# Combining the generalized likelihood uncertainty estimation (GLUE) and the Bayesian model averaging (BMA) to account for conceptual model uncertainty in groundwater modelling

**RODRIGO ROJAS[1], LUC FEYEN[2] & ALAIN DASSARGUES[1, 3]**

1 *Hydrogeology, Applied geology and mineralogy, Department Geography-Geology, Katholieke Universiteit Leuven, Celestijnenlaan 200 E, 3001 Heverlee, Belgium*
rodrigo.rojasmujica@geo.kuleuven.be

2 *Land management and natural hazards unit, Institute for Environment and Sustainability, DG Joint Research Centre, European Commission, TP261, 21020 Ispra (Va), Italy*

3 *Hydrogeology and environmental geology, Department of Architecture, Geology, Environment, and Civil Engineering (ArGEnCo), Université de Liège, B.52/3 Sart-Tilman, B-4000 Liège, Belgium*

**Abstract** Conceptual model uncertainty is one of the most difficult problems to deal with in the practice of groundwater modelling. In recent years, several methodologies, based on the construction and calibration of alternative models, have been proposed to face this problem. In this article, a more general and flexible approach than those previously developed is described. We achieve this combining the generalized likelihood uncertainty estimation (GLUE) and the Bayesian model averaging (BMA) methodologies. Implementing the GLUE methodology ensures that a large set of acceptable simulators, i.e., conceptual models and parameter sets, are included in the analysis, therefore, avoiding compensation of the conceptual model errors and biased parameter estimates. Implementing the BMA approach allows the inclusion of previous knowledge about the system and the obtaining of consensus multi-model predictions. Preliminary results show that the approach provides a general and flexible framework to account for predictive uncertainty due to the specification of alternative conceptual models.
**Keywords** GLUE, BMA, multi-model prediction, Monte Carlo methods

## INTRODUCTION

With increasing human and climate pressures on groundwater resources, accurate and reliable predictions of groundwater flow and pollutant transport are essential. However, typically, the geological structure is partially known and data of subsurface properties are sparse and prone to error. Consequently, incomplete process representation, errors in the definition of initial and boundary conditions, and errors in the model parameters render the groundwater model predictions uncertain.

Over the last decades, considerable efforts have been put in developing methods to obtain optimal groundwater parameter values and in quantifying predictive uncertainty associated with them. However, the major weakness of parameter-calibration approaches is that all sources of uncertainty are attributed to parameter errors, therefore, ignoring conceptual model uncertainty by confining the suite of plausible conceptual models to a single hydrological concept. This often leads to uncertainty analyses that are under-dispersive and prone to statistical bias (Neuman, 2003).

These concerns have motivated researchers to consider multi-model methods, which seek to obtain consensus predictions from a set of plausible models. The weights to aggregate multi-model outputs can be equal or can be determined through regression-based approaches (e.g., Abrahart and See, 2002; Georgakakos, 2004). However, the weights in such combinations are not connected to model performance and can take any arbitrary real value. Bayesian Model Averaging (BMA) (Draper, 1995; Hoeting et al., 1999), on the other hand, weights the predictions of competing models by their posterior model probability, which represents each model's relative skill in the training period. Hence, BMA weights are tied directly to individual model performance. Neuman (2003) proposed a Maximum Likelihood Bayesian Model Averaging (MLBMA) method to assess the joint predictive distribution of several competing models. MLBMA is an approximation of BMA that relies on maximum likelihood parameter estimation which expands around these estimates through non-exhaustive Monte Carlo simulation. Nevertheless, the mayor weaknesses of the multi-model parameter-calibration approaches are that compensation of conceptual model errors may occur through calibration of biased parameters, and that many parameter sets may perform equally well but provide different results (Refsgaard et al., 2006).

An alternative methodology that rejects the idea of a unique optimal simulator is the Generalised Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992; Beven, 1993). GLUE is based on the concept of equifinality, which acknowledges that there exist many combinations of model structures and parameter sets that provide (equally) good reproductions of the observed system response. Even though equifinality arises because of the combined effects of errors in the forcing data, system conceptualization, measurements and parameter estimates, as yet, it has only been applied considering a single deterministic conceptual model (Refsgaard et al., 2006), thereby, neglecting model structural uncertainty.

