Journal of Hydrology (2006) **330**, 368–381

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology

Pietro Mantovan [a], **Ezio Todini** [b,*]

[a] *Department of Statistics, University of Venice, Italy*
[b] *Department of Scienze della Terra e Geologico-Ambientali, University of Bologna, Via Zamboni, 67, 40126 Bologna, Italy*

Received 11 October 2004; received in revised form 10 April 2006; accepted 11 April 2006

KEYWORDS
Bayesian analysis;
Coherence;
Inconsistency;
Inference process;
Value of a statistical experiment;
GLUE

Summary The aim of the paper is to demonstrate the incoherence, in terms of Bayesian inference, of the generalized likelihood uncertainty estimation (GLUE) approach, introduced by Beven and Binley in 1992. This results into a reduced capacity of the technique to extract information, in other words to ''learn'', from observations. The paper also discusses the implications of this reduced learning capacity for parameter estimation and hydrological forecasting uncertainty assessment, which has led to the definition of the ''equifinality'' principle. The notions of coherence for learning and prediction processes as well as the value of a statistical experiment are introduced. These concepts are useful in showing that the GLUE methodology defines a statistical inference process, which is inconsistent and incoherent.

Introduction and scope

In the last decade, increased scientific interest had been shown on parameter estimation as well as on the assessment of parameter and forecasting uncertainty in hydrological models. As pointed out by Vrugt et al. (2005), hydrological models conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed through relatively simple mathematical equations with model parameters that often do not represent directly measurable entities and must therefore be estimated using input and output measurements. This inevitably leads to uncertain parameter estimates (and consequently to uncertain forecasts) due to structural errors in the model schematisation, errors in the input (such as for instance rainfall, temperature, and upstream inflows) and output measurements (generally downstream discharges), errors in the initial conditions of several state variables (such as for instance soil moisture content, the extent of saturated areas, and the accumulated snowpack). In addition, the parameter estimation phase is generally carried out only based upon the downstream residuals, without really accounting for their temporal correlation as well as their non–stationarity.

* Corresponding author. Tel.: +39 051 2094537; fax: +39 051 2094522.
E-mail address: todini@tin.it (E. Todini).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2006.04.046

If the statistical properties of all these errors and residuals are not properly accounted for, biased parameter estimates are obtained together with wrong assessments of parameter estimate uncertainty as well as of forecasting uncertainty (Vrugt et al., 2005). Several methods have appeared in the literature, which seek to provide meaningful confidence bands for parameter estimates in conceptual non-linear watershed models. These approaches range from classical Bayesian techniques (Kuczera and Parent, 1998; Thiemann et al., 2001; Vrugt et al., 2003a; Kavetski et al., 2003), to the pseudo-Bayesian (Beven and Binley, 1992; Freer et al., 1996), set theory (Keesman, 1990; Klepper et al., 1991; van Straten and Keesman, 1991; Vrugt et al., 2001), multiple criteria (Gupta et al., 1998; Yapo et al., 1998; Boyle et al., 2000; Vrugt et al., 2003b), and recursive model and parameter identification techniques (Thiemann et al., 2001; Young, 2001; Vrugt et al., 2002; Wagener et al., 2003) to account for non-stationarities in model residuals. Unfortunately, most of the previously mentioned approaches tend to consider the model uncertainty to be mostly produced by parameter uncertainty and use, as a measure of fit, a function based on the model residuals. Essentially, the model is assumed to be correct, and the parameter estimated conditional on this assumption. To overcome this problem and to incorporate a better treatment of the input, output, parameter and model structural uncertainties in hydrologic modeling, Vrugt (2004), Vrugt et al. (2005) proposed a simultaneous parameter optimization and data assimilation (SODA) method, which combines the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm (Vrugt et al., 2003a), with an ensemble Kalman filter (Evensen, 1994). Although the SODA approach seems to be of great interest, it is the belief of the authors that further research should be fostered into a more formal Bayesian approach that would incorporate model, parameter, initial conditions uncertainty as well as non-stationarity and correlation into the model identification and parameter estimation process.

Following this line of thought, and for the sake of clarity, the present paper investigates certain aspects of one of the previously mentioned techniques, the ''generalised likelihood uncertainty estimation'' (GLUE) method introduced by Beven and Binley (1992). The reason for this focus is that GLUE has been interpreted as a Bayesian approach (or more precisely as an extension of the Bayesian approach) and widely used over the past ten years to analyse and estimate predictive uncertainty, particularly in hydrological applications (Beven and Binley, 1992; Beven, 1993; Romanowicz et al., 1994; Freer et al., 1996; Franks and Beven, 1997; Zak et al., 1997). However, most of the above mentioned authors, as well as many users all over the world, are apparently not aware that GLUE is inconsistent with the Bayesian inference process, which inevitably leads to large overestimation of uncertainty, both for the parameter estimates and the resulting simulation forecasts. Users are apparently attracted by the ease of use of GLUE, which provides in a single ready-made package, simple understandable ideas, workable inference techniques and a spurious sensation of freedom in relation to the strict rationale of statistical modelling and Bayesian inference.

Parameter estimation and/or predictive uncertainty?

Before entering into a discussion of the GLUE approach, it is necessary to clarify a fundamental concept fully embedding the Bayesian approach. In the classical Bayesian approach, the parameters of a model are not necessarily considered as the representations of physically meaningful quantities which have true (albeit unknown) values, but rather temporary ''dummy'' or ''convenient'' quantities of uncertain nature (on which all our uncertainty is projected) to be marginalized out by means of their posterior probability density, which is obtained from observations via the Bayesian inference process.

Therefore, it is worthwhile clarifying the objective of our uncertainty assessment. Is ''parameter estimation'' our main objective, or is it ''prediction''?

If the objective is ''parameter estimation'', we implicitly assume that our parameters have a true albeit unknown value to be found, and this can be estimated once the posterior density has been derived, either in terms of the maximum likelihood value or as the expected value. But the parameter estimation does not necessarily reflect the original Bayesian idea.

When the objective is ''prediction'', the Bayesian approach ''does not require'' the estimation of a specific parameter value, but rather the estimation of ''their entire posterior probability density'' that expresses our uncertainty after sampling the observations. Maybe, it is worthwhile bearing in mind that a ''deterministic'' quantity can be represented as a delta of Dirac over its value, while an uncertain quantity requires either of the descriptions of its probability distribution.

To clarify this concept, let us consider the following observable random vector $\mathbf{y}_i = (y_1, y_2, \ldots, y_c)^{\mathrm{T}}$, $c \geqslant 1$, $i = 1, 2, \ldots, n$, of predictive interest, such as for instance, for $c = 1$, a time series of discharges or water stages at a downstream section in a river, sampled at constant time steps $i = 1, \ldots, n$, jointly observable with the predictor vector $\mathbf{x}_i = (x_1, x_2, \ldots, x_k)^{\mathrm{T}}$, $k \geqslant 1$, $i = 1, 2, \ldots, n$, such as for instance, for $k = 1$, a time series of precipitation. Under the conditional stochastic independence, let $f(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{x}_i)$ be the conditional probability density function of the observable vector $\mathbf{y}_i$ conditioned on the covariate vector $\mathbf{x}_i$ and the parameter vector $\boldsymbol{\theta}$. The conditional probability density function $f(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{x}_i)$ implicitly summarizes the modelled dependence of the observable random vector $\mathbf{y}_i$ from the covariate vector $\mathbf{x}_i$ by the parameter vector $\boldsymbol{\theta}$. This relation may be explained, for instance, by the conditional expected value:

$$E(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{x}_i) = \mu(\boldsymbol{\theta}, \mathbf{x}_i) = (\mu_1(\boldsymbol{\theta}, \mathbf{x}_i), \mu_2(\boldsymbol{\theta}, \mathbf{x}_i), \ldots, \mu_c(\boldsymbol{\theta}, \mathbf{x}_i))^{\mathrm{T}},$$

$i = 1, 2, \ldots, n$, where the conditional mean model $\boldsymbol{\mu}(\boldsymbol{\theta}, \mathbf{x}_i)$ is identified by the parameter vector $\boldsymbol{\theta}$, the size of which is somehow related to the complexity of the conditional mean model and, as previously stated, not necessarily encapsulating a physical meaning.

More explicitly, but in an equivalent way, the observable random vector $\mathbf{y}_i$ can be assumed to be defined by a structural part $\boldsymbol{\mu}(\boldsymbol{\theta}, \mathbf{x}_i)$ modelling directly the expected dependence of the observable random vector $\mathbf{y}_i$ from the

covariate vector $\mathbf{x}_i$ by the parameter vector θ composed with an additive or multiplicative random error. From the distribution of the random error, follows the conditional distribution of the observable random vector $\mathbf{y}_i$.

