Deliverable 18

Model assessment in integrated procedures for environmental impact evaluation: software prototypes

December 2001

Version: 1
Security: Int.
Status: Final
Prepared by: M. Ratto, A. Saltelli
Joint Research Centre of European Commission,
Institute for the Protection and Security of the Citizen, Ispra
TABLE OF CONTENTS

1 INTRODUCTION .................................................................................................................................. 4
   1.1 Contribution of the present deliverable to the overall goals of the IMPACT project .......................................................................................................................... 4
   1.2 List of publications ........................................................................................................................ 6

2 INTRODUCTION TO THE GLUEWIN CODE .............................................................................. 7
   2.1 Summary of the GSA-GLUE methodology ................................................................................... 9
      2.1.1 Short description of GLUE ..................................................................................... 9
      2.1.2 Short description of variance-based GSA ...................................................................... 10
      2.1.3 Combined GSA-GLUE approach .............................................................................. 12

3 GLUEWIN USER’S MANUAL ........................................................................................................ 13
   3.1 Installation of GLUEWIN .......................................................................................................... 13
   3.2 GLUEWIN Main Window ............................................................................................................. 15
      3.2.1 Define the type of input ............................................................................................... 16
      3.2.2 Format of data files ....................................................................................................... 17
      3.2.3 Default likelihood measures ....................................................................................... 19
      3.2.4 The command frame ..................................................................................................... 21
   3.3 The "Dotty plots" window ........................................................................................................... 23
      3.3.1 Figure commands .......................................................................................................... 23
      3.3.2 The File menu .............................................................................................................. 26
      3.3.3 The Likelihood menu ..................................................................................................... 27
      3.3.4 The Variable menu ........................................................................................................ 28
   3.4 The "Uncertainty analysis" window ............................................................................................. 30
      3.4.1 Uncertainty analysis of model outputs at fixed time points .......................................... 30
      3.4.2 Figure commands for the 'Fixed time' mode .................................................................... 31
      3.4.3 Analysis of the posterior distribution of input factors ..................................................... 32
      3.4.4 Uncertainty analysis of the whole Time Series ............................................................... 33
      3.4.5 Figure commands for the 'Time series' mode ............................................................... 34
      3.4.6 The Type menu .............................................................................................................. 35
3.4.7 The Posterior menu ....................................................................................... 36
3.4.8 Other menus ................................................................................................. 36

3.5 The "Sensitivity analysis" window ...................................................................... 37
3.5.1 Figure commands ......................................................................................... 38
3.5.2 Figure menus ................................................................................................. 39

3.6 Transform likelihoods ........................................................................................ 40
3.6.1 Transform options ......................................................................................... 40
3.6.2 Figure commands ......................................................................................... 43
3.6.3 The Likelihood menu .................................................................................... 45

3.7 Combine likelihoods .......................................................................................... 46
3.7.1 Combine figure commands ........................................................................... 47
3.7.2 Combine figure menus .................................................................................. 47

3.8 Save likelihood files ............................................................................................ 49

3.9 Filter output variables ....................................................................................... 50
3.9.1 Filter figure commands ................................................................................. 50
3.9.2 The Variable menu ........................................................................................ 51

3.10 Save filtered data .............................................................................................. 52

3.11 Analyse the posterior distributions of input factors .......................................... 53
3.11.1 Bootstrap sub-panel ..................................................................................... 53
3.11.2 View sample space sub-panel .................................................................... 54

3.12 The DEMO files .................................................................................................. 58

4 CONCLUDING REMARKS ........................................................................... 60

5 REFERENCES ........................................................................................... 61

- 3 -


1 Introduction

1.1 Contribution of the present deliverable to the overall goals of the IMPACT project

Procedures for the appraisal of environmental impact include the application of mechanistic models and statistical procedures. The main goal of JRC activities in framework of the IMPACT project is to identify and apply Uncertainty and Sensitivity Analysis (UA-SA) techniques for the assessment of the models used in such procedures.

The role of UA-SA methods in framework of the IMPACT project is two-fold:

✓ the intermediate step in the merging of mechanistic models and statistical procedures, in which the original model is reduced by allowing unimportant factors to be fixed or eliminated;

✓ the tool for making more 'transparent' the final environmental appraisal, in which a global uncertainty estimation is performed and the relative importance of different hypothesis, assumptions, model structures/statistical procedures, as well as of parameter/data uncertainty on such appraisal is quantified.

The theoretical and methodological principles of the UA-SA techniques useful in the IMPACT project have been presented in Deliverables 16-17 (Dec. 2000) [1,2]. In particular the combined use of the Bayesian uncertainty estimation technique called GLUE [3, 4] and of Global Sensitivity Analysis techniques seemed the most effective approach, since it allows performing an UASA-based model assessment, in which all the information included observed trends of natural systems can be used [5].

In the present deliverable a software prototype is presented, in which advanced GSA methodologies and the Bayesian uncertainty estimation are embedded in a user-friendly workspace, with the aim of making the methodology developed within IMPACT easily applicable for potential end-users.
The user-friendly workspace makes use of the already available SIMLAB software, developed by the Applied Statistics sector of JRC [6]. SIMLAB provides an interface for UA/SA applications: generation of Monte Carlo samples designed for different types of SA (factor screening, regression based SA, FAST, extended FAST, etc.), uncertainty and sensitivity analysis of the model output.

Moreover, a new code called GLUEWIN has been developed, which integrates the GLUE-type Bayesian uncertainty estimation in the 'classical' GSA framework provided by SIMLAB. The new features of GLUEWIN allow an easier time-domain model output analysis and the inclusion of the information of observed patterns, which is missing in the SIMLAB workspace. In other words, GLUEWIN allows extending the application of the available UA/SA methodologies in a time series-normalisation framework. The GLUEWIN code has been written in MATLAB™ language but can be also run in a stand-alone mode.

Main features of the GLUEWIN software are:
1) the possibility of defining likelihood weights for model runs;
2) the use of likelihood weights for the uncertainty estimation of model predictions in a time domain;
3) the analysis of the posterior distributions of the uncertain parameters obtained from conditioning to observations (marginal distributions, covariance structure, bootstrap sampling from the posterior joint pdf, 2D projections of the joint pdf);
4) 'visual' SA through scatter plots and cumulative distributions (Smirnov type);
5) an entirely compatible data interface with SIMLAB: e.g. files of likelihood weights created with GLUEWIN can be directly loaded by SIMLAB for GSA-GLUE analysis.

