COMMENTS ON GLUE

Renata Romanowicz and Keith Beven

Lancaster University, Lancaster, LA1 4YQ, United Kingdom
E-mail: R.Romanowicz@lancaster.ac.uk

Abstract: The paper presents an application of GLUE (Generalised Likelihood Uncertainty Estimation) methodology to the problem of estimating the uncertainty of predictions produced by environmental models. The methodology is placed in a wider context of different approaches to inverse modelling and, in particular, a comparison is made with Bayesian estimation techniques based on explicit structural assumptions about model error. Using a simple example of a rainfall-flow model, different evaluation measures and their influence on the prediction uncertainty and confidence limits are demonstrated.

Keywords: Bayesian inference, Deterministic environmental models, Generalised Likelihood Uncertainty Estimation (GLUE), Model determination.

1. INTRODUCTION

The Generalised Likelihood Uncertainty Estimation (GLUE) technique [1] was introduced partly to allow for the possible equifinality (non-uniqueness, ambiguity or non-identifiability) of parameter sets during the estimation of model parameters in over-parameterised models. The technique has been applied to a variety of environmental problems [2]. Its popularity results from the very few assumptions that it requires and the simplicity of the approach when used in practical applications. GLUE assumes that, in the case of large over-parameterised models, there is no inverse solution and, hence, that the estimation of a unique set of parameters, which optimise goodness-of fit-criteria given the observations, is not possible. The technique is based on the estimation of the weights or probabilities associated with different parameter sets, based on the use of a subjective likelihood measure to derive a posterior probability function, which is subsequently used to derive the predictive probability of the output variables. In [3] a statistically motivated, more formal equivalent of GLUE was developed. The idea was to explicitly derive the likelihood function based on the error between the observed outputs and those simulated by the model. This formal approach is equivalent to Bayesian statistical estimation: it requires assumptions about the statistical structure of the errors. GLUE is usually applied by directly likelihood weighting the outputs of multiple model realisations (either deterministic or stochastic, defined by sets of parameter values within one or more model structures) to form a predictive distribution of a variable of interest. Prediction uncertainties are then related to variation in model outputs, without necessarily adding an additional explicit error component. There is thus an interesting question as to whether an appropriate choice of likelihood measure can result in similar results from the two approaches.

There are a number of possible measures of model performance that can be used in this kind of analysis. The only formal requirements for use in a GLUE analysis are that the likelihood measure should increase monotonously with increasing performance and be zero for models considered as unacceptable or non-behavioural. Application-oriented measures are easily used in this framework. Measures based on formal statistical assumptions, when applied to all model realisations (rather than simply in the region of an “optimal” model)

should give results similar to a Bayesian approach when used within a GLUE framework [3], but the assumptions made (additive Gaussian errors in the simplest cases) are not always easily justified in the case of nonlinear environmental models with poorly known boundary conditions (see the discussion in [4]). In this paper, we shall explore the influence of the choice of observation based likelihood weights on the predictive uncertainty of the model.

2. RELATION BETWEEN GLUE AND STATISTICAL APPROACHES: DISCUSSION OF LIKELIHOOD MEASURES

There is a question as to how far GLUE can be consistent with formal statistical approaches since it weights the predictions of different feasible models without necessarily using an explicit error model, albeit that the likelihood weights are determined from a calculation that depends in some way on the model errors. In a Bayesian framework, the inverse problem is usually posed in the context of the observation equation:

$$y_t = g(u_0,...,u_t,\xi_0,...,\xi_t,\zeta_t,\theta,t) \tag{1}$$

or assuming an additive error model:

$$y_t = g(u_0,...,u_t,\xi_0,...,\xi_t,\theta,t) + \zeta_t \tag{2}$$

where y_t denotes the observed model output; u_t is model input, θ denotes vector of model parameters and the errors on the inputs, ξ_t, and the measurement error ζ_t are not known, and, in the general case, may be non-Gaussian (and might indeed be required to compensate for model structural error). The Bayesian approach allows detailed studies to be targeted at the most informative areas and variables (e.g. [5]). It also allows parameter and error estimates to be updated as more observations become available: for example, in the case of Gaussian normal assumptions (see e.g. [6]), the Bayesian estimator of the state variables in a linear, stochastic, dynamic system is the famous *Kalman Filter* (KF: [7]); while the Bayesian estimator of the constant or stochastic, time variable parameters in a linear-in-the-parameters regression model is the classic recursive least squares estimation algorithm. Environmental models are not normally that simple, and it might be difficult to formulate an appropriate error model. In that case there might be advantage in the use of a nonparametric representation of the errors ([8], [9], [10]).

