# An approach for matching accuracy and predictive capability in hydrological model development

**F. FENICIA[1,2], H. H. G. SAVENIJE[2] & L. HOFFMANN[1]**

1 *Centre de Recherche Public Gabriel Lippmann, Belvaux, Luxembourg*
f.fenicia@tudelft.nl

2 *Delft University of Technology, Delft, The Netherlands*

**Abstract** Hydrological applications are often unique. Each case study is different from the other, both because of the purpose of the application, and because of the variability of nature. Our tools, including models and diagnostic techniques, are often too rigid to adapt to each new requirement. This impacts our ability to address complex water-related problems both in engineering and in research. This paper shows the advantages of a flexible model structure in application to a case study. A variety of models are generated, which are applied to a headwater catchment in Luxembourg. The models are evaluated with an adapted GLUE methodology. GLUE has been often criticized for the subjective choices involved in its application, such as the selection of a discriminating threshold to separate "good" and "bad" models. Here we introduce a non-arbitrary flexible threshold which is automatically determined at a selected accuracy of prediction, with the objective of balancing predictive capability and parameter uncertainty. These tools aim at facilitating the understanding of the system's behaviour, which is essential to assess the impact of human activities on water and *vice versa*.

**Key words** hydrological models; FLEX; GLUE; Luxembourg

## 1 INTRODUCTION

A better understanding of the hydrological cycle is a research problem that has direct practical implications (Kovács, 1981). More specifically, problems associated with floods, droughts and water pollution require a better understanding of aspects related to storage dynamics, flow pathways and water transport. Hydrological models are important tools for the understanding and interpretation of the system behaviour, which ultimately serves water management and policy.

Hydrological models are useful when they provide the right answers for the right reasons (Kirchner, 2006). For this purpose, the choice of a hydrological model needs to be substantiated by meeting a number of requirements (Wagener *et al.*, 2001). It is desirable that the model provides a realistic representation of the system; it is necessary that model parameters are well identifiable; and it is necessary that model predictions are precise and accurate.

These requirements may not be necessary for all applications. However, in general, they improve the utility of models both in research and in practice. Essentially, they help in placing confidence in a given model, which can then be used both as a tool to interpret and understand catchment behaviour and as an instrument for planning and decision making.

What exactly characterizes a successful application of a conceptual model has been widely debated in the literature. The objective of model realism has motivated the development of physically-meaningful model applications (e.g. Atkinson *et al.*, 2002; Seibert & McDonnell, 2002; McDonnell *et al.*, 2007). The issue of parameter identifiability or "equifinality" has inspired several model diagnostic approaches (Gupta *et al.*, 2008) including the GLUE methodology (e.g. Freer *et al.*, 1996; Beven, 2006). The need for precise and accurate prediction has inspired different calibration frameworks ( Gupta *et al.*, 1998; Bates & Campbell, 2001).

Although there has been increasing attention towards these problems, we still lack the material and methods to build models that satisfy these main requirements. We struggle with the problem that each application is different from another, and therefore requires *ad hoc* solutions. This problem not only depends on the purpose of applications, it emerges from the intrinsic variability of nature (Beven, 2000).

The problem of "uniqueness of place" (Beven, 2000) indicates that it is necessary to be flexible in our approach and in the solutions to the modelling problem. Current model structures are often "fixed", *a priori* conceived representations of reality. Hence, they have little chance of being appropriate in a complex and heterogeneous world. Similarly, the approaches for model evaluation need to flexible enough to adapt to the requirements.

In this paper, we propose an application of a flexible model structure, to demonstrate its usefulness for understanding catchment behaviour. This model is based on the FLEX model of Fenicia *et al*. (2008). The model evaluation framework is an extension of the GLUE methodology. This methodology has been often criticized for the degrees of subjectivity that it involves. We propose a way to reduce this subjectivity by setting objective performance criteria, with the objective of balancing parameter uncertainty and predictive uncertainty.

The application of multiple model structures has an additional advantage over the application of a single model. It allows the user to learn more about the system's behaviour through the comparison of different model structures and their performance.

