CHAPTER 4

The Generalized Likelihood Uncertainty Estimation methodology

Calibration and uncertainty estimation based upon a statistical framework is aimed at finding an optimal set of models, parameters and variables capable of simulating a given system.

There are many possible sources of mismatch between observed and simulated state variables (see section 3.2). Some of the sources of uncertainty originate from physical randomness, and others from uncertain knowledge put into the system. The uncertainties originating from physical randomness may be treated within a statistical framework, whereas alternative methods may be needed to account for uncertainties originating from the interpretation of incomplete and perhaps ambiguous data sets.

The GLUE methodology (Beven and Binley 1992) rejects the idea of one single optimal solution and adopts the concept of equifinality of models, parameters and variables (Beven and Binley 1992; Beven 1993). Equifinality originates from the imperfect knowledge of the system under consideration, and many sets of models, parameters and variables may therefore be considered equal or almost equal simulators of the system. Using the GLUE analysis, the prior set of models, parameters and variables is divided into a set of non-acceptable solutions and a set of acceptable solutions. The GLUE methodology deals with the variable degree of membership of the sets. The degree of membership is determined by assessing the extent to which solutions fit the model, which in turn is determined

by subjective likelihood functions. By abandoning the statistical framework we also abandon the traditional definition of uncertainty and in general will have to accept that to some extent uncertainty is a matter of subjective and individual interpretation by the hydrologist. There are strong parallels between uncertainty in a Fuzzy set ruled system and uncertainty in the GLUE methodology. Fuzzy logic is an alternative or supplement to the classical probabilistic framework in situations where very little information is available, and such information as there is tends to be ambiguous and vague. Considering the sources of mismatch between observed and simulated state variables (see section 3.2), it can be argued that the mismatch is to a great extent due to vague and ambiguous interpretations.

The GLUE methodology consists of the 3 steps described below (Fig. 4.1).

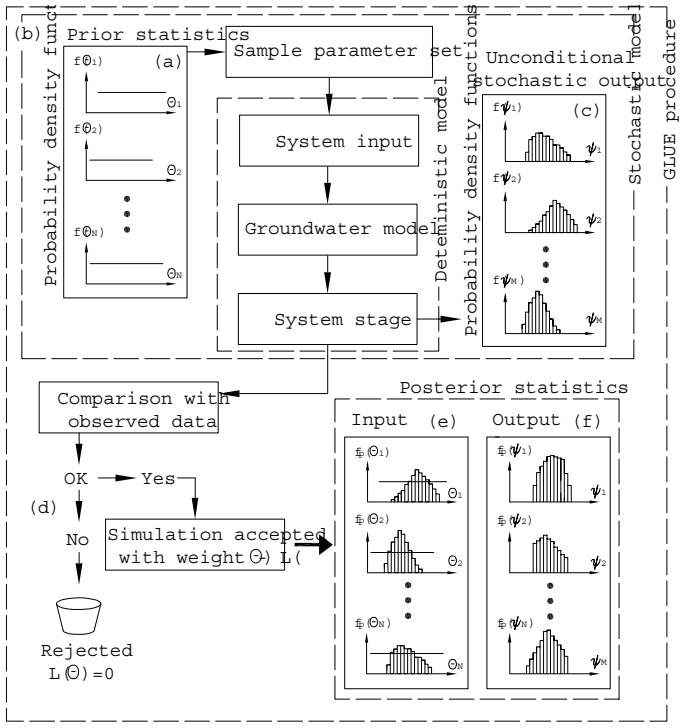


Figure 4.1: The GLUE procedure. (a) prior statistics, (b) stochastic modelling, (c) unconditional statistics of system state variables, (d) evaluation procedure (e) posterior parameter likelihood functions and (f) likelihood functions for system state variables

Step 1 is to determine the statistics for the models, parameters and variables

that, prior to the investigation, are considered likely to be decisive for the simulation of the system **(a)**. Typically quite wide discrete or continuous uniform distribution is chosen - reflecting the fact that there is little prior knowledge of the uncertainties arising from models, parameters and variables. In principle all available knowledge can be put into the prior distributions.

Step 2 is a stochastic simulation **(b)** based on the models, parameters and variables defined in step 1. The Monte Carlo or Latin Hypercube method (Appendix A) may be used to do a random sample of the parameter sets. Step 2 gives us an unconditional estimate of the statistics of any system state variable **(c)**.