Measures of goodness of fit applied in order to compare different nonparametric probability density functions (or log likelihood ratios) include:

(i) Mean Square Error (MSE),

$$MSE = E[\hat{f}(x) - f(x)]^2 = Var\{\hat{f}(x)\} - [E[\hat{f}(x)] - f(x)]^2, \tag{3}$$

which combines both variance and bias of the estimates;

(ii) averaged Predictive Squared Error (PSE), related to MSE:

$$PSE = E[\hat{f}(x) + \varepsilon - f(x)]^2 = MSE + \sigma^2; \tag{4}$$

where σ^2 denotes the prediction error variance; and

(iii) Mean Integrated Square Error (MISE):

$$MISE = E[\sum_{i=1}^{N} (\hat{f}(x_i) - f(x_i))^2]. \tag{5}$$

Instead of L_2 norm used in the above definitions, we can use L_∞ norm: $\sup_x |\hat{f}(x) - f(x)|$; L_1 norm: $\sum_{i=1}^{N} |\hat{f}(x_i) - f(x_i)|$ or even L_p norm: $\sum_{i=1}^{N} |\hat{f}(x_i) - f(x_i)|^p$. The influence of these different norms on the estimates of probability density function is discussed in [9].

[8] discusses the use of a Generalised Gaussian distribution in inverse modelling. It is defined as the normalised probability density f(x) with fixed L_p norm estimator of dispersion: $\sum_{i=1}^{n} n |x_i - x_0|^p f(x_i) = (\sigma_p)^p$ (discrete case), which has the minimum information content (widest spreading) and has the form:

$$f_p(x) = \frac{p^{1-1/p}}{2\sigma_p \Gamma(1/p)} \exp\left[-\frac{1}{p} \frac{|x - x_0|^p}{(\sigma_p)^p} \right] \tag{6}$$

where $\Gamma(.)$ denotes Gamma function and x_0 is the centre of f(x) in the L_p norm sense.

For p=2 $f_2(x)$ is the Gaussian function with mean x_0 and standard deviation σ_2. The Generalised Gaussian distribution covers a range of distributions from the symmetric exponential to a box-car distribution for $p = \infty$. In the case of general exponential distribution family, the log-likelihood based on (2) is equal to the sum of the functions of errors $x - x_0$ and depending on the error structure (in this case the parameter p), we shall get the criteria related to L_1, L_2 or L_p norm.

In [3], it is assumed that, for a general, nonlinear model, the distribution of errors is Gaussian with unknown mean and variance and the log-likelihood function has a sum of squared errors form, equivalent to the Nonlinear Least Squares approach. In this particular case, the equivalence of the measure of fit between the model output and observations and assumed error structure follows from the equivalence between likelihood and least square approaches for the mean of the distribution of independent errors ([11], [10]). Following this approach, the predictive distribution of output variables y_t, modelled by Eq.2, conditioned on the calibration data **z** is given by (discrete case):

$$P(y_t < y \mid \mathbf{z}) = \sum_{\theta} \sum_{\phi} P(y_t < y \mid \theta, \phi) f(\theta, \phi \mid \mathbf{z}) \tag{7}$$

where $f(\theta, \phi \mid \mathbf{z})$ is a posterior likelihood function for the parameters:

$$f(\theta, \phi \mid \mathbf{z}) \sim \exp(-\sum_{t=1}^{T} (z_t - g_t(\theta) - \mu)^2 / 2\sigma^2); \tag{8}$$

$\phi = (\mu, \sigma)$ denotes a vector of statistical model parameters and $P(y_t < y \mid \theta, \phi)$ is a standard normal distribution function N(0,1).

