

Soft combination of local models in a multi-objective framework

F. Fenicia^{1,2}, D. P. Solomatine³, H. H. G. Savenije², and P. Matgen¹

¹Public Research Center – Gabriel Lippmann, Luxembourg

²Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft Univ. of Technology, The Netherlands

³UNESCO-IHE Institute for Water Education, Delft, The Netherlands

Received: 9 January 2007 – Published in Hydrol. Earth Syst. Sci. Discuss.: 19 January 2007

Revised: 30 July 2007 – Accepted: 25 October 2007 – Published: 22 November 2007

Abstract. Conceptual hydrologic models are useful tools as they provide an interpretable representation of the hydrologic behaviour of a catchment. Their representation of catchment's hydrological processes and physical characteristics, however, implies a simplification of the complexity and heterogeneity of reality. As a result, these models may show a lack of flexibility in reproducing the vast spectrum of catchment responses. Hence, the accuracy in reproducing certain aspects of the system behaviour may be paid in terms of a lack of accuracy in the representation of other aspects.

By acknowledging the structural limitations of these models, we propose a modular approach to hydrological simulation. Instead of using a single model to reproduce the full range of catchment responses, multiple models are used, each of them assigned to a specific task. While a modular approach has been previously used in the development of data driven models, in this study we show an application to conceptual models.

The approach is here demonstrated in the case where the different models are associated with different parameter realizations within a fixed model structure. We show that using a “composite” model, obtained by a combination of individual “local” models, the accuracy of the simulation is improved. We argue that this approach can be useful because it partially overcomes the structural limitations that a conceptual model may exhibit. The approach is shown in application to the discharge simulation of the experimental Alzette River basin in Luxembourg, with a conceptual model that follows the structure of the HBV model.

1 Introduction

Conceptual hydrological models consist of an ensemble of fluxes and storages representing relevant processes and key zones of catchment response. In the field of hydrological research, these models are useful tools for two main reasons. First, they are based on a reasonable representation of the major hydrological processes, which enables an interpretation of the real behaviour of the catchment. Second, their data requirement and computational demand is limited, which makes them easy to apply and to operate.

Conceptual models represent certain abstraction of reality, which results in a simplification of the complexity and heterogeneity of the real world. This simplification is justified as the complex process interaction at small scales can be represented by simple analytical approaches at larger scales (Sivapalan, 2003; Dooge, 2005). It has been suggested that this may be due to the self-organizing capacity of large systems (Savenije, 2001). However, it is often the case that simple models display a lack of flexibility in capturing the dynamic and time varying nature of hydrological responses (Wagener et al., 2003).

In order to improve model accuracy, one solution can be to develop the model further, in such a way that more processes are included (Fenicia et al., 2007). This approach, which has the advantage of enabling a better understanding of the system through a process of testing the effects of additional modelling assumptions, is time consuming and may be limited by our ability of understanding catchment behaviour through an analysis of its response.

A second possibility consists of using several models instead of one to better characterize the various conditions that influence the catchment hydrological behaviour. This approach, which is here investigated, is based on the idea that an integration of the results obtained by different models provides a more comprehensive and accurate representation of

Correspondence to: F. Fenicia
(fenicia@lippmann.lu)

catchment response than what can be obtained using a single model. The number works published on this topic while the discussion version of this article has been on line document the increasing interest in this approach (e.g. Marshall et al., 2006, 2007; Ajami et al., 2006, 2007; Vrugt and Robinson, 2007).

Multi-model approaches have been widely used in hydrological modelling in different frameworks and for different purposes. One objective is the estimation of conceptual model uncertainty. In this context, an ensemble of models is generated by multiple realizations from one or more model structures. Model simulations are eventually weighted or averaged or used to derive statistics of model outputs. The assessment of model uncertainty is the purpose of the GLUE framework (Beven, 1993; Beven and Freer, 2001), and of other approaches such as model and multi-model ensembles (Georgakakos et al., 2004; McIntyre et al., 2005). Most recently, approaches based on Bayesian model averaging (BMA) methods have been successfully applied in this field (Duan et al., 2007; Vrugt and Robinson, 2007; Ajami et al. 2007).

A second objective is the improvement of model accuracy. In this context, it is recognized that some models can be more accurate than others in reproducing different aspects of the system response. One possibility to take advantage of this aspect is to simulate the system response through models of different types, and use weighing procedures that attempt to retrieve the individual strengths of each model in simulating the system response. Following this approach, Shamseldin et al. (1997, 2007); Xiong et al. (2001); Abrahart and See (2002); Ajami et al. (2006); Duan et al. (2007), propose different combination methods to integrate the outcomes of different models. They show that in general the discharge estimates obtained by combining different models are more accurate than those obtained from any single model used in the combination. Recently, BMA methods also proved to be useful in this context (Duan et al., 2007; Vrugt and Robinson, 2007).

In order to improve model accuracy, instead of combining the outputs of models that aim at simulating the whole range of system response, it is possible use models that are directly built and calibrated on different event types or data sequences (Jordan and Jacobs, 1994; Zhang and Govindaraju, 2000; See and Openshaw, 2000; Hu et al., 2001; Hsu et al., 2002, Solomatine and Xue, 2004, Wang et al., 2006; Jain and Srinivasulu, 2006; Marshall et al., 2006, 2007; Corzo and Solomatine, 2007). In this approach, the distinctive role of different models in reproducing the system response is explicitly recognized from the beginning of the model development. See and Openshaw (2000) show the application of different neural networks built on different event types. Hsu et al. (2002) present a method of reproducing the catchment response through multiple linear local models which are built for specific flow conditions. Wang et al. (2006) used a combination of ANNs for forecasting flow: different networks were

trained on the data subsets determined by applying either a threshold discharge value, or clustering in the space of inputs (lagged discharges only but no rainfall data, however). Jain and Srinivasulu (2006) apply a mixture of neural networks and conceptual techniques to model the different segments of a decomposed flow hydrograph. Solomatine and Xue (2004) show an application of data-driven models M5 model trees and neural networks in a flood-forecasting problem, consisting of a combination of models locally valid for particular hydrologic conditions represented by specific regions of the input-output space. Corzo and Solomatine (2007) used several methods of baseflow separation, build different models for base and excess flow and combine these models ensuring optimal overall model performance. Marshall et al. (2006, 2007) introduced a framework known as hierarchical mixture of experts, where different models are applied at different times with a probability that depends on the hydrologic state of the catchment. The approach is similar to Bayesian Model Averaging (Duan et al., 2007; Vrugt and Robinson, 2007). However, in this case models may be developed specifically for different aspects of the catchment response (Marshall et al., 2007).

Approaches where different models are developed to perform similar modelling operations can be classified as “ensemble” strategies. The last approach corresponds to a “modular” strategy, as different models are developed to perform different tasks.

