Computer Physics Communications 136 (2001) 212–224

Computer Physics Communications

www.elsevier.nl/locate/cpc

# Sensitivity analysis in model calibration: GSA-GLUE approach

M. Ratto *, S. Tarantola, A. Saltelli

*Institute for Systems, Informatics and Safety, European Commission, Joint Research Centre, Via E. Fermi 1, TP 361, 21020 Ispra, Italy*

Received 9 January 2001

**Abstract**

A new approach is presented applicable in framework of model calibration to observed data. The approach consists of a combination of the Generalized Likelihood Uncertainty Estimation technique (GLUE) and Global Sensitivity Analysis (GSA). The method is based on multiple model evaluations. The GSA is a quantitative, model independent approach and is based on estimating the fractional contribution of each input factor to the variance of the model output, also accounting for interaction terms. In GLUE, the model runs are classified according to a likelihood measure, conditioning each run to observations. In calibration procedures, strong interaction is observed between model parameters, due to model over-parameterization. The use of likelihood measures allows an estimate of the posterior joint pdf of parameters. By performing a GSA to the likelihood measure, input factors mainly driving model runs with good fit to data are identified. Moreover GSA allows highlighting the basic features of the interaction structure. Any other tool subsequently adopted to represent in more detail the interaction structure, from correlation coefficients to Principal Component Analysis to Bayesian networks to tree-structured density estimation, confirms the general features identified by GSA. © 2001 Elsevier Science B.V. All rights reserved.

*PACS:* 02.70.-c; 02.70.Lq; 07.05.Tp

*Keywords:* Global Sensitivity Analysis; Uncertainty analysis; Likelihood measures; Model calibration; Model uncertainty; Parameter Interaction

## 1. Introduction

This work addresses the problem of model calibration to observed data for distributed models. Generally speaking, mathematical models are characterized by a certain degree of uncertainty, resulting both from uncertainty in modeled processes and observation errors, and the structural and numerical errors of the mathematical model.

Good modeling practice requires the modeler to provide an evaluation of the confidence in the model predictions; possibly assessing the uncertainties associated with the outcome (response) of the model itself. Uncertainty Analysis (UA) and Sensitivity Analysis (SA) are therefore prerequisites for model building in any field where models are used.

Model uncertainty can be accounted for by the application of parametric uncertainty methodologies and conditioning model predictions on observations. In the last decade, a method based on the concept of Bayesian Inference for uncertainty estimation, has been used in hydrology as the Generalized Likelihood Uncertainty Estimation Technique (GLUE) [1,2]. The GLUE technique is as an extension of the Generalized Sensitivity Analysis methodology, which has now come to be called Regional Sensitivity Analy-

* Corresponding author.
*E-mail address:* marco.ratto@jrc.it (M. Ratto).

0010-4655/01/$ – see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0010-4655(01)00159-X

sis (RSA), by Spear and Hornberger [3,4]. GLUE has been developed from an acceptance of the possible equifinality of models, i.e. different sets of model factors/structures, lumped under the term 'input factors' in this work, may be equally likely as simulators of the real system. It works with multiple sets of factors, typically via Monte Carlo sampling, and applies likelihood measures to estimate the predictive uncertainty of the model. Model realizations are weighted and ranked on a likelihood scale via conditioning on observations and the weights are used to formulate a cumulative distribution of predictions. Applying the RSA terminology, model structures/factor sets with almost zero likelihood can be classified as non-behavioral and rejected.

In the RSA-GLUE framework, the basic role of SA is also clear. In general, SA is aimed at establishing how the variation in the model output can be apportioned to different sources of variation, in order to establish how the given model depends upon the information fed into it. When one is mainly interested in the predictive uncertainty, a sensitivity analysis can help in better explaining the model structure and the main sources of model output uncertainty. Additionally, in a calibration problem, a quantitative SA able to account for conditioning on observations, can provide useful information about the model internal structure and, above all, of the interaction structure among model factors resulting from the conditioning itself. The determination of an interaction structure between model factors, in fact, is a typical feature of the RSA-GLUE classification of model realizations, where (complex) structured factor subsets having similar likelihood of being simulators of reality are identified. Spear at al. (1994) [5] found that the application of conventional multivariate statistics like principal component analysis to analyse interaction is not very revealing and Spear (1997) [6] showed also that complex parametric interactions do not become evident from looking at univariate marginal probability densities. So, the deep inspection of this interaction structure is a challenging aspect of RSA and the development of suitable methodologies is still an open problem. The performance of a Global Sensitivity Analysis (GSA) conditioned to observations (i.e. a SA of the likelihood measure to model factors), as well as the evaluation of the effectiveness of GSA for the analysis of the interaction structure is the main goal of the present paper.

In the GLUE approach, factors are never considered independently but as sets of values. The likelihood measure for each model realization is associated with a particular set of factors, conditioned on the observed data. From the methodological point of view, when a detailed SA is to be performed on such kind of 'output', some peculiar aspects have to be taken into account:

- non monotonic input–output mapping;
- high level of interaction between input factors.

The former aspect is mostly due to the form of likelihood measures, inherently non-monotonic. The latter is connected to model equifinality or over-parameterization, i.e. many different combinations of input factors give the same model performance when conditioning on observations. These aspects pose severe constraints about the SA methods to be applied. In particular, it can be expected that MC regression based methodologies are poor for this kind of study and that global, model independent or model free techniques should be used.

In the RSA of Spear and Hornberger, a classification algorithm is applied to the model output, resulting in a classification of each model run as behavioral or non-behavioral. The factor sets leading to the result are stored according to the behavioral outcome. Subsequently all factor vectors are analyzed to determine the degree to which the *a priori* distributions separate under the behavioral mapping. The separation, or the lack thereof, forms the basis of the generalized sensitivity analysis.

