

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

M. Ratto¹, P. C. Young^{2,3}, R. Romanowicz², F. Pappenberger^{2,4}, A. Saltelli¹, and A. Pagano¹

¹Joint Research Centre, European Commission, Ispra, Italy

²Centre for Research on Environmental Systems and Statistic, Lancaster University, Lancaster, UK

³Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia

⁴European Centre for Medium-Range Weather Forecasts, Reading, UK

Received: 7 April 2006 – Published in Hydrol. Earth Syst. Sci. Discuss.: 25 September 2006

Revised: 14 March 2007 – Accepted: 29 March 2007 – Published: 3 May 2007

Abstract. In this paper, we discuss a joint approach to calibration and uncertainty estimation for hydrologic systems that combines a top-down, data-based mechanistic (DBM) modelling methodology; and a bottom-up, reductionist modelling methodology. The combined approach is applied to the modelling of the River Hodder catchment in North-West England. The top-down DBM model provides a well identified, statistically sound yet physically meaningful description 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. These characteristics are defined inductively from the data without prior assumptions about the model structure, other than it is within the generic class of nonlinear differential-delay equations. The bottom-up modelling is developed using the TOPMODEL, whose structure is assumed a priori and is evaluated by global sensitivity analysis (GSA) in order to specify the most sensitive and important parameters. The subsequent exercises in calibration and validation, performed with Generalized Likelihood Uncertainty Estimation (GLUE), are carried out in the light of the GSA and DBM analyses. This allows for the *pre-calibration* of the the priors used for GLUE, in order to eliminate dynamical features of the TOPMODEL that have little effect on the model output and would be rejected at the structure identification phase of the DBM modelling analysis. In this way, the elements of meaningful subjectivity in the GLUE approach, which allow the modeler to interact in the modelling process by constraining the model to have a specific form prior to calibration, are combined with other more objective, data-based benchmarks for the final uncertainty estimation. GSA plays a major role in building a bridge between the hypothetico-deductive (bottom-up) and inductive (top-down) approaches

and helps to improve the calibration of mechanistic hydrological models, making their properties more transparent. It also helps to highlight possible mis-specification problems, if these are identified. The results of the exercise show that the two modelling methodologies have good synergy; combining well to produce a complete joint 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 considered in the paper is developed specifically as a vehicle for flow and flood forecasting (although the generality of DBM modelling means that a simulation version of the model could be developed if required); 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 hydrology and hydraulic modelling (see discussion in Pappenberger and Beven (2006) and Pappenberger et al. (2006a)). 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).

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

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. Global sensitivity analysis (GSA) 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. (2006b, c, 2007) 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. Tang et al. (2006) also provide an interesting review of GSA methods within the context of hydrologic modelling.

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 GSA 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 approach is inductive: the models are derived directly from the data based on a generic class of dynamic models (here, differential-delay equations or their discrete-time equivalents). In particular, the DBM approach is based on the statistical identification and estimation of 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 a unified point of view, combining the bottom-up, reductionist TOPMODEL with the top-down, DBM approach, which plays the role of an “independent” benchmark. Note that, in this respect, we do not pursue any competitive comparison, to judge or prove one type of model to be “better” than the other. Our choice of DBM modelling as the right candidate for benchmarking, is based on its capability of providing a meaningful description of observations, in addition to its simplicity, ease of interpretation and rapidity of implementation. We will also highlight the role of GSA in building a bridge between the two approaches by assisting in the improved calibration of reductionist models and highlighting both their weaknesses and strengths. Initially, however, we will compare and contrast the two modelling approaches used in the paper.

2 The Different Modelling Approaches

Data-based Mechanistic (DBM) modelling is a methodological *approach* to model synthesis from time-series data and it is not restricted to rainfall-flow modelling; nor, indeed, to the specific rainfall-flow model identified in the present paper. Its application to 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 and the prior references therein).

DBM models are 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. Essentially, DBM modelling involves four stages that exploit advanced methods of time series analysis: data-based identification of the model structure based on the assumed generic model form; 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.

In the case of the *specific* DBM model used in the present paper, 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 initially in a non-parametric, 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 normally 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, constant coefficient, dynamic 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. This STF model is the discrete-time equivalent of a continuous-time differential-delay equation and it could be identified in this form, if required (e.g. Young, 2005).

In the estimation stage of the DBM modelling, the non-parametric SDP nonlinearity is parameterised in a parsimonious (parametrically efficient) manner. Normally, this parameterisation is in terms of a power law or an exponential function, although this depends upon the results of the non-parametric identification, as discussed in the later example. The approach used in this paper has been developed in the previously published DBM modeling of rainfall flow processes. Here, the STF model and the power law parameters, are estimated simultaneously using a special, nonlinear least squares optimisation 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 optimisation 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 characterises 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 normally consist of a “quick” flow component, with a short residence time (time constant) that seems to account for the surface and near surface processes; and a “slow” component with a long residence time, that often can be associated with the replenishment of the groundwater storage and the consequent displacement of groundwater into the river channel. Of course, it must be emphasised that these components are unobserved (not measured directly) and so their identification is done via 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.

The situation is quite different in the case of TOPMODEL (Beven and Kirkby, 1979), which has the specific model structure that is assumed a priori and so constrains the estimation of the component flows. Consequently, the flow decompositions of the DBM model and TOPMODEL are quite different. Moreover, in both cases, the interpretation of the decomposed flow components is subjective, depending not only on the model form but also on how the modeller views the unobserved flow components within the hydrological context.

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 for TOPMODEL and most other 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.

It must be emphasized again that DBM modelling is a statistical *approach* to model building and any specific DBM model will depend upon the objectives of the modelling exercise. For instance, the DBM model for the River Hodder, as considered in this paper, is intended primarily for use in data assimilation and flow forecasting (see Young, 1998, 2003), rather than simulation (see later). However, it does define the dominant dynamic characteristics of the catchment, as inferred from the data alone, and so these can be compared with those inferred by the alternative TOPMODEL. The latter model is obtained from the rainfall-flow data by a process

of hypothesis and deduction (i.e. the hypothetico-deductive approach of Karl Popper), with the model evaluation assisted by the application of GSA and with the uncertainty estimation provided by the DBM analysis used as a benchmark within the GLUE calibration. In order to help clarify this combined approach, the calibration and validation analyses are carried out using the same data sets.

In combining and comparing the models and their associated modelling methodologies (see e.g. Young, 2002), the inductive and relatively objective procedure behind the DBM modelling implies that the sensitivity analysis in DBM 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. In particular, the process of model structure identification ensures that the model is identifiable and minimal: i.e. it is the simplest model, within the generic model class, that is able to adequately explain the data. As such, the identified DBM model provides a parametrically efficient (parsimonious) description of the data in which all the parameters, by definition, are important in sensitivity terms, thus negating the need for GSA.

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 GSA, which is used to specify the most sensitive and important parameters. The subsequent exercises in calibration and validation, as performed here by GLUE, are carried out in the light of this GSA and the DBM results.

It is important to note that the DBM model developed in the present paper and TOPMODEL are quite different representations of the rainfall flow dynamic behaviour. Although unusual in some ways, the DBM model quite closely resembles conventional, conceptual hydrological models that use a “bucket” analogy. The idea of an “effective rainfall” input is common in such models: indeed, the catchment storage model used for effective rainfall generation in the DBM simulation model of Young (2003) can be related to the conventional Antecedent Precipitation Index (API) approach. Moreover, the impulse response of its STF component is directly interpretable in terms of the unit hydrograph; while the parallel decomposition of this STF reveals first order dynamic elements that can be interpreted in conventional “bucket” and mass conservation terms.

