 and it forms part of the larger River Ribble catchment area of 456 km^2. The Hodder rises in the Bowland Fells and drains a thinly populated, entirely rural catchment and, together with the River Calder, it joins the River Ribble just downstream of a flow gauging station at Henthorn. There is a reservoir in the catchment which alters the water regime. During the summer months water abstractions from the reservoir strongly influence the flow. Only the rainfall and flow data are used in the DBM modelling. The January–March 1991 rainfall-flow data have been used for the identification and estimation (calibration). The second part of the data (October–December) has been used for the validation. The catchment is predominantly covered by grassland. During summer time, the flow is affected by abstractions from the reservoir situated in the catchment. Therefore, for simplicity and to make the model comparisons more transparent, only Autumn–Winter months are used for the modelling.

Following the DBM modelling procedure outlined in Sect. 2, the first "identification" stage of the modelling confirms previous research and suggests that the only significant SDP nonlinearity is associated with the rainfall input. This "effective rainfall" nonlinearity is identified from the data in the following form:

$$ue_t = f(y_t) \times r_t$$

where ue_t denotes effective rainfall, r_t denotes rainfall and $f(y_t)$ denotes an estimated non-parametric (graphical) SDP function in which, as pointed out in Sect. 2, the flow y_t is acting as a surrogate for the soil moisture content in the catchment (catchment storage). This non-parametric estimate is plotted as the dash-dot line in Fig. 1, with the standard error bounds shown dotted. The shape of the graphical SDP function suggests that it can be parameterised as either a power law or an exponential growth relationship. However, in accordance with previous research (see above references), the more parsimonious power law relationship $f(y_t)=y_t^{\beta}$ is selected for the subsequent "parametric estimation" stage of the DBM modelling.

The best identified linear STF model between the effective rainfall $ue_t=r_t y_t^{\beta}$ and flow is denoted by the usual triad, as [2 2 0]; i.e. a TF model with 2nd order denominator, 2nd order numerator and no advective time delay. This results in the following, final DBM model between the rainfall r_t and flow y_t:

$$\begin{aligned} ue_t &= r_t \times y_t^{0.38} \\ y_t &= \left[\frac{0.0716}{1-0.9788z^{-1}} + \frac{1.0299}{1-0.8645z^{-1}}\right] ue_t + \xi_t \quad \mathrm{var}(\xi_t) = 0.124\hat{y}_t^2 \end{aligned} \tag{8}$$

where ξ_t represents the estimated noise on the relationship, which is modelled as a heteroscedastic process with variance proportional to the square of the estimate of flow, $\hat{y}_t^2$.

In order to reveal the most likely mechanistic interpretation of the model, the [2 2 0] STF relationship in the model (8) is decomposed, as shown, into a parallel form that reveals the major identified flow pathways in the rainfall-flow system: namely, a "slow" component whose flow response is shown in the upper panel of Fig. 2, with a residence time (time constant) equal to about 46 h; and a "quick" flow component, with a residence time of about 7 h, whose flow response is shown in the lower panel of Fig. 2. The partition of flow between these quick and the slow components, respectively, is identified from the decomposed model parameters as 0.63 to 0.37. In other words, 63% of the total flow appears to arrive quickly from surface and near surface processes; while 37% appears to be due to the displacement of groundwater into the river channel (base flow). This model has a Coefficient of Determination, R_T^2=0.924 (i.e. 92.4% of the measured flow variance is explained by the simulated model output), which is a substantial improvement on the linear model (R_T^2=0.84), even though it involves the addition of only the single power law parameter.

Finally, Fig. 3 compares the output of the DBM model (8) with the measured flow: the shaded area shows the double SE bands, which are equivalent to approximately 95% confidence bands under the assumption of Gaussian errors. Note that we have not carried out any Monte Carlo Simulation (MCS) analysis in connection with the model (8), in order to emphasise that this is not essential in DBM modelling. Without such analysis, the DBM modelling is very computationally efficient, normally involving only a very short period of time. This contrasts with the TOPMODEL analysis, considered in the next section, where Monte Carlo Simulation analysis is an essential aspect of the GLUE and uncertainty analysis. Quite often, however, MCS analysis is used as an adjunct to DBM modelling, in order to evaluate various additional aspects of the model: e.g. the uncertainty associated with the "derived parameters": e.g., the residence times and partitioning percentages of the decomposed model (see e.g. Young, 1999); the effects of stochastic input noise; and the relative effects of the estimated parametric uncertainty (as quantified by the covariance matrix associated with the model parameters) and the stochastic inputs (as quantified by the variances of the input disturbances and measurement noise) on the flow output. In the latter connection, note that the output noise ξ_t is often modelled as an AutoRegressive (AR) or AutoRegressive, Moving Average (ARMA) process to account for autocorrelation in the model residuals. However, for simplicity, we have omitted this in the present example and restricted the noise modelling to the important heteroscedastic effects.

5 TOPMODEL calibration for Hodder catchment

The main aim of the TOPMODEL modelling (Beven and Kirkby, 1979), in this example, is the estimation of the uncertainty associated with the flow predictions for the Hodder catchment, using the same data as those used in the previous section. The choice of TOPMODEL is justified by its simple structure and its mechanistic interpretation: it has a total of 9 parameters but only 4 of these are considered in the associated sensitivity analysis described here (as well as uncertainties associated with the input and output observations). TOPMODEL bases its calculations of the spatial patterns of the hydrological response on the pattern of a topographic index for the catchment, as derived from a Digital Terrain Model (DTM). We have chosen the SIMULINK version of TOPMODEL, described in Romanowicz (1997), because of its clear, modular structure. This model has already been applied in similar Monte Carlo settings by Romanowicz

and Macdonald (2005). The saturated zone model is assumed to be nonlinear, with the outflow $Q_b(t)$ calculated as an exponential function of a catchment average soil moisture deficit S, i.e.,

$$\frac{dS(t)}{dt} = Q_b(t) - Q_v(t) \quad Q_b(t) = Q_0 \exp(-S(t)/m) \tag{9}$$

where $Q_0 = SK0 \exp(-\lambda)$ is the flow when $S(t)=0$; $Q_v(t)$ denotes the recharge to the saturated zone; $SK0$ is a soil transmissivity parameter; m is a parameter controlling the rate of decline in transmissivity with increasing soil moisture deficit; and λ is the mean value of the topographic index distribution in the catchment (Beven and Kirkby, 1979). Other parameters control the maximum storage available in the root zone (LRZ) and the rate of recharge to the saturated zone (KS). The calibration exercise includes the analysis of uncertainties associated with the input and output observations (i.e., rainfall and flow), the model structure and its parameters.

We assume here that rainfall observations are affected by a biased measurement, accounted for by a multiplicative noise ($input$ in Table 1) of ±20% that is applied to rainfall data at each Monte Carlo Simulation (MCS) run. Flow observation uncertainty is included in the choice of the error model structure, which is assumed to follow a fuzzy trapezoidal membership function. The breakpoints of the trapezoid are given by the array $[A, B, C, D] = [1-4\delta, 1-\delta, 1+\delta, 1+4\delta]$, where δ denotes the characteristic width of the trapezoid. Breakpoints for each time location are determined by multiplication e.g. for a breakpoint at time t, $A_t = \text{flow}_t^{\text{obs}}(1-4\delta)$, implying a heteroscedastic error structure. The width δ of the trapezoidal membership function is assumed to be uncertain as well, in a range between [0.1% and 20%]. This will allow the sensitivity of the calibration to the assumptions about the error magnitude to be evaluated.