## 2 STUDY AREA

The study area is the Huewelerbach catchment in Luxembourg. The catchment area is 2.7 km$^2$, and its lithology is dominated by sandstone on top of an impermeable layer. The sandy soil layer is characterized by a high infiltration capacity, which allows the Huewelerbach to maintain a very stable baseflow regime. The riparian zone, however, is located on marls, and constitutes a fast runoff producing area. The dominant land use consists of forest and grassland.

Forcing data used for model evaluation are precipitation, potential evaporation and discharge. Evaporation is estimated based on temperature using the Hamon equation (Hamon & Belt, 1973). The data time step is 1 h and the calibration record runs from 1 July 2003 until 31 August 2005.

## 3 METHODOLOGY

### 3.1 Model description

The modelling framework is based on the FLEX model of Fenicia *et al*. (2008). We here develop a complex (redundant) structure, which is represented in Fig. 1. From this structure, it is possible to generate simpler model architectures as subsets of the more complex structure.

The model is characterized by five reservoirs. IR accounts for interception, RR represents a riparian zone or an impervious zone directly connected to the stream, UR is the unsaturated soil reservoir, FR accounts for the fast component of discharge, and SR accounts for baseflow.

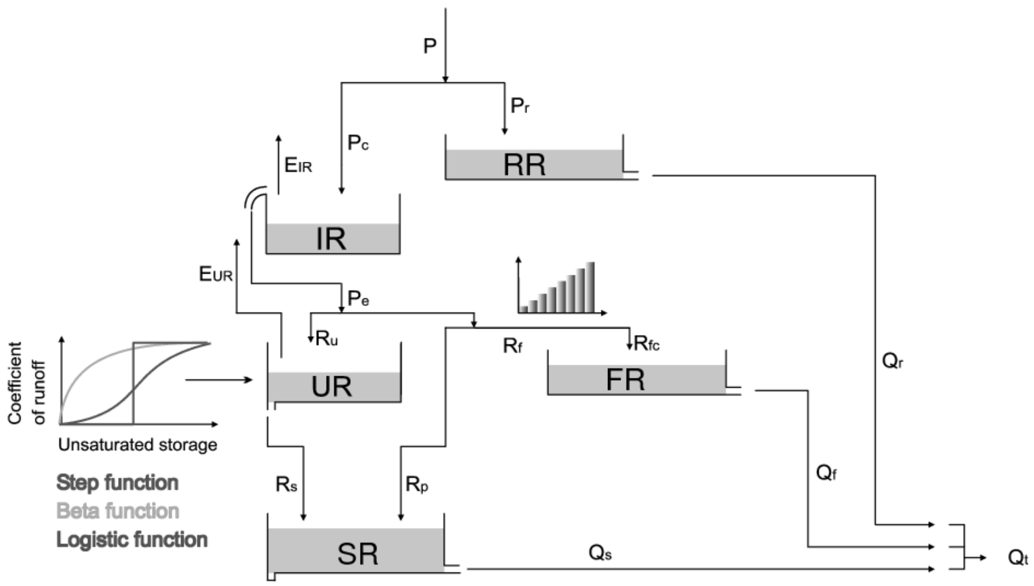


**Fig. 1** Schematic representation of the complete FLEX model structure from which simpler structures have been generated.

Precipitation $P$ is partitioned into a fraction, $f$, that goes to RR, and a fraction $(1 - f)$ that reaches IR, from which it can evaporate at a rate, $E_p$, as long as there is water available. Water that exceeds the threshold, $I_{max}$, is routed through UR, from which it is partitioned in what is stored in UR and what flows to other reservoirs. The partitioning is determined through a runoff coefficient $C_r$, which is a function of the storage in UR ($S_u$) and can assume different expressions. We considered a step function, a logistic function, and a power function. The step function represents a threshold behaviour, meaning that all water infiltrates into UR if $S_u$ is less than the maximum storage, $S_{u,\max}$, and it is routed to subsequent reservoirs otherwise. The logistic function can be interpreted as a smooth threshold. The power function is used in the HBV model (Lindstrom *et al.*, 1997) and in most of its derivations (among others, HyMod, Vrugt *et al.*, 2003; TAC, Uhlenbrook & Leibundgut, 2002). The energy that is not consumed in the interception process is available for transpiration from UR, which is moisture constrained through a parameter $L_p$.

