Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Reducing the uncertainty of flood forecasts using multi-objective optimization algorithms for parameter estimation

YAN WANG, CHRISTIAN GATTKE & ANDREAS SCHUMANN

Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr-University Bochum, D-44801 Bochum, Germany

yan.wang-2@rub.de

Abstract The focus of this study was to characterize the extrapolation uncertainty resulting from different calibration strategies. A single-objective parameter estimation based on Monte Carlo simulations as well as a multi-objective optimization, are employed for the calibration. The extrapolation uncertainties that were obtained with these methods are evaluated with an extreme flood event. The results demonstrate that a unique parameter set, suitable for the entire hydrograph, does not exist. Utilization of a multi-objective optimization approach proved that the considerable uncertainty regarding model extrapolation originates from structural model inadequacies, which cause an inability of the model to reproduce all aspects of the hydrograph equally well with a single parameter set. It is suggested to use a multi-objective optimization strategy, which utilizes a problem-oriented definition of the performance measures to reduce the prediction uncertainty of peak flows. This method has been applied for flood modelling in Germany.

Key words flood forecasts; multi-objective optimization; extrapolation uncertainty

INTRODUCTION

Semi-distributed or even distributed hydrological models are used more and more for flood forecasts (Wagener & McIntyre, 2005). As the number of parameters of these models is usually high, the user faces the problem of model parameterization. Several parameters have to be calibrated comparing the computed and observed discharge. This applies to conceptual as well as to physically-based models. Often different parameter sets have to be used if such models are applied for continuous or event-based simulations. Various methods to identify an optimum parameter set have been developed during the past decades. Efficient global optimization strategies such as the Shuffled Complex Evolution algorithm (SCE; Duan *et al.*, 1993) are well established tools in hydrological modelling. The goal of optimization is to define a feasible and unique parameter set that yields best fits of the observations of the hydrological target variable, e.g. discharge. However, the results of optimization depend mainly on the chosen objective function. Another issue of importance is parameter equifinality. Equally good results in terms of the objective functions can be achieved by many different combinations of parameter values (Beven & Freer, 2001). To address this problem, simulation-based approaches that explore the feasible parameter space are nowadays widely used to assess the parameter uncertainties. Set theoretical approaches such as the Generalized Likelihood Uncertainty Estimation technique (GLUE, Beven & Binley, 1992) are based on random sampling procedures followed by a classification into non-behavioural and behavioural parameter sets. The latter are used to estimate uncertainty bounds for the model simulations/predictions. These procedures tend to result in considerable uncertainties in predicting peak flows with regard to the objective function (e.g. Beven & Binley, 1992; Harlin & Kung, 1992). An alternative to the random sampling strategy are multi-objective optimization approaches. These are based on the assumption that a model is incapable of reproducing all aspects of a hydrograph equally well with a single parameter set (Gupta *et al.*, 1998). Thus, calibration of hydrological models should be treated as a multi-objective optimization problem, which can be handled by a set of Pareto-optimal solutions.

Due to the nonlinearity of hydrological models, different sets of conceptual model parameters show considerable differences in the modelling results if the model has to be extrapolated, e.g. to provide flood forecasts. Within this study, the uncertainties of the extrapolation of a hydrological model to an extreme flood event are used to compare different parameter estimation procedures. Parameter sets are obtained from (a) ensembles of randomly sampled, so called equifinal parameters, and (b) multi-objective optimization and utilisation of Pareto-optimal parameters. Two different performance criteria were applied (Nash-Sutcliffe and the root mean square error) and different hydrological periods (ascent and recession) were used for the multi-objective optimization.

