WATER RESOURCES RESEARCH, VOL. 45, W12422, doi:10.1029/2009WR007706, 2009

On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information

H. C. Winsemius,[1] B. Schaefli,[1] A. Montanari,[2] and H. H. G. Savenije[1]

Received 5 January 2009; revised 25 June 2009; accepted 5 August 2009; published 31 December 2009.

[1] This paper presents a calibration framework based on the generalized likelihood uncertainty estimation (GLUE) that can be used to condition hydrological model parameter distributions in scarcely gauged river basins, where data is uncertain, intermittent or nonconcomitant. At the heart of this framework is the conditioning of the model parameters such as to reproduce key signatures of the observed data within some limits of acceptability. These signatures are either based on hard or on soft information. Hard information signatures are defined as signatures for which the limits of acceptability may be objectively derived from the distribution of long series of observed values, and which effectively constrain the model parameters. Soft signatures are less effective in parameter conditioning or their limits of acceptability cannot be objectively derived. During random parameter sampling, parameter sets are accepted as equally likely if they meet all the hard limits of acceptability. This results in an intermediate parameter distribution, which can be used to reduce the sampling limits. Then, the soft information may be introduced in a second constraining step to reach a final parameter distribution. The modeler can use the final results as a guideline for a further search for information, possibly from new observations yet to collect. In an application of the framework to the Luangwa catchment in Zambia, three information signatures are retrieved from a data set of old discharge time series and used to condition the parameters of a daily conceptual rainfall-runoff model. We performed two independent calibration experiments with two significantly different satellite rainfall estimates as model input. The results show consistent parameter distributions and considerable reduction of the prior parameter space and corresponding output realizations. These results illustrate the potential of the proposed calibration framework for predictions in scarcely gauged catchments.

Citation: Winsemius, H. C., B. Schaefli, A. Montanari, and H. H. G. Savenije (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, *45*, W12422, doi:10.1029/2009WR007706.

1. Introduction

[2] The formulation of hydrological models to assess water resources, flood and drought risk, and effects of man-made and climatic change in river basins is jeopardized when hydrological data are absent or scarce. In fact, the limited availability of ground data induces ill-quantifiable uncertainty in model outputs. This problem often restrains water resource managers from further investigation while there is a great need for hydrological models in these ungauged basins. Initiatives to solve the afore-mentioned problems have been reported under the umbrella of the research initiative Predictions in Ungauged Basins (PUB) [*Sivapalan*, 2003], launched in 2003 by the International Association of Hydrological Sciences. Since then, work has been done on new measurement techniques [e.g., *Uhlenbrook and Wenninger*, 2006; *Selker et al.*, 2006; *Westhoff et al.*, 2007], new modeling frameworks [e.g., *Reggiani et al.*, 1998; *Zhang et al.*, 2006; *Lee et al.*, 2007; *Schymanski et al.*, 2008], use of new and soft data sources [e.g., *Seibert and McDonnell*, 2002; *Franks*, 2007; *Immerzeel and Droogers*, 2008; *Winsemius et al.*, 2008], new optimization techniques [e.g., *Vrugt et al.*, 2003; *Kuczera et al.*, 2006] optimal use of available data [e.g., *Atkinson et al.*, 2002; *Wagener*, 2003; *Montanari and Toth*, 2007; *Schaefli and Zehe*, 2009] and improvement of model structures by means of multi-informative optimization [*Vaché and McDonnell*, 2006; *Son and Sivapalan*, 2007; *Fenicia et al.*, 2008] that may benefit prediction of streamflow in ungauged basins. A general consensus has been reached about the opportunity to profit from any hydrological information that might be available in these cases. *Seibert and Beven* [2009] for instance, show what can be done with very limited available discharge measurements in model calibration. Other authors showed that effectively combining information from different independent (sometimes highly uncertain) data sources can result both in constraining of hydrological model parameters and in enhancement of the model structure [*Winsemius*

[1]Department of Water Management, Delft University of Technology, Delft, Netherlands.
[2]Faculty of Engineering, University of Bologna, Bologna, Italy.

0043-1397/09/2009WR007706$09.00

et al., 2006; *Klees et al.*, 2007; *Fenicia et al.*, 2008]. To this end, multiobjective calibration techniques have been proposed [e.g., *Vrugt et al.*, 2003]. However, these can only be applied when dealing with well-defined objective functions. In the presence of limited and uncertain information, formal calibration methods may be impossible to apply. In fact, a typical problem of many ungauged or scarcely gauged river basins in the world is that data are scarce, not accurate, intermittent, nonconcomitant and collected at different timescales, leading to the fact that it is not clear how one can integrate their nonconventional information content.

[3] In this paper, a framework is proposed to integrate both hard and soft information delivered by available hydrological observations, which makes use of the Generalized Likelihood Uncertainty Estimation (GLUE) [*Beven and Binley*, 1992] within a limit of acceptability approach [*Beven*, 2006]. A well known debate is ongoing within the hydrological community about the opportunity of using ''formal statistical'' versus ''informal'' methods for model calibration and uncertainty analysis [*Mantovan and Todini*, 2006; *Beven et al.*, 2007, 2008; *Beven*, 2008]. The latter category of methods includes GLUE, which has been often criticized for being subjective. We generally agree about the subjectivity of GLUE [*Montanari*, 2005] but nevertheless we are convinced that GLUE is potentially useful for reducing the uncertainty of hydrological model parameters in conditions of data scarcity, when a formal statistical assessment would not be reliable. In order to strengthen the applicability of GLUE in ungauged situations, we propose a new framework which fixes limits of acceptability on ''signatures'' which are potentially present in any hydrological data, even when the data is subject to uncertainty, nonconcomitance or other problems related to the ungauged nature of the studied river basin.

[4] The use of signatures from time series in model conditioning shows important developments in recent literature. *Vogel and Sankarasubramanian* [2003] showed that statistical signatures of time series reveal significant and hydrologically meaningful information, which is not conveyed by regularly used least squares performance criteria. *Gupta et al.* [2008] and *Yilmaz et al.* [2008] proposed to use a set of hydrologically meaningful signatures to evaluate model performance besides less meaningful residual based methods. *Yilmaz et al.* [2008] provide an extensive overview of potentially useful signatures in model calibration. Such signatures have consequently been used in calibration of hydrological models [e.g., *Pokhrel et al.*, 2009; *Herbst et al.*, 2009; *Bulygina et al.*, 2009] and model regionalization [e.g., *Yadav et al.*, 2007; *Oudin et al.*, 2008]. Direct calibration or regionalization may however not be applicable in many basins because the conventional data sets needed, are not available. Therefore this paper focuses on the use of any information present in nonconventional data sets available for the basin itself.

[5] Section 2 describes the essentials of the framework and how limits of acceptability on the information signatures are derived. In Section 3, the framework is demonstrated in an example case study where the parameters of a rainfall-runoff model of the scarcely gauged Luangwa basin in Zambia are conditioned. We employed a number of information signatures from old discharge records to constrain the model parameters, while running the model with new satellite rainfall data. A validation of the method is shown in Section 4. Finally, we discuss how the outcomes of this study can be employed to find what additional information could contribute effectively to further reduce parameter uncertainty in Section 5 and summarize conclusions in Section 6.

2. A General Framework for Integrating Hard and Soft Hydrological Information

2.1. Problem Definition

[6] We focus on the problem of calibrating a rainfall-runoff model for an ungauged basin, where some information is available to identify one or several objectives for the model itself. For instance, the objectives could be to reach a satisfactory fit of assigned behaviors of the hydrograph (recession curve, peak flows, etc.). Let us assume that the objectives can be specified in the form of target values to be optimized during model calibration. The target values here represent the information content that we retrieve from the data and are possibly affected by significant uncertainty induced by data scarcity or poor data quality. In fact, we assume that the user does not have a conventional data set available, namely concomitant and long series of input and output data observed at the required time step. This implies that a traditional optimization procedure (either single-objective or multiobjective), for instance based on least squares, cannot be carried out.

[7] Nonetheless, our principal goal here is to calibrate hydrological models under the aforementioned ungauged circumstances. Generally speaking, in the ungauged case, formal statistical parameter inference is of limited use since there is not enough data to apply such methods in a rigorous way. In fact, a residual time series is either not available or highly dependent on uncertain input. Either way, a modeler is not able to test the statistical hypotheses underlying any formal likelihood measure.

[8] In this context, GLUE offers promising perspectives to perform model calibration and uncertainty analysis based on any available information that may provide a measure of goodness of fit of model response. This is the basic concept of Bayesian model inference, where subjective probabilities are used to assess the plausibility of model output. A crucial issue is to define the probabilities within GLUE. These should be derived based on demonstrable valid statements. Valid statements are first of all required to be consistent with the probability calculus. We explain here below how we define probabilities based on the concept of limits of acceptability for the model output [*Beven*, 2006].