The flux that does not infiltrate into UR is partitioned into $R_f$ which reaches FR and $R_s$ (representing preferential recharge) through a coefficient $D$. $R_f$ is convoluted through a triangular transfer function, and then routed through FR which can be a linear or nonlinear reservoir. SR and RR are both linear reservoirs. Model equations are summarized in Table 1.

By including or excluding different components, and changing constitutive relations, we have generated 16 different model structures, which are described in Table 2.

The model has been implemented using the explicit Euler modelling scheme, which is still predominant in hydrological applications. In future work we will reformulate the Flex model for using different, more accurate, time stepping schemes.

## 3.2 GLUE revisited

The GLUE methodology (e.g. Freer *et al.*, 1996; Beven, 2006) is based on the concept of *equifinality*, which encourages the acceptance of many feasible descriptions of reality. In practice, the GLUE methodology is based on the following steps. After specifying feasible ranges of model parameters, a high number of parameter sets is generated via uniform sampling. The performance of each trial is assessed through a likelihood measure (e.g. the Nash & Sutcliffe coefficient). Only

**Table 1** Description of model components.

| Equations | Description |
|---|---|
| $P_r = fP$ | Precipitation on RR |
| $P_c = (1 - f)P$ | Precipitation on the rest of the catchment |
| $E_{IR} = E_p$ (if $S_i > 0$), 0 (if $S_i = 0$) | Evaporation from IR |
| $P_e = P_c$ (if $S_{IR} = S_{I,\max}$), 0 (otherwise) | Effective precipitation |
| $C_r$ = case s: 0 (if $S_u < S_{u,\max}$), 1 (otherwise): case p: $(S_u/S_{u,\max})^{\beta p}$ case l: $\dfrac{1}{1+\exp\left(\dfrac{-S_u/S_{fc}+1/2}{\beta_l}\right)}$ | Coefficient of runoff |
| $E_{UR} = (E_p - E_{IR})\min(1, S_u/(Lp\, S_{u,\max}))$ | Transpiration from UR |
| $R_u = (1 - C_r)P_e$ | Infiltration into UR |
| $R_p = C_r D P_e$ | Preferential flow |
| $R_f = P_e - R_u - R_s$ | Flux to FR |
| $R_{fc} = R_f \times \mathrm{f}(N_{\mathrm{lag},f})$ | $R_f$ convoluted through the transfer function |
| $R_s = (S_u/S_{u,\max})P_{\max}$ | Percolation |
| $Q_r = K_r S_r$ | Outflow from RR |
| $Q_f = K_f S_f^{\alpha}$ | Outflow from FR |
| $Q_s = K_s S_s$ | Outflow from SR |
| $Q_{tot} = Q_r + Q_f + Q_s$ | Total discharge |

**Table 2** Schematic description of the 16 model structures. $N_{str}$ is the model structure identification number, $N_{par}$ is the number of parameters, IR–SR are the names given to the reservoirs, $C_r$ is the coefficient of runoff defined by a step (s), logistic (l) or power (p) function. The symbols v and - show the presence or absence of a component or parameter. Parameter $\alpha = 1$ when FR is linear.

| $N_{str}$ | $N_{par}$ | IR | RR | UR | FR | SR | $C_r$ | $I_{\max}$ | $L_p$ | $S_{u,\max}$ | $\beta_p$ | $\beta_l$ | $P_{\max}$ | $D$ | $N_{\text{lag},f}$ | $K_f$ | $\alpha$ | $K_s$ | $f$ | $K_r$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | - | - | v | v | - | s | - | v | v | - | - | - | - | v | v | - | - | - | - |
| 2 | 5 | - | - | v | v | - | s | - | v | v | - | - | - | - | v | v | v | - | - | - |
| 3 | 6 | - | - | v | v | - | l | - | v | v | - | v | - | - | v | v | v | - | - | - |
| 4 | 6 | - | - | v | v | - | p | - | v | v | v | - | - | - | v | v | v | - | - | - |
| 5 | 7 | v | - | v | v | - | l | v | v | v | - | v | - | - | v | v | v | - | - | - |
| 6 | 7 | v | - | v | v | - | p | v | v | v | v | - | - | - | v | v | v | - | - | - |
| 7 | 8 | - | v | v | v | - | l | - | v | v | - | v | - | - | v | v | v | - | v | v |
| 8 | 9 | v | v | v | v | - | l | v | v | v | - | v | - | - | v | v | v | - | v | v |
| 9 | 7 | - | - | v | v | v | l | - | v | v | - | v | v | - | v | v | - | v | - | - |
| 10 | 7 | - | - | v | v | v | l | - | v | v | - | v | - | v | v | v | - | v | - | - |
| 11 | 8 | v | - | v | v | v | l | v | v | v | - | v | - | v | v | v | - | v | - | - |
| 12 | 8 | - | - | v | v | v | l | - | v | v | - | v | v | v | v | v | - | v | - | - |
| 13 | 8 | - | - | v | v | v | p | - | v | v | v | - | v | v | v | v | - | v | - | - |
| 14 | 7 | - | - | v | v | v | s | - | v | v | - | - | v | v | v | v | - | v | - | - |
| 15 | 9 | v | - | v | v | v | l | v | v | v | - | v | v | v | v | v | - | v | - | - |
| 16 | 11 | v | v | v | v | v | l | v | v | v | - | v | v | v | v | v | - | v | v | v |