Parameter uncertainty is taken care of in the choice of the parameter ranges for the parameters *SK0*, *m*, *LRZ*, *KS*, as required for the MCS analysis. The influence of the model structure uncertainty may be accounted for *via* a random sample from different model structures (e.g., TOPMODEL with and without dynamic contributing areas). In this paper, we restrict the analysis to parametric and observational uncertainty.

Note that, to some extent, the specification of the above input uncertainty information, as required for the MCS analysis, is subjective and will depend upon the experience and prior knowledge of the analysis. Consequently, these specifications may well be adjusted following the investigation of the initial MCS results, as discussed later.

In this example, the MCS analysis involves 1024 runs of TOPMODEL, sampling parameters from the ranges specified in Table 1 using quasi-random LPTAU sequences (Sobol' et al., 1992). In the GLUE approach, the predictive probability of the flow takes the form:

$$P(\hat{y}_t < y|z) = \sum_{\{i: y_t < y|\vartheta_i, z\}} f(\vartheta_i|z) \tag{10}$$

where $f(\cdot)$ denotes the likelihood weight of parameter sets $\Omega = \{\vartheta_1, \ldots, \vartheta_n\}$, in which n is the number of MC samples, conditioned on the available observations z.

The next step is to investigate the weights that account for both prediction and parameter structure related errors. The trapezoidal membership function, at each discrete time t and each element ϑ_i, $i=1, \ldots n$ of the MC sample, takes the values:

$$\begin{array}{lll} f_t(\vartheta_i|z_t) = 1 & if & B_t < y_t < C_t \\ f_t(\vartheta_i|z_t) = \frac{y_t - A_t}{B_t - A_t} & if & A_t < y_t < B_t \\ f_t(\vartheta_i|z_t) = \frac{y_t - D_t}{C_t - D_t} & if & C_t < y_t < D_t \\ f_t(\vartheta_i|z_t) = 0 & if & y_t < A_t, y_t > D_t \end{array} \tag{11}$$

where $A_t = z_t(1-4\delta_i)$, $B_t = z_t(1-\delta_i)$, $C_t = z_t(1+\delta_i)$, $D_t = z_t(1+4\delta_i)$. The likelihood weight for each parameter set is, therefore:

$$f(\vartheta_i|z) \sim \sum_{t=1}^{T} f_t(\vartheta_i|z_t) \tag{12}$$

In the present case study, the confidence bounds resulting from the GLUE methodology will depend on the width δ of the trapezoidal membership function. This function

gives zero weight for the simulations falling outside the range $z_t(1 \pm 4\delta)$ at any time (i.e. outside a relative error bound of 400δ%); while it will be maximum for those falling within the range $z_t(1 \pm \delta)$ at any time (i.e. within a relative error bound of 100δ%). The values of δ used in the MCS (Table 1) range from a very strict rejection criterion (simulations are rejected if the relative error at any time is larger than ±0.4%) to a very mild one (simulations are given the maximum weight when the relative error at any time is smaller than ±20%, while they are rejected only when they fall outside the ±80% bound). This initially wide range will undergo a revision in the second part of this example, based on conditioning to the observations and on sensitivity analysis results. However, as discussed later, δ is not the only parameter strongly driving the uncertainty bounds, with m playing an important role as well.

Figure 5 shows the calibration result for the flow predictions of TOPMODEL. The predictions describe 90.5% of the observed flow variations, a little less than that obtained using the DBM approach (92.3%: see Fig. 3). On the validation data (Fig. 8), the model predictions describe 74.7% of the flow variations, a similar result to DBM (75%: see Fig. 4). On the other hand, looking at the results in more detail, two major differences appear:

1. the 95% uncertainty bound of the calibrated TOPMODEL at the peaks of flow is about twice the bound of the DBM model;

2. the partition of the output flow of the TOPMODEL into two components, interpreted as "surface runoff" and "saturated and groundwater" flow, differs from the DBM's partitioning (Figs. 6–7). The "saturated and groundwater" flow accounts almost entirely for the whole flow (95%), leaving only a 5% to a quick component, which merely adjusts the peaks. Moreover, the dynamical pattern of major TOPMODEL component cannot be related to any particular flow pathway. In contrast, the DBM partition is well defined (and perhaps more informative, although this is likely to depend on the view of the hydrologist), attributing about 2/3 of the flow to the quick (surface) component. Note also that the DBM model predictions of the

peak flows are, in general, rather better than those of TOPMODEL.

5.1 Sensitivity analysis: is it possible to reduce volatility of predictions?

As discussed in Sect. 3, sensitivity analysis can play a key role in model calibration. In particular, to reduce volatility of predictions, it is clear that we have to refer to the Factors Prioritisation setting from which one can identify strategies for the reduction of the uncertainty predictions of the TOPMODEL, acting on the parameters having the highest main effects. To do so, we perform the sensitivity analysis for the likelihood weights (Ratto et al., 2001). In Fig. 9 we can see the main effect indices, computed on the same 1024 runs used for calibration and applying the State Dependent Regression method by Ratto et al. (2004, 2006), as well as the Elementary Effect Test (EET), computed using only an additional 168 model runs. The main effects tell us that the uncertainty in the width of the trapezoidal membership function dominates, followed by the parameter m, which is the only structural parameter having a non-zero effect. Input rainfall noise also plays a significant role. Looking at the EET indices, which are cheap proxies of the total effects as discussed in Sect. 3, we can see that, among the other structural parameters, only *KS* plays a significant role through interactions, while *SKO* and *LRZ* are not influential as far as the flow predictions are concerned (Factors Fixing setting). A detailed sensitivity analysis of each flow component suggests that *KS* acts through the so-called "surface" flow, mostly in the periods of low flow (details of the SA not shown here). In the periods of high flow, on the other hand, the model output displays a large sensitivity to small values of m: for $m<0.005$, the peaks of "surface" flows can reach extremely large values with respect to the average flows predicted. This suggests a way of restricting the support of m, in order to reduce the prediction uncertainty.

In addition to the above, let us consider the scatter plots of the likelihood weights versus the input factors (Fig. 10). Looking at the plot for δ, we can draw the following conclusions for the prior bound of δ[0.001–0.2]

1. the lower bound of the range seems to be rejected by the data: in fact, for values smaller than about 0.01, the width of the trapezoid is too narrow, in such a way that most of the model runs falling in this region are rejected as being unlikely;

2. the upper bound seems to be too wide: already at 0.1, the sum of the likelihood weights is allowed to reach values around 1000. For higher values, the upper bound of the likelihood weights almost reaches a plateau; while the lower bound rises in such a way that, at 0.2, the likelihood weights are never smaller than 400. This implies that, whatever the parameterisation of the model, the predictions always fall within the trapezoid. This could be the reason for the uncertainty bounds appearing too wide.