2.2. Description of the Calibration Framework

[9] The calibration framework that we propose is based on GLUE and is graphically presented in Figure 1. Generally, GLUE establishes whether or not for a given model a parameter set is feasible, by evaluating on some objective function that conveys information from the data. The target value to reach during model conditioning is limited by a generally subjectively chosen threshold of the objective function. In the framework, proposed in this paper, we establish multiple objectives along with their related target values from any information derived from any available observations. The model output should satisfactorily reproduce the targets in order to be accepted as a plausible

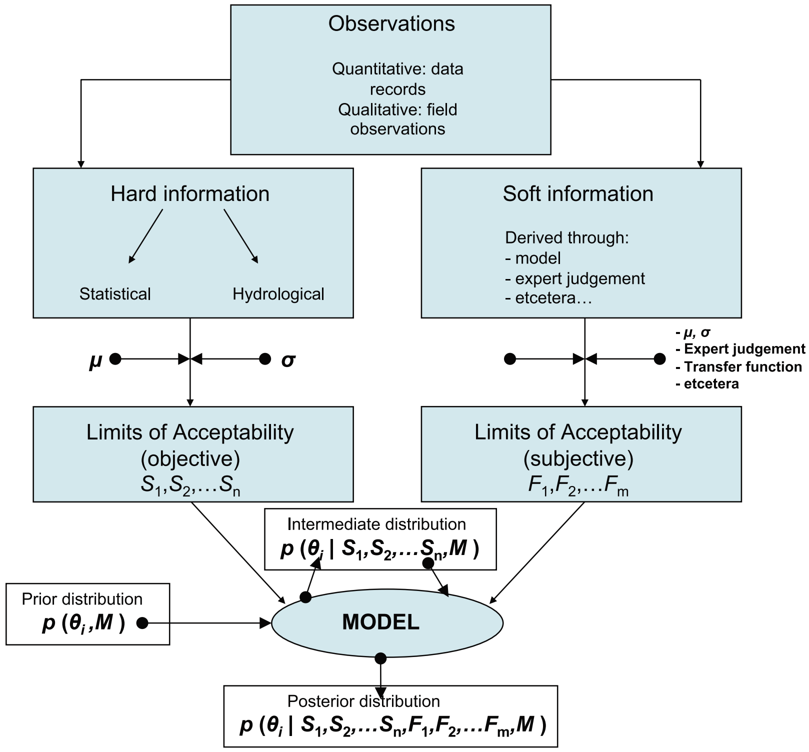

Figure 1. A schematic diagram of the proposed framework.

realization of the process. Such targets are estimated based on both statistical and hydrological signatures of available river discharge data or any other observations. Many examples of such signatures are given by *Yilmaz et al.* [2008] and *Yadav et al.* [2007].

[10] Other than in classical cases, we define target values in the form of "limits of acceptability" of the derived information signatures [*Beven*, 2006]. All models that meet the targets are accepted as equally likely. The limits of acceptability approach was recently suggested by *Beven* [2006] and provides an alternative to fixing threshold values for informal likelihood measures within GLUE to identify behavioral solutions [*Beven and Freer*, 2001]. Acceptability limits should somehow be estimated from an uncertainty assessment in the evidential data, which is the subject of a later section in this paper.

[11] In detail, the framework consists of the following steps.

[12] 1. First the modeler searches for information content in the form of signatures in any data that is readily available (many examples of signatures are provided by *Yadav et al.* [2007]). The signatures do not necessarily have to be fully independent, meaning that the information conveyed by them could be partially overlapping.

[13] 2. The information is then subdivided into hard and soft information. Hard information is such that a) it has a related target value which has a significant impact on the parameter conditioning of the hydrological model; and (b) limits of acceptability can be objectively defined for the target value itself. This last point is a crucial step as the choice for limits of acceptability may impose a strong control on the outcome. In a later section, we describe how to deal with the subjectivity of this choice. We consider soft information to be of a complementary nature, which means that a) the information is less effective to condition the model parameters and/or b) its uncertainty cannot be objectively quantified. The distinction between hard and soft data may imply some subjectivity which is difficult to completely remove when dealing with ungauged basins. Hard information can be further categorized in statistical properties, such as the autocorrelation of the river flows, and hydrological information, such as the time to peak, the water balance and the shape of a recession curve [*Merz and Blöschl*, 2008a, 2008b].

[14] 3. Monte Carlo simulations are performed while using the limits of acceptability of the hard information to establish whether or not a randomly sampled parameter set is feasible given the information content. Any model that performs within the chosen limits of acceptability is accepted as equally likely, because there usually is not enough information to assign a likelihood to the accepted models. Having sampled a satisfactory number of parameter sets, for each parameter the posterior density may be derived according to Bayes' theorem

$$p(\theta_i|S_1, S_2, \cdots, S_n) = \frac{L(S_1, S_2, \cdots, S_n|\theta_i)p(\theta_i)}{p(S_1, S_2, \cdots, S_n)}, \qquad (1)$$

where θ_i is a parameter of the given model, $S_1, S_2, \cdots, S_n$ are the given targets, $L(S_1, S_2, \cdots, S_n|\theta_i)$ is the probability of meeting the targets conditioned on the parameter θ_i, $p(\theta_i)$ is the prior probability for θ_i and $p(S_1, S_2, \cdots, S_n)$ is the probability of meeting the targets. In practice, $L(S_1, S_2, \cdots, S_n|\theta_i)$ is estimated from the empirical density of the accepted parameter sets. $p(S_1, S_2, \cdots, S_n)$ is defined as a normalization constant, so that the posterior distribution $p(\theta_i|S_1, S_2, \cdots, S_n)$

has a cumulative value of unity, such as suggested by *Beven and Binley* [1992].

[15] 4. Based on analysis of the intermediate parameter distributions, a new Monte Carlo simulation is performed, now using the intermediate results as prior. This allows for a computationally efficient subsequent constraining. Now, the soft information with related limits of acceptability is included for further constraining.

[16] 5. Analysis of the results provides the modeler insights into which part of the parameter space, and, consequently, which outputs are well constrained by means of the included information and what constraints are still lacking. The user may then decide what information has the potential to further condition the parameters and, if deemed feasible, to collect this information.

[17] 6. After the collection of new information, the procedure may be repeated to update the parameter distributions with new targets.

[18] One does not necessarily need to perform a two-stage optimization and therefore the distinction between hard and soft information is not strictly necessary. In fact, a simultaneous assessment of the achievement of all the hard and soft targets is in principle possible. The advantages of performing a two stage procedure are the following. First of all, if there is a significant difference among the constraining powers of hard and soft targets, the two-stage procedure is computationally more efficient. Due to the hard information the parameter space is first reduced to a smaller region where the soft information, which may be less efficient in constraining the parameters, can be more profitably used. Second, with a two stage procedure the soft information can be used, according to the user needs, for instance to refine the posterior distribution of selected parameters only. Finally, the intermediate parameter distribution can inform the user which parameters are poorly constrained by the hard targets and what information is needed to further condition these parameters. This is essential because any collection and processing of new data is costly and time consuming.

[19] The framework cannot generalize the used signatures as these are strongly dependent on available data, the dominant processes and related timescales and nonlinearities of the basin studied and even the user's perception, which may be influenced by the objective of the model itself. However, this is also the case in any classical calibration approach, where the modeler decides on the objective function used. Instead, the generality lies in the method to reduce the subjectivity of rejection criteria for models, and the iterations that the modeler can go through after the first results have been obtained. The first results can act as a learning tool to discover what additional information is required for further constraining.

2.3. Definitions of the Limits of Acceptability

[20] A crucial part in this framework is the identification of the limits of acceptability of the target values. It is a delicate step of the analysis because it is usually subjective. For the hard information, we propose a procedure for eliminating part of the subjectivity, which can be applied when a multiyear observation is available. Accordingly, the limits of acceptability are based on the analysis of the interannual variability of each target value. In this way one can get an objective indication about the uncertainty affecting the evaluation of the given target.

[21] In detail, we can compute the identified target value from the data for a number of hydrological years, therefore obtaining a sample of targets. Then, we transform the sample to the Gaussian distribution by using the normal quantile transform (NQT) [see *Kelly and Krzysztofowicz*, 1997; *Krzysztofowicz and Kelly*, 2000]. By assuming that the underlying random variable is stationary we can construct a 95% confidence interval for each transformed target value to define the related limits of acceptability in the Gaussian domain (i.e. $\mu \pm 1.96\sigma$, where μ and σ are respectively the mean and standard deviation of the transformed sample). Finally, by applying the inverse of the NQT we obtain the 95% confidence interval for each target. When interpolation is not possible, the inverse of the NQT is computed numerically, by linear extrapolation of the tails of the NQT. In this study, we selected the upper and lower 3 points of the NQT. More details about the computation of the NQT and its inverse are given by *Montanari and Brath* [2004]. We assume that the signature is a stationary process. Of course, this assumption does not hold if significant changes in the catchment have occurred during the observation period.

[22] Given that the limits of acceptability are defined on the basis of the interannual variability of the corresponding target values, during model calibration the objective function should be evaluated for each hydrological year for which the model has been run. Only the parameter sets for which the simulation satisfies the limits of acceptability in each simulated hydrological year should be retained.

3. Application to the Luangwa River, Zambia

[23] In order to demonstrate the proposed framework in detail we refer to a real world case study of the Luangwa river basin for which only scarce hydrological data is available. The Luangwa basin (Figure 2) is a relatively pristine and remote area of about 150 000 $(\mathrm{km})^2$, located in Zambia, Southern Africa. The Northern, most upstream part of the basin is mountainous and is subject to many locally generated flash floods. The downstream parts consist of sandy/loam soils (among which black cotton soils) covered by typical tropical savanna vegetation such as Miombo, Mopane [*Frost*, 1996] and acacia species. Many of these lower areas are interspersed with wetlands, locally called "dambos," which are also used for land cultivation. The northeastern boundary (the Muchinga escarpment) consists of densely forested pristine wetland areas, having a different hydroclimatology from the low lying savannas. Temperatures on the escarpment are much lower and given the type of vegetation present, these areas have a higher capability of retaining moisture during the dry season than the lower savannah regions. This was also shown by modeling of the evaporation regime in the work by *Winsemius et al.* [2008]. The annual rainfall in the catchment is around 1000 mm per year. Rainfall is concentrated in one wet season from November until April.