The scope of the Bayesian inference is to derive from the use of the historical observations $(\mathbf{Y}_n, \mathbf{X}_n)$, $\mathbf{Y}_n = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)^T$, $\mathbf{X}_n = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^T$, a posterior probability density for the parameter vector θ, namely $g(\theta|\mathbf{Y}_n, \mathbf{X}_n)$, with which it will then be possible to marginalize out the conditionality on the parameters. This involves integrating over Θ, the entire domain of existence of the parameters, the conditional probability density function $f(\mathbf{y}_p|\theta, \mathbf{x}_p)$ of the observable vector $\mathbf{y}_p$ conditioned on the covariate vector $\mathbf{x}_p$ and the parameter vector θ that identifies the specific model, to derive the predictive probability density for the pth observation:

$$f(\mathbf{y}_p|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{x}_p) = \int_{\Theta} f(\mathbf{y}_p|\theta, \mathbf{x}_p) g(\theta|\mathbf{Y}_n, \mathbf{X}_n)\, \mathrm{d}\theta, \quad (1)$$

which can be used to describe the predictive uncertainty both in ''hindcast'' mode, when $p \leqslant n$, and in ''forecast'' mode for $p > n$.

From Eq. (1), note that the dependence of the predictive probability density on the uncertain parameters has been marginalized by taking its expected value on the basis of the posterior parameter density. A similar expression results if θ belongs to a discrete parametric space. Of course, the predictive probability density can be extended to more that one next observation.

The concept on which Beven and Binley's (1992) approach is based, where ''alternative models'' (namely the same model with alternative parameters) are considered to have the same predictive value (the concept of ''equifinality''), reflects this concept up to a certain point, in the sense that, when dealing with non-linear models, the typical hydrological approach of using a model with fixed parameter values may lead to large predictive biases, while by expressing Eq. (1) in discrete form, one should use an ensemble of model predictions (one for each parameter vector realization) averaged with its derived posterior probability mass. Unfortunately, GLUE on the one hand does not provide a correct expression for Eq. (1) and, on the other hand, produces extremely ''flat'' posterior probability distributions that lead to unrealistically wide predictive uncertainties, as opposed to a correct application of the Bayesian inference approach, which leads to peakier posterior densities and smaller predictive uncertainties. As a matter of fact, one is supposed to make the hypothesis of ''equifinality'' at the onset of the Bayesian inference process, namely that all the models (one per parameter vector realisation) have the same informative value, due to our prior lack of knowledge (typically expressed by a uniform prior distribution for the parameters). But then, by means of the observations, one aims to produce peakier posterior densities; so the real scope of the whole Bayesian process is to achieve ''inequifinality'' where some of the models are more likely to be correct than others. The most important aspect to be borne in mind (which should not be confused with the concept of ''equifinality'') is that, in order to avoid predictive biases, following Eq. (1), one has to take into account and use all the model predictions, namely one per parameter vector realization (and not only the one relevant to the most likely parameter values), and all the predictions have then to be averaged using the derived posterior probability function to marginalize out the uncertainty due to the lack of knowledge of the actual parameter values to be used.

Therefore, when dealing with the derivation of the predictive probability, which is the main object of modelling (and in particular of hydrological modelling), the estimation of the parameter values is a false problem, since the full predictive uncertainty information will be encapsulated into the derived posterior probability density of the parameters given the observations.

This brings us to the reason for this paper. As it will be demonstrated in the following sections, GLUE does not allow for a correct estimation of the posterior probability density, since the choice of the ''less formal likelihoods'' highly reduces its capability to extract information from the observations and consequently it is not advisable to use GLUE for estimating the predictive probabilities on the basis of the resulting posterior probability density.

The GLUE approach

As already noted, the generalised likelihood uncertainty estimation technique was introduced by Beven and Binley (1992) starting from concepts that, although expressed in different terms, are quite similar to the Bayesian ones. According to Beven et al. (2000), GLUE ''represents an extension of Bayesian or fuzzy averaging procedures to less formal likelihood or fuzzy measures''. The concept of ''the less formal likelihood'', which represents the main element of differentiation with the Bayesian inference, is a remarkable aspect of the GLUE methodology. In theory, this concept was introduced to overcome the need to advance detailed distribution functions of the observable variables and/or of errors even in complex situations, such as when there is more than one source of errors, when the explicative models provided are complex or when the number of parameters used to realize a process of learning is high.

Unfortunately, as will be shown in the sequel, the use of such less formal likelihoods results in GLUE loosing the learning properties of the Bayesian inferential approach, and so the GLUE methodology is incoherent and inconsistent with the statistical inference process. The loss of the learning properties prevents the possibility of reducing the uncertainty down to a lower limit imposed by the structural properties of the model and the other sources of errors, by increasing the number of observations. This inevitably results in very ''flat'' parameter posterior densities, a fact that may have led to the emergence of the ''equifinality'' principle. In reality, a proper Bayesian approach would eventually start from a flat non-informative prior distribution with the aim at reaching an ''inequifinality'' condition, namely a ''peaky'' posterior density over the most likely parameter sets, by adding the sampling information provided by an increasing number of observations.

Although, according to Beven and Binley (1992), GLUE was formulated in order to overcome uncertainties relevant to several sources of errors: ''error due to poorly defined

boundary conditions and input data; error associated with measurements used in model calibration; and error due to deficiencies in model structure''; in reality it does not account for them directly, but only indirectly through the statistical properties of the model residuals. GLUE is in fact conditional on model structure, on input errors and on initial conditions, which are not taken as uncertain. Uncertainty is only considered to be generated by the lack of knowledge of model parameter values as well as possibly by the predictand observation errors, with the implicit assumption of additive errors, ''perhaps after a suitable transformation'' (Beven and Freer, 2001), which can be summarised in Eq. (2), similar to what is generally done in the Bayesian approach:

$$\mathbf{y}_i = \boldsymbol{\mu}(\theta, \mathbf{x}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \ldots, n; \ \ \theta \in \Theta; \tag{2}$$

where $\mathbf{y}_i$ is the predictand at step i, $\mathbf{x}_i$ is the predictor vector incorporating all the input observations up to step i, θ is one of the possible choices of the parameter vector and $\boldsymbol{\varepsilon}_i$ is an additive error term which will incorporate the measurement errors on the predictand, the model structure errors as well as all the errors generated by the lack of knowledge of the ''true'', if any, or optimal parameter set. In reality, the choice of the additive error formulation given in Eq. (2) is not binding when the main objective is the derivation of the predictive probability distribution given by Eq. (1).

Conversely, when the objective is ''parameter'' estimation, several problems may occur. Firstly, the formulation of Eq. (2) does not guarantee that all the parameters can be properly estimated (the concept of ''observability'' that can be found in Gelb, 1974, and in several papers such as for instance in Gupta and Sorooshian, 1983, and Sorooshian and Gupta, 1983, may not be fulfilled), which implies that other sources of information should be used, not necessarily limiting ourselves to a downstream cross section water stage or discharges of predictive interest. For instance, the recent availability of distributed satellite measurements could be of great relevance by providing information on the extent of saturation surfaces or (hopefully in a nearby future) on the soil moisture content, both highly related to the parameters of the distributed hydrological models. Secondly, by not directly accounting for the variability induced by the other sources of errors, such as for instance the measurement errors in $\mathbf{x}_i$, the input variables, the estimated parameter values, due to the generally non-linear links between parameters and measurements in the hydrologic models, will inevitably be biased.

Therefore, two aspects must be clearly borne in mind all throughout the paper:

(i) the focus of this work is the derivation of the predictive probability of a predictand (Eq. (1)) and
(ii) the parameters, although generally considered as ''deterministic'' in a physically meaningful model of given structure, encapsulate ''all our uncertainties'', thus becoming ''uncertain'' quantities. This implies that it is not necessary to estimate specific parameter values (i.e., as the expected value or the most likely one), but rather their posterior probability distribution, which is the essential information required for assessing the predictive probability.

More details on points (i) and (ii) related to the inductive reasoning in statistical inference and its mathematical characterization can be found in de Finetti (1975, Chapters 11 and 12).

The basic formulation

The GLUE formulation closely follows the classical Bayesian inference approach, from which it mainly differs in the choice of the likelihood to be used, the so called ''less formal likelihood'', as well as in the discrimination between the ''behavioural'' and the ''non-behavioural'' parameters.

The main steps in the GLUE methodology are

(1) the subjective choice of a parameters prior probability distribution;
(2) the use of a subjective ''less formal likelihood'';
(3) the selection of the behavioural parameters;
(4) the derivation of the parameters posterior probability distribution via Monte Carlo sampling;
(5) the derivation of the predictive probability distribution.

All these points will be briefly described in the sequel.

It is worthwhile noting that no parameter estimation is included in GLUE. This is consistent with the objective of correctly assessing the predictive uncertainty, given a specific model structure, on the assumption that the parameter estimates are dummy quantities, incorporating not only the true (if any) unknown parameter value, but also the uncertainty deriving from all the possible sources (model schematization, input and output data measurement errors, etc.).