models (parameter sets and model structure combinations) that provide a likelihood measure reaching a minimum threshold are retained as "behavioural". The others are discarded as "non-behavioural".

The GLUE methodology has often been criticized for the subjectivity involved in the choice of the likelihood measure and the discriminating threshold (e.g. Montanari, 2005; Xiong & O'Connor, 2008). These studies have shown that both parameter uncertainty and model prediction bands are sensitive to the choice of the threshold values. These criticisms have motivated an extension of GLUE towards a "limits of acceptability" approach (Blazkova & Beven, 2009; Liu *et al.*, 2009), which, instead of applying a behavioural threshold to the likelihood measure, specifies an uncertainty band on the model predictions. Behavioural models are then those that provide predicted variables that fall within the limits of acceptability.

While this approach reduces the subjectivity of previous GLUE implementations (the limits of acceptability could be determined through an analysis of the rating curve), it is not difficult to anticipate some drawbacks from the adjustments that have been proposed. First, the limits of acceptability, which need to be specified prior to the use of a model, have to accommodate both input and output errors. This is a challenging operation, as the effect of input error (e.g. rainfall uncertainty) on the observed output (e.g. discharge) is often difficult to quantify. Second, the pattern of the observations may be more informative than the exact values. An error band that comprises the observed variables may hide important information.

In this application, we propose a different approach to limit GLUE's subjectivity. The approach consists of adopting a "moving" threshold to the likelihood measure, which can be adjusted depending on the model's ability to describe the observations (Fig. 2). Specifically, we first define an observation uncertainty band around the observed hydrograph, based on output error analysis. This step is similar to the definition of limits of acceptability around the observed variables. Subsequently, GLUE is applied and prediction uncertainty bands are evaluated based on the definition of a performance measure and a corresponding behavioural threshold. The main difference to GLUE is that this performance threshold is determined so that the interception between the observation and prediction uncertainty bands covers a predefined proportion of the observations (Fig. 2).

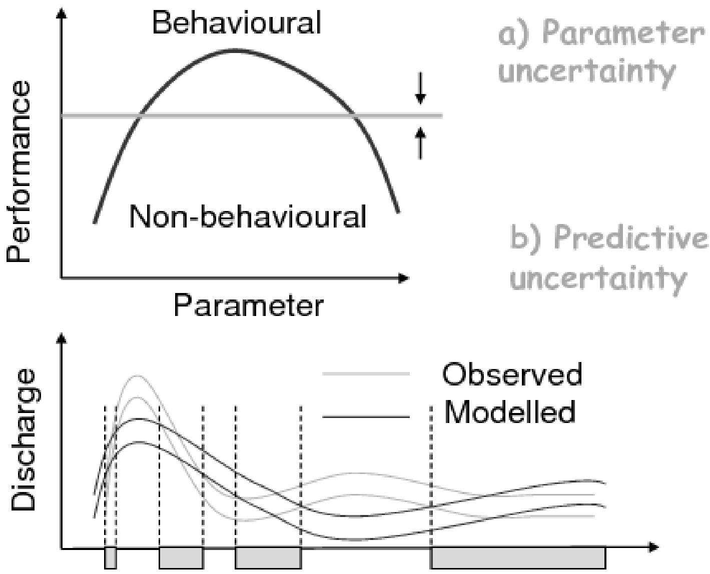


**Fig. 2** Schematic diagram of model evaluation approach.

## 4 RESULTS

We divide the results section into three subsections where we separately illustrate the outcomes of the study for the three basins with respect to: (i) accuracy measured by the selected objective function, (ii) uncertainty in model parameters and model response, and (iii) model realism. In the discussion section, the results are interpreted and discussed.