In step 3 an evaluation procedure **(d)** is carried out for every single simulation performed in step 2. Simulations and thus parameter sets are rated according to the degree to which they fit observed data. If the simulated state variables are "close" to the observed values the simulation is accepted as having a given likelihood $L(\boldsymbol{\theta}|\boldsymbol{\psi})$, whereas if the considered simulated state variables are unrealistic the simulation is rejected as having zero likelihood.

In this way a likelihood value is assigned to all accepted parameter sets (zero for rejected sets and positive for accepted sets). The direct result of this is a discrete joint likelihood function (DJPDF) for all the models, parameters and variables involved. The DJPDF can only be illustrated in two, maximum three, dimensions, and likelihood scatter plots are often used to illustrate the estimated parameters, see e.g. Fig. 5.7. In Fig. 4.1 the models, parameters and variables $\theta_1, ..., \theta_i, ..., \theta_N$ are considered independent, the likelihood is projected onto the parameter axis, and discrete density functions **(e)** are presented, see section 4.3. Discrete likelihood functions for all types of system state variables can likewise be constructed **(f)**.

4.1 Likelihood measures

Likelihood is a measure of how well a given combination of models, parameters and variables fits, based on the available set of observations. The likelihood measure thus describes the degree to which the various acceptable solutions are members of the set, i.e. their degree of membership.

The calculation of the likelihood of a given set of models, parameters and variables is the key feature of the GLUE methodology, and in this respect GLUE differs from the classical methods of calibration and uncertainty estimation. As will be seen in what follows a wide range of likelihood measures are suggested - all with different qualities. There are no definitive rules for choosing a certain likelihood measure. Some personal preferences are however mentioned in section 4.2.

The likelihood measure consists, in this thesis, of three elements: 1) a rejection level that indicates whether the acceptance criteria are fulfilled or not, 2) a point likelihood measure that sums up the degree of model fit in the individual observation points and 3) a global likelihood measure that is an aggregation of all the point likelihood measures.

Often the rejection level is implicitly given in the point likelihood function, and occasionally the rejection level, the point likelihood measure and the global likelihood measure are all gathered in one function.

The likelihood functions presented below in Fig. 4.2 are based on a combination of the likelihood functions derived from the classical statistical framework and from GLUE, and the Fuzzy logic literature.

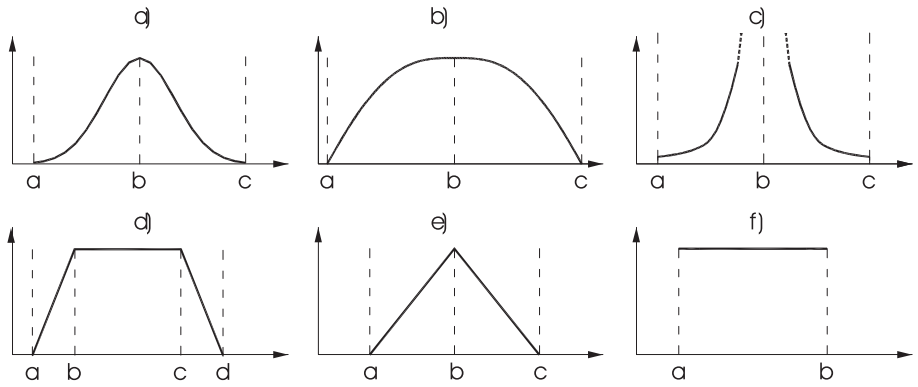


Figure 4.2: a) Gaussian likelihood function , b) model efficiency likelihood function, c) inverse error variance likelihood function, d) trapezoidal likelihood function, e) triangular likelihood function and f) uniform likelihood function

4.1.1 Traditional statistical likelihood measures

Gaussian likelihood function

The Gaussian likelihood function, Fig. 4.2a, is often used in a classical statistical framework. The residuals are assumed to be Gaussian and the likelihood equals the probability that the simulated value, $\psi_i(\boldsymbol{\theta})$, equals the observed value, ψ_i^*:

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_i^*}} e^{-\left(\frac{\left(\psi_i^* - \psi_i(\boldsymbol{\theta})\right)^2}{2\sigma^2_{\psi_i^*}}\right)} \tag{4.1}$$

or for N_{obs} observations

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = (2\pi)^{\frac{N_{obs}}{2}} \mid C_{\boldsymbol{\psi}^*} \mid^{-\frac{1}{2}} e^{\left(\frac{1}{2}(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta}))^T C^{-1}_{\boldsymbol{\psi}^*}(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta}))\right)} \tag{4.2}$$

where, as in a statistical framework $\sigma_{\psi_i^*}$ and $C_{\boldsymbol{\psi}^*}$ symbolise the unknown standard deviation and covariance of observed state variables - often approximated by the expected standard deviation and covariance of observed state variables. Eq. 4.2 corresponds to the product inference function (section 4.1.4) of Eq. 4.1, given independent observations.