In this work, we combine GLUE with BMA to explicitly account for uncertainty that originates from errors in the model conceptualization, forcing data and parameter values. Within the GLUE framework, we explore the global likelihood response surface of a suite of possible combinations of plausible model structures, forcing data and parameter values in order to select those simulators that perform well. For each model structure, the posterior model likelihood is obtained by integrating the likelihood measures over the retained simulators for that model structure. The posterior model likelihoods are subsequently used in BMA to weight the predictions of the competing models when assessing the joint predictive uncertainty.

## METHODOLOGY FOR INTEGRATED UNCERTAINTY ASSESSMENT

### Generalized Likelihood Uncertainty Estimation (GLUE) methodology

GLUE is a Monte Carlo simulation technique based on the concept of equifinality (Beven and Binley, 1992; Beven and Freer, 2001, Beven, 2005). It rejects the idea of a single correct representation of a system in favour of many acceptable or behavioural simulators, i.e. conceptual model and parameter sets. For each simulator a likelihood measure is defined based on the degree of correspondence between simulated and observed records of system response. Simulators that perform better than a subjectively chosen threshold criteria are retained and used to estimate likelihood weighted predictions. Ensemble predictions are based on the predictions of the retained set of simulators, weighted by their respective rescaled likelihood. Different likelihood functions to estimate the likelihood measures can be implemented.

Expanding on the notation of Beven and Freer (2001), let us consider a set of plausible model structures $\mathbf{M}=(M_1,..,M_k,..,M_K)$, with parameter vector $\mathbf{\Theta}=(\theta_1,..,\theta_l,..,\theta_L)$, input variable vector $\mathbf{\Upsilon}=(Y_1,..,Y_m,..,Y_M)$, observed variable vector $\mathbf{Z}^*=(Z^*_1,..,Z^*_n,..,Z^*_N)$ and simulated variable vector $\mathbf{Z}=(Z_1,..,Z_n,..,Z_N)$. $L_{kl}(M_k(\theta_l|\mathbf{\Upsilon},\mathbf{Z}^*))$ represents the associated likelihood measure of the $k$-th model structure and the $l$-th parameter vector, conditioned on both the input data vector $\mathbf{\Upsilon}$ and the observations vector $\mathbf{Z}^*$.

## Bayesian Model Averaging (BMA)

BMA is a statistical procedure that infers consensus predictions by weighting predictions from competing models based on their relative skill, with predictions from better performing models receiving higher weights than worse performing models. Following the notation of Hoeting et al., (1999), if $\Delta$ is a quantity to be predicted, the BMA predictive distribution of $\Delta$, given a set of data $\mathbf{D}$ and a finite and discrete ensemble of conceptual models $\mathbf{M}=(M_1,M_2,..,M_k,..,M_K)$, is given by

$$p(\Delta \mid \mathbf{D}) = \sum_{k=1}^{K} p(\Delta \mid \mathbf{D}, M_k)\, p(M_k \mid \mathbf{D}) \tag{1}$$

where $p(\Delta|\mathbf{D},M_k)$ is the posterior distributions of $\Delta$ and $p(M_k|\mathbf{D})$ is the posterior model probability, both for the $k$-th model structure. This latter term reflects how well model $k$ fits the observed data $\mathbf{D}$ and can be computed using Bayes' rule (2)

$$p(M_k \mid \mathbf{D}) = p(\mathbf{D} \mid M_k)\, p(M_k) \Big/ \sum_{k'=1}^{K} p(\mathbf{D} \mid M_{k'})\, p(M_{k'}) \tag{2}$$

where $p(M_k)$ is the prior probability and $p(\mathbf{D}|M_k)$ is the integrated likelihood of model $M_k$, respectively. Variance of the BMA prediction of $\Delta$ is given by (Draper, 1995)

$$Var[\Delta \mid \mathbf{D}] = \sum_{k=1}^{K} Var[\Delta \mid \mathbf{D}, M_k]\, p(M_k \mid \mathbf{D}) + \sum_{k=1}^{K} \left(E[\Delta \mid \mathbf{D}, M_k] - E[\Delta \mid \mathbf{D}]\right)^2 p(M_k \mid \mathbf{D}) \tag{3}$$

From equation (3) it is seen that the variance of the BMA prediction consists of two terms, the first representing the within-model variance and the second representing the between-model variance.