The assumptions on the parameters prior to probability distribution

In order to initiate the GLUE analysis, one has to assume a prior distribution for the parameters from which all the parameter sets θ_j, $j = 1, 2, \ldots, m$, are drawn using a Monte Carlo type extraction.

The probability distribution assumed for θ is invariably a non-informative multi-uniform distribution, where the parameters are initially assumed independent with uniform marginal distributions, mostly describing their range of existences, which is generally possible if the parameters are physically meaningful. Although the real benefit of the Bayesian approach stems from the possibility of starting the inferential process from informative priors, the choice of such a distribution is not strictly binding when dealing with hydrological models within the frame of the Bayesian inference approach, particularly if the number of parameters is not large. In fact, in the case of hydrological models the number of observations used to derive the posterior distribution is generally so large that, provided that the inference mechanism works properly and the number of parameters is not excessive, the dependence on the prior distribution becomes marginal.

Unfortunately, as it will be demonstrated, this property does not apply when using the less formal likelihoods defined in GLUE.

Less formal likelihood functions

If the classical likelihood functions (Box and Tiao, 1973) were assumed, there would be scarce or very little difference between GLUE methodology and the Bayesian approach to the statistical inference. In the Bayesian statistical inference, the likelihood function is directly connected (proportional) to the probability density function of the observable random vector $\mathbf{y}_i$, conditional to the knowledge of the parameter vector $\boldsymbol{\theta}$ and the predictor vector $\mathbf{x}_i$ (proper likelihood).

These probabilistic laws adopted to explain the observable random vector, with the definition of an error model, are generally aimed at modelling the conditional expected value and variance. They follow the physical laws governing the phenomenon of the observable random vector $\mathbf{y}_i$ (water levels, discharges, etc.) in its relation to the predictor vector $\mathbf{x}_i$ (precipitation, upstream inflows, etc.).

Alternatively, the ''less formal likelihood'' functions introduced in the GLUE methodology are conceived as functions of a synthesis of errors between the observed values and the values provided by the deterministic model.

To make the process steps clearer, a real-valued observable random variable y_i is considered in the following.

Given the acquired sampling observations $(\mathbf{y}_n, \mathbf{X}_n)$, expressed in terms of the notation introduced in the second section, the following sampling values are considered for the observable variable:

The sampling mean:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The sampling variance (sampling mean square error in the absence of a model) is

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

The sampling residual variance is

$$s_n^2(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \mu(\boldsymbol{\theta}, \mathbf{x}_i)]^2,$$

namely, the sampling mean square error estimated on the basis of the values given by the model identified by the parameter vector $\boldsymbol{\theta}$.

In a series of papers, the following less formal likelihood functions of $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Theta$, with Θ being the support (the domain of existence) of the parameter vector $\boldsymbol{\theta}$, have been introduced and used in the literature.

For Nash and Sutcliffe efficiency criterion with shaping factor $N \geqslant 1$ (see Freer et al., 1996):

$$\mathscr{L}_1(\boldsymbol{\theta}; \mathbf{y}_n, \mathbf{X}_n) \triangleq \begin{cases} \left[1 - s_n^2(\boldsymbol{\theta})/s_n^2\right]^N & \text{if } \boldsymbol{\theta} : s_n^2(\boldsymbol{\theta}) \leqslant s_n^2, \\ 0 & \text{if } \boldsymbol{\theta} : s_n^2(\boldsymbol{\theta}) > s_n^2. \end{cases}$$

For inverse error variance with shaping factor $N \geqslant 1$ (see Beven and Binley, 1992):

$$\mathscr{L}_2(\boldsymbol{\theta}; \mathbf{y}_n, \mathbf{X}_n) \triangleq [s_n^2(\boldsymbol{\theta})]^{-N}.$$

For exponential transformation of error variance with shaping factor $N \geqslant 1$ (see Freer et al., 1996):

$$\mathscr{L}_3(\boldsymbol{\theta}; \mathbf{y}_n, \mathbf{X}_n) \triangleq \exp\{-N s_n^2(\boldsymbol{\theta})\},$$

as well as many others.

In particular, the ''less formal likelihood'' based on the Nash and Sutcliffe criterion plays a critical role in the GLUE methodology: although not substantiated by facts, it is believed to be connected to the equifinality property of the models in hydrological and environmental applications, as well as with the GLUE approach itself (see Beven et al., 2000, section ''Choosing a likelihood measure''). The possibility to introducing such less formal likelihood functions prevents the formulation of precise distribution functions of the observable variables and/or of the errors in complex situations — the latter due to the presence of many sources of error, to the complexity of the explicative models considered and to the high number of parameters with which to build a learning process. Unfortunately, as it will be shown later, the use of less formal likelihood functions generally leads to paradoxical inferential results.

Note that the above — mentioned less formal likelihoods are not defined to represent probability density functions, as usually defined in classical or Bayesian statistics, but rather their integral, the probability distribution function, and they are forced to be limited between 0 and 1 via renormalization of the obtained values.

In the sequel, without loss of generality, if not otherwise specified, a ''less formal likelihood'' is understood as one of the above three less formal likelihoods. As it will be shown later, the use of such less formal likelihood functions is itself a questionable aspect of the inferential methodology under review.

Behavioural parameters

Another important element in GLUE is the distinction between behavioural and non-behavioural parameters, which in GLUE terminology are also called ''models''. A behavioural parameter is that specification of the vector $\boldsymbol{\theta}$ whose likelihood value exceeds a certain threshold. The non- behavioural parameters are not taken into account in the inference process. Among the behavioural parameters, the attention is focused on a subset of effective parameters. An effective parameter is that specification of the vector $\boldsymbol{\theta}$ whose likelihood value is high. The introduction of such a distinction in GLUE is in fact a consequence of the choice of the less formal likelihoods that do not guarantee that their domain of existence is positive. This means that all the parameters inducing negative values (or more in general values below a threshold) of the less formal likelihood will be considered non-behavioural and dropped as done with $\mathscr{L}_1(\boldsymbol{\theta}; \mathbf{y}_n, \mathbf{X}_n)$. In reality, there is nothing wrong with using prior information to exclude points from the parameter space, but this should not be done on the basis of the likelihood, as in GLUE.

Although the distinction between behavioural and non-behavioural parameters may also lead to biased results, for the sake of clarity this problem will not be discussed in this paper in order to concentrate the attention of the reader on the most critical points.

The derivation of the parameter posterior probability distribution via Monte Carlo sampling

In GLUE, similarly to what is generally done in the unconjugate Bayesian inference approach, the parameter posterior probability distribution is obtained through the use of Monte Carlo simulation techniques. The assumption of a uniform prior probability density function for the parameter vector on a limited support facilitates such simulations, and only the behavioural parameters and the corresponding values of the likelihood function are considered. In the case of GLUE, it is important to note that forecasts are affected only by the set of models taken as ''behavioural'', and as previously pointed out, this conditioning to behavioural parameters may lead to forecasting bias.

Following the GLUE procedure, a multi-uniform prior probability distribution on the discrete space Θ is assumed for the behavioural parameters θ, and a large number m of parameter sets $\theta_j \in \Theta$, $\forall j = 1, m$ are generated at random. The value of the less formal likelihood (generally chosen among the ones introduced in section ''Less formal likelihood functions'') is then estimated for each specific parameter set θ_j and the results are then presented as ''dotty plots'' after renormalizing the values between 0 and 1 in order to obtain a function which would look similar to a probability distribution function. Note that, as mentioned in section ''Less formal likelihood functions'', given the choice of the less formal likelihoods, the GLUE procedure provides a probability distribution function (not a probability density), which is then forced to be limited between 0 and 1 via re-normalization of the obtained values.

The derivation of the predictive probability distribution

In the GLUE literature (Beven and Freer, 2001; Beven et al., 2000), a predictive probability distribution is given. Unfortunately, even if the likelihood measure is a proper likelihood, it is not difficult to show that the predictive probability given in the GLUE literature does not represent the predictive probability. As clearly pointed out by Krzysztofowicz (1999), the predictive probability for the pth observation of interest, which is the probability of the observable value y_p (not the model forecast) to be greater or equal to a specific value y, can be obtained by marginalizing the effect of the model uncertainty by means of the posterior parameter distribution, to obtain:

$$f(y_p|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{x}_p) = \int_{\Theta} f(y_p|\theta, \mathbf{x}_p) g(\theta|\mathbf{Y}_n, \mathbf{X}_n)\, \mathrm{d}\theta, \tag{3}$$

which involves the need of estimating $P(y_p \leqslant y|\theta, \mathbf{x}_p)$ the probability of $y_p \leqslant y$ conditional to the model forecast (or hindcast). A similar expression is given if θ belongs to a discrete parametric space. Of course, the predictive probability distribution can be expressed for more than one next observation.

It must be acknowledged that an equation similar to Eq. (3), given the multi-uniform prior, has been introduced as the predictive probability estimate given in discrete form in a more recent paper (Eq. (2) in Romanowicz and Beven, 2003).