The N-dimensional Gaussian likelihood function (4.2) is a function that depends on the number of observations. As the number of observations increases, so does the likelihood of the best simulations, until finally ($N_{obs} \rightarrow \infty$) all likelihood is ascribed to the single best simulation. The likelihood function yields parameter and uncertainty estimates that are similar to those achieved within a statistical framework when this is applied to well-posed linear models with Gaussian errors and the estimate implicit assumes that the model is error free.

The Gaussian likelihood function is defined from $-\infty$ to ∞ and thus no rejection level is implicitly given. In order to reduce the number of simulations, it will often be appropriate to introduce a rejection level (a and c on Fig. 4.2a), e.g. at three times the standard deviation.

4.1.2 Traditional GLUE likelihood measures

Model efficiency function

The model efficiency function, Fig. 4.2b, is given as (Beven and Binley 1992)

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \left(1 - \sigma_\varepsilon^2/\sigma_0^2\right); \quad \sigma_\varepsilon^2 \geq \sigma_0^2 \Rightarrow L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = 0 \tag{4.3}$$

where

$$\sigma_\varepsilon^2 = \frac{1}{N_{obs}} \left(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})\right)^T \mathbf{V} \left(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})\right) \tag{4.4}$$

is the weighted variance of the residuals and σ_0^2 is the weighted variance of the observations. Here $\mathbf{V}$ is a weight matrix.

The likelihood equals one if all residuals are zero, and zero if the weighted variance of the residuals is larger then the weighted variance of the observations.

Inverse error variance function

Beven and Binley (1992) have suggested a function based on the inverse error variance with shaping factor N, Fig. 4.2c:

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \left(\sigma_\varepsilon^2\right)^{-N} \tag{4.5}$$

This function concentrates the weights of the best simulations as N increases. For $N \to \infty$ all weight will be on the single best simulation and for small values of N all simulations will tend to have equal weight.

4.1.3 Fuzzy likelihood measures

A point observation of the i^{th} system state variable, ψ_i^*, and a computed value of the same system state variable, $\psi_i(\boldsymbol{\theta})$ are considered. In the set of all possible values of ψ_i, a subset, Ψ_i, is defined where the transition between membership and non-membership is gradual. The likelihood - or, in Fuzzy terms, the degree of membership - is maximum for simulated state variables that belong completely to Ψ_i; elsewhere it is between 0 and the maximum value. In Fuzzy logic Ψ_i is called a fuzzy set and the likelihood (degree of membership) is described by the likelihood function (membership function), $Ł_{\Psi_i}$. The likelihood function can in principle be an arbitrary, non-symmetric and biased function. The trapezoidal, triangular and uniform likelihood functions are typical Fuzzy logic membership functions where the likelihood or degree of membership is evaluated through relatively simple functions.

First the point likelihood measures are described, and then the point likelihood measures are combined through the so-called inference functions.

Trapezoidal likelihood function

The trapezoidal likelihood function, Fig. 4.2d, is given as

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \frac{\psi_i(\boldsymbol{\theta}) - a}{b-a} I_{a,b}(\psi_i(\boldsymbol{\theta})) + I_{b,c}(\psi_i(\boldsymbol{\theta})) + \frac{d - \psi_i(\boldsymbol{\theta})}{d-c} I_{c,d}(\psi_i(\boldsymbol{\theta})) \tag{4.6}$$

where

$$
\begin{aligned}
I_{a,b} &= \begin{cases} 1 & \text{if } a \leq \psi_i(\boldsymbol{\theta}) \leq b \\ 0 & \text{otherwise} \end{cases} \\
I_{b,c} &= \begin{cases} 1 & \text{if } b \leq \psi_i(\boldsymbol{\theta}) \leq c \\ 0 & \text{otherwise} \end{cases} \\
I_{c,d} &= \begin{cases} 1 & \text{if } c \leq \psi_i(\boldsymbol{\theta}) \leq d \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Triangular likelihood function

The triangular likelihood function, Fig. 4.2e, is given as

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \frac{\psi_i(\boldsymbol{\theta}) - a}{b-a} I_{a,b}(\psi_i(\boldsymbol{\theta})) + \frac{c - \psi_i(\boldsymbol{\theta})}{c-b} I_{b,c}(\psi_i(\boldsymbol{\theta})) \tag{4.7}$$

where

$$
\begin{aligned}
I_{a,b} &= \begin{cases} 1 & \text{if } a \leq \psi_i(\boldsymbol{\theta}) \leq b \\ 0 & \text{otherwise} \end{cases} \\
I_{b,c} &= \begin{cases} 1 & \text{if } b \leq \psi_i(\boldsymbol{\theta}) \leq c \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Uniform likelihood function

The uniform likelihood function, Fig. 4.2f, is a special case of the trapezoidal likelihood function where $a = b$ and $c = d$.