Based on these considerations, we conduct a second calibration step, by shrinking the support for δ to the range [0.01, 0.1] and the support for m to [0.005, 0.03] (Monte Carlo filtering). Such modifications are quite in order since, as pointed out previously, the specification of the original input uncertainties is, to some extent, subjective. Moreover, the revision of δ (measurement error) from an initially wide interval is one possible way to conduct a calibration of the measurement error; another one being to pre-calibrate δ based on DBM confidence bounds, which provides a statistically reliable estimate of the measurement error. Figures 11–12 show the updated calibration and validation exercises. The portion of flow variance now explained by the new calibrated model is 89.1%, only slightly less than in the initial setting. The uncertainty bound is now comparable to the one obtained in the DBM analysis, i.e. we have managed to obtain less volatile model predictions. How significantly does this affect the fitting performance? In the present step, 9.9% of the data fall outside the 95% bound, whereas this was 4.7% for the previous case. This is exclusively due to a larger number of time periods where the model uncertainty bound is above the data, so the results are on the "safe" side.

The surface/subsurface partitioning is also hardly modified: surface runoff is now 4.4% of the overall flow. Cutting the high tails from the distribution of surface flows did not alter the partitioning too much, meaning that the high tails of Fig. 6 are not supported by the data. The validation step confirms the results: 76.6% of the data is now described by model predictions, i.e. a larger amount than before. The uncertainty bounds in this case are also similar to DBM, with the drawback of a larger portion of data below the bounds: 14.4% versus the 6.4% of the initial case.

In this second step, the effect of the different uncertainty sources is more balanced: m has now the same importance as δ; moreover, *KS* also acquires a small main effect (see Fig. 13). Balanced main effects are usually a sign of better identification. Lack of identification is still present for *SKO* and *LRZ*, however, which can have virtually any value in their prescribed prior ranges without significantly affecting the model performance in predicting outflow (over-parameterisation or equifinality). This is an example of what is called, in sensitivity analysis, "fit for purpose". We actually know that both of these parameters should have an effect on flow, as they show threshold-like behaviour (Romanowicz and Beven, 2006). However, this behaviour has no significant impact on the fit of observations for the current case study, implying the irrelevance of *SKO* and *LRZ* for the purposes of outflow predictions. Of course, there can be other data sets or other "behvavioural" classifications that may highlight the role of *SKO* and *LRZ*.

5.2 Sensitivity analysis: is other partitioning possible?

As pointed out previously, the interpretation of the partitioning and decomposed flows of rainfall-flow models is open to subjective judgement. Consequently, the 5% "surface runoff" of TOPMODEL may not be considered any less physically acceptable than the much greater percentage of quick flow suggested by the DBM model. For example, there are some physical reasons why we might expect fast subsurface runoff in the wet soils of the Hodder. However, the relatively objective nature of the DBM analysis and its association of the partitioning with the estimated steady state gains and residence times of the DBM model, is persuasive and, we believe, provides a very reasonable decomposition. Consequently, purely as an illustrative exercise to further demonstrate how sensitivity analysis can assist in model evaluation, it is instructive to use it to

investigate whether other partitions are possible within the uncertainty specifications of the TOPMODEL. This exercise is, in practice, dual to the one considered in the previous section.

Sensitivity analysis results show that the portion of surface flow can rise to 30–40% only for very small values of m, i.e. in the complementary support adopted to reduce the volatility of model predictions. We tried to calibrate the TOPMODEL by restricting m in the range [0.001–0.003]. The calibration failed, with 93% of MC runs having negative R_T^2 values and huge volatility of predictions. This considered, we filtered the MC sample, keeping only the 7% runs with positive R_T^2. This strategy resulted in R_T^2=52% (Fig. 14), with the contribution of surface flow in the range 10–20% (Figs. 15–16). Looking at Fig. 14, we can see that the confidence bounds are narrower than in the previous calibration, with a large part of flow under-predicted in periods of small flow; while the flow peaks tend to be over-predicted. On the other hand, considering Figs. 15–16, we can see that the dynamical features of the two flow components of the model show greater similarity: both of them present spikes in correspondence to rainfall, while the slow dynamic component is still present for the "subsurface" flow during rainfall periods. The root store is still important in smoothing the predictions, but the slow and fast mechanisms are closely related in TOPMODEL, and increasing the "surface" component, also makes the quick response of the "subsurface" flow larger.

Finally, the regionalised sensitivity analysis (Smirnov test) of the 7% of calibrated runs, compared to the initial 100% of runs, gave a significant effect for all input factors, except for *SKO*. Table 2 shows the restricted ranges identified by the filtering and calibration procedure. With these results, GSA is able to highlight all the available "resources" that the TOPMODEL can exploit in order to get a reasonable calibration when m is restricted to assure a more balanced partitioning between "surface" and "sub-surface" processes (if this is considered reasonable requirement: we have only carried out this analysis for illustrative purpose).

6 Discussion

We have applied the bottom-up, reductionist approach (physically based, semi-distributed rainfall-runoff model TOPMODEL) and the alternative top-down, DBM approach to rainfall-flow modelling in the Hodder catchment. The DBM modelling approach gives a well identified representation of the data, with a good explanation of the flow data in calibration and validation, as well as a clear mechanistic interpretation arising from the identified linear, second order STF between effective rainfall and output flow. The output flow of the DBM model can be further decomposed into a "slow" (37%) and a "quick" (63%) component, with a suggested physical interpretation that these are associated with groundwater and surface processes, respectively.

The results of the TOPMODEL calibration and validation are not too far from the DBM ones in terms of explaining the data. However, two issues have arisen in our initial sensitivity analysis (i.e. with the initially specified nominal parameters and uncertainty bounds) that reveal differences between the TOPMODEL and the DBM model: (i) the large volatility of TOPMODEL predictions, with confidence bounds twice that of the DBM model; and (ii) the behaviour of the two output flow components, with one flow component accounting for most of the output flow, while the other component has comparatively little impact on the model behaviour.

Since these differences exist, they have provided the justification for an illustrative exercise, in which the TOPMODEL properties have been analysed and the calibration revised in the light of the sensitivity analysis results. In particular, GSA helped in obtaining less volatile predictions, at the price of slightly worse fitting properties (the data are below confidence bounds for a larger number of time periods), by acting on the prior supports of the two most relevant parameters: m and δ. This also allowed us to show an example of using GSA for model simplification: *SKO* is clearly highlighted as irrelevant for calibration purposes and, hence, it can be fixed at a nominal value. GSA also highlighted the parameter ranges, that produce a more balanced partitioning between the two output flows (if this is considered desirable); and it played a role in

obtaining a reasonable calibration under such a restriction. However, while the partitioning is now closer to the DBM model results, it still does not clearly split in slow and quick dynamics, with the "subsurface" processes still displaying a mixture between the two dynamical modes. Moreover, the calibration under such conditions is much more difficult and explains a significantly smaller portion of the data. So it is impossible to bring the two models into reasonable agreement as regards the flow partitioning, which is not all that surprisng since the models have significant structural differences.

This last result shows that the TOPMODEL structure would have to undergo a more substantial change if the modelling of slow and fast response of the catchment were to be the aim of the modelling. However, TOPMODEL represents one particular hypothesis about the nature of the hydrological processes involved in the transfer of rainfall into river flow, namely the assumption of exponential dependence between the soil moisture storage in the catchment and outflow. And so the partitioning is a natural consequence of this hypothesis and may well be acceptable, as pointed out in the earlier discussion on this topic (see Sect. 5.2). Of course, TOPMODEL's main advantage lies in the ability to estimate the spatial distribution of soil moisture in the catchment.

