

Comparison of different approaches to quantify the reliability of hydrological simulations

C. Gattke and A. Schumann

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

Received: 17 January 2007 – Revised: 11 April 2007 – Accepted: 4 May 2007 – Published: 16 May 2007

Abstract. The focus of this study was to compare different uncertainty estimation approaches to evaluate their ability to predict the total amount of uncertainty in hydrological model predictions. Three different approaches have been compared. Two of them were based on Monte-Carlo sampling and the third approach was based on fitting a probability model to the error series of an optimized simulation. These approaches have been applied to a lumped and a semi-distributed model variant, to investigate the effects of changes in the model structure on the uncertainty assessment. The probability model was not able to predict the total amount of uncertainty when compared with the Monte-Carlo based approaches. The uncertainty related to the simulation of flood events was systematically underestimated.

1 Introduction

Hydrological models are common tools for water resources planning and management. They are used to e.g. predict water balances or extreme events (floods and droughts), to extrapolate discharge time-series or to evaluate management strategies. Despite the efforts devoted to model development within the past decades, there is still significant uncertainty associated with hydrological simulations (Montanari and Brath, 2004). This applies to both conceptual and physically based models. The uncertainty in hydrological model output stems from three main causes: (i) the model uncertainty, which denotes incompatibilities between the structures represented in the model and the structures present in the hydrological system, (ii) the parameter uncertainty and (iii) the data uncertainty, which applies to the input data (e.g. precipitation and temperature) as well as to the data used for calibration (e.g. discharge). Recognizing the problem of un-

certainty, many different approaches to quantify the reliability of hydrological simulations have been proposed in recent years (Montanari and Brath, 2004).

Conventional “point” methods, such as first-order analysis, calculate the mean and variance of the predicted variable, based on the mean and variance of uncertain inputs and/or parameters. They rely on the assumption that a single linearization of the system performance function at the central values of the basic variables is representative of the statistical properties of system performance over the complete range of basic variables (Melching, 1995). These point-estimation methods are limited by the assumption of approximate linearity of the model (Gupta et al., 2005), which is often not suitable for hydrological models. Due to this problem, simulation based approaches that explore the feasible parameter space and/or consider an uncertainty range for the data sets are commonly used for uncertainty assessment nowadays. The most used simulation based approach for inferring the global simulation uncertainty is probably the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992). A third option to estimate the simulation uncertainty is to analyze the statistical properties of the model error series that occurred in reproducing observed discharge data. An example for such an approach is the Meta-Gaussian model, recently presented by Montanari and Brath (2004).

However, only few studies compare different uncertainty estimation approaches (e.g. Montanari and Brath, 2004; McIntyre et al., 2002) to evaluate their ability to predict the total amount of uncertainty in hydrological model predictions. The goal of the study presented here was to carry out such a comparison between three different approaches (two simulation-based and a stochastic error model). The three uncertainty estimation approaches have been applied to two model variants with different spatial discretizations to investigate the effects of changes in the model structure on the uncertainty assessment.

Correspondence to: C. Gattke
(christian.gattke@rub.de)

2 Model structures and data

The two model variants used in this study were implemented following the concept of the HBV-96 model (Lindström et al., 1997). One significant difference lies in the generation of surface runoff. Within the original HBV model all water from the soil routine that does not evaporate recharges into the upper response box. In contrast to this, for the model applications presented here, surface runoff can occur from saturation excess and is routed to the stream network via simple linear storage. The triangular weighting function to smooth the generated discharge has been excluded, owing to daily time step discretization and detailed structuring of the area under investigation into subbasins. The Muskingum method was used for the flood routing within the river reaches. The first model variant used lumped representations of all vertical model components within the subbasins. The second model variant used semi-distributed representations of the interception layer (depending on land use distribution), the root zone layer (combination of land use and soil types) and the upper response layer (distribution of soil types). The recession constants of the upper response boxes were scaled depending on transmissivity and average distance to the stream network. Implementation of both variants was based on an object-oriented framework designed for adaptive development of hydrological models (Gattke and Pahlow, 2006). The study was carried out for a mesoscale catchment located at the upper Werra river in Mid-East Germany (Meiningen, 1167 km²). The catchment was divided into 39 subbasins. A six-year period (Nov 1991–Oct 1997) of observed discharge and meteorological data was available. The issue of appropriate data lengths for model identification has been investigated by a variety of studies, the general result being that the required length mainly depends on data quality, model complexity and climatic variability (Wagener and McIntyre, 2005). Considering the results of Wagener and McIntyre who used a three-year calibration period to compare different objective functions, the six-year period used here was deemed to be adequate. The simulations were carried out at a daily time-step. The maximum peak discharge observed in this period was 181 m³/s (April 1994). The average discharge was 14.5 m³/s. The average annual precipitation depth over the basin area was 838 mm/yr.