The GLUE methodology is closer in philosophical terms to a nonparametric approach. The critical difference is that posterior distributions for predicted variables are estimated directly from the outputs of a set of multiple acceptable or behavioural models, rather than from a model and an additive residual model. In [2] various likelihood measures are presented and applied to different environmental problems. These include measures similar in nature to

(6) (but also fuzzy measures and binary measures that can be used to exclude some models as non-behavioural). Romanowicz and Beven [3] have also shown how a formal error model may be used within the GLUE framework by evaluating likelihood weights over both model and error model parameters. GLUE allows that the likelihood surface may be very complex in form because of complex parameter interactions. It is the set of parameters that produces a behavioural model for a given input sequence, and there may be no well-defined posterior distribution for individual model parameters.

In the example that follows, we concentrate on two issues: (i) definition of a likelihood measure for use within GLUE with a well-defined scaling parameter based on the dispersion of the errors to control the width of prediction limits; (ii) the influence of the choice of different likelihood measures on the model predictive uncertainty.

3. THE MODELLING PROBLEM: APPLICATION OF RAINFALL-FLOW MODEL TO THE CAN VILA CATCHMENT, N. E. SPAIN

To illustrate the proposed methodology, we consider a rainfall-flow model for a set of rainfall-flow data. The study catchment area of Can Vila is situated in Spain, in the Valcebre catchment of the South-East Pyrenees ([12], [13]). The catchment is 0.56 km^2 in area and is partly covered by Pinus Sylvestris. Rainfall and flow measurements at 20 min. time steps were made available for winter 1995/96 and summer 1997 events. We shall use only winter observations in this study with hourly time step.

The data were modelled using two forms of the mechanistic rainfall-runoff model TOPMODEL ([14]). The SIMULINK version of TOPMODEL, described in [15], bases its calculations of the spatial patterns of hydrological response on the pattern of a topographic index for the catchment derived from a Digital Terrain Model (DTM). The time series data used by the model are the rainfall, runoff and evaporation averaged over the catchment. The model has a modular structure. The saturated zone model is assumed to be non-linear with the outflow $Q_b(t)$ calculated as an exponential function of a catchment average soil moisture deficit S_3 as:

$$\frac{dS_3}{dt} = Q_b(t) - Q_v(t)$$

$$Q_b(t) = Q_0 \exp(-S_3(t)/m) \qquad (9)$$

where Q_0=SK0 $e^{-\lambda}$ is the flow when $S_3(t)=0$. and $Q_v(t)$ denotes the recharge to the saturated zone. SK0 is a soil transmissivity parameter, m is a parameter controlling the rate of decline in transmissivity with increasing soil moisture deficit and λ is the mean value of the topographic index distribution in the catchment (see [16]). Other parameters control the maximum storage available in the root zone (LRZ) and the rate of recharge to the saturated zone (KS).

In the first step the MC sensitivity analysis was performed using the full version of TOPMODEL and January 1996 rainfall-flow data. Following an initial sensitivity analysis the parameter ranges were chosen to ensure that the range of the simulations covers the observations. 10000 simulations were then performed varying the four TOPMODEL parameters according to prior distributions shown in Table 1.

Table 1. Parameter distributions applied in MC analysis of TOPMODEL

	distribution	Min value	Max value	mean	std
SKO	uniform	10	500	251	141
m	uniform	0.003	0.03	0.017	0.0075
LRZ	Log-uniform	1.e-4	0.01	0.0147	0.023
KS	Log-uniform	1.e-15	0.01	0.0003	0.0012

4. ESTIMATION OF DISCHARGE PREDICTION LIMITS

Two methods were then applied to estimate prediction limits: (i) a formal likelihood function based on an assumed error model, and (ii) a non-formal GLUE approach with the likelihood weights proposed by Eq. 12-13.

In both approaches, following [3], we used a multiplicative error model to account for the usual tendency of rainfall-model errors to increase with increasing magnitude of the prediction. Thus:

$$\zeta_t = \log(\delta_t) = \log(Q_{obs,t}) - \log(Q_{sim,t}(\theta)) \quad (10)$$

where $Q_{obs,t}$ denotes the observation of flow at time t and $Q_{sim,t}(\theta)$ denotes the simulated flow for a given model run, depending on parameter set θ.