TOPMODEL has quite different *nonlinear* hydrograph dynamics, being formulated as the nonlinear differential equation where the outflow is calculated as an exponential function of the changing catchment average soil moisture deficit (see later, Eq. (9)). As a result, TOPMODEL has nonlinear recession behaviour that is dependent on the soil moisture deficit and so will change as the deficit changes. In contrast to this, the recession part of the underlying hydrograph, that

characterises the linear STF component of the DBM model, is of a conventional form (in the case of a second order STF, a combination of two decaying exponential curves) because the only significant nonlinearity identified from the rainfall-flow data by the DBM analysis is connected with the effective rainfall transformation at the input to the model. Of course, another of TOPMODEL’s differences, in relation to the specific DBM model used in the present paper, lies in its ability to exploit Digital Terrain Map (DTM) data and to estimate the spatial distribution of soil moisture in the catchment.

Finally, it seems to us that the use of the elements of relative objectivity inherent to the DBM approach provides a useful contribution within the GLUE context. GLUE introduces elements of meaningful subjectivity, so allowing the modeller to interact in the modelling process by constraining the model to have a specific form prior to calibration. This is of course, both a strength and a weakness, and it is achieved by relaxing some elements of full Bayesian estimation. One qualifying element of our paper is that we provide an objective benchmark for uncertainty prediction, which, in conjunction with GSA, allows us to *pre-calibrate priors* for the GLUE calibration of TOPMODEL in order to eliminate dynamical features that have little effect on the model output and would be rejected by the DBM approach.

3 Global sensitivity analysis and model calibration

Since some aspects of sensitivity analysis are not broadly known, this section will provide an outline of the main topics in GSA that are important within the context of the present paper, i.e. with particular attention to the links with calibration. All the GSA methods mentioned here will be subsequently applied for the calibration exercise of the TOPMODEL. Any mathematical or computational model can be formalised as a mapping $Y=f(X_1, \dots, X_k)$, where X_i are the uncertain input factors. 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 GSA, 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, pointing out their role in the context of calibration.

3.1 Variance based methods

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

$$V_i = V[E(Y|X_i)] \quad (1)$$

and the total effect

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

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:

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

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

$$\sum_{i=1}^k S_i = 1 \quad (4)$$

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 $\sum_{j=1}^k (S_j) \leq 1$ and these models are characterised by the existence of interaction effects, leading to the most general variance decomposition scheme (Sobol', 1990),

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \sum_i \sum_{j>i} \sum_{l>j} S_{ijl} + \dots S_{12\dots k} = 1 \quad (5)$$

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, $S_i \leq S_{Ti}$ 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 sensitivity settings that are extremely relevant in the calibration context: factors prioritisation and factors fixing.

– 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 prioritization” 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 .

– Factors fixing (FF)

Another aim of GSA 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 most clearly identifiable 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: such a situation flags a clearly identifiable influence on the output and the ana-

lyst 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 realisations 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 a 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 realisations 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 orders, or even simply of the S_i , S_{Ti} couple, is the most intensely researched topic in GSA. 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: $N_{tot}=Nk+2$, 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 the absolute value of derivatives over the space of factors. Elementary effects are defined as

$$EE_i^j = \frac{|Y^j(\Delta_i^j) - Y^j|}{\Delta_i^j} \quad (6)$$

with,

$$Y^j(\Delta_i^j) = Y(X_1^j, X_2^j, \dots, X_{i-1}^j, X_i + \Delta_i^j, X_{i+1}^j, \dots, X_k^j)$$

$$Y^j = Y(X_1^j, X_2^j, \dots, X_k^j)$$

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:

$$EET_i = \sum_{j=1}^r EE_i^j, \quad (7)$$

where r is the sample size for each factor, usually of the order of 10, for an overall cost of $r(k+1) \ll N_{tot}$. 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, 2005; Campolongo et al., 2006).

The Elementary Effect Test is a development and refinement of the Morris (1991) method. It departs from “Morris” original idea in two ways (Campolongo et al., 2006):

- the sampling strategy of Morris is improved as follows: once the number of trajectories r has been chosen, we generate a number $r^* \gg r$ of trajectories applying “standard” Morris (1991) design and retaining for

model evaluation the subset r that provides the best exploration of the input space, i.e. to avoid oversampling of some levels and undersampling of others that often occurs in the standard Morris procedure;

- the standard “Morris” procedure computes the mean (μ) and standard deviation (σ) of elementary effects, while we only compute the mean of the absolute values of the elementary effects (EET). For screening purposes, this is actually the only measure needed, since this alone is able to provide negligible input factors ($EET \approx 0$) while, in the standard Morris approach, one has to look at the bi-dimensional plot in the (μ, σ) plane and selectively screen the parameters lying towards the origin. Of course, μ and σ can be used to make some guess about non-linearities or interaction effects, but this goes beyond the scopes of screening applied in the current paper.

4 A practical example

The data set used in this example contains hourly data for rainfall, evaporation and flow observations for the River Hodder at Hodder Place, North West England, 1991. The River Hodder has a catchment area of 261 km² and it forms part of the larger River Ribble catchment area of 456 km². 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 on a seasonal basis.

Data from the Hodder catchment have been analysed before using DBM modelling: first, with the objective of adaptive forecasting, using the flow measurement as a surrogate for catchment storage (Young, 2002); and second, as a simulation model (Young, 2003), where the catchment storage is modelled as a first order dynamic system with rainfall as the only input. In both cases the data set used is quite short: it consists of 720 and 480 hourly samples, respectively, with the rainfall series based on the Thiessen polygon average of 3 tipping bucket rain gauges. Based on these data, the latter model, in particular, achieves excellent validation performance, with the model output explaining 94% of the observed flow, only marginally less than the 95% achieved in calibration.

In the present example, we have chosen a data set that is somewhat longer, with 1200 hourly samples, but the rainfall measurements are from a single rain gauge. Probably as a result of this and other inconsistencies in the data due to a reservoir operation (see later discussion in Sect. 4.1), the rainfall-flow data used for calibration and validation exhibit some problems from a modelling standpoint and, as we shall see, the validation performance of both the DBM model and TOPMODEL is considerably worse than that achieved in the previous DBM modelling studies. However, this has

the virtue of allowing us to see how the modelling methodologies function when confronted with data deficiency and how our proposed combined approach, involving GSA, can provide useful results, even when using relatively poor data such as these, which are similar to those often experienced in practical hydrological applications.

Only the rainfall-flow data are used in the DBM modelling; while the TOPMODEL calibration uses the DTM and evaporation data, in addition to the rainfall-flow measurements. The January–March 1991 rainfall-flow data are used for the identification and estimation (calibration) of both models; while the second part of the data (October–December) is used for the validation. The catchment is predominantly covered by grassland and, 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, it will be noted that only Autumn–Winter months are used for the modelling. Of course, both models would require further evaluation, based on a much more comprehensive data set, before being used in practice.

4.1 DBM rainfall-flow model for Hodder catchment

Following the DBM modelling procedure outlined in Sect. 2, the first model structure 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).