The code has been applied in the development of the test cases of IMPACT involving GSA: in particular we applied it for the GSA of the WAMPUM model by GKSS in the Elbe River case study. Moreover, the analysis of the Transfer Function (TF) model identified by ULANC partners for the same system has been also performed applying the GLUEWIN code [17,8].
1.2 List of publications

Paper presented in international conferences and symposia.


Papers published in the open literature and in preparation:


2 Introduction to the GLUEWIN Code

GLUEWIN is a code designed for analysing the output of Monte Carlo runs when empirical observations of the model output are available and implements the combined GSA-GLUE methodology identified and applied by JRC within the IMPACT project [1,5]. In particular, it is carefully designed to allow easy analysis of time dependent model outputs (e.g. for time series normalisation or analysis of temporal trends in the state of the environment).

A Monte Carlo analysis is based on performing multiple model evaluations with probabilistically selected model input, and then using the results of these evaluations to determine both the uncertainty in model predictions (uncertainty analysis, UA) and the input variables that give rise to this uncertainty (global sensitivity analysis, GSA). Moreover, in the GLUE framework, observations for the model output are available, so that the uncertainty estimation should also account for conditioning to observations. In the same manner, GSA must account for this conditioning.

In general, a Monte Carlo analysis involves five steps. In the first step, a range and distribution are selected for each input factor. These selections will be used in the next step in the generation of a sample from the input factors. If the analysis is primarily of an exploratory nature, then rather crude (e.g., uniform and log-uniform, triangular) distribution assumptions may be adequate. In the second step, a sample is generated from the ranges and distributions specified in the first step. The result of this step is a sequence of sample elements. In the third step, the model is evaluated for each sample element. In essence, these model evaluations create a mapping from the inputs to the results that can be studied in subsequent uncertainty and sensitivity analysis. In the fourth step, the results of the third step are used as the basis for an uncertainty analysis. The fifth and final step is the sensitivity analysis.

GLUEWIN takes advantage of the already available software SIMLAB 1.1, developed by the JRC staff [6]. SIMLAB provides an interface for UA/SA applications: generation of Monte Carlo samples designed for different types of SA (factor screening, regression based SA, FAST, extended FAST, etc.), uncertainty and sensitivity analysis of the model output. SIMLAB is a 'general' UA/SA tool and all aspects regarding conditioning to
observations and Bayesian analysis are missing. Such additional tools are provided by GLUEWIN.

GLUEWIN implements only the fourth and, partially, the fifth steps of the Monte Carlo procedure and assumes that the user is able to generate a sample for the model parameters (e.g. using SIMLAB), to run the model and to generate a sample of the model output. Once the first three steps of the Monte Carlo procedure are implemented, GLUEWIN enables to perform Bayesian uncertainty and some few kinds of Global Sensitivity Analysis. Moreover, if the sample has been generated with SIMLAB, it is possible to apply the advanced GSA tools available in SIMLAB for completing the GSA-GLUE analysis. In the Bayesian framework, GLUEWIN can be also used to create new samples for the Monte Carlo analysis, by applying a bootstrap strategy.

The layout of the code is inspired by the freeware GLUE package by K. Beven (the author of the GLUE methodology), available at the Lancaster University web site on the page [9]:

http://www.es.lancs.ac.uk/hfdg/glue.html

The downloadable version of the GLUE package is a teaching tool and has strong limitations: it accepts up to 6 input parameter, 6 basic likelihood measures, 10 predicted variables and 1000 Monte Carlo simulations. Only single predicted variables are handled and not time series of predictions. The GLUEWIN code developed by JRC, on the other hand, has no limitations in the number of parameters, of likelihoods and predicted variables and in the number of Monte Carlo simulations. Moreover, the GLUEWIN code handles time series of predictions. The only limitations are given by the PC memory on which the code is run. GLUEWIN has additional tools, such as the analysis of the posterior distribution of the input factors, a bootstrapping tool, and can link with the SIMLAB software, for GSA purposes.
2.1 **Summary of the GSA-GLUE methodology**

2.1.1 **Short description of GLUE**

In the last decade, a method based on the concept of Bayesian Inference for uncertainty estimation, has been used in hydrology as the Generalised Likelihood Uncertainty Estimation Technique (GLUE) \[3,4\]. The GLUE technique is an extension of the Generalised Sensitivity Analysis methodology, which has now come to be called Regional Sensitivity Analysis (RSA), by R.C Spear and G.M. Hornberger \[10,11\]. 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 realisations 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-behavioural and rejected.

The GLUE procedure is based upon making a large number of runs of a given model with different sets of factor values, chosen randomly from specified 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. Different possible likelihood measures have been defined in literature \[3,4,12-13\], one of them is:

\[
L(\theta_i \mid Y) \propto \exp\left(-\frac{\sigma_i^2}{\sigma_0^2}\right)
\]

where

\[
\sigma_i^2 = \frac{1}{Nobs} \sum_{j=1}^{Nobs} \left(\hat{Y}(t_j) - Y(t_j)\right)^2
\]

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.

2.1.2 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. [14].

When using variance based techniques [15], the SA is based on estimating the fractional contribution of each input factor to the variance of the model output. In order to calculate the sensitivity indices for each factor, the total variance $V$ of the model output is decomposed as

$$ V = \sum_i V_i + \sum_{i \neq j} V_{ij} + \sum_{i \neq j \neq m} V_{ijm} + \ldots + V_{12...k} $$  \hspace{1cm} (3)

where

$$ V_i = V(E(Y|X_i = x_i^*)) $$  \hspace{1cm} (4)

$$ V_{ij} = V(E(Y|X_i = x_i^*, X_j = x_j^*)) - V(E(Y|X_i = x_i^*)) - V(E(Y|X_j = x_j^*)) $$  \hspace{1cm} (5)

and so on. In the above formulas, $Y$ denotes the output variable, $X_i$ denotes an input factor, $E(Y|X_i = x_i^*)$ denotes the expectation of $Y$ conditional on $X_i$ having a fixed value $x_i^*$, and $V$ stands for variance over all the possible values of $X_i$. The same variance decomposition underlies the theory of experimental design [16]. The decomposition is unique if the $X_i$ are independent from each other. The sensitivity index $S_i$ for the factor $X_i$ is defined as $V_i / V$. The reason for that is intuitive: if the inner mean $E(Y|X_i = x_i^*)$ varies considerably with the selection of a particular value $x_i^*$ for $X_i$, while all the effects of the $X_j$'s, $j \neq i$ are being averaged, then surely factor $X_i$ is an
influential one. Estimation procedures for $S_i$ are the Fourier Amplitude Sensitivity Test, FAST [17], the method of Sobol’ [18], and others [19].