Calibration results and sensitivity analysis of the TOPMODEL show that, for the Hodder catchment, the DBM model explains the data a little better and, in the view of the present authors, provides a more informative description of the data because of its clear decomposition of the flow into quick and slow components. Of course, one should expect that, most of the time, the statistically more efficient nature of the DBM top-down modelling methodology should normally provide a better explanation of the data. However, this will depend on the catchment characteristics and the ability of the DBM analysis to identify a good model within its assumed generic model structure (i.e. nonlinear differential equations or their discrete-time equivalents).

On the other hand, one might also expect that a mechanistic interpretation, such as that of TOPMODEL, arising from a process of hypothesis and deduction, should be helpful in providing a better description of the physical processes involved in the transference of rainfall to flow. If this is not happening to the satisfaction of the modeller, a model revision should be taken into account. In such an exercise, DBM modelling can be of assistance, since it can guide the modeller in revising the model specification. For example, the modeller might wish to identify such advection-dispersion mechanisms that have characteristic times similar to those estimated in the DBM model and base the model revision on this. In this connection, the effective rainfall nonlinearity that is a feature of the DBM approach, is represented by the concept of contributing areas in TOPMODEL. Moreover, TOPMODEL has nonlinear dynamics of a different type to those of the DBM model, hence its equivalence with DBM model cannot be established directly and such exercises in model modification may not be straightforward (although they can be much easier in the case of other rainfall-flow models, such as the IHACRES model of Jakeman et al., 1990, which has a clear similarity to the DBM model structure).

In conclusion, we have seen that GSA can be of assistance in improving the calibration of deterministic, bottom-up, hydrological models and making their properties more transparent. Although sensitivity analysis cannot solve the structural problems in models, it can only help in highlighting them. It is also clear that the computationally much less intensive DBM analysis provides a useful input into subsequent TOPMODEL sensitivity analysis and calibration, raising critical questions about the results obtained in the initial sensitivity analysis and leading to modifications in subsequent sensitivity and uncertainty analysis. In other words, the two modelling methodologies, one inductive and the other hypothetico-inductive, have good synergy; combining well to produce a complete modelling approach that has the kinds of checks-and-balances required in practical data-based modelling of rainfall-flow systems.

Such a combined top-down/bottom-up approach also produces two types of model that are suitable for different kinds of application. In particular, the DBM model provides a very quick and efficient analysis of the rainfall-flow data, revealing important characteristics of the catchment-scale response, such as the nature of the effective rainfall nonlinearity and the partitioning of the effective rainfall into different flow pathways. In this form, the DBM model can provide a vehicle for flow and flood forecasting (e.g. Young, 2002; Romanowicz et al., 2004, Romanowicz et al., 2006). However, its effec-

tive rainfall module needs to be modified into a more conceptual form (see e.g. Young, 2003) if it is to be used for simulation modelling applications. On the other hand, TOPMODEL, when suitably calibrated (and perhaps modified depending on the views of the modeller) in the light of the DBM and GSA results, immediately provides a simulation model with a variety of potential applications, in areas such as catchment management and planning.

References

Beck, M. B., Ravetz, J. R., Mulkey, L. A., and Barnwell, T. O.: On the Problem of Model Validation for Predictive Exposure Assessments, Stochastic Hydrology and Hydraulics, 11, 229–254, 1997.

Beven, K. J. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992.

Beven, K. J. and Kirkby, M.J.: A physically based variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43–69, 1979.

Hall, J. W., Tarantola, S., Bates, P. D., and Horritt, M. S.: Distributed sensitivity analysis of flood inundation model calibration, J. Hydraulic Eng.-Asce, 131, 117–126, 2005.

Hornberger G. M. and Spear R. C.: An approach to the preliminary analysis of environmental systems, J. Environ. Manage., 7, 7–18, 1981.

Jakeman, A. J., Littlewood, I. G., and Whitehead, P. G.: Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, J. Hydrol., 117, 275–300, 1990.

Li, G., Wang, S. W., and Rabitz, H.: Practical approaches to construct RS-HDMR component functions, J. Phys. Chem., 106, 8721–8733, 2002.

Li, G., Hu, J., Wang, S.-W., Georgopoulos, P. G., Schoendorf, J., and Rabitz, H.: Random Sampling-High Dimensional Model Representation (RS-HDMR) and Orthogonality of Its Different Order Component Functions, J. Phys. Chem. A, 110, 2474–2485, 2006.

Oakley, J. E. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach, J. R. Statist. Soc. B, 66, 751–769, 2004.

Pappenberger, F., Iorgulescu, I., and Beven, K. J.: Sensitivity Analysis based on regional splits and regression trees (SARS-RT), Environmental Modelling and Software, 21(7), 976–990, 2006a.

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J.-B., Pfister, L., and de Fraipont, P.: Influence of uncertain boundary conditions and model structure on flood inundation predictions, Adv. Water Resour., in press, available at http://dx.doi.org/10.1016/j.advwatres.2005.11.012 since 19 May 2006, 2006b.

Ratto, M., Tarantola, S., and Saltelli, A.: Sensitivity analysis in model calibration: GSA-GLUE approach, Computer Physics Communication, 136, 212–224, 2001.

Ratto, M., Tarantola, S., Saltelli, A., and Young, P. C.: Accelerated estimation of sensitivity indices using State Dependent Parameter models, Proc. 4th Int. Conf. on Sensitivity Analysis of Model Output (SAMO 2004), Santa Fe, New Mexico, March 8–11, edited by: Hanson, K. M. and Hemez, F. M., electronic proceedings, 61–70, 2004.

Ratto, M., Saltelli, A., Tarantola, S., and Young, P. C.: Improved and accelerated sensitivity analysis using State Dependent Parameter models, EUR Technical Report, Joint Research Centre of the European Commission, 2006.

Romanowicz, R. J.: A MATLAB implementation of TOPMODEL, Hydrol. Process. 11, 1115–1129, 1997.

Romanowicz, R. J. and Macdonald, R.: Modelling uncertainty and variability In environmental systems, Acta. Geophys. Pol., 53, 401–417, 2005.

Romanowicz, R., Beven, K. J., and Young, P. C.: Assessing the risk of flooding in real time, Proc. ACTIF Conference on Quantification, reduction and dissemination of uncertainty in flood forecasting, Delft, Netherlands, http://www.actif-ec.net/Workshop2/papers/ACTIF_S1_06.pdf; reviewed online publication, 2004.

Romanowicz, R. J. and Beven, K.: Comments on Generalised Likelihood Uncertainty Estimation, Reliability Engineering and System Safety, in press, available at http://dx.doi.org/10.1016/j.ress.2005.11.030 since 6 January 2006, 2006.

Romanowicz, R., Beven, K. J., and Young, P. C: Data assimilation and adaptive forecasting of water levels in the river Severn catchment, UK, Water Resour. Res., W06407, doi:10.1029/WR004373, 2006.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Sensitivity analysis for chemical models, Chemical Reviews, 105, 2811–2828, 2005.

Saltelli, A. and Tarantola, S.: On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal, J. Am. Stat. Assoc., 97(459), 702–709,

2002.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M.: Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, John Wiley and Sons, 2004.

Sobol', I. M.: Sensitivity estimates for nonlinear mathematical models, Matematicheskoe Modelirovanie, 2, 112–118, 1990 (in Russian), [Transl. Sensitivity analysis for non-linear mathematical models. Mathematical Modelling & Computational Experiment, 1, 407–414, 1993.]

Sobol' I. M., Turchaninov, V. I., Levitan, Y. L., and Shukhman, B. V.: Quasirandom sequence generators, Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, 1992.