### 4.1 Model accuracy

Model accuracy with respect to hydrograph simulation is represented through the objective function $F_{NS}$, directly related to the Nash and Sutcliffe coefficient, $C_{NS}$:

$$F_{NS} = \frac{\sum_i \left(Q_{o,i}^T - Q_{m,i}^T\right)}{\sum_i \left(Q_{o,i}^T - \overline{Q}_{o,i}^T\right)} = 1 - C_{NS} \tag{1}$$

where $i$ is the current observation, the subscripts $o$ and $m$ stand for observed and modelled, $Q^T$ is discharge after an eventual transformation $T$, the overbar indicates an average over the observation period. The transformation $T$ applied to the discharge is as follows: $Q^T = \ln(Q + \varepsilon)$ where $\varepsilon = 10^{-3}$. The log-transformation enhances the error on low flows. Lower values of $F_{NS}$ indicate better performance.

Figure 3 shows the performance of the 16 model structures on the Huewelerbach catchment. It is generally believed that more complex models, disposing of more degrees of freedom, have a larger ability of fitting the data. Figure 3 demonstrates that this is not always the case, showing that parsimonious models can perform better than more complex ones. This underlines the importance of model concept over model complexity.

The performance of the models in the Huewelerbach catchment differs substantially. Here it clearly appears that models 1–6, all characterized by a “horizontal” structure, without an explicit description of the groundwater system, perform poorly. This shows that in this catchment it is necessary to specify a groundwater component. Models 7 and 8, which are characterized by a FR and a RR reservoir, perform better. However they are outperformed by subsequent models. In models 7 and 8, FR will act as a groundwater reservoir, while RR will tend to simulate the peaks. Structure 14 performs badly due to the threshold function adopted to describe the partitioning of rainfall in UR. It appears that in this catchment smoother functions work better.

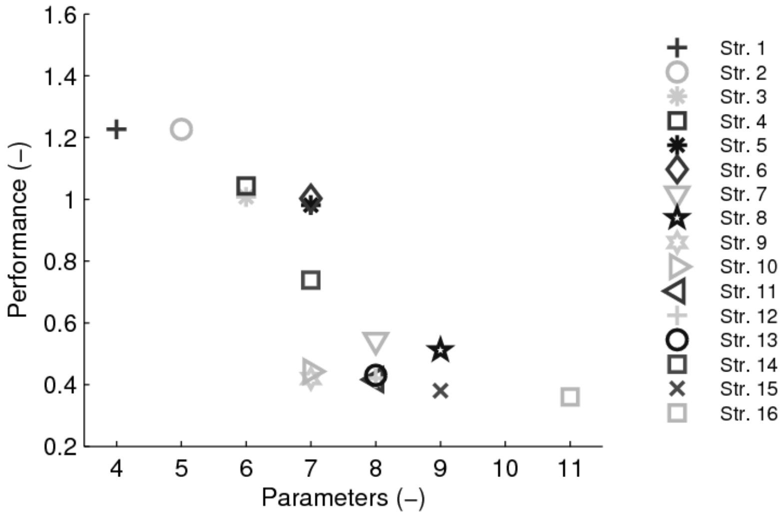


**Fig. 3** Comparison of model performance on the Huewelerbach catchment.

## 4.2 Parameter uncertainty

As illustrated in Section 3, we have quantified parameter uncertainty by scaling the likelihood threshold. This threshold was determined in such a way that the intersection between the observation discharge band and the simulation discharge band, obtained by calculating the 5–95% quantiles of the simulated discharge distribution, covers a certain percentage $P_Q$ of the observations. $P_Q$ has been set at 90%.