We then applied the error model (2) with the assumption for the vector error $\zeta_t \sim N(\mu, \Sigma)$; where μ denotes the unknown mean of the errors and $\Sigma = I\sigma^2$ is the covariance matrix. The observation sets for the conditioning of the estimates were chosen in such way that the correlation between the observations could be neglected. Eq. 7 can be used to estimate the predictive uncertainty of the model both for the calibration and the validation stages, under the assumption that the distribution of errors remains the same during the validation stage. Fig. 1 presents the predictions together with 95% confidence limits for the calibration (upper panel) and validation (lower panel) periods.

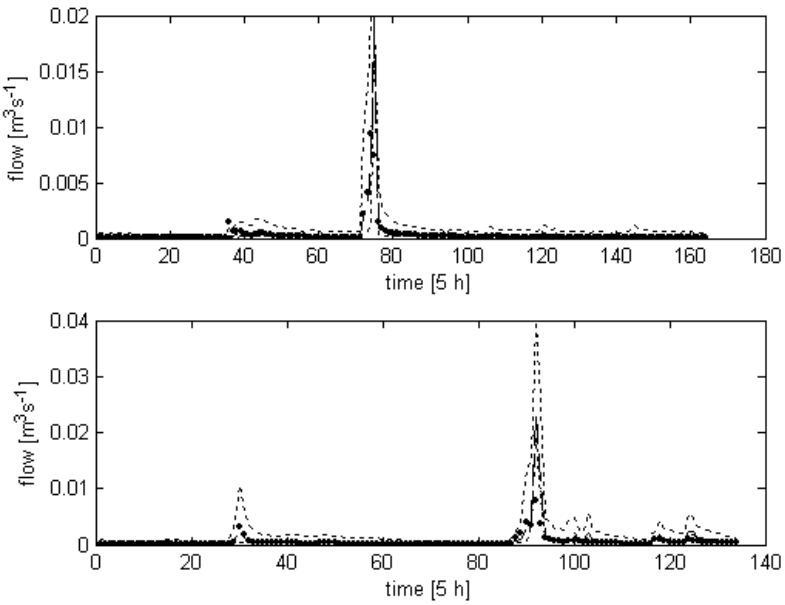

Figure 1. Flow predictions together with 95% confidence limits for Can Villa catchment; upper panel: calibration stage – December 1995, lower panel: validation event January 1996; dashed lines denote the 95% confidence limits, dots represent the observations; formal approach. Note change in discharge scale.

In the GLUE approach the prediction equation takes the form:

$$P(\hat{y}_t < y \mid \mathbf{z}) = \sum_i \{f(\theta_i \mid \mathbf{z}) \mid \hat{y}_t < y)\} \tag{11}$$

We look for the weights $f(\theta_i \mid \mathbf{z})$, which will account for both prediction and parameter/structure related errors. By analogy with (7-8) we assume the form:

$$f(\theta_i \mid \mathbf{z}) = \exp(-\sum_{t=1}^{T} (\log(Q_{t,sim}(\theta_i) - \log(Q_{t,obs}))^2 / \sigma^2) \tag{12}$$

In the formal approach, with an explicit error model, σ^2 is the variance of the prediction error based on the observations. The optimal value of this variance may be derived from the likelihood function (8).

In the non-formal GLUE approach, however, we can treat σ^2 as an additional scaling parameter reflecting our lack of knowledge of the true information content of the residuals ζ_t in constraining the uncertainty in the model predictions. One possible form for this scaling is to take the sum of the variances of the simulated flows over all the behavioural models and all time steps as an estimate, such that:

$$\sigma^2 = \sum_{t=1}^{T} \text{var}(\log(Q_{t,sim}(\theta)) \tag{13}$$

This will increase the dispersion of the resulting posterior likelihoods (relative to the formal case) to account for the predictive model uncertainty without making additional assumptions about the model error structure. This scaling could also be made time-dependent, since the cumulated variance of simulated flows at each time step can be calculated over some specified process memory, in a way similar to allowing a variable kernel in kernel density estimation ([9]) but without making any assumptions about the form of the likelihood surface.

The resulting predictions together with 95% confidence bands for the calibration period are shown in Figure 2.