Spear, R. C., Grieb, T. M., and Shang, N.: Factor Uncertainty and Interaction in Complex Environmental Models, Water Resour. Res., 30, 3159–3169, 1994.

Young, P. C.: Recursive Approaches to Time-Series Analysis, Bull. Inst. Maths Appl., 10, 209–224, 1974.

Young, P. C.: A general theory of modelling for badly defined dynamic systems, in: Modeling, Identification and Control in Environmental Systems, edited by: Vansteenkiste, G. C., 103–135, North Holland: Amsterdam, 1978.

Young, P. C.: "Data-based mechanistic modelling of environmental, ecological, economic and engineering systems", Environ. Modelling and Software, 13, 105–122, 1998.

Young, P. C.: Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis. Computer Physics Communication, 117, 113–129, 1999.

Young, P. C.: Advances in real-time flood forecasting, Philosophical Trans. Royal Society, Phys. Eng. Sci., 360, 1433–1450, 2002.

Young, P. C.: Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale, Hydrol. Process., 17, 2195–2217, 2003.

Young, P. C. and Beven, K. J.: Data-based mechanistic modelling of rainfall-flow nonlinearity, Environmetrics, 5, 335–363, 1994.

Young, P. C., Spear, R. C., and Hornberger, G. M.: Modeling Badly Defined Systems: Some Further Thoughts, Proceedings SIMSIG Conference , SIMSIG: Canberra, 1978.

Young, P. C., Parkinson S. D., and Lees M.: Simplicity out of complexity: Occam's razor revisited, J. Appl. Stat., 23, 165–210, 1996.

Table 1. Parameter distributions applied in MC calibration of TOPMODEL.

Parameters	Distribution	Min.	Max.	Mean	Std
input	uniform	0.8	1.2	1	0.116
SKO	uniform	10 000	40 000	25 000	8600
m	uniform	0.001	0.03	0.0157	0.008
LRZ	uniform	1×10^{-4}	0.04	0.02	0.012
KS	log-uniform	1×10^{-7}	0.1	0.007	0.017
δ	uniform	0.001	0.2	0.1	0.575

Table 2. Restricted ranges after filtering the MC sample to assure $R_T^2>0$ when the support of m is restricted to assure more balanced partitioning between surface and sub-surface processes.

Parameters	New ranges
Input	[0.8–0.95]
m	[0.0025–0.003]
LRZ	[0.015–0.04]
KS	[1.e^{-4}–1.e^{-3}]
δ	[0.1–0.2]

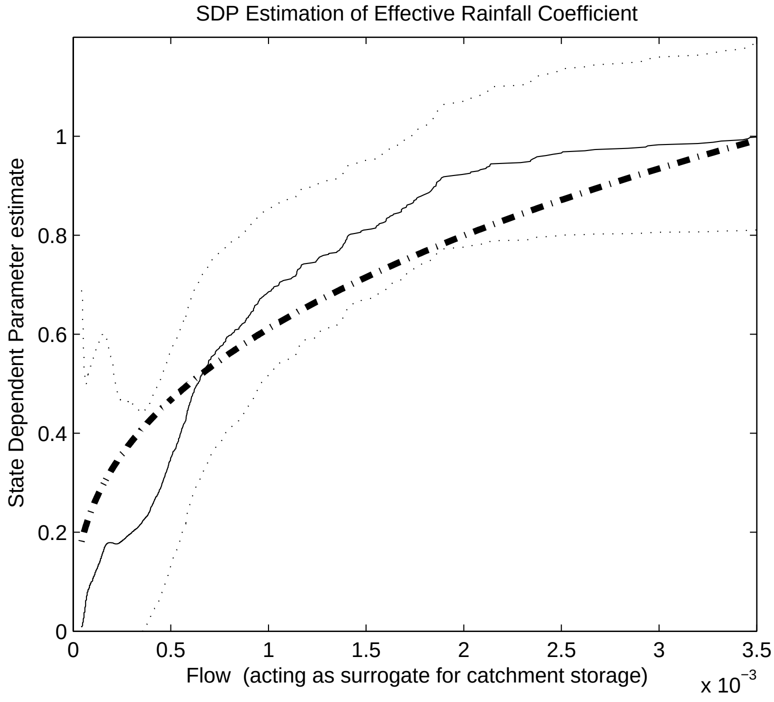

Fig. 1. This graph shows the estimated non-parametric SDP relationship (dash-dot line) between measured flow and the effective rainfall parameter, where the flow is acting as a surrogate for catchment soil moisture. The 95% confidence bands are shown by the dotted lines and the full line is a parametric estimate of the SDP relationship obtained in the later estimation stage of the modelling using a power law approximation.

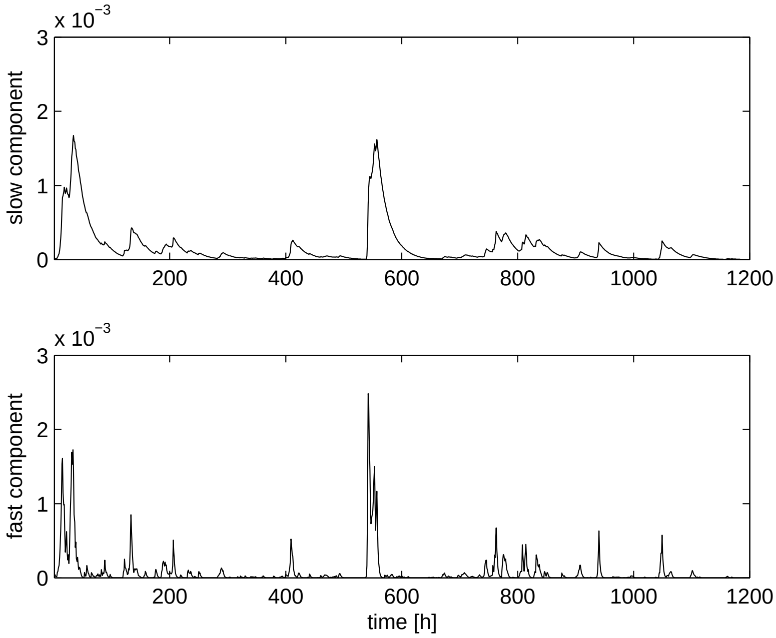

Fig. 2. The estimated slow (upper panel) and quick (lower panel) flow components of the DBM model output. The most likely interpretation of these component flows is that the slow component represents the groundwater (baseflow) effects and the quick component represents the surface and near-surface process effects.