Another SA approach applied in the past has been to evaluate the marginal distribution of likelihood for each factor by integrating across the factor space [2]. A simplified approach is given by 'visual' SA based on scatter plots of the likelihood measure vs. single factors. When significant patterns are detected, a pronounced influence on model predictions can be concluded. Moreover, subsets of better model performance can be singled out for such influent factors.

Both the RSA and the scatter plots by Romanovicz et al. address well the problem characteristics, above all as far as the non-monotony is concerned. However, they are quantitatively poor (both) and are not very efficient in the case of strong interaction. A possible extension of RSA for the study of factor interaction has been presented in Hornberger and Spear, 1981 [7], based on the diagonalization of the correlation matrix of the input factor sub-sample under the behaviour classification. More recently, Spear et al. [5] provided a further extension of RSA, consisting of a tree-structured density estimation technique to characterize the complex interaction in the portion of the factor space rising successful simulation. As a result, the factor space can be partitioned into small, densely populated regions and relatively large, sparsely populated regions.

In order to improve the sensitivity analysis aspects of model calibration and uncertainty prediction, in the present paper we apply variance-based Global Sensitivity Analysis methods (GSA) (see Archer et al. [8] for a review). Variance based methods are based on the decomposition of the model output variance into a sum of terms depending on single factors and on interaction terms of increasing order. They are quantitative methods and work without any restriction about monotony or additivity of the model. The only requirement is that all what we desire to know about the model output is captured by its variance. Application of variance-based methods allows the determination not only of main effect of input factors (equivalent to RSA or scatter plots) but also of the total effect of each factor in combination with all the others. Such a quantification is very useful, since it allows a classification of factors according to, e.g.:

- factors with high main effect: such factors affect model output singularly, independently of interaction;
- factors with small main effect but high total effect: such factors influence the model output mainly through interaction;
- factors with small main and total effect: such factors have a negligible effect on the model output and can be fixed at a nominal value.

The first class of factors can be detected also with the other methodological approaches (RSA, scatter plots, regression analysis), while the second class could be qualitatively evaluated through the extension of the RSA in Hornberger and Spear [7] (in case of second order interactions) and in Spear et al. [5].

## 2. The method

The methodology consists of a combination of the GLUE technique with variance based GSA (extended FAST, Importance measures, Sobol' indices, etc.) [9].

### 2.1. Short description of GLUE

A thorough description of can be found in Refs. [1, 2]. The GLUE procedure is based upon making a large number of runs of a given model with different sets of factor values, chosen randomly form specified factor distributions. Different sets of initial, boundary conditions or model structures can also be considered. On a basis of comparing predicted and observed responses, each set of factor values is assigned a likelihood of being a simulator of the system. The definition of the likelihood measure is matter in the GLUE framework and the uncertainty prediction can strongly depend on that definition. In a Bayesian framework, this is connected to how errors in the observations and in the model structure are represented by a statistical model. However, such a differentiation is not relevant as far as the application of GSA is concerned, which is compatible to any definition of the likelihood measure. Hence, for the sake of simplicity and without loss of generality, among the different possible likelihood measures [1,2,10], the following is used here:

$$L(\theta_i|Y) = \left(\frac{1}{\sigma_i^2}\right)^N, \tag{1}$$

where

$$\sigma_i^2 = \frac{1}{2Nobs}\sum_{j=1}^{Nobs}\left(\widehat{Y}_i(t_j) - Y(t_j)\right)^2 \tag{2}$$

is the mean squared difference between predictions and observations for the $i$th factor set.

Rescaling of the likelihood measures such that the sum of all the likelihood values equals 1 yields a distribution function for the factor sets. From this, the uncertainty estimation can be performed, by computing the model output cumulative distribution, together with prediction quantiles.

An interesting feature of this approach is that interaction between factor values is reflected implicitly in the likelihood measure associated with the factor sets, so that no hypothesis about the correlation structure is necessary in defining the *prior* distributions of the model factors.

The GLUE methodology allows also combining or updating likelihood measures, by applying the Bayes theorem [1].

### 2.2. Short description of variance-based GSA

A thorough description of sensitivity analysis methods, including linear regression, correlation analysis, importance measures, variance-based and screening methods, can be found in Saltelli et al. [9].

In variance-based methods the output variance $V(Y)$ can be decomposed in the sum of a top marginal variance and a bottom marginal variance [11]. Specifically,

$$V(Y) = V\big[E(Y|U)\big] + E\big[V(Y|U)\big], \tag{3}$$

where $U$ is a group of one or more elements $X_i$. The top marginal variance from $U$ is the expected reduction of the variance of $Y$ in case $U$ becomes fully known and is fixed at nominal values, whereas other inputs remain variable as before. The bottom marginal variance from $U$ is defined as the expected value of the variance of $Y$ in case all inputs but $U$ become fully known, $U$ remaining as variable as before.

The main effect or first order sensitivity index $S_i$, representing the sensitivity of $Y$ to the factor $X_i$, is defined as the top marginal variance divided by the total variance, where the subset $U$ reduces to the single factor $X_i$:

$$S_i = \frac{V[E(Y|X_i = x_i^*)]}{V(Y)} \tag{4}$$

and represents the average output variance reduction that can be achieved when $X_i$ becomes fully known and is fixed. Estimation procedures for $S_i$ are the Fourier Amplitude Sensitivity Test, FAST [12], the method of Sobol' [13], and others [14]. Higher order sensitivity indices, which quantify the sensitivity of the model output to interactions among subsets of factors, can be estimated using similar formulae. For instance, the second order sensitivity index $S_{ij}$, representing the sensitivity of $Y$ to the interaction between $X_i$ and $X_j$, is

$$S_{ij} = \frac{V[E(Y|X_i = x_i^*,\ X_j = x_j^*)]}{V(Y)} - \frac{V[E(Y|X_i = x_i^*)] - V[E(Y|X_j = x_j^*)]}{V(Y)}. \tag{5}$$

From the definitions in Eqs. (4) and (5), a complete series development of the output variance can be achieved:

$$1 = \sum_i S_i + \sum_{i<j} S_{ij} + \sum_{i<j<m} S_{ijm} + \cdots + S_{12\ldots k}, \tag{6}$$

where higher order terms are defined in a similar way to Eq. (5). Given that the estimation of each sensitivity index, be it $S_i$, $S_{ij}$ or higher order, might require a significant number of model executions, the analysis is rarely carried further after the computation of second order indices (their number is $k(k-1)$), as the related computational load might be impracticable.