Bayesian incoherence and inconsistent learning of GLUE

The required properties of a Bayesian inference process

The most important knowledge element in the Bayesian learning process is the amount of information extracted from one or several experiments, which will be here called the ''value'' of the experiment(s). This ''value'' reflects the amount that a rational agent should be prepared to pay to acquire new data.

There are two basic requirements that guarantee the success of a Bayesian inference process: the first one is the non-decreasing value with the increasing number of experiments (for instance, the number of sampling observations); the second one is the equivalence of the experimental value in terms of the number of experiments, irrespective of the order in which the experiments have been carried out or whether the results are presented in sequential or batch form.

It will be shown that the inferential process based on the GLUE less formal likelihoods fails to guarantee these essential properties.

Consistency and coherence in learning

A natural requirement for a statistical inference process is the increasing value of experiment, with the increasing number n of sampling observations of the sampling experiment with results $(\mathbf{y}_n, \mathbf{X}_n)$.

In the Bayesian inferential approach, this property is expressed by the Bayesian consistency, that is, by the asymptotic properties of the posterior distribution, namely the convergence in probability of the sequence of random variables $\{\theta_n\}$ with probability distribution function $G_n(\theta|(\mathbf{y}_n, \mathbf{X}_n))$ (see Bernardo and Smith, 1994, 5.3 Asymptotic analysis, p. 285 and following, and bibliographic references). Assuming that the ''true model'', for instance, the parameter vector θ_0, is included in the set of models (parameters) taken into consideration, under regular conditions, and referring to the above-mentioned bibliography, the sequence of random variables $\{\theta_n\}$ converges in probability to θ_0. The convergence of the predictive probability density function $f(y_p|(\mathbf{y}_n, \mathbf{X}_n), \mathbf{x}_p)$ to the limiting conditional probability density function $f(y_p|\theta_0, \mathbf{x}_p)$ follows.

As it will be verified in the following sections, the use of the less formal likelihoods generally proposed and used in GLUE does not ensure the desired asymptotic consistency property. In order to demostrate it, the property of ''coherence'' for a parametric inference process will be introduced since an inferential process which is not coherent is also inevitably non-consistent.

The coherence property is here introduced to formalize the assumption that the experimental value of the experimental observable matrix $(\mathbf{y}_{n+m}, \mathbf{X}_{n+m})$ is higher than the included experimental observable matrix $(\mathbf{y}_n, \mathbf{X}_n)$ or $(\mathbf{y}_m, \mathbf{X}_m)$;

this is considered for each number n and $m (n \geqslant 1,\ m \geqslant 1)$, with $\mathbf{y}_{n+m} = (\mathbf{y}_n^{\mathrm{T}}, \mathbf{y}_m^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{X}_{n+m} = (\mathbf{X}_n^{\mathrm{T}}, \mathbf{X}_m^{\mathrm{T}})^{\mathrm{T}}$.

In order to make a complete analysis of the observed cases, we assume $n \geqslant 0$, $m \geqslant 0$. For either $n > 0$ or $m > 0$, without loss of generality, the following probability distribution functions (pdf) for the parameter vector θ are considered:

- prior: $g_0(\theta)$;
- posterior: $g_{n+m}(\theta|\mathbf{y}_{n+m}, \mathbf{X}_{n+m})$.

The marginal probability distribution function of the observable vector $\mathbf{y}_{n+m}$ (predictive probability distribution function) is

$$F(\mathbf{y}_{n+m}|\mathbf{X}_{n+m}) = E_{G_0(\theta)}\{F(\mathbf{y}_{n+m}|\theta, \mathbf{X}_{n+m})\},$$

in correspondence with $\mathbf{X}_{n+m}$, where $E_{G_0(\theta)}\{d(\theta)\}$ denote the expected value of the measurable function $d(\cdot)$ of the random vector θ, calculated by considering the prior probability distribution function $G_0(\theta)$, corresponding to the prior density (or probability) function $g_0(\theta)$.

Let $d[g_{n+m}(\cdot|\mathbf{y}_{n+m}, \mathbf{X}_{n+m}), g_0(\cdot)]$ be a *non-negative* real-valued and *convex* function, in the following said *discrepancy* between posterior and prior pdf, where $g_0(\cdot)$ is fixed and $g_{n+m}(\cdot|\mathbf{y}_{n+m}, \mathbf{X}_{n+m})$ is the argument variable of the function.

For instance, *Chi-square discrepancy* and *Kullback–Leibler divergence* between posterior and prior probability densities, as well as the *Frobenious norm* of the difference between the posterior and the prior probability density function may be considered (for the use of the Kullback–Leibler divergence in Bayesian statistics, see for instance Bernardo and Smith, 1994).

Let $d[g_{n+m}, g_0]$ be the expected discrepancy:

$$d[g_{n+m}, g_0] = E_{F(\mathbf{y}_{n+m}|\mathbf{x}_{n+m})}\{d[g_{n+m}(\cdot|\mathbf{y}_{n+m}, \mathbf{X}_{n+m}), g_0(\cdot)]\}$$

that defines and measures the *value of the experiment* corresponding to the observable vector $\mathbf{y}_{n+m}$.

$E_{F(\mathbf{y}_{n+m}|\mathbf{x}_{n+m})}\{d(\mathbf{y}_{n+m})\}$ denote the expected value of the measurable function $d(\cdot)$ of the random vector $\mathbf{y}_{n+m}$, calculated by considering the marginal probability distribution function $F(\mathbf{y}_{n+m}|\mathbf{X}_{n+m})$.

Of course, we assume the conditions of regularity which ensure the finiteness of the above mean value.

The following definition of coherence in learning for a statistical process of inference is proposed:

Definition 1. A statistical process of inference on parameter vector θ is defined as coherent in learning if

$$d[g_{n+m}, g_0] > \max\{d[g_n, g_0], d[g_m, g_0]\}, \tag{4}$$

for all $n \geqslant 1$, $m \geqslant 1$.

The condition of coherence in learning corresponds to the request of the experimental value $d(n) = d[g_n, g_0]$ to be *increasing* monotone in n, for $n \geqslant 1$. It should be noted that the expected discrepancy $d[g_n, g_0]$ measures the value of the experiment concerning the observable vector $\mathbf{y}_n$, considering the observable vector $\mathbf{y}_n$ subordinate to the matrix $\mathbf{X}_n$.

The parametric inference processes that violate the condition of coherence in learning are called incoherent.

Now the sampling result $\mathbf{y}_n$ might be understood as the union of results $(\mathbf{y}_n, \mathbf{y}_m)$, as $\mathbf{y}_m$ varies among all possible results. Thus the results $(\mathbf{y}_n, \mathbf{y}_m)$ define a partition of $\mathbf{y}_n$. The coherence in learning still requires increasing sampling information as the partition of $\mathbf{y}_n$ is refined.

It can be proved that

Proposition 1. *The Bayesian process of inference on parameter vector θ with a proper likelihood function is coherent in learning according to the above definition.*

Proposition 2. *The Bayesian process of inference on parameter vector θ with the GLUE less formal likelihood functions is incoherent in learning according to the above definition.*

Predictive coherence may also be considered, in this case predictive probability density functions (or predictive probability functions) obtained before and after sampling are compared. It can be shown that Bayesian process of parametric inference with proper likelihood is coherent also in prediction. Equivalence between learning and predictive coherence for Bayesian process of parametric inference can be proved. Coherence is necessary for consistency.

The proof of the above results can be found in Mantovan and Todini, 2004.

In section ''A simulated experiment'' Propositions 1 and 2 are verified on the basis of a simulated experiment.

Equivalence between batch and sequential learning

In the Bayesian inference process, the equivalence between batch, sequential learning as well as the sequential order descends from the basic definition and the symmetrical nature of the proper likelihood

$$\begin{aligned}\mathscr{L}(\theta; \mathbf{y}_{m+n}, \mathbf{X}_{m+n}) &= f((\mathbf{y}_{m+n}, \mathbf{X}_{m+n})|\theta) = f((\mathbf{y}_m, \mathbf{X}_m), (\mathbf{y}_n, \mathbf{X}_n)|\theta) \\ &= f((\mathbf{y}_m, \mathbf{X}_m)|\theta, (\mathbf{y}_n, \mathbf{X}_n)) f((\mathbf{y}_n, \mathbf{X}_n)|\theta) \\ &= f((\mathbf{y}_n, \mathbf{X}_n)|\theta, (\mathbf{y}_m, \mathbf{X}_m)) f((\mathbf{y}_m, \mathbf{X}_m)|\theta).\end{aligned}$$

If this property is not guaranteed, there is no uniqueness and therefore equivalence of the experimental value for a given number of experiments. The less formal likelihoods used in GLUE violate this property, which makes the proposed Bayesian updating (see for example Eq. (3) in Romanowicz and Beven, 2003) non-unique and consequently of dubious validity.

It is not difficult to acknowledge that when using the less formal likelihoods, the batch and the sequential learning are not equivalent for samples of dimensions n and m, respectively. This property is here demonstrated, without loss of generality, for the case of independent samples.