MODEL AND DATA

Hydrological model

The GIS-based hydrological model ArcEGMO (Becker *et al.*, 2002) is used in this study. It consists of three sub-models with different spatial resolutions: (a) a semi-distributed runoff generation model, (b) a lumped runoff concentration model, and (c) a distributed river routing model. The model concept is shown in Fig. 1. Eight

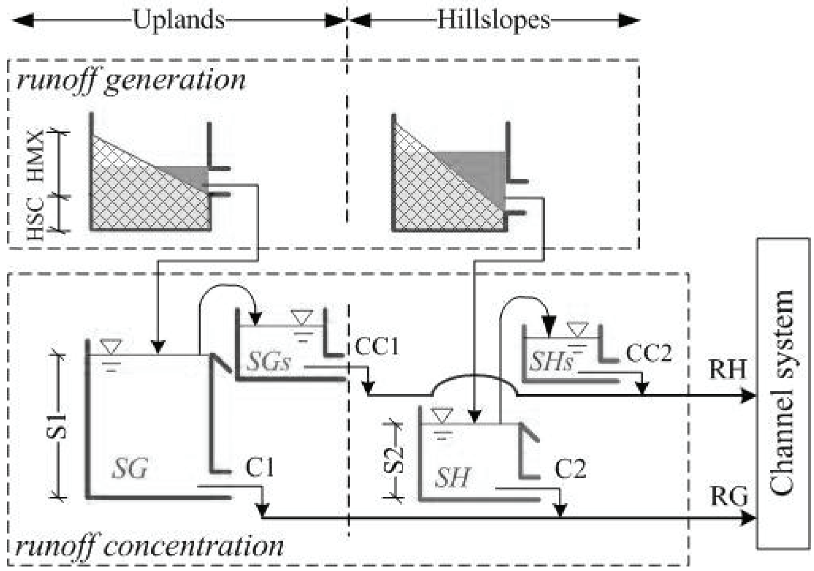

Fig. 1 Conceptual diagram of ArcEGMO.

Table 1 Parameters of ArcEGMO.

Parameters	Units	Type	Spatial units	Description	Initial value	Lower limit	Upper limit
HSC	mm	Semi-distributed	Hydrotope	min. soil storage capacity	-	-	-
HMX	mm	Semi-distributed	Hydrotope	max. soil storage capacity	-	-	-
C1	day	Lumped	Uplands	storage coefficient of SG	150	30	270
CC1	day	Lumped	Uplands	storage coefficient of SGs	8	1.6	16
S1	mm	Lumped	Uplands	storage capacity of SG	120	25	215
C2	day	Lumped	Hillslopes	storage coefficient of SH	5	1	10
CC2	day	Lumped	Hillslopes	storage coefficient of SHs	1	0.2	1.8
S2	mm	Lumped	Hillslopes	storage capacity of SH	20	4	35

parameters have to be determined by the user (see Table 1). The runoff generation sub-model uses hydrotopes to consider the spatial variability of the land use and soil related parameters HSC (minimum storage capacity) and HMX (maximum storage capacity). These two parameters define the total runoff volume. They were estimated from GIS-data in pre-processing. Here, the six remaining model parameters which define the shape of the hydrograph, are investigated. For the description of the runoff concentration processes, the catchment is partitioned into two types of contributing areas supplying mainly slow (uplands) and quick (hillslopes) runoff. If the minimum storage capacity HSC is exceeded during a rainfall event, water can percolate into the storages SG and SH (see Fig. 1). The amount of percolation depends on the distribution function of the maximum storage capacity HMX. Both reservoirs (SG and SH) drain accordingly to their storage coefficients (CC1 and CC2, respectively) into the channel system, forming a slow runoff component. If the storage capacities S1 or S2 are exceeded, water enters two additional reservoirs (SGs and SHs) with significant smaller storage coefficients, describing an accelerated runoff component. In total, the model structure contains thresholds and interacting flow paths which make the parameter estimation more difficult. In this study we investigate headwaters, thus channel routing is not considered here.

Data

The study was carried out for a watershed in the Ore Mountains region of eastern Germany (Fig. 2). It has an area of 363 km^2, with 28% uplands area and 72% hillslopes. Five summer events with an hourly time step were chosen for calibration, shown in Fig. 3. The August 2002 flood (event number 6 in Fig. 3) was used for validation. The flood events differ significantly in magnitude, with the flood in August 2002 being an extreme event. It was chosen for validation to demonstrate the capability of the model to be extrapolated to events that differ considerably from the calibration range. The initial state for each event was calculated from long-time simulation with daily time step and has subsequently been adjusted manually.