## Combining GLUE and BMA

Combining the GLUE and BMA methods involves the following steps:

1. Based on prior knowledge, a suite of alternative conceptual models is proposed (e.g., see Neuman and Wierenga, 2003).
2. Realistic prior ranges are defined for the input and parameter vectors under each plausible model structure.
3. Likelihood functions and rejection criteria are defined.
4. For the suite of alternative conceptual models, input and parameter values are sampled from the prior ranges.
5. A likelihood measure is calculated for each simulator based on the agreement between the simulated and observed values. Simulators that are not in agreement with the selected rejection criterion are discarded by setting their likelihood to zero.
6. For each conceptual model $M_k$ a subset $A_k$ of simulators with likelihood $p(\mathbf{D}|\theta_l,M_k) \approx L_{kl}(M_k(\theta_l|\mathbf{\Upsilon},\mathbf{Z}^*)$ is retained. For that, steps 4-6 are repeated until the hyperspace of possible simulators is adequately sampled, i.e., when the

conditional distributions of predicted state variables converge to a stable distribution for each of the conceptual models $M_k$.

7. The integrated likelihood of each conceptual model $M_k$ (eq. 2) is approximated by summing the likelihood weights of the retained simulators in subset $A_k$, or $p(\mathbf{D}|M_k) \approx \sum_{l \in A_k} L_{kl}\left(M_k\left(\theta_l \mid \Upsilon, \mathbf{Z}^*\right)\right)$.
8. The posterior model probabilities are then obtained by normalizing the integrated model likelihoods such that they sum up to one,

$$p\left(M_k \middle| \mathbf{D}\right) \approx \sum_{l \in A_k} L_{kl}\left(M_k\left(\theta_l \middle| \mathbf{Y}, \mathbf{Z}^*\right)\right) \Big/ \sum_{j=1}^{K} \sum_{l \in A_k} L_{jl}\left(M_j\left(\theta_l \middle| \mathbf{Y}, \mathbf{Z}^*\right)\right) \quad (4)$$

9. After normalization of the likelihood weighted predictions under each individual model, a multi-model prediction is obtained with equation (1) using the weights obtained in (4).

## Implementation of the methodology

For illustrative purposes, a hypothetical setup similar to the reference case described in Poeter and Anderson (2005) was used (Fig. 1). The spatial distribution of the hydraulic conductivity was generated using the sequential Gaussian simulation (sgsim) algorithm (Deutsch and Journel, 1998) following an exponential model. Simulation of steady state flow employed MODFLOW-2000 (Harbaugh et al., 2000) using a uniform recharge of $4 \times 10^{-5}$ m $d^{-1}$, a specified head at the west and two river sections. A well pumped 1000 $m^3$ $d^{-1}$ from the lowermost layer. An evapotranspiration zone was defined with a surface elevation at 43 m, evapotranspiration rate of $1.37 \times 10^{-3}$ m $d^{-1}$ and extinction depth of 5 m.