Higher order sensitivity indices, responsible for interaction effects among factors, are rarely estimated in computational experiments, as in a model with $k$ factors the total number of indices (including the $S_i$’s) that should be estimated is as high as $2^k - 1$. This problem is sometimes referred to as the curse of dimensionality. However interactions may have a strong impact on the output uncertainty especially when $k$ is large and factors are varied on a wide scale, as often happens in numerical modelling.

A method, which is able of accounting for interactions and simultaneously coping with the curse of dimensionality, is the extended FAST [20]. The extended FAST can yield estimates of the total sensitivity indices. $S_{Ti}$ defined as the sum of all the indices ($S_i$ and higher orders) where a given factor $X_i$ is included. This concentrating in one single term all the effects involving $X_i$. For additive models, $S_i = S_{Ti}$ for all the factors $X_i$. The estimation of the total sensitivity indices $S_{Ti}$ makes the analysis affordable from a computational point of view, as we only need $k$ total indices for decomposing quantitatively the output variance $V$. Furthermore, the extended FAST allows the simultaneous evaluation of the first and total effect indices. The estimation of the pair $(S_i, S_{Ti})$ is important to appreciate the difference between the impact of factor $X_i$ alone on $Y$ ($S_i$) and the overall impact of factor $X_i$ through interactions with the others on $Y$ ($S_{Ti}$). For a 3-factor model, the three total sensitivity indices are:

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{123}$$
$$S_{T2} = S_2 + S_{12} + S_{23} + S_{123}$$
$$S_{T3} = S_3 + S_{13} + S_{23} + S_{123}$$

(6)

where now each $S_{k,d_2,...,d_i}$ is simply $\frac{V_{k,d_2,...,d_i}}{V}$. Clearly the $V_{k,d_2,...,d_i}$ add up to $V$, and the corresponding $S_{k,d_2,...,d_i}$ add up to one; this is not true for the $S_{Ti}$’s, but a normalisation factor can be introduced.
Variance based methods such as Sobol’ and the extended FAST display a number of attractive features for SA:

- Model independence: the sensitivity measure is model independent. It works for non-linear and non-additive models, unlike methods based on linear regression such as the standardised regression coefficients [21].
- The measure captures the influence of the full range of variation of each factor.
- The measure captures interaction effects; this can be a crucial issue for a design problem, or for a risk analysis study.
- The methods can treat “sets” of factors as one single factor.

2.1.3 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.

3.1 Installation of GLUEWIN

The GLUEWIN code is written in MATLAB™ language (Release 12), but is also available in a stand-alone mode, obtained from the MATLAB C/C++ Compiler Toolbox. GLUEWIN is a collection of 67 functions, providing Graphical User Interface for loading data, perform the GSA-GLUE analysis, view results, save data and graphs.

The list of routines is given in Table 1. Users need to copy the GLUEWIN files in a separate directory. In the case of the MATLAB version, the GLUEWIN directory should be put in the MATLAB path.

<table>
<thead>
<tr>
<th>about.m</th>
<th>opensam.m</th>
<th>setFigUncfix.m</th>
</tr>
</thead>
<tbody>
<tr>
<td>bayes.m</td>
<td>outfix.m</td>
<td>setFigUncpost.m</td>
</tr>
<tr>
<td>bootstrap.m</td>
<td>outtime.m</td>
<td>sobcom0.m</td>
</tr>
<tr>
<td>boxplot.m</td>
<td>pick.m</td>
<td>sobolGraph.m</td>
</tr>
<tr>
<td>calcBayes.m</td>
<td>reject.m</td>
<td>sobolTable.m</td>
</tr>
<tr>
<td>changeDefault.m</td>
<td>rejout.m</td>
<td>spop0.m</td>
</tr>
<tr>
<td>changeNameGlue.m</td>
<td>restoreDefault.m</td>
<td>spop1.m</td>
</tr>
<tr>
<td>changeNameLik.m</td>
<td>sa.m</td>
<td>spop1_andrea.m</td>
</tr>
<tr>
<td>changeNameOutput.m</td>
<td>saveCum.m</td>
<td>spop_classic0.m</td>
</tr>
<tr>
<td>combs0.m</td>
<td>saveUnc.m</td>
<td>spop_andrea.m</td>
</tr>
<tr>
<td>dottycomp.m</td>
<td>savevars.m</td>
<td>stat.m</td>
</tr>
<tr>
<td>dottyplo.m</td>
<td>scatbarPlo.m</td>
<td>transfor.m</td>
</tr>
<tr>
<td>dottyvar.m</td>
<td>senscomp.m</td>
<td>transout.m</td>
</tr>
<tr>
<td>echoCoord.m</td>
<td>senslik.m</td>
<td>ua.m</td>
</tr>
<tr>
<td>endGlue.m</td>
<td>sensvar.m</td>
<td>uncfix.m</td>
</tr>
<tr>
<td>fileMenu.m</td>
<td>setFigBayes.m</td>
<td>uncfixedcomp2.m</td>
</tr>
<tr>
<td>glue2simlab.m</td>
<td>setFigBootstrap.m</td>
<td>uncpost.m</td>
</tr>
<tr>
<td>glue2sob.m</td>
<td>setFigGluewin.m</td>
<td>uncser.m</td>
</tr>
<tr>
<td>gluewin.m</td>
<td>setFigSensVar.m</td>
<td>uncsercomp2.m</td>
</tr>
<tr>
<td>likeli1.m</td>
<td>setFigTrans.m</td>
<td>viewRun.m</td>
</tr>
<tr>
<td>openfile.m</td>
<td>setFigTransout.m</td>
<td>whist.m</td>
</tr>
<tr>
<td>openout.m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Moreover, in order to enable to link to the SIMLAB software, the following further files are needed:

- glue_path.dat
- simlab.lnk

The first contains information on the location of the GLUEWIN program, while the second is a Windows link to the SIMLAB executable file, which should be already installed on the system. The first file is generated when installing the GLUEWIN package.

