Jour. Korean For. Soc. Vol. 95, No. 5, pp. 569~579 (2006)

JOURNAL OF KOREAN FOREST SOCIETY

Evaluation of Multi-criteria Performances of the TOPMODEL Simulations in a Small Forest Catchment based on the Concept of Equifinality of the Multiple Parameter Sets

Hyung Tae Choi*, Kyongha Kim, Jae-Hong Jun, Jae-Yun Yoo and Yong-Ho Jeong

Department of Forest Environment, Korea Forest Research Institute, Seoul 130-712, Korea

Abstract : This study focuses on the application of multi-criteria performance measures based on the concept of equifinality to the calibration of the rainfall-runoff model TOPMODEL in a small deciduous forest catchment. The performance of each parameter set was evaluated by six performance measures, individually, and each set was identified as a behavioral or non-behavioral parameter set by a given behavioral acceptance threshold. Many behavioral parameter sets were scattered throughout the parameter space, and the range of model behavior and the sensitivity for each parameter varied considerably between the different performance measures. Sensitivity was very high in some parameters, and varied depending on the kind of performance measure as well. Compatibilities of behavioral parameter sets between different performance measures also varied, and very few parameter sets were selected to be used in making good predictions for all performance measures. Since different behavioral parameter sets with different likelihood weights were obtained for each performance measure, the decision on which performance measure to be used may be very important to achieve the goal of study. Therefore, one or more suitable performance measures should be selected depending on the environment and the goal of a study, and this may lead to decrease model uncertainty.

***Key words* :** *equifinality, TOPMODEL, multi-criteria performance measures, behavioral parameter set, sensitivity, compatibility, likelihood weight, deciduous forest catchment*

Introduction

The rainfall-runoff process, which is an important compartment of the hydrological system, is very complex considering the large number of factors involved and their variability in time and space. Hydrological modeling is a powerful technique to represent the rainfall-runoff process in various physical or mathematical forms using known or assumed functions expressing the various components of a rainfall-runoff response (Ndiritu and Daniell, 1999). In the last half-century there have been hundreds of hydrological response models, each with their own attributes and shortcomings, developed by many different researchers. Furthermore, with the current rapid developments within computer technology and hydrology, the application of computer based hydrologic models is only likely to increase in the near future (Loague and VanderKwaak, 2004).

The distributed hydrological models aim to better represent the spatio-temporal variability of hydrological characteristics governing the rainfall-runoff response at the catchment scale (Vieux *et al.*, 2004). One of the distributed hydrological models used commonly is TOPMODEL, which is a quasi-physically based semi-distributed hydrological model (Beven and Kirby, 1979; Beven, 2001, Beven and Freer, 2001a).

Most physically based distributed models have parameters which are effective at the scale of the computational elements. In order for a rainfall-runoff model to have practical use or be useful for hypothesis testing, it is necessary to select appropriate values for the model parameters. Unfortunately, it is not normally possible to estimate the effective values of parameters by either prior estimation or measurement, even given intensive series of measurements of parameter values. Therefore, parameter values must be calibrated for individual applications (Refsgaard and Knudsen, 1996; Refsgaard, 1997; Freer, 1998; Beven, 2001).

In general, the process of parameter calibration has involved some form of determination of a parameter set that gives a simulation that adequately matches the

*Corresponding author
E-mail: choiht@paran.com

observation. However, many calibration studies in the past have revealed that while one optimum parameter set could often be found, there would usually be a multitude of quite different parameter sets that can produce almost equally good simulation results. Recognition of multiple acceptance of parameter sets results in the concept of equifinality of parameter sets. Indeed, to focus attention on a rejection of the concept of the optimal model in favor of multiple possibilities for producing simulations that are acceptable simulators in some sense, this idea has been called elsewhere equifinality. The concept of equifinality has many applications as a new working paradigm for parameter calibration and uncertainty estimation of hydrological models (Beven, 2002; Beven and Freer, 2001b; Freer *et al.*, 2003).

In addition, in the general case of rainfall-runoff modeling with multiple storm sequences, it might be difficult to assess model performance using a single likelihood measure, because the form of the distribution of uncertain predictions varies markedly over the range of streamflow and the appropriate error structure might vary with both of type of data and the model parameter set (Freer *et al.*, 2003). It may often be the case that the available data are not adequate to allow identification of complex models and/or that a single performance measure (objective function) is not adequate to properly take into account the simulation of all the characteristics of a system used. Thus, the multi-criteria or multi-objective methods using multiple objective functions or other data in addition to rainfall-runoff data may allow more robust analyses of models, and aid hypothesis testing of competing model structures (Gupta *et al.*, 1999; Beven, 2001; Madsen *et al.*, 2002; Freer *et al.*, 2003; Uhlenbrook and Siebert, 2004).

In this study, the multi-criteria performance measures based on the concept of equifinality were used for calibration of the rainfall-runoff model TOPMODEL. The focus was particularly on the identification and demonstration of the equifinality of behavioral parameter sets for different performance measures in the mechanical modeling of complex environmental systems. Results specific to TOPMODEL can contribute, with additional experience, to the vigorous debate on applications of environmental models.