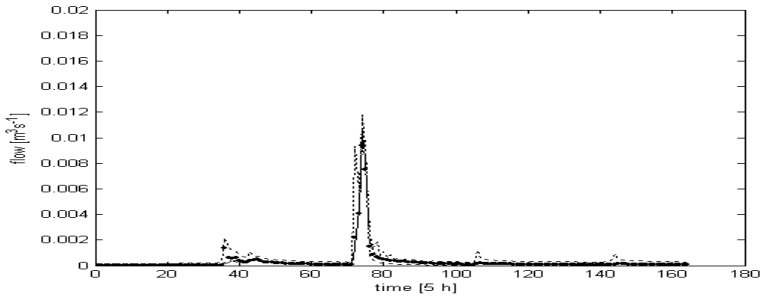

Figure 2. Flow predictions together with 95% confidence limits for Can Villa catchment; December 1995, dashed lines denote the 95% confidence limits, dots represent the observations; non-formal GLUE approach.

Comparison with the Fig. 1 shows that, in this case, the simplified GLUE method gives smaller over-prediction and better represents the observations. Results for a further evaluation period, using the same set of behavioural models and likelihood weights determined for the calibration period, are shown in Fig. 5, upper panel. In this case the GLUE method also gives very good results.

5. INFLUENCE OF THE CHOICE OF OBSERVATION SETS ON THE MODEL PREDICTIVE UNCERTAINTY: COMPARISON OF DIFFERENT GOODNESS OF FIT CRITERIA

The availability and quality of observations is often a major constraint on the identifiability of environmental models. In addition, different prediction problems might require different types of model evaluation. In the case of rainfall-flow models, there is usually sufficient amount of observations available but input errors and model structural errors can give rise to complex error structures for any model run, including heteroscedasticity, nonstationarity and correlation. We can attempt to model these complexities (as in [3]) but experience suggests that less formal methods can still provide useful prediction bounds. In what follows we shall compare the uncertainty predictions obtained when different (reduced) observation sets are chosen for the conditioning of the predictions and different norms are used to evaluate the likelihood weights. We shall use the non-formal GLUE approach in this comparison with L_2 norm in (12) replaced by L_1 norm and also we shall compare the use of observations from the whole time period with conditioning only on time steps with peak discharge observations (global MISE and local MSE goodness of fit criterion) used for the conditioning (as these criteria may use different norms).

As an example, Fig. 3 shows the results of conditioning on the two highest peak values only in terms of the resulting cumulative density functions (cdfs) for the parameters integrated over all the behavioural parameter sets for likelihood weights based on an L_1 norm (dashed

lines). These are compared with the cdfs obtained using likelihood weights based on L_2 norm and the same type of conditioning (dotted lines) and the likelihoods with scaling (12) conditioned on the whole range of observations (solid lines).

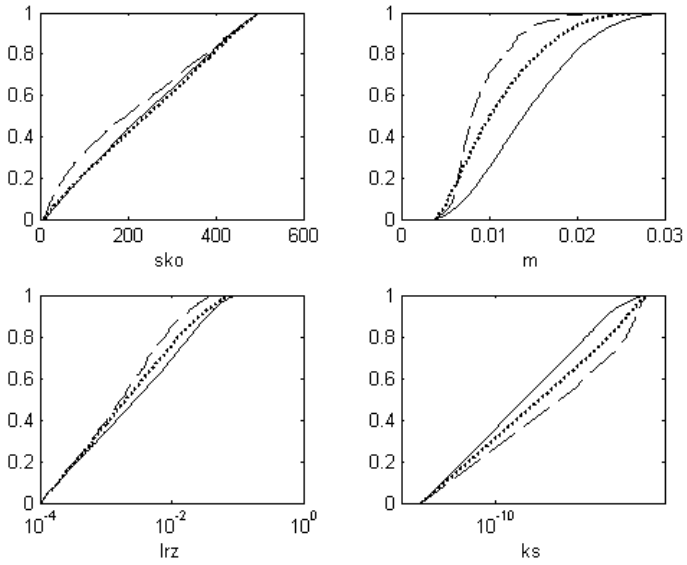

Figure 3. Comparison of the cdfs for the parameters obtained using two criteria for the observation based weights; continuous lines correspond to the weights (12) conditioned on the whole dataset (MISE L_2 criterion); dashed lines show cdfs for the 2 peak values of flow, norm L_1; dotted lines show cdfs for the 2 peak values, L_2 norm.