The non-parametric estimate of the state-dependent function $f(y_t)$ is plotted as the full 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 $f(y_t) = g y_t^\beta$ is selected for the subsequent “parametric estimation” stage of the DBM modelling, where g is a scale factor selected to make the the SDP coefficient $f(y_t) = 1.0$ at maximum flow.

The best identified linear STF model between the effective rainfall $ue_t = g y_t^\beta \times r_t$ 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

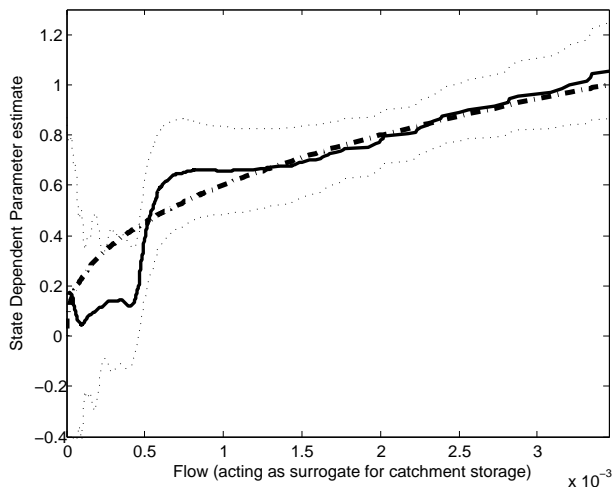


Fig. 1. This graph shows the estimated non-parametric SDP relationship (full line) between measured flow and the effective rainfall parameter, where the flow is acting as a surrogate for catchment soil moisture. The 95% confidence bands are shown by the dotted lines and the dash-dot line is a parametric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

the following, nonlinear DBM model between the rainfall r_t and flow y_t :

$$\begin{aligned} ue_t &= 10.258 y_t^{0.411} \times r_t \\ \hat{x}_t &= \left[\frac{0.0045}{1-0.9903z^{-1}} + \frac{0.1269}{0.86821z^{-1}} \right] ue_t \\ y_t &= \hat{x}_t + \xi_t; \quad \text{var}(\xi_t) = 0.042\hat{x}_t^2 \end{aligned} \quad (8)$$

where ξ_t represents the estimated noise on the relationship which, as shown, is modelled as a (state-dependent) heteroscedastic process with variance proportional to the square of the estimated flow, \hat{x}_t^2 , and z^{-1} is the backward shift operator: i.e. $z^{-s} y_t = y_{t-s}$.

Concerning the expression $ue_t = f(y_t) \times r_t$, it is worth noting that, in its present, purely surrogate role, the flow series is *not* being used as an additional input to the model (i.e. in addition to the rainfall) in the normal sense: rather, at each time point, it simply affects the value of the state-dependent gain coefficient $f(y_t)$ that multiplies the rainfall in its conversion to effective rainfall. Therefore, no dynamic system is involved as there would be if one used the modelled flow rather than the measured flow in this equation. The resulting effective rainfall occurs on the same time point as the real rainfall and just the level of rainfall at this time is changed. In other words, *no flow feedback* occurs at all in the model.

However, this exploitation of the flow measurement as a surrogate measure of soil moisture content in the catchment does restrict the present model's use to defining the dominant dynamic characteristics of the catchment, as well as for data assimilation and flow forecasting applications. It cannot be used for simulation, although an alternative DBM

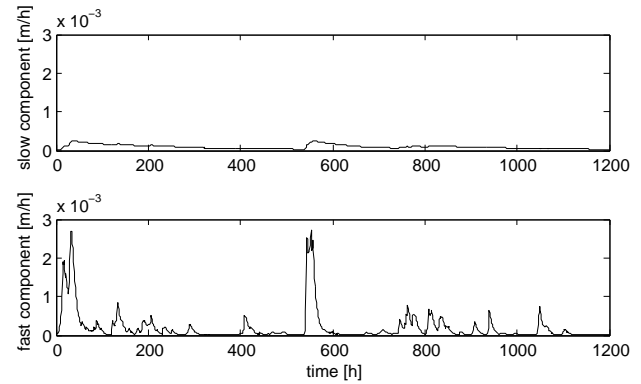


Fig. 2. The estimated slow (upper panel) and quick (lower panel) flow components of the DBM model output. The most likely interpretation of these component flows is that the slow component represents the “groundwater” (baseflow) effects and the quick component represents the “surface and near-surface”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

simulation model has been developed by introducing a parsimonious model for catchment storage s_t (e.g. Young, 2003). This replaces the flow y_t in the above definition of ue_t , and, if required, a similar model could involve temperature and/or evapo-transpiration data, when these are available (see e.g. the rainfall-flow models developed by Whitehead and Young, 1975; Jakeman et al., 1990).

The noise ξ_t identified in DBM modelling represents that part of the data not explained by the model. It can be quantified in various ways: e.g. simply by its variance; by its variance in relation to that of the model output; by a heteroscedastic process, where the variance is modulated in some state-dependent manner, as in the present paper; by a coefficient of determination; by a stochastic model, if such a model is applicable; by its probability distribution; or by its spectral properties. Of course, it can be due to various reasons: measurement inaccuracies, limitations in the model as a representation of the data, the effects of unmeasured inputs, etc. Sometimes, if it has rational spectral density, this noise can be modelled stochastically, e.g. by an Auto-Regressive (AR) or an Auto-Regressive, Moving Average (ARMA) process, but this is not always applicable when dealing with real data (see later). Consequently, the “instrumental variable” identification and estimation methods used for DBM modelling are robust in the face of noise that does not satisfy such assumptions (see e.g. Young, 1984).

In order to reveal the most likely mechanistic interpretation of the model, the [2 2 0] STF relationship in the model (8) is decomposed by objective continued fraction expansion (calculus of residues) 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 103 h; and a “quick” flow component, with a residence time of 7.1 h, whose flow response is

Table 1. Synthetic table of main calibration and validation results.

	Calibration R_T^2	Validation R_T^2
DBM	0.928	0.79
DBM (CT)	0.917	0.838
TOPMODEL (base)	0.905	0.747
TOPMODEL (pre-calib.)		
case (a)	0.883	0.766
case (b)	0.880	0.762
case (c)	0.53	< 0

shown in the lower panel of Fig. 2. However, it should be noted that, while the quick residence time is estimated well, with narrow uncertainty bounds (95 percentile range between 7.2 h and 6.9 h), the slow residence time is poorly estimated with a skew distribution (95 percentile range between 81 h and 136 h). This is due to the residence time being relatively large in relation to the observation interval of 1200 h, so that the information in the data on this slow mode of behaviour is rather low.

The partition of flow between these quick and the slow components, respectively, is identified from the decomposed model with about 68% of the total flow appearing to arrive quickly from surface and near surface processes; while about 32% appears as the slow-flow component, probably due, at least in part, to the displacement of groundwater into the river channel (base flow). The decomposition and its associated partitioning are a natural mathematical decomposition of the estimated model which has a nice interpretation in both dynamic systems and in hydrological terms (see e.g. Young, 2005). It is a function of two mathematical properties of model: the eigenvalues, which define the residence times of the “stores” in the parallel pathways; and the steady state gains of these stores, that define how much flow passes through each store¹. However, it must be emphasised that the uncertainty associated with the partitioning in this particular example is high because of the uncertainty in the estimation of the slow residence time.