Materials and Methods

1. Study site and data used

The study site herein considered is a forested experimental catchment, located in the Gwangnung experimental forests within the Korea National Arboretum (see Figure 1). This catchment is referred to the deciduous

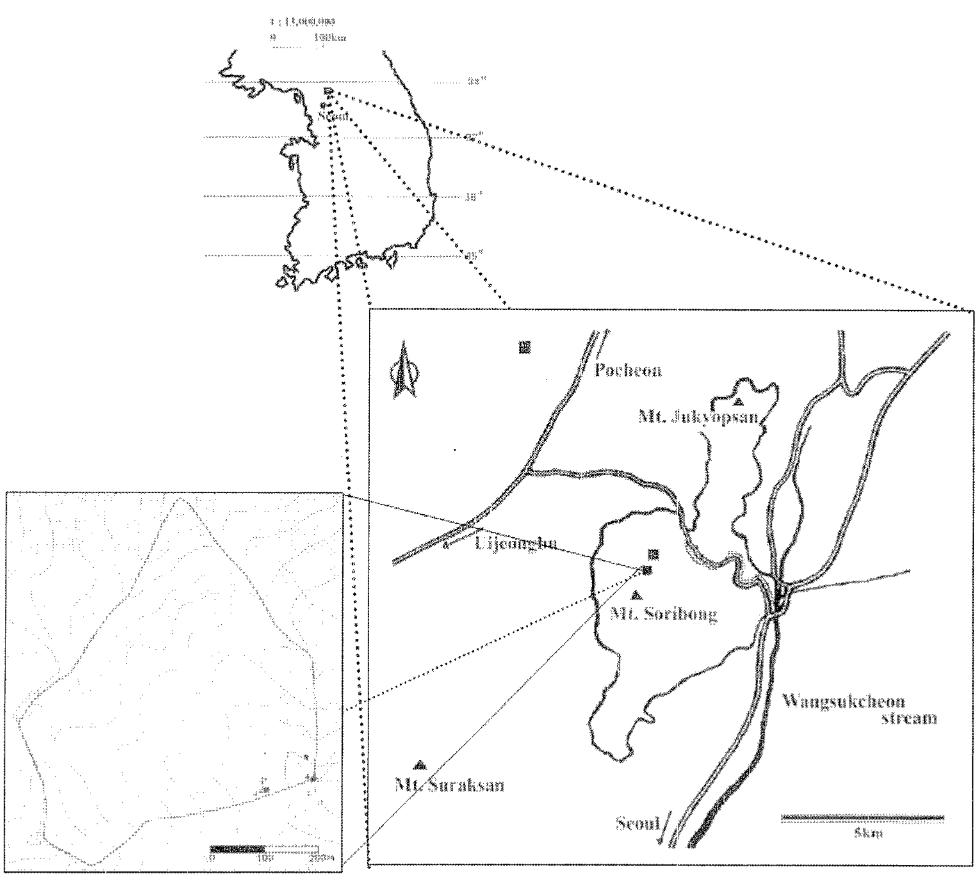

Figure 1. Location and Topography of the deciduous forest catchment.

forest catchment according to the main forest type to be established. The area of the catchment is 22.0 ha, covered with dominantly mature natural deciduous stands consisted of *Quercus serrata* and *Carpinus laxiflora* predominance (mean age, about 80 years). The study site is underlain by gneiss, and characterized by a high relief intensity, with elevation ranging from 260 to 470 m above sea level. Mean annual precipitation is approximately 1,433 mm, generating mean annual discharges of about 841 mm.

Hourly stream flow and rainfall data of the deciduous forest catchment observed from April to October, 2005 were used for model calibration. The recorded rainfall totaled 1,546.2 mm during this period, and a runoff of 1,051.4 mm was observed during the same period with a runoff rate of 68.0%. The recorded maximum rainfall and runoff values for 1 hour within the period were 48.9 mm/hr and 4.8 mm/hr, respectively. The potential evapotranspiration was calculated by the FAO Penman-Monteith equation using weather information of the Dongduchun Weather Station, and modified by the comparison with the actual evapotranspiration data obtained within the study site. Hourly values of the potential evapotranspiration were derived by applying the observed hourly rates of the actual evapotranspiration.

2. The hydrological model

TOPMODEL is a rainfall-runoff model in which distributed predictions of catchment response are made based on a simple theory of hydrological similarity of points in a catchment (Beven and Kirby, 1979; Beven *et al.*, 1995; Beven, 2001; Beven and Freer, 2001a). In this version of the model, the hydrological similarity comes from the use of the topographic index as $\ln(\alpha/\tan\beta)$, where β is the area draining through a point from upslope and tan â is the local slope angle (Beven, 1997).

TOPMODEL was originally developed to predict the rainfall-runoff relationship, and to describe the spatial pattern of storm flow generation in upland humid temperate climate catchments. It has the advantages that the computational burden of the model is greatly reduced relative to a fully distributed model and the number of parameters required to run the model can be kept small, reducing the possibilities of overparameterization. In most of these cases it has been found that, after calibration of the parameters, TOPMODEL provides good simulations of stream discharges (Beven, 2001).

In TOPMODEL, total runoff is generally calculated as the sum of two major flow components: saturation excess overland flow from variable contributing areas and subsurface flow from the saturated zone of the soil. And, infiltration excess overland flow component can also be included if suggested by the catchment soil and rainfall characteristics (Beven *et al.*, 1994; Franchini *et al.*, 1996).

The deciduous forest catchment is located in a humid temperate climate zone, and has a long distinct wet season and steep slopes. This wet environment and sloping terrain can be considered as a suitable test environment for the TOPMODEL.

The version of TOPMODEL used in this study is based on the original assumptions of an exponential decline of transmissivity with depth or storage deficit. Digital Terrain Model (DTM) of the study area with grid size of 10 m was used to calculate topographic indices for the TOPMODEL. A detailed description of the TOPMODEL can be found in Beven *et al.* (1995) or Beven (2001).

3. Multi-criteria performance measures

The equifinality thesis focuses attention on a search for multiple acceptable model parameter sets that would give reliable simulations and which should therefore be used in making predictions with a given model. The procedure is based upon making a large number of simulation runs of a model with different parameter sets, chosen randomly from the specified ranges for each parameter by Monte Carlo simulation (Beven and Binley, 1992). In most case, parameter sampling is carried out using non-informative uniform sampling without prior knowledge of individual parameter distributions other than a feasible range of values (Beven and Freer, 2001b).

Table 1 shows the parameters to be considered in the Monte Carlo simulation in this study, together with their respective ranges. These ranges were thought feasible for the study site on the basis of previous studies (Beven, 1997; Beven and Freer, 2001b). Each parameter value is drawn uniformly and independently from within the ranges, and in total 100,000 sets were chosen to

Table 1. Parameter ranges used in the Monte Carlo Simulations.

