



Fuzzy committees of specialized rainfall-runoff models: further enhancements and tests

N. Kayastha¹, J. Ye¹, F. Fenicia², V. Kuzmin³, and D. P. Solomatine^{1,4}

¹UNESCO-IHE Institute for Water Education, Delft, the Netherlands

²Centre de Recherche Public – Gabriel Lippmann, Belvaux, Luxembourg

³Russian State Hydrometeorological University, St. Petersburg, Russia

⁴Water Resources Section, Delft University of Technology, Delft, the Netherlands

Correspondence to: N. Kayastha (n.kayastha@unesco-ihe.org)

Received: 29 November 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 15 January 2013

Revised: 4 October 2013 – Accepted: 9 October 2013 – Published: 12 November 2013

Abstract. Often a single hydrological model cannot capture the details of a complex rainfall–runoff relationship, and a possibility here is building specialized models to be responsible for a particular aspect of this relationship and combining them to form a committee model. This study extends earlier work of using fuzzy committees to combine hydrological models calibrated for different hydrological regimes – by considering the suitability of the different weighting function for objective functions and different class of membership functions used to combine the specialized models and compare them with the single optimal models.

components of flow hydrograph that correspond to different flow regimes. These models can be then combined providing a more comprehensive and accurate representation of catchment processes. Such models are referred to as multi-models, or committee models.

Multi-model approaches are not new in hydrological modelling – examples are the early works of Keefer and McQuivey (1974), Todini and Wallis (1977), Bruen (1985) and Becker and Kundzewicz (1987), who built piece-wise linear models instead of the overall linear hydrological model. Cavadias and Morin (1986) aggregated several watershed models which were considered by WMO (1986) for intercomparison of their model performances. Juemoe et al. (1987) combined a conceptual model and a statistical model, which is known as synthesized constrained linear systems model. This model was developed by a combination of the Xinanjiang model (Zhao, 1977) and the constrained linear system model (Todini and Wallis, 1977). McLeod et al. (1987) combined three models, namely, transfer function noise model, periodic autoregressive model and conceptual model for flow forecast. Since then various authors have been exploring various approaches to identification of different hydrological regimes and the ways of combining specialized models, both process-based and data driven (e.g. Shamseldin et al., 1997; Abrahart and See, 2002; Solomatine and Xue, 2004; Antil and Tape, 2004; Solomatine, 2006; Oudin et al., 2006; Ajami et al., 2006; Fenicia et al., 2007; Nasr and Bruen, 2008; Cullmann et al., 2008; Toth, 2009).

1 Introduction

Conceptual rainfall-runoff models are based on fluxes and storages representing relevant hydrological processes, and one of the challenges is to identify a set of parameters characterizing the behaviour of time-varying streamflows in a catchment. In lumped models the parameters cannot be measured directly due to the dimensional and scaling problems (Beven, 2000). These are calibrated based on the measurement of meteorological forcing data to produce model predictions that are as close as possible to the observed discharge data using some degree of expertise and experience. Typically this approach focuses on a single model using the best single set of parameters. However the model produced by one best set of parameters might not equally well describe the characteristic of the hydrological processes for all ranges of flow, and multiple models can be built from different

This paper continues to explore and improve the dynamic combination “fuzzy committee” method outlined in Solomatine (2006) and further developed and tested in Fenicia et al. (2007). Weights assigned to each specialized model’s output are based on optimally designed fuzzy membership functions, and they may be different at every time step depending on the current value of flow. In the present paper we test the performance of committee models that use several weighting schemes in objective functions for calibration of specialized models, as well as different membership functions to combine models. We also test their performance on test data sets. The method is tested on three catchments, and two approaches of optimization are used: (i) multi-objective optimization non-dominated sorted genetic algorithms (NSGA II) by Deb et al. (2002) to find Pareto-optimal solutions of specialized models and (ii) a single objective optimization – genetic algorithm (GA) (e.g. see Goldberg, 1989) and adaptive cluster covering algorithm (ACCO) by Solomatine (1999) are used to calibrate the single specialized models and single optimal models.

2 Methodology

2.1 Lumped conceptual modelling

A simplified version of HBV model (Lindström et al., 1997; Fenicia et al., 2007) is used for this study. This is a lumped conceptual hydrological model which describes hydrological processes at catchment scale. The model comprises subroutines for snow accumulation and melt, soil moisture accounting procedure, routines for runoff generation, and a simple routing procedure. The model has 13 parameters; however, only 9 parameters are effectively used when there is no snowfall.

2.2 Building specialized models

We can build several sub-models instead of using only one single model to characterize better the various regimes which represent the catchment hydrological behaviour. The sub-models are also called “specialized models”. One of the approaches of multi-modelling has been implemented in this study. The details of such an approach, previously adopted by Fenicia et al. (2007), are briefly outlined below and complemented by the possibilities of its further improvement. We considered high flows and low flows as distinctive regimes, or states of the system behaviour. Our aim was to reproduce the system response during both regimes accurately. In order to evaluate the performance of the single hydrological model in both conditions, the two weighted objective functions are used, where one is stressing the model error with respect to low-flow simulation, and the other stressing the model error with respect to high flows.

The two objective functions are defined as follows:

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

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

where n is total number of time steps, $Q_{s,i}$ is simulated flow for the time step i , and $Q_{o,i}$ is observed flow for the time step i . The two weighting functions W_{LF} and W_{HF} allow for placing the stronger weight on the low or on the high portions of the hydrograph. As a result, $RMSE_{LF}$ places stronger weight on low-flow errors and weaker weight on high-flow errors than $RMSE_{HF}$. (Please note that values $RMSE_{LF}$ and $RMSE_{HF}$ cannot be compared to each other and to the values of RMSE because of difference in weighting; this is important when viewing the resulting plots.) The types of the weighting functions (schemes) together with their parameters will be referred to further as WStype, and corresponding equations and figures are given below.