The results from Fig. 3 show significant differences of posterior distributions of parameters when different observation sets and likelihood weights are used. However, the results were less sensitive to the use of different norms (L_1, or L_2). The resulting prediction limits shown in Fig. 4 are also affected by the choice of the observation sets, with confidence limits for the conditioning on the 2 peak observations of flow better following the peak values but over predicting low flows.

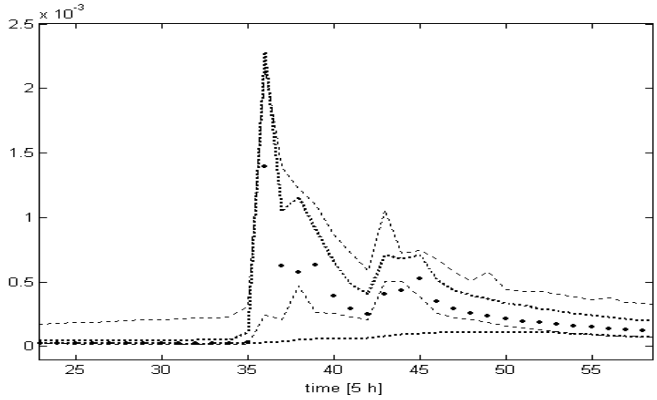

Figure 4. Comparison of confidence levels obtained using two criteria for the calibration period, December, 1995; the thick dashed lines correspond to 95% confidence limits obtained from MISE L_2 criterion (5); thin dashed lines show 95% confidence limits for 2 peak values of flow, L_2 norm, and dots denote the observed flow.

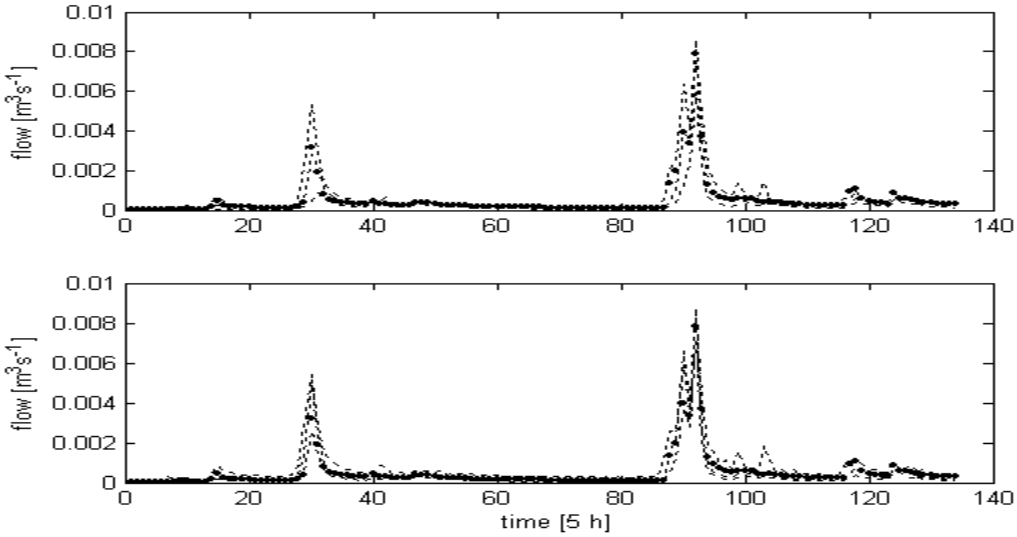

Figure 5. Validation stage: predicted flow (median) (--) and 95% confidence limits (:); big dots denote the observations; upper panel: weights derived from the sum of square errors; lower panel: weights derived from two peak values; January 1996. Note change in discharge scale from Fig. 1

Fig. 5 shows the application of the derived sampling scheme and likelihood weights to a new validation period in January 1996. Comparison of these results with the predictions for the same validation period obtained using formal approach (Fig. 1, lower panel) shows that the non-formal approach provides reasonably good predictions for high flows and is much better for low flows, where the formal approach overestimated the observed values.