The investigation of higher order effects is computationally cheaper if *total sensitivity indices* are employed. The total sensitivity index $S_{Ti}$ for the factor $X_i$ collects in one single term all the interactions involving $X_i$. It is defined as the average output variance that would remain as long as $X_i$ stays unknown (i.e. the bottom marginal variance with $U$ grouping all factors but $X_i$):

$$S_{Ti} = \frac{E[V(Y|X_{-i} = x_{-i}^*)]}{V(Y)}. \tag{7}$$

The term $X_{-i}$ indicates all the factors but $X_i$. The usefulness of the $S_{Ti}$ is in that they can be computed without necessarily evaluating the single indices $S_{ijlm\ldots}$, thus making the analysis affordable from a computational point of view [15].

Estimating the pair $(S_i, S_{Ti})$ is important to appreciate the difference between the impact on $Y$ of the factor $X_i$ alone (the $S_i$) and the overall impact on $Y$ of factor $X_i$ through interactions with the others (the $S_{Ti}$). Such property is particularly interesting in a calibration framework, where high order interaction are usually encountered. Efficient estimators of the pair $(S_i, S_{Ti})$ are provided by variance-based techniques such as the extension of the Fourier Amplitude Sensitivity Test (FAST) [16] and the Sobol' method [15].

### 2.3. Combined GSA-GLUE approach

The way of combining GSA and GLUE is straightforward. It is necessary that the sample generated for the GLUE analysis is designed also for the computation of variance-based sensitivity indices. So a Sobol' sample or a FAST sample should be used. In this way, by applying the same set of model runs, predictive uncertainty can be estimated, sensitivity indices computed and bootstrapping performed.

## 3. Chemical experiment case study

A very simple chemical system has been considered as a first case study, consisting of the observation of the time evolution of an isothermal first order irreversible reaction in a batch system $A \to B$. Such a system is described by the following model equation:

$$\frac{dy_B}{dt} = ky_A, \tag{8}$$

where

$$y_i = \frac{n_i}{n_T} = \frac{n_i}{n_A^0 + n_B^0}, \quad i = A, B,$$

is the dimensionless concentration, $k$ is the chemical kinetics rate constant.

The solution to this ordinary differential equation leads to:

$$y_B(t) = 1 + \left(y_B^0 - 1\right)\exp(-kt). \tag{9}$$

An experiment has been simulated, by considering the following operating conditions:

$$\begin{aligned} &k = k_\infty \exp(-E/RT),\\ &k_\infty = 2.5\text{e}5\ \text{s}^{-1},\\ &E/R = 5000\ \text{K},\\ &T = 300\ \text{K},\\ &[k = 0.014\ \text{s}^{-1}],\\ &y_B^0 = 0.1. \end{aligned} \tag{10}$$

To simulate observations, a normally distributed error has been added to the analytical behaviour, with zero mean and standard deviation 0.05. The pseudo-experiment considered for the present test case is shown in Fig. 1.

Three factors have been considered in the GSA-GLUE study (see Table 1). This case study is designed to represent the typical case of an over-parameterized model, with a complex structure, in which the underlying interaction between factors is not elementarily detectable. In fact, the strong interaction between $k_\infty$ and $E$ is a well-known feature in the estimation of chemical rate constants (see, e.g., [17]).

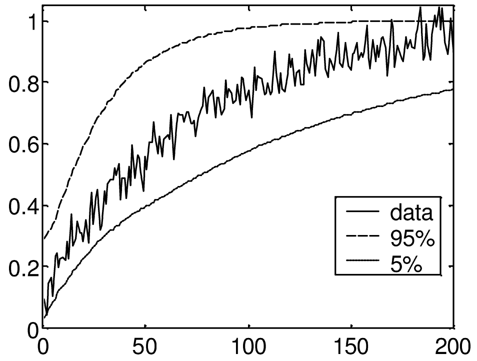


Fig. 1. Experimental time series and 5 and 95% confidence bounds for the output $y_B$, using the likelihood measure $1/\sigma^2$.

Table 1
Input factors used for the GSA-GLUE analysis and distributions

| Factor | Distribution |
|---|---|
| $k_\infty$ | U[0, 5e5] |
| $E$ | U[4500, 5500] |
| $y_B^0$ | U[0, 0.3] |

An $LP_\tau$ sample of size 2048 [18] was generated to estimate Sobol' sensitivity indices (first and total effect). Two model outputs have been considered: the physical output $y_B(t)$ and the likelihood measure with $N = 1, 4$. By increasing $N$, we give a much higher weight to good runs, while most runs are classified as 'unlikely'.

### 3.1. GLUE analysis

In Fig. 1 the confidence bound (5 and 95%) of the output $y_B$ is shown, obtained by using the above defined likelihood measure with $N = 1$. This is an example of use of GLUE for the prediction uncertainty.