3.1. Data Availability

[24] A daily river discharge time series is available near the outlet of the basin at the bridge on the Great East Road (see Figure 2). About 20 years of daily records are therein available from 1956 until 1976. Data were collected by observing the river stage that was subsequently converted to

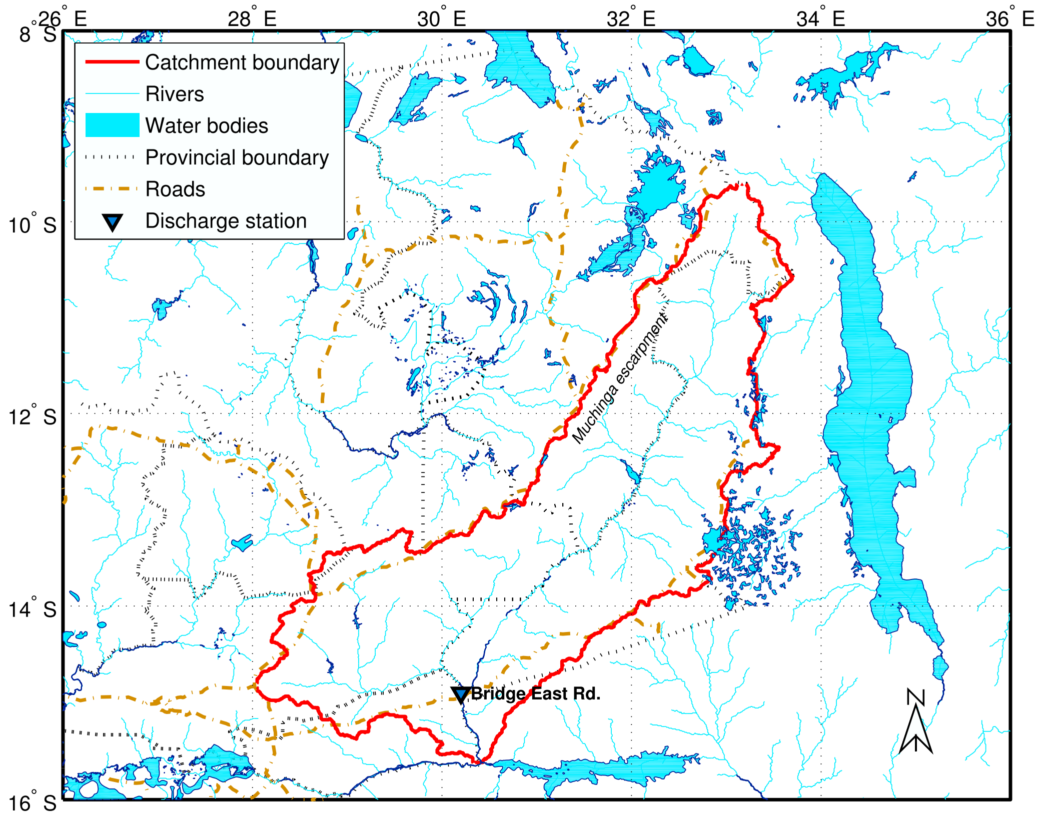

Figure 2. Luangwa basin, located in Southern Africa. The study area is plotted in red. The large water bodies and larger dambos are indicated in blue. The tarred road network is also plotted to emphasize the remoteness of the area. All smaller untarred roads inside the basin are only accessible in the dry season. At Great East Road Bridge, a long old time series of daily discharge data is available.

river discharge by means of a rating curve. Some of the years are affected by gaps or unreliable data. A subset of 16 hydrological years was selected (from October to September, 1956–1973, with exception of 1960), with only a minor amount of missing values, that were linearly interpolated.

[25] A considerable amount of monthly rainfall data was found in the Global Historical Climatology Network v. 2, for the period 1956–1973 (GHCN) [*Vose et al.*, 1992]. Although these records were originally based on daily rainfall, collected by the meteorological department of Zambia in some of the larger towns surrounding the basin, these numbers are no longer available to the public. Therefore, submonthly rainfall is not available over the same period as the available discharge records. Instead, two satellite rainfall estimates at finer timescale for the period 2002–2006 have been considered for this study: product 3B42 of the Tropical Rainfall Measuring Mission (TRMM) [*Huffman et al.*, 2007], which is available at 3 hour intervals, and the CPC/Famine Early Warning System (FEWS) daily estimates [*Herman et al.*, 1997]. Both data sets are merged rainfall estimates from different satellite instruments, among which microwave imagers and cold cloud duration from geostationary satellites. The TRMM satellite is the first satellite carrying a space-based radar instrument. Figure 3 shows a comparison of the two rainfall estimates, lumped over the whole catchment. From Figure 3 (top), there appears to be discrepancies over the years, which may be due to uncertainties in retrieval algorithms, instrumental errors, low availability of ground stations and aliasing problems related to the overpass frequency of the satellites used. However, there is a remarkable resemblance in the spectral properties (Figure 3, bottom). All available data, along with their observation period and observation time intervals have been summarized in Table 1.

3.2. Model Setup

[26] A lumped conceptual model has been preliminarily identified for the Luangwa river. It has been derived by modifying the HBV model [*Lindström et al.*, 1997]. For the purpose of this study, the structure was slightly simplified into the structure schematized in Figure 4 to obtain a more parsimonious tool in terms of involved parameters. The model structure's storage compartments consist of an interception store, a soil moisture store S_u [L] and two flow generating stores S_q [L] and S_s [L]. Interception I is limited, either by a threshold to net precipitation P_n [L T^{-1}], represented by parameter D [L T^{-1}], by the amount of rainfall P [L T^{-1}] or by the amount of potential evaporation E_p [L T^{-1}]. Therefore, interception can be computed as

$$I(t|D) = \min\left[P(t), D, E_p(t)\right], \quad (2)$$

where θ is the parameter set. Then, net rainfall is estimated as

$$Pn(t|D) = P(t) - I(t|D). \quad (3)$$

Throughfall is transferred to an unsaturated soil zone, corresponding to the HBV soil zone, with storage S_u, whose outgoing fluxes are controlled by 3 parameters, referred to

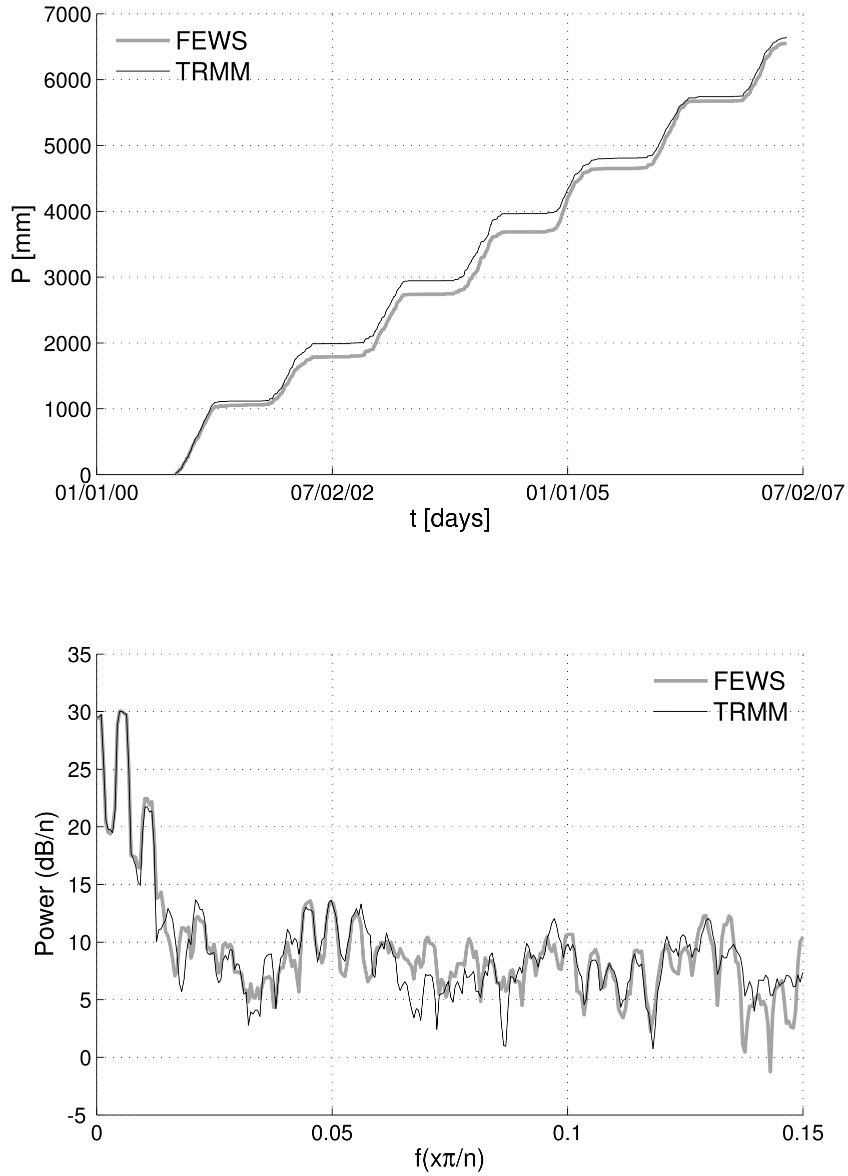

Figure 3. (top) Accumulated rainfall from FEWS and TRMM over the period 2000–2007. (bottom) Periodogram of FEWS and TRMM data.

as $S_{\max}$ [L], B (dimensionless) and l_p (dimensionless) (they are equivalent to the abbreviations "FC," "BETA" and "LP," respectively, as used by *Lindström et al.* [1997]). In detail, outgoing fluxes are transpiration T_a [L T^{-1}] and fraction of recharge r_c (dimensionless). The former is computed as

$$T_a(t|\theta) = \min\left(\frac{S_u(t|\theta)}{S_{\max} l_p}, 1\right) T_p(t), \tag{4}$$