It is clear from this analysis that the partitioning of the DBM model, as identified in this manner, is inferred from the data by the DBM modelling methodology and it is not imposed on the model in any way by the modeller, as in other approaches to rainfall-flow modelling. For example, the interested reader might compare the HyMOD model of Moradkhani et al. (2005) of the Leaf River, where the model structure is assumed by the modellers, with the DBM model of Young (2006), which is based on the same data set but where a similar but subtly different structure is inferred from the data.

¹Another feedback decomposition is possible (identifiable) but this is rejected because it has no clear physical interpretation.

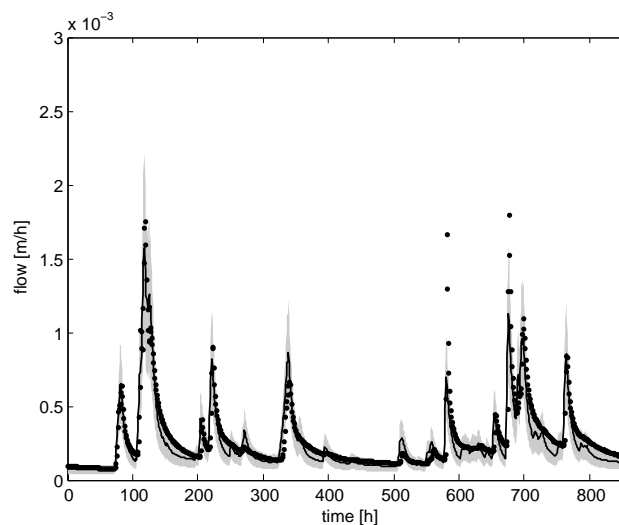


Fig. 3. Validation of the DBM model on October–November 1991 period, 79% of output variation explained; dots denote the observations, simulations are marked by a solid black line, shaded area denotes 95% confidence bands.

In Table 1, the main results in calibration and validation for the DBM model and TOPMODEL are synthesised. The DBM model (8) has a Coefficient of Determination, $R_T^2=0.928$ (i.e. 92.8% of the measured flow variance is explained by the simulated model output), which is a substantial improvement on the linear model ($R_T^2=0.84$), where it involves the addition of the single power law parameter. Clearly, the use of observed flow in the definition of effective rainfall feeds some information about output flow in the linear STF (see earlier discussion). However, similar performance is obtained when the catchment storage s_t is modelled from rainfall data, without any use of flow data (Young, 2003). This confirms that it is a sensible identification of the input nonlinearity and its associated effective rainfall that allows the model to describe well the observed system, not the use of flow data per se. Finally, the estimated noise ξ_t is identified as a heteroscedastic, high order process.

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: the 95% confidence bands, as shown later in Fig. 3, for example, are a function of both the estimated parametric uncertainty and the residual heteroscedastic noise ξ_t . Without the MCS analysis, the DBM modelling is very computationally efficient, normally involving only a very small computational cost (a few seconds of computer time). This contrasts with the TOPMODEL analysis, considered in the next section, where computationally much more demanding Monte Carlo Simulation analysis is an essential aspect of the GLUE calibration and uncertainty analysis. Consequently, it is easy to use the DBM modelling for the role suggested in this paper.

Table 2. Parameter distributions applied in MC calibration of TOPMODEL.

For each parameter, the top line shows the ranges for the base calibration exercise, while modified ranges for pre-calibration cases (a,b,c) are specified, when applicable.

Parameters	Distribution	Min.	Max.	Mean	Std
<i>input</i>	uniform	0.8	1.2	1	0.116
case (c)	uniform	0.8	1	0.9	0.0577
<i>SK0</i>	uniform	10 000	40 000	25 000	8600
<i>m</i>	uniform	0.001	0.03	0.0157	0.008
cases (a, b)	uniform	0.005	0.03	0.0175	0.0072
case (c)	uniform	0.0025	0.003	0.00275	1.44e-4
<i>LRZ</i>	uniform	1.e-4	0.04	0.02	0.012
case (c)	uniform	0.015	0.04	0.0275	0.0072
<i>KS</i>	log-uniform	1.e-7	1.e-3	1.08e-4	2.05e-4
case (c)	log-uniform	1.e-4	1.e-3	3.9e-4	2.49e-4
δ	uniform	0.001	0.2	0.1	0.0575
case (a)	uniform	0.01	0.2	0.105	0.0548
case (b)	fix	0.15	0.15	-	-
case (c)	uniform	0.1	0.2	0.15	0.0289

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), as well as the effects of stochastic input noise. Typically, the uncertainty associated with the DBM flow response is mainly accounted for by model output error, which in DBM modelling is considered as output “measurement noise” (see earlier).

The model validation results are shown in Fig. 3 and Table 1: here only 79.0% of measured flow variation is explained by the DBM model. This is a little better than TOPMODEL which explains 74.7% of the output variation (see later Sect. 4.2). Interestingly, the performance of the DBM model is improved if the continuous-time (CT) version of the model (see e.g. Young, 2005) is identified and estimated using the continuous time version (RIVC) of the RIV algorithm in the CAPTAIN Toolbox. Although the calibration $R_T^2=0.917$ is then marginally less than for the discrete-time DBM model, the validation $R_T^2=0.838$ is improved.

The main difference between the models is that the DBM model validates well on the peak flows, while TOPMODEL does not do so well in this regard, although having somewhat better performance than the DBM model at lower flows. Both of these results are not particularly good, however, in comparison with previous DBM rainfall-flow modelling results based on Hodder data (see previous discussion), where 94% of the output variation is explained by the model during validation. It is also worse than other typical DBM modelling results (see the cited references on DBM modelling), where

85–95% of the flow is normally explained by the model during validation.

As pointed out previously, this poor performance is almost certainly due to data deficiencies in the form of poor, single gauge rainfall observations and other inconsistencies in the flow measurements arising from the reservoir operations. Note that, in this example, the high level of uncertainty of the slow flow component (see earlier comments) has a deleterious effect on the validation performance of the DBM model because the validation error is particularly sensitive to errors in this slow flow component. Also, it is interesting to note that, if the DBM model calibration is performed on the validation data set, only 84% of flow variation is explained by the model with the same structure. Compared with this, the above validation figures (79.0% and 74.7%) are not as bad as they look at first sight. However, this reinforces our comments about the deficiencies in the data, particularly, it would appear, the validation data set.

Despite these limitations in the data, the modelling results are satisfactory for the illustrative purpose of the present paper (i.e. presentation of the two different approaches to rainfall-flow modelling and how they may be combined) since both models are subject to exactly the same data deficiency and the results reveal interesting comparative aspects of the models.

4.2 TOPMODEL calibration for Hodder catchment

In the present context, the main aim of the TOPMODEL (Beven and Kirkby, 1979) 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 sec-

tion. 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 calibration and the associated sensitivity analysis described here, in addition to uncertainties pertaining to 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.,

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

where $Q_0 = SK_0 \exp(-\lambda)$ is the flow when $S(t)=0$; $Q_v(t)$ denotes the recharge to the saturated zone; SK_0 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 2) of $\pm 20\%$ 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 (for a similar application see Pappenberger and Beven, 2004). 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 the A breakpoint at time t , $A_t = y_t(1-4\delta)$, implying a heteroscedastic error structure. The width δ of the trapezoidal membership function is assumed to be uncertain as well, in a range between $[0.1\% \text{ and } 20\%]$. 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 SK_0 , 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.