### 3.2. Global sensitivity analysis

In Fig. 2 scatter plots are shown for the likelihood measure vs. the three factors, while results of the Sobol' SA vs. time are shown in Figs. 3 and 4

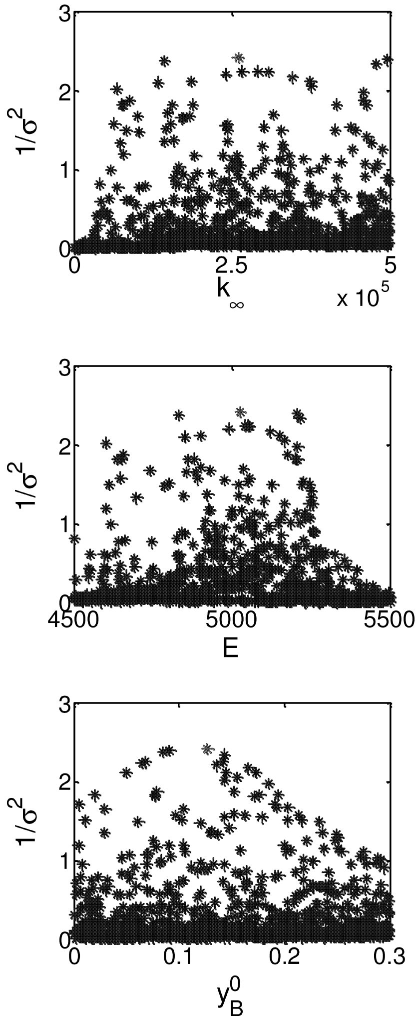


Fig. 2. Scatter plots of the likelihood measure $1/\sigma^2$ vs. $k_\infty$, $E$, $y_B^0$.

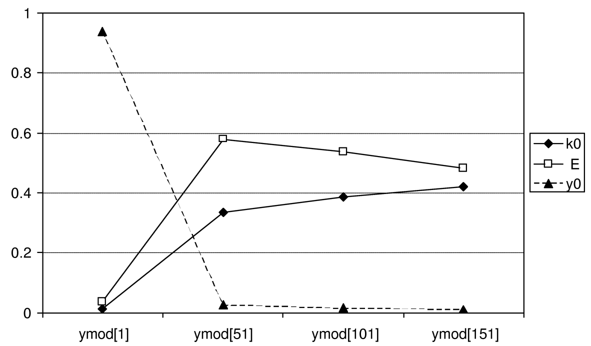


Fig. 3. Sobol' 1st order sensitivity indices for the output $y_B(t)$.

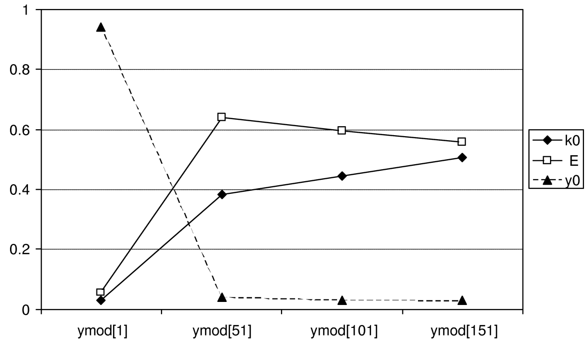


Fig. 4. Sobol' total order sensitivity indices for the output $y_B(t)$.

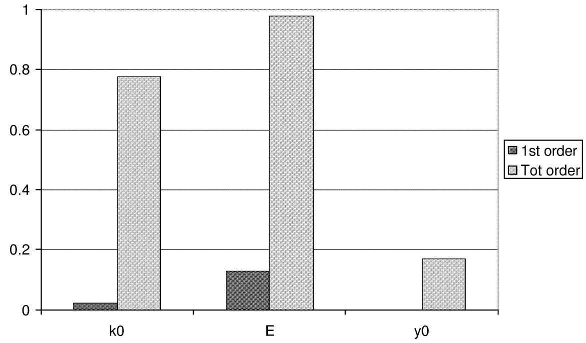


Fig. 5. Sobol' sensitivity indices for the likelihood measure $(1/\sigma^2)$.

for the physical output $y_B(t)$ and in Fig. 5 for the likelihood measure. Scatter plots provide the same type of information as main effect sensitivity indices. In fact, the conditional variance defining main effects in Eq. (4) can be 'visualized' in scatter plots. The inner expectation is the mean of the likelihood measure at fixed values of the factor $x_i$, i.e. by considering vertical 'slices' in the scatter plot. If such a mean varies strongly by varying $x_i$, the main effect is high. Accordingly we must see a clear pattern in the scatter plot by moving along the abscissas.

### 3.2.1. Analysis of the physical output

The numeric values of the Sobol' indices are shown in Tables 2 and 3. Sobol' sensitivity indices show a trend in time where the initial condition $y_B^0$ is important for the very initial time period, while the factors of the chemical rate constant prevail for the rest of the simulation. The sum of the first order indices is never smaller than 0.86. By considering the total effect sensitivity indices, a very slight increase

Table 2
Sobol' 1st order indices of $y_B(t)$ at different times

| | $t=1$ | $t=51$ | $t=101$ | $t=151$ |
|---|---|---|---|---|
| $K_\infty$ | 0.00431 | 0.35322 | 0.42291 | 0.46856 |
| $E$ | 0.02897 | 0.56321 | 0.49595 | 0.42532 |
| $Y_B^0$ | 0.94262 | 0.02841 | 0.01603 | 0.01167 |

Table 3
Sobol' total order indices of $y_B(t)$ at different times

| | $t=1$ | $t=51$ | $t=101$ | $t=151$ |
|---|---|---|---|---|
| $K_\infty$ | 0.02397 | 0.40774 | 0.4765 | 0.54324 |
| $E$ | 0.04509 | 0.61671 | 0.55044 | 0.50147 |
| $Y_B^0$ | 0.94789 | 0.03302 | 0.0188 | 0.01263 |

Table 4
Sobol' sensitivity indices using $N=1$

| | 1st order | Tot order |
|---|---|---|
| $k_\infty$ | 0.02289 | 0.77898 |
| $E$ | 0.13028 | 0.97787 |
| $y_B^0$ | 0.00063 | 0.17242 |

in the absolute values with respect to the first order sensitivity indices is detected. This implies that few interaction is revealed by the analysis of the physical output, which simply singles out the importance of *both* kinetic factors.