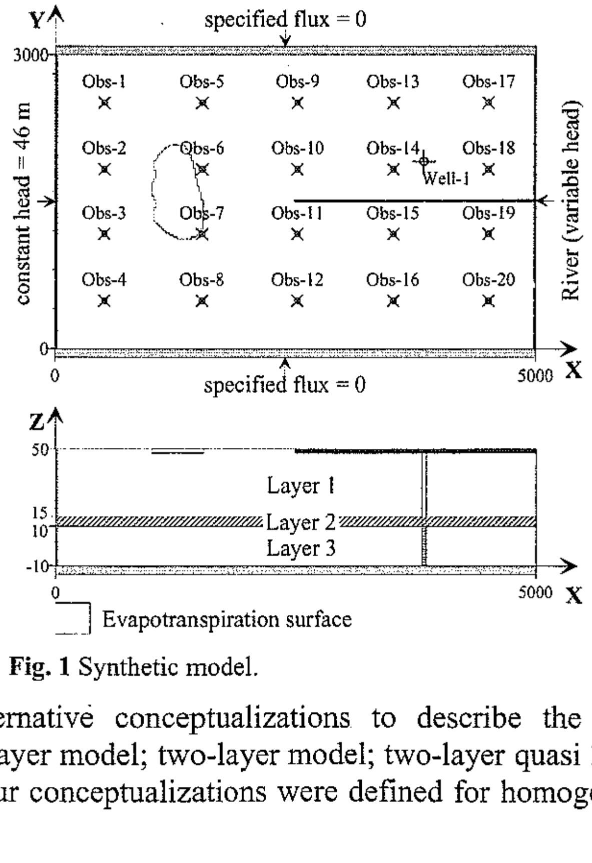


**Fig. 1** Synthetic model.

Eight alternative conceptualizations to describe the synthetic example were defined: one-layer model; two-layer model; two-layer quasi 3D model; and three-layer model. All four conceptualizations were defined for homogeneous and heterogeneous conditions. Non-informative priors for the alternative models and for the unknown inputs and parameters were adopted. Parameter and input vectors were sampled from the prior distributions using a Latin Hypercube Sampling (LHS) scheme. A rejection threshold of 2.5 m was defined corresponding to a maximum allowable deviation at any of the 20 observation wells. Traditional Gaussian (GAUSS) likelihood function (Jensen, 2003), model efficiency (MODEFF) likelihood function (Freer and Beven, 1996; Feyen et al., 2001), and a Fuzzy-type triangular (TRIANG) likelihood function (Jensen, 2003) were used to estimate the likelihood measures.

## PRELIMINARY RESULTS

Figures 2 and 3 present the results for the three likelihood functions. The total variance (Eq. 3) has been estimated for six groundwater budget terms and 20 observation wells. Both components of the total variance are expressed as a percentage of the total variance. From Fig. 2, between-model variance ranges from 3% up to 45% of the total variance, depending on the groundwater budget term and on the likelihood function. For the groundwater heads (Fig. 3) between-model variance ranges from 0.5% up to 48%.

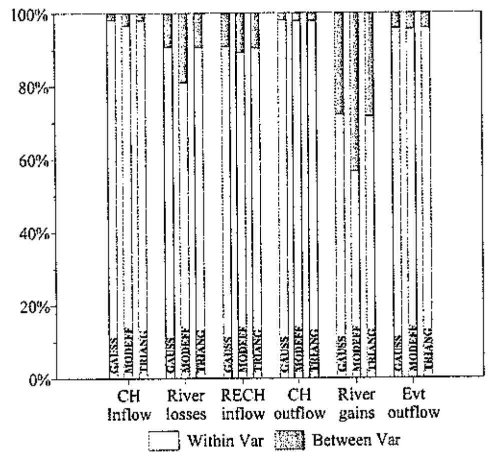


**Fig. 2** Total variance estimated using equation (3) for six groundwater budget terms.

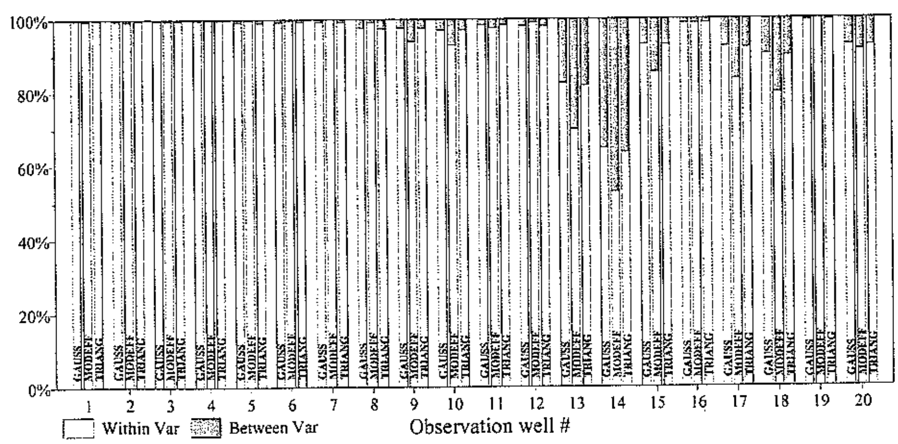


**Fig. 3** Total variance estimated using equation (3) for 20 observation wells of Fig. 1.

Although some differences are observed between the variance estimations made with different likelihood functions, the pattern is identical, i.e., all three likelihood functions identify the same groundwater budget terms and observation wells with the largest between-model variance. It is important to note that the largest predictive uncertainty due to conceptual model uncertainty is observed for the state variables river discharges and recharge inflow, which were not used in the estimation of the likelihood measures. From Fig. 3, a large influence of the pumping well is observed in the estimation of the predictive uncertainty due to alternative conceptual models. In this sense, estimation of likelihood measures could be easily conditioned on quantitative and/or qualitative information about the pumping well through the inclusion of, e.g., stepwise or open-close likelihood functions.