Figure 4 shows the cumulative distribution functions (cdf) of model parameters for the Huewelerbach catchment. The steeper the line of the cdf, the more identifiable is the corresponding model parameter. It is interesting to observe the performance of structures 1–8 in relation to parameter uncertainty (Fig. 4). These structures miss a groundwater component. As $F_{NS}$ is not so sensitive to the correct simulation of low flows, FR will tend to simulate the baseflow, acting as a groundwater reservoir. The steep lines in the subplot of $K_f$, which indicates the timescale of FR, correspond to these models. Models 7 and 8 perform better than previous models. These models include RR, which acts as a fast reacting reservoir, while FR continues to represent the baseflow component. The partitioning of rainfall in the two reservoirs is constant, depending on parameter $f$. The performance of these two models is worse than subsequent models, indicating that such a description of processes can be improved, as it will be explained later.

## 4.3 Model predictive uncertainty

The hydrograph uncertainty bands of selected model architectures on the Huewelerbach catchment are represented in Fig. 5. Structure 3 does not include a groundwater component. The single reservoir FR is not able to capture the dynamics of observed discharge in this catchment. As a result, the hydrograph uncertainty band is extremely large when compared to that of subsequent models. Structures 1 to 6 (not shown in the figure for reason of clarity) follow a similar behaviour. Structure 7 (Str. 8 has a similar behaviour) has a better performance. However, in general, it under-predicts peaks during wet conditions, and it over-predicts peaks during dry conditions. This is due to the constant partition of flow between a fast (RR) and a slow (FR) component, which is represented by the parameter $f$. A soil-moisture dependent flow partition, as represented by $C_r$, results in a better performance, as shown by structures 10 and 11. Structure 11 performs slightly better than Structure 10 (6%) due to the inclusion of an interception component, which has the effect of levelling out secondary peaks due to small isolated rainfall events.

Structure 16 seems to be overly complex, with parameter $K_r$ not well identifiable (Fig. 4). The hydrograph uncertainty band does not reduce sensibly, and model performance improves only slightly.