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \begin{cases} 1 & \text{if } a < \psi_i^* - \psi_i(\boldsymbol{\theta}) < b \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

4.1.4 Inference functions

The overall combination of the individual point likelihood (degree of membership) for the observation points is assembled through the so-called degree of fulfilment (DOF) (Dubois and Prade 1980), which, in this context, is the overall likelihood value for the simulation - a global likelihood measure, $L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right)$. A classification of aggregation operators used in Fuzzy rules systems is given in Zimmermann (1991), p. 40-41, and some relevant operators are given below (Dubois and Prade 1980; Zimmermann 1991):

Product inference

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \prod_{i=1}^{N_{obs}} L\left(\boldsymbol{\theta}|\psi_i^*\right) \tag{4.9}$$

The product inference is very restrictive - if one observation is outside the Fuzzy set, Ψ, (i.e. rejected) the global likelihood will be zero. As N_{obs} increases, the global likelihood response surface becomes steeper and steeper and as $N_{obs} \rightarrow \infty$ all except the single best simulation will have negligible likelihood.

Min. inference

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \min_{i=1,\ldots,N_{obs}} L\left(\boldsymbol{\theta}|\psi_i^*\right) \tag{4.10}$$

The min. inference is as restrictive as the product inference function but the global likelihood response surface is more flat.

Max. inference

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \max_{i=1,\ldots,N_{obs}} L\left(\boldsymbol{\theta}|\psi_i^*\right) \tag{4.11}$$

The max. inference is the least restrictive inference function. The likelihood is evaluated from the observation point with the best agreement. If just one observation is inside the Fuzzy set (i.e. accepted), then the simulation is accepted.

Weighted arithmetic mean inference

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \omega_i L\left(\boldsymbol{\theta}|\psi_i^*\right) \tag{4.12}$$

where ω_i is the weight on the i^{th} observation.

As in the case of max. inference, the inclusion of just one observation within the accepted set will result in acceptance of the simulation. The response surface for the arithmetic mean inference is very flat.

Geometric mean inference

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \sqrt[N_{obs}]{\prod_{i=1}^{N_{obs}} L\left(\boldsymbol{\theta}|\psi_i^*\right)} \tag{4.13}$$

The geometric mean inference is as restrictive as the product and min. inference, but the likelihood response surface is less steep. The function is independent of the number of observations.

The way that a likelihood calculation might be performed when different types of observation data are available is illustrated in example 4.1 below.

Example 4.1 *A stationary groundwater model is constructed for a river catchment. The model is calibrated to a summer situation. The following observations are available:*

- *Head observations in 16 wells. From initial studies the standard error on the observed heads is estimated to be 1.5 m. Trapezoidal likelihood functions are applied. Fig. 4.3(a)*
- *Median value of annual minimum discharge observations at one station in "Large Creek". The estimation error is assumed to be Gaussian with a standard error of 10 % of measured discharge. The rejection level is three times standard error. Fig. 4.3(b)*
- *A local farmer has stated that "Little Creek dries out every summer". We do not rely totally on this statement and formulate a likelihood function that gradually decreases from 0 l/s to 2.0 l/s. Fig. 4.3(c)*
- *Information from the local waterworks indicates that so far abstraction well no. 12 has never dried out. Low hydraulic conductivities may result in the closing of abstraction wells in the numerical model. Seen in the light of the information given above, every simulation where the abstraction is closed must be unrealistic, and consequently the likelihood is set at zero. Fig. 4.3(d)*

In all, 19 observations are available and they are combined into an global simulation likelihood measure by an inference rule, e.g. weighted arithmetic mean, Eq. 4.12, or geometric mean inference, Eq. 4.13. Alternatively two or more rules can be combined, e.g. Eq. 4.14.