Table 1. Summary of the Available Hydrological Information

Variable	Data Source	Observation Interval	Start	End
Discharge	Local authority	Daily	1956	1973
Rainfall	GHCN v. 2 (ground data)	Monthly	1956	1973
Rainfall	TRMM 3B42 (satellite based)	3 hourly	1998	ongoing
Rainfall	FEWS (satellite based)	Daily	2001	ongoing

where T_p [L T^{-1}] is the amount of potential evaporation left over after evaporation of intercepted water i.e. $E_p(t) - I(t|D)$. And r_c is given by

$$r_c(t|\theta) = \left(\frac{S_u(t|\theta)}{S_{\max}}\right)^B. \tag{5}$$

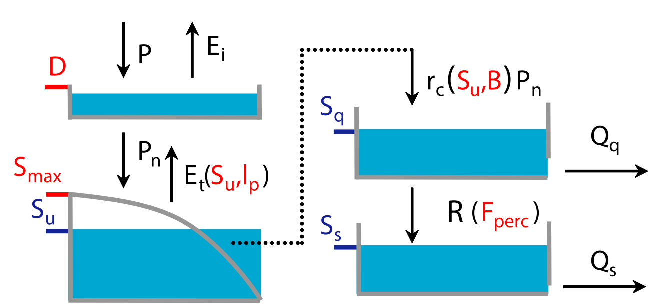

Figure 4. Structure of the modified HBV model.

Table 2. Uniform Prior Parameter Ranges for the Modified HBV Model

Parameters	Description	Unit	min	max
D	Interception threshold	mm day^{-1}	2 (fixed)	2 (fixed)
$S_{\max}$	Unsaturated zone capacity	mm	100	2500
l_p	Moisture stress fraction	-	0.1	1
B	Runoff generation power shape	-	1	5
F_{perc}	Maximum percolation rate	mm day^{-1}	0.01	4
K_q	Reciprocal of fast reservoir residence time	day^{-1}	0.01	0.5
K_s	Reciprocal of slow reservoir residence time	day^{-1}	0.001	0.1

Recharge ($r_c P_n$) is transferred to an upper zone with storage S_q. Streamflow is generated from this zone, assuming that it behaves as a linear reservoir with reciprocal of the residence time K_q [T^{-1}]. Q_q represents the fast flow generated from water bodies or seasonal wetlands (dambos). Finally, a lower zone with storage S_s [L] (conceptualizing groundwater) receives a maximum amount of percolation R [L T^{-1}] from the upper zone, determined by the parameter F_{perc} [L T^{-1}], according to the relationship

$$R(t|\theta) = \min\left(S_q(t|\theta)/\mathrm{d}t, F_{perc}\right). \tag{6}$$

[27] This zone also behaves as a linear reservoir, contributing to the base flow (Q_s) with one parameter K_s [T^{-1}] representing the reciprocal of the average residence time.

[28] Table 2 provides a list of the model parameters along with the respective prior range. This latter was based on previous experiences, also described by *Winsemius et al.* [2008].

[29] In order to apply the calibration framework herein proposed, after analyzing the available hydrological data, we identified the following objectives, along with the related target values, to be used to drive parameter estimation.

[30] 1. Shape of the recession curve of the hydrograph. Hard hydrological information.

[31] 2. Spectral properties of nonconcomitant daily river flows. Hard statistical information.

[32] 3. Monthly water balance estimates based on old monthly averaged records of rainfall and stream flow. Soft hydrological information.

[33] The computation of the above target values along with the related limits of acceptability is reported in Section 3.3.

3.3. Target Values and Limits of Acceptability

3.3.1. Shape of the Recession Curve

[34] The low flow behavior during flow recession periods is a property that is insensitive to rainfall forcing. In many cases the receding limb of the hydrograph can be described by a linear storage-discharge relationship and therefore generally plots as a straight line after a log transform of the discharge. The offset of this straight line is dependent on the initial storage condition. *Lamb and Beven* [1997] showed that recession periods may be combined into a master recession curve, which consequently may be used in model calibration, specifically for low flows. *Fenicia et al.* [2006] used this concept in a stepwise calibration of a conceptual model, showing that deviation from the straight line is caused by percolation and capillary rise. Depending on the climatic conditions, one may find fixed recession periods. Particularly in the tropics, seasonally defined dry periods without significant rainfall in each hydrological year can be identified by the modeler. Furthermore, in these periods the discharge contribution has a quite limited range, which mainly depends on the offset of the recession curve within the dry season period.

[35] In previous applications, recession curve analysis was primarily used to construct a master recession curve without considerations of uncertainties herein. To construct limits of acceptability, a number of sampled slopes can be derived from a number of recession periods. It is hypothesized here that the variability in the slope of the recession curve can be caused by uncertainties during river flow measurement [*Di Baldassarre and Montanari*, 2009], variability in the spatial distribution of soil moisture in the catchment at the beginning of the recession period or other natural variability, unaccounted for in the model structure.

[36] Within the present application, given that the climate over the Luangwa river basin is characterized by one dry period per year, the slope of the recession curve and the average discharge within recession periods was computed on a yearly basis from the daily river flow data collected in the period 1961 until 1972. Consequently, the limits of acceptability were constructed based on the NQT of the distribution of yearly sampled slopes and discharge contributions, using the method described in Section 2.3. A 95% confidence interval has been used. We assume that a parameter set is behavioral if it produces a river discharge simulation whose yearly slope and average discharge of the recession curve fall within the related limits of acceptability.

3.3.2. Spectral Properties of the Daily River Flows

[37] A relevant hydrological signature is provided by the spectral density function of discharge time series. *Montanari and Toth* [2007] described a maximum likelihood calibration procedure for rainfall-runoff models based on matching the spectral density of the modeled and observed discharge [see also *Whittle*, 1953]. *Montanari and Toth* [2007] proved that the Whittle likelihood is a powerful measure for model performance under the assumption of stationarity. Moreover, the estimator can be applied when the observed rainfall forcing over the basin is not concomitant with the available river discharge record.

[38] Let us assume that the spectral properties of daily river discharges are largely constrained by a lag-1 autoregressive process, i.e. the periodogram of the discharge time series is for a large part determined by the mean value, μ_Q, standard deviation, σ_Q and lag-1 autocorrelation coefficient, $\rho_1(Q)$ of the observed series. Therefore, μ_Q, σ_Q and $\rho_1(Q)$ allow us to define a three-element target vector to resemble the spectral calibration given by *Montanari and Toth* [2007]. Again, the related limits of acceptability can be estimated from the interannual variability of the targets (see Section 2.3). We assume that a parameter set is behavioral if it produces a river discharge simulation of which these

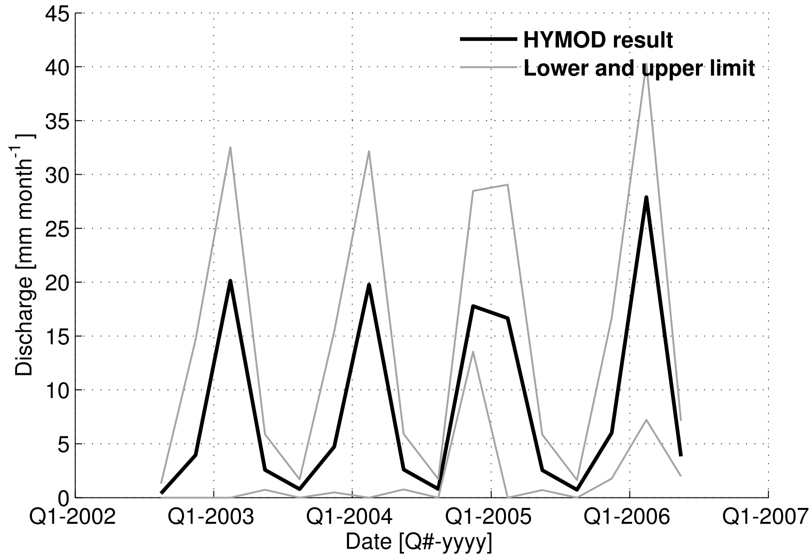

Figure 5. Limits of acceptability of seasonal-averaged discharge for the period 2002–2006 based on the HYMOD model, running at monthly time steps, using FEWS satellite rainfall as input.

yearly statistics fall within the related limits of acceptability for each year of the model simulation.

3.3.3. Monthly Water Balance

[39] We now turn to the analysis of a soft information signature, that is delivered by an auxiliary rainfall-runoff model at monthly timescale. GHCN (see Section 3.1) provides monthly ground station rainfall data in the period 1956–1973, during which daily river discharges at the basin outlet are also available. GHCN provides monthly rainfall for many parts of the world in historical periods so these data are potentially available in many basins, therefore allowing the computation of monthly mean areal rainfall over the catchment. These rainfall estimates were used to calibrate an auxiliary rainfall-runoff model running at monthly timescale, to reproduce estimates of mean monthly discharge for any period for which monthly rainfall is available. The daily time step model can then be constrained toward reproducing the discharges provided by the monthly auxiliary model (for more information about the use of auxiliary models, see *Seibert* [2001] and *Schaefli and Gupta* [2007]).

[40] As auxiliary model, we used a modification of the 5-parameter rainfall-runoff model HYMOD [*Boyle*, 2000], which was automatically calibrated with the self-adaptive differential evolution algorithm [*Brest et al.*, 2006] on the time period 1956–1973. The first year was used as a spin-up period.