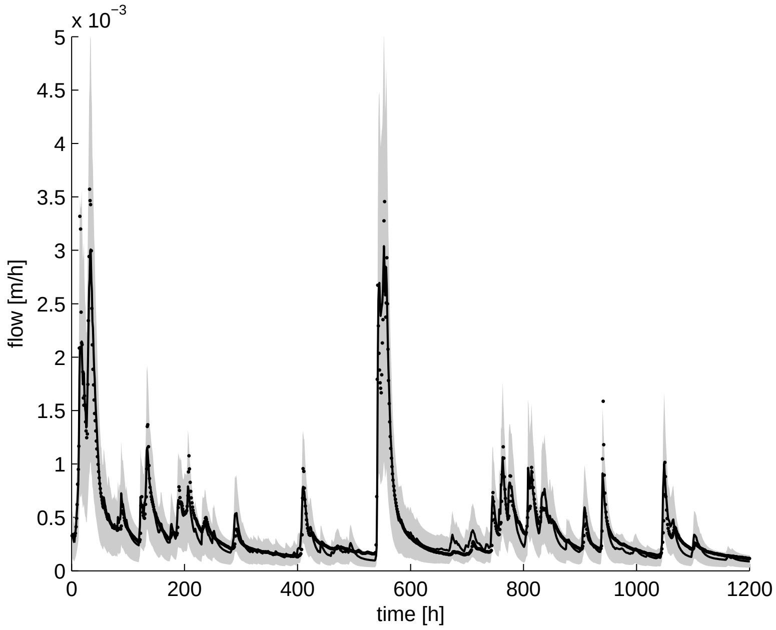

Fig. 3. Calibration of the data: 94% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

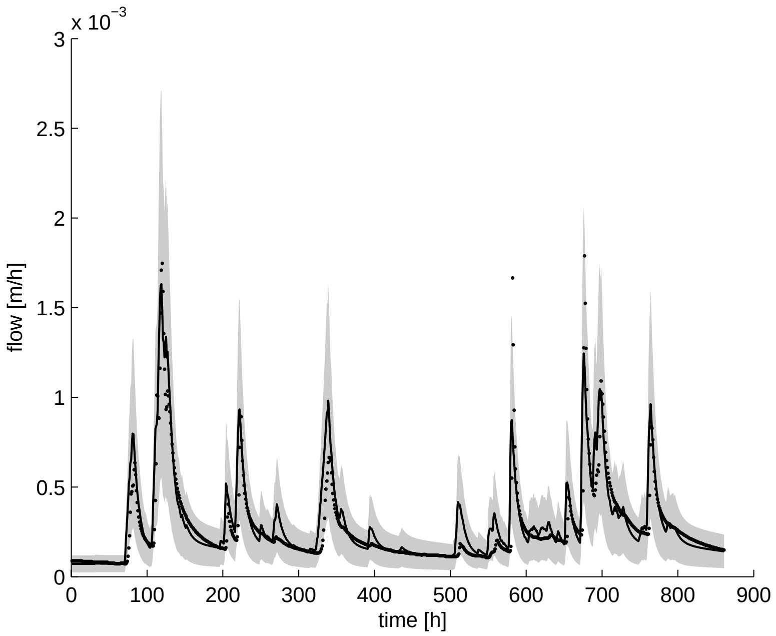

Fig. 4. Validation of the model on October–November 1991 period, 75% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

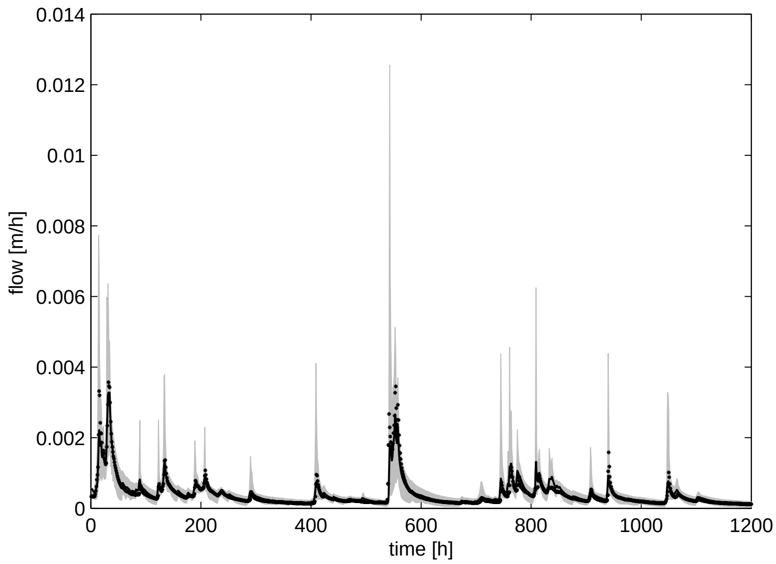

Fig. 5. Calibration of the TOPMODEL: 90.5% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

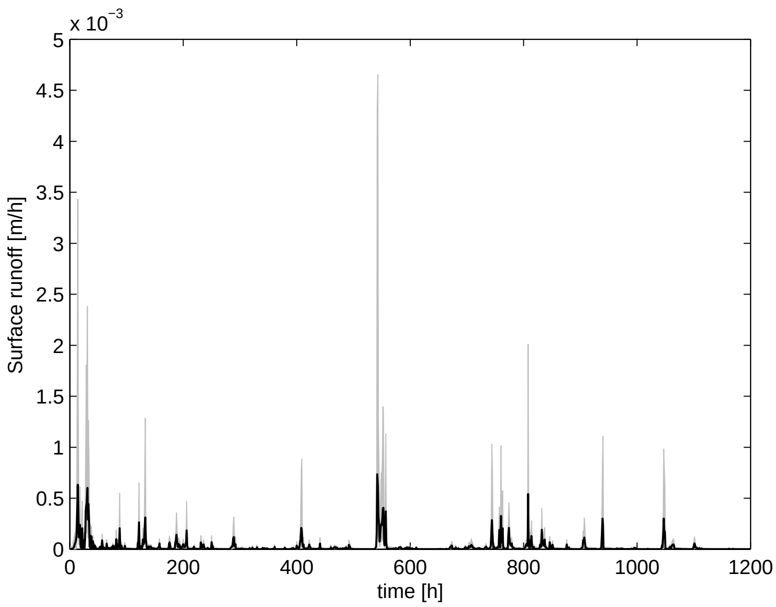

Fig. 6. Uncertainty estimation of the surface runoff of the TOPMODEL (5.2% of the overall flow); simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

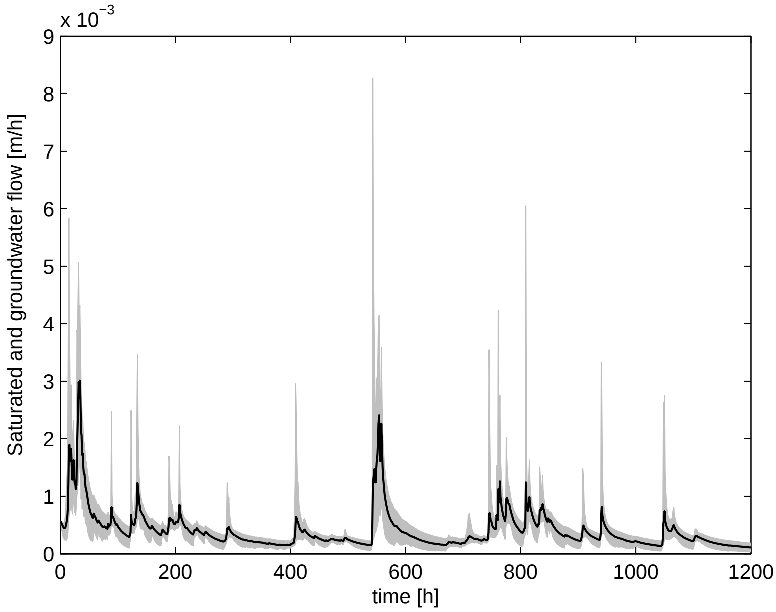

Fig. 7. Uncertainty estimation of the saturated and groundwater flow of the TOPMODEL (94.8% of overall flow); simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