For $\mathscr{L}_1$, the Nash and Sutcliffe criterion is based on the less formal likelihood, and assuming without loss of generality that $\mathscr{L}_1 > 0$, $\forall \theta \in \Theta$, $N \geqslant 1$,

$$\mathscr{L}_1(\theta; \mathbf{y}_{n+m}, \mathbf{X}_{n+m}) \neq \mathscr{L}_1(\theta; \mathbf{y}_n, \mathbf{X}_n)\mathscr{L}_1(\theta; \mathbf{y}_m, \mathbf{X}_m)$$

given that

$$[1 - s^2_{n+m}(\theta)/s^2_{n+m}]^N \neq [1 - s^2_n(\theta)/s^2_n]^N [1 - s^2_m(\theta)/s^2_m]^N.$$

For $\mathscr{L}_2$, the inverse error variance is based on the less formal likelihood:

$$\mathscr{L}_2(\boldsymbol{\theta};\mathbf{y}_{n+m},\mathbf{X}_{n+m}) \neq \mathscr{L}_2(\boldsymbol{\theta};\mathbf{y}_n,\mathbf{X}_n)\mathscr{L}_2(\boldsymbol{\theta};\mathbf{y}_m,\mathbf{X}_m)$$

given that

$$[s^2_{n+m}(\boldsymbol{\theta})]^{-N} \neq [s^2_n(\boldsymbol{\theta})]^{-N}[s^2_m(\boldsymbol{\theta})]^{-N}.$$

For $\mathscr{L}_3$, the exponential transformation of error variance is based on the less formal likelihood:

$$\mathscr{L}_3(\boldsymbol{\theta};\mathbf{y}_{n+m},\mathbf{X}_{n+m}) \neq \mathscr{L}_3(\boldsymbol{\theta};\mathbf{y}_n,\mathbf{X}_n)\mathscr{L}_3(\boldsymbol{\theta};\mathbf{y}_m,\mathbf{X}_m)$$

given that

$$\exp\{-Ns^2_{n+m}(\boldsymbol{\theta})\} \neq \exp\{-Ns^2_n(\boldsymbol{\theta})\}\exp\{-Ns^2_m(\boldsymbol{\theta})\}.$$

The non-equivalence of batch and sequential learning is also acknowledged in Beven (2001, p. 253), where the following statement is made:

> ''... It has the disadvantage that if a period of calibration is broken down into smaller periods, and the likelihoods are evaluated and updated for each period in turn, for many likelihood measures the final likelihoods will be different from using a single evaluation for the whole period.''

Unfortunately, this is not a technical disadvantage, but a basic requirement to an inference process.

A simulated experiment

The *abc* hydrological model

The *abc* model presented by Fiering (1967) primarily for didactic purposes was deemed useful to show the consequences of the incoherence and inconsistency of the GLUE statistical inferential processes based on less formal likelihoods. The schematical description of the model structure is given in Fig. 1. The extremely simplified model is based on a function which partitions rainfall into evapotranspiration, surface runoff and aquifer recharge. The aquifer is represented by a simple linear storage directly connected to the stream.

Two basic state equations can be formulated to represent the *abc* model:

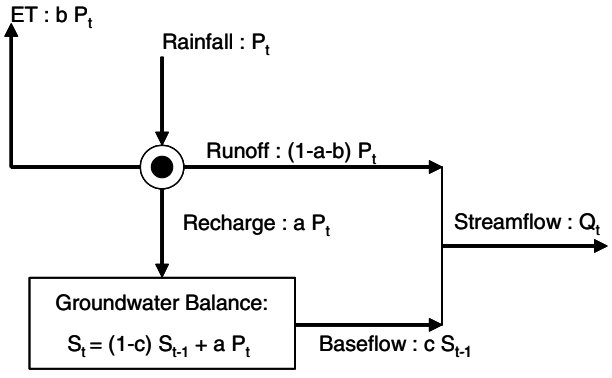

Figure 1 Schematic representation of the *abc* model.

$$\begin{cases} Q_t = (1-a-b)P_t + cS_{t-1}, \\ S_t = (1-c)S_{t-1} + aP_t, \end{cases}$$

where P_t is the rainfall input, Q_t is the output discharge and S_t the aquifer storage at any given time t.

There are three model parameters: a, b and c on top of an additional parameter, S_0, representing the initial aquifer storage condition. The model, whose parameters are poorly identified when calibrated to streamflow data only, was used in a previous work (Kuczera and Parent, 1998) to assess parameter uncertainty and the merit of the Metropolis algorithm for sampling from the parameter posterior distribution.

The test case

In this study, following the previous work of Kuczera and Parent (1998), the same *abc* ''true'' model parameter values were used (Table 1) and, only for the sake of clarity but without loss of generality, the posterior parameter distribution was derived for parameters a and b, where a is the recharge coefficient and b is the evapotranspiration coefficient, while taking parameters c and S_0 as known with their value fixed at the ''true'' value, namely $c = 0.05$ and $S_0 = 500$.

In order to preserve some realistic features of rainfall, the input to the *abc* model was generated, without lack of generality, using a correlated Poisson–Gamma autocorrelated process, while the corresponding output values, the streamflow discharges, were obtained by running the *abc* model using the true values of the parameters given in Table 1.

So far, no measurement or structural errors were introduced. Therefore, in order to take into account the sources of uncertainty, under the assumption that parameters c and S_0 are known, the *abc* model equations were rewritten in the following exponentially weighted (smoothed) process:

$$Q_t = (1-a-b)P_t + a\sum_{k=1}^{t-1}(1-c)^{t-k-1}P_k + c(1-c)^{t-1}S_0,$$

$$t = 1,2,\ldots,n,$$

which can be thought of as the following linear regression model:

$$[Q^o_t - c(1-c)^{t-1}S_0] = \left[(1-a-b)P_t + a\sum_{k=1}^{t-1}(1-c)^{t-k-1}P_k\right] + \varepsilon_t,$$

$$t = 1,2,\ldots,n.$$

Without loss of generality, the structural part

$$\left[(1-a-b)P_t + a\sum_{k=1}^{t-1}(1-c)^{t-k-1}P_k\right]$$

Table 1 The *abc* model parameters

Parameter	True value
a	0.10
b	0.75
c	0.05
S_0	500

is combined with the measurement error ε_t in additive way. In the following, ε_t was generated according to an autocorrelated *AR*(1) Gaussian error model:

$$\varepsilon_t = \alpha\varepsilon_{t-1} + \eta_t, \quad t = 1, 2, \ldots,$$

with $\alpha = 0.8$; $\varepsilon_0 \triangleq \text{Normal}(0, \delta^2)$, with $\delta^2 = 8$; and $\eta_t \triangleq \text{Normal}(0, \sigma^2)$, with $\sigma^2 = 8$ and $E\{\eta_t\eta_j\} = 0$, $\forall t \neq j$, $\forall j \geqslant 1$; $E\{\eta_t\varepsilon_0\} = 0$, $\forall\ t \geqslant 1$.

The generation of the additive noise allowed computation of the ''observed'' streamflow discharges Q_t^o, $t = 1, 2, \ldots, n$.

For the error ε_t given by the above *AR*(1) model, as t diverges, we just have:

$$\text{Var}(\varepsilon_t) = \sigma^2/(1-\alpha^2);$$
$$\text{Corr}(\varepsilon_t, \varepsilon_{t+k}) = \alpha^k, \quad k = 1, 2, \ldots$$

The autocorrelated Gaussian error model is usually adopted when the r.v. y_i, $i = 1, 2, \ldots, n$, is observable in subsequent equispatial time instants.

This model has also been widely used in hydrological studies (see, for instance, Romanowicz et al., 1994, 1996).