6. CONCLUSIONS

Due to the stochastic nature of the variables influencing the deterministic model of the physical process at study, the predictions of the model output should also be considered to be stochastic. We may have some information about the probability distribution of different model variables, but usually environmental models are highly nonlinear and poorly defined, so it may be difficult, or even impossible to obtain the solution of the related inverse problem. This paper is meant as the bridge between formal and non-formal approaches to estimation of hydrological models. We presented a short discussion of statistical methods and their applicability to nonlinear, multidimensional and uncertain processes and pointed out that may be some justification for using a less formal approach such as GLUE.

Our results indicated that use of different criteria for evaluation of likelihood weights influences the shape of the resulting posterior distributions of the parameters but does not influence so much the uncertainty bands for the predictions. This is consistent with past experience reported, for example, in [2] and [17]. In order to obtain the control over the uncertainty limits we should use a suitable scaling parameter for the likelihood weights as well as a suitable choice of the observations for the conditioning of the probabilities for a particular application.

In future work we shall show the influence of the choice of time variable observation window on the uncertainty of model predictions and the way in which the results from different model structures can be integrated within the GLUE framework.

Acknowledgements

This study constitutes the preliminary modelling results in a collaboration with Francesc Gallart, Pilar Llorens, Jérôme Latron, of the Insitute of Earth Sciences Jaime Almera, Barcelona, Spain who are thanked for supplying the data.

REFERENCES

1. K. J. Beven and A. Binley. The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes.* 1992, 6: 279-298.
2. K. J. Beven, J. Freer, B. Hankin and K. Schulz. The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems. In *Nonlinear and Nonstationary Signal Processing*, W J Fitzgerald, R L Smith, A T Walden and P C Young (Eds). CUP, 115-151, 2000.
3. R. J. Romanowicz K. J. Beven and J. Tawn. Evaluation of Predictive Uncertainty in Nonlinear Hydrological Models Using a Bayesian Approach. In: *Statistics for the Environment 2, Water Related Issues* ,eds V. Barnett and K. F. Turkman, 297-315, Wiley, Chichester, 1994.
4. K. J. Beven and P. C. Young. Comment on Bayesian Recursive Parameter Estimation for Hydrologic Models by M Thiemann, M Trosset, H Gupta and S Sorooshian, *Water Resources Research,* 2003: 9(5), doi: 10.1029/2001WR001183.
5. G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Wiley, Chichester, 1992.
6. P. C. Young. Applying Parameter Estimation to Dynamic Systems. parts l, and 2 *Control Engineering*, 1969: 16, 10/11, 119-125/118-124.
7. R. E. Kalman. A new approach to linear filtering and prediction problems. ASME Transactions Journal Basic Engineering 1960: 83-D, 95-108.
8. A. Tarantola. *Inverse Problem Theory*. Elsevier; 1987
9. D. W. Scott. *Multivariate Density Estimation*. Wiley, New York, 1992.
10. A. B. Owen. *Empirical likelihood*. New York: Chapman&Hall/CRC; 2001.
11. R. J. Romanowicz, K. J. Beven and J. Tawn. Bayesian calibration of flood inundation models, in M G Anderson and D E Walling (Eds) *Floodplain Processes*, 333-360. Wiley, Chester, 1996.
12. F. Gallart, J. Latron , P. Llorens and D. Rabada. Hydrological functioning of Mediterranean mountain basins in Vallcebre, Catalonia: Some challenges for hydrological modelling, *Hydrological Processes*. 11 (9): 1263-1272, 1997.
13. F. Gallart, P. Llorens, J. Latron and D. Regues. Hydrological processes and their seasonal controls in a small Mediterranean mountain catchment in the Pyrenees, *Hydrology and Earth System Sciences*. 6 (3): 527-537, 2002.
14. K. J. Beven and M. J. Kirkby. A physically based variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*. 1979: 24, 1, 43-69.
15. R. J. Romanowicz. A MATLAB implementation of TOPMODEL, *Hydrological Processes*, 1997: 11, 1115-1129.
16. K. J. Beven. *Rainfall-runoff modelling: the primer*. Wiley, Chichester, 2001.
17. K. J. Beven and J. Freer. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *J. Hydrology*. 2001; 249: 11-29.