[41] The model allowed us to reconstruct the monthly discharges at the basin outlet for the period 2002–2006, the run time period of the daily modified HBV model for which daily satellite rainfall observations are available that can be averaged to monthly values. The estimated monthly discharges were then used to constrain the daily modified HBV model simulation. The reasons why we consider this type of information as soft are the following: first of all the monthly water balance is less effective for parameter constraints compared with the two previous target values; second, the limits of acceptability have to be constructed through an auxiliary model, rather than directly from available observations. Furthermore, the accuracy of the rainfall used for the calibration of the auxiliary model may be strongly related to the density of the rain gauge network. The estimates may not prove accurate enough to warrant hard constraints.

[42] Limits of acceptability for the monthly discharges were computed by dividing the hydrological year into 4 seasons: November–January, February–April, May–June and July–October. For each of these seasons, we computed the season-averaged residuals of the modified HYMOD model, over the calibration period 1956–1973. The inter-annual variability of these residuals allowed us to estimate the related 95% confidence limits around the predicted value for each season. Figure 5 shows the seasonal limits for the period 2002–2006 and the case that FEWS satellite rainfall is used. Any model realization of the daily modified HBV model that gives any seasonally averaged discharge estimate outside the limits of acceptability presented here, has therefore been rejected.

[43] The derived limits of acceptability of each target value are given in Table 3.

3.4. Presentation of the Results

[44] Three million parameter sets have been sampled from the prior uniform distributions over the ranges given in Table 2. It turned out that D was quite insensitive with

Table 3. Limits of Acceptability Based on the Normal Quantile Transform[a]

Description	Lower Limit	Upper Limit
Recession: slope (day^{-1})	0.0055	0.014
Recession: mean contribution (m^3 s^{-1})	50.6	146
$\rho_1(Q)$	0.968	0.994
σ_Q (m^3 s^{-1})	269	1943
Water balance: Q, Nov–Jan (mm month^{-1})	−4.24	+10.69
Water balance: Q, Feb–Apr (mm month^{-1})	−20.7	+12.4
Water balance: Q, May–Jun (mm month^{-1})	−1.84	+3.3
Water balance: Q, Jul–Oct (mm month^{-1})	−0.91	+0.92

[a]The water balance limits of acceptability are dependent on the output of the monthly HYMOD auxiliary model. Therefore, only the deviation (+/−) from the modeled output is given. Figure 5 shows the time varying limits of acceptability in case FEWS rainfall is used as input.

respect to the selected targets. Therefore we decided to fix the value of D. Literature suggests to use a value between 1 and 5 mm day^{-1} for Southern Africa [e.g., *de Groen and Savenije*, 2006]. *Pitman* [1973] concluded that a value of 1.5 mm day^{-1} should be adequate for many river basins in South Africa. The large size of the basin and full spatial averaging of the rainfall used, suggests that a relatively low value should be sufficient, given the nonlinear behavior of this process. We selected a value of D equal to 2 mm day^{-1}. To emphasize the sensitivity of the results with respect to the rainfall input, we performed two calibration experiments by using the two available rainfall forcings (FEWS and TRMM). Therefore, for each parameter set the modified HBV model was run twice. Model run time was from 1 November 2000 (the start date of availability of the FEWS rainfall data set) until 30 September 2006. The first 23 months of model run time were used as spin-up time. This leaves precisely 4 years for model evaluation (i.e. 1 October 2002 until 30 September 2006). The simulations that satisfied the limits of acceptability for all the considered target values were retained as behavioral.

[45] First, three posterior parameter distributions were derived for each of the two calibration experiments, by considering only one of the objectives in turn. This was done in order to inspect the sensitivity of the model parameters to each objective. The resulting marginal posterior parameter distributions are given in Figure 6 for recession, spectral properties and monthly water balance, respectively. Figure 6 (top) presents the results using FEWS rainfall, while Figure 6 (bottom) presents the case with TRMM rainfall. Figure 6 displays smoothed (for display purposes) histograms over a total of 50 bins, being normalized so that the integral of the density over the 50 bins equals 1.

[46] It is clear that the individual objectives have a limited capacity to constrain the parameter space. The routing parameters F_{perc}, K_q and K_s are clearly conditioned, all by a different information signature. F_{perc} is well conditioned by the water balance, while K_q is well constrained by the spectral properties. Intuitively, K_s is well conditioned by the recession objective. In fact, K_s is in this model exactly equivalent to the slope of the log transform of the recession curve, which means that this parameter is automatically cut off at the upper and lower limit of acceptability. The original full prior range ($0.001 < K_s < 0.1$) is therefore not plotted.

[47] The other three parameters that determine how rainfall is partitioned into evaporation and streamflow, generally exhibit less sensitivity to the objectives used and their posterior distributions obtained through the different objectives are somewhat contradicting. Consider for instance $S_{\max}$, which, according to the objective related to spectral properties, shows a much lower mode than the ones obtained by the other objectives. This is not surprising as the different objectives focus on different behavior of the hydrograph. There are also some slight disagreements between the use of FEWS and TRMM rainfall estimates which may be an effect of their uncertainty (see Figure 3).

[48] Figure 7 shows the resulting marginal parameter distributions in histogram form, conditioned on the hard information (black dash-dotted line) and conditioned on all considered information (blue continuous line), obtained by application of equation (1). Again the top plots present results generated with FEWS rainfall while the bottom plots present results generated with TRMM rainfall. If all objectives are used, about 99.94% of all model realizations had to be rejected but clear constraints could be found for the parameters F_{perc}, K_q and K_s. It is evident that the constraint on the water balance results in a slight shift in the distribution of parameters $S_{\max}$ and B. Analysis of the relation between l_p and B reveals correlation (not shown here) which suggests the possible presence of equifinality. It is important to note that the final posterior parameter distributions are quite insensitive to the rainfall estimate used, although the rainfall estimate is significantly different (see Figure 3).

[49] The effect of the different constraints on the variability of the outputs becomes apparent in Figure 8. The graphs show what we refer to as "plausibility intervals," which reflect the band of the model output encompassed by the behavioral parameter sets. These intervals should not be regarded as confidence intervals because there is still subjectivity involved and only parameter uncertainty is considered [*Montanari*, 2007]. Introduction of the hard data results in a considerable reduction of variability in the discharge outputs. Furthermore, it can be seen that the soft information is capable of reducing the discharge uncertainty considerably as well. Finally, it can be seen from Figure 8 (bottom) that the evaporation regime is only marginally constrained by the targets used. It means that the modeled evaporation regime is insensitive to any of the targets considered for model calibration. Both the fact that the water balance related parameters are generally poorly identifiable and the fact that evaporation shows almost no sensitivity to our calibration efforts, indicates that additional information should be sought to better constrain these parameters. Given the results, it is likely that information on the soil moisture regime (i.e. conditioning parameters $S_{\max}$, l_p and B, and constraining the evaporation) may provide an additional strong constraint.

4. Validation

[50] For this particular basin, we do not have enough data to perform a meaningful split sample model validation. In fact, concurrent rainfall and runoff data are not available to check the reliability of the model to reproduce out-of-sample river flow records. This is a common situation in ungauged basins. However, in view of the available information, it is possible to assess the model reliability in reproducing the statistical properties, used for calibration, during a validation period.

[51] In detail, we ran all the accepted parameter sets for a 2 year time span outside the calibration period (October 2006 until September 2008) and computed for each year the hard targets, i.e. the recession slope, recession contribution, lag-1 autocorrelation, mean and standard deviation from the simulated discharge. Furthermore, the auxiliary rainfall-runoff model has been run for the period 2006–2008 to provide limits of acceptability of the seasonal discharge, as described in Section 3.3. Hence, a validation could be made by evaluating how well the accepted models were able to reproduce in a new independent time series the hydrograph behaviors as identified by the limits of acceptability of the signatures. This analysis has been performed with both the FEWS rainfall and TRMM rainfall product.