### 3.2.2. *Analysis of the likelihood measure*

Numeric values of the Sobol' sensitivity indices are shown in Table 4. In this case, the sensitivity behaviour is strongly modified. Scatter plots are a representation of the marginal posterior distribution for the input factors. By considering the scatter plots, no clear trend can be seen for the chemical rate factors, while for the initial condition a trend is detected only for the high values of likelihood, but no subset can be singled out for $y_B^0$, where the likelihood measure has only high values. Also applying the Hornberger and Spear RSA approach (not shown here), no appreciable separation in the factor distribution is detected, when separating runs with high likelihood/small likelihood. This is reflected in the first order sensitivity indices, which are much smaller than the main effects for the physical output. From the analysis of the main effect (RSA, scatter plots, first order indices) we can conclude that no single factor drives the model to be more 'behavioral' and that interaction mainly characterizes model calibration. On the basis of the main effect, it is not possible to get any information about the interaction structure.

By analyzing the total effect indices, very high sensitivity is detected for the chemical kinetics factors, implying that the behavioral runs are driven by an interaction between them. On the other hand, the influence of the initial condition is small also in terms of total effect. So, total effect indices provide a quantitative completion of the sensitivity analysis, allowing the identification of the factors that are mostly interacting.

### 3.2.3. *First conclusions about GSA*

From these results, one may conclude that:
(1) the initial condition can be judged as unimportant, since it has the smallest total effect;
(2) the chemical rate factors mainly drive the model fit to the experimental data, since they have the highest main and total effects;
(3) on the other hand, the chemical rate factors cannot be precisely estimated, since the absolute values of the first order indices are small, leaving the main contribution to the output variance to interaction terms;
(4) the high difference between main and total effects implies that the model is over-parameterized.

We claim that items (1)–(4) can be taken as general rules for the analysis of model calibration to experimental data. In particular, it is very interesting the deep relationship singled out in items (3) and (4) between the difference between total and first order sensitivity indices, the indeterminacy of the optimization (estimation) problem, the interaction structure of the input factors in the posterior pdf conditioned on the observations.

### 3.2.4. *SA on model output and on its likelihood: what differences?*

By comparing results in the previous sections, it is evident that the input–output structure is much more complicated when using the likelihood than considering the physical output. In particular, we should generally expect that the likelihood function is non-monotonic with respect to the input factors and that more interactions are reflected by the use of the likelihood. This implies some restriction as far as the SA tools to be applied: specifically, only variance-based methods are suitable, since they are model free, they are able to deal with non-monotonic behaviour and to reveal interaction terms.

Finally, in the present test case, we knew that there exist an interaction between $E$ and $k_\infty$, but only by analyzing the likelihood we could appreciate it. This implies that SA for the likelihood gives very useful information for model calibration.

#### 3.2.5. *Further analysis*

The further work on this simple system aims at verifying the statements above and to get further information from the analysis of the likelihood measure. Specifically:

- A deeper study of the interaction structure. This will show how any tool applied to represent in more detail the interaction structure confirms the basic features identified by the GSA. Moreover some ideas are also shown on how improving the use of the information provided by the likelihood measure for the analysis of the interaction structure. In Spear's RSA, the model outcome classification is based on a separation in the behavioral/unbehavioral classes. So, no likelihood measure is defined for the different sets of input factors and the inspection of the interaction structure can be performed only based on the sub-sample of the behavioral runs (correlation matrix, PCA, tree-structured density estimation). On the other hand, in the GLUE approach, the likelihood measure provides very useful additional information about the posterior joint pdf, which can be effectively used to analyse the interaction structure.
- The modification of the set of observations in order to modify the interaction structure. This will allow verifying that the GSA results change according to the new structure.

### 3.3. *Analysis of the interaction structure*

This analysis aims at studying in more detail the properties of the posterior joint pdf of the input factors. Moreover, this analysis allows a direct verification of the effectiveness of GSA in highlighting factors driving 'behavioral' runs. Advanced methods are available for the analysis of the interaction structure, such as the tree-structured density plots by Spear et al. [5] or the construction of Bayesian networks in the space of the input factors [20]. However, in this simple case study, the application of more simple techniques is effective. Moreover, the information provided by the likelihood measure will be applied. The first idea is to study the covariance structure of the posterior distribution of the input factors obtained by applying the likelihood measure.

By normalizing likelihood measures we obtain weights such as:

$$\sum_{i=1}^{n} w_i(\mathbf{x}) = 1, \tag{11}$$

where $n$ is the sample size and $\mathbf{x}$ is the vector of input factors $\mathbf{x} = (x_1, x_2, \ldots, x_m)$.

The properties of the posterior marginal distributions can be evaluated as follows:

$$\widehat{E}(x_j) = \sum_{i=1}^{n} x_{ij} w_i(\mathbf{x}) = \hat{\mu}_j, \tag{12}$$

$$\widehat{V}(x_j) = \sum_{i=1}^{n} x_{ij}^2 w_i(\mathbf{x}) - \hat{\mu}_j^2 = \hat{\sigma}_j^2 \tag{13}$$

and defining the new standardized factors:

$$\tilde{x}_j = \frac{x_j - \hat{\mu}_j}{\hat{\sigma}_j}, \tag{14}$$

the correlation coefficients can be estimated as:

$$\hat{\rho}_{jk} = \sum_{i=1}^{n} \tilde{x}_{ij} \tilde{x}_{ik} w_i(\mathbf{x}) \in (-1, 1). \tag{15}$$

Such correlation coefficients would allow evaluating the pair-wise interaction structure, which is usually not observable from the total sensitivity indices. This analysis is similar to the approach presented by Hornberger and Spear [7] but with a major difference: in RSA, correlation structure is analyzed on the behavioral subset, while here all runs are used by applying different weights. This allows using the whole information.