## CONCLUSIONS

We presented an approach for global uncertainty assessment in groundwater modelling. It combines the Generalised Likelihood Uncertainty Estimation and Bayesian Model Averaging to account for uncertainty in the predictions that arise from errors in model structure, forcing data and parameter estimates. The method is flexible because (i) there is no restriction on the diversity of conceptual models or on the level of uncertainty in the forcing data or parameters that can be included; (ii) it allows for different ways of expressing the likelihood of a simulator (including a formal Bayesian one) based on the distribution of the residuals, hence allowing different types of knowledge to be incorporated (quantitative as well as qualitative); and (iii) it is Bayesian in nature, which provides a formal framework to incorporate previous knowledge about the model structures and parameters, or to update the estimates should new information become available.

**Acknowledgments** The authors thank the financial support provided to the first author in the framework of the PhD IRO-scholarships of the Katholieke Universiteit Leuven.

## REFERENCES

Abrahart, R. & See, L. (2002) Multi- model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. *Hydrology and Earth System Sciences* **6**(4), 655-670.

Beven, K. & Binley, A. (1992) The future of distributed models – model calibration and uncertainty prediction. *Hydrological Processes* **6**(3), 279-283.

Beven, K. (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Res.* **16**(1), 41-51.

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

Beven, K. (2005) A manifesto for the equifinality thesis. *J. Hydrology* **320**(1-2), 18-36.

Bredehoeft, J. (2005) The conceptualization model problem – surprise. *Hydrogeol. J.* **13**(1), 37-46.

Deutsch, C. & Journel, A. (1998) *GSLIB Geostatistical software library and user's guide.* Oxford University Press.

Draper, D. (1995) Assessment and propagation of model uncertainty. *J. R. Stat. Soc., Ser. B* **57**(1), 45-97.

Feyen, L., Beven, K., De Smedt, F., Freer, J. (2001) Stochastic capture zone delineation within the GLUE-methodology: conditioning on head observations. *Water Resour. Res.* **37**(3), 625-638.

Freer, J. & Beven, K. (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resour. Res.* **32**(7), 2161-2173.

Georgakakos, K., Seo, D., Gupta, H., Schaake, J., Butts, M. (2004) Characterizing streamflow simulation uncertainty through multimodel ensembles. *J. Hydrology* **298**(1-4), 222-241.

Harbaugh, A., Banta, E., Hill, M., McDonald, M. (2000) MODFLOW-2000, U.S. Geological Survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. *U.S. Geol. Surv. Open File Rep., 00-92, 121 pp.*

Hoeting, J., Madigan, D., Raftery, A., Volinsky, C. (1999) Bayesian model averaging: A tutorial. *Stat. Sci.* **14**(4), 382-417.

Jensen, J. (2003) Parameter and uncertainty estimation in groundwater modelling. PhD thesis. *Department of Civil Engineering, Aalborg University, Series Paper No 23.*

Neuman, S. (2003) Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environ. Res. Risk Assess.* **17**(5), 291-305.

Neuman, S. & Wierenga, P. (2003) A comprehensive strategy of hydrogeologic modelling and uncertainty analysis for nuclear facilities and sites. *Report NUREG/CR-6805, US Nucl. Regul. Comm., Washington, D.C.*

Poeter, E. & Anderson, D. (2005) Multimodel ranking and inference in ground water modelling. *Ground Water* **43**(4), 597-605.

Refsgaard, J., van der Sluijs, J., Brown, J., van de Keur, P. (2006) A framework for dealing with uncertainty due to model structure error. *Adv. Water Res.* **29**(11), 1586-1597.