In terms of the choice of the parameters prior to distribution, following the GLUE procedure, a uniform prior probability distribution on a discrete space for parameters (a,b) is taken. Bearing in mind the continuity of mass constraint, which requires that $a + b \leqslant 1$, the following discrete space for parameters (a, b) was considered:

$$\Theta = \{(a_i, b_j) : a_i = i/m, i = 1, 2, \ldots, m;$$
$$b_j = j/m, j = 1, 2, \ldots, (m-i)/m\}; \quad m = 100.$$

The proper likelihood function, in this case, given the autocorrelated *AR*(1) Gaussian error model, is the following:

$$\mathscr{L}(a_i, b_j) \triangleq \frac{\exp\left(-\frac{1}{2}\left[\mathbf{Q}_n^o - \boldsymbol{\mu}(a_i, b_j, \mathbf{P}_n)\right]'\boldsymbol{\Sigma}^{-1}\left[\mathbf{Q}_n^o - \boldsymbol{\mu}(a_i, b_j, \mathbf{P}_n)\right]\right)}{(2\pi)^{n/2}[\det(\boldsymbol{\Sigma})]^{1/2}},$$

where $\mathbf{Q}_n^o = (Q_1^o, Q_2^o, \ldots, Q_n^o)'$; $\mathbf{P}_n = (P_1, P_2, \ldots, P_n)'$; $\boldsymbol{\mu}(a_i, b_j, \mathbf{P}_n) = (\mu_1(a_i, b_j, \mathbf{P}_1), \mu_2(a_i, b_j, \mathbf{P}_2), \ldots, \mu_n(a_i, b_j, \mathbf{P}_n))'$; $\mu_t(a_i, b_j, \mathbf{P}_t) = [(1-a-b)P_t + a\sum_{k=1}^{t-1}(1-c)^{t-k-1}P_k] + c(1-c)^{t-1}S_0$; $i = 1, 100$; $j = 1, 100$; $\boldsymbol{\Sigma} = [\sigma^2/(1-\alpha^2)]R(\alpha)$, $0 < |\alpha| < 1$, $\sigma^2 > 0$; where $R(\alpha)$ is the so-called first order Markov matrix (see Basilevsky, 1983, p. 221):

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \ldots & \alpha^{n-2} & \alpha^{n-1} \\ \alpha & 1 & \alpha & \ldots & \alpha^{n-3} & \alpha^{n-2} \\ \alpha^2 & \alpha & 1 & \ldots & \alpha^{n-4} & \alpha^{n-3} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha^{n-2} & \alpha^{n-3} & \alpha^{n-4} & \ldots & 1 & \alpha \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \ldots & \alpha & 1 \end{bmatrix}.$$

The inverse matrix of $\boldsymbol{\Sigma}$ is a symmetric Jacobian matrix which can readily be computed (see Basilevsky, 1983, p. 222).

The posterior probability function is then updated as the sample size increases both using the exact likelihood given and with the less formal likelihoods proposed in GLUE.

Therefore, the sample size will be equal to $n(i) = 50 \cdot i$, where i is the number of sample periods taken into account to determine the posterior probability function, $i = 1, 2, \ldots, 10$.

The sample size on which the experiment was carried out is therefore equal to $n(i) = 50 \cdot i$, where $i = 1, 2, \ldots, 10$ is the number of sample extensions successively taken into account to determine the posterior probability function.

Results of the experiment

A simple comparison

A sample time series, of length $n + m = 200$ of rainfall inputs was generated together with the corresponding observable streamflow discharges, given $s_0 = 500$, $c = 0.05$, $a = 0.10$, $b = 0.75$. The posterior probability function is then obtained considering the first $n = 100$ observations and then updated considering the first observations jointly with the next $m = 100$ observations.

Figure 2(a) and (b) shows the profile of the posterior probability function (posterior pf), when the Nash and Sutcliffe (N—S) efficiency criterion is used, at sample sizes $n = 100$ in Figure 2(a) and $n + m = 200$ in Figure 2(b). Of course, the shapes of the two posterior pf are different, but no significant change is shown with the increased sample dimension. Increased fixed values of sample size change the shape of the posterior probability function, but they do not lead to consistency in the Bayesian sense. Similar results are obtained for the other two less formal likelihoods (inverse error variance and exponential transformation of error variance likelihood).

Figure 2(c) and (d) shows the profile of the posterior probability function, when the exact likelihood (EL) is used, at the same two sample sizes $n = 100$ in Figure 2(c) and $n + m = 200$ in Figure 2(d). Coherent significant changes are shown with the increased sample dimension. Increased fixed values of sample size change the shape of the posterior probability function and they lead to consistency.

It is worthwile mentioning the fact that the scope of this paper is not to demonstrate that when using the exact likelihood in the Bayesian inferential process, the posterior probability function is consistent. This is a well-known result. The scope of this paper is to demonstrate that the consistency property does not occur with the GLUE proposed and widely used less formal likelihood.

Since the (N—S) posterior probability is very flat, different unit measure scales are used for the geometrical representation of the posterior probabilities. To make the comparison between the two kinds of posterior probability functions more easy, in Figure 3(a)—(d) the posterior probability contour lines are plotted. Of course, the outlined inconsistency of a Bayesian inference process with a less formal likelihood is highlighted. It can be noted that the (N—S) posterior probability with larger sample size $n + m = 200$ is more flat than the other one with sample size $n = 100$. This is strange but possible with the three less formal likelihoods considered.

Simulations and value of the experiments

For further illustration, 100 independent sample time series, each of length $N = 500$ of rainfall, were generated together with the corresponding observable streamflow discharges, given $s_0 = 500$, $c = 0.05$, $a = 0.10$, $b = 0.75$.

For each time series pair, the posterior probability function was successively updated at the sample size $n(i) = 50 \cdot i$,

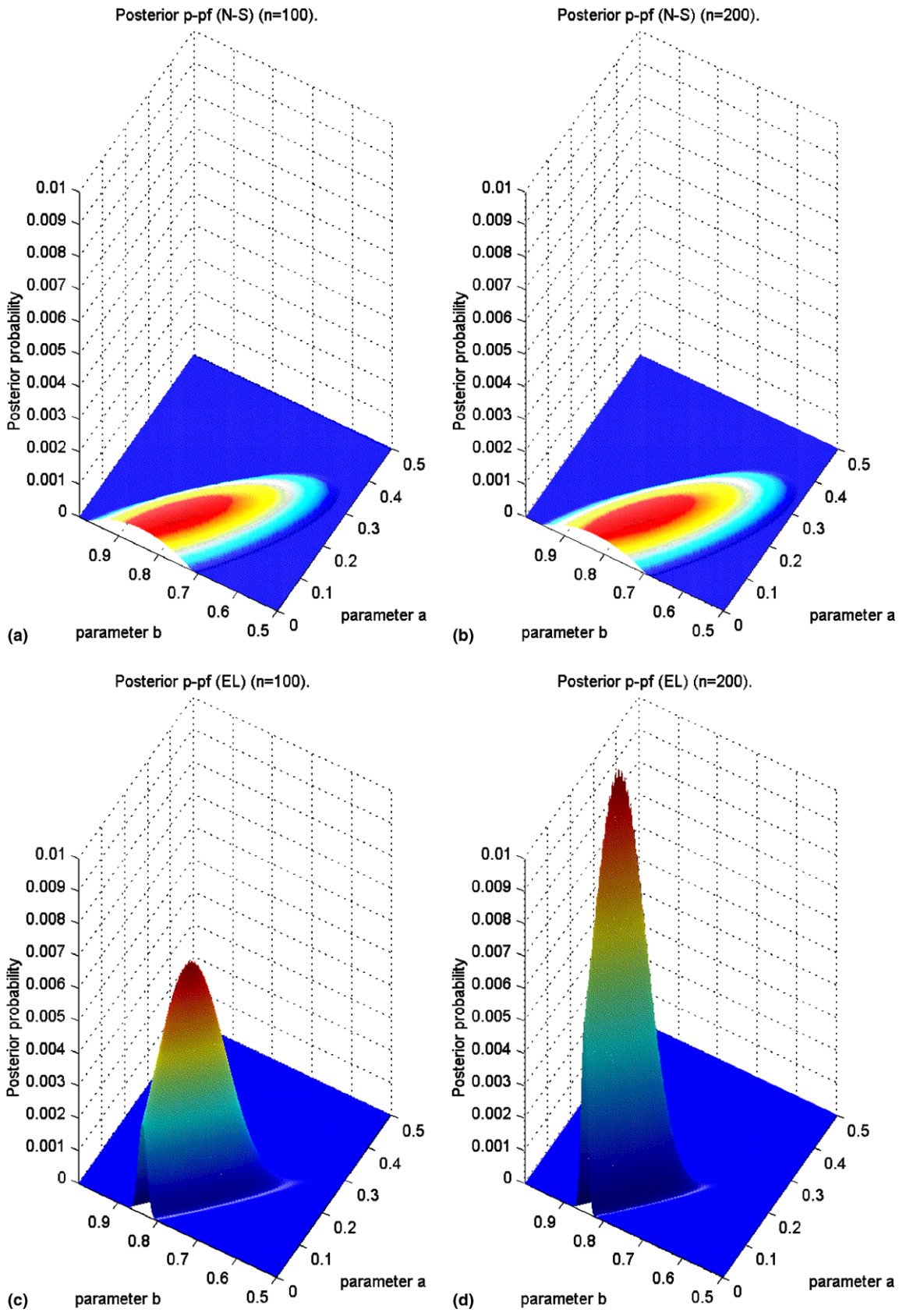

Figure 2 Posterior probability profile — (N—S): (a) $n = 100$ and (b) $n + m = 200$; (EL): (c) $n = 100$ and (d) $n + m = 200$.