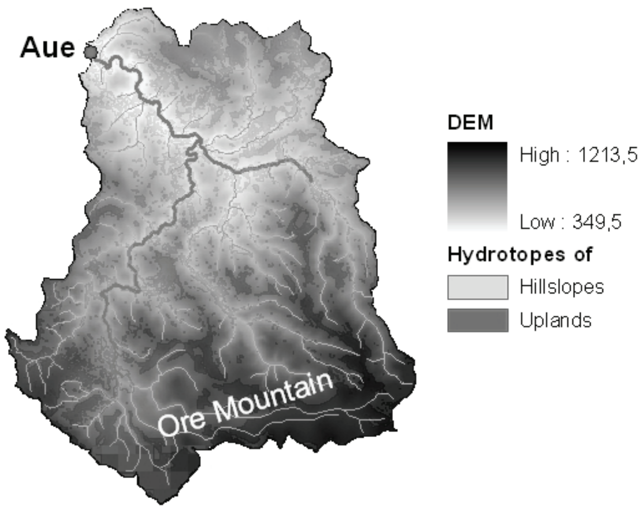

Fig. 2 Aue catchment in the Ore Mountains region of eastern Germany

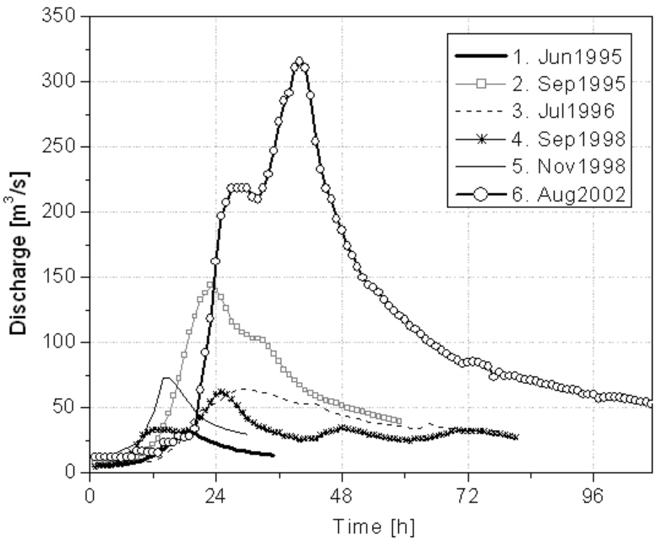

Fig. 3 Selected flood events with hourly time step: 1.–5. used for calibration, 6. used for validation.

PARAMETER ESTIMATION STRATEGIES COMPARED IN THIS STUDY

Four different calibration methods were analysed and compared (Table 2). Due to the reasonable assumption that the catchment response is inherently different during rainfall periods and periods without rain (Boyle *et al*., 2000), different parameter sets may become necessary to simulate both aspects of the hydrograph. Thus, objective functions quantifying the goodness-of-fit of the entire hydrograph as well as different parts of the hydrograph (i.e. ascent and recession periods) should be evaluated. Figure 4 demonstrates the partitioning of the observed hydrograph.

The first two cases employed here (cases 1 and 2 in Table 2) were based on Monte Carlo simulations (10 000 parameter sets, uniform sampling) using different

Table 2 Different calibration cases considered in this study.

Nr.	Tool	Performance measure	Solutions for evaluation	Objective function OF
Case 1	Monte Carlo	*NS*	100 from 10 000	Single OF: Entire hydrograph
Case 2	Monte Carlo	*RMSE*	100 from 10 000	Single OF: Entire hydrograph
Case 3	MOCOM	$RMSE_A$, $RMSE_R$	100	OF1: Ascent period; OF2: Recession period
Case 4	MOCOM	$RMSE_A$, $RMSE$	100	OF1: Ascent period; OF2: Entire hydrograph

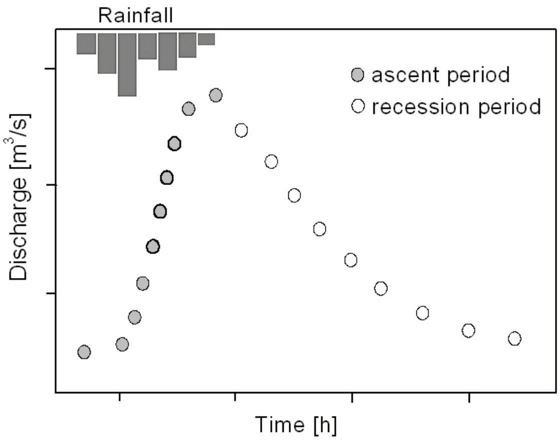

Fig. 4 Partitioning of the observed hydrograph into ascent and recession periods (edited from Boyle *et al.*, 2000).