WStype	$W_{LF,i}$	$W_{HF,i}$	Eq.	Figures
I	$= (l)^N$	$= (h)^N$	(3)	Fig. 1a
II	$= \begin{cases} 0, & \text{if } l > \alpha \\ (1-l*(1/2-\alpha))^N, & \text{if } l \leq \alpha \end{cases}$	$= \begin{cases} 1, & \text{if } h > \alpha \\ (h/\alpha)^N, & \text{if } h \leq \alpha \end{cases}$	(4)	Fig. 1b
III	$= \begin{cases} 0, & \text{if } l > \alpha \\ (1-l*(1/2-\alpha))^N, & \text{if } l \leq \alpha \end{cases}$	$= \begin{cases} 1, & \text{if } h > \alpha \\ 0, & \text{if } h \leq \alpha \end{cases}$	(5)	Fig. 1c
IV	$= \begin{cases} 0, & \text{if } l > \alpha \\ (1-l*(1/2-\alpha)), & \text{if } l \leq \alpha \end{cases}$	$= \begin{cases} 1, & \text{if } h > \alpha \\ 0, & \text{if } h \leq \alpha \end{cases}$	(6)	Fig. 1d

In above equations the variables l and h are calculated as

$$l = \frac{Q_{o,max} - Q_{o,i}}{Q_{o,max}} \quad h = \frac{Q_{o,i}}{Q_{o,max}}, \quad (7)$$

where $Q_{o,max}$ is maximum observed flow, N is power value (for the experiment in this study we considered only 1, 2 or 3), and α is the threshold for selecting weights of flows (it was chosen to be 0.75). Note that both N and α can be also subjected to optimization, but in this study it was not done. By computing both objective functions over the whole range of discharges, both functions constrain the model to fit the entire hydrograph for WStype I where α parameter is not used. However for WStype with α parameter, W_{LF} excludes high flows from computation of objective function if the condition is $l > \alpha$. In the same way, W_{HF} excludes low flows if $h \leq \alpha$ for WStype III and IV.

One potential issue related to the scaling formulas (Eq. 7) is worth mentioning: $Q_{o,max}$ is the maximum for calibration data, but this of course does not guarantee that it will not be superseded in the future when model is in operation (or when simulating operation by using verification data). The quadratic function will still handle values above 1, but if the calibration maximum is exceeded considerably, then the high flow will be given disproportionately high weights, and low flows disproportionately low ones. A solution could be in using a bit wider range for scaling.

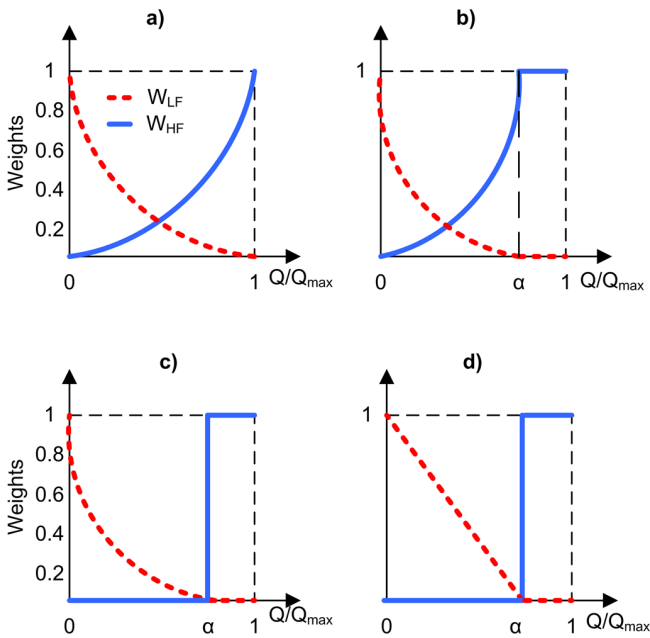


Fig. 1. (a) WStype I, weighting scheme for objective functions studied in Fenicia et al. (2007). (b) WStype II, (c) WStype III, and (d) WStype IV; additionally, these three weighting schemes were attempted in the latest experiments.

2.3 Combining specialized models

The specialized models are built under the conditions of different regimes of catchment hydrological responses and are combined using the appropriate combining scheme. However the issue is how to handle the compatibility at the boundaries between the two different specialized models. One of the possible ways is to use a soft weighting scheme that switches smooth transition between boundaries. The contribution of each specialized model is based on using a fuzzy membership function – the so-called “fuzzy committee” described by Solomatine (2006). In this weighting scheme we initially used trapezoidal functions parameterized by the two transitional parameters $[\gamma, \delta]$ (see Fig. 2a): the membership function (weight) of low-flow model is assigned 1 when the relative flow is below the parameter γ , then starts to decrease in the proximity of the region boundary when the relative flow is between γ and δ ; it decreases to zero beyond the boundary when the relative flow is above δ . Similarly, the membership function of the high-flow model follows the opposite logic. These membership functions for the two specialized models are described by Eqs. (9) and (10). The outputs of models are multiplied by the weights that depend on the value of flow and then normalized (Eq. 8). So the overall committee model is defined as follows:

$$Q_{c,i} = (m_{LF} \cdot Q_{LF,i} + m_{HF} \cdot Q_{HF,i}) / (m_{LF} + m_{HF}) \quad (8)$$

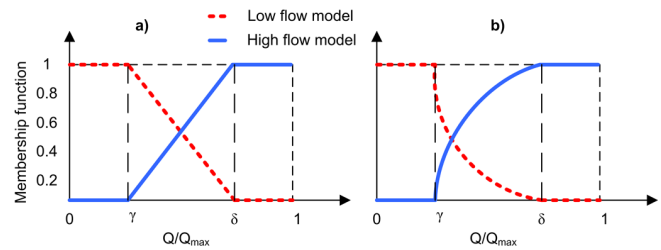


Fig. 2. (a) A typical fuzzy membership function used to combine the single specialized models (MFtype A), (b) a class of membership functions for high and low-flow models tested in the new experiments (MFtype B).

$$m_{LF} = \begin{cases} 1, & \text{if } h < \gamma \\ 1 - (h - \gamma) / (\delta - \gamma)^N, & \text{if } \gamma \leq h < \delta \\ 0, & \text{if } h \geq \delta \end{cases} \quad (9)$$

$$m_{HF} = \begin{cases} 0, & \text{if } h < \gamma \\ (h - \gamma) / (\delta - \gamma)^{1/N}, & \text{if } \gamma \leq h < \delta \\ 1, & \text{if } h \geq \delta \end{cases} \quad (10)$$

where m_{LF} and m_{HF} are membership functions (denoted also as MFtype) for the two specialized models, $Q_{LF,i}$ and $Q_{HF,i}$ are simulated high and low flows for the time step i ; γ and δ are threshold for high and for low flows respectively, N is power value which was used to smooth between the models (if $N = 1$, MFtype is referred to as Type A (see Fig. 2a), and for $N = 2$ or more – as Type B (see Fig. 2b)).