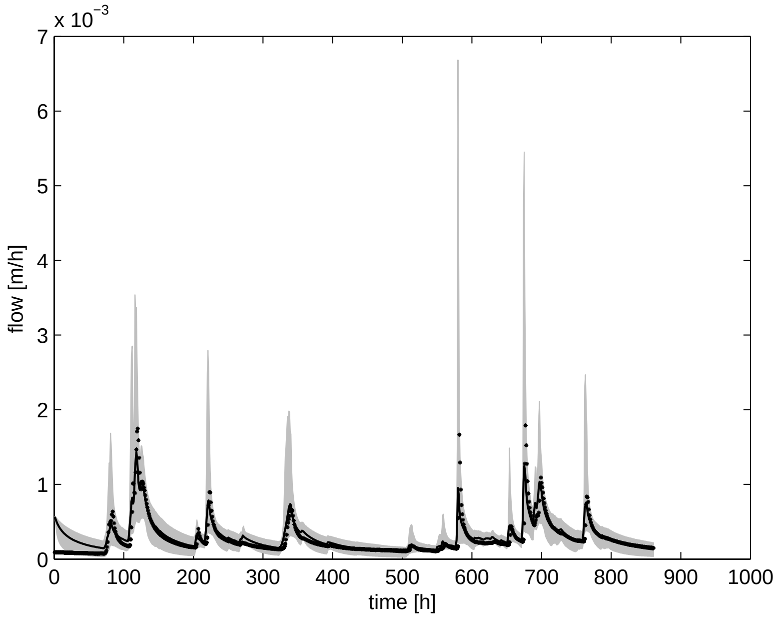

Fig. 8. Validation of the TOPMODEL: 74.7% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

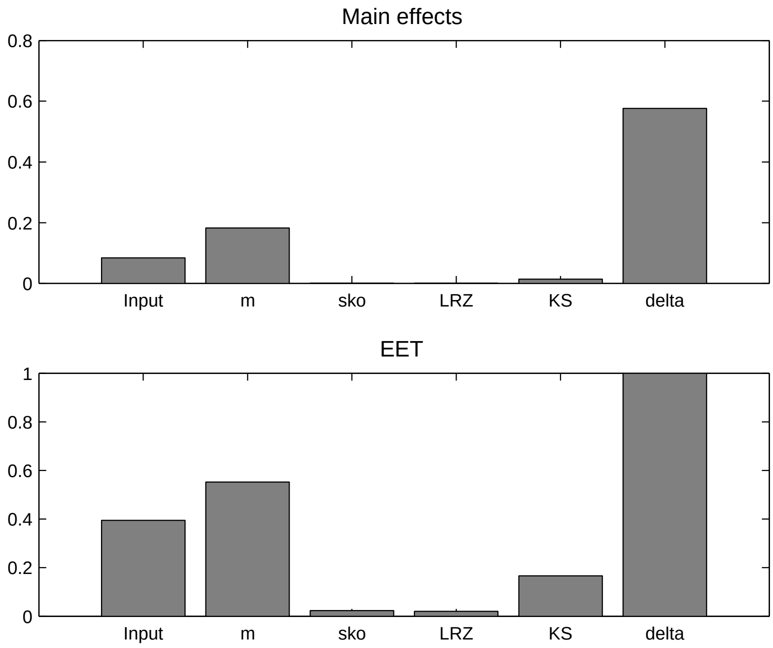

Fig. 9. Sensitivity indices of the likelihood weights versus model parameters. Upper panel: main effects. Lower panel: EET indices

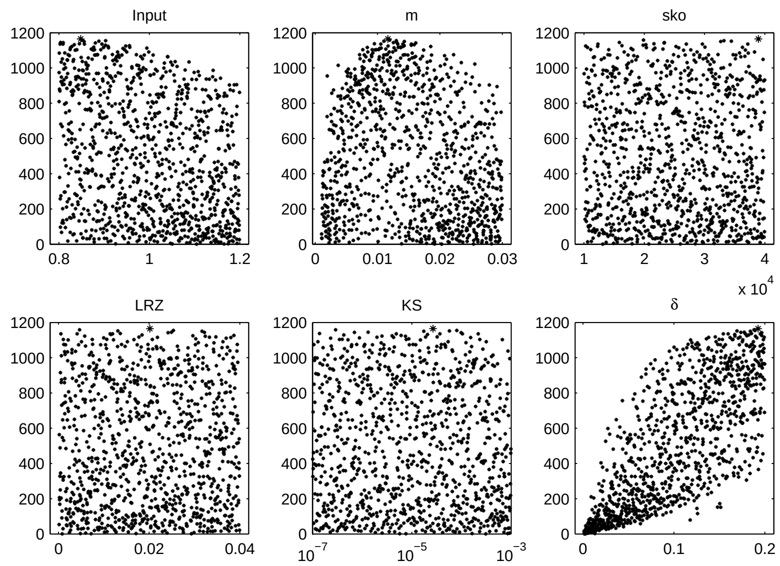

Fig. 10. Scatter plots of the sample of likelihood weights versus uncertain input factors.

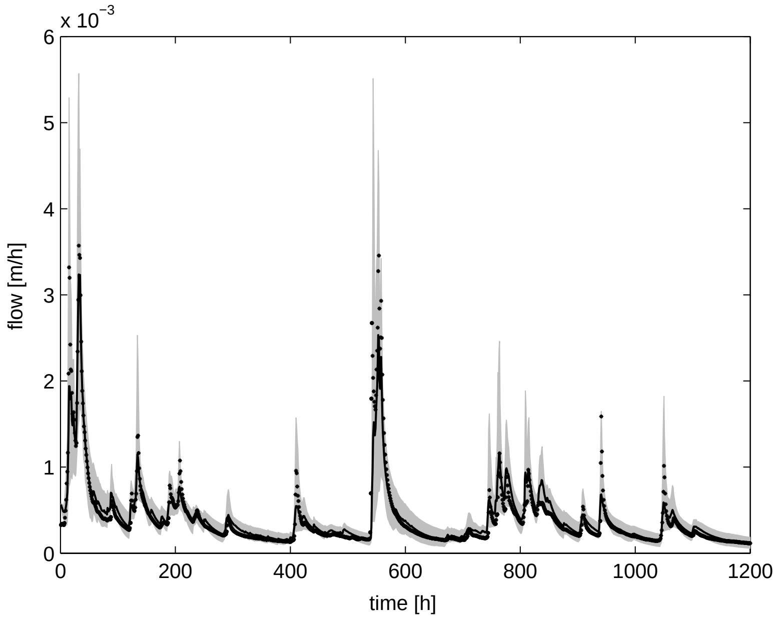

Fig. 11. Revised calibration of the TOPMODEL, to reduce volatiliy of predictions: 89.1% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands. "Surface" flow partition: 4.4% of the overall flow.