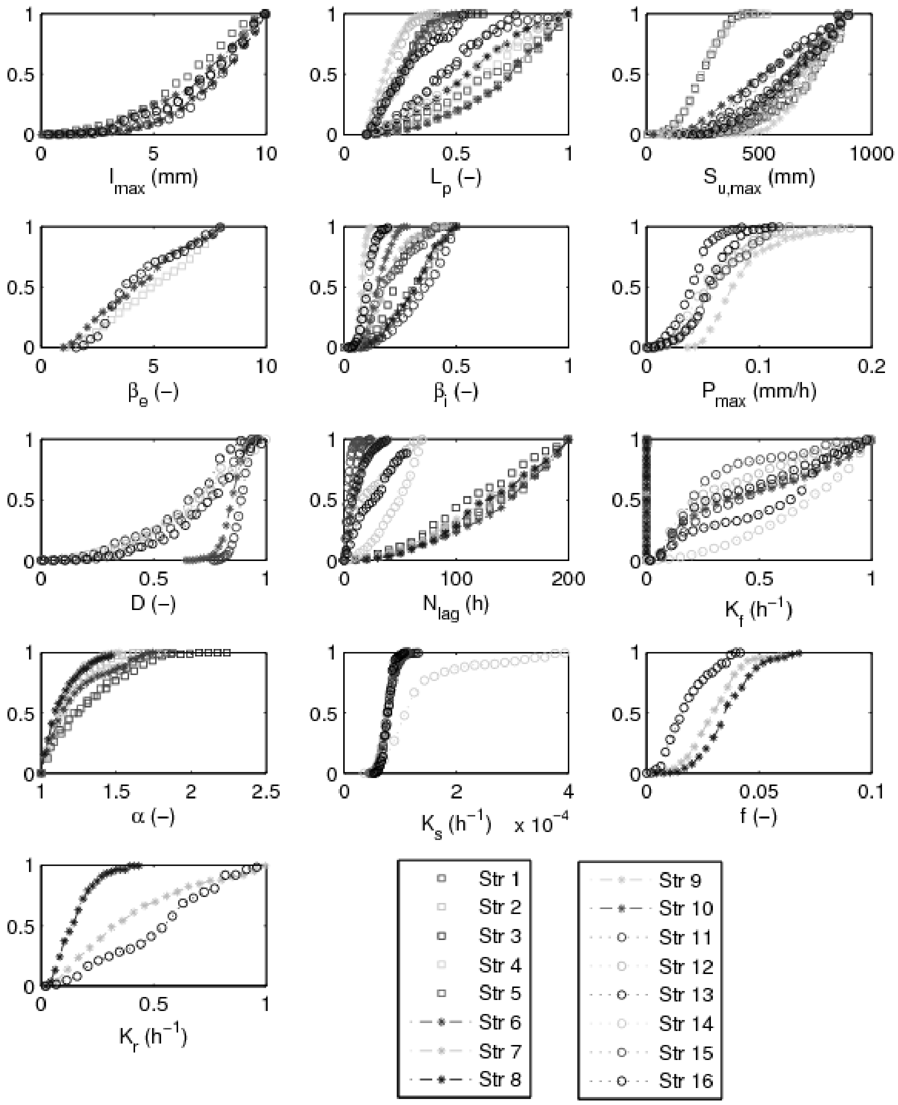


**Fig. 4** Parameter uncertainty for the Huewelerbach catchment.

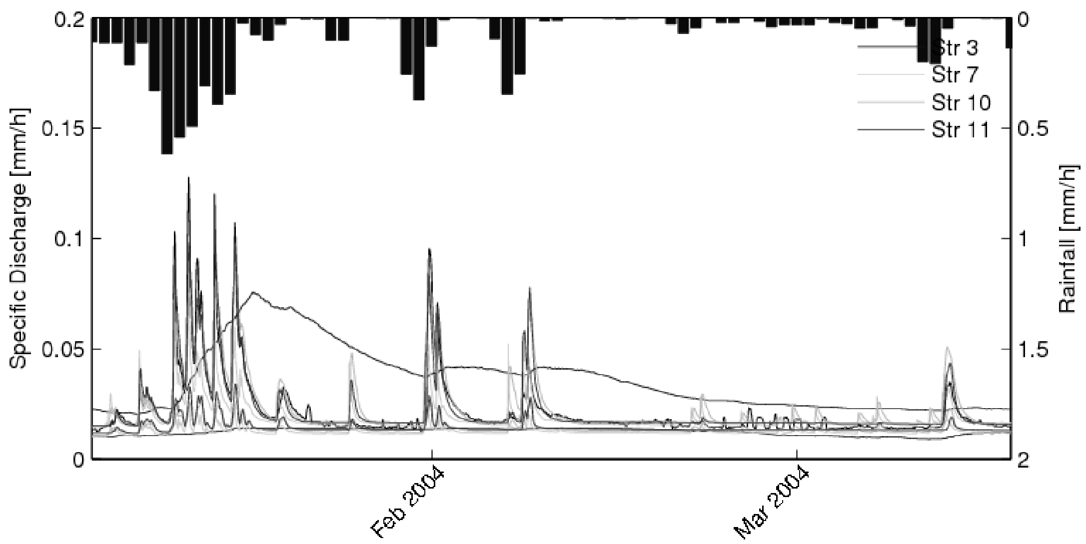


**Fig. 5** Comparison of the hydrograph uncertainty bands of selected model structures on Huewelerbach catchment.

## 5 DISCUSSION

The application of multiple models on a given catchment is an improvement with respect to the traditional practice of applying a single, preconceived model. First of all, through a flexible modelling approach it may be possible to fulfil what are perceived to be the main requirements that an appropriate modelling application should demonstrate (Wagener *et al.*, 2001). These are: (i) an appropriate balance between model complexity and data availability, which should result in an accurate and precise model output while providing well identifiable model parameters, (ii) a model structure which is realistic in terms of process representation, that is, which is in accordance with the perceived functioning of the catchment, and (iii) a model configuration that fits the purpose of the application. In addition, through the application of multiple model structure it is possible to have an overall view on the relative importance of different processes and components, which allows a better understanding of catchment behaviour.

The GLUE methodology has been often criticized for its subjective decisions (e.g. Montanari, 2005; Xiong & O'Connor, 2008). We propose some adjustments to this methodology, which reduce the subjectivity about the selection of a "behavioural" threshold to the likelihood function. These modifications allow more meaningful estimations of prediction limits and parameter uncertainty.