Building committee models consists of the following steps. First the two optimal specialized models – model 1 for low flow ($Q_{LF,i}$) and model 2 for high flow ($Q_{HF,i}$) – are sought using optimization (minimizing $RMSE_{LF}$ for model 1 and $RMSE_{HF}$ for model 2); this can be done by solving a single-objective optimization problem separately for these two models, or by multi-objective optimization for two objective functions $RMSE_{LF}$ and $RMSE_{HF}$. After that model 1 and model 2 are combined using the two membership functions m_{LF} and m_{HF} whose two parameters δ and γ are found by optimization to ensure the lowest classical root mean squared error (RMSE) of the committee model (Fig. 2). The resulting model is subsequently verified (tested) on the verification (test) data set, and compared to the single hydrological model (which is optimized by a single-objective optimization algorithm) based on RMSE as an objective function and in addition evaluated using Kling–Gupta efficiency index (KGE) proposed by Gupta et al. (2009). They discuss limitations of using a single criterion like Nash–Sutcliffe efficiency (NSE) (or closely related mean squared error, MSE) and show that it can be decomposed into the three components – representing correlation, variability, and bias. These authors also suggest aggregating this information into one formula (representing distance in the space of the three mentioned criteria to the ideal point, which is zero). They state

Table 1. The summary of the runoff data for Alzette catchment in Luxembourg, Leaf catchment in USA, and Bagmati catchment in Nepal.

Statistical properties	Complete data	Calibration data	Verification data
Alzette (area = 288 km ²)			
Period (day/month/year hour)	7/29/2000, 12:00–8/6/2002, 07:00	7/29/2000, 12:00–8/6/2001, 07:00	8/6/2001, 08:00–8/6/2002, 07:00
Number of data points	17 720	8960	8760
Average (m ³ s ⁻¹)	4.64	5.55	3.70
Minimum (m ³ s ⁻¹)	0.45	0.59	0.45
Maximum (m ³ s ⁻¹)	51.41	51.41	31.15
Standard deviation (m ³ s ⁻¹)	5.35	5.52	5.00
Leaf (area = 1924 km ²)			
Period (day/month/year)	28/07/1951–21/09/1961	28/07/1951–25/07/1957	26/07/1957–21/09/1967
Number of data points	3717	2190	1527
Average (m ³ s ⁻¹)	28.28	23.02	35.81
Minimum (m ³ s ⁻¹)	1.56	1.56	2.92
Maximum (m ³ s ⁻¹)	2.38	549.35	1313.91
Standard deviation (m ³ s ⁻¹)	64.48	47.37	82.51
Bagmati (area = 3500 km ²)			
Period (day/month/year)	01/01/1988–31/12/1995	01/01/1991–30/06/1995	01/01/1988–28/02/1991
Number of data points	2922	1767	1155
Average (m ³ s ⁻¹)	150.0	150.8	148.6
Minimum (m ³ s ⁻¹)	5.1	5.1	7.7
Maximum (m ³ s ⁻¹)	5030.0	5030	3040
Standard deviation (m ³ s ⁻¹)	271.2	280.5	256.40

however that “the primary purpose of this study was not to design an improved measure of model performance, but to show clearly that there are systematic problems inherent with any optimization that is based on mean squared errors (such as NSE). The alternative criterion KGE was simply used for illustration purposes.” We decided that this statement is too modest and followed the other authors (e.g. Pechlivanidis et al., 2011; Coron et al., 2013) that explicitly use KGE as one of the model performance measures. An additional advantage of using KGE is that it is a relative measure permitting inter-catchment comparison. Using RMSE along with KGE, we run the risk of having an overlap (since KGE is in fact constructed of components of NSE and hence RMSE), but still consider this to be useful to do to follow an adopted hydrological practice, and to allow for having a better judgement about various facets of model performance including an explicit absolute measure of model error.

The equations of RMSE and KGE are given below.

$$\text{RMSE} = \sqrt{\frac{1}{n} \left(\sum_{i=1}^n (Q_{s,i} - Q_{o,i})^2 \right)}, \quad (11)$$

where $Q_{o,i}$ is the observed discharges for the time step i , $Q_{s,i}$ is the simulated discharges (single optimal or committee models) for the time step i , and n is the number of observations.

$$\text{KGE} = 1 - \sqrt{(c - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}, \quad (12)$$

where c is the linear cross-correlation coefficient between Q_o and Q_s , α is a measure of variability in the data values (equal to the standard deviation of Q_s , over the standard deviation of Q_o), and β is equal to the mean of Q_c , over the mean of Q_o . KGE is subject to maximization with an ideal value at unity.

3 Results and discussion

Three catchments, namely, Alzette catchment in Luxembourg, Leaf River catchment in USA and Bagmati catchment in Nepal, are selected for case study. The summary statistics and records of data for calibration and verification of catchments are presented in Table 1. These data cover multiple-year periods (except Alzette), all seasons and multiple peak flows. Ideally, we have to try to split data into statistically similar sets (coverage of seasons, number and size of peaks, variance, mean, etc.). Of course in this type of splits of hydrological data, one is constrained by the wish to keep data in contiguous blocks (to be able to plot the time-series data such as hydrographs), so the calibration and verification data sets almost always have statistical differences.

The experiment follows the one used in an earlier study (Fenicia et al., 2007) where the Alzette catchment was considered, and only calibration data were considered for building the models without further validation. We present here two additional catchments (Leaf and Bagmati) with both calibration and verification periods, and compare the overall

Table 2. The ranges of model parameters.

Parameters	Units	Descriptions	Ranges used in calibration (optimization)		
			Alzette	Leaf	Bagmati
FC	(mm)	Maximum soil moisture content	100–450	100–400	50–500
LP	(–)	Limit for potential evapotranspiration	0.3–1	0.1–1	0.3–1
ALFA	(–)	Response box parameter	0.1–1	0–2	0–4
BETA	(–)	Exponential parameter in soil routine	0.1–2	1.0–4	1.0–6
K	(mm h ⁻¹)	Recession coefficient for upper tank	0.005–0.5	0.05–0.5	0.05–0.5
K4	(mm h ⁻¹)	Recession coefficient for lower tank	0.001–0.1	0.01–0.3	0.01–0.3
PERC	(mm h ⁻¹)	Percolation from upper to lower response box	0.01–1	0–5	0–8
CFLUX	(mm h ⁻¹)	Maximum value of capillary flow	0–0.05	0–1	0–1
MAXBAS	(h)	Transfer function parameter	8–15	2–6	1–3

Note: The unit d (day) is used for Leaf and Bagmati catchments instead of h (hour).

model performance when using different weighing schemes for objective functions (Fig. 1) and different membership functions (Fig. 2). First two months of calibration data are considered as the warming-up in Leaf and Bagmati catchment. However for Alzette catchment we used the hourly data set of only one year for the calibration period and one year for verification. To somehow compensate for the lack of data in calibration, we allocated 168 h of data for the warm-up period.