The approach introduced here can be attributed to the latter case. We in fact adopt a modular strategy based on the “fuzzy committee” approach (Solomatine, 2006) to characterize different aspects of a stream hydrograph. However, while previous works are based on purely data-driven models, the present work focuses on conceptual model structures and it is set in a multi-objective framework. The approach consists in calibrating a conceptual model with respect to different objectives (Gupta et al., 1998), representing model performance towards different aspects of the simulation, and in combining the best performing models associated to each objective in such a way that the strength of each individual model used in the combination is exploited. This approach attempts at improving the global accuracy of the simulation overcoming possible limitations in the model structures.

The approach is demonstrated using a conceptual model that follows the structure of the well-known HBV model (Lindström et al., 1997). The model is analysed with respect to its ability of reproducing the rainfall-discharge behaviour of a catchment in Luxembourg, with particular reference to accurate reproduction of the high and low flows behaviour. Multi-objective optimization with respect to two defined objectives representing model performances for the selected hydrograph characteristics shows that there are several solutions (the so-called “Pareto-optimal” set of solutions) that simultaneously optimize the selected criteria. These solutions represent a trade-off between the selected objectives and show that individual optimal models are better in

matching different aspects of the observed hydrograph.

The two best performing models associated with the selected hydrograph characteristics (in this case high flows or low flows) are subsequently weighted together using a fuzzy combining scheme. The paper concludes with a discussion on advantages, limitations, and physical significance of the proposed approach.

2 Problem formulation

In this work, we use the following definitions: “global” model is the model that aims at reproducing the full range of system response through a single description of reality; “local” model is the model that aims at reproducing a specific aspect of the system behaviour, which we call “event”; “composite model” is the model that provides the description of the full range of system response through a combination of local models; we call “model” both the structure and its realization through a given parameter set. The process of developing a “composite” model by means of aggregating multiple “local” models, each of which is specialized in simulating a certain aspect of the system response, can require a series of operations, summarized hereafter.

- *Events selection.* Within a modular approach, which presumes switching between different models, these events should correspond to different aspects of the system behaviour. Consequently, they should refer to different ranges or different time periods of a certain measured variable. As an example, Abrahart and See (2000) use a data decomposition based on season, Jain and Srinivasulu (2006) and Boyle et al. (2000) separate the hydrograph in different segments based on physical consideration on underlying processes, Corzo and Solomatine (2007) employ baseflow separation algorithms to differentiate between high and low flows. While the type and number of events may be based on physical considerations (e.g. Jain and Srinivasulu, 2006), it can also be performed through the help of automatic procedures such as Self Organizing Map models (e.g. Abrahart and See, 2000; Hsu et al. 2002) or model trees (e.g. Solomatine and Xue, 2004). In principle, the number of events should not be too high, in order to avoid a too fragmented description of the system response, which could also reduce the global efficiency for periods outside the calibration period.
- *Model selection.* The selected events could be represented by models of the same nature or of different nature (e.g. conceptual, physically based, data driven). As an example, Jain and Srinivasulu (2006) use conceptual and data driven models to simulate different segments of a flow hydrograph. They found that in the considered case study models of conceptual type performed better

than data driven ones in reproducing hydrograph recession.

- *Objective function definition.* Objective functions express the quality of the simulation in numerical form by aggregating model residuals in time. Different functions may enhance the error in simulating different aspects of the simulation while neglecting or downplaying the error in simulating other aspects. Since the use of a single objective function may result in a loss of information contained in the observed data (Gupta et al., 1998), the use of multiple functions in the assessment of model performance is becoming increasingly more popular.
- *Model calibration.* As model parameters most often do not refer to measurable quantities, they have to be inferred by calibration (Gupta et al., 1998). Hence, the local models associated with the different events have to be calibrated (or trained) to optimize the selected objective functions.
- *Model combination.* The local models are finally reintegrated into one composite model. Several combination techniques have been introduced in the literature. Shamsledin et al. (1997) were the first to analyze different combination methods to integrate the results of different models. They applied three different combination methods (the Simple Model Average method, the Weighted Average Method and the Artificial Neural Network method) to the outputs of five rainfall runoff models, reporting that the results of the model combination was superior to that of any single prediction. Subsequent studies analyzed and compared a variety of alternative combination techniques (Xiong et al., 2001; See and Openshaw, 2000; Abrahart and See, 2002; Solomatine, 2006; Ajami et al., 2006; Shamsledin et al., 2007). A general consensus of these works is that multi model predictions are superior to single model predictions. The advantage of one combination method with respect to another may depend on the application. Abrahart and See (2002), for example, determined that neural network combination techniques perform better for stable hydrologic regimes, while fuzzy probabilistic mechanism generated superior outputs for flashier catchments with extreme events.

2.1 Model structure description

The model used in this application is a lumped conceptual model that follows the structure of the HBV-96 model (Lindström et al., 1997), of which we keep the same list of symbols. In this study the model was run with an hourly time step. The model structure consists of routines for soil moisture accounting, runoff response, and a routing procedure (Fig. 1). The structure is composed of three storage components: a soil moisture reservoir, an upper reservoir, and a

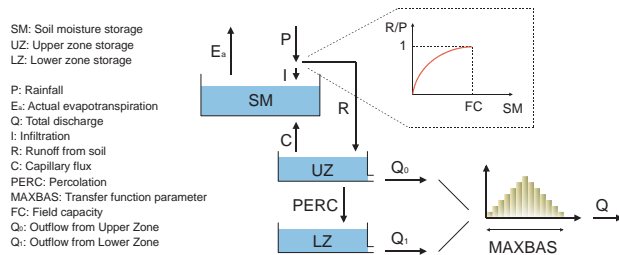


Fig. 1. HBV model schematic diagram

Table 1. Model parameters and corresponding units.

Parameter name	Description	Units
FC	Maximum soil moisture storage	mm
LP	Limit for potential evaporation	–
β	Non linear runoff parameter	–
PERC	Percolation rate	mm/h
CFLUX	Maximum capillary rate	mm/h
α	Non linear response parameter	–
K_1	Upper storage coefficient	mm/h
K_2	Lower storage coefficient	mm/h
MAXBAS	Transfer function length	h

lower reservoir. The output from the lower and upper reservoir is combined and routed through a triangular transfer function.

The soil moisture routine represents the runoff generation and involves three parameters, β , FC and LP . The proportion of precipitation that produces direct runoff is related to the soil moisture by the following relation:

$$\frac{R}{P} = \left(\frac{SM}{FC} \right)^\beta \quad (1)$$

Where P (mm/h) is the total rainfall, R (mm/h) is the direct runoff, SM (mm) is the storage of the soil moisture reservoir, FC (mm) is the maximum soil moisture storage, and $\beta(-)$ is a parameter accounting for non linearity. The remaining part is added to the soil moisture storage.