3 Uncertainty assessment

Three different uncertainty estimation approaches (two simulation-based and a stochastic error model) were used in this study: a) 0.05- and 0.95-quantiles of the simulated discharge values, b) the quantiles estimated with the GLUE likelihood-weighting procedure and c) the confidence intervals obtained by application of the Meta-Gaussian model of Montanari and Brath (2004). Montanari and Brath presented an approach to estimate the uncertainty of a calibrated hydro-

logical model. They make use of a Meta-Gaussian model in order to estimate the probability distribution of the model error conditioned by the simulated discharge. Model residues and model simulations are considered as realisations of two stochastic processes. The standard normal quantile transform (NQT) is used to make the marginal probability distributions of both Gaussian. The normalised series are assumed to be stationary and ergodic. This approach can be applied to models with a varying complexity. Rejecting the concept of an optimum parameter set, Beven and Binley (1992) developed their generalized likelihood uncertainty estimation (GLUE) method. First, a large number of parameter sets is generated randomly via sampling from assumed prior parameter distributions (usually uniform). After performing the simulation trials the parameter sets are weighted with an arbitrary chosen goodness-of-fit criterion (likelihood measure). This criterion is employed to differentiate behavioural and non-behavioural parameter sets and to reject the latter ones. The likelihood weights of the accepted parameter combinations are rescaled to produce a cumulative sum of one and used to construct a cumulative distribution function of the simulated discharge values. This allows uncertainty bounds to be derived at every time step.

Both the GLUE approach and the Meta-Gaussian model are supposed to consider all sources of uncertainty implicitly. In case of the GLUE procedure this occurs through the likelihood weighting (Beven and Freer, 2001). To verify this assumption the 0.05- and 0.95-quantiles of the simulated discharge values were also included in the comparison shown here. The Nash-Sutcliffe efficiency (NSE) was used as likelihood measure (Nash and Sutcliffe, 1970). The rejection criterion shows considerable effects on the simulation range and therefore on the possibility that observed values are lying outside of this range (Montanari, 2005). Thus this criterion should be chosen carefully on the one hand to avoid a simulation range too narrow and on the other hand to avoid blurring of the models predictive performance by an extremely wide simulation range. Here the 3rd quartile of the simulations has been used as rejection criterion. In total, 10 000 simulations with parameter sets obtained by randomly generated, uniformly distributed values were performed with both model variants (lumped and semi-distributed). The simulations with the maximum NSE values were selected to estimate the model uncertainty by means of the Meta-Gaussian approach.

4 Results

The results obtained for the maximum NSE and the 3rd quartile values were 0.905 and 0.787 for the lumped model variant and 0.919 and 0.792 for the semi-distributed variant respectively. Hence simulation runs with NSE less than 0.787 (lumped) and 0.792 (semi-distributed) had been rejected.