The ranges of HBV model parameters for optimization are given in Table 2. We produced the specialized models (the best single model specialized on high flows and low flows) which are optimized by multi- and single-objective optimization algorithms. The identified best sets of parameters for different models are given in Table 3. It is worth mentioning that in this paper we use only one model structure (HBV), but the use of several calibration methods and several error functions results in several model parameterizations, or instantiations. For simplicity we will be speaking of several models, meaning actually the same model structure but its several parameterizations.

We present results of calibration using several optimization algorithms: NSGA-II, GA and ACCO. The reasons for using different optimization algorithms for calibration are as follows: (1) the initially used GA appeared to be quite slow (in terms of the required model runs), so we decided to test the use of faster algorithms as well; (2) to cross-check one by another since they both use randomization of initial population and this affects results. We have also tried the use of a stepwise line search (SLS) algorithm (Kuzmin et al., 2008), which showed its effectiveness and efficiency, and its working depends less on the randomized generation of the initial population but uses rather some a priori estimates; these tests however were not finalized, so we decided not to present the intermediate results in this comparison.

In each experiment a committee model is compared with the single optimal model which is calibrated by two different single-objective optimization algorithms (GA and ACCO).

The best single models specialized for low and high flows respectively (found by NSGA-II, GA and ACCO) are used in the committee model. The points denoted “committee models” correspond to the model parameterizations generated during the exhaustive search for the best γ and δ ensuring the lowest RMSE. We also tested a committee model which is built by combining the specialized models, and compared against the single optimal model for all catchments. In Tables 4 and 5 these models are denoted as $Q_c(\text{ACCO})$, $Q_c(\text{GA})$ and $Q_c(\text{NSGA-II})$. Interestingly, the committee model is better on both objective functions (RMSE and KGE) than the single optimal models for all case studies.

The graphs of Pareto-optimal of single specialized models (calibrated by NSGA-II), committee models and single optimal models (calibrated by ACCO and GA) for Leaf catchment are shown in Fig. 3 (Kayastha and Solomatine, 2013). In this plot one may see the 20205 single specialized models presented as they were generated during multi-objective optimization process by using WStype I weighting scheme. The 70 model parameterizations identified as the best (Pareto-optimal) are presented as well (by the darker points). The Pareto-optimal models in calibration are not necessarily the best in verification – this is of course no surprise; what is however important is to stress that their performance is not too much lower than that of calibration data. Among the Pareto-optimal models, the best single model specialized on low flows obtained 11.01 in RMSE_{LF} (all errors are in $\text{m}^3 \text{s}^{-1}$), and it can be seen from Figure 3b that this model is also very close to be the best in verification with RMSE_{LF} being 18.33. However, the best single model specialized on high flows (RMSE_{HF} is 6.08) is not too bad, but not the best in verification (its RMSE_{HF} is 8.42 and in Fig. 3b it is easy to see that there are many models with lower RMSE_{HF}).

The committee model resulted in the classical RMSE of 15.63 in calibration and 25.23 in verification. To represent this model in Fig. 3 (solid square), we had to calculate for it the corresponding RMSE_{LF} and RMSE_{HF} , and we did it

Table 3. Sets of parameters identified by different optimization algorithms.

Catchments	Models		FC	LP	ALFA	BETA	K	K4	PERC	CFLUX	MAXBAS
Alzette	ACCO	SO	284.83	0.26	0.06	0.65	0.02	0.01	0.16	0.04	10.96
		LF	356.34	0.46	0.10	0.42	0.02	0.00	0.14	0.10	13.48
		HF	414.48	0.19	0.30	0.49	0.00	0.03	0.97	0.01	8.51
	GA	SM	309.97	0.35	0.03	0.72	0.03	0.01	0.27	0.01	11.45
		LF	255.11	0.46	0.07	0.98	0.03	0.01	0.23	0.05	12.62
		HF	338.84	0.56	0.06	0.95	0.01	0.02	0.89	0.00	8.37
	NSGA-II	LF	253.24	0.16	0.07	0.54	0.02	0.00	0.13	0.00	9.49
		HF	253.25	0.34	0.07	0.52	0.02	0.01	0.14	0.00	9.54
Leaf	ACCO	SM	272.11	0.29	0.30	1.57	0.27	0.26	2.27	0.62	6.04
		LF	303.49	0.14	0.42	1.14	0.08	0.04	0.47	0.91	4.94
		HF	230.27	0.16	0.62	1.08	0.08	0.28	0.00	0.96	5.99
	GA	SM	349.80	0.64	0.65	2.29	0.07	0.14	0.65	0.97	5.99
		LF	313.84	0.22	0.26	1.24	0.14	0.05	0.66	1.00	5.06
		HF	285.21	1.00	0.72	1.91	0.05	0.26	1.83	0.99	6.00
	NSGA-II	LF	301.88	0.36	0.37	1.95	0.14	0.24	1.07	0.89	5.57
		HF	274.26	0.90	0.45	2.27	0.15	0.26	1.24	0.85	5.86
Bagmati	ACCO	SM	371.88	0.67	0.07	1.01	0.49	0.14	7.78	0.20	2.57
		LF	448.05	0.79	0.13	1.04	0.37	0.06	7.88	0.29	2.53
		HF	445.04	0.65	0.08	1.05	0.46	0.23	7.79	0.67	2.94
	GA	SM	430.26	0.59	3.85	1.04	0.50	0.08	8.00	0.01	2.99
		LF	301.36	0.83	0.22	1.05	0.22	0.06	7.99	0.10	2.39
		HF	453.50	0.58	0.07	1.23	0.49	0.03	0.05	0.00	2.91
	NSGA-II	LF	364.46	0.75	0.16	1.05	0.33	0.07	7.94	0.11	2.48
		HF	370.66	0.66	0.10	1.06	0.37	0.10	7.17	0.17	2.74

SM: single hydrological model (single optimal model-optimized by a single-objective optimization algorithm based on the classical RMSE); LF and HF: low-flow model and high-flow model (optimized by a single-objective optimization algorithms and multi-objective optimization based on the $RMSE_{LF}$ and $RMSE_{HF}$).