The model does not include the process of interception, and the transpiration from vegetation is combined with the evaporation from intercepted water into a total evaporation term. Actual total evaporation (E_a , mm/h) is calculated from potential total evaporation (E_p , mm/h) according to the following formula:

$$E_a = E_p \cdot \min \left(1, \frac{SM}{FC \cdot LP} \right) \quad (2)$$

Where $LP(-)$ is the fraction of FC above which the evaporation reaches its potential level. Direct runoff R enters the upper reservoir, and the lower reservoir is filled by a constant percolation rate (PERC, mm/h) as long as storage in the

upper reservoir is available. Capillary flux from the upper reservoir to the soil moisture reservoir is calculated according to the following equation:

$$C = CFLUX \cdot \left(1 - \frac{SM}{FC} \right) \quad (3)$$

Where the parameter CFLUX (mm/h) represents the maximum flux rate. Outflow from the upper reservoir is expressed as

$$Q_1 = K_1 \cdot UZ^{1+\alpha} \quad (4)$$

Outflow from the lower reservoir is expressed as

$$Q_2 = K_2 \cdot LZ \quad (5)$$

Where UZ (mm) and LZ (mm) are the storage states of the upper and lower reservoirs respectively, K_1 (mm/h) and K_2 (mm/h) are storage coefficient, and α is a parameter accounting for non linearity.

The outlets from the two reservoirs are finally added and routed through a transfer function with base defined by the parameter MAXBAS (h) (Fig. 1). The model has a total of nine calibration parameters, which are summarized in Table 1.

2.2 Events selection and objective functions

In the present application, we considered high flows and low flows as distinctive states of the system behaviour. Our aim was to accurately reproduce the system response during both events. In order to evaluate the performance of the “global” HBV model in both conditions, we used two objective functions, one enhancing the model error with respect to low flow simulation, and the other enhancing model error with respect to high flows.

The two functions are defined as follows:

$$N_{HF} = \sqrt{\frac{1}{n} \left(\sum_{i=1}^n (Q_{s,i} - Q_{o,i})^2 \cdot w_{HF,i} \right)} \quad (6)$$

$$N_{LF} = \sqrt{\frac{1}{n} \left(\sum_{i=1}^n (Q_{s,i} - Q_{o,i})^2 \cdot w_{LF,i} \right)} \quad (7)$$

Where:

$$w_{HF,i} = \left(\frac{Q_{o,i}}{Q_{o,max}} \right)^2 \quad (8)$$

$$w_{LF,i} = \left(\frac{Q_{o,max} - Q_{o,i}}{Q_{o,max}} \right)^2 \quad (9)$$

And:

- n : total number of time steps
- $Q_{s,i}$: simulated flow for the time step i
- $Q_{o,i}$: observed flow for the time step i
- $Q_{o,max}$: maximum observed flow

The two weighing functions w_{HF} and w_{LF} allow placing a stronger weight on the low or on the high portions of the hydrograph (Fig. 2). As a result, N_{LF} places a stronger weight on low flows errors and a weaker weight on high flows errors than N_{HF} . By computing both objective functions over the whole range of discharges, both functions constrain the model to fit the entire hydrograph.

3 Model calibration

The model is calibrated following a standard framework of multi-objective analysis which, for hydrological models has been introduced by Gupta et al. (1998). This framework adopts the notions of Domination and Pareto-optimality, which are hereafter recalled.

We use the term solution to mean a parameter set x_i . Each solution x_i is associated to a number of objective function values $N_j(x_i)$ ($j=1..m$, m =number of objectives), expressing the performance of the model. Lower values of $N_j(x_i)$ indicate better performance.

- A solution x_1 is said to dominate another solution x_2 when x_1 is better than x_2 in at least one objective (meaning $N_j(x_1) < N_j(x_2)$ for at least one value of j), and not worse than x_2 in any of the others (meaning $N_j(x_1) \leq N_j(x_2)$ for all values of j).
- The Pareto-optimal set of solutions is composed of those solutions that are not dominated by any solution of the feasible search space.

The outcome of a multi-objective problem in such a framework consists in the Pareto-optimal set of solutions. This set in general consists of more than one solution. The existence of more than one solution indicates that the objective functions are conflicting to each other, meaning that an optimal performance in one objective is “paid” in terms of sub-optimal performances in the others. This has been demonstrated adopting different models and various types of objective functions (e.g. Yapo et al., 1998; Boyle et al., 2000; Vrugt et al., 2003).

When applied to hydrological models, the existence of multiple optimal solutions can be related to a systematic component of the modelling error (Gupta, 1998), which is normally attributed to model structural inadequacies (Gupta et al., 1998, Vrugt et al., 2003). While it is plausible to think that other sources of error may contribute to this component, such as data distortion caused by incorrect rating curves or boundary conditions, it is also reasonable, when no other information is available, to put more confidence in the data than in the model, and therefore try to build models that represent the observations as correctly as possible. In this sense, the existence of multiple Pareto-optimal solutions can be regarded as a failure of the model structure. It in fact indicates that the model is not able to simultaneously represent the full

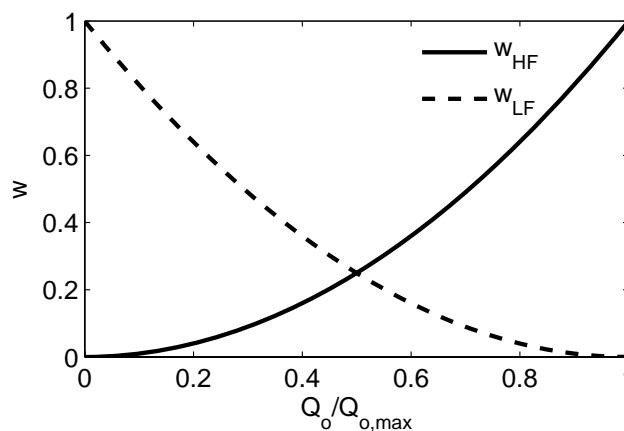


Fig. 2. Weight distribution for objective functions.

variability of catchment responses. All Pareto-optimal solutions are “equally important”, in a sense that it is difficult to prefer one solution over the other without any further information about the problem. The different solutions, however, are not “equifinal”, in the sense given to this term by Beven (1993). Every solution has its strengths and limitations in describing the different aspects of the observed system behaviour, as expressed by the selected objective functions. This observation can be used by trying to combine different optimal solutions in such a way that the individual strengths of each solution are exploited.

3.1 Combining scheme

Calibration of model parameters with respect to the two selected objective functions results in a set of Pareto-optimal solutions, which represents the performance of the global HBV model. The Pareto-optimal front is delimited by the two best models that minimize each of the individual objectives N_{LF} and N_{HF} . These local models were named HBV_{LF} and HBV_{HF} .

The two local models were then combined with an appropriate weighting procedure to generate a composite model HBV_C that aims at reproducing the whole range of discharges exploiting the best parts of each local model.