Parameter	Description	Parameter Range
SZM [m]	The parameter of the exponential transmissivity function	0.002 ~ 0.1
$\ln T_0$ [m^{2}h^{-1}]	Effective lateral saturated transmissivity	2.33 ~ 3.4
T_d [mh^{-1}]	Unsaturated zone time delay per unit deficit	0.5 ~ 20
SR_{INIT} [m]	The initial storage deficit in the root zone	0 ~ 0.2
SR_{MAX} [m]	The soil profile storage available for transpiration	0.01 ~ 0.2

drive the TOPMODEL.

The performance of each independent random parameter set is evaluated by a quantitative measure of performance or likelihood measure, and each set is classified as a behavioral or non-behavioral parameter set by a chosen behavioral acceptance threshold or rejection criteria. The likelihood value is associated with a parameter set and it will reflect all sources of error (including model structure, inputs and observations) and any effects of the covariation of parameter values on model performance implicitly (Beven and Binley, 1992; Beven and Freer 2001a, b; Blazkova and Beven, 2004).

It is obvious that the choice of performance measures and the behavioral acceptance threshold are important in the multi-criteria likelihood evaluation. In this methodology, choosing which performance measures to use is necessarily subjective unless the user is prepared to make strong assumptions about the nature of the errors (see discussion in Beven, 2005). The uncertainty estimates will depend on the definition of the performance measures, and the boundary between behavioral and non-behavioral parameter sets may not be clear. However, some subjectivity in the choice of the likelihood measure would seem to be unavoidable for cases where the model residuals do not confirm to a simple structure, and there is evidence that the choice of the behavioral acceptance threshold may not be as critical as previously thought in the application (Beven and Binley, 1992; Ratto *et al.*, 2001; Freer *et al.*, 2003). The performance measures and behavioral acceptance threshold used in this study are defined in Table 2. The different definitions of performance measures shown in Table 2 were chosen to reflect their sensitivity to different hydrologic characteristics of the simulated period. M_{EFF} is biased towards reflecting large errors associated with peak discharges, M_{LOG} is biased towards recession flows, M_{BIAS} is the bias for the simulation period and M_{WI}, M_{CM} and M_{SAE} are compromises between M_{EFF} and M_{LOG}

Results and Discussion

1. Variability in the behavioral parameter distributions

The behavioral parameter sets were identified separately for each performance measure, by application of their corresponding behavioral acceptance thresholds as defined in Table 2. Figure 2 shows the dotty plots of likelihood values for selected TOPMODEL parameters from Monte Carlo simulations of the deciduous forest catchment conditioned on the 2005 discharge period. In Figure 2, each dot represents one run of the model from a Monte Carlo experiment using 10,000 simulations with different randomly chosen parameter values from the ranges defined in Table 1, and horizontal lines mean thresholds identifying behavioral parameter sets for each performance measure; all of parameter sets give different predictions, and dots over the line (in cases of M_{EFF} M_{LOG} M_{WI} and M_{CM}), dots between both lines (in case of M_{BIAS}) and dots below the line (in case of M_{SAE}) are classified as behavioral simulations.

In Figure 2, some plots show peaky distributions of likelihood values, indicating there are narrow, specific ranges of parameter values having better performance, but most plots show that for each parameter there are a lot of good simulations across a wide range of parameters for each performance measure. As shown in many of plots in Figure 2, behavioral parameter sets identified

Table 2. Definition of different performance measures and behavioral acceptance thresholds.

Performance measure	Description	Formula	Threshold
M_{EFF}	Nash-Sutcliffe efficiency	$1-\sigma_{\varepsilon}^2/\sigma_0^2$	0.8
M_{LOG}	Log transformed Nash-Sutcliffe efficiency	$1-\sigma_{\log\varepsilon}^2/\sigma_{\log 0}^2$	0.8
M_{WI}	Wilmot Index of Agreement	$1-\frac{\Sigma_{t=1}^{N}(Q_{obs(t)}-Q_{sim(t)})^2}{\Sigma_{t=1}^{N}(\lvert Q_{sim(t)}-\overline{Q_{obs(t)}}\rvert)+\lvert Q_{obs(t)}-\overline{Q_{obs(t)}}\rvert}$	0.8
M_{CM}	Chiew and McMahon Index	$1-\frac{\Sigma_{t=1}^{N}(\sqrt{Q_{obs(t)}}-\sqrt{Q_{sim(t)}})^2}{\Sigma_{t=1}^{N}(\sqrt{Q_{obs(t)}}-\sqrt{\overline{Q_{obs(t)}}})^2}$	0.8
M_{BIAS}	Cumulative error	$\Sigma_{t=1}^{N}(O_{obs(t)}-Q_{sim(t)})$	± 5% Discharge
M_{SAE}	Cumulative Absolute error	$\Sigma_{t=1}^{N}\lvert Q_{obs(t)}-Q_{sim(t)}\rvert$	30% Discharge

σ_{ε}^2 is the error variance (log transformed $\sigma_{\log\varepsilon}^2$), σ_0^2 is the variance of the observations (log transformed $\sigma_{\log 0}^2$), and $Q_{sim(t)}$ are $Q_{obs(t)}$ the simulated streamflow and the observed streamflow at timestep *t*, respectively. *N* is the number of time steps.