(running the batch file install.bat for the stand-alone version, pinstall.bat for the MATLAB version). The second one must be created by the user in the GLUEWIN root directory.

To run GLUEWIN in MATLAB, the GLUEWIN directory must be put in the MATLAB path. Typing <gluewin> at the command prompt, the program starts.
3.2 **GLUEWIN Main Window**

To start GLUEWIN, the user can either type the `gluewin` command at the MATLAB prompt or start the `gluewin.exe` stand-alone executable file in Windows. Starting GLUEWIN prompts the following window.

![GLUEWIN starting panel](image)

**Figure 1.** GLUEWIN starting panel.

Input data needed are:

- a sample matrix for the input parameters (dimension \( n \times k \), where \( n \) is the number of Monte Carlo runs and \( k \) is the number of factors),
- a matrix of model outputs (dimension \( n \times m \times t \), where \( m \) is the number of model outputs, \( t \) is number of the time points at which model output is computed, if \( t=1 \), the model output is a scalar),
- a matrix of observations of the model output (dimension \( n_{obs} \times m_{obs} \times t_{obs} \), where \( n_{obs} \) is the number of experimental sets available, \( m_{obs} \leq m \) is the number of model outputs
for which observations are available, \( t_{obs} \) is the number of time values at which observations are taken; 

✓ a matrix of likelihood weights (dimension \( n \times l \), where \( l \) is the number of likelihood measures computed).

### 3.2.1 Define the type of input

User can choose among three modes of feeding data to the program, in the upper frame of the window (‘Define type of input’).

1. *Likelihood is given*: this case applies when likelihood measures are computed by the model together with the model output, so that a file of likelihood weights is already available. In this case the matrix of observations is an optional input to the program, but not necessary (likelihood measures are already computed). Clicking on the 'Likelihood is given' button, the loading of all four types of ASCII-file is enabled.

2. *Measurements are given*: this case applies when the model computes no likelihood measure. In this case, GLUEWIN loads the matrix of model outputs and observations, and computes default likelihood functions to enable the user starting the analysis. Clicking on the 'Measurements are given' button, the loading of a likelihood file is not possible.

3. *Saved analysis*: this case applies when the user has already run GLUEWIN for a particular case study and has saved all the data in a MATLAB binary mat-file (extension *.mat), by clicking on the 'Save analysis' button in the 'Command frame' in the right-hand part of the window. This option allows a much quicker data loading, with respect to ASCII files. Clicking on the 'Load saved analysis' button, enables only the loading of one mat-file (upper button in 'Select data to analyse' frame), containing all information saved in a previous GLUEWIN session.

Once the type of input has been selected, clicking on the four buttons in the 'Select data to analyse' frame, allows the user to browse the computer drives to find the files to load. When all required files have been chosen, click the 'OK' button to load the files. After
files are loaded, the buttons in the command frames in the right-hand part of the window are enabled.

### 3.2.2 Format of data files

For the first two modes of feeding data, the user must supply data in ASCII files. The format of such data is the same as for the SIMLAB software.

**Sample file**

If the sample has been created with SIMLAB, GLUEWIN can load it. The format of the SIMLAB sample file has 4 flags before the matrix of data (the second is the sample dimension, the third is the number of parameters)

```
0
500
3
0
-0.9924140755   -0.1212273205   -0.9982218419
0.2443174323   -0.4752461143   -0.9748393885
0.402405846   0.2402597332   -0.4863455815
...       ...       ...
...       ...       ...
```

If the user has created is own sample file, this should have the following format, as required for SIMLAB in the case of external samples: the sample dimension, the number of factors, the list of factor names, and the matrix of data.

```
64
3
fac1
fac2
fac3
-3.090232264   -3.090232264   -3.090232264
3.82550629e-007   3.82550629e-007   3.82550629e-007
-0.6729169549   0.6729172578   -0.6729169549
0.6729172578   -0.6729169549   0.6729172578
-1.146713641   0.3179800352   1.146714693
0.3179800352   -1.146713641   -0.3179804109
-0.3179804109   0.3179800352   0.3179800352
1.146714693   1.146714693   -1.146713641
...       ...       ...
...       ...       ...
-0.83909755938   0.9442102089   0.1175329646
2.129808813   -0.4441455026   -1.666758199
```
Model results file

In the case of scalar output, the format of the file must be the following: the number of model outputs, the list of model output names, the message "time=no", the number of model runs, the matrix of model outputs.

```
3
outvar1
outvar2
outvar3
time = no
64
1.146714693  1.146714693  -1.146713641
-1.5270442   1.527045317   0.4877169824
0.1569936491 -0.1569936491 -0.8848267589
-0.487717548  -0.8848267589  1.527045317
0.8848260558  0.4877169824  -0.1569936491
-0.8848267589 -0.487717548  -0.487717548
0.4877169824  0.8848260558  0.8848260558
-0.1569936491  0.1569936491  -1.5270442
...  ...  ...
...  ...  ...
```

In the case of time dependent output, the format for the file should be the following: the number of model outputs, the list of model output names, the message "time=yes", the number of model runs. Then, for each model run, the cardinality must be quoted as RUN 0 ... RUN (n-1), then the number of time points for which the model outputs have been computed and finally the matrix of model outputs itself. In the first column of the matrix the time values are given (i.e. \( m + 1 \) columns).

```
1
KP
time = yes
4
RUN 0
2
0.1 1.37318
0.177828 1.4313
RUN 1
2
0.1 0.560303
0.177828 0.641186
...  ...
...  ...
```

Different model runs can have different numbers of time points, so this format allows dealing with simulation models with adaptive/variable time step. In the case when
adaptive/variable time step has been applied, after reading the file, GLUEWIN interpolates (linear interpolation) all data, in order to have common time co-ordinates for all model runs. Interpolation is made based on the model run having the largest number of time steps.

*Observations file*

The format of the observation file is the same as the model output file (either scalar or time dependent). Instead of the number of model runs, users must quote the number of experimental sets. If more than one model output is given, the observed variables must be ordered by column in the same order as the model outputs. Moreover, if observations are not available for all model outputs, the model outputs must be ordered in such a way that outputs, for which observations are available, are placed in the first columns, while outputs without observations are placed in the far-right columns.