Several approaches to combine the individual outcomes of different models have been introduced. This combination can be straight-forward (Jain and Srinivasulu, 2006) implying a switch between different models at different time steps, but can also involve some kind of weighing (See and Openshaw, 2000; Xiong et al., 2001; Abrahart and See, 2002; Solomatine, 2006). Model weighing can improve simulation results, and avoid unrealistic discontinuities in the simulated system behaviour. The combining scheme that was used to weight the contributions of each local model makes use of a fuzzy attribution of weights, according to an approach termed “fuzzy committee” by Solomatine (2006). The approach is similar to the Takagi-Sugeno fuzzy combination

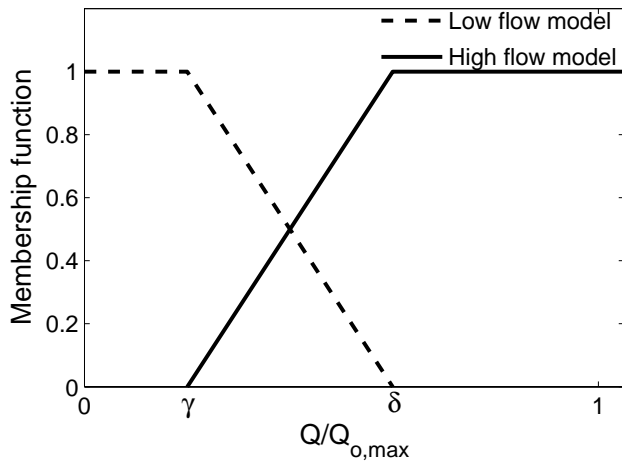


Fig. 3. Combining scheme for local models.

framework (Xiong et al., 2001). However, while in their study it is applied to integrate the results of a model “ensemble”, in this case it is used in a “modular” framework, that is, to integrate models that are specifically developed for different flow conditions. This approach has been selected among others because it suits the idea that the definition of a catchment “state”, which characterizes the type of runoff generation mechanisms, is very qualitative and uncertain.

The HBV_{LF} model is assumed to be accurate in simulating low flow events, but may be not accurate in simulating high flow events. Vice versa, the HBV_{HF} model is assumed to be more accurate during high flows than during low flows. In order to express this difference in the degree of believability of the outputs of the two models, each model was associated with a certain membership function. The degree of membership associated with the low flow model is 1 when the simulated flow is below the threshold γ , it decreases linearly when the flow is between the thresholds γ and δ , and it is 0 when the flow is above the threshold δ (Fig. 3). The degree of membership of the high flow model follows a symmetric behaviour. Membership functions for the two local models are described in equations 10 and 11; γ and δ were named threshold for high flows and for low flows respectively and are expressed as a fraction of the maximum observed discharge $Q_{o,max}$.

$$m_{LF}(Q) = \begin{cases} 1, & \text{if } Q/Q_{o,max} < \gamma \\ 1 - \frac{Q/Q_{o,max} - \gamma}{\delta - \gamma}, & \text{if } \gamma \leq Q/Q_{o,max} < \delta \\ 0, & \text{if } Q/Q_{o,max} \geq \delta \end{cases} \quad (10)$$

$$m_{HF}(Q) = \begin{cases} 0, & \text{if } Q/Q_{o,max} < \gamma \\ \frac{Q/Q_{o,max} - \gamma}{\delta - \gamma}, & \text{if } \gamma \leq Q/Q_{o,max} < \delta \\ 1, & \text{if } Q/Q_{o,max} \geq \delta \end{cases} \quad (11)$$

The outputs of the two models were finally combined according to Eq. (12). This weighing approach allows a smooth

combination of the two models, and avoids discontinuities in the reproduction of the system response.

$$Q_C(t) = \frac{m_{LF}(Q_{LF}) \cdot Q_{LF}(t) + m_{HF}(Q_{HF}) \cdot Q_{HF}(t)}{m_{LF}(Q_{LF}) + m_{HF}(Q_{HF})} \quad (12)$$

Where Q_{LF} and Q_{HF} are the outputs of the local low flow (HBV_{LF}) and high flow (HBV_{HF}) models, and Q_C is the output of the composite (HBV_C) model. Note that the weighting schemes shown on Fig. 2 and 3, at first sight similar, serve different purpose: the first one is used to stress low/high flows in the objective function used to calibrate two separate models, and the second one is responsible for ensuring smooth compatibility between the models.

4 Case study

4.1 Study area and data description

The study area is within the experimental Alzette river basin, which is located for the large part in the Grand-Duchy of Luxembourg. For model calibration, three years of hourly data from Hesperange, a gauging station placed along the course of the Alzette River upstream of Luxembourg-city, were used.

Catchment size is 288 km², and land cover is composed of cultivated land (27%) grassland (26%), forest land (29%) and urbanized land (18%). Lithology is mainly represented by Marls and Marly-Sandstones on the left bank tributaries and Limestones on the right bank tributaries of the Alzette River.

The rainfall-runoff behaviour of these units is quite different. Marls areas are characterized by impermeable bedrock, therefore rainfall water, after losses for evaporation or transpiration, reaches the stream mostly as saturated subsurface flow that develops at the interface between the weathered zone and the underlying bedrock areas. When the weathered zone becomes saturated, or during heavy rainfall events, surface runoff occurs.

In limestone areas a large part of rainfall water infiltrates and after subtraction of losses percolates to the groundwater aquifer, which is capable of storing and releasing large quantities of water. The response to rainfall of Marl areas is faster and characterized by larger volumes of water than that of limestone areas. Moreover, the large part of the base-flow during prolonged dry periods is mostly sustained by the limestone aquifer.

The basin is instrumented by several rain gauges including tipping-buckets and automatic samplers measuring at a time step which does not exceed 20 min. Hourly rainfall series were calculated by averaging the series at the individual stations with the Thiessen polygon method. Daily potential evaporation was estimated through the Penman-Monteith equation (Monteith, 1965). The meteorological variables needed to compute the evaporative loss were measured at

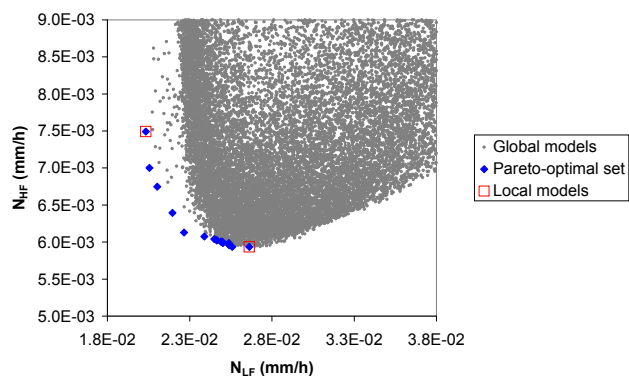


Fig. 4. Scatter plot of model simulations in the objective space.

the Luxembourg airport meteorological station. Hourly estimates were then calculated distributing the daily amounts through a sine function. The use of an hourly time step is justified considering that the average concentration time of the basin is of the order of a few hours.