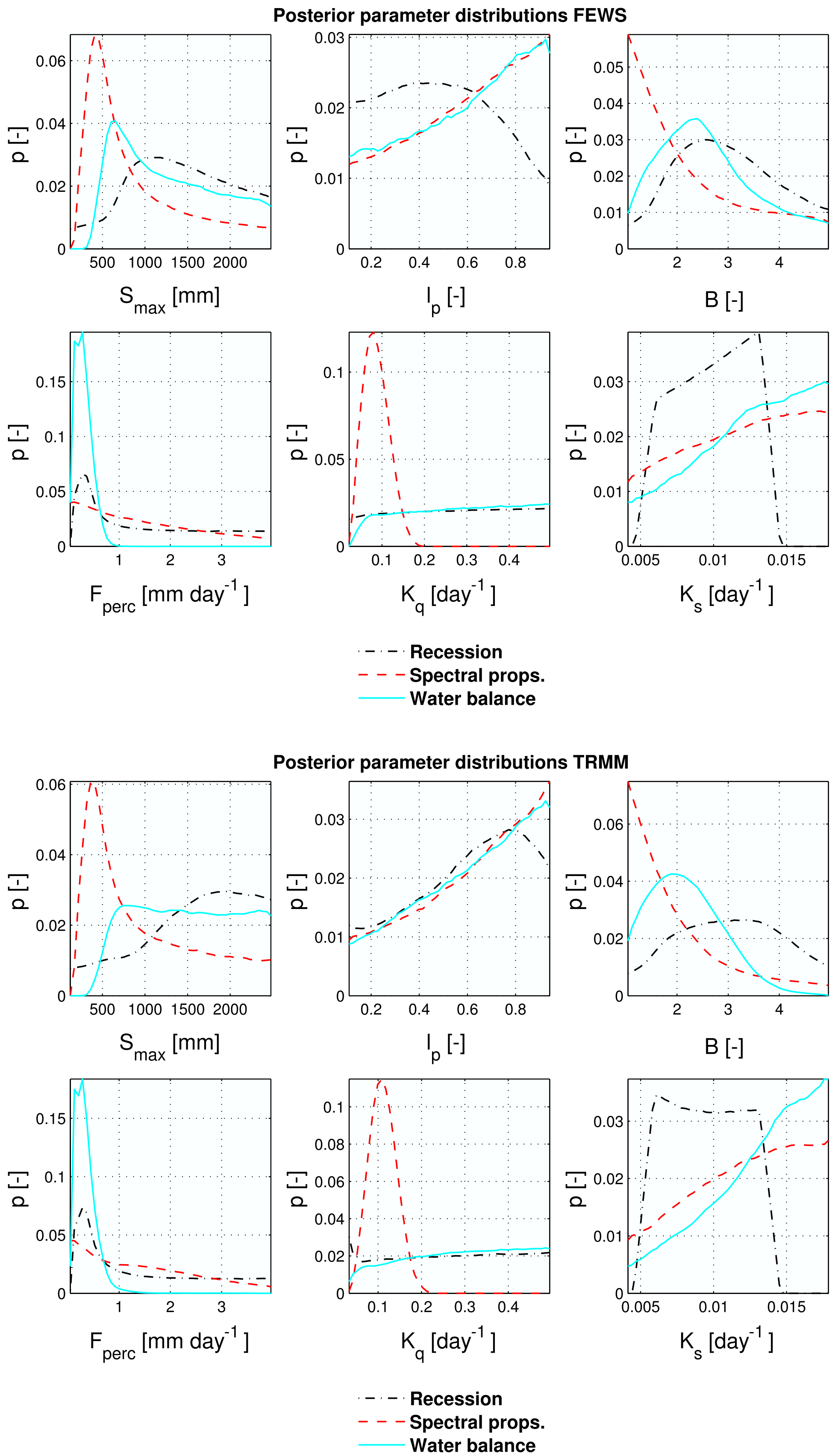

Figure 6. Effect of individual targets on the posterior parameter distribution using (top) FEWS and (bottom) TRMM.

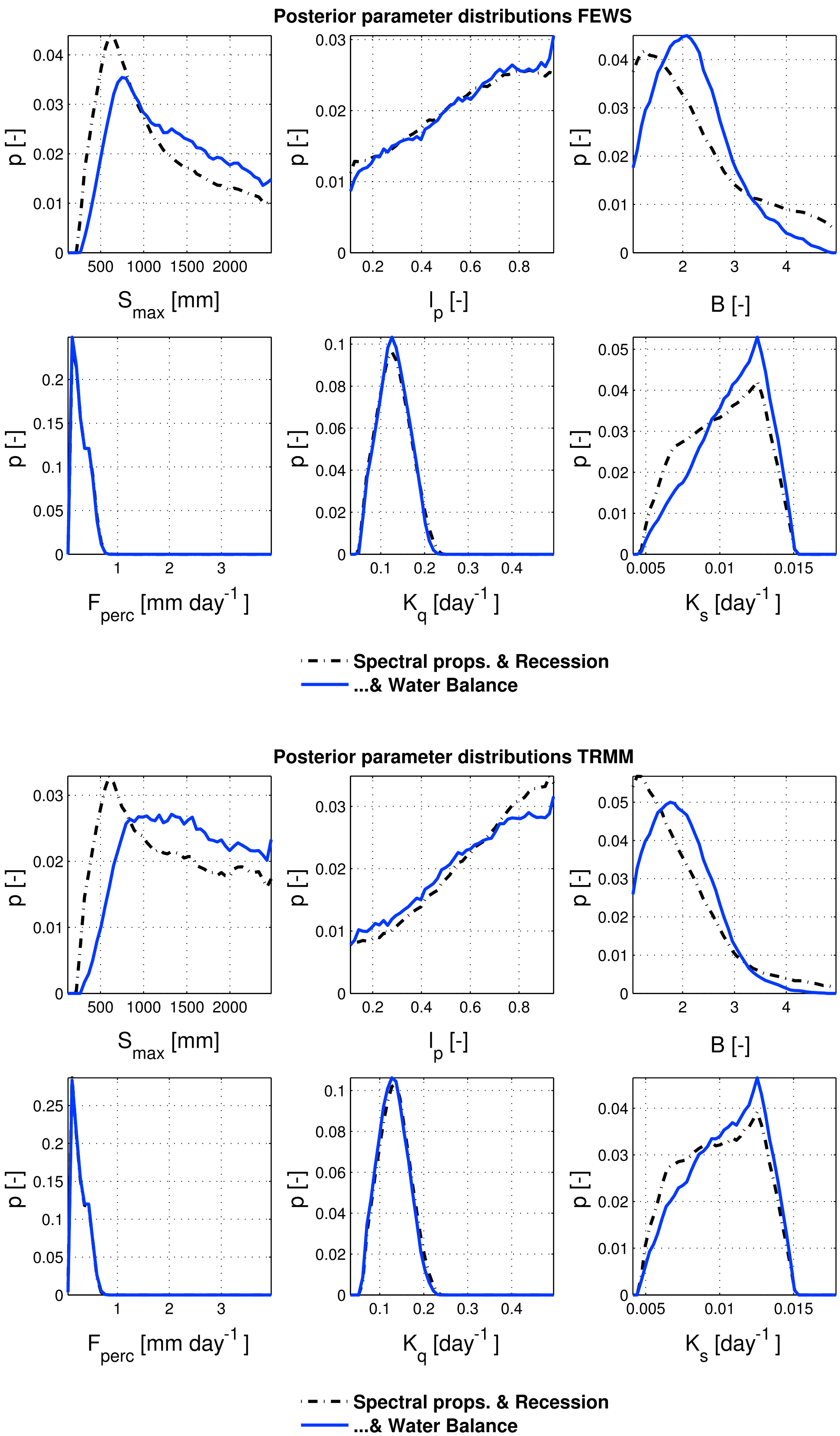

Figure 7. The effect of combinations of targets on the posterior parameter distribution using (top) FEWS and (bottom) TRMM.

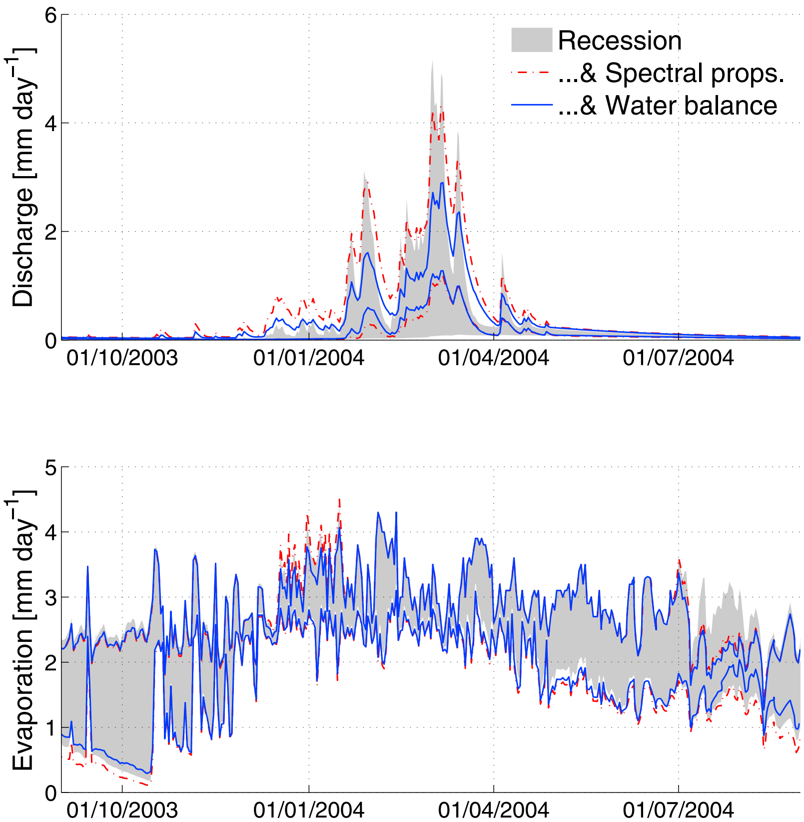

Figure 8. Parameter plausibility intervals of the output with multiple constraints: (top) discharge and (bottom) evaporation. The graphs show 5% and 95% plausibility intervals based on posterior parameter distributions.

[52] The results are presented in Figure 9 as normalized values with respect to the upper and lower limit of acceptability of each target. It demonstrates that most of the parameter sets satisfy all targets in each year. In detail, some inconsistencies were found for the lag-1 autocorrelation in the year 2007–2008 in the case FEWS rainfall is used as input, for which only 68% of the accepted parameter sets remain within limits. Moreover, the limits of acceptability on the seasonal water balance were fully met by 78% of the model realizations when TRMM rainfall was used, and by 97% when FEWS was used. This demonstrates some of the differences present in the two rainfall estimates and the related modeling uncertainty.

5. Discussion

5.1. Limitations of the Method

[53] A considerable limitation of our proposed framework is that the user needs multiyear information related to each objective to infer the limits of acceptability for the hard information. Moreover, the question remains how reliable the limits of acceptability are when only a limited number of years is available to estimate the variability of the target values. Nonetheless, this method may prove to be useful in many cases. For instance, it is well known that in Africa many hydrometric measurements were in fact taken from colonial times until the mid 70s, which resulted in the availability of quite long historical time series of river flows. Many of them are affected by missing values, or are not associated to concurrent rainfall observations, but nevertheless could be well used within the framework proposed here. Another possible limitation of the proposed approach is the assumption of stationarity that we introduced in order to compute statistical properties of river flows along a possibly extended time span.