In our test case, the matrix shown in Table 5 gives the posterior correlation structure. The posterior distribution confirms the interaction between the kinetic factors highlighted by the GSA. When high values of

Table 5
Estimate of the correlation matrix of the posterior joint distribution

| | $k_\infty$ | $E$ | $y_B^0$ |
|---|---|---|---|
| $k_\infty$ | 1 | 0.6682 | −0.0387 |
| $E$ | 0.6682 | 1 | 0.0901 |
| $y_B^0$ | −0.0387 | 0.0901 | 1 |

the correlation coefficients are detected, they also suggest a way to reduce the input factor space, in particular if the coefficient is positive, the couple of factors acts in the model as a quotient/difference, if it is negative they act as a product/sum. In the case under analysis, the positive sign correctly reveals the quotient interaction of $k_\infty$ and $E$. This is a clarification of what we claimed in the previous paragraphs: GSA allows a general, quantitative, model free identification of basic features of the interaction structure. On the other hand, it does not allow a complete representation of such a structure. Such a representation can be drawn applying other tools, which, on the other hand, require the introduction of more stringent assumptions about the interaction structure and have a less general applicability. In all cases, such representations confirm GSA results (in this case the interaction between the kinetic factors) and GSA, therefore, is a 'common denominator' to them.

### *3.3.1. Principal component analysis*

Another easy way for the representation of the interaction structure is provided by the performance of a Principal Component Analysis (PCA). Let us imagine to have a sample from the 3-dimensional posterior joint probability distribution $(k_\infty, E, y_B^0)$. PCA of this sample consists of determining the eigenvalues and eigenvectors of the correlation matrix of the sample (see, e.g., [19]). Eigenvectors provide a transformation matrix from the space of the (standardized) original factors to the principal components. Such components will have null correlation under the posterior pdf. Diagonalization of that matrix above determined matrix provides the transformation between the original (standardized) factors to the principal components in the posterior joint pdf. This is shown in Table 6.

PC1 is the main component and accounts for the 56% of the total variance in the joint posterior factor space. PC1 is dominated by the kinetic factors, and essentially is the sum of $k_\infty$ and $E$. This component can be seen as an evaluation of the sub-space of all the possible solutions for the under-determined calibration problem. In other words the direction determined by PC1 is the direction along which the likelihood has, on average, the smallest variation. This is also shown in the scatter plot of Fig. 6. The principal components give us an evaluation of the equifinality sub-space of the input factors. If we find a factor set leading to a 'behavioral' run and we move along PC1, we would have the highest probability, on average, to find behavioral runs, i.e. we are moving in the equifinality sub-space. In particular, by observing the coefficients for $k_\infty$ and $E$ in PC1, given a 'good' factor set, the model quality is not changed, on average, if $k_\infty$ and $E$ are augmented/decreased by the same fraction of their standard deviation.

Table 6
Principal components of the posterior joint pdf

| | PC1 | PC2 | PC3 |
|---|---|---|---|
| | 55.67% | 33.67% | 10.65% |
| $k_\infty$ | 0.7034 | 0.1325 | 0.6983 |
| $E$ | 0.7087 | −0.0553 | −0.7034 |
| $y_B^0$ | 0.0546 | −0.9896 | 0.1329 |

Table 7
Estimate of first order sensitivity indices for principal components

| PC1 | PC2 | PC3 |
|---|---|---|
| 0.0224 | 0.0074 | 0.5816 |

The same conclusion can be obtained from another point of view, i.e. analyzing PC3. Along PC3 (the one with the smallest eigenvalue), the likelihood measure has the greatest variation (see Fig. 6). This implies that, for a good model calibration, the first factor to be fixed should be PC3. However, by comparing PC1 and PC3, it can be noted that to fix PC3, almost exactly correspond to move along PC1. So, as before, we obtain that an approximated evaluation of the structure of the equifinality sub-set is given by the sub-space spanned by PC1. This is also reflected by the main effect sensitivity indices, which are clearly dominated by PC3 (see Table 7).

Sensitivity indices have been evaluated by applying the definition given in Eq. (4), where the conditional expectations are computed by dividing the domain

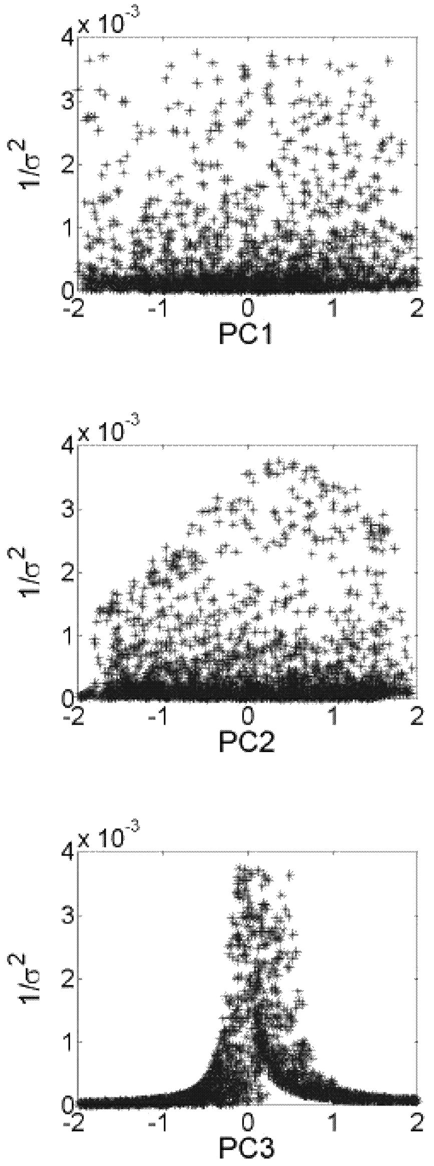


Fig. 6. Scatter plots for the principal components in the space defined by the posterior joint pdf.

of each component in bins of equal probability. For each bin, the expected value is computed. Finally, the variance of the expectations is calculated. The central limit theorem suggests that bins of equal probability can be defined assuming that principal components have normal distributions.