4.2 Multi-objective calibration

The global HBV model was calibrated according to the two objectives N_{HF} and N_{LF} , previously defined. The problem was posed in a multi-objective framework and solved by determining the Pareto-optimal set of solutions. In order to efficiently sample the parameter space, the Multi-Objective Shuffled Evolution Metropolis University of Arizona (MOSCEM-UA) algorithm has been used (Vrugt et al., 2003).

The MOSCEM-UA algorithm begins with a random sequence of s points sampled throughout the feasible parameter space. For each point the set of objective functions is evaluated. The sequence is partitioned into q complexes, and in each complex a parallel sequence is launched. New candidate points from each complex are generated from a multivariate normal distribution with mean at the current draw of the sequence and covariance matrix calculated based on the history of each sequence. The sequences evolve based on a Metropolis-type acceptance criterion. The algorithm proceeds until a specified maximum number of iterations m is reached.

The MOSCEM-UA has three algorithmic parameters that have to be specified by the user: s , q , and m . No theoretic guidelines exist in determining these parameters; however, a good criterion is to use a number of complexes that is at least equal to the number of parameters.

Parameter bounds were determined by analysing the results of an initial run of the algorithm on a wide parameter space. The selected parameter bounds are reported in Table 2. The algorithm was run with the following parameters: $s=2000$, $q=10$, $m=50\,000$.

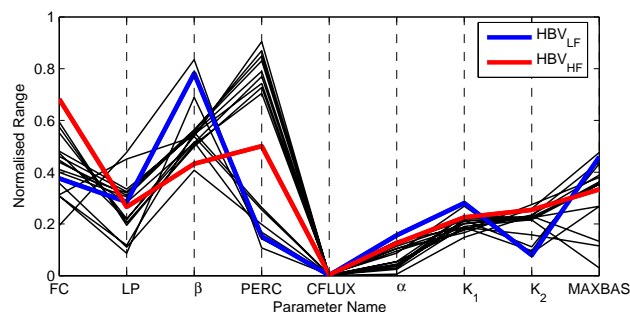


Fig. 5. Normalized parameter plot. Pareto-optimal solutions from Fig. 4 are shown. The two thicker lines represent the local models, which are the optimal solutions with respect to each objective function (N_{LF} and N_{HF}).

The outcome of the optimization algorithm is presented in Fig. 4 and Fig. 5. Figure 4 shows the objective function values corresponding to the evaluated parameter sets together with the set of Pareto-optimal solutions and the optima corresponding to each individual objective. This plot clearly illustrates a trade-off in the selected objectives, and reveals the inability of the model to match the selected aspects of the hydrograph simultaneously.

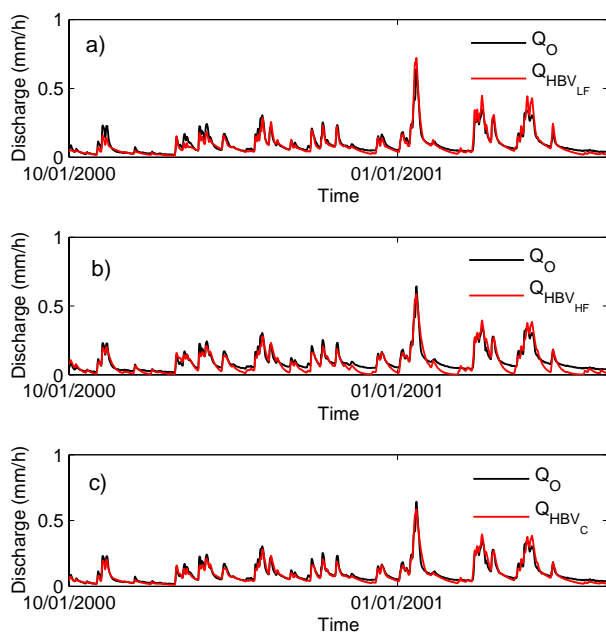
The variation of the Pareto-optimal parameter sets is shown in Fig. 5. The parameter values are normalised with respect to the upper and lower bounds given in Table 2, so that the feasible range of all parameters is between 0 and 1. Each line on the plot represents one parameter set. The figure gives a visual indication on the relation between the initial feasible parameter range, and the parameter range that corresponds to the optimal solutions. Conclusions about large or small variability of parameter values would not be meaningful, as the extent of the optimal range displayed in the figure clearly depends on the initial lower and upper limits that are selected.

While moving from one solution to another on the Pareto-optimal front (Fig. 4), the corresponding parameter values may show a trend from one extreme to another (Fig. 5). The existence of such a trend reveals potential deficiencies in the model structure. This behaviour is significant for the parameters β , which accounts for non linearity of the rainfall-direct runoff relation, and K_2 , representing the storage coefficient of the lower reservoir.

As an example, the parameter K_2 shows higher values when calibrated towards the high flows, and lower values when calibrated towards the low flows. If we assume that the correct value for this parameter is what corresponds to the low flow calibration (as this parameter determines the behaviour of the lower reservoir which mostly affects low flow simulation), we can conclude that the calibration of the model with respect to high-flows will result in overestimating this parameter. As a result, the lower reservoir in the optimal high flow model will empty faster than it should in order to compensate for errors in the simulation of other processes.

Table 2. Parameter ranges for model optimization and optimal values for local models.

Parameter name	Units	Lower Bound	Upper Bound	HBV _{LF}	HBV _{HF}
<i>FC</i>	mm	200	450	2.94 E+02	3.71 E+02
<i>LP</i>	–	0.01	1	2.92 E-01	2.73 E-01
β	–	0.01	2	1.57 E+00	8.71 E-01
PERC	mm/h	0.01	1	1.60 E-01	5.06 E-01
CFLUX	mm/h	0	0.1	1.77 E-04	3.56 E-04
α	–	0	0.5	7.81 E-02	6.28 E-02
K_1	mm/h	0.001	0.1	2.88 E-02	2.33 E-02
K_2	mm/h	0.001	0.1	8.89 E-03	2.62 E-02
MAXBAS	h	7	15	1.07 E+01	9.67 E+00

**Fig. 6.** comparison (a) local low-flow model; (b) local high-flow model; (c) composite model.

This behaviour is also evident on the hydrographs presented in Fig. 6a and b. Figure 6b shows that the best performing model with respect to N_{HF} (HBV_{HF}) is characterized by steeper recessions than observed, while Fig. 6a, representing the best model with respect to N_{LF} (HBV_{LF}), shows a better agreement with the observations during recession periods. The parameter sets corresponding to the two local models HBV_{HF} and HBV_{LF} are shown in Table 2.

4.3 Local models combination and “composite” model

The local models were combined according to the procedure described in Sect. 2.4. The combining scheme aims at integrating the strengths of each individual local model in reproducing specific characteristics of the simulation and

it requires the selection of two discharge thresholds: γ and δ (Fig. 3). These two thresholds can be selected based on knowledge of the system behaviour, or can be assessed automatically to minimize the error of the composite model.