When the 'Measurements are given' button has been clicked, likelihood measures are computed, based on mean square errors and absolute errors. To compute such errors, model outputs are interpolated on the time co-ordinates of the observations.

*Likelihood file*

The format of the likelihood file is the same as for a scalar model output.

*Saved analysis file*

This must be a mat-file saved in a previous session of GLUEWIN, after clicking on the button 'Save analysis'.

### 3.2.3 Default likelihood measures

When the 'Measurements are given' button has been clicked, GLUEWIN computes default measures. GLUEWIN allows users to derive other types of measures using commands available in the various sub-windows. The default measures are computed as follows.

*Scalar output with multiple experiments*

The default likelihood function computed is defined as:
where

\[ x_i \] is the \( i \)-th element of the sample of the model input factors.

\[ y = (y^{(1)}, ..., y^{(i)}, ..., y^{(n_{obs})}) \] is the vector of observations of the scalar variable \( y \).

\[ \hat{y}(x_i) \] is the \( i \)-th model run.

\[ \text{std}(\hat{y} - y^{(j)}) \] is the weighting factor in the sum of square errors, equal to the standard deviation of the model errors with respect to the \( j \)-th observations.

**Time dependent output with one set of experiments**

Two basic measures are computed: the mean square error \( \text{MSE} \) and the mean absolute error \( \text{MAE} \):

\[
\text{MSE}(x_i) = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{y}_i(x_i) - y_t \right)^2
\]

\[
\text{MAE}(x_i) = \frac{1}{T} \sum_{t=1}^{T} |\hat{y}_i(x_i) - y_t|
\]

where \( T \) is the number of time points for which observations are available.

Moreover a default likelihood weight is computed:

\[
L(x_i | y_1, ..., y_T) = \exp \left( - \frac{\text{MSE}(x_i)}{\min(\text{MSE})} \right)
\]
Time dependent output with multiple sets of experiments

Two basic measures are computed: the mean square error $MSE$ and the mean absolute error $MAE$, defined as the weighted sum over the $n$ observation sets of the mean square errors and the mean absolute errors.

\[
MSE(x_i \mid y^{(1)}..., y^{(j)}..., y^{(n_{obs})}) = \sum_{j=1}^{n_{obs}} \frac{MSE(x_i \mid y^{(j)})}{\text{std}(MSE(X \mid y^{(j)}))}
\]

\[
MAE(x_i \mid y^{(1)}..., y^{(j)}..., y^{(n_{obs})}) = \sum_{j=1}^{n_{obs}} \frac{MAE(x_i \mid y^{(j)})}{\text{std}(MAE(X \mid y^{(j)}))}
\]

(10)

where the weights are given by the standard deviation of all the model errors with respect to the $j$-th set of data, $X$ is the whole sample matrix, and

\[
MSE(x_i \mid y^{(j)}) = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{y}_t(x_i) - y_t^{(j)} \right)^2
\]

\[
MAE(x_i \mid y^{(j)}) = \frac{1}{T} \sum_{t=1}^{T} |\hat{y}_t(x_i) - y_t^{(j)}|
\]

(11)

are the mean square error and the mean absolute error for the $j$-th set of observations.

3.2.4 The command frame

Once all data have been loaded, the buttons in the command frame become active. The GLUEWIN main window remains always open and active even when other specific windows are opened using the command buttons. This enables the user to start different types of analysis.

Dotty plots

Clicking on this button a new window is opened showing scatter plots of likelihood measures or model outputs (at fixed time) versus the various input parameters. If a 'Dotty plots' window was already opened, clicking on the button makes the 'Dotty plots' window visible on the top of the desktop.

Uncertainty analysis
Clicking on this button a new window is opened for the uncertainty analysis. If an 'Uncertainty analysis' window was already opened, clicking on the button makes the 'Uncertainty analysis' window visible on the top of the desktop.

Sensitivity analysis
Clicking on this button a new window is opened for qualitative sensitivity analysis. If a 'Sensitivity analysis' window was already opened, clicking on the button makes the 'Sensitivity analysis' window visible on the top of the desktop.

SIMLAB
With this button, a SIMLAB session is opened.

Sobol' indices
With this button, the computation of Sobol' sensitivity indices can be performed. The Sobol' indices are not currently computed with SIMLAB. This is a temporary button and will be eliminated as soon as the Sobol' method will be implemented in SIMLAB. To compute Sobol' indices requires that the user has generated a suitable sample for the Sobol' method.

Save analysis
With this button, the user can save into a mat-file (binary format) all the data available in the current GLUEWIN session (including all modifications to likelihood, filtering of model outputs, bootstrapping, etc., performed during the session).

Close
With this button, all GLUEWIN windows are closed and the GLUEWIN session is ended.
3.3 The "Dotty plots" window

The 'Dotty plots' window shows scatter plots for likelihood weights and output variables at fixed time points. By default, the figure shows the scatter plots for the first likelihood vs. the first 6 parameters (if $k>6$). The maximum value of the likelihood/variable is highlighted with a red dot. In the lower part of the figure, some information is given: the name of the likelihood/variable plotted, the range of the likelihood/variable values and the time range if the output variable is time dependent.

![Dotty plots window of the GLUEWIN code.](image)

Figure 2. Dotty plots window of the GLUEWIN code.

3.3.1 Figure commands

Change the scale of the scatter plot

Clicking with the mouse right-button inside the axes area (but NOT on the dots), activates a menu for changing the scale of the plot, with four choices: normal, semilog-x, semilog-y, log-log.
View single run

Clicking with the mouse right-button on the dots of one subplot, allows highlighting with a red circle the particular combination of parameter values in all scatter graphs (see red circles in Figure 2). Moreover a menu is opened reporting the number of the run corresponding to the selected dot. Clicking on the Run number, or choosing another custom run, opens a new figure showing the result of that run.

Edit the name of the input factors

Each subplot of the figure shows the name of the parameter. Clicking with the mouse right-button on the parameter name allows editing it.
**Edit the name of the plotted likelihood/variable**

Clicking with the mouse right-button on the name of the plotted likelihood/variable allows editing it.