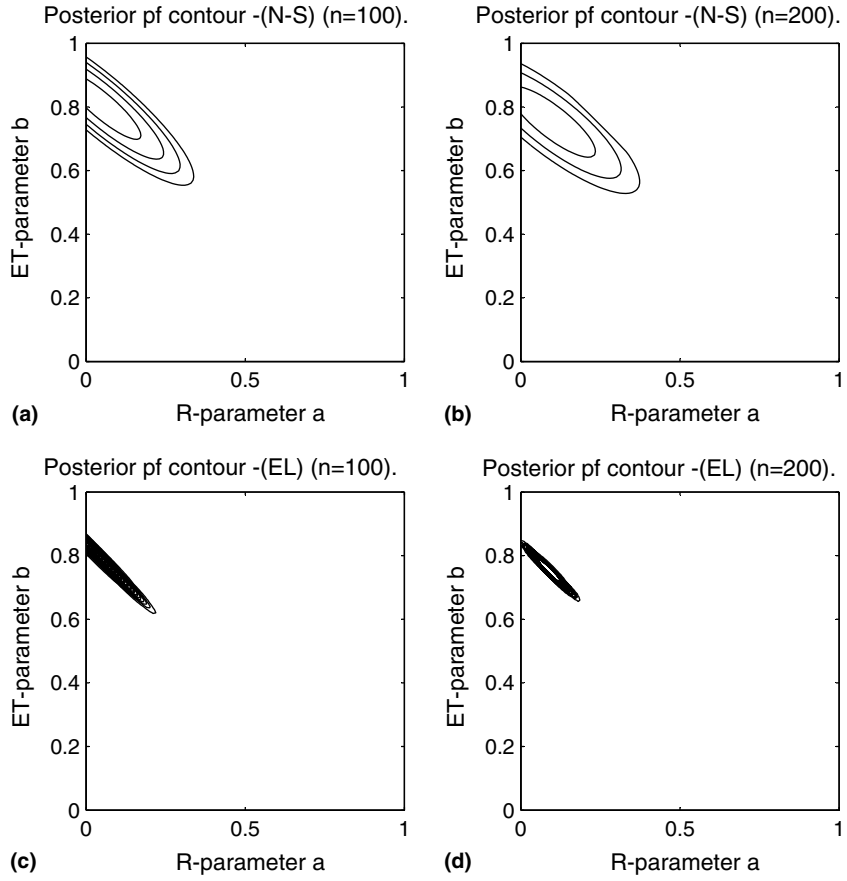

Figure 3 Posterior pf contour — (N—S): (a) $n = 100$ and (b) $n + m = 200$; (EL): (c) $n = 100$ and (d) $n + m = 200$.

where $i = 1, 2, \ldots, 10$ is the number of sample extensions successively taken into account to determine the posterior probability function.

The inconsistency of a Bayesian inference process with a less formal likelihood can still be shown by the discrepancy function as a specified measure of the value of the experiment.

At every update step i, given the posterior probability function $p(\boldsymbol{\theta}|\mathbf{y}_{n(i)}, \mathbf{X}_{n(i)})$, the Euclidean distance between the posterior probability function $p(\boldsymbol{\theta}|\mathbf{y}_{n(i)}, \mathbf{X}_{n(i)})$ and the prior one $p_0(\boldsymbol{\theta})$:

$$d(n(i)) = \left[\sum_{\boldsymbol{\theta}\in\Theta}(p(\boldsymbol{\theta}|\mathbf{y}_{n(i)}, \mathbf{X}_{n(i)}) - p_0(\boldsymbol{\theta}))^2\right]^{1/2},$$

was computed, $i = 1, 2, \ldots, 10$. Then the 5th, 50th and 95th percentiles of the distance $d(n(i))$ were determined on the basis of the 100 independent sample time series generated, at the step of updating $i = 1, 2, \ldots, 10$.

Fig. 4(a) shows the 5th, 50th and 95th percentiles of the distance between the posterior probability function and the prior one as the number of sample extensions taken into account to determine the posterior probability distribution increases, when the Nash and Sutcliffe efficiency criterion is used (Percentiles of the posterior—prior (p—p) distance). As one can see, no increasing value of the experiment, estimated by the mean of the distribution with the above percentiles, is observed for this less formal likelihood. Similar results are obtained for the other two less formal likelihoods. Of course, similar results are obtained with the Chi-square discrepancy and the Kullback—Leibler divergence.

Fig. 4(b) shows the 5th, 50th and 95th percentiles of the distance between the posterior probability function and the prior one as the sample size increases, with $i = 1, 2, \ldots, 10$, when the exact likelihood is used. Figure 4(b) shows increasing experimental values estimated by the mean of the distribution with the above percentiles, as the sampling dimension of the posterior probability updating grows.

Total mean squared error loss

Moreover, since the model generating the data is known, at every update step i, given the posterior probability function $p_{n(i)}(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y}_{n(i)}, \mathbf{X}_{n(i)})$, the posterior total mean squared error (MSE) loss was computed:

$$E_{P_{n(i)}(\boldsymbol{\theta})}\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\} = \sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathsf{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)p(\boldsymbol{\theta}|\mathbf{y}_{n(i)}, \mathbf{X}_{n(i)}),$$

with $\boldsymbol{\theta}^{\mathsf{T}} = (a, b)$, $\boldsymbol{\theta}_0^{\mathsf{T}} = (a_0, b_0) = (0.10, 0.75)$, $i = 1, 2, \ldots, 10$.

Then the 5th, 50th and 95th percentiles of the posterior total mean squared error loss $E_{P_{n(i)}(\boldsymbol{\theta})}\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\}$ were determined on the basis of the 100 independent sample time series generated, at the step of updating $i = 1, 2, \ldots, 10$.

Fig. 5(a) shows the 5th, 50th and 95th percentiles of the posterior total MSE loss $E_{P_{n(i)}(\boldsymbol{\theta})}\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\}$, for

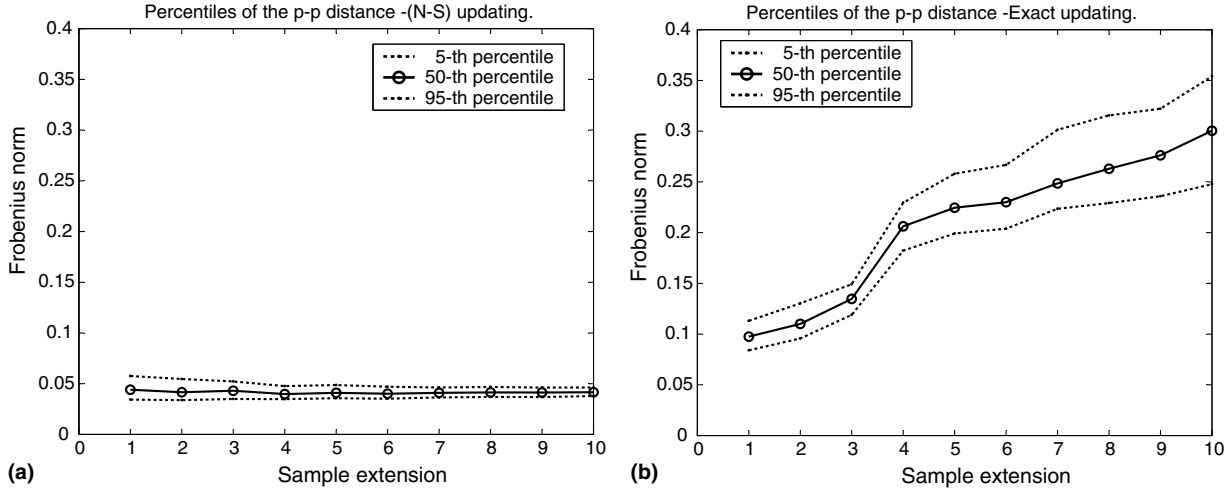

Figure 4 Percentiles of the p—p distance — (a) N—S updating and (b) exact updating.

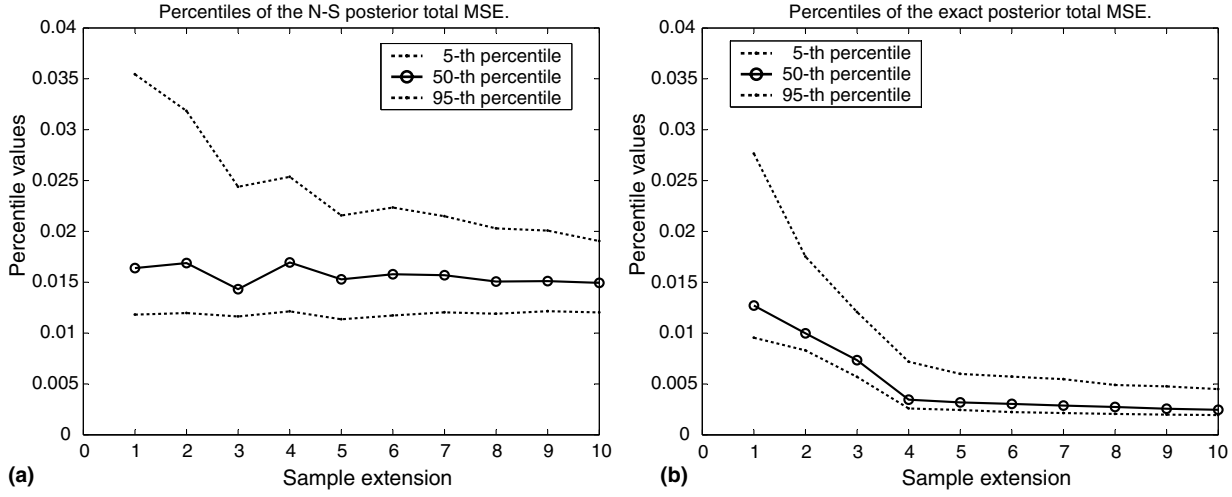

Figure 5 Percentiles of the N—S posterior total MSE (a) and exact posterior total MSE (b).