Table 4. The performances of single optimal models (optimized based on classical RMSE) and committee models of various catchments. Committee models are assembled by combination of the weighting scheme WStype I and membership function MFtype A.

Catchments	Models	Membership function		$RMSE_{HF}$		$RMSE_{LF}$		RMSE		KGE	
		δ	γ	Cal.	Ver.	Cal.	Ver.	Cal.	Ver.	Cal.	Ver.
Alzette	Q_s (ACCO)	n.a.	n.a.	0.97	1.23	2.11	1.71	2.37	2.39	0.88	0.86
	Q_s (GA)	n.a.	n.a.	0.99	1.2	2.03	2.01	2.31	2.42	0.89	0.88
	Q_c (ACCO)	0.50	0.25	0.53	0.81	1.65	1.48	2.10	2.06	0.91	0.89
	Q_c (GA)	0.60	0.40	0.56	0.78	1.71	1.35	2.19	2.15	0.90	0.90
	Q_c (NSGA-II)*	0.50	0.30	0.50	0.86	1.48	1.48	1.99	2.07	0.93	0.89
Leaf	Q_s (ACCO)	n.a.	n.a.	8.97	14.31	11.84	18.81	17.56	26.76	0.87	0.83
	Q_s (GA)	n.a.	n.a.	7.96	11.70	11.71	19.64	17.36	26.58	0.88	0.84
	Q_c (ACCO)*	0.39	0.37	5.88	9.00	10.55	18.74	15.63	25.23	0.91	0.85
	Q_c (GA)	0.51	0.50	5.82	8.91	10.86	19.38	15.76	24.88	0.90	0.85
	Q_c (NSGA-II)	0.50	0.49	5.61	9.05	10.85	18.59	16.05	23.86	0.90	0.88
Bagmati	Q_s (ACCO)	n.a.	n.a.	29.55	85.33	38.84	87.54	101.01	112.42	0.87	0.90
	Q_s (GA)	n.a.	n.a.	32.42	86.18	34.16	92.38	101.69	116.72	0.86	0.88
	Q_c (ACCO)	0.61	0.49	40.09	69.31	72.69	65.09	95.96	109.38	0.87	0.90
	Q_c (GA)	0.57	0.47	18.93	39.75	77.29	74.55	94.39	110.29	0.89	0.91
	Q_c (NSGA-II)*	0.50	0.47	26.94	48.34	82.42	81.67	94.16	109.72	0.87	0.91

Q_s (ACCO) and Q_s (GA): single hydrological models (SMs); Q_c (ACCO), Q_c (GA), and Q_c (NSGA-II): committee models (CMs) (bold: best CMs, italics: best SMs and *: best model for a catchment, n.a.: not available).

Table 5. Performance of committee models for the Leaf catchment, with possible combinations of the various weighting schemes and membership functions.

Models	Weighted function		Membership function			RMSE _{HF}		RMSE _{LF}		RMSE		KGE	
	WStype	<i>N</i>	MFtype	δ	γ	Cal.	Ver.	Cal.	Ver.	Cal.	Ver.	Cal.	Ver.
<i>Q_s</i> (ACCO)	n.a.	n.a.	n.a.	n.a.	8.97	14.31	11.84	18.81	n.a.	17.56	26.76	0.87	0.83
<i>Q_s</i> (GA)	n.a.	n.a.	n.a.	n.a.	7.96	11.70	11.71	19.64	n.a.	<i>17.36</i>	<i>26.58</i>	<i>0.88</i>	<i>0.84</i>
<i>Q_c</i> (ACCO)	I	2	A	0.39	0.37	5.88	9.00	10.55	18.74	15.63	25.23	0.91	0.85
	II	2	A	0.45	0.44	7.27	10.47	9.71	17.86	16.01	24.38	0.89	0.85
	III	2	A	0.65	0.14	1.02	3.23	9.56	17.26	15.60	24.52	0.92	0.88
	IV	2	A	0.56	0.55	1.13	3.00	11.47	19.64	16.20	25.68	0.87	0.86
	I	2	B	0.39	0.38	5.88	9.00	10.55	18.74	15.63	25.26	0.90	0.85
	II	2	B	0.45	0.44	7.27	10.47	9.71	17.86	16.03	24.38	0.89	0.85
	III	2	B	0.94	0.15	1.02	3.23	9.56	17.26	15.67	24.72	0.92	0.88
	IV	2	B	0.56	0.55	1.13	3.00	11.47	19.64	16.20	25.68	0.87	0.85
<i>Q_c</i> (GA)	I	2	A	0.51	0.50	5.82	8.91	10.86	19.38	15.76	24.88	0.90	0.85
	II	2	A	0.66	0.14	7.36	9.84	9.81	18.93	16.13	25.81	0.92	0.86
	III	2	A	0.99	0.16	0.99	2.77	9.67	18.36	16.53	24.67	0.92	0.87
	IV	2	A	0.99	0.30	1.01	2.85	11.50	19.84	16.60	23.96	0.92	0.87
	I	2	B	0.99	0.15	5.82	8.91	10.86	19.38	16.30	25.56	0.93	0.86
	II	2	B	0.87	0.16	7.36	9.84	9.81	18.93	16.22	25.58	0.93	0.87
	III	2	B	0.99	0.31	0.99	2.77	9.67	18.36	16.47	24.34	0.92	0.87
	IV	2	B	0.99	0.42	1.01	2.85	11.50	19.84	16.55	24.06	0.92	0.87
	I	1	B	0.42	0.41	9.43	13.06	12.44	20.56	15.96	24.04	0.91	0.89
	I	3	B	0.99	0.23	3.64	6.95	9.30	17.08	16.50	25.53	0.91	0.86
<i>Q_c</i> (NSGA-II)	I	2	A	0.50	0.49	5.61	9.05	10.85	18.59	16.05	23.86	0.90	0.88
	II	2	A	0.50	0.49	7.31	9.98	9.70	17.45	15.71	23.85	0.91	0.88
	III	2	A	0.86	0.47	1.08	2.86	10.13	17.76	17.36	23.41	0.91	0.90
	IV	2	A	0.86	0.45	1.08	3.10	11.62	20.01	16.76	23.97	0.90	0.87
	I	2	B	0.50	0.29	5.61	9.05	10.85	18.59	16.45	23.96	0.90	0.88
	II	2	B	0.50	0.15	7.31	9.98	9.70	17.45	16.71	23.95	0.91	0.88
	III	2	B	0.99	0.49	1.08	2.86	10.13	17.76	17.29	23.46	0.91	0.91
	IV	2	B	0.99	0.46	1.08	3.10	11.62	20.01	16.71	23.97	0.91	0.88
	I	1	A	0.38	0.36	9.59	12.76	12.91	20.46	16.58	23.86	0.91	0.91
	I	3	A	0.50	0.49	4.19	7.88	9.60	17.07	15.96	23.79	0.90	0.88