<table>
<thead>
<tr>
<th>Dotty plots for</th>
<th>Variable:</th>
<th>Change model output name</th>
</tr>
</thead>
<tbody>
<tr>
<td>With range from</td>
<td>0</td>
<td>to: 17.73</td>
</tr>
<tr>
<td>Time range from</td>
<td>0</td>
<td>to: 29</td>
</tr>
</tbody>
</table>

**'Close' button**

With this command the 'Dotty plots' figure is closed.

**'Next' button**

If $k>6$, the button 'Next 6' is enabled which allows the user to loop over all $k$ factors (see Figure 2).

**Change the time value of the scatter graph**

When a variable with time dependent output is plotted, an editable field will be available in which the user can set a time value and by clicking the 'OK' button the scatter graph for that time value will be plotted (see Figure 2).

**Change the sample**

If a filtering of the output variable has been done, with which a reduced sample has been obtained, a menu is activated on the top of the figure, which allows choosing the sample to use for creating graphs.
3.3.2 The File menu

The File menu of the figure, allows the performance of three commands: Save as, Export and Print.

Save As…

This command allows saving the Figure as a MATLAB .fig file. The figure is saved in a way that all GLUEWIN GUI commands are removed and only graphs and text are saved. All text fields are saved in MATLAB as 'text' and not as 'uicontrol'. This allows MATLAB users to edit easily graphs and text.

Export …

This command allows saving the Figure in the following formats: bitmap (.bmp), enhanced metafile (.emf), encapsulated postscript (.eps), postscript(.ps) and MATLAB (.fig). In the latter case, text fields are not modified, but are saved as viewed in the figure.
In this case users cannot edit text fields double clicking on the field, but must edit them as 'uicontrols'.

Print
This command allows printing the Figure.

3.3.3 The Likelihood menu
This menu allows the execution of the following commands: change the likelihood to be plotted in the figure, transform the likelihood, combine likelihood weights, save likelihood values on an ASCII file.

Change likelihood to plot
These commands (in the upper part of the menu) allow choosing the particular likelihood to be plotted. Since the likelihood menu is active also when an output variable is plotted, this command also allows switching from the plot of variables to the plot of likelihoods.

Transform
This command opens a new window, which allows transforming the likelihood values and save the transformations (see Section 3.6).
Combine

This command opens a new window, which allows combining likelihood weights and save combinations (see Section 3.7).

Save for GSA-GLUE

This command opens a new window, which allows choosing likelihood weights to be saved on an ASCII file in SIMLAB format (see Section 3.8). This allows using SIMLAB for GSA purposes.

3.3.4 The Variable menu

This menu allows the execution of the following commands: change the model output to be plotted in the figure, filter the model outputs, save filtered data in an ASCII file.

Change variable

These commands (in the upper part of the menu) allow choosing the particular output variable to be plotted. Since the variable menu is active also when a likelihood function is plotted, this command also allows switching from the plot of likelihoods to the plot of variables.

Filter

This command opens a new window, which allows filtering the model outputs and save the filtered samples (see Section 3.9).
Save filtered samples

This command opens a new window, which allows choosing filtered data to be saved on an ASCII file in SIMLAB format (see Section 3.10). Three files will be created containing the filtered matrix of input factors, likelihood weights and model outputs. All these files are saved in SIMLAB format.
3.4 The "Uncertainty analysis" window

The 'Uncertainty analysis' window allows performing the Bayesian uncertainty estimation using the GLUE methodology. It is the core of the GLUEWIN code. It has three modes: uncertainty analysis of model outputs at fixed time points (the default), analysis of the posterior distribution of the input factors, analysis of the whole time series (only for time dependent model outputs). The three modes can be selected using the 'Type' menu of the Figure.

3.4.1 Uncertainty analysis of model outputs at fixed time points

The first default view, plots the uncertainty analysis of the first output variable given the first likelihood measure at the initial time point.

![Uncertainty plot at a fixed time value](image)

**Figure 3.** Uncertainty plot at a fixed time value.
The density plot

The density plot of the model prediction is constructed with 20 bins. The histogram is in normalised units. The height of the bars is given by the sum of the likelihood weights of the model runs falling in that bin. The figure also shows the mean as a vertical red line and, if observations are available, the observed values at that time point are also shown by means of vertical green lines.

The cumulative density plot

The plot shows the cumulative sum of the likelihood weights vs. the ranked model output values. The red dots show the 5% and 95% quantiles, the mean is identified by a vertical blue line, while observations, if available, are shown through vertical green lines.

3.4.2 Figure commands for the 'Fixed time' mode

Save cumulative distribution to file

Clicking with the mouse right-button on the cumulative distribution curve prompts a command allowing the user to save in an ASCII file the data for the distribution. Mean, mode, median, variance and 90% confidence bound are also saved.

Edit the name of the analysed variable

In the lower part of the Figure, the name of the analysed variable is shown. Clicking with the mouse right-button on the name allows editing it.


**Edit the name of the plotted likelihood**

In the lower part of the Figure, the name of the likelihood used to perform the Bayesian uncertainty estimation. Clicking with the mouse right-button on the name allows editing it.

'Close' button

With this command the 'Uncertainty analysis' figure is closed.

### 3.4.3 Analysis of the posterior distribution of input factors

This window has exactly the same layout (Section 3.4.1) and the same window commands (Section 3.4.2) as for the analysis of model outputs at fixed time points and shows the analysis of the posterior distribution of the input factors. However, instead of the 'Variable' menu it has a 'Posterior' menu (see Section 3.4.7). This mode of the Uncertainty Figure allows performing a Bayesian calibration of input parameters, upon conditioning to observations.
3.4.4 Uncertainty analysis of the whole Time Series

This window shows the Bayesian uncertainty estimation for the whole time series. The graph shows the mean and 90% confidence bound of the model prediction. Moreover, if available, observations are also shown with red dots.

**Figure 4.** Uncertainty plot for the analysis of posterior distributions of input factors.
3.4.5 Figure commands for the 'Time series' mode

Save the uncertainty bound to file

Clicking with the mouse right-button on the confidence bound opens a menu for saving the cumulative bounds in an ASCII file. The program saves time values, mean, 5% and 95% quantiles and observed values if present.