In this simple case study, the study of the correlation matrix and the performance of a PCA give us a quite clear description of the interaction structure defined by the equifinality sub-space of the input factors. However, as already explained in the introduction [5], in more complex cases, it is often verified that the interaction structure cannot be summarized mainly into second order terms and PCA is not informative. So it cannot be generally expected that PCA is effective.

In such cases, a deeper comprehension of the interaction structure needs more powerful techniques, such as the tree-structured density estimation technique by Spear et al. [5], or the use of Bayesian networks [20]. In all cases, GSA results will be confirmed by any particular representation.

### 3.3.2. Bootstrapping

Further inspection in the posterior joint pdf can be obtained by a bootstrapping procedure. Each factor set can be sampled with a frequency proportional to the weight assigned by applying the likelihood measure (Russian roulette). As a result, a sample of the posterior joint pdf is obtained by using the same runs of the previous analyses. By considering $N = 1$, we obtain a new sample with the correlation matrix in Table 8, matching very well the evaluation performed in the previous section:

On the other hand, by considering $N = 4$, the correlation matrix of Table 9 is obtained, with a much larger correlation between kinetic factors.

Table 8
Correlation matrix of the new sample obtained with bootstrapping, with $N = 1$ in the likelihood measure

| | $k_\infty$ | $E$ | $y_B^0$ |
|---|---|---|---|
| $k_\infty$ | 1.0000 | 0.6615 | −0.0266 |
| $E$ | 0.6615 | 1.0000 | 0.0898 |
| $y_B^0$ | −0.0266 | 0.0898 | 1.0000 |

Table 9
Correlation matrix of the new sample obtained with bootstrapping, with $N = 4$ in the likelihood measure

| | $k_\infty$ | $E$ | $y_B^0$ |
|---|---|---|---|
| $k_\infty$ | 1.0000 | 0.9433 | 0.0041 |
| $E$ | 0.9433 | 1.0000 | 0.0762 |
| $y_B^0$ | 0.0041 | 0.0762 | 1.0000 |

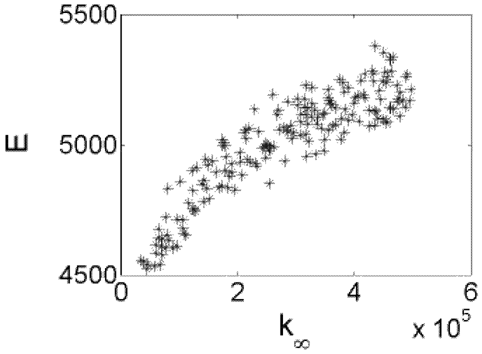


Fig. 7. Projection of the joint posterioir pdf defined by $(1/\sigma^2)^4$ onto the plane $(k_\infty, E)$.

Bootstrapping is not only useful for determining correlation, but, e.g., to observe the projection of the posterior pdf onto planes in the factors space. For example it is interesting to look at the projection of the posterior joint pdf onto the plane $(k_\infty, E)$ in Fig. 7. The non-linear shape of the scatter plot indicates that the effective internal combination of $k_\infty$ and $E$ inside the model is more complicated than the simple relationships $(E - k_\infty) = \text{const}$ or $(E/k_\infty) = \text{const}$ (i.e. linear relationships). In fact, the right relationship should be $[\log(k_\infty)/E] = \text{const}$.

Qualitatively similar projections could also be obtained by applying a Monte Carlo filtering. However, likelihood measures allow using all the runs and therefore a larger amount of information. Finally, it is worth noting that the bootstrapping approach can have a general application, whatever the method used for the representation of the interaction structure.

#### *3.3.3. Conclusions about the representation of the interaction structure*

The performance of a GSA on the likelihood measure does not necessarily provide a complete representation of the interaction structure. The clear representation of the interaction structure is something additional with respect to the performance of a GSA and can be in some cases a formidable task. This task usually requires the use of computationally intensive methods and/or the formulation of hypotheses about the interaction structure and the introduction of a certain degree of arbitrariness for such representation.

On the other hand, the application of a GSA provides a quantitative evaluation about fundamental aspects of the calibration problem, such as:

- which factors are important for calibration, i.e. are somehow conditioned by observations;
- the degree of complexity of the interaction structure;
- which factors are involved in the interaction structure.

Such information has a general validity, since it is obtained without assumptions about the model structure and/or the error structure. So, even if it does not provide a complete representation of the interaction structure, GSA reveals some general and basic properties of such a structure, which are common to any more detailed representation and which are not affected by any “modeler’s prejudice”.

### *3.4. Further analysis by varying temperature in the data set: fewer interactions in the model*

Let us now consider the same chemical system, but assume that 9 sets of observations are available at 9 different temperatures: in particular, we considered five measurements at each temperature for a total of 45 observations. The new pseudo-experimental data are shown in Fig. 8. It is assumed that the temperature of each observation is known so as the model contains always three factors for calibration. The likelihood measure is the inverse of the mean square difference between the model and experiments over the nine time series.

Sobol’ sensitivity indices have been computed for the likelihood measure of the model simulations and are shown in Table 10. Posterior joint pdf correlation matrix is shown in Table 11.

As expected, when the temperature range of the different experimental measurements is varied significantly, the interaction between the kinetic factors is strongly reduced. Correspondingly the absolute values of the first order sensitivity indices become much larger, summing up to almost one. Under the particular operating conditions chosen, the influence of the two kinetic factors does not split uniformly, but concentrates on $k_\infty$. This is not a general result. What has to be expected in general is the decrease in the interaction of the model as a whole. This means that the ‘posterior’ pdf structure can be described elemen-

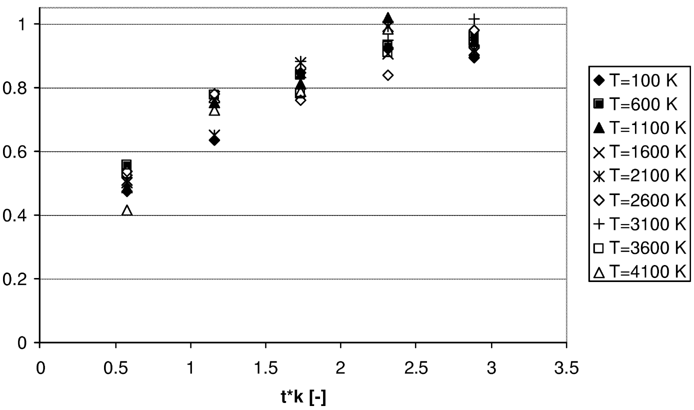


Fig. 8. Experimental data at 9 different temperature vs. dimensionless time-scale.