$$\begin{aligned} L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \quad & \omega_{head}\tfrac{1}{16}\textstyle\sum_{i=1}^{16} L_{h_i^*}(h_i(\boldsymbol{\theta}))\cdot\omega_{q_1}L_{q_1^*}(q_1(\boldsymbol{\theta})) \\ & \cdot\omega_{q_2}L_{q_2^*}(q_2(\boldsymbol{\theta}))\cdot\omega_{Abs}L_{Abs_2^*}(Abs(\boldsymbol{\theta})) \end{aligned} \tag{4.14}$$

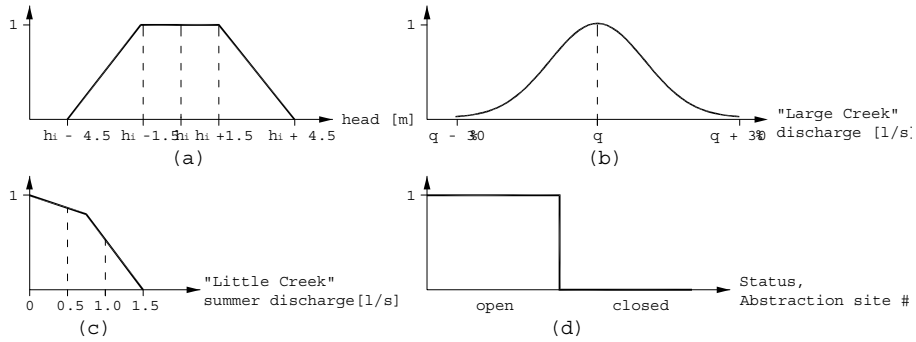


Figure 4.3: Examples of different likelihood functions

where ω_{head}, ω_{q_1}, ω_{q_2} and ω_{Abs} are the weight on observed head data, observed discharge in "Little Creek", observed discharge in "Large Creek" and waterworks observation respectively. $L_{h_i^*}(h_i(\boldsymbol{\theta}))$, $L_{q_1^*}(q_1(\boldsymbol{\theta}))$, $L_{q_2^*}(q_2(\boldsymbol{\theta}))$ and $L_{Abs_2^*}(Abs_2(\boldsymbol{\theta}))$ are likelihood functions for head data, discharge in "Little Creek", discharge in "Large Creek" and the abstraction respectively.

4.2 Designing the likelihood measure

The GLUE methodology is aimed at finding possible sets of models, parameters and variables which produce a model output that is in agreement with observations. The likelihood measure reflects the degree to which we accept the simulated output to deviate from observations due to the numerous error sources.

The first step in the construction of the likelihood function is to analyse possible sources of mismatch between observed and simulated state variables. Section 3.2 is a description of the different types of observation data and a description of the different sources of mismatch between observed and simulated values. Section 3.2 may be used as a guideline in estimating the expected standard errors of observation. In reviewing the possible errors, the hydrologist is forced to consider what is included in the model and what is not. E.g. if the purpose of the model is to model small-scale point pollution, small-scale heterogeneity is very important and consequently has to be modelled in such a way that the error contribution from ignored small-scale heterogeneities will be very small.

In the opinion of the author the estimated expected error should be closely related to the likelihood measure. The rejection level may be three times the expected standard error, reflecting a very low probability of larger errors: see

Chapters 5 and 6

The second step in the calculation of the likelihood measure is the combination of the individual point likelihood measures into a global likelihood measure.

The aim of the point likelihood measures is to account for all expected uncertainty, and in the author's opinion therefore the simulation can be accepted only if all point observations are accepted - no simulated state variables can be tolerated outside the rejection level. If this results in an over-restrictive likelihood measure, the point likelihood measures, and thus the expected errors, should be reconsidered, and if there is no objective reason for increasing the amount of expected error the model should be reconsidered.

The min. inference, the product inference and the geometric mean inference function fulfil the requirement listed above (all point likelihood measures have to be positive in order to accept the simulation).

The geometric mean inference function is attractive because the likelihood measure is independent of the number of observations. This means that the uncertainty estimate does not improve if the number of observations is doubled. This behaviour contrasts with the classical regression framework, where it is assumed that the estimation error is reduced as the number of observations increases. Actually, the maximum likelihood estimate for N independent parameters is the product inference of the independent maximum likelihood estimate.

The reason why the geometric mean inference function is found attractive lies within the error sources. From section 3.2 it can be seen that the main error contributions (scale errors) do not disappear as the number of observations increases, and neither should the uncertainty of the model outcome.