The value of $\alpha = 0.75$ used in WStype II, III and IV (bold: best CMs, italics: best SMs).

using the same type of weighing scheme (WStype I) which was used in calibration of specialized models. In the same way we presented the single optimal models identified using the two single objective optimization methods (ACCO and GA represented by a circle and a star respectively). It can be seen that the committee models are closer to the ideal point than the other single optimal models, and this means that the committee model's performance is the highest of any of the single models. Plots for Bagmati and Alzette are not presented here due to space limitations.

The performances of the best single models specialized on high and low flows (RMSE_{HF} and RMSE_{LF}) on various catchments are presented in Table 4. However, again, it should be noted that RMSE_{LF}, and RMSE_{HF} cannot be compared since they use different formulas. RMSE_{LF} values are even higher than those of RMSE_{HF} – the reason is that the

number of low flows is much higher than of high flows, and the denominator (total number of observations) in both formulas is the same.

In Leaf catchment we tested all possible combinations of different weighting schemes types and classes of membership functions, the results of which are presented in Table 5. Noticeably, all committee models improved their performances in verification in comparison to the single hydrological models which were optimized by single objective optimization. However, in the other two catchments Alzette and Bagmati, the number of experiments was smaller, but in all of them the committee models demonstrated the higher performance in both calibration and test data sets.

Table 4 reports the performance of committee models and single-optimal models calibrated by ACCO and GA for each catchment. For the Bagmati catchment, the

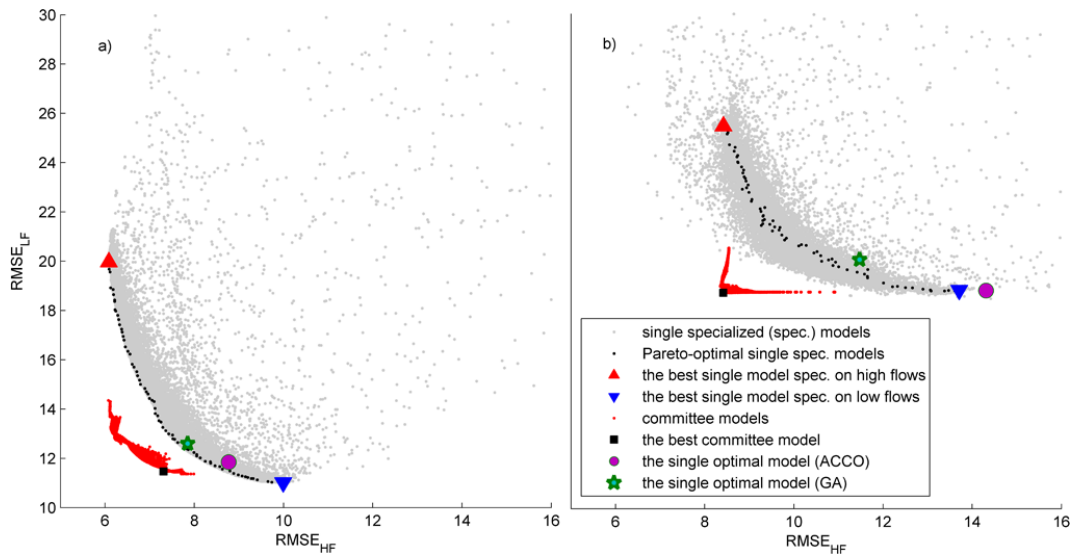


Fig. 3. The identified sets of Pareto-optimal parameterizations of single specialized models (optimized by NSGA-II), committee models and single optimal models, calibrated by ACCO and GA in Leaf catchment: (a) calibration data set and (b) verification data set (model parameterizations from (a) are used).

RMSE calculated by single optimal model in calibration is $101.01 \text{ m}^3 \text{ s}^{-1}$, and verification is $112.42 \text{ m}^3 \text{ s}^{-1}$. However, it can be noticeably improved by the committee models and obtained around $94.16\text{--}95.96 \text{ m}^3 \text{ s}^{-1}$ in calibration and $109.38\text{--}110.29 \text{ m}^3 \text{ s}^{-1}$ in verification. The RMSE of single model produced $26.76 \text{ m}^3 \text{ s}^{-1}$ in verification period for Leaf catchment. However, when new types of weighting and membership functions were used, RMSE dropped to $23.41 \text{ m}^3 \text{ s}^{-1}$ (see Table 5).

The visual plots of the committee models which are built from the combination of the two specialized models for high and low flows with respect to the hydrograph simulations are represented in Fig. 4. It can be observed that the committee model combines the best features of the specialized models.

Our experiments have led to one important observation related to using the weighting function for objective functions (Fig. 1 and Eqs. 3–7) in calibration of specialized models quadratic function we used earlier (Fenicia et al., 2007) was in fact the first guess that it will allow for distinguishing the low and high flows. In our latest experiments it appeared, quite expectedly, that other function (for example, cubic) may work better in calibration period.

It is useful to mention an issue of overfitting that was raised by one of the reviewers and reiterated by the editor. This problem (typically addressed in data-driven modelling) may lead to a decrease of accuracy in operation (which can be detected during validation), and there are several ways of dealing with it, for example trying to limit the complexity of a model, and/or using cross-validation data set to control the calibration process. The committee model includes a number of parameters that increase accuracy of the overall model. However, this contributes to increasing the complexity of the

model and may lead to overfitting. However the number of parameters added is very small (only two – δ and γ in the membership function), so an increase in complexity is also small, and so we see no need in using regularization. Early-stopping strategy cannot be used since process of training is pure exhaustive search, and not iterative, so it cannot be stopped prematurely. Alternatively, a cross-validation set can be used; however, it would require effort in finding the optimal splitting of the data set into three rather than two sets, with the smaller subsets for each purpose, and subsequent analysis of the impact of having smaller sets for calibration and testing. So in this paper we have not used cross-validation but can recommend doing so in future research if the size of data would allow for this.