$i = 1, 2, \ldots, 10$, when the Nash and Sutcliffe efficiency criterion is used. As it can be seen, no consistent process is reflected by this less formal likelihood. Similar results are obtained for the other two less formal likelihoods.

Fig. 5(b) shows the 5th, 50th and 95th percentiles of the posterior total MSE loss $E_{P_{n(i)}(\theta)}\{\|\theta - \theta_0\|^2\}$, for $i = 1, 2, \ldots, 10$, when the exact likelihood is used. Figure 5(b) shows decreasing total MSE loss values as the sampling size of the posterior probability updating grows.

Figs. 2—5 illustrate coherence and consistency of the inference process with appropriate likelihood as well as incoherence and inconsistency of the inference process with a less formal likelihood. The results do not depend on the complexity of the structural part of the additive error model considered or on the model used to generate predictors. Similar results are obtained changing the hyperparameters $(\alpha, \delta^2, \sigma^2)$ of the model.

Conclusions and recommendations

The GLUE methodology can be considered as a simplified Bayesian inferential procedure conducted using likelihood functions which are derived in a non-probabilistic way. This methodology is attractive because there is no need for detailed distribution functions of the observable variables and /or of errors even in complex situations, such as when there is more than one source of errors; when the explicative models provided are complex; when the number of parameters used to realize a process of learning is high.

On the other hand, if a less formal likelihood function is adopted, some fundamental properties are lost in the statistical process of parametric inference. Such properties are the coherence of the inference process and the equivalence between sequential inference process and block inference process (which can be easily demonstrated when conditional independence is assumed). The coherence property is directly connected with Bayesian consistency: incoherence entails inconsistency. Consistency is a required fundamental property for an inference process. In addition, the loss of the equivalence between batch and sequential learning not only violates the principle of compound probability, which is the basic principle on which the whole Bayesian learning process is founded, but also it dispenses with the capacity to run statistical inference sequentially, which is very useful for real time forecasting, where the probability distribution functions are updated step by step with the

newly available measurement of the observable random variables.

Even if an appropriate likelihood function is difficult to define, likelihood functions corresponding to an incoherent inferential process are not acceptable. The analysis of residuals together with model assessment and selection on the basis of its prediction capability may allow to chose an appropriate likelihood function and model. The predictive performance may be useful to evaluate the quality of a selected model.

Indeed, other solutions can be found within a non-parametric inferential approach and/or a robust statistic inferential approach. A relatively robust approach (although time correlation of residuals is neglected) can for instance be developed by transforming data into a Gaussian space via the normal quantile transform (Van der Waerden, 1952, 1953; Kelly and Krzysztofowicz, 1997; Liu et al., 2005).

References

Basilevsky, A., 1983. Applied Matrix Algebra in the Statistical Sciences. North-Holland, New York.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. John Wiley & Sons, Chichester.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16, 41–51.

Beven, K.J., 2001. Rainfall-Runoff Modelling: The Primer. John Wiley & Sons, Ltd, Chichester, 360 pp.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Proces. 6, 279–298.

Beven, J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249 (1–4), 11–29.

Beven, K.J., Freer, J., Hankin, B., Schulz, K., 2000. The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems. In: Fitzgerald, W.J. et al. (Eds.), Nonlinear and Nonstationary Signal Processing. Cambridge Univ. Press, New York, pp. 115–151.

Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods. Water Resour. Res 36, 3663–3674.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference Statistical Analysis. Addison-Wesley-Longman, Reading, MA.

de Finetti, B., 1975Theory of Probability, vol. 2. John Wiley & Sons, London.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143–10162.

Fiering, M.B., 1967. Streamflow Synthesis. Harvard University Press, Cambridge, MA.

Franks, S.W., Beven, K.J., 1997. Bayesian estimation of uncertainty in land surface–atmosphere flux predictions. J. Geophys. Res. 102 (D20), 23991–23999.

Freer, J., Beven, K.J., Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resour. Res. 32, 2161–2173.

Gelb, A. (Ed.), 1974. Applied Optimal Estimation. MIT Press.

Gupta, V.K., Sorooshian, S., 1983. Uniqueness and observability of conceptual rainfall-runoff model parameters: the percolation process examined. Water Resour. Res. 19 (1), 269–276.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrogic models: multiple and non-commensurable measures of information. Water Resour. Res. 34, 751–763.

Kelly, K.S., Krzysztofowicz, R., 1997. A bivariate meta-Gaussian density for use in hydrology. Stochast. Hydrol. Hydraul. 11, 17–31.

Kavetski, D., Franks, S.W., Kuczera, G., 2003. Confronting input uncertainty in environmental modelling. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibratoin of watershed models, Water Science and Application 6. Amer. Geophys. Union, Washington DC, pp. 48–68.

Keesman, K.J., 1990. Set theoretic parameter estimation using random scanning and principal component analysis. Math. Comput. Simul. 32, 535–543.

Klepper, O., Scholten, H., van de Kamer, J.P.G., 1991. Prediction uncertainty in an ecological model of the Oosterschelde Estuary. J. Forecast. 10, 191–209.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res. 35 (9), 2739–2750.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. J. Hydrol. 211, 69–85.

Liu, Z., Martina, M.V.L., Todini, E., 2005. Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment. Hydrol. Earth Syst. Sci. (HESS) 9, 347–364.

Mantovan P., Todini E., 2004. Coherence of the statistical inference process in hydrological forecasting. Technical Report, University of Venice, Venice. Available from <www.dst.unive.it/~mantovan>.

Romanowicz, R., Beven, K., 2003. Estimation of flood inundation probabilities as conditioned on event inundation maps. Water Resour. Res. 39 (3), 1073. doi:10.1029/2001WR001056.

Romanowicz, R., Beven, K.J., Tawn, J., 1994. Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach. In: Barnett, V., Turkman, K.F. (Eds.), Statistics for the Environment 2: Water-Related Issues. John Wiley, New York, pp. 297–315.

Romanowicz, R., Beven, K.J., Tawn, J., 1996. Bayesian calibration of flood inundation models. In: Anderson, M.G., Walling, D.E., Bates, P.D. (Eds.), Floodplain Processes. John Wiley, New York, pp. 333–360.

Sorooshian, S., Gupta, V.K., 1983. Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. Water Resour. Res. 19 (1), 260–268.

Thiemann, M. et al, 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resour. Res. 37 (10), 2521–2535.

Van der Waerden, B.L., 1952. Order tests for two sample problem and their power. Indagat. Math. 14, 253–458.

Van der Waerden, B.L., 1953. Order tests for two sample problem and their power. Indagat. Math. 14, 303–316.

van Straten, G., Keesman, K.J., 1991. Uncertainty propagation and speculation in projective forecasts of environmental change: a lake eutrophication example. J. Forecast. 10, 163–190.

Vrugt, J.A., Weerts, A.H., Bouten, W., 2001. Information content of data for identifying soil hydraulic parameters from outflow experiments. Soil Sci. Soc. Am. J. 65, 19–27.

Vrugt, J.A., Bouten, W., Gupta, H.V., Sorooshian, S., 2002. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. Water Resour. Res. 38 (12), 1312. doi:10.1029/2001WR001118.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003a. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour. Res. 39 (8), 1201. doi:10.1029/2002WR001642.

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

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resour. Res. 41, W01017. doi:10.1029/2004WR003059.

Vrugt, J.A., 2004. Towards improved treatment of parameter uncertainty in hydrologic modeling. Ph.D. Thesis of University of Amsterdam, the Netherlands, ISBN:90-76894-46-9.

Wagener, T., McIntyire, N., Lees, M.J., Wheather, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrol. Proc. 17, 455—476.

Yapo, P.O., Gupta, H.V., Soroosian, S., 1998. Multi-objective global optimization for hydrologic models. J. Hydrol. 204, 83—97.

Young, P.C., 2001. Data-based mechanistic modeling and validation of rainfall-flow processes. In: Anderson, M.G., Bates, P.G. (Eds.), Model Validation. John Wiley & Sons Ltd, Chichester, UK, pp. 117—161.

Zak, S.K. et al, 1997. Uncertainty in the estimation of critical loads: a practical methodology. Soil Water Air Pollut. 98, 297—316.