Following the GLUE analysis a validation of all observation points should be performed. From the accepted simulations the probability density functions for the simulated values in the observations points can be found, and the majority of the observations should be within the 95% prediction interval. A poor validation indicates that the likelihood measure is too restrictive and that not all sources of uncertainty are accounted for. See sections 5.6.8 and 6.4.

4.3 Bayesian updating of prior parameter distributions

Following the GLUE analysis the likelihoods are known in a number of discrete points in the space of models, parameters and variables. The posterior likelihood functions for the models, parameters and variables involved can be found from

Bayes' theorem

$$L_p\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \frac{L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) L\left(\boldsymbol{\theta}\right)}{\int L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) L\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta}} \tag{4.15}$$

where $L_p\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right)$ is the posterior likelihood distribution for models, parameters and variables and $L\left(\boldsymbol{\theta}\right)$) is the prior likelihood/probability distribution for models, parameters and variables.

Let us for example assume the we have N_{acc} acceptable parameter sets with likelihood $L\left(\boldsymbol{\theta}_1|\boldsymbol{\psi}^*\right), ..., L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*\right) ..., L\left(\boldsymbol{\theta}_{N_{acc}}|\boldsymbol{\psi}^*\right)$ and from the joint prior likelihood /probability distribution we have corresponding prior likelihood at the same points in parameter space $L\left(\boldsymbol{\theta}_1\right), ..., L\left(\boldsymbol{\theta}_i\right) ..., L\left(\boldsymbol{\theta}_{N_{acc}}\right)$ The posterior likelihood of the points considered in the space of models, parameters and variables is

$$L_p\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*\right) = \frac{L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*\right) L\left(\boldsymbol{\theta}_i\right)}{\sum_{i=1}^{N_{acc}} L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*\right) L\left(\boldsymbol{\theta}_i\right)} \tag{4.16}$$

It can be shown that in the case of uniform prior distributions the posterior likelihood equals the GLUE computed likelihood, $L_p\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*\right) = L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*\right)$.

4.4 An example

In example 3.1, p. 42, it was argued that both head and river inflow observations were necessary in order to make the calibration of q and T unique. The GLUE methodology does not set restrictions on the basis of uniqueness - non-uniqueness will simply result in a larger range of possible parameter values.

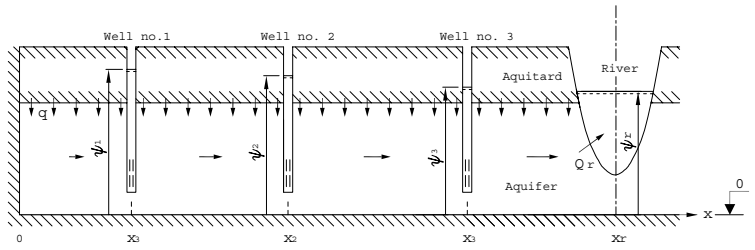
The GLUE methodology is applied to example 1.1, p. 42, with the parameters presented in Fig. 4.4.

A Monte Carlo simulation is performed with 20,000 random realisations of q and T. Each realisation results in an estimate of h_2, h_3 and Q_r. Q_r is found as the total amount of water infiltrated into the aquifer, $Q_r = q \cdot 1000m \cdot 1m$

We now want to use the "observations" of h_2^*, h_3^* and Q_r^* in order to calculate the likelihood of each of the 20,000 simulations. h_2^*, h_3^* and Q_r^* are found from Eq. 1.3 with the parameters:

q $= 400$ mm year^{-1}

T $= 5.\ 10^{-4}$ m^2 s^{-1}



$x_1 = 250$ m $\quad x_2 = 500$ m $\quad x_3 = 750$ m $\quad x_r = 1000$ m $\quad \psi_r = 20$ m
$q = \mathrm{U}[200,600]$ (mm year^{-1}) $\quad \log_{10} T = \mathrm{U}[\text{-}3,\text{-}4]$ ($\log_{10}$ (m^2 s^{-1}))

Figure 4.4: Groundwater flow problem and parameters. $U[\cdot]$ denotes uniform distribution

and error of -0.1 m, 0.7 m and 7.34 10^{-7} m^3 s^{-1} are added to h_2^*, h_3^* and Q_r^*, respectively in order to represent observation errors and model errors. This yields

$h_2^* \;\; = 29.4$ m

$h_3^* \;\; = 25.5$ m

$Q_r^* \;\; = 1.2710^{-4}$ m^3 s^{-1}

Prior to the simulation the expected standard error in the observations is estimated at 0.3 m on the head observations and 10 % of the observed river inflow.