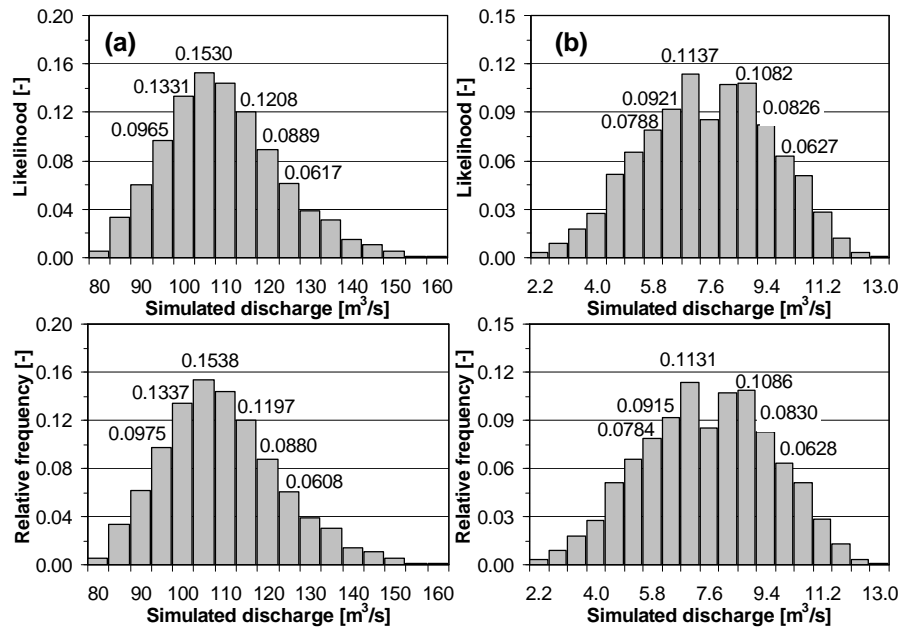


Fig. 1. Posterior distribution of the GLUE likelihood weighted model predictions (top) and the relative frequency of the simulated discharge values (bottom), (a) extreme winter flood (30 Jan 1995), (b) single daily discharge value out of a low-flow period during summer (12 Aug 1995).

4.1 Comparison of the simulation-based prediction ranges

The comparison of the prediction quantiles estimated with GLUE and the quantiles of the simulated discharge values confirmed, as expected, that these are not identical, i.e. the 90% GLUE prediction limits do not comprise 90% of the simulated discharge values. Overall the differences between both quantile ranges were negligible. The average absolute difference of the compared prediction limits over the entire simulation period of six years was less than 0.03 m³/s for both model variants. This corresponds to a difference of 0.1% for the upper and lower limit. This consistency arises from the fact that the posterior distribution of the likelihood weighted model predictions, as derived from the GLUE procedure and the relative frequency of the simulated discharge values, are in good agreement. Figure 1 exemplifies this for two simulation time steps with different flow magnitudes. Shown are the peak of an extreme winter flood (30 Jan 1995, Fig. 1a) and a single daily discharge value out of a low-flow period during summer (12 Aug 1995, Fig. 1b). Furthermore, this example confirms that the distributions have characteristics with changing shape and variance over time as demonstrated by Beven and Freer (2001). The distributions are clearly non-Gaussian for the peak-flow. More substantial differences between the prediction ranges do not occur until 90% of all simulations out of the Monte-Carlo sample (this corresponds to a threshold of about 0.3 for the NSE) are accepted as behavioural and are included in the uncertainty analysis. However, this would be an implausible

Table 1. Average uncertainty ranges (in m³/s) over the six-year period for different magnitude classes of the observed discharge (MQ = average discharge = 14.5 m³/s).

Model structure Approach	Lumped		Semi-distributed	
	GLUE	Meta-Gauss	GLUE	Meta-Gauss
Overall	9.5	12.9	10.0	11.6
>3 MQ	44.4	21.6	40.8	28.4
<MQ	6.1	15.0	6.6	13.4

choice. Thus, further discussion is restricted to the comparison of the GLUE method and the Meta-Gaussian model of Montanari and Brath (2004).

4.2 Comparison of simulation-based prediction ranges and the confidence intervals estimated with the Meta-Gaussian Model

Figure 2 compares the observed discharge values over the six-year period that lie outside the different uncertainty ranges for both model variants. The Median GLUE output has been chosen as a representative for the simulation range. In both cases the observed discharge was systematically underestimated in the high flow regime by the Median GLUE. This is due to skewness of the GLUE simulation range at time steps with high-flows, as shown in the previous section. For