To some extent and in contrast to the DBM modelling, the specification of the above uncertainty on the input parameters, as required for the MCS analysis, is subjective and will depend upon the experience and prior knowledge of the analyst. 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 2 using quasi-random LPTAU sequences (Sobol' et al., 1992). The sample size requirements of GLUE applications are mainly linked to the convergence of the cumulative distribution functions (Pappenberger et al., 2005; Saltelli et al., 2004, p. 153–155) and thus to the “success rate”, i.e. the percentage of samples that provides good fit. Given the relatively small number of parameters in this case, the sample size was sufficient for the current application. Moreover, we also kept the computational cost at the minimum to show that careful pre-calibration based on sensitivity analysis results allows for a reduction of the computational costs of GLUE exercises, by improving the “success rate” (see Sect. 4.2.1 below).

In the GLUE approach, the predictive probability of the flow takes the form:

$$P(\hat{x}_t < y|\Omega) = \sum_{\{i:\hat{x}_t < y|\vartheta_i, \Omega\}} f(\vartheta_i|\Omega) \quad (10)$$

where \hat{x}_t are simulated outputs and $f(\cdot)$ denotes the likelihood weight of parameter sets $\Theta = \{\vartheta_1, \dots, \vartheta_n\}$, in which n is the number of MC samples, conditioned on the available observations $\Omega = \{y_1, \dots, y_T\}$.

The next step is to investigate the weights. The trapezoidal membership function, at each discrete time t and each element ϑ_i , $i=1, \dots, n$ of the MC sample, takes the values:

$$\begin{aligned} f_t(\vartheta_i|y_t) &= 1 && \text{if } B_t < \hat{x}_t < C_t \\ f_t(\vartheta_i|y_t) &= \frac{\hat{x}_t - A_t}{B_t - A_t} && \text{if } A_t < \hat{x}_t < B_t \\ f_t(\vartheta_i|y_t) &= \frac{\hat{x}_t - D_t}{C_t - D_t} && \text{if } C_t < \hat{x}_t < D_t \\ f_t(\vartheta_i|y_t) &= 0 && \text{if } \hat{x}_t < A_t, \hat{x}_t > D_t \end{aligned} \quad (11)$$

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

$$f(\vartheta_i|\Omega) \sim \sum_{t=1}^T f_t(\vartheta_i|y_t) \quad (12)$$

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 $y_t(1\pm 4\delta)$ at any time (i.e. outside a relative error bound of 400%); while it will be maximum for those falling within

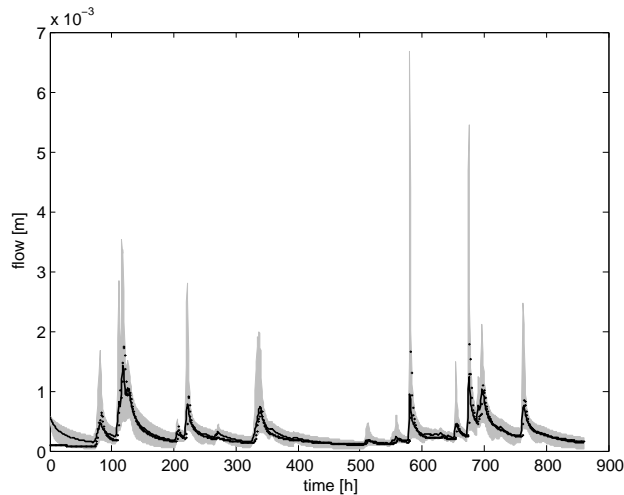


Fig. 4. Validation of the TOPMODEL: 74.7% of output variation explained; dots denote the observations, simulations (posterior mean) are marked by a solid black line, shaded area denotes 95% confidence bands.

the range $y_i(1 \pm \delta)$ at any time (i.e. within a relative error bound of $100\delta\%$). The values of δ used in the MCS (Table 2) range from a very strict rejection criterion (simulations are rejected if the relative error at any time is larger than $\pm 0.4\%$) to a very mild one (simulations are given the maximum weight when the relative error at any time is smaller than $\pm 20\%$, while they are rejected only when they fall outside the $\pm 80\%$ bound). This initially wide range will undergo a revision in the second part of this example, based on sensitivity analysis results and taking DBM uncertainty estimation as a benchmark for adjusting (pre-calibrating) the priors in a less subjective way than this initial set-up. However, as discussed later, δ is not the only parameter strongly driving the uncertainty bounds, with m playing an important role as well.

The calibration flow predictions of TOPMODEL explain 90.5% of the observed flow variations (see Table 1), not significantly different from that obtained using the DBM approach (92.8%). Note that TOPMODEL utilises evaporation measurements in addition to the rainfall and flow measurements used in the DBM model. The model residuals are heteroscedastic and highly coloured, even more so than the DBM model residuals. On the validation data (Fig. 4), the model output explains 74.7% of the flow variations, a little worse than DBM (79.0%: see Fig. 3). 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 peak flows is about twice the bound of the DBM model. The latter uncertainty bounds seem more reasonable, given the nature of the observations around the peak flows.

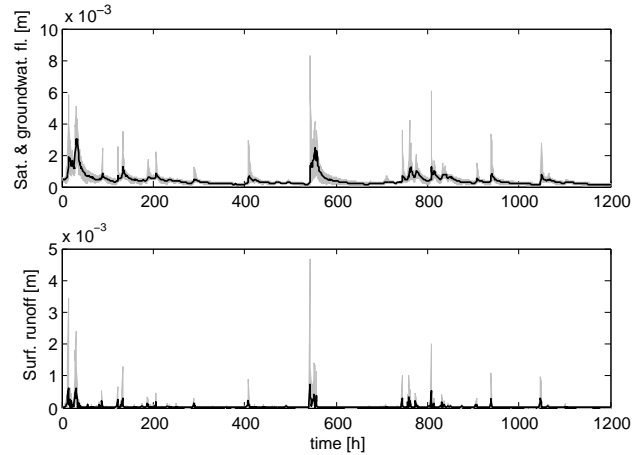


Fig. 5. Uncertainty estimation of the discharge from the saturated subsurface store (upper panel) and of the “surface runoff” (lower panel) of the TOPMODEL; simulations (posterior mean) are marked by a solid black line, shaded area denotes 95% confidence bands.

2. The partition of the output flow of the TOPMODEL into two components, deriving from rainfall on the dynamic saturated contributing areas (denoted here as “surface runoff”) and from the discharge from a nonlinear saturated subsurface store, differs from the DBM’s partitioning, as discussed previously, and is shown in Fig. 5. The saturated contributing area accounts for only 5% of the predicted discharge, with “quick” dynamical features merely adjusting the flow peaks, while the remainder is contributed by the nonlinear subsurface store.