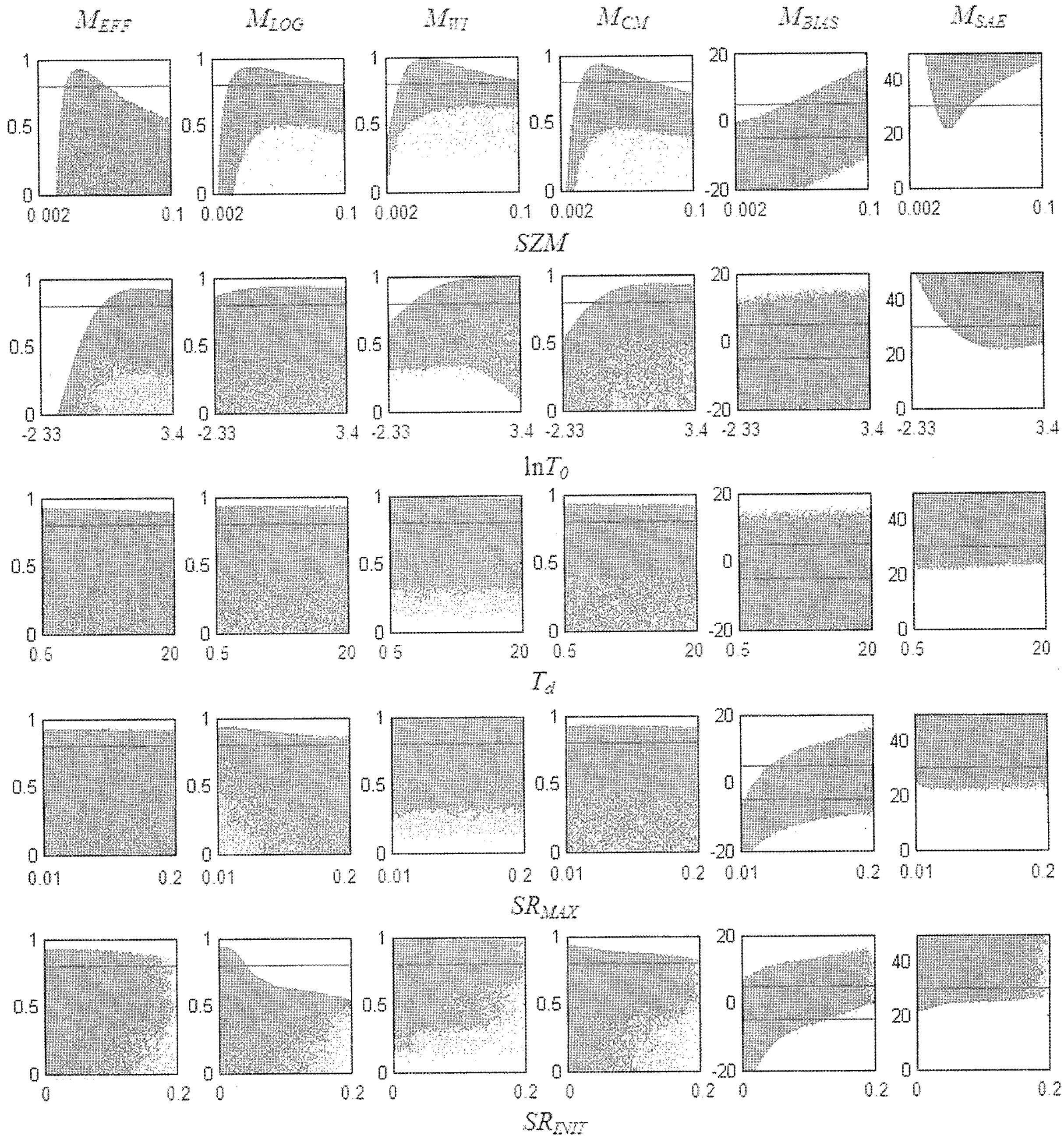

Figure 2. Scatter plots of likelihood values for TOPMODEL parameters from Monte Carlo simulations of the deciduous forest catchment conditioned on the 2005 discharge period using six performance measures defined in Table 2. Each dot represents one simulation with a likelihood weight calculated by a given performance measure, and horizontal lines mean thresholds identifying behavioral parameter sets for each performance measure; dots over the line (in cases of M_{EFF}, M_{LOG}, M_{WI} and M_{CM}), dots between both lines(in case of M_{BIAS}) and dots below the line (in case of M_{SAE}) are classified as behavioral simulations.

by a given threshold for each performance measure are distributed across the whole range of each parameter except for the parameter of the exponential transmissivity function, SZM and effective lateral saturated transmissivity, $\ln T_0$ with relatively peaky likelihood distributions.

The numbers of behavioral parameter sets for each performance measure are shown in Table 3, varying depending on the performance measures, and relatively large numbers of parameter sets are identified as behavioral for performance measures, M_{BIAS} and M_{WI}. Relatively high acceptable rates of behavioral parameter sets may indicate that the hydrological model and initial ranges of parameters might be appropriately chosen.

Figure 3 shows the variations in the behavioral parameter distributions for each performance measure, and Figure 4 illustrates the likelihood response surfaces between the TOPMODEL parameters, conditioned on the 2005 discharge period of the deciduous forest catch-

Table 3. The number of behavioral parameter sets for each performance measure.

	Performance measure					
	M_{EFF}	M_{LOG}	M_{WI}	M_{CM}	M_{BIAS}	M_{SAE}
The number of behavioral parameter sets	8,724	18,744	52,003	22,595	47,916	9,773

*Initial population of parameter sets to drive the TOPMODEL is 100,000.

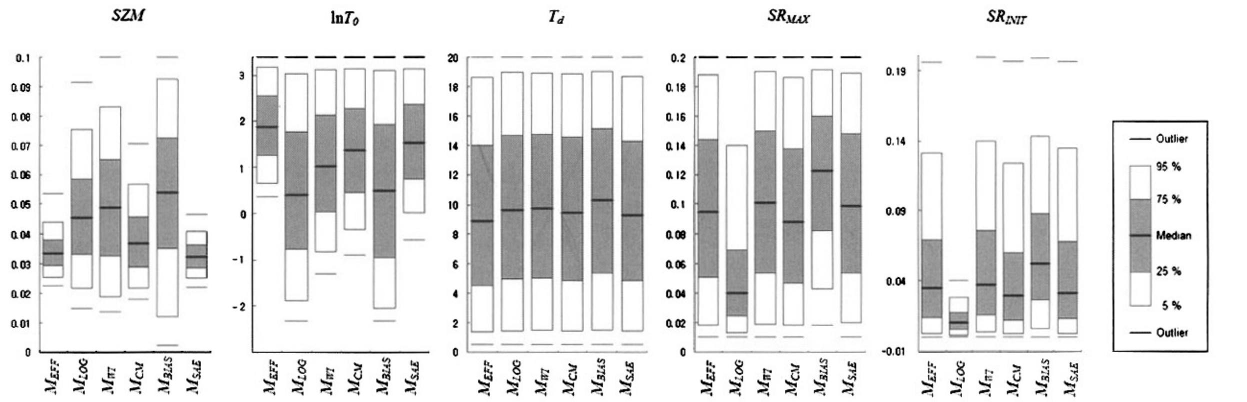

Figure 3. Variations and medians in the behavioral parameter distributions for each performance measure for each TOPMODEL parameter, conditioned on the 2005 discharge period of the deciduous forest catchment.

ment. Behavioral parameter sets with higher model performance are in the bright area in Figure 4.