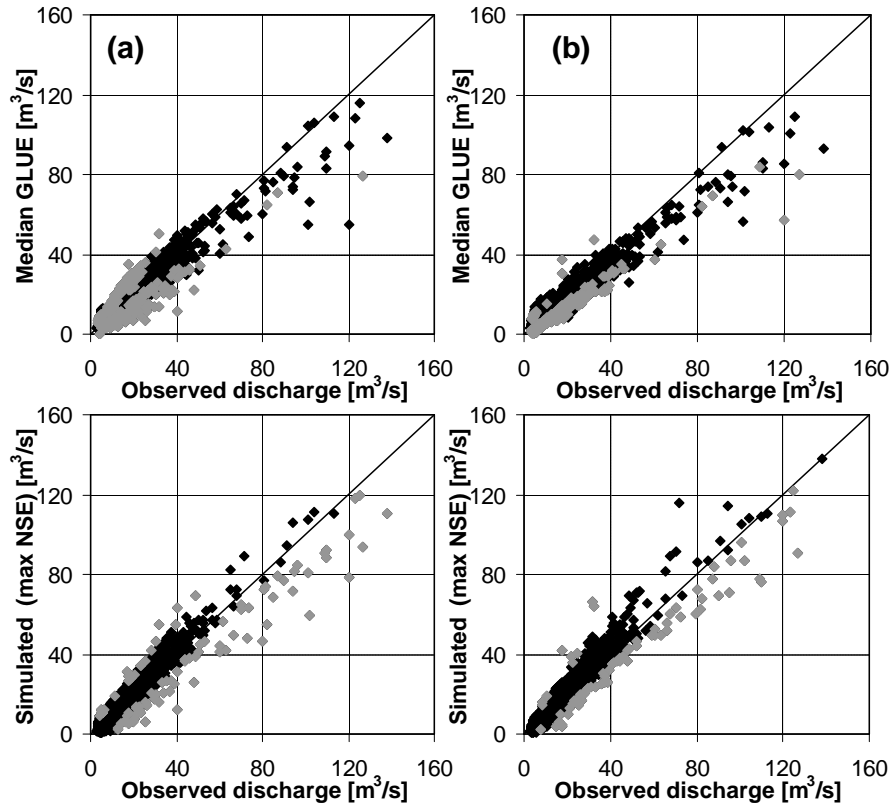


Fig. 2. Dotty plots of Median GLUE vs observed discharge (top) and simulated vs observed discharge (bottom) for the six-year period. Grey dots indicate observed discharge values lying outside the uncertainty ranges of GLUE (top): (a) lumped model variant (18.8% of overall values, 12.5% of values >3 MQ), (b) semi-distributed model variant (16.1%, 12.5%) and the Meta-Gaussian model (bottom): (a) lumped model variant (4.6%, 46.6%), (b) semi-distributed model variant (3.8%, 42.0%).

the lumped model 18.8% of the overall observed discharge values and 12.5% of the high-flow values (>3 times of average discharge MQ) were lying outside the 90% GLUE simulation range (see Fig. 2a, top). The model performs with lowest accuracy for the low-flows and smaller flow peaks during the summer period, which are mostly underestimated. This can be interpreted as a systematic underestimation of the soil moisture dynamics for dry conditions due to lumped model representation of the runoff generating processes. Application of the semi-distributed structure resulted in better model performance, illustrated by the reduced scatter in the mid- and low-flow region (see Fig. 2b, top). Thus, the percentage of observed values lying outside the prediction ranges was reduced to 16.1, while the percentage of the high-flow values remained unchanged.

The analysis of the uncertainty ranges estimated with the Meta-Gaussian model leads to different results. The amount of observed values lying outside the uncertainty ranges reduces to 4.6% (lumped) and 3.8% (semi-distributed). On the other hand, the amount of high-flow values lying outside is increased to 46.6% and 42.0% respectively (see Fig. 2a and b, bottom). This is due to opposite trends regarding the quan-

tification of the uncertainty for different regions of the flow regime. Table 1 summarizes the average uncertainty ranges estimated for different magnitude classes of the observed discharge. The uncertainty in the low-flow regions (observed values smaller than average discharge MQ) estimated by the Meta-Gaussian model is about two times higher than the uncertainty obtained with the GLUE approach. The situation is approximately reversed for the high flows (observed values >3 MQ). Figure 3 compares the dependency of the uncertainty range on the magnitude of the simulated discharge. In case of the Meta-Gaussian model the basic hypothesis that the cross dependence between the simulated discharge and the model error is governed by a normal linear equation leads to a monotonically increasing relation. In contrast to this, the relationships for uncertainty ranges estimated with GLUE exhibit an increasing trend, but this is connected with considerable scatter. This may be due to different hydrological situations (e.g. ascent or recession period of the hydrograph, influence of snowmelt etc.).