The trapezoidal point likelihood function is used in the evaluation of h_2^*, h_3^* and Q_r^*, see Fig. 4.5.

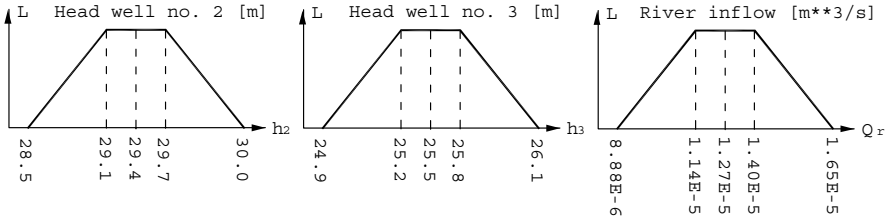


Figure 4.5: Likelihood functions for h_2, h_3 and Q_r

Three point likelihood values, $L_{h_{2,i}}$, $L_{h_{3,i}}$, $L_{Q_{r,i}}$, are calculated on the basis of $h_{2,i}$, $h_{3,i}$ and $Q_{r,i}$ and the global likelihood for the i^{th} simulation is calculated using the geometric mean inference function.

Two scenarios are considered:

a) Only head observations are used in the calculation of the global likelihood:

$$L_i\left(\psi_2^*, \psi_3^* | q_i, T_i\right) = \sqrt{L_{h_{2,i}} L_{h_{3,i}}} \tag{4.17}$$

In this scenario 1,800 of the 20,000 simulations are accepted.

b) Head and river inflow observations are used in the calculation of the global likelihood:

$$L_i\left(\psi_2^*, \psi_3^*, Q_r^* | q_i, T_i\right) = \sqrt[3]{L_{h_{2,i}} L_{h_{3,i}} L_{Q_{r,i}}} \tag{4.18}$$

Here 1,400 of the 20,000 simulations are accepted.

In Figure 4.6 the parameter respond surface for scenarios **a** and **b** is presented.

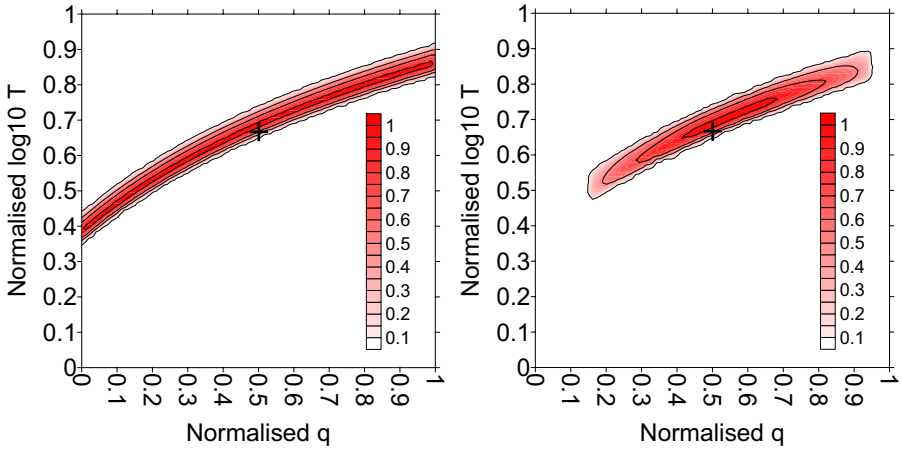


Figure 4.6: Likelihood surfaces in normalised parameter space. ***a)*** *h_2 and h_3 have been used in the calculation of likelihood surface.* ***b)*** *h_2, h_3 and Q_r have been used in the calculation of the likelihood surface. The cross indicates the parameter set used in the calculation of the "observed" values.*

Non-uniqueness is recognised in scenario **a** where only head data are used in the GLUE analysis. If we look at the response surface/curve at a given value of q it is seen that the band of possible T values is quite narrow, but when we look at the total variation of T for all values of q the band is much wider.

In scenario **b** both head and river data are included and the band of possible q values is narrower than in case **a**.

For both scenarios the "true" parameter solution is among the accepted solutions, but not in the region with maximum likelihood. This is due to errors introduced on the observations. If we remove the errors the true parameter solution will fall on the line with maximum likelihood.

The posterior parameter likelihood distributions are identical to the parameter likelihood distribution, because the prior likelihood distribution of q and T is uniform.