4.2.1 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. We have seen that the uncertainty from the calibrated TOPMODEL is much larger than for the DBM model (notice the difference in the vertical axis scales of the figures caused by this). Taking DBM as a relatively objective benchmark for uncertainty estimation, we would like to analyse under which conditions the volatility of predictions can be reasonably reduced. This is a typical example for the Factors Prioritisation setting, where one can identify strategies for the reduction of the uncertainties 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 Table 3, first column, we show 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); while in Table 3, second column we show the Elementary Effect Test (EET), computed using only an additional 168 model runs. The main effects tell us that the

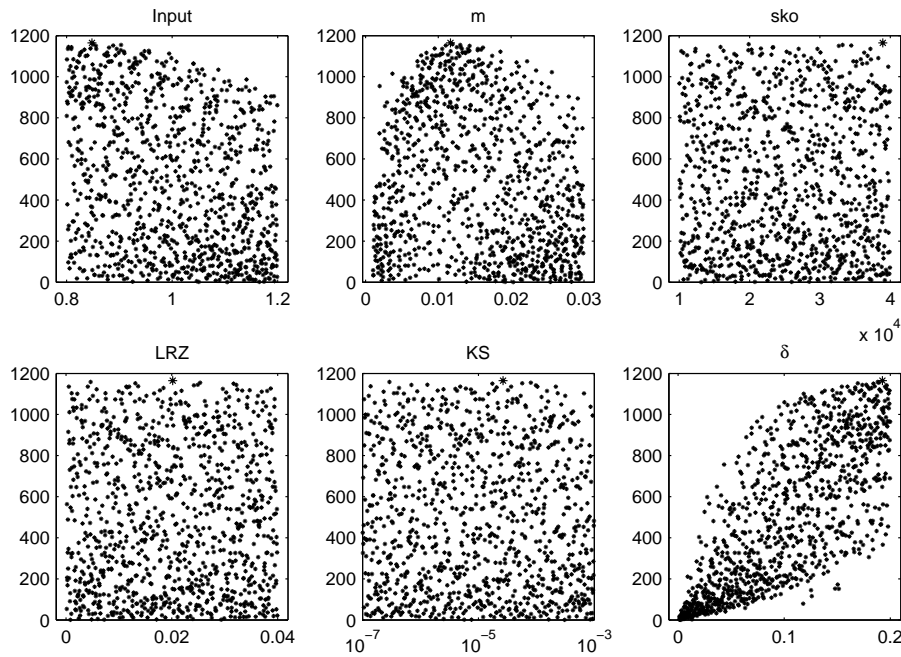


Fig. 6. Scatter plots of the sample of likelihood weights versus uncertain input factors.

uncertainty in the width of the trapezoidal membership function dominates, followed by the parameter m , which is the only physical 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 , controlling the recharge from the root zone to unsaturated zone, has influence on the model mostly in the periods of low flow (details of the SA are 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 $m < 0.005$, flow peaks related to “surface runoff” 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. 6). Looking at the plot for δ , we can see that 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.

Based on these considerations, we conduct a second calibration step, by shrinking the support for δ to the range [0.01, 0.2] and the support for m to [0.005, 0.03] (this is an example of application of the Monte Carlo filtering technique). Such

Table 3. Sensitivity analysis of the TOPMODEL: first two columns show the main effect sensitivity indices and the elementary effect tests for the initial calibration step. The third column shows the main effects for the pre-calibrated case discussed in Sect. 4.2.1, to reduce volatility of predictions.

Parameters	S_i	EET_i	S_i , case (a)	S_i , case (b)
<i>input</i>	0.0842	0.3946	0.123	0.304
<i>m</i>	0.1823	0.5526	0.214	0.49
<i>SKO</i>	0.0004	0.0236	0.0002	0.0004
<i>LRZ</i>	0.0003	0.0204	0.0002	0.0021
<i>KS</i>	0.0140	0.1664	0.03	0.103
δ	0.5760	1.0000	0.53	–

modifications are quite in order since, as pointed out previously, the specification of the original input uncertainties is rather subjective. Moreover, this revision of δ (measurement error) from an initially wide interval is one possible way to conduct a calibration of the measurement error (denoted as case (a) in Table 1 and Table 2); another one being to pre-calibrate δ based on DBM confidence bounds, which provides a statistically reliable estimate of the measurement error, when no direct information on the magnitude of such an error is available. In the latter case, one can consider the DBM heteroscedastic noise variance estimation in (8), which implies a noise standard deviation of 0.205 times the output flow, i.e. a 95% bound of $\pm 0.41\hat{x}_t$. With the fuzzy trapezoidal membership function used here, this roughly corresponds to fixing $\delta \approx 0.15$. The results of the associated

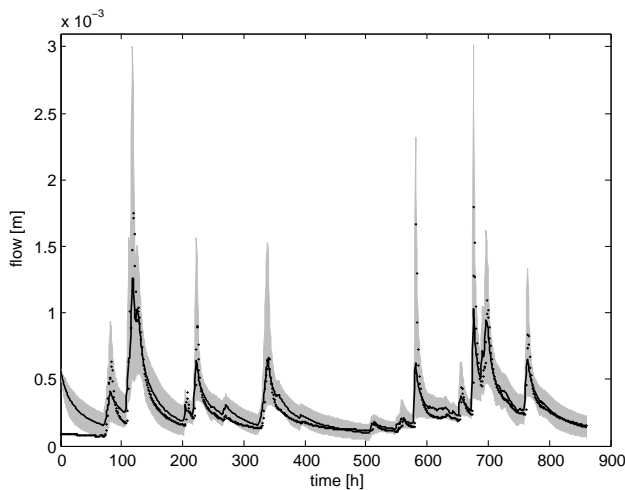


Fig. 7. Revised validation of the TOPMODEL, to reduce volatility of predictions, case (a): 76.6% of output variation explained; dots denote the observations, simulations (posterior mean) are marked by a solid black line, shaded area denotes 95% confidence bands. Surface flow partition: 3.5% of the overall flow.

calibration and validation exercises are denoted as case (b) in Table 1 and Table 2.

Figure 7 shows the updated validation exercise for case (a): the portion of flow variance now explained by the new calibrated model is 88.3% (Table 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, 10.4% 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 partitioning is also hardly modified: “surface runoff” is now 4.3% of the overall flow. Cutting the high tails from the uncertainty distribution of surface runoff flows did not alter the partitioning too much, meaning that the high tails in the upper panel of Fig. 5 are not supported by the data. The validation step confirms the results (Table 1): 76.6% of the data is now explained by model predictions, i.e. a slightly larger amount than before. Although this is not a statistically significant improvement, there is clearly no degradation in the results, and this is an indication of the validity of the approach proposed in this paper. The uncertainty bounds in this case are also similar to DBM, with the drawback of a larger portion of data below the bounds in periods of small flow: 14.3% versus the 6.4% of the initial case.

The calibration/validation results for case (b) are almost identical to case (a) (see Table 1). This shows that, using DBM results for the GLUE calibration, can allow for the calibration of the measurement error in a reasonable manner, thus reducing the number of uncertain parameters.

In both of these new calibration steps, the effect of the different uncertainty sources is different. For case (a), δ remains the most important parameter, but its sensitivity decreases, while the sensitivity of m increases. Moreover, KS also acquires a small main effect (see Table 3, third column). For case (b), on the other hand, we can see that fixing δ based on DBM estimation allows other sensitivity patterns to be revealed more clearly and in a more balanced way, and balanced main effects are usually a sign of better identification. In particular, we have that, among the structural parameters, *both* m and KS have a significant effect on model performance. 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, this is not a general result: it does not mean that there are not other “behavioural” classifications or other data sets that may highlight the role of SKO and LRZ .