For the Huewelerbach catchment, the application of our methodology has allowed a better understanding of its behaviour. By adding and removing reservoirs, it was possible to appreciate the fundamental role of a groundwater component. The choice of different relations in the soil moisture component has allowed appreciation of the role of soil moisture on water partitioning between a quick and a slow flow component.

Overall, we believe that a move away from the single model–single catchment working paradigm may improve research in hydrology. In particular, it may help in catchment classification, as it may allow a better understanding in differences and similarities of the functional behaviour of different catchments.

## 6 CONCLUSIONS

This work deals with the application of 16 different model structures to a headwater watershed in Luxembourg. Through this exercise we show the importance of moving away from the "single model–single catchment" working paradigm. The application of multiple models allows identifying models that reflect a balance between their complexity and data availability. In addition, it favours a better understanding of catchment behaviour. These models were evaluated with a modified GLUE methodology, which allows a flexible selection of the behavioural threshold. This adjustment allows matching accuracy and predictive capacity in hydrological modelling. These tools are useful for an improved understanding of the system behaviour, and ultimately for water management and engineering applications.

## REFERENCES

Atkinson, S. E., Woods, R. A. & Sivapalan, M. (2002) Climate and landscape controls on water balance model complexity over changing timescales. *Water Resour. Res.* **38**(12), 1314.

Bates, B. C. & Campbell, E. P. (2001) A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall–runoff modeling. *Water Resour. Res.* **37**(4), 937–947.

Beven, K. (2000) Uniqueness of place and process representations in hydrological modelling. *Hydrol. Earth System Sci.* **4**(2), 203–213.

Beven, K. (2006) A manifesto for the equifinality thesis. *J. Hydrol.* **320**(1-2), 18–36.

Blazkova, S. & Beven, K. (2009) A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resour. Res.* **45**, W00B16.

Fenicia, F., Savenije, H. H. G., Matgen, P. & Pfister, L. (2008) Understanding catchment behavior through stepwise model concept improvement. *Water Resour. Res.* **44**(1), W01402.

Freer, J., Beven, K. & B. Ambroise, B. (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resour. Res.* **32**(7), 2161–2173.

Gupta, H. V., Sorooshian, S. & Yapo, P. O. (1998) Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resour. Res.* **34**(4), 751–763.

Gupta, H. V., Wagener, T. & Liu, Y. Q. (2008) Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrol. Processes* **22**(18), 3802–3813.

Hamon, W. R. & Belt, G. H. (1973) Energy balance-resistance model for computing evapotranspiration. *Trans. Am. Geophys. Union* **54**(4), 270–270.

Kirchner, J. W. (2006) Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* **42**(3), W03S04.

Kovács, G. (1981) *Seepage Hydraulics*. Elsevier Scientific Pub. Co., Amsterdam, The Netherlands.

Lindstrom, G., Johansson, B., Persson, M., Gardelin, M. & Bergstrom, S. (1997) Development and test of the distributed HBV-96 hydrological model. *J. Hydrol*. **201**(1-4), 272–288.

Liu, Y. L., Freer, J., Beven, K. & Matgen, P. (2009) Towards a limits of acceptability approach to the calibration of hydrological models: extending observation error. *J. Hydrol.* **367**(1-2), 93–103.

McDonnell, J. J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J. & Weiler, M. (2007) Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resour. Res.* **43**(7), W07301.

Montanari, A. (2005) Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall–runoff simulations. *Water Resour. Res.* **41**(8), W08406.

Seibert, J. & McDonnell, J. J. (2002) On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resour. Res.* **38**(11), 1241.

Uhlenbrook, S. & Leibundgut, C. (2002) Process-oriented catchment modelling and multiple-response validation. *Hydrol. Processes* **16**(2), 423–440.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. (2003) A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **39**(8), 1201.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. & Sorooshian, S. (2001) A framework for development and application of hydrological models. *Hydrol. Earth System Sci.* **5**(1), 13–26.

Xiong, L. & O'Connor, K. M. (2008) An empirical method to improve the prediction limits of the GLUE methodology in rainfall–runoff modeling. *J. Hydrol.* **349**(1-2), 115–124.