performance measures, the Nash-Sutcliffe coefficient *NS* in case 1:

$$NS = \frac{1}{M}\sum_{i=1}^{M}\left[1-\sum_{j=1}^{N}\left(Q_{obs,j}-Q_{sim,j}\right)^2 \Big/ \sum_{j=1}^{N}\left(Q_{obs,j}-\overline{Q}_{obs}\right)^2\right] \tag{1}$$

and the overall root mean square error *RMSE* in case 2:

$$RMSE = \frac{1}{M}\sum_{i=1}^{M}\left[\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(Q_{obs,j}-Q_{sim,j}\right)^2}\right] \tag{2}$$

where M is the number of simultaneously simulated events, N is the total number of observations during an event, j is the time step and Q is discharge. The subscripts *sim* and *obs* indicate the simulated and observed values. To minimize the predictive uncertainty, a restrictive criterion for model performance has been used. Thus, only the 100 best solutions (1% of overall simulations) with respect to the performance measures were accepted as behavioural for further evaluation.

In cases 3 and 4, a multi-objective optimization procedure was applied. The multi-objective calibration tool MOCOM (Yapo & Gupta, 1998) was used to estimate 100 Pareto optimal solutions between two different objective functions: (a) ascent *vs* recession period in case 3, and (b) ascent period *vs* entire hydrograph in case 4. The model performance was assessed in cases 3 and 4 for different parts of the hydrograph separately. The suffix *A* indicates the *RMSE* of the rising limb of the hydrograph (ascent period) and the suffix *R* the falling limb (recession period).