Figures 4 and 5 compare the uncertainty ranges for one year out of the six-year period. Due to the moderate performance of the lumped model in the summer period (the

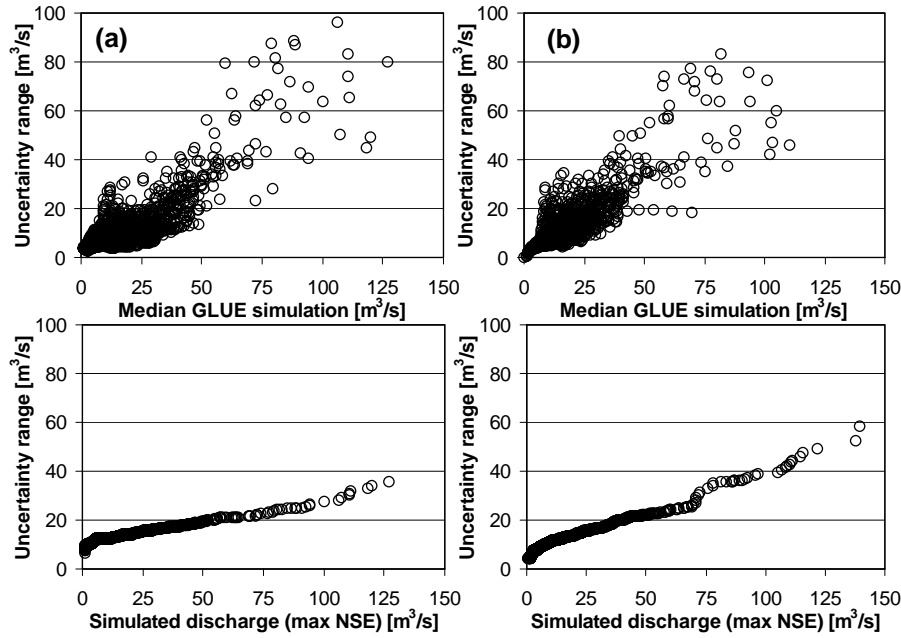


Fig. 3. Dependency of the uncertainty ranges of GLUE (top) and the Meta-Gaussian model (bottom) on the magnitude of the simulated discharge, (a) lumped model variant, (b) semi-distributed model variant.

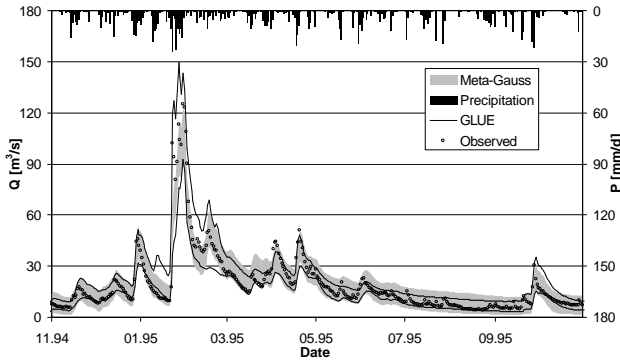


Fig. 4. Comparison of the uncertainty ranges for one year out of the six-year period (lumped model variant).

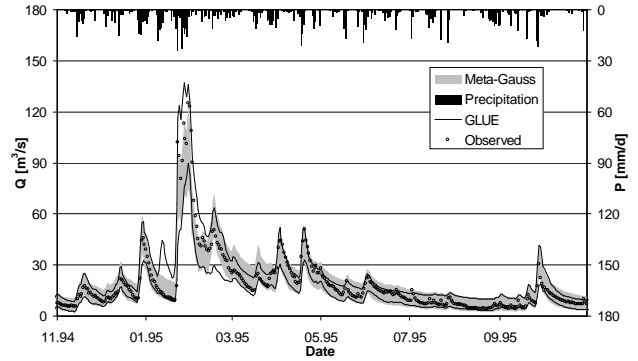


Fig. 5. Comparison of the uncertainty ranges for one year out of the six-year period (semi-distributed model variant).

small flow fluctuations cannot be reproduced, as indicated by the GLUE prediction ranges) the Meta-Gaussian model tends to overestimate the uncertainty, compared to the results obtained with GLUE. On the other hand the uncertainty ranges show stronger conformity for the semi-distributed model. However, the uncertainty connected with the simulation of the extreme winter event seems to be underestimated in both cases by the Meta-Gaussian model.