As shown in Figure 3, the range of model behavior for each parameter varied considerably between the different performance measures, i.e. the median values and ranges of the behavioral parameter distributions are highly variable for each performance measure for some parameters. Treated individually, the *SZM* constraints the model responses most, and the $\ln T_0$, SR_{MAX} and SR_{INIT} are also the parameter that may constraint the model responses relatively.

In addition, Figure 4 illustrates that different parts of the parameter space are optimized for different performance measures, i.e. sometimes a quite different optimal parameter set or different behavioral parameter sets distributions may be obtained according to the kind of performance measures. Therefore, using a single performance measure for the calibration of a hydrological model may lead to increase model uncertainty.

2. Sensitivity analysis of behavioral parameter distributions

Sensitivity analysis is a general methodology used to evaluate the sensitivity of model output to changes in model input, i.e. the rate of change of the response function relative to the input parameters. It is also closely linked to Uncertainty Analysis, where concern shifts to the evaluation of the uncertainty on the model response as a result of uncertainties on the model input parameters (parametric uncertainty) and on the model from itself (structural uncertainty).

Figure 5 presents cumulative marginal likelihood distribution plots for each parameter from behavioral parameter sets with likelihood weights for each performance measure, conditioned on the 2005 discharge period of the deciduous forest catchment. Those parameters showing a strong deviation from the original uniform distribution, which is stretched out along the imaginary diagonal line from the left-low corner to right-up corner of each plot, may be considered the most sensitive in that they have been most strongly conditioned by the model evaluation process. Those that are still uniformly distributed across the same parameter ranges show less sensitivity (Hornburger and Spear, 1981). Such plots must be interpreted with care, however. The visual impression will depend on the original range of parameters considered, while the value of a parameter that continues to show a uniform marginal distribution may still have significance in the context of a set of values of the other parameters. Experience suggests that fixing the value of such parameters may constrain the performance of the model too much (Beven and Freer, 2001b).

As Figure 5 shows, the parameter of the exponential transmissivity function, *SZM* shows strongest sensitivity for most performance measures except for M_{BIAS}. Also, effective lateral saturated transmissivity, $\ln T_0$, the initial storage deficit in the root zone, SR_{INIT} and the soil profile storage, SR_{MAX} are identified as more sensitive. According to the sensitivity of parameters by each performance measure, the sensitivities based on M_{EFF}, M_{LOG} and M_{SAE} are stronger than those of other measures. It may be considered that this may result from the oversensitivity to outliers (the largest errors), which is more strongly