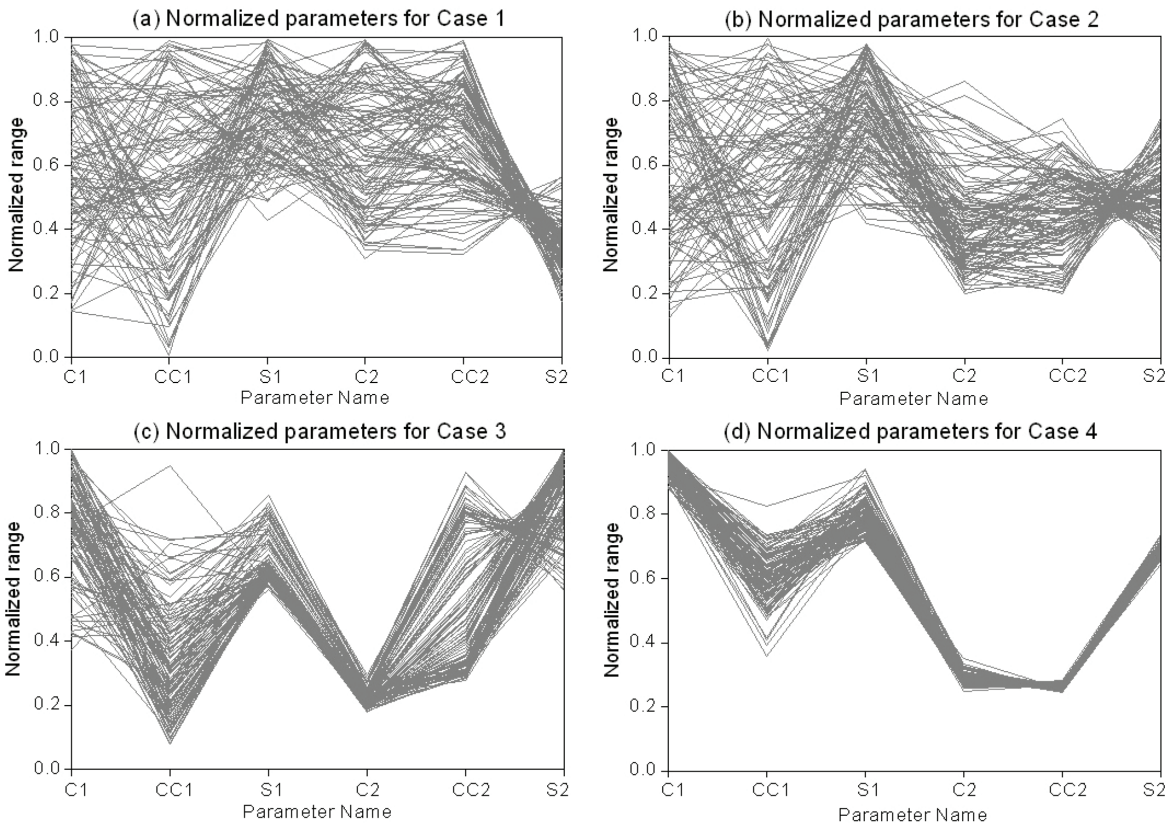

Fig. 5 Normalized parameters for the 100 solutions obtained with the different calibration methods: (a) Case 1, (b) Case 2, (c) Case 3, and (d) Case 4.

Table 3 Normalized parameter ranges (0.05–0.95 quantiles, supplementing Fig. 5).

	C1	CC1	S1	C2	CC2	S2
Case 1	0.211–0.948	0.113–0.954	0.543–0.972	0.369–0.975	0.422–0.941	0.212–0.494
Case 2	0.218–0.950	0.075–0.927	0.479–0.961	0.244–0.715	0.237–0.659	0.357–0.700
Case 3	0.460–0.984	0.116–0.657	0.587–0.806	0.187–0.270	0.295–0.850	0.615–0.984
Case 4	0.900–0.996	0.482–0.718	0.726–0.896	0.265–0.321	0.249–0.276	0.667–0.720

RESULTS

Parameter ranges

Figure 5(a)–(d) shows the variability in the estimated parameter values across the 100 parameter sets that were accepted as behavioural (cases 1 and 2) and the Pareto optimal solutions (cases 3 and 4), using a normalized parameter plot. Each grey line represents one of these parameter sets. The variability ranges between the 0.05–0.95 quantiles for each parameter are shown in Table 3. It should first be noted that the multi-objective automatic calibration results in a significant reduction of the parameter range (comparison of cases 1–2 with 3–4). Secondly, it is remarkable that the parameters CC2 and S2 differ significantly in their ranges between all cases. These parameters are responsible for interflow from the fast-reacting reservoir, connected with the hydrotopes of the hillslopes. The hillslopes cover a large percentage of the entire catchment area (72%), hence their corresponding parameters CC2 and S2 are much more sensitive than the other model parameters.

Characterization of extrapolation uncertainty: validation with extreme flood event

The extreme flood event in 2002 was used to compare the extrapolation uncertainty resulting from different calibration strategies. For the Monte Carlo simulations the 100 parameter sets with the highest value of the performance measure (*NS* or *RMSE*) were accepted. The Pareto-fronts derived from MOCOM were also specified by 100 parameter sets. To compare the resulting uncertainty in extrapolation, confidence bands based on the 0.05 to 0.95 quantiles were used. The results are shown in Fig. 6(a)–(d). In the following these results are compared in pairs.

Case 1 *vs* Case 2 In both cases, single-objective calibration with different performance measures was used. The *NS* coefficient in case 1 indicates an overall agreement between observed and simulated time series with less preference for higher values. The *NS* (Nash-Sutcliffe) coefficient can be seen as a transformed and normalised measure of the overall *RMSE* in case 2 (normalised with respect to the variance of the observed hydrograph, according to Madsen, 2000). The *RMSE* tends more to minimize peak flow errors. It was proved by visual comparison for the validation between Fig. 6(a) (case 1 with NS) and 6(b) (case 2 with *RMSE*) that the Fig. 6(b) demonstrates a closer fit to the observed data in terms of peak flow. Furthermore, a strong deviation between observed and simulated values during the recession period is evident in both cases. Because both the peak and timing are more important for flood forecasting, the *RMSE* should be preferred as a performance measure.