Figure 5. Uncertainty figure for a time series.
Change plot type

Clicking with the mouse right-button on the observed data, opens a menu which allows changing the format of the plot of observations and average prediction from dots to bars.

3.4.6 The Type menu

This menu allows switching from the analysis of model output at fixed time to the analysis of the posterior distribution of the input factors to the analysis of the whole time series (only in the case of time dependent model outputs).
3.4.7 The Posterior menu

This menu is enabled when the mode 'Posterior' is chosen from the 'Type' menu. The upper commands of this menu allow users to choose the input factor to be analysed. Moreover, the 'Bootstrap' command, allows a further analysis of the joint posterior distribution (see Section 3.11).

3.4.8 Other menus

The Uncertainty analysis window has also the File menu, the Likelihood menu and the Variable menu. Their properties have been already described for the Dotty plots window in Sections 3.3.2, 3.3.3, 3.3.4.
3.5  The "Sensitivity analysis" window

This window allows a 'visual' sensitivity analysis as firstly proposed by Spear and Hornberger [10 22]. The elements of the sample of likelihood measures /model outputs obtained with the Monte Carlo runs are divided into two classes: the 50% highest values and the 50% smallest values. By mapping these classes backwards onto the space of input parameters, we obtain the subsets of parameter which provide 'high' values of the model output and the subset giving 'low' values of the model output. By plotting the marginal cumulative distributions of these subsets, it is possible to perform a sensitivity analysis. If the cumulative distributions for the two classes of a given parameter tend to split, this implies that the parameter is important. Alternatively, if the parameter values under the two classes do not split but appear both to be from the original distribution, then it can be stated that the parameter is unimportant.

![Sensitivity analysis window](image)

**Figure 6.** Sensitivity analysis figure for a time dependent output.
For example, in Figure 6 parameter 'par4' is clearly important, parameters 'par1', 'par2', 'par3', 'par6' are intermediately important, while 'par5' is unimportant.

3.5.1 Figure commands

Change the number of classes

Clicking with the mouse right-button on the grey area of the figure, a menu is opened, which allows the user to chose the number of classes into which samples should be subdivided. The legend in the right-hand side of the plot allows interpreting the results, indicating the 'ranking' of the different classes. Classes are ranked and plotted in descending order: so the "Set 1 (High)" cumulative distribution in Figure 6 corresponds to the class with the highest values of the likelihood/model output.
Edit the name of the input factors
Each subplot of the figure shows the name of the parameter. Clicking with the mouse right-button on the parameter name allows editing it.

Edit the name of the plotted likelihood/variable
Clicking with the mouse right-button on the name of the analysed likelihood/variable allows editing it.

'Close' button
With this command the 'Sensitivity analysis' figure is closed.

'Next' button
If \( k > 6 \), the button 'Next 6' is enabled which allows the user to loop over all \( k \) factors.

Change the time value of the analysed model output
When a variable with time dependent output is plotted, an editable field will be available in which the user can set a time value and by clicking the 'OK' button the scatter graph for that time value will be plotted (see Figure 6).

Change the sample
If a filtering of the output variable has been done, with which a reduced sample has been obtained, a menu is activated on the top of the figure, which allows to choose the sample to use for make the sensitivity analysis. If a reduced sample is used, obtained with a filtering procedure, the subdivision into classes is made only for the filtered elements.

3.5.2 Figure menus
The 'Sensitivity Analysis' figure has the same menus as the 'Dotty plots' figure. See Sections 3.3.2, 3.3.3, and 3.3.4 for details.
3.6 **Transform likelihoods**

This panel allows transforming the available likelihood weights.

![GLUEWIN panel for transformation of likelihood functions.](image)

**Figure 7.** GLUEWIN panel for transformation of likelihood functions.

**3.6.1 Transform options**

Five types of transformations are available.

**Normalise**

This transformation allows re-scaling all likelihood values to get unit cumulative sum. This transformation does not affect uncertainty computations, since in the Uncertainty panel, normalised likelihood measures are always used, i.e. if the user supplies GLUEWIN with non-normalised likelihood measures, in the Uncertainty panel they are in all case normalised, in such a way that their cumulative sum is one.
**Raise to power**

Clicking this option, all positive likelihood values are raised to a user-supplied power. The user can enter the value of the exponent in the editable field to right-hand side of the button. If non-positive likelihood values are present, the transformation is performed for a filtered sample with only positive values. This option is useful if the user wants to change the relative weight of model runs.

If the power is larger than 1, the new likelihood will be more 'peaked' than the original one, i.e. a smaller number of runs will have likelihood values significantly larger than 0, while the number of runs with almost zero likelihood will increase. This is a way to make a more 'severe' conditioning to observations, and also implies smaller confidence bounds in the Uncertainty analysis.

If the power is smaller than 1, the new likelihood will be less peaked, and a larger number of runs will have likelihood values significantly larger than 0. This also implies larger confidence bounds in the Uncertainty analysis.

If the likelihood function being transformed is actually a cost function (e.g. a sum of squared errors, SSE), this option allows transforming the cost function into a likelihood function, by setting a negative exponent. In this way, large SSE values will result into small transformed values and vice-versa, obtaining following type of 'empirical' likelihood functions proposed by Beven and Binley [3]:

\[
L(\theta_i \mid Y) \propto \left(\sigma_i^2\right)^{-N}
\]  

(12)
Take exponential

This option allows the following transformation of the original likelihood $L(\theta_i \mid Y)$:

$$L_{\text{new}}(\theta_i \mid Y) = \exp(\alpha \cdot L(\theta_i \mid Y))$$

(13)

where $\alpha$ is a constant supplied by the user in the editable files to the right-hand side of the button.

This transformation allows, for example, creating the likelihood function defined in equation (11) starting from SSE values provided by the user in the likelihood file. To do so it is sufficient to set the value $\alpha$ as:

$$\alpha = -\frac{1}{\min(L(\theta_i \mid Y))}$$

(14)

Take log values

This option allows creating a new likelihood defined as

$$L_{\text{new}}(\theta_i \mid Y) = \ln(L(\theta_i \mid Y))$$

(15)

The transformation applies only for positive values of the original likelihood.