5 Conclusions

The main objective of the study presented here was to compare different uncertainty estimation approaches to evaluate their ability to predict the total amount of uncertainty in hydrological model predictions. Three different approaches have been compared. Two of them were based on Monte-Carlo sampling and the third approach was based on fitting a probability model to the error series of an optimized simulation (i.e. the one with the best model performance out of the Monte-Carlo sample). These approaches have been applied to a lumped and a semi-distributed model variant, to investigate the effects of changes in the model structure on the

uncertainty assessment. The 0.05- and 0.95-quantiles of the simulated discharge values and the corresponding quantiles estimated with the GLUE likelihood-weighting procedure exhibited only negligible differences. This result may require verification, e.g. the application of different likelihoods measures (including *real* likelihood functions) and/or different model structures. However, experiments that included more simulations from the Monte-Carlo sample for the uncertainty assessment suggested, that the likelihood weighting leads to a smoothing of the predicted uncertainty range. Thus, the assumption that the GLUE approach handles implicitly any effects of errors in the model structure and the data appears questionable. Particularly the latter may require explicit consideration, e.g. by application of a range (or different realizations) of the input data (e.g. Haberlandt and Gattke, 2004). The Meta-Gaussian model was not able to predict the total amount of uncertainty when compared with the GLUE approach. The uncertainty related to the simulation of flood events was systematically underestimated. On the other hand, the different estimation of uncertainty in the summer period for both model variants (lumped and semi-distributed) strongly indicated a structural inadequacy of the lumped model. However, the quantity of the difference was not justified by the GLUE approach.

Edited by: K.-E. Lindenschmidt

Reviewed by: Y. Wang and T. Weichel

References

- Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, 1992.
- Gattke, C. and Pahlow, M.: Using object oriented methods for adaptive hydrological model development and uncertainty estimation, in: *Proceedings of the 7th International Conference on Hydroinformatics*, edited by: Gourbesville, P., Cunge, J., Guinot, V., and Liong, S.-Y., 2, 1317–1324, 2006.
- Gupta, H. V., Beven, K. J., and Wagener, T.: Model calibration and uncertainty estimation, in: *Encyclopedia of hydrological sciences*, edited by: Anderson, M. G. and McDonnell, J. J., et al., John Wiley & Sons Ltd., Chichester, 2005.
- Haberlandt, U. and Gattke, C.: Spatial interpolation vs. simulation of precipitation for rainfall-runoff modelling – a case study in the Lippe river basin, in: *Hydrology: Science and practice for the 21st century*, edited by: Webb, B., Acreman, M., Maksimovic, C., Smithers, H., and Kirby, C., *Proceedings of the British Hydrological Society International Conference*, 1, 120–127, 2004.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 model, *J. Hydrol.*, 201, 272–288, 1997.
- McIntyre, N., Wheater, H., and Lees, M.: Estimation and propagation of parametric uncertainty in environmental models, *J. Hydroinformatics*, 4(3), 177–198, 2002.
- Melching, C. S.: Reliability estimation, in: *Computer models of watershed hydrology*, edited by: Singh, V., *Water Resources Publications*, 69–118, 1995.
- Montanari, A.: Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41(8), W08406, doi:10.1029/2004WR003826, 2005.
- Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 40(1), W01106, doi:10.1029/2003WR002540, 2004.
- Nash, J. E. and Sutcliffe, A. Y.: River flow forecasting through conceptual models 1. A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Wagener, T. and McIntyre, N.: Identification of rainfall-runoff models for operational applications, *Hydrolo. Sci. J.*, 50(5), 735–751, 2005.