Case 2 *vs* Case 3 This is a comparison between single- and multi-objective calibrations. In case 3, $RMSE_A$ and $RMSE_R$ are associated with two non-commensurable objective functions reflecting the model performances in the ascent and recession periods. The Pareto-optimal solutions correspond to various trade-offs among the two objectives (ascent and recession period). The performance for the ascent period (which is most important for flood forecasting) cannot be improved along this frontier without worsening the model performance during the recession period. The incompatibility between the two objectives results in considerable uncertainty regarding the extrapolation results (Fig. 6(c)), which in turn are comparable to the results for cases 1 and 2. This can be interpreted as an indication of structural model inadequacies, which cause an inability of the model to reproduce both aspects of the hydrograph equally well with a single parameter set. Furthermore, it seems that the Pareto-optimal solutions cannot ensure the optimal performance during peak flow, which however, was reached by single-objective calibration in case 2 (Fig. 6(b)). This is due to the restriction to a 90% range of the simulated runoff values, i.e. the multi-objective calibration identified fewer solutions matching the peak flow than the Monte Carlo simulations.

Case 3 *vs* Case 4 Both are multi-objective calibrations with different objective functions. Case 3 has two non-commensurable objective functions $RMSE_A$ and $RMSE_R$. In case 4 the ascent period is used for both criteria. This procedure provides the same result as a weighting of the ascending period in multi-objective optimization.

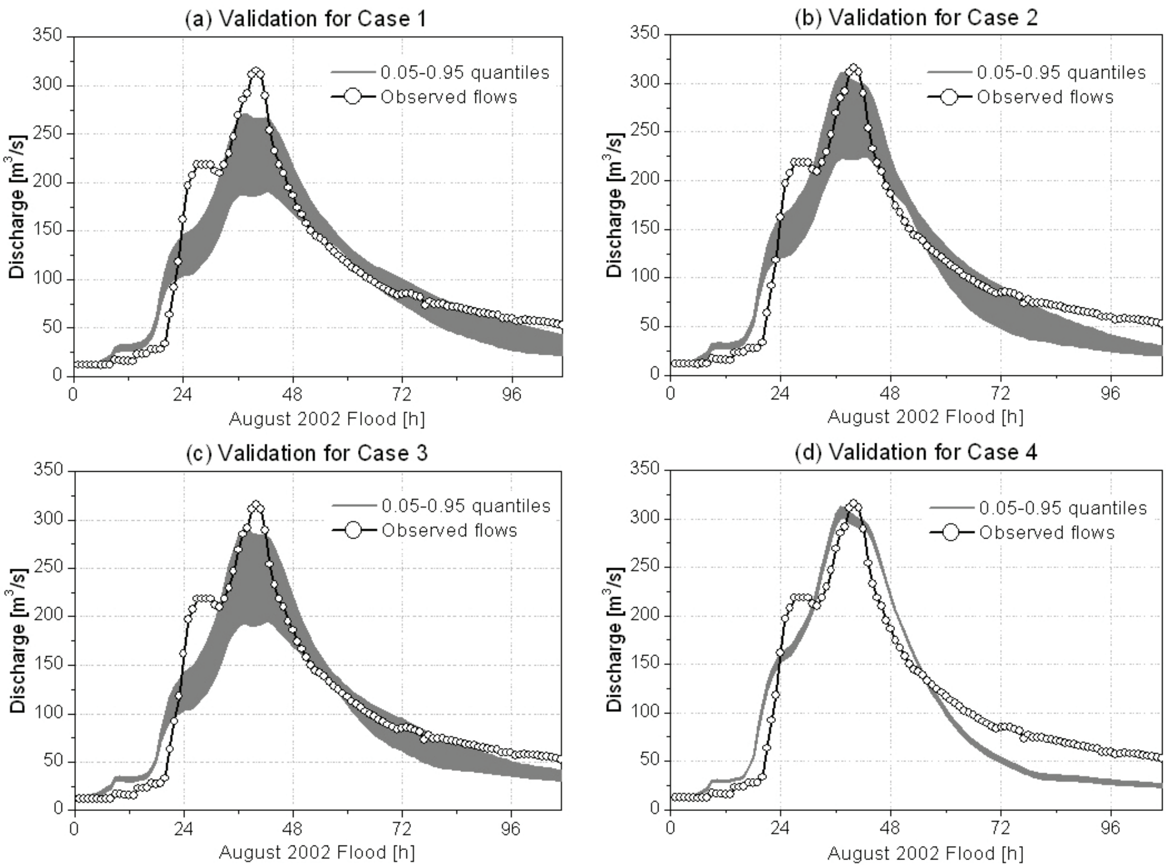

Fig. 6 Validation with the different parameters: (a) Case 1, (b) Case 2, (c) Case 3, and (d) Case 4.