Manual selection of thresholds could be based on the ground of a physical understanding of the behaviour of the catchment. In this case, the thresholds could represent switches in catchment behaviour that correspond, for example, to changes in contributing areas related to catchment storage, in channel properties related to the water level in the stream. This evidence was not the case of this study, therefore the thresholds have been initially selected by a visual inspection of model performances across the range of observed discharges. A procedure to perform automatic tuning of the thresholds is described in the next paragraph.

Analysis of the observed rainfall and flow led us to a conclusion that it would be reasonable to choose the following thresholds for flow: $Q=0.12$ mm/h for high flows and $Q=0.07$ mm/h for low flows. As the maximum discharge in the calibration period is 0.64 mm/h, this results in $\gamma=0.11$ and $\delta=0.17$. Performances of the composite model with respect to the hydrograph simulations are represented in Fig. 6. Figure 6a shows the performances of the low flow local model, Fig. 6b shows the performance of the high flow local model, and Fig. 6c shows the performance of the composite model developed from the combination of the two local models. It is possible to observe visually that the composite model incorporates the best features of both local models, considerably improving the overall accuracy.

Model performance in term of the calibration objectives is presented in Fig. 7. The solution corresponding to the composite model lies beyond the Pareto-optimal set, showing that the composite model improves the accuracy of the simulations. Figure 8 shows the performance of the Pareto-optimal global models and of the composite model with respect to indicators of more common use in hydrological modelling. Specifically, we compare the performance of the models with respect Nash and Sutcliffe coefficient (N_{NS}) and with respect to the Root Mean Square Error (N_{RMSE}), which is given by

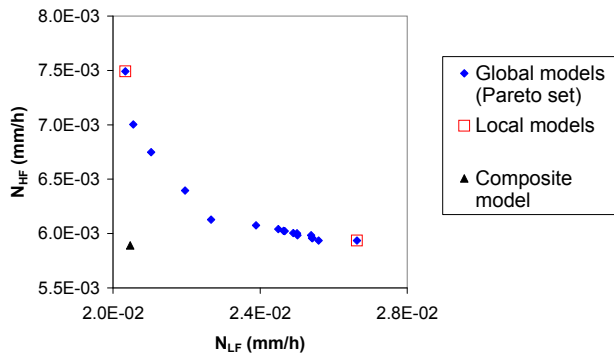


Fig. 7. Comparison of the performances of global, local (HBV_{HF}, HBV_{LF}) and composite (HBV_C) models.

Eqs. (6) or (7) for a weight w of one. It is possible to note that the composite model outperforms the global models also with respect to these objective functions.

It is also noteworthy that the performance of some Pareto-optimal models is higher than that of both single best solutions corresponding to each individual criterion. This is due to the fact that the selected objectives calculate the error with respect to the whole range of flows, even with different weights. An improvement in low flow description, for example, given the same performances in high flows, reduces the total error as calculated by N_{HF} .

4.4 Automatic tuning of the thresholds

When no evidence exists in determining the thresholds corresponding to changes in the system behaviour, these thresholds can be calculated by trying to maximise the performance of the model. With this purpose, a sequence of thresholds was generated on a grid in the (γ, δ) space, and the Pareto-optimal set of solutions corresponding to different values of the thresholds has been calculated. Results are represented in Fig. 9. It is possible to observe that even the employed simple type of search improves the global model accuracy.

The thresholds values corresponding to the Pareto solutions are represented in Fig. 10. With respect to the manually selected values, the Pareto values are smaller for the low flow threshold γ , and larger for the high flow threshold δ , which enlarge the area where both models are evaluated and weighted.

5 Discussion

The synthetic view of reality that is incorporated in conceptual hydrological models often does not allow a simultaneous optimal representation of different aspects of the system behaviour. To overcome this problem, a modular approach to hydrologic simulation has been presented. This approach allows for different models to operate simultaneously, each of them developed to reproduce a specific aspect of the system

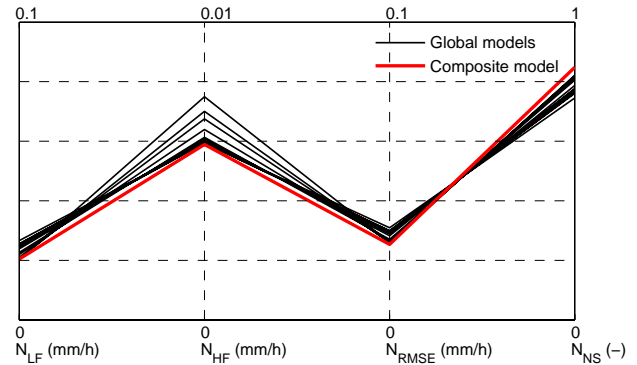


Fig. 8. Comparison of the performances of global and composite models with respect to additional objective functions. The composite model outperforms any global model.

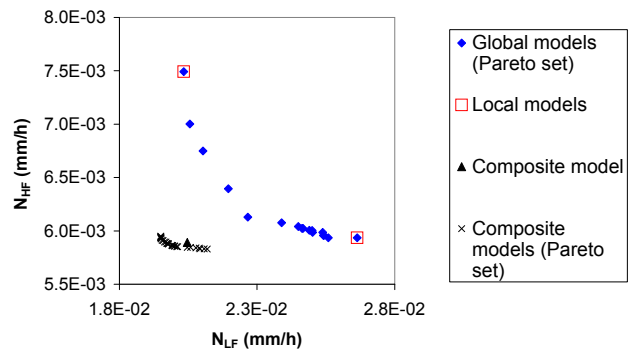


Fig. 9. Comparison of the performances of global models, local models (HBV_{HF}, HBV_{LF}), composite model (HBV_C) with manually selected thresholds, and composite models with automatic tuning of thresholds.

behaviour. The various models are then combined through an appropriate weighing procedure, which produces a composite representation of the catchment behaviour. The combining scheme exploits the strengths of each individual model in a synergistic manner.

The presented method allows for different parameter sets of a fixed model structure, but, in principle, could be applied allowing for different model structures too (e.g. conceptual, physically based, data driven). Specifically, we build separate models for high flow and low flow simulation, which are subsequently combined through a soft combination approach. This approach follows the Takagi-Sugeno fuzzy combination framework (Xiong et al., 2001). However, in contrast to Xiong et al. (2001), the procedure is here applied to combine models that are explicitly derived to express different aspects of the system behaviour. Results show that the composite model reaches a higher overall accuracy than what can be obtained using any global model.