4 Conclusions and direction for further work

In this study we presented further improvements to a fuzzy committee approach – one possible way to improve the hydrological model prediction involving a combination of model outputs obtained by differently parameterized models with the same model structure. The two best single models specialized in low flows and high flows, which were found by optimization based on two objective functions RMSE_{LF} and RMSE_{HF} respectively, are combined using membership function to form the committee model and this model compared to the single hydrological model based on RMSE as an objective function.

The major findings of this study are as follows:

- By conducting a number of experiments, we can confirm the initial results reported earlier in Fenicia et

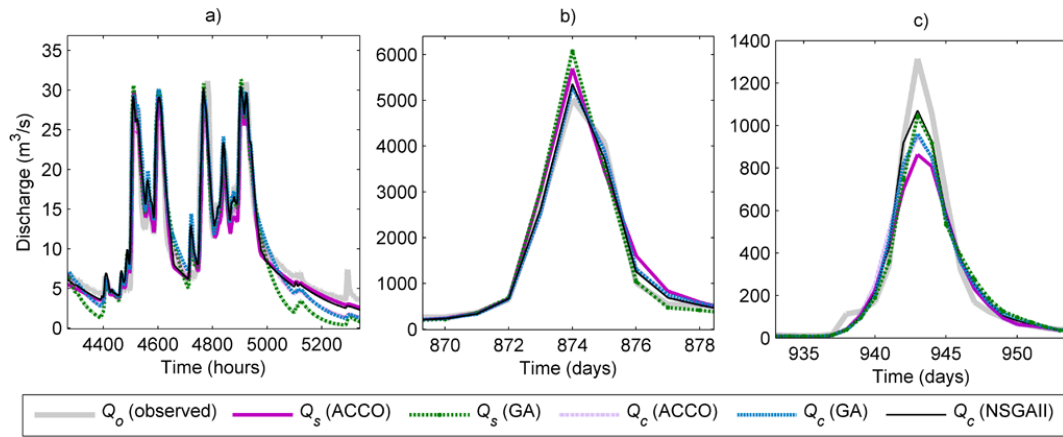


Fig. 4. A fragment of a hydrograph generated from various models: Q_o – observed discharge, Q_s – model identified by single-objective optimization (ACCO and GA), Q_c – committee model (ACCO, GA and NSGA-II), (a) Alzette (31/01/2002 08:00:00–18/03/2002 03:00:00), (b) Bagmati (20/5/1990–28/5/1990), and (c) Leaf (13/02/1960–08/03/1960).

al. (2007) that the combination of two specialized models indeed provides a method leading to a higher model of the resulting model. In three case studies we could reproduce the situation shown in Fig. 3 when the fuzzy committee model is better on both objective functions than the single model(s). In calibration a committee model is always better than the single model, independent of the values of parameters MFtype and WStype (however we have to optimize δ and γ).

- We found that the best committee model identified by calibration, when tested on verification data, outperforms the best single model (identified by calibration) in all case studies (Table 4). We see this as the most important conclusion from this study.
- We conducted a thorough investigation of the impact of MFtype and WStype on model performance using the Leaf catchment data (Table 5). We cannot suggest the “universal” best set of parameters MFtype and WStype applicable for any case study: in calibration all of them led to models better than the single one, but in verification performance of models using different MFtype and WStype slightly differs for different cases (however it is still higher than that of the single model).
- There is an interesting effect concerning direct optimization of parameters γ and δ . It appeared that in most experiments after optimization these parameters obtained very close values, which means that there is a very narrow region where specialized models “work together”. Potentially this may lead to situations when a minor change in average flows will force the committee model to produce relatively large changes in outputs.

- Results for the Alzette case study should be considered with care since the data set was limited (one year for calibration and one year for verification).

Further development and application of the presented approach is seen in the following.

- It would be useful to explore possible interactions between the parameters of the weighting functions and the shapes and parameters of the membership functions.
- The committee model may be sensitive to the choice of these parameters determining the shapes of the weighting and membership functions, and indeed exploring the use of more robust optimization methods for their identification may ensure higher robustness of the committee model.
- More accurate comparison between performances of various models can be made if the problem of overfitting is addressed (for example by cross-validation during model calibration and stopping the optimization process earlier to ensure minimum error on cross-validation) instead of the “deep” optimization of the model on calibration set and using the single validation set.

An aspect related to the model performance metrics which needs attention as well is worth mentioning. We used the same objective functions for different magnitude of flows which originate from statistical theory. However the nature of this metric (e.g. RMSE) is basically oriented towards high flows and may not be suited for low flows. Therefore the performance measure can be in the form of transformed metric (e.g. transformed RMSE) to calibrate a low-flow model (e.g. van Werkhoven et al., 2009; Willems, 2009; Kollat et al.,

2012), or being a multi-scale objective function taking into account errors at various timescales (Kuzmin et al., 2008).

Further developments are foreseen in improving the weighting schemes involving hydrological states and various combinations of variables influencing the streamflow (for example those presented by Oudin et al., 2006; Kim et al., 2006; Corzo and Solomatine, 2007a, b; Marshall et al., 2007; Jeong and Kim, 2009; Fernando et al., 2012). Combining these approaches will hopefully lead to the techniques for discovering various regimes in the time series representing the modelled system – and this would allow for optimal combination of domain (hydrologic) knowledge incorporated in models with automatic machine learning or time-series analysis routines.

Acknowledgements. Part of this work was financially supported by the enviroGRIDS and WeSenseIt projects supported by the FP7 Programme of the European Commission, and by the grant no. 14.B.37.21.1977 “Ensemble forecasting of floods for decision making during extreme hydrological events” of the Russian Ministry of Education and Science. The authors acknowledge the three anonymous referees for providing very valuable comments. We are especially grateful to the editor E. Toth, for her helpful and constructive suggestions.