Because further reduction of prediction uncertainty in peak flow and timing is expected, case 4 improves the calibration strategy with additional impact on the ascent period. The objective functions of case 4 are $RMSE_A$ and $RMSE$. The overall $RMSE$ comprises the $RMSE_A$ in the ascent period. This means, both objective functions have considered the fit of the ascent period. Comparison of Fig. 6(c) (case 3) and Fig. 6(d) (case 4) clearly shows that the loss of objectivity of Pareto-optimal solutions leads to a better prediction accuracy for peak flows.

CONCLUSIONS AND OUTLOOK

The study presented here compares different parameter estimation strategies and the corresponding extrapolation uncertainty. Four different calibration methods were analysed and compared. Two of them were based on uniform random sampling using different performance measures, and the other two on multi-objective optimization using different combinations of objective functions. Due to the restriction on non-dominated Pareto-optimal solutions (Gupta *et al.*, 1998) the multi-objective optimization leads to significantly smaller ranges for the parameter values, i.e. the parameters appear more identifiable. Comparing the results obtained by random sampling (case 1 *vs* case 2) *RMSE* was shown to exceed *NS* as a performance measure, in particular for peak flow and timing prediction, both crucial in flood forecasting. The multi-objective optimization combining the performance measures for the ascent and the recession periods (case 3) indicates that the considerable uncertainty regarding the extrapolation results in the case of random sampling using a single performance measure (often referred to as parameter equifinality), originate from structural model inadequacies, which cause an inability of the model to reproduce both aspects of the hydrograph equally well with a single parameter set. Case 4 (combination of performance measures for the ascent period and the entire hydrograph) demonstrated that a multi-objective optimization, utilizing a problem-oriented definition of the performance measures, can lead to a significant reduction of the extrapolation uncertainty.

Each calibrated parameter can be characterized by its properties, e.g. uncertainty and sensitivity (Haimes, 1998). These properties were illustrated with Fig. 5. It has been found that parameters with a small variability range have high sensitivity and small uncertainty. In addition, the different position of the parameter ranges in Fig. 5(c)–(d) of cases 3–4 provides further information for the parameter identification. Such information and the representative characteristics of each parameter will be investigated in future work and to develop an effective parameter tracking scheme for real time flood forecasting system.

REFERENCES

Becker, A., Kloecking, B., Lahmer, W. & Pfuetzner, B. (2002) The hydrological modelling system Arc/EGMO. In: *Mathematical Models of Large Watershed Hydrology* (ed. by V. P. Singh & D. K Frevert), 321–384. Water Resources Publ., Highlands Ranch, Colorado, USA.

Beven, K. & Binley, A. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Processes* **6**, 279–298.

Beven, K. & Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249**, 11–29.

Boyle, D. P., Gupta, H. V. & Sorooshian, S. (2000) Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resour. Res.* **36**(12), 3363–3674.

Duan, Q., Gupta, V. K. & Sorooshian, S. (1993) A shuffled complex evolution approach for effective and efficient global minimization. *J. Optimization Theory and Applications* **76**(3), 501–521.

Gupta, H. V., Sorooshian, S. & Yapo, P. O. (1998) Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resour. Res.* **34**(4), 751-764.

Haimes, Y. Y. (1998) The uncertainty sensitivity index method. In: *Risk Modelling, Assessment, and Management.* (ed. by A. P. Sage), Chapter 6.6, 252–258. John Wiley & Sons, New York, USA.

Harlin, J. & Kung, C. S. (1992) Parameter uncertainty and simulation of design flood in Sweden. *J. Hydrol.* **137**, 209–230.

Madsen, H. (2000) Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.* **235**, 276–288.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1998) Multi-objective global optimization for hydrologic models. *J. Hydrol.* **204**, 83–97.

Wagener, T. & McIntyre, N. (2005) Identification of rainfall–runoff models for operational applications. *Hydrol. Sci. J.* **50**(5), 735–751.