The “modular” approach has been increasingly applied to various hydrological problems in the past decade. Mostly, this approach has been applied in the development of purely

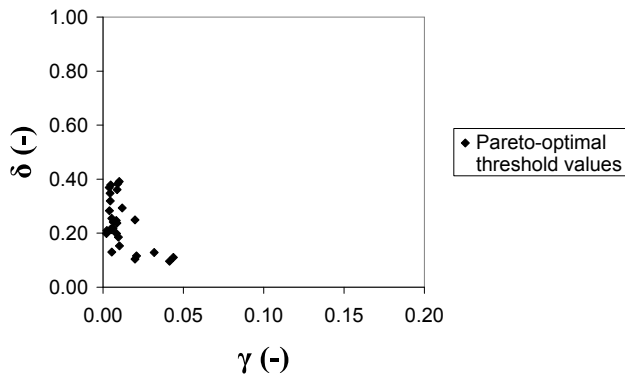


Fig. 10. Combination of thresholds corresponding to the Pareto-optimal composite models.

data-driven models (Jordan and Jacobs, 1994; Zhang and Govindaraju, 2000; See and Openshaw, 2000; Abrahart and See, 2002; Hu et al., 2001; Hsu et al., 2002, Solomatine and Xue, 2004, Corzo and Solomatine, 2007). Recently, its potential has been explored in the field of conceptual modelling (Marshall et al., 2007). In our work we use this approach to overcome the limitations of an HBV-type conceptual model in reproducing simultaneously various aspects of the system response. The advantage of using conceptual versus data-driven models is that conceptual models provide an interpretable representation of reality, which may be more difficult to obtain with data-driven models.

In contrast to Marshall et al. (2007) where the model weighing process is set in a Bayesian framework, we here adopt a fuzzy approach to combine model predictions. Strict Bayesian approaches in fact require assumptions on data errors that are difficult to justify in practical applications. Moreover, while Marshall et al. (2006, 2007) show the benefit of integrating different outcomes of a three parameter conceptual model, which due to its under parameterization has an obvious lack of predictive capability, we here use a nine parameter version of a well known conceptual model, which has an higher potential to adapt to the observations. Our intention is in fact to demonstrate how a modular approach can help to improve the level of accuracy of a model formulation which is commonly used in hydrologic applications.

While the practical utility of a multi model approach as here proposed relies on an improvement of the simulation accuracy, the physical implications involved require interpretation and justification. The switching between different models, in fact, implies an alternation between different views or descriptions of reality. If the natural system is not modified by natural phenomena or artificial activities, it may seem physically inconsistent to represent it by means of separate descriptions.

The hydrological processes involved in the rainfall-runoff transformation, however, are extremely complex and characterized by a high degree of spatial and temporal variability.

Despite at large scale catchment behaviour can be expressed by simple laws (Savenije, 2001; Sivapalan, 2003; Dooge, 2005), there may be several factors influencing catchment response that a model concept does not represent. As a result, the catchment response can be visibly different from its model representation. The different “personalities” that a catchment may display are determined by a series of phenomena and processes that can be in general identified, but that is difficult to separate and quantify.

The different interacting causes of variability in hydrological behaviour include:

- *Seasonality effects.* As a consequence of vegetation or biologic activities, aspects like land cover or macropore distribution in the top-soil vary, affecting processes such as interception, infiltration, pathways of water on the soil surface and in the weathered zone. As a result of the changing catchment properties, the catchment response may be strongly season dependent.
- *Environmental forcing conditions.* Forcing conditions influence the amount and distribution of water in the soil, determining the catchment hydrological “state” which drives catchment response. With changes in hydrological states, such as from low to high flow or from dry to wet conditions, the compartments of the catchment that contribute to discharge (e.g. saturated and unsaturated zone, near stream saturated areas) vary dynamically, leading to different domains of formation and integration of the hydrological processes.
- *Non linear behaviour.* The occurrence of several hydrological processes is characterized by highly non linear, threshold-like behaviour. Groundwater levels, rainfall intensities, soil moisture conditions control the occurrence of processes such as surface runoff or rapid subsurface flow, and can trigger the contribution of different compartments of the catchment.

A multi-model approach can implicitly take into account the variability in hydrologic behaviour that is not explicitly considered in the realization of a single model structure. By allowing different models to operate for the simulation of different aspects of the system response it is implicitly recognized that a single model cannot explain by itself the full variability of catchment responses. In the specific case, where the different models are represented by individual parameter sets, it can be assumed that model parameters, depending on the particular stage of the simulation, describe different behaviour of the catchment expressing different processes.

In conceptual modelling it is typically assumed that model parameters, if not physically based or clearly related to catchment attributes, are representative of inherent properties of the catchment, and therefore not supposed to vary (Wagner et al., 2003). However, model parameter may have different values depending on aspects such as the length of the calibration period or the performance measure used for calibration.

A reason for this could be non-stationarity of errors in forcing conditions such as rainfall. In absence of information on data errors, we here regard this problem as a symptom of model structural errors.

Understanding where the model fails, and where the catchment shows a certain “personality” that is different than what is estimated a-priori, can guide towards a better understanding of the system. When building a model, in fact, we use a possible representation of the most relevant processes and their interrelation. The analysis of the performance of the model represents a possibility to test the hypotheses made. In this sense, identifying a switching between different states can clarify triggers or thresholds in the catchment behaviour, helping to guide model refinement and providing new understanding that could be a base for field research.

It can be concluded that the presented approach can be seen as an effective way to improve model accuracy by representing different aspects of the system behaviour by differently parameterized models. The multi-objective framework makes it possible to perform the detailed analysis of the models’ performance and to construct an optimal model structure. The use of a “fuzzy committee” allows for soft combination of local models and prevents discontinuities between the model predictions. The approach is quite universal and can be used to combine different types of models, from conceptual to data-driven ones.

We remain confident that the ultimate goal of hydrologic research has to be the understanding of catchment behaviour. Multi-model approaches may be useful in practical applications as they improve accuracy in model predictions, and in research to represent and clarify the non linear behaviour of catchment response. However, they do not have to be considered as a shortcut to bypass process knowledge and conceptual thinking. When this is the case, they should be regarded as black box approaches that yield little increased understanding on the physical system.

6 Future research

The proposed approach involves several subjective decisions that influence the results of the analysis. As described in Sect. 2, these decisions include the selection of events that characterize different aspects of the system behaviour, the choice of the model structures that are used to simulate the selected events, the choice of objective functions and calibration strategy, and the choice of a combining scheme that aggregates the local models into a composite model.

Ideally, the effects of every choice should be justified based on an argument that proves its value with respect to other alternatives. Clearly, this is a difficult task, as it expands enormously the number of possibilities to be investigated. However, it would be interesting to compare this approach to others (e.g. Ajami et al., 2006; Marshal et al., 2006, 2007; Ajami et al., 2007; Vrugt et al., 2007), and evaluate the

strengths and limitations of each. This requires the effort of a comparative work, which could be an objective for future research.

In order to improve the proposed approach, a first challenge is to complement it with algorithms aimed at discovering various regimes in the time series representing the modelled system; this would allow for optimal combination of domain (hydrologic) knowledge incorporated in models with the automatic machine learning or time series analysis routines. A second challenge is to implement different states of catchment behaviour directly within the model structure, in order to obtain a comprehensive description of the overall catchment behaviour within a single representation of reality. The switching between different models causes in fact a loss of continuity between model internal states. This however, could complicate the application of the approach.