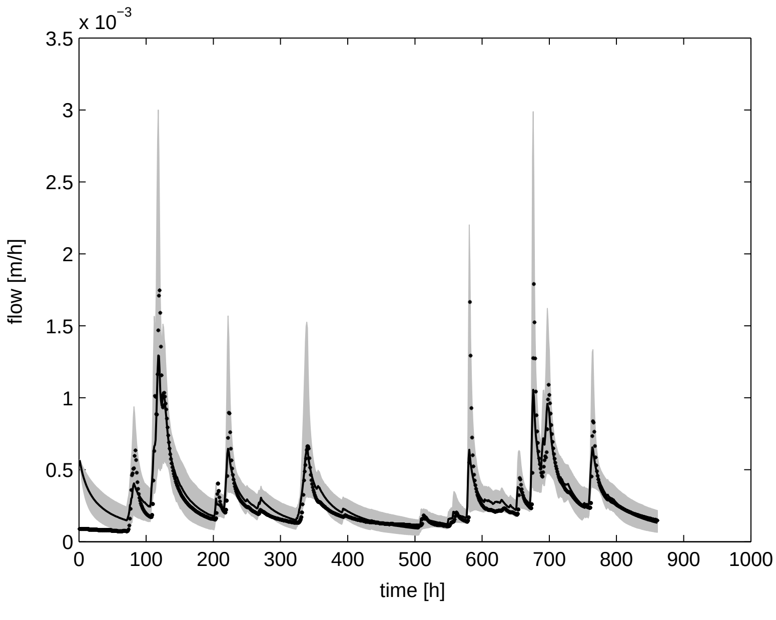

Fig. 12. Revised validation of the TOPMODEL, to reduce volatility of predictions: 76.6% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands. Surface flow partition: 3.5% of the overall flow.

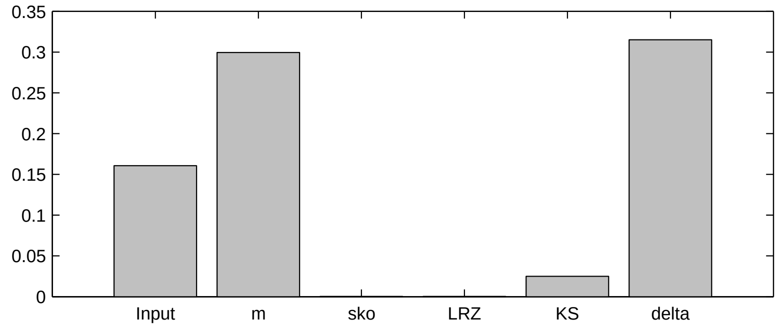

Fig. 13. Sensitivity analysis of the likelihood weights versus input factors, in the revised calibration step, to reduce volatility of predictions.

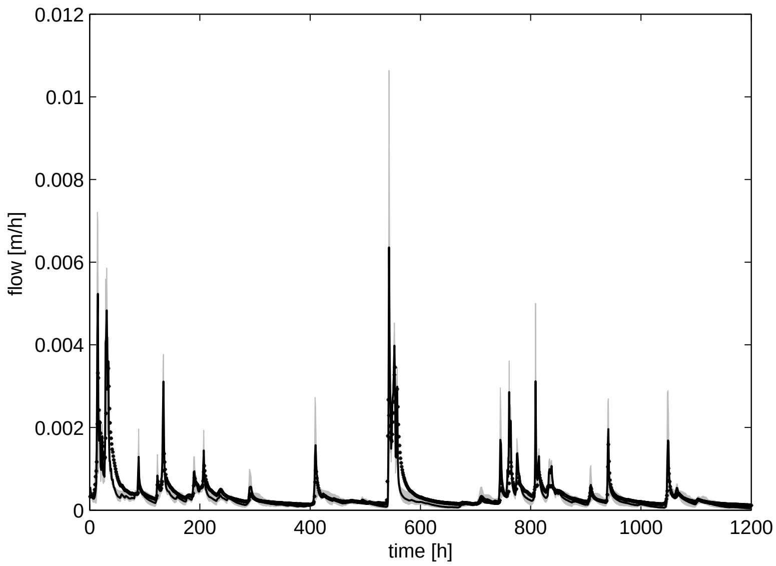

Fig. 14. Calibration of the TOPMODEL (R_T^2=52%), when the support of m is restricted to assure more balanced partitioning between surface and sub-surface processes and the MC sample is filtered to assure positive R_T^2.

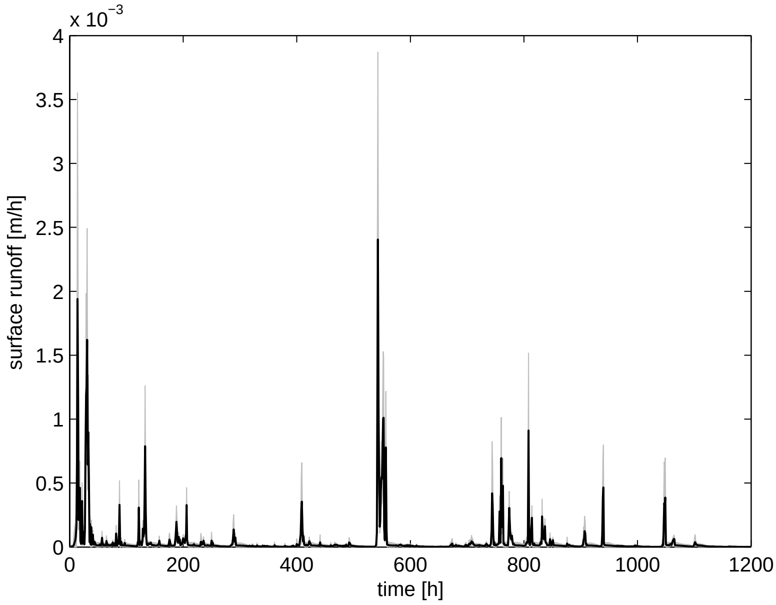

Fig. 15. Surface runoff in the calibrated TOPMODEL (10–20% of the overall flow), when the support of m is restricted to assure more balanced partitioning between surface and sub-surface processes and the MC sample is filtered to assure positive R_T^2.

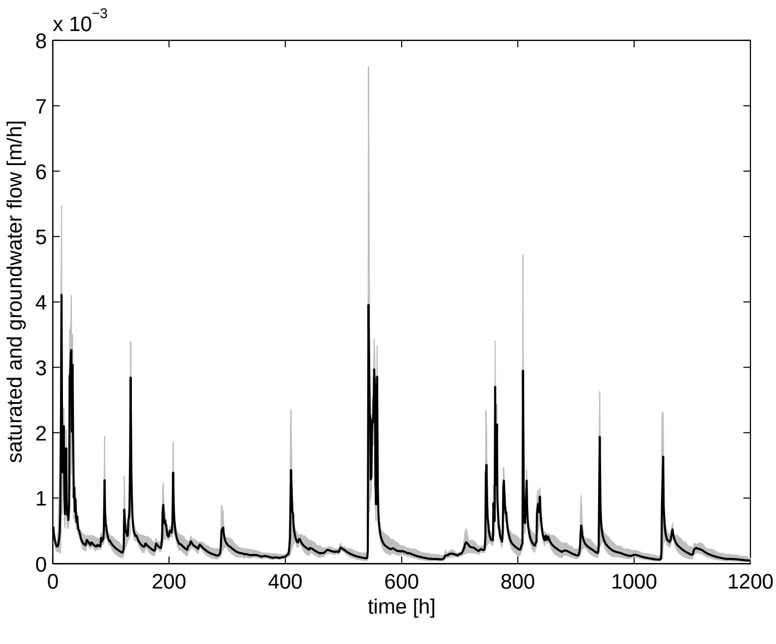

Fig. 16. Saturated and groundwater flow in the calibrated TOPMODEL (80–90% of the overall flow), when the support of m is restricted to assure more balanced partitioning between surface and sub-surface processes and the MC sample is filtered to assure positive R_T^2.