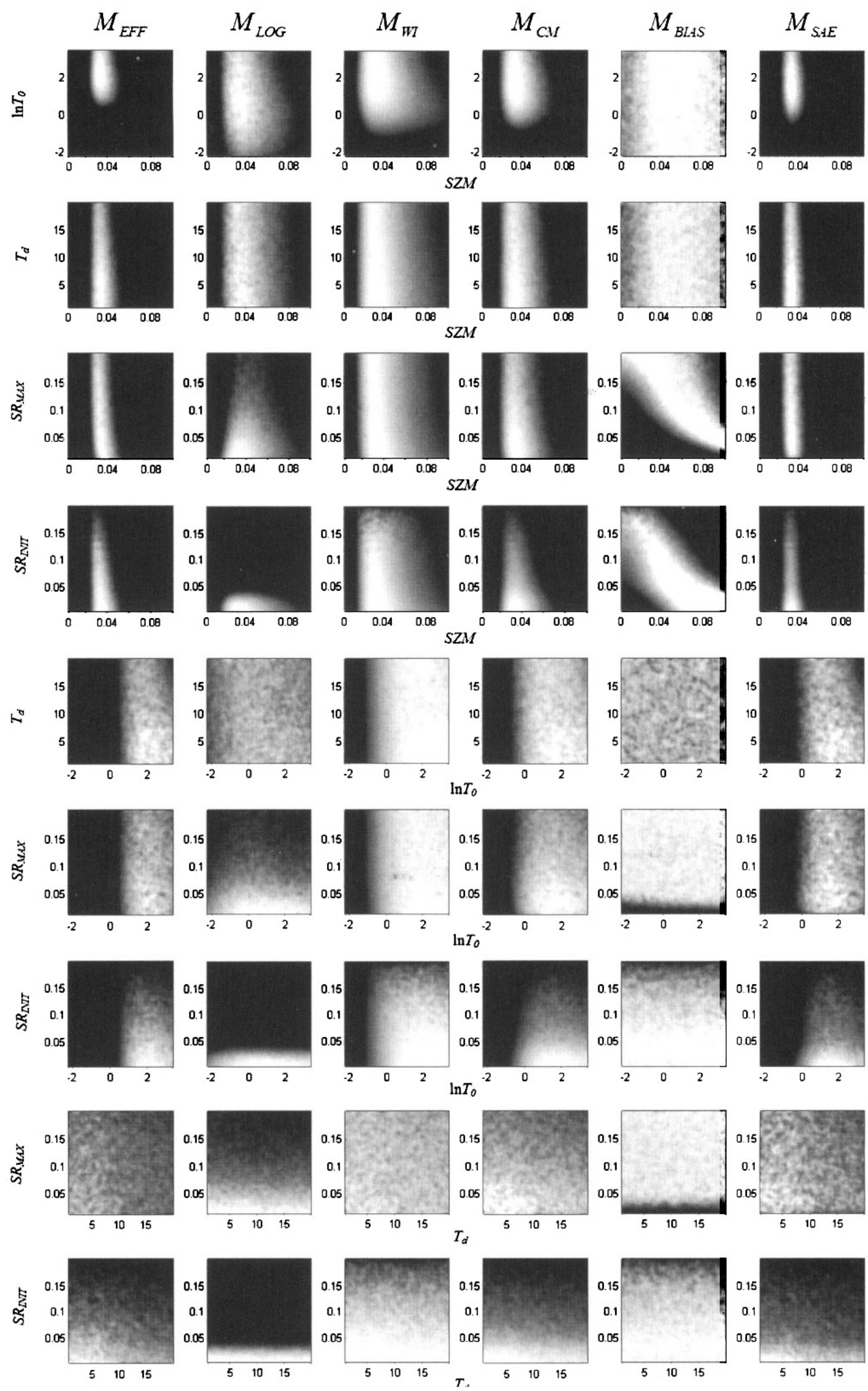

Figure 4. Likelihood response surfaces between the major parameters of TOPMODEL, conditioned on the 2005 discharge period of the deciduous forest catchment. (Behavioral parameter sets with higher model performance are in the white zone.)

revealed in M_{EFF}, M_{LOG} and M_{SAE}. In general, correlation-based measures such as the performance measures used in this study are more sensitive to outliers in the given variables, and this high sensitivity to outliers leads to the result in that the performance measures are biased towards the extreme events (Legates and Davis, 1997).

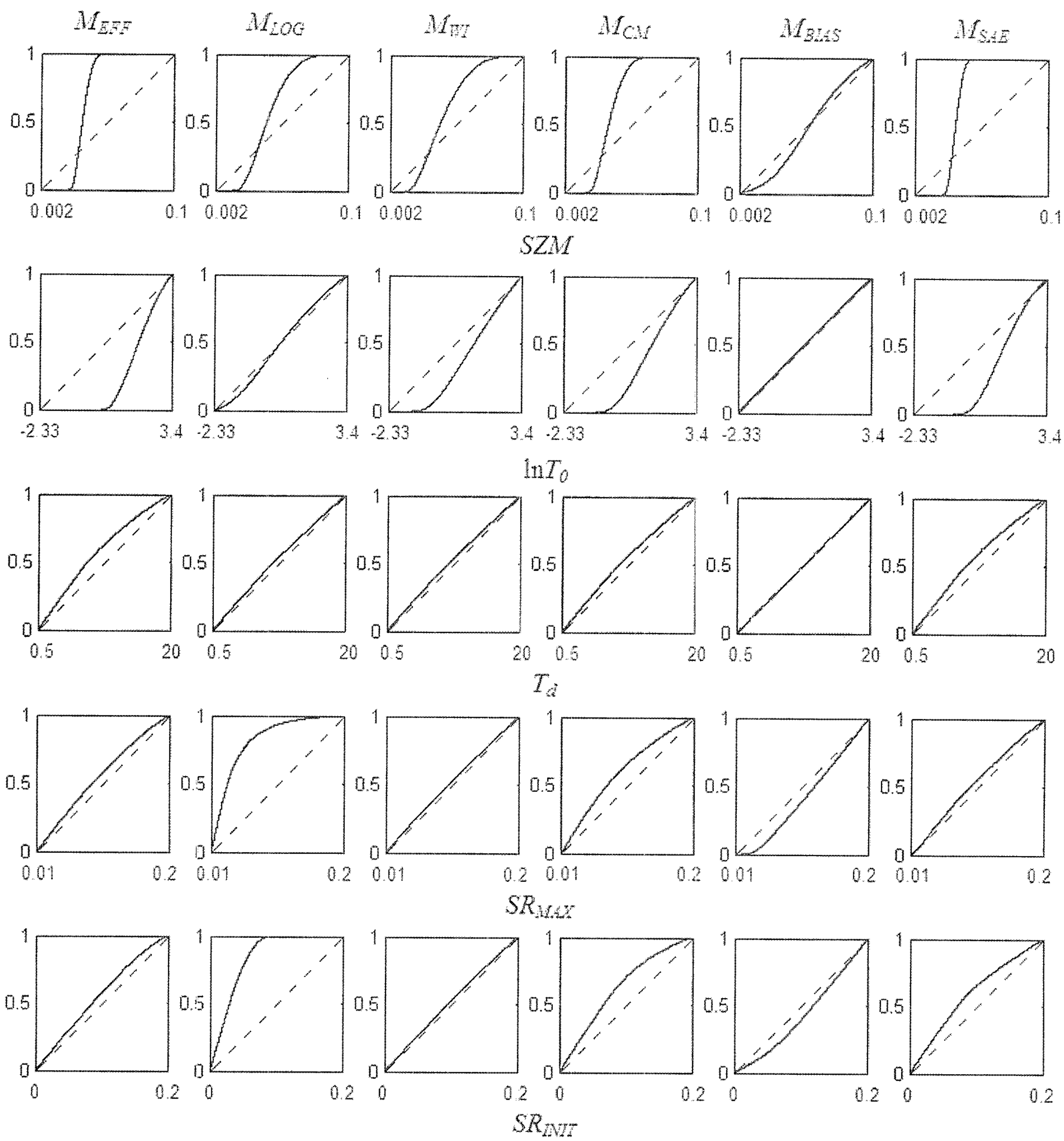

Figure 5. Cumulative marginal likelihood distributions for each parameter from behavioral parameter sets with likelihood weights for each performance measure using the 2005 discharge period of the deciduous forest catchment. Dotty diagonal lines in plots mean original uniform distributions of likelihood weights.