Edited by: E. Toth

References

- Abrahart, R. J. and See, L.: 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, doi:10.5194/hess-6-655-2002, 2002.
- Ajami, N. K., Duan, Q., Gao, X., and Sorooshian, S.: Multi-model combination techniques for hydrological forecasting: Application to Distributed Model Inter-comparison Project results, *J. Hydrometeorol.*, 7, 755–768, 2006.
- Anctil, F. and Tape, D. G.: An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition, *J. Environ. Eng. Sci.*, 3, 121–128, 2004.
- Becker, A. and Kundzewicz, Z. W.: Nonlinear flood routing with multilinear models, *Water Resour. Res.*, 23, 1043–1048, 1987.
- Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, *Hydrol. Earth Syst. Sci.*, 4, 203–213, doi:10.5194/hess-4-203-2000, 2000.
- Bruen, M.: Black box modelling of catchment behaviour, Ph.D. thesis, National University of Ireland, Dublin, 1985.
- Cavadias, G. and Morin, G.: The Combination of Simulated Discharges of Hydrological Models, Application to the WMO Inter-comparison of Conceptual Models of Snowmelt Runoff, *Nordic Hydrol.*, 17, 21–32, 1986.
- Coron, L., Andréassian, V., Perrin, C., Bourqui, M., and Hendrickx, F.: On the lack of robustness of hydrologic models regarding water balance simulation – a diagnostic approach on 20 mountainous catchments using three models of increasing complexity, *Hydrol. Earth Syst. Sci. Discuss.*, 10, 11337–11383, doi:10.5194/hessd-10-11337-2013, 2013.
- Corzo, G. and Solomatine, D. P.: Baseflow separation techniques for modular artificial neural network modelling in flow forecasting, *Hydrol. Sci. J.*, 52, 491–507, 2007a.
- Corzo, G. and Solomatine, D. P.: Knowledge-based modularization and global optimization of artificial neural network models in hydrological forecasting, *Neural Networks*, 20, 528–536, 2007b.
- Cullmann, J., Krauß, T., and Philipp, A.: Enhancing flood forecasting with the help of processed based calibration, *Phys. Chem. Earth*, 3, 1111–1116, 2008.
- Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, 6, 182–197, 2002.
- Fenicia, F., Solomatine, D. P., Savenije, H. H. G., and Matgen, P.: Soft combination of local models in a multi-objective framework, *Hydrol. Earth Syst. Sci. Discuss.*, 4, 91–123, doi:10.5194/hessd-4-91-2007, 2007.
- Fernando, A., Shamseldin, A., and Abrahart, R.: Use of Gene Expression Programming for Multimodel Combination of Rainfall-Runoff Models, *J. Hydrol. Eng.*, 17, 975–985, 2012.
- Goldberg, D. E.: Genetic algorithms in search, optimization and machine learning, Reading, MA, Addison-Wesley, 1989.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, 2009.
- Jeong, D. I. and Kim, Y.-O.: Combining single-value streamflow forecasts – A review and guidelines for selecting techniques, *J. Hydrol.*, 377, 284–299, 2009.
- Juemou, W., Ruifang, Z., and Guanwu, X.: Synthesised Constrained Linear System (SCLS), *J. Hydraul. Eng.*, 7, Beijing, 1987.
- Kayastha, N. and Solomatine, D. P.: Graphical materials from studies on committee modeling and uncertainty analysis report, UNESCO-IHE, Institute for Water Education, Delft, available at: www.hydroinformatics.org/sol/papers/KSGraphics.pdf (last access: 5 May 2013), 2013.
- Keefer, T. N. and McQuivey, R. S.: Multiple linearization flow routing model, *J. Hydraul. Div., ASCE*, 100, 1031–1046, 1974.
- Kim, Y.-O., Jeong, D., and Ko, I. H.: Combining Rainfall-Runoff Model Outputs for Improving Ensemble Streamflow Prediction, *J. Hydrol. Eng.*, 11, 578–588, 2006.
- Kollat, J. B., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resour. Res.*, 48, 3520, doi:10.1029/2011WR011534, 2012.
- Kuzmin, V., Jun, S. D., and Koren, V.: Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search, *J. Hydrol.*, 353, 109–128, 2008.
- 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 mixture of experts framework, *Hydrol. Process.*, 21, 847–861, 2007.
- McLeod, A. I., Noakes, D. J., Hipel, K. W., and Thompstone, R. M.: Combining hydrologic forecast, *J. Water Resour. Plann. Manage.*, 113, 29–41, 1987.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I, A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.

- Nasr, A. and Bruen, M.: Development of neuro-fuzzy models to account for temporal and spatial variations in a lumped rainfall-runoff model, *J. Hydrol.*, 349, 277–290, 2008.
- Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., and Michel, C.: Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations, *Water Resour. Res.*, 42, W07410, doi:10.1029/2005WR004636, 2006.
- Pechlivanidis, I., Jackson, B., McMillan, H., and Gupta, H.: Using an informational entropy-based metric as a diagnostic of flow duration to drive model parameter identification, *Global NEST J.*, 14, 325–334, 2012.
- 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.
- Solomatine, D. P.: Two strategies of adaptive cluster covering with descent and their comparison to other algorithms, *J. Global Optim.*, 14, 55–78, 1999.
- 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. J., and Rizzoli, A. E., Burlington, USA, July, 2006.
- 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. Hydrol. Eng.*, 6, 491–501, 2004.
- Todini, E. and Wallis, J. R.: Using CLS for Daily or Longer Period Rainfall-Runoff Modelling, in: *Mathematical Models for Surface Water Hydrology*, edited by: Ciriani, T. A., Maione, U., and Wallis, J. R., John Wiley & Sons, London, 149–168, 1977.
- Toth, E.: Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting, *Hydrol. Earth Syst. Sci. Discuss.*, 6, 897–919, doi:10.5194/hessd-6-897-2009, 2009.
- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Sensitivity guided reduction of parametric dimensionality for multi-objective calibration of watershed models, *Adv. Water Resour.*, 32, 1154–1169, 2009.
- WMO: Intercomparison of Models of Snowmelt Runoff. Operational Hydrology Report No. 23, WMO Report No. 646, World Meteorological Organization, Geneva, Switzerland, 1986.
- Willems, P.: A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models, *Environ. Modell. Softw.*, 24, 311–321, 2009.
- Zhao, R. J.: Flood forecasting method for humid regions of China. East China College of Hydraulic Engineering, Nanjing, China, 1977.