4.2.2 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 and it is expected that the partitioning of the two models will be different because it is constrained in TOPMODEL by the assumptions built into its formulation, namely by the surface contributing areas. Consequently, TOPMODEL’s decomposition may not be considered any less physically acceptable than the one suggested by the DBM model. 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 (although this is likely to depend on the view of the hydrologist and not all readers will agree). For instance, if we consider the observed flow response, there seems to be an appreciable base-flow component and the DBM model estimate of the “slow-flow” component seems consistent with what one might expect in this regard. On the other hand, TOPMODEL’s partitioning, based on predicted contributing areas, gives quite a different form of partitioning, with part of the fast response provided by the nonlinear subsurface store. Consequently, purely as an illustrative exercise, it might be instructive to use sensitivity analysis to investigate whether other partitions are possible within the uncertainty specifications of the TOPMODEL. This helps to further demonstrate how GSA can assist in model evaluation

and in revealing the essential differences between different model mechanisms. In particular, this will exemplify the capability of MCF techniques in constraining models into specified behavioural characteristics, *if this is desirable* according to the modeller's views. This exercise is, in practice, the dual to that considered in the previous section.

The MCF procedure adopted for this purpose was to target more balanced partitioning, starting from the base MC sample used in Sect. 4.2, i.e., in this case, the behavioural classification requires higher portion of surface runoff, *irrespective* (initially) of the implications this may have on the model fit. This MCF analysis shows that the portion of surface flow can be increased to 30–40% only for very small values of m . Remembering the results in previous Sections, this turns out to be the range of m values displaying the smallest GLUE fuzzy weights, as well as the complementary support adopted to reduce the uncertainty of TOPMODEL predictions. This already anticipates that, while it is possible to change the partitioning in TOPMODEL, this may be rejected by a reduced ability of the model to explain the data. Nonetheless, we attempted to calibrate the TOPMODEL by restricting m in the range [0.001–0.003], to verify that there are not other features in TOPMODEL that allow the model to fit the data in such a constrained situation. The calibration failed, with 93% of MC runs having negative R_T^2 values and huge uncertainty bands. This considered, we tried a final MCF step, in which we kept only the 7% runs with positive R_T^2 (denoted as case (c) in Table 1 and Table 2). This strategy resulted in an unsatisfactory $R_T^2=53\%$ in calibration, with the contribution of surface flow in the range 10–20%. The confidence bounds on the flow response are narrower than in the previous calibrations, with a large part of flow under-predicted in periods of low flow; while the flow peaks tend to be over-predicted. Also, the dynamical features of the two flow components of the model still appear too similar: both of them present clear rapid response to rainfall episodes and increasing the “surface” component, also makes the response of the “subsurface” flow larger. Finally, the validation results are totally unacceptable ($R_T^2 < 0$).

The above analysis provides an instructive academic example of how MCF techniques can be used to investigate whether the parameters of a mathematical model can be modified to yield specific behavioural characteristics (partitioning in this case). However, for the present TOPMODEL case-study this severely impairs the model's ability to explain the observed data. As a result, and not surprisingly given the nature of TOPMODEL, the partitioning objective fails and the possibility of changing the model behaviour in this sense is rejected.

5 Discussion

This paper has described a modelling framework for rainfall-flow modelling that combines bottom-up, reductionist mod-

elling with top-down, parsimonious modelling. This combined approach, which exploits sensitivity analysis to build a bridge between the two types of model, has been illustrated by its application to intentionally selected poor data from the Hodder catchment in the North West of England, using the physically based, semi-distributed rainfall-runoff model TOPMODEL to exemplify the bottom-up approach; and DBM modelling to exemplify the top-down alternative.

The DBM model yields a well identified representation of the Hodder data, with good (92.8%) explanation of the flow data in calibration, as well as a clear mechanistic interpretation arising from the identified linear, second order stochastic transfer function between effective rainfall and output flow. In particular, the output flow of the DBM model can be further decomposed into a “slow” (32.3%) and a “quick” (67.7%) component, with a suggested physical interpretation that these are associated with “groundwater” and “surface/near surface” processes, respectively. In terms of explaining the flow data, the calibration results from the TOPMODEL analysis are not all that different from those of the DBM model but the flow decomposition is much different (see below). As expected from the data-set chosen, the validation results for both models are not particularly good, with the DBM model explaining 79% of the validation flow variation and TOPMODEL yielding a little less than this, almost certainly because of the data deficiencies mentioned above and discussed in Section 4 of the paper.

Two issues arise from the initial sensitivity analysis of TOPMODEL (i.e. with the initially specified nominal parameters and uncertainty bounds) that reveal potentially important differences between the stochastic and dynamic characteristics of the TOPMODEL and the DBM model. First, the large volatility of TOPMODEL predictions, with confidence bounds twice that of the DBM model; and second, the behaviour of the two output flow components in the TOPMODEL, with the subsurface component accounting for most of the output flow, while the surface component has comparatively little impact on the model output flow.

As regards the high volatility of the TOPMODEL predictions, GSA shows that less volatile performance is possible, without significantly affecting the explanation of the data (the R_T^2 slightly improves while 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 shows how the combined use of DBM and GSA results can be helpful for model identification and simplification: (i) when δ is fixed according to DBM estimation (case (b)), identification of m and KS is strongly enhanced; (ii) SKO is clearly highlighted as irrelevant for calibration purposes and, hence, it can be fixed at a nominal value. In relation to flow partitioning, GSA confirms that it is impossible to bring the two models into reasonable agreement as regards the flow partitioning, which is not all that surprising since the models have significant structural differences: DBM's inference is based on *dynamical* features identified

from the data, like residence times; while TOPMODEL's is based on the concept of dynamic contributing areas in space.

One of these differences lies in the nature of the nonlinearity. As in the present example, the application of the relatively objective State-Dependent Parameter (SDP) estimation procedure to other rainfall-flow data considered previously in the cited references, suggests that the nonlinear effects are quite often, although not always, connected with the rainfall input. In the case of TOPMODEL, the nonlinearity is specified a priori in the model and, in the present example, it mainly resides internally or at the output. Its input nonlinearity was evaluated carefully and found to have a quite minor effect in this particular case, only becoming active at all in a fairly narrow range of small flow values. This suggests the need for more research on this topic in order to reach a clearer picture on the nature of nonlinearity in rainfall-flow processes and it is hoped to pursue this in the future.

The paper also demonstrates how GSA can assist in improving the calibration of deterministic, bottom-up, hydrological models, by making their properties more transparent and in building a bridge between them and the DBM, or other data-based modelling results. It is also clear that the computationally much less intensive DBM analysis provides a useful input into TOPMODEL sensitivity analysis and calibration, raising critical questions about the results obtained in the initial calibration step and leading to useful modifications in the subsequent steps denoted as cases (a) and (b) in Sect. 4.2.1. In other words, the two modelling methodologies, one inductive and the other hypothetico-deductive, have good synergy; combining well to produce a complete, critical 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: the *specific* DBM-type model considered in the present paper has advantages in the context of real-time forecasting; while TOPMODEL has advantages within a semi-distributed, simulation context.