Table 10
Sobol' sensitivity indices obtained performing the GSA-GLUE analysis with the new observation data set (Fig. 8)

| | 1st order | Tot order |
|---|---|---|
| $k_\infty$ | 0.63072 | 0.68821 |
| $E$ | 0.07971 | 0.11696 |
| $y_B^0$ | 0.29506 | 0.3192 |

Table 11
Correlation matrix of the posterior joint pdf obtained performing the GSA-GLUE analysis with the new observation data set (Fig. 8)

| | $k_\infty$ | $E$ | $y_B^0$ |
|---|---|---|---|
| $k_\infty$ | 1 | 0.0366 | −0.0432 |
| $E$ | 0.0366 | 1 | 0.0079 |
| $y_B^0$ | −0.0432 | 0.0079 | 1 |

tary, as a summation of first order effects. Moreover, in this case, estimation capability can also be more precisely assessed, as in the classical factor estimation problems. In the present case, the Arrhenius pre-exponential factor will be very well estimated, while the activation energy is not well determined, not because of under-determination, but because it does not affect significantly the 'objective' function.

## 4. Conclusions

In the present paper, a new methodological approach is presented applicable in the calibration of models with distributed output. The methodology consists of a combination of the GLUE and GSA methodologies. GSA allows a quantitative assessment of model factors mainly driving model behavioral runs. The use of GLUE, through the definition of a likelihood measure for each model run, allows the performance of a GSA conditioned to observations. The likelihood measure provides an estimate of the posterior joint pdf of the input factors and its in-depth analysis allows a description of the interaction structure between factors, connected to model over-parameterization. GSA allows a quantitative decomposition of the likelihood variance with respect to the input factors, including high order terms. Factors providing negligible contributions to the likelihood variation can be clearly identified, allowing the modeler to exclude them from the calibration procedure and to fix them at a nominal value. On the other hand, factors having a significant impact on the likelihood measure (either as a main effect or as a total effect in interaction with all the other factors) have to be accounted in calibration, since they are able to drive behavioral runs of the model.

Model over-parameterization usually implies that factors important for calibration hardly have an effect identifiable through elementary structures. On the other hand, a highly complex interaction structure is usually present. The analysis of the interaction structure is a challenging problem and a general method for assessing the posterior joint pdf is hardly to be identified. Some degree of arbitrariness in the construction of such tree-structures [5] or Bayesian networks [20] cannot be avoided. Global SA can be very useful in this context, since it provides quantitative criteria for choosing 'leading' factors based on main and total effect. Such criteria do not necessarily provide a direct, complete representation of the interaction structure. However, the advantage of variance based GSA is that it makes few assumptions on the structure of the errors and of the input–output mapping. So, GSA results can be taken as a common denominator to all other tools applied to represent the interaction structure.

## Acknowledgements

This work has been partially funded by European Commission, General Directorate Information Society, IST Program, through contract IST-1999-11313 (IMPACT project).

## References

[1] K.J. Beven, A. Binley, Hydrological Processes 6 (1992) 279–298.
[2] R. Romanowicz, K.J. Beven, J. Tawn, in: V. Barnett, F. Turkman (Eds.), Statistics for the Environment 2, Water Related Issues, Wiley, New York, 1994, pp. 297–315.
[3] G.M. Hornberger, R.C. Spear, Water Res. 14 (1980) 29–42.
[4] R.C. Spear, G.M. Hornberger, Water Res. 14 (1980) 43–49.
[5] R.C. Spear, T.M. Grieb, N. Shang, Water Resources Res. 30 (1994) 3159–3169.
[6] R.C. Spear, Environmental Modelling Software 12 (1997) 219–228.
[7] G.M. Hornberger, R.C. Spear, J. Environmental Management 12 (1981) 7–18.
[8] G. Archer, A. Saltelli, I.M. Sobol', J. Statist. Comput. Simulation 58 (1997) 99–120.
[9] A. Saltelli, K. Chan, M. Scott (Eds.), Sensitivity Analysis, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, 2000.
[10] R. Romanowicz, H. Higson, I. Teasdale, Environmetrics 11 (2000) 351–371.
[11] M.J.W. Jansen, W.A.H. Rossing, R.A. Daamen, in: G. van Straten (Ed.), Predictability and Nonlinear Modelling in Natural Sciences and Economics, 1994, pp. 334–343.
[12] R.I. Cukier, C.M. Fortuin, K.E. Schuler, A.G. Petschek, J.H. Schaibly, J. Chem. Phys. 59 (1973) 3873–3878.
[13] I.M. Sobol', Mathematical Modelling Comput. Experiment 1 (1993) 407–414.
[14] R.L. Iman, S.C. Hora, Risk Anal. 10 (1990) 401–406.
[15] T. Homma, A. Saltelli, Rel. Eng. Syst. Safety 52 (1996) 1–17.
[16] A. Saltelli, S. Tarantola, K.P.-S. Chan, Technometrics 41 (1999) 39–56.
[17] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York, 1974.
[18] I.M. Sobol', USSR Comput. Math. Math. Phys. 7 (1967) 86–112.
[19] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, New York, 1981.
[20] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman Publishers, San Mateo, CA, 1988.