M_{EFF} is necessarily biased towards reflecting the larger errors associated with high flows based on its definitions, particularly in the case of model timing errors. It thus may appear that the behavioral parameter sets identified by M_{EFF} for each calibration year are strongly reflecting hydrologic characteristics of the high flow periods in each calibration year.

3. Relationships and compatibilities of behavioral parameter sets between different performance measures

Analysis of the relationships between multi-criteria performance measures can give a greater understanding of the model dynamics and potentially help in the future development of the model structure. The use of multi-criteria performance measures increases our ability to perform a test on various model structure hypotheses. Relationships among different performance measures for the behavioral simulations are shown in Figure 6 as dotty plots, where each point signifies a behavioral model simulation. Figure 6 shows that correlations between performance measures of the behavioral simulations are quite variable, often having a lot of scatter for one performance measure, when compared with another perfor-

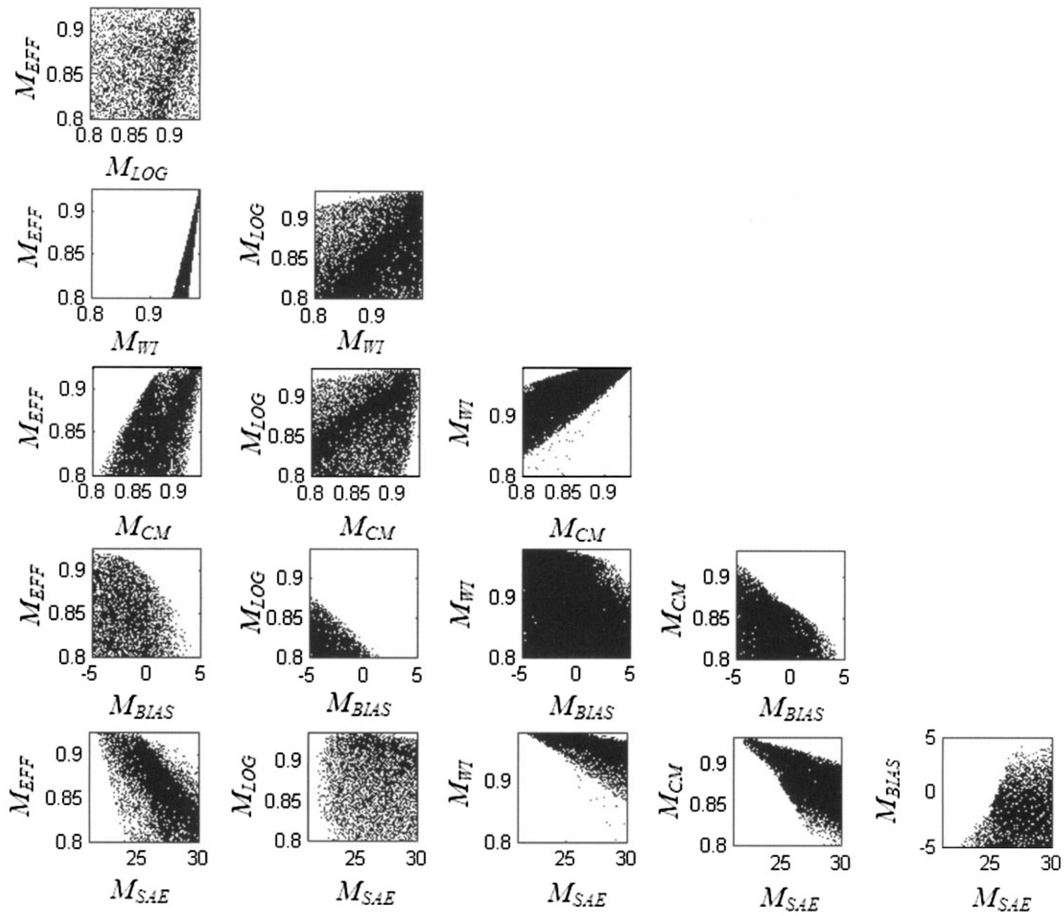

Figure 6. The relationships between the likelihood weights of each compatible parameter set associated with each performance measure.

mance measure. A comparison of the results for M_{BIAS} with the other performance measures suggests that the model as currently formulated generally over-predicts total streamflow (negative M_{BIAS}) while obtaining good simulations of the larger storm events.

In addition, to examine the relationships between the behavioral parameter sets for the different performance measures, compatibilities of parameter sets between two different performance measures were investigated for all combinations of performance measures. A particular parameter set can be identified as compatible if it features in two or more different distributions of behavioral parameter sets. As a result, the behavioral simulations identified by one performance measure are not necessarily behavioral for another performance measure. These characteristics between behavioral parameter sets for different performance measures can be revealed in Table 4, which summarize the compatibilities between different performance measures. Most of the behavioral parameter sets for M_{EFF} and M_{SAE} exist in the behavioral parameter sets for M_{WI} and M_{CM}, and most of the behavioral parameter sets for M_{CM} can also function as behavioral parameter sets for M_{WI}. Relatively high compatibilities between

Table 4. The number of compatible behavioral parameter sets between each performance measure.

		Performance measures				
		M_{LOG}	M_{WI}	M_{CM}	M_{BIAS}	M_{SAE}
Performance measures	M_{EFF}	3,734	8,724	8,724	4,189	7,480
	M_{LOG}		14,912	10,464	3,091	4,364
	M_{WI}			22,588	29,604	9,771
	M_{CM}				10,801	9,763
	M_{BIAS}					4,465

M_{EFF}, M_{WI}, M_{CM} and M_{SAE} may imply that their behavioral parameter sets come from similar areas in the parameter space, and reflect their similarity in the definitions of the performance measures. In contrast, behavioral simulation parameter sets of M_{LOG} and M_{BIAS} differ markedly from those for the other performance measures. Also, behavioral parameter sets between the M_{LOG} and M_{BIAS} differ quite markedly from each other with the lowest compatibility.