[54] The methodology cannot prevent the occurrence of some subjectivity. For instance, it is up to the modeler what model structure to select, what information signatures to retrieve from the data, and to a certain extent, which information to consider "hard" and which "soft." Subjectivity in model calibration cannot be completely eliminated in the proposed approach. We believe this is unavoidable when modeling ungauged basins. In fact, the need for expert (subjective) knowledge increases with decreasing amount and reliability of the available information. However, we would like to underline that the framework we propose reduces the subjectivity of GLUE.

[55] Although the authors are aware of these limitations, we feel that the proposed framework allows the user to fairly objectively profit from any available observations to calibrate a hydrological model. We believe this is a valuable opportunity for ungauged basins.

5.2. What New Information Should Be Considered?

[56] The results show that the targets used in the Luangwa case study encapsulate information that allows us to constrain the model parameters. Clearly the constraining abilities of a target value depend on the hydroclimatological behavior of the catchment. For instance, in flash flood dominated catchments, the shape of the recession

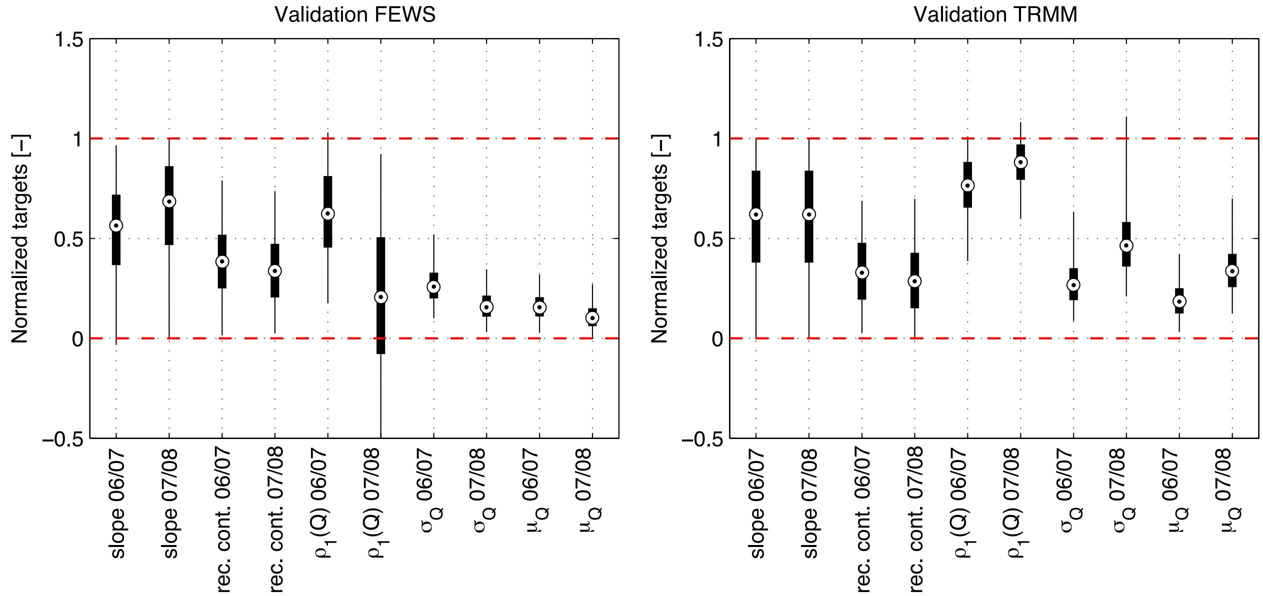

Figure 9. Performance of accepted models during validation. The box plots show for each signature and for each simulated year (x axis, 2006–2007 and 2007–2008) how well the accepted parameter sets remain within limits of acceptability. The box indicates the upper and lower quartiles, and the whiskers show upper and lower values found in all simulations. For display purposes, the targets have been normalized, so that the lower limits resemble the value 0 and the upper limits resemble the value 1. The limits of acceptability are shown by red dashed lines.

curve may not provide a strong control (there may be no base flow at all), while the spectral properties may form a more significant constraint. Based on the intermediate and final analysis of the parameter distributions, the modeler can identify additional information that may be useful to support the user in planning which additional data to collect. For instance, in the Luangwa River case study the posterior distribution of $S_{\max}$, l_p and B appears poorly constrained while l_p and B are correlated. Therefore a supplementary information source should be sought that can provide constraints on other fluxes than discharge. An independent soft data source that can be considered is for instance GRACE gravity information [e.g., *Tapley et al.*, 2004], which provides estimates of large-scale water storage variability. When properly postprocessed [*Klees et al.*, 2007], they may be applied when a river basin is at least 200 000 (km)^2 and has a regular shape. Models that do not obey the large-scale monthly change in storage, observed by GRACE may then be rejected. Unfortunately, the Luangwa basin is too small to apply this data source already, although solutions and spatial filters, used to reduce noise, are improving (e.g., anisotropic optimal filters [*Klees et al.*, 2008]). Other data, which could provide valuable information for further conditioning are satellite-based evaporation estimates [e.g., *Bastiaanssen et al.*, 1998]. These data can typically provide information about the spatiotemporal variability of the evaporation regime within the catchment, which was shown to be poorly constrained by the chosen discharge-related objectives. Unfortunately uncertainty of satellite-based evaporation estimates in natural catchments is hard to quantify. Therefore they may be considered as soft data [*Seibert and McDonnell*, 2002; *Winsemius et al.*, 2008] and limits of acceptability should be imposed in a different way than proposed here. This will be the subject of study in a future paper.

5.3. On the Value of Qualitative Observations

[57] In the description of our calibration framework, we have suggested the use of qualitative observations. Although we do not use qualitative information in our case study, it is important to identify the potential use of such information in ungauged conditions. *Seibert and McDonnell* [2002] show a case study in a well gauged catchment where many qualitative informations are used in a fuzzy acceptability framework because qualitative information is inherently soft. In principle, a modeler can use qualitative information in two ways: (a) by superimposing an updated prior parameter distribution; or (b) by superimposing a constraint on the model output. In the work by *Seibert and McDonnell* [2002], both approaches are used, dependent on the information considered, in a conceptual model structure that well simulates the physical properties of the catchment. In the ungauged case however, it is more likely that parameters will have to compensate for misconceptions in the model structure. This means that a modeler should be aware that model parameters in ungauged basins will have fewer physical meaning than in gauged basins, which means that it is usually safer to constrain the model output, which will consequently reflect on the parameter inference, than directly constrain the prior parameter distribution, which imposes strong assumptions on the physical representativeness of the parameters themselves.

6. Conclusions

[58] In this paper, we have proposed a calibration framework based on GLUE, which can be applied to ungauged

catchments. The framework enables parameter conditioning under the circumstances of highly uncertain data, and nonavailability of residual time series. Instead it uses information signatures with related limits of acceptability as calibration targets. The general subjectivity of GLUE is significantly reduced by means of an objective selection of acceptability limits. The framework allows for the integration of hard and soft information in the parameter conditioning process.

[59] In short, it consists of the following steps: a) From available rainfall and discharge observations, hard and soft information signatures are extracted. For hard information, calibration targets, in the form of limits of acceptability, are constructed based on the year-to-year variability of the signature present in the data. Because of the limited number of yearly samples usually available under data scarcity, a Normal Quantile Transform approach is used to construct 95% confidence intervals in the Gaussian domain. For soft information, the limits are either retrieved from a transfer model, if such a model was used to obtain the information, or they are based on the modeler's prior expert knowledge. b) An intermediate marginal parameter distribution is derived for each parameter through Monte Carlo sampling, where models, obeying all the hard targets are accepted as equally likely and all others are assigned a zero probability. c) Further sampling in a reduced parameter space is performed, now including soft information. d) Based on the posterior density of the parameters and the variability in the associated output realizations, the modeler may decide which information is expected to provide an additional strong constraint on the parameter space and if it is feasible to collect this information. e) The modeler includes, after collection, information retrieved from the newly observed variables (not necessarily discharge) in a further constraining step.

[60] In a case study, a daily rainfall-runoff model for the Luangwa catchment in Zambia has been conditioned using the framework. We have demonstrated the use of three information signatures, that are expected to be relatively insensitive to the absolute accuracy of the data used: the shape of the recession curve, the spectral properties of the discharge process and, as soft information, monthly estimates of the water balance through a calibrated auxiliary monthly water balance model, based on available concomitant monthly rainfall and discharge records. Two calibration experiments with our framework have been conducted, using two different satellite-based rainfall estimates, both generating considerably different rainfall amounts. The resulting parameter distributions show clear constraints in the parameters. The results of the two experiments are consistent with each other. While the number of behavioral discharge regimes is seriously reduced by the introduction of the information signatures, the water balance of the soil moisture, and consequently the evaporation regime remained marginally constrained. This indicates that additional independent information of the soil moisture or evaporation regime would be needed to further reduce parameter uncertainty.

[61] We believe the approach presented here is a valuable tool for prediction in ungauged basins. In fact it provides the modeler with a stepwise and objective procedure to condition parameter optimization on his perception and a guide to further data collection, based on the information at hand.

[62] **Acknowledgments.** Eduardo Martins, Murugesu Sivapalan, Hoshin Gupta and an anonymous referee are acknowledged for providing insightful reviews of the manuscript. We are very grateful for the expert knowledge of, and cooperation and friendship with the staff of the Zambia Meteorological Department, the Department of Water Affairs of the Ministry of Energy and Water Development, and the Zambia Wildlife Authority. A special thanks goes to our good friend Michael Mutale for all his logistic arrangements and provision of contacts in Zambia. Finally, we are grateful for the financial support by the Netherlands Organisation for Scientific Research (NWO).