If we look at the likelihood distribution curves for ψ_1 for scenarios **a** and **b** we see that they are very similar, Fig. 4.7

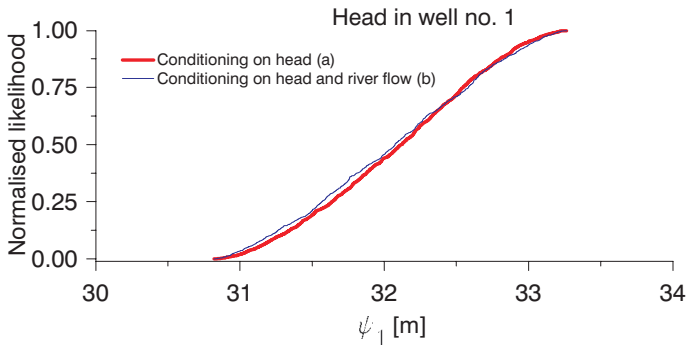


Figure 4.7: Likelihood distribution curves ψ_1.

This indicates that the predictive uncertainty of ψ_1 is mainly influenced by the head observations. If the head rejection criteria are tightened (less expected error in ψ_2 and ψ_3 or a different likelihood function) then there will be less predictive uncertainty in ψ_1. However, this does not mean that predictive capability is invariant to Q_r in general.

Fig. 4.8 presents the likelihood distribution curves of the average Darcy velocity in the aquifer.

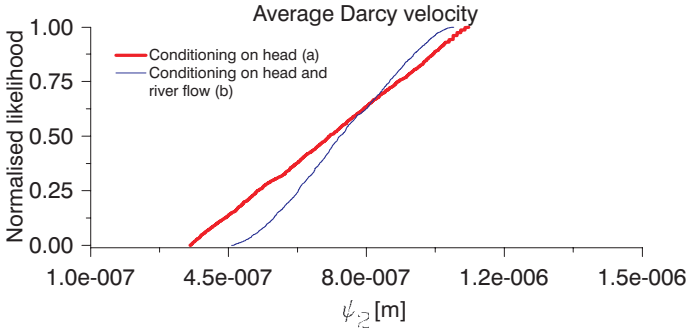


Figure 4.8: Likelihood distribution curves ψ_2.

Scenario **a** results in a significantly larger uncertainty in flow velocities in the aquifer than scenario **b**.

4.5 Generation of random parameter sets

The aim of the GLUE methodology is to find regions in the parameter space resulting in acceptable simulations. Random search methods such as the Monte Carlo method and the Latin Hypercube method have been used in the search in most GLUE applications. These methods are in many ways ineffective because regions of interest often only constitute a small fraction ($< 1\%$) of the prior-defined space of models, parameters and variables. The response surface however is often very complex, with multiple local maxima, valleys and plateaus in a high dimensional parameter space. This makes more intelligent search methods complicated and in some cases inefficient.

In the Gjern setup presented in Chapter 6 an attempt was made to reject certain parameter sets prior to the simulation simply by examining the likelihood in the surrounding region of the parameter set in the parameter space. A similar procedure was used in the original Beven and Binley (1992) study.

The procedure was 1) to generate a parameter set, 2) to interpolate the likelihood value from the surrounding, already simulated, parameter sets, 3) to add a distance-related error to the interpolated value (the closer the point is to the previously sampled parameter sets, the more certain is the interpolation and vice versa) and 4) to simulate the parameter set if the likelihood value was above a certain level. To start with almost all parameter sets were simulated because of the sparse representation, but once a few millions parameter sets had been simulated, up to 60 % of new parameter sets were rejected in advance. There were however no computational benefits from this, due to the costs of interpolating among millions of parameters sets in an 11-dimensional space.

4.6 Concluding remarks

This chapter describes the GLUE methodology that has become a central part of this ph.d. thesis. The use of likelihood functions to evaluate model fit is the key feature of the GLUE methodology. As a supplement to the traditional likelihood measures a number of subjective likelihood measures are introduced and it is thus accepted that the GLUE methodology does not yield uncertainty measures comparable to those produced within the classical statistical framework, but rather offers a statistical measure relating to the subjective impressions of the hydrologist involved. In section 4.2 a few guidelines regarding the design of the likelihood measure have been suggested.

In the following two chapters the GLUE methodology is applied to a synthetic groundwater model and to a region aquifer system.

The main questions to be answered in these two chapters are:

(i) Is it possible from a computational point of view to conduct a GLUE analysis on a typically stationary groundwater model application?

(ii) Is it possible to use the guidelines presented in section 4.2 to design likelihood measures that yield reasonable results?