Figure 6 and Table 4 support the multi-criteria calibration argument. It shows that results which are seen as acceptable by one criterion fail with other criteria computed. Moreover, when the performance measures were combined over all performance measures, only 176 of the parameter sets were found that are behavioral over all measures. 176 parameter sets are only 0.18% of the initial population of the parameter sets used in this study. Therefore, an evaluation based on only one performance measure is unlikely to be reliable (even when one has put some effort into investigating one specific method).

Conclusions

In this study, the multi-criteria performance measures based on the concept of equifinality of behavioral model simulations were used for calibration of the rainfall-runoff model, TOPMODEL. Totally 100,000 parameter sets uniformly sampled by Monte Carlo Simulations from the ranges for each TOPMODEL parameters defined in Table 1 were applied in TOPMODEL, and hourly stream flow and rainfall data observed from April to October, 2005 in the deciduous forest catchment located in the Gwangnung experimental forests were used for model calibration.

The performance of each parameter set was evaluated and identified with 6 different performance measures against behavioral acceptance thresholds defined for each performance measure, and the results were analyzed focused on the variability and relationship between the behavioral parameter distributions according to the definitions of performance measures.

The results demonstrate that there are many acceptable parameter sets scattered throughout the parameter space, all of which are consistent in some sense with the calibration data, and the range of model behavior for each parameter varied considerably between the different performance measures. Sensitivity was very high in some parameters, and varied depending on the kind of performance measure. Compatibilities of behavioral parameter sets between different performance measures also varied, and a very small minority of parameter sets could produce reliable predictions regardless of the kind of performance measures (at least, for the performance measures used in this paper). Especially, the results indicate that using a single performance measure for the calibration of a hydrological model may lead to an increase in model uncertainty. Therefore, careful consideration should be given to the choice of performance measure appropriate to the characteristics of used model and data and the purpose of study.

Differences in the behavioral parameter distributions according to the performance measures may be directly caused by the definitions of performance measures. However, it also should be considered that the effects of model nonlinearity, covariation of parameter values and errors in model structure, input data or observed variables may be taken into account in the nonlinearity of the response of acceptable model.

The performance of the parameter set can be used to produce the likelihood-weighted marginal parameter distributions for individual parameters, and the likelihood weighted model simulations can be used to estimate prediction quantiles in a way that allows that different models may contribute to the ensemble prediction interval at different time steps and that the distributional form of the predictions may change from time to time step.

Literature Cited

1. Beven, K. J. 1997. TOPMODEL: a critique. Hydrological Processes 11(3): 1069-1085.
2. Beven, K.J. 2001. Rainfall-Runoff Modeling. The Primer, Wiley, Chichester. pp. 360.
3. Beven, K.J. 2002. Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences 458(2026): 2465-2484.
4. Beven, K.J. 2005. A manifesto for the equifinality thesis. Journal of Hydrology, in press.
5. Beven, K.J. and Binley, A. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6: 279-298.
6. Beven, K.J. and Freer, J. 2001a. A dynamic TOPMODEL. Hydrological Processes 15, 1993-2011.
7. Beven, K.J. and Freer, J. 2001b. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249: 11-29.
8. Beven, K.J. and Kirkby, M.J. 1979. A physically-based variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24(1): 43-69.
9. Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J. 1995. TOPMODEL. In: Singh, V.P., Editor, 1995. Computer Models of Watershed Hydrology, Water Resource Publications, Colorado, 627-668.
10. Beven, K.J., Quinn, P.J., Romanowicz, R., Freer, J., Fisher, J. and Lamb, R. 1994. TOPMODEL and GRIDATB: A Users Guide to the Distribution Versions (94.01). TR110/94, 43 pp., CRES.
11. Blazkova, S. and Beven, K. 2004. Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. Journal of Hydrology 292: 153-172.
12. Franchini, M., Wendling, J., Obled, C. and Todini, E. 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. Journal of Hydrology 175: 293-338.
13. Freer, J. 1998. Uncertainty and calibration of conceptual rainfall runoff models. Ph.D. Thesis, Lancaster

University.

14. Freer, J.E., Beven, K.J. and Peters, N.E. 2003. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A.N. and Turcotte, R., Editors, 2003. *Calibration of watershed models*. AGU, Water Science and Application Series, Washington pp. 346.
15. Gupta, H.V., Bastidas, L.A., Sorooshian, S., Shuttleworth, W.J. and Yang, Z.L. 1999. Parameter estimation of a land surface scheme using multicriteria methods. Journal of Geophysical Research 104(D16): 19,491-19,503.
16. Hornberger, G.M. and Spear, R. 1981. An approach to the preliminary analysis of environmental systems. Journal of Environmental Management 12: 7-18.
17. Legates, D.R., and Davies, R.E. 1997. The continuing search for an anthropogenic climate change signal: Limitations of correlation-based approaches. Geophys. Res. Lett. 24: 2319-2322.
18. Loague, K. and VanderKwaak, J.E. 2004. Physics-based hydrologic response simulation: platinum bridge, 1958 Edsel, or useful tool. Hydrological Processes 18: 2949-2956.
19. Madsen, H., Wilson, G. and Ammentorp, H.C. 2002. Comparison of different automatic strategies for calibration of rainfall-runoff models. Journal of Hydrology 261: 48-59.
20. Ndiritu J.G. and Daniell, T.M. 1999. Assessing model calibration adequacy via global optimization. Water SA 25(3): 317-326.
21. Ratto, M., Tarantola, S. and Saltelli, A. 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. Computer Physics Communications 136: 212-224.
22. Refsgaard, J.C. 1997. Parameterisation, calibration and validation of distributed hydrological models, Journal of Hydrology 198: 69-97.
23. Refsgaard, J.C. and Knudsen, J. 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32(7): 2189-2202.
24. Uhlenbrook, S. and Siebert, A. 2004. On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model. Environmental Modelling & Software, in press.
25. Vieux, B.E., Cui, Z. and Gaur, A. 2004. Evaluation of a physics-based distributed hydrologic model for flood forecasting. Journal of Hydrology 298: 155-177.

(Received June 20, 2005; Accepted September 11, 2006)