Within a forecasting context, TOPMODEL is not intended for forecasting applications. On the other hand, the DBM model discussed in the present paper was designed specifically for forecasting and it can be enhanced in such applications by the incorporation of data assimilation procedures: i.e. by the introduction of real-time state and parameter updating (e.g. Young, 2006) based on the most recent runoff observations. This is particularly easy in the case of the DBM-type models and an example of such a system has been described recently by Romanowicz et al. (2006) and Young et al. (2006).

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 dynamic response, the form of the effective rainfall nonlinearity and the partitioning of the effective rainfall into differ-

ent flow pathways. This can be useful in a number of ways. For instance, a mechanistic characterisation of the rainfall-flow process, such as that of TOPMODEL, which arises 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, provided the user accepts the premises on which it is based. However, if this is not happening to the satisfaction of the modeller, then the possibility of a model revision needs to be taken into account. In such model revision exercises, DBM modelling can be of assistance, since, with the help of GSA tools, it can guide the modeller in revising the model specification or structure.

For example, in a wider context to that considered in the present paper, the modeller might wish to identify, within a conceptual model, advection-dispersion mechanisms that have characteristic times similar to those estimated in the DBM model and base the model revision on this. Or again, the DBM model with a flow-dependent effective rainfall nonlinearity is normally obtained very quickly from the data analysis and yet it can often provide a better validated explanation of rainfall-flow data than alternative conceptual models. As a result, this estimate can be utilised for evaluating alternative catchment storage equations in "hybrid-metric-conceptual" (HMC) models (see Wheater et al., 1993) that include effective rainfall generation: for example, models such as IHACRES (Jakeman et al., 1990) and HYMOD (Vrugt et al., 2005).

The DBM model discussed in the present paper has been developed specifically for flow and flood forecasting (e.g. Young, 2002; Romanowicz et al., 2004; Romanowicz et al., 2006). However, the effective rainfall module in the DBM model needs to be modified into a form that does not use the flow measurement as a surrogate measure of catchment storage if it is to be used for simulation modelling applications. Here, the effective rainfall nonlinearity is dependent on an unobserved or "latent" catchment storage state obtained, for instance, from some conceptualised catchment storage equation, as in Young (2003). In more general terms, it is possible to conceive alternative DBM models of an HMC type that utilise other data, such as evapotranspiration information or even DTM data, as in TOPMODEL. These alternative DBM models emphasise the fact that the DBM modelling is much more than the specific DBM hydrological model considered in this paper: it is *not* limited to the synthesis of forecasting models, but an objective-orientated *approach* to modelling general stochastic, dynamic systems. In this context, it is hoped that an evaluation of how alternative DBM-type simulation models may perform in relation to TOPMODEL within a simulation context may lead to improvements in both model types.

In conclusion, it is clear that there is never a single, unique model of any real system and we believe that a primary function of hydrological systems analysis should be to select the model *that best suits the nature of the defined modelling ob-*

jectives, be they forecasting, simulation or a combination of both. In this context, exercises such as those described in this paper, which try to critically compare and combine different models of the rainfall-flow process (but *not* to judge or prove one model to be “better” than the other), can be extremely useful and revealing. The present paper shows that conceptual model analysis (represented here by TOPMODEL) can be more “corroborated” and robust if it is benchmarked and possibly updated in the light of some “independent” form of data-based analysis, such as that provided by DBM modelling.

Edited by: J. C. Refsgaard

References

- Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, *Environ. Modell. Software*, 22, 1509–1518, 2007.
- Beven, K. J. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, 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. Hydraul. 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.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, 28, 135–147, 2005.
- Morris, M. D.: Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161–174, 1991.
- 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. and Beven, K. J.: Functional Classification and Evaluation of Hydrographs based on Multicomponent Mapping, *International Journal of River Basin Management*, 2(2), 1–9, 2004.
- Pappenberger, F. and Beven, K. J.: Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resour. Res.*, 42, 5302, doi:10.1029/2005WR004820, 2006.
- Pappenberger, F., Beven, K. J., Horritt, M., and Blazkova, S.: Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations, *J. Hydrol.*, 302(1–4), 46–69, 2005.
- Pappenberger, F., Harvey, H., Beven, K., Hall, J., and Meadowcroft, I.: Decision tree for choosing an uncertainty analysis methodology: a wiki experiment, *Hydrol. Processes*, 20, 3793–3798, 2006a.
- 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, 2006b.
- 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.*, 29, 1430–1449, 2006c.
- Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R., and Matgen, P.: Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations, *Hydrol. Earth Syst. Sci.*, 11, 739–752, 2007, <http://www.hydrol-earth-syst-sci.net/11/739/2007/>.
- 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, in: *Proc. 4th Int. Conf. on Sensitivity Analysis of Model Output (SAMO 2004)*, Santa Fe, New Mexico, 8–11 March, 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. Processes*, 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*, 91, 1315–1321, 2006.
- Romanowicz, R., Young, P.C., and Beven, K. J.: 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., Chan, K., and Scott, M. (Eds.): *Sensitivity Analysis*, John Wiley and Sons publishers, Probability and Statistics series, 2000.
- Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Sensitivity analysis for chemical models, *Chem. Rev.*, 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.
- Tang, Y., Reed, P., Wagener, T., and van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, 11, 793–817, 2007, <http://www.hydrol-earth-syst-sci.net/11/793/2007/>.
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004WR003059 2005.
- Wheater, H. S., Jakeman, A. J., and Beven, K. J.: Progress and directions in rainfall-run-off modelling, in: *Modelling Change in Environmental Systems*, edited by: Jakeman, A. J., Beck, M. B. and McAleer, M. J., Chapter 5, Wiley: Chichester, 1993.
- Whitehead, P. G. and Young, P. C.: A dynamic stochastic model for water quality in part of the Bedford-Ouse river system, in: *Modeling and Simulation of Water Resource Systems*, edited by: Vansteenkiste, G., North Holland, Amsterdam, p. 417–430, 1975.
- Young, P. C.: Recursive Approaches to Time-Series Analysis, *Bull. Inst. Maths Appl.*, 10, 209–224, 1974.
- Young, P. C.: *Recursive Estimation and Time-Series Analysis*, Springer-Verlag, Berlin, 1984.
- Young, P. C.: Data-based mechanistic modelling of environmental, ecological, economic and engineering systems, *Environ. Modell. 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.: Data-based mechanistic modelling and validation of rainfall-flow processes, in: *Model Validation: Perspectives in Hydrological Science*, edited by: Anderson, M. G. and Bates, P. D. John Wiley, Chichester, 117–161, 2001.
- 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. Processes*, 17, 2195–2217, 2003.
- Young, P. C.: Rainfall-runoff modeling: Transfer Function models, in: *Encyclopedia of Hydrological Sciences*, edited by: Anderson, M. G. John Wiley, Hoboken, N. J., Vol. 3, part II, 1985–2000, 2005.
- Young, P. C.: Data-based mechanistic modelling and river flow forecasting, in: *Proceedings 14th International Federation on Automatic Control (IFAC) Symposium on System Identification, Newcastle, NSW, 756–761*, 2006.
- Young, P. C., Romanowicz, R., and Beven, K. J.: Data-based mechanistic modelling and real-time adaptive flood forecasting, in: *Proceedings of Institution of Civil Engineers and British Hydrological Society Workshop on Real-Time Flood Forecasting: Developments and Opportunities*, City Conference Centre, London, 14 November 2006.
- 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.