References

Atkinson, S. E., R. A. Woods, and M. Sivapalan (2002), Climate and landscape controls on water balance model complexity over changing timescales, *Water Resour. Res.*, *38*(12), 1314, doi:10.1029/2002WR001487.

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag (1998), A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation, *J. Hydrol.*, *212–213*, 198–212.

Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*, 18–36.

Beven, K. (2008), On doing better hydrological science, *Hydrol. Processes*, *22*, 3549–3553, doi:10.1002/hyp.7108.

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

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

Beven, K., P. Smith, and J. Freer (2007), Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini, *J. Hydrol.*, *338*, 315–318.

Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, *354*, 15–32.

Boyle, D. (2000), Multicriteria calibration of hydrological models, Ph.D. thesis, Dep. of Hydrol. and Water Resour. Univ. of Ariz., Tucson.

Brest, J., S. Greiner, B. Boškoviá, M. Mernik, and V. umer (2006), Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems, *IEEE Trans. Evol. Comput.*, *10*(6), 646–657.

Bulygina, N., M. McIntyre, and H. Wheater (2009), Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, *Hydrol. Earth Syst. Sci.*, *13*, 893–904.

de Groen, M. M., and H. H. G. Savenije (2006), A monthly interception equation based on the statistical characteristics of daily rainfall, *Water Resour. Res.*, *42*, W12417, doi:10.1029/2006WR005013.

Di Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: A quantitative analysis, *Hydrol. Earth Syst. Sci.*, *13*, 913–921.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2006), Is the groundwater reservoir linear? Learning from data in hydrological modeling, *Hydrol. Earth Syst. Sci.*, *10*, 139–150.

Fenicia, F., J. J. McDonnell, and H. H. G. Savenije (2008), Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resour. Res.*, *44*, W06419, doi:10.1029/2007WR006386.

Franks, S. W. (2007), Integrating models, methods and measurements for prediction in ungauged basins, in *PUB Kick-Off Meeting*, edited by D. Schertzer et al., *IAHS Publ.*, *309*, 13–21.

Frost, P. (1996), The ecology of miombo woodlands, in *The Miombo in Transition: Woodlands and Welfare in Africa*, edited by B. Campbell, chap. 2, pp. 11–58, Cent. for Int. For. Res., Bogor, Indonesia.

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*, 3802–3813, doi:10.1002/hyp.6989.

Herbst, M., H. V. Gupta, and M. C. Casper (2009), Mapping model bahaviour using self-organizing maps, *Hydrol. Earth Syst. Sci.*, *13*, 395–409.

Herman, A., V. B. Kumar, P. A. Arkin, and J. V. Kousky (1997), Objectively determined 10-day African rainfall estimates created for famine early warning systems, *Int. J. Remote Sens.*, *18*(10), 2147–2159.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*, 38–55.

Immerzeel, W. W., and P. Droogers (2008), Calibration of a distributed hydrological model based on satellite evapotranspiration, *J. Hydrol.*, *349*, 411–424.

Kelly, K. S., and R. Krzysztofowicz (1997), A bivariate meta-Gaussian density for use in hydrology, *Stochastic Hydrol. Hydraul.*, *11*(1), 17–31.

Klees, R., E. A. Zapreeva, H. C. Winsemius, and H. H. G. Savenije (2007), The bias in GRACE estimates of continental water storage variations, *Hydrol. Earth Syst. Sci.*, *11*, 1227–1241.

Klees, R., E. A. Revtova, B. C. Gunter, P. Ditmar, E. Oudman, H. C. Winsemius, and H. H. G. Savenije (2008), The design of an optimal filter for monthly GRACE gravity models, *Geophys. J. Int.*, *175*(2), 417–432, doi:10.1111/j.1365-246X. 2008.03922.x.

Krzysztofowicz, R., and K. S. Kelly (2000), Hydrologic uncertainty processor for probabilistic river stage forecasting, *Water Resour. Res.*, *36*(11), 3265–3277.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Toward a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*(1–2), 161–177.

Lamb, R., and K. J. Beven (1997), Using interactive recession curve analysis to specify a general catchment storage model, *Hydrol. Earth Syst. Sci.*, *1*, 101–113.

Lee, H., E. Zehe, and M. Sivapalan (2007), Predictions of rainfall-runoff response and soil moisture dynamics in a microscale catchment using the CREW model, *Hydrol. Earth Syst. Sci.*, *11*(2), 819–849.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergstrom (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, *201*, 272–288.

Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, *330*, 368–381.

Merz, R., and G. Blöschl (2008a), Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, *44*, W08432, doi:10.1029/2007WR006744.

Merz, R., and G. Blöschl (2008b), Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, *44*, W08433, doi:10.1029/2007WR006745.

Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, *41*, W08406, doi:10.1029/2004WR003826.

Montanari, A. (2007), What do we mean by "uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrol. Processes*, *21*, 841–845.

Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, *40*, W01106, doi:10.1029/2003WR002540.

Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, *Water Resour. Res.*, *43*, W05434, doi:10.1029/2006WR005184.

Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, *44*, W03413, doi:10.1029/2007WR006240.

Pitman, W. V. (1973), A mathematical model for generating monthly river flow from meteorological data in southern Africa, *Rep. 2*, Hydrol. Res. Unit, Dep. of Civ. Eng., Univ. of Witwatersrand, Johannesburg, South Africa.

Pokhrel, P., K. K. Yilmaz, and H. V. Gupta (2009), Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, doi:10.1016/j.jhydrol.2008.12.004, in press.

Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (1998), A unifying framework for watershed thermodynamics: Balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, *Adv. Water Resour.*, *22*, 367–398.

Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrol. Processes*, *21*, 2075–2080.

Schaefli, B., and E. Zehe (2009), Hydrological model performance and parameter estimation in the wavelet-domain, *Hydrol. Earth Syst. Sci.*, *13*, 1921–1936.

Schymanski, S. J., M. Sivapalan, M. L. Roderick, J. Beringer, and L. B. Hutley (2008), An optimality-based model of the coupled soil moisture and root dynamics, *Hydrol. Earth Syst. Sci.*, *12*(3), 913–932.

Seibert, J. (2001), On the need for benchmarks in hydrological modeling, *Hydrol. Processes*, *15*, 1063–1064, doi:10.1002/hyp.446.

Seibert, J., and K. J. Beven (2009), Gauging the ungauged basin: How many discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, *13*, 883–892.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.

Selker, J., N. van de Giesen, M. Westhoff, W. Luxemburg, and M. B. Parlange (2006), Fiber optics opens window on stream dynamics, *Geophys. Res. Lett.*, *33*, L24401, doi:10.1029/2006GL027979.

Sivapalan, M. (2003), Prediction in ungauged basins: A grand challenge for theoretical hydrology, *Hydrol. Processes*, *17*, 3163–3170.

Son, K., and M. Sivapalan (2007), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resour. Res.*, *43*, W01415, doi:10.1029/2006WR005032.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004), GRACE measurements of mass variability in the Earth system, *Science*, *305*, 503–505.

Uhlenbrook, S., and J. Wenninger (2006), Identification of flow pathways along hillslopes using electrical resistivity tomography (ERT), in *Predictions in Ungauged Basins: Promise and Progress*, edited by M. Sivapalan et al., *IAHS Publ.*, *303*, 15–20.

Vaché, K. B., and J. J. McDonnell (2006), A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resour. Res.*, *42*, W02409, doi:10.1029/2005WR004247.

Vogel, R. M., and A. Sankarasubramanian (2003), Validation of a watershed model without calibration, *Water Resour. Res.*, *39*(10), 1292, doi:10.1029/2002WR001940.

Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid (1992), The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data, *Rep. ORNL/CDIAC-53, NDP-041*, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wagener, T. (2003), Evaluation of catchment models, *Hydrol. Processes*, *17*, 3375–3378.

Westhoff, M. C., H. H. G. Savenije, W. M. J. Luxemburg, G. S. Stelling, N. C. Van de Giesen, J. S. Selker, L. Pfister, and S. Uhlenbrook (2007), A distributed stream temperature model using high resolution temperature observations, *Hydrol. Earth Syst. Sci.*, *11*(4), 1469–1480.

Whittle, P. (1953), Estimation and information in stationary time series, *Ark. Mat.*, *2*(23), 423–434.

Winsemius, H. C., H. H. G. Savenije, A. M. J. Gerrits, E. A. Zapreeva, and R. Klees (2006), Comparison of two model approaches in the Zambezi River basin with regard to model confidence and identifiability, *Hydrol. Earth Syst. Sci.*, *10*, 339–352.

Winsemius, H. C., H. H. G. Savenije, and W. G. M. Bastiaanssen (2008), Constraining model parameters on remotely sensed evaporation: Justification for distribution in ungauged basins?, *Hydrol. Earth Syst. Sci.*, *12*(4), 1403–1413.

Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, *30*, 1756–1774, doi:10.1016/j.advwatres.2007.01.005.

Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resour. Res.*, *44*, W09417, doi:10.1029/2007WR006716.

Zhang, G. P., H. H. G. Savenije, F. Fenicia, and L. Pfister (2006), Modeling subsurface stormflow with the representative elementary watershed (REW) approach: Application to the Alzette River Basin, *Hydrol. Earth Syst. Sci.*, *10*, 937–955.

A. Montanari, Faculty of Engineering, University of Bologna, I-40126 Bologna, Italy.

H. H. G. Savenije, B. Schaefli, and H. C. Winsemius, Department of Water Management, Delft University of Technology, Stevinweg 1, NL-2628 CN, Delft, Netherlands. (h.c.winsemius@tudelft.nl)