7 Conclusions

This paper presents a modular approach to overcome the limitations displayed by simple conceptual models in reproducing simultaneously different aspects of the system response. Modular approaches have been already introduced in the literature. However, previous studies focused on the development of purely data-driven models, while in this paper we evaluate the approach in the field of conceptual modelling.

We considered an HBV-type conceptual model, and we evaluated its ability of reproducing high flow and low flow behaviour of the catchment. We performed a multi-objective optimization between the selected calibration objectives, and we showed the limitations of the model in capturing simultaneously both aspects of the system response. Subsequently, we identified the two models that optimize each calibration objective, and we combined them through a fuzzy weighing approach. We showed that the “composite” model, obtained by a combination of individual “local” models, outperforms any “global” model, which aims at reproducing the whole range of catchment response through a single parameter set.

The approach is straight forward, easy to apply, and useful for problems that require an accurate simulation of the catchment response. We also think that the use of different models to reproduce a single system may have its physical justification, as the catchment may display different “personalities” during various regimes, which are difficult to capture within a single conceptualization of reality.

The approach has its limitations, which may be object of further investigation. In particular, the approach involves several subjective decisions whose effect has not been examined. Some of these decisions, such as the identification of different regimes of catchment behaviour, could be helped by the use of automatic procedures. The approach could be compared to others to assess its relative value. These aspects can be object of future research.

Edited by: R. Abrahart

References

- Abrahart, R. J. and See, L.: Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments, *Hydrol. Process.*, 14, 2157–2172, 2000.
- Abrahart, R. J. and See, L. M.: Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments, *Hydrol. Earth Syst. Sci.*, 6, 655–670, 2002, <http://www.hydrol-earth-syst-sci.net/6/655/2002/>.
- Ajami, N. K., Duan, Q., Gao, X. and Sorooshian, S.: Multi-model combination techniques for hydrological forecasting: Application to Distributed Model Intercomparison Project results, *J. Hydrometeorol.*, 7, 755–768, 2006.
- Ajami, N. K., Duan, Q. and Sorooshian, S.: An Integrated Hydrologic Bayesian Multi-Model Combination Framework: Confronting Input, Parameter and Model Structural Uncertainty in Hydrologic Prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745, 2007.
- Beven, K. J.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41–51, 1993.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249, 11–29, 2001.
- Boyle, D. P., Gupta H. V., and Sorooshian S.: Towards improved calibration of hydrologic models: combining the strengths of manual and automatic methods, *Water Resour. Res.*, 36, 3663–3674, 2000.
- Corzo, G. and Solomatine, D. P.: Baseflow separation techniques for modular artificial neural network modelling in flow forecasting, *Hydrol. Sci. J.*, 52, 491–507, 2007.
- Dooge, J. C. I.: Bringing it all together, *Hydrol. Earth Syst. Sci.*, 9, 3–14, 2005, <http://www.hydrol-earth-syst-sci.net/9/3/2005/>.
- Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S.: Multi-Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging, *Adv. Water Resour.*, 30, 1371–1386, 2007.
- Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, *Water Resour. Res.*, 43, W03434, doi:10.1029/2006WR005098, 2007.
- Georgakakos, K. P., Seo, D. J., Gupta, H., Schaake, J., and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, 298, 222–241, 2004.
- Gupta H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information, *Water Resour. Res.*, 34, 751–763, 1998.
- Hsu, K. L., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B.: Self-organizing linear output map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis, *Water Resour. Res.*, 38, 1302, doi:10.1029/2001WR000795, 2002.
- Hu, T. S., Lam, K. C., and Ng, S. T.: River flow time series prediction with a range-dependent neural network, *Hydrol. Sci. J.*, 46, 729–745, 2001.
- Jain, A. and Srinivasulu, S.: Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques, *J. Hydrol.*, 317, 291–306, 2006.
- Jordan, M. I. and Jacobs, R. A.: Hierarchical mixture of experts and the EM algorithm, *Neural Comput.*, 6, 181–214, 1994.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, 201, 272–288, 1997.
- Marshall, L., Nott, D., and Sharma, A.: Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework, *Hydrol. Process.*, 21, 847–861, 2007.
- Marshall, L., Sharma, A., and Nott, D.: Modeling the catchment via mixtures: Issues of model specification and validation, *Water Resour. Res.*, 42, W11409, doi:10.1029/2005WR004613, 2006.
- McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289, 2005.
- Monteith, J. L.: Evaporation and the environment, *Symp. Soc. Exp. Biol.*, 19, 205–234, 1965.
- Savenije, H. H. G.: Equifinality, a blessing in disguise?, *Hydrol. Process.*, 15, 2835–2838, 2001.
- See, L. and Openshaw, S.: A hybrid multi-model approach to river level forecasting, *Hydrolog. Sci. J.*, 45, 523–536, 2000.
- Shamseldin, A. Y., O'Connor, K. M., and Liang, G. C.: Methods for Combining the Output of Different Rainfall-Runoff Models, *J. Hydrol.*, 197, 203–229, 1997.
- Shamseldin, A. Y., O'Connor, K. M. and Nasr, A. E.: A comparative study of three neural network forecast combination methods for simulated river flows of different rainfall-runoff models, *Hydrol. Sci. J.*, 52, 896–916, 2007.
- Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?, *Hydrol. Process.*, 17, 1037–1041, 2003.
- Solomatine D. P. and Xue, Y.: M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the Huai River in China, *J. Hydr. Eng.*, 6, 491–501, 2004.
- Solomatine D. P.: Optimal modularization of learning models in forecasting environmental variables, *Proc. of the iEMSS 3rd Biennial Meeting: “Summit on Environmental Modelling and Software”*, edited by: Voinov, A., Jakeman, A., Rizzoli, A., et al., Burlington, USA, July, CD ROM, <http://www.iemss.org/iemss2006/sessions/all.html>, 2006, 2006.
- Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, 39, 1214, doi:10.1029/2002WR001746, 2003.
- Vrugt, J. A. and Robinson, B. A.: Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, 43(1), W01411, doi:10.1029/2005WR004838, 2007.
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modeling: dynamic identifiability analysis, *Hydrol. Process.*, 17, 455–476, 2003.
- Wang, W., van Gelder, P., Vrijling, J. K., and Ma, J.: Forecasting daily streamflow using hybrid ANN models, *J. Hydrol.*, 324(1–4), 383–399, 2006.
- Xiong, L., Shamseldin, A. Y., and O'Connor, K. M.: A nonlinear combination of the forecasts of rainfall-runoff models by the

- first-order Takagi-Sugeno fuzzy system, *J. Hydrol.*, 245(1), 196–217, 2001.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-Objective Global Optimization for Hydrologic Models, *J. Hydrol.*, 204, 83–97, 1998.
- Zhang, B. and Govindaraju, R. S.: Prediction of watershed runoff using Bayesian concepts and modular neural networks, *Water Resour. Res.*, 36, 753–762, 2000.