Apply threshold

This option allows filtering the sample of the Monte Carlo sample choosing threshold values for the likelihood function. Threshold values can be entered in the editable fields below the button.
Apply

Once a transform has been defined, click on the 'Apply' button to see information on the range of the new transformation and, in the case of a threshold, the proportion of the available Monte Carlo runs that are retained. The program also enters a default name for the transformed likelihood. The 'Apply' command does not save the transformed likelihood. This allows the user to make a 'rough' check of the transformation before saving, based on the minimum and maximum values and on the filtering rate. Moreover, the user can change the default name entered by the program, before saving.

Saving the transformation is allowed by clicking either on the 'Save' or on the 'Replace' buttons, which are enabled by the 'Apply' button.

3.6.2 Figure commands

Apply
**Choose the transform name**

The transformed name can be edited and changed by the user before the 'Save' or 'Replace' commands.

**Replace**

If the Transform option is entered with an existing transform chosen, that transform can be edited and applied, then replaced by clicking on the 'Replace' button. The new transform name will then appear under the Likelihoods Menu Item. The 'Replace' button is inactive if the likelihood being transformed is one of those provided by the user in the Likelihood file, i.e. it is not possible to destroy the original information provided by the user.

The 'Replace' command allows making transformations in series, without the need of saving all the intermediate transformations, e.g. one may want to normalise the original likelihood, then rise to the power 2 and finally take the exponential multiplied by -1. If only the last transformation is of real interest, clicking on the Replace button allows only
a single likelihood to be added to the Likelihood menus in the different panels of GLUEWIN. The replacement is only temporarily saved (the new likelihood is only visible in Likelihoods menu item of the 'Transform' panel). Only clicking on the 'OK' button the replacements will be permanent.

Save
The 'Save' button allows adding a new likelihood weight, defined by the transformation performed clicking on the 'Apply' button. The new likelihood is only temporarily saved (the new likelihood is only visible in the 'Transform' panel). Only clicking on the 'OK' button the transformations will be permanent.

Cancel
The 'Cancel' button exits the 'Transform' panel ignoring every transformation performed in the last session allowing the user to discard any Replacement or transformation performed in the last 'Transform' session.

OK
The 'OK' button exits the 'Transform' panel making all transformations and replacements permanent.

3.6.3 The Likelihood menu
The likelihood menu allows choosing the likelihood measure being transformed.
3.7 Combine likelihoods

The need to combine different likelihood measures arises in a number of different circumstances including [9]:

- combining likelihood measures for different types of model evaluation (such as one measure calculated for the prediction of discharges and one calculated for the prediction of water table or soil moisture levels);
- updating an existing likelihood estimate with a new measure calculated for the prediction of a new set of observations.

Most cases can be expressed as successive combinations of likelihoods, where a prior likelihood estimate is updated using a new likelihood measure to form a posterior likelihood. For different types of combination are allowed:

- **Multiplicative (Bayes):**
  \[ L_p(\theta | Y) = L_o(\theta) L(\theta | Y) \]

- **Weighted Addition:**
  \[ L_p(\theta | Y) = w_o L_o(\theta) + w_i L(\theta | Y) \]

- **Fuzzy Union:**
  \[ L_p(\theta | Y) = \max[L_o(\theta), L(\theta | Y)] \]

- **Fuzzy Intersection:**
  \[ L_p(\theta | Y) = \min[L_o(\theta), L(\theta | Y)] \]

where \( L_p(\theta | Y) \) is a posterior likelihood for parameter set \( \theta \), \( L_o(\theta) \) is the prior likelihood, and \( L(\theta | Y) \) is the likelihood calculated for the current data set \( Y \). \( L_o(\theta) \) and \( L(\theta | Y) \) in use could be different likelihoods calculated for the same data support \( Y \). Weights \( w_0 \) and \( w_1 \) are entered by the user in the fields to the right hand side of the button.

Bayes equation represents only one way of combining likelihoods. It has a number of characteristics that may, or may not, be attractive in model evaluation. Since it is a multiplicative operation if any evaluation results in a zero likelihood, the posterior likelihood will be zero regardless of how well the model has performed previously. This
may be considered as an important way of rejecting non-behavioural models, it may cause a re-evaluation of the data for that period or variable; it may lead to the rejection of all models.

The successive application of Bayes equation will tend to gradually reduce the impact of earlier data and likelihoods relative to later evaluations. This may be an advantage if it is thought that the system is changing over time; it has the disadvantage that if a period of calibration is broken down into smaller periods and the likelihoods are evaluated and updated for each period in turn, for many likelihood measures the final likelihoods will be different from using a single evaluation for the whole period (except for some very specific likelihood measures based on an exponential transformation of an error measure).

Bayes equation is not the only way of combining likelihoods. A simple weighted addition might be considered more appropriate in some cases. Further forms of combination come from fuzzy set theory. Taking a fuzzy union will emphasise the best performance of each model or parameter set over all the measures considered, taking a fuzzy intersection will emphasise the worst performance. In particular if any of the measures is zero, taking a fuzzy intersection will lead to the rejection of that model as non-behavioural.

### 3.7.1 Combine figure commands

The figure commands of the 'Combine' panel have the same meaning as in the 'Transform likelihood' panel (see Section 3.6.2). Only the 'Replace' button has some slight behaviour. To avoid ambiguity, the 'Replace' button works only on the 'Likelihood 1' Menu Item.

### 3.7.2 Combine figure menus

The two menus 'Likelihood 1' and 'Likelihood 2' allow choosing the two likelihoods to be combined. To avoid ambiguity, only the Likelihood selected in menu 'Likelihood 1' can be replaced using the 'Replace' button.

GLUE: Combine likelihood measures

Likelihood measures being combined: lik-O2-1997 lik-CHL-1997

Note that both measures will be scaled to consume1 before combination

Define combination

- Bayesian Multiplication
- Weighted Addition
- Fuzzy Union (normalised maximum)
- Fuzzy Intersection (normalised minimum)

Combination name: BM: lik-O2-1997*lik-CHL-1997

Minimum value: 1.713e-0168
Maximum value: 1.852e+007

Number of Parameter Sets remaining after combining: 7168 100%

Apply  Replace  Save  Cancel  OK
3.8 Save likelihood files

This panel allows saving a user defined subset of likelihoods in an ASCII file in SIMLAB format, allowing direct use of the SIMLAB software for GSA. Likelihood measures are firstly picked up using the Likelihoods Menu Item and then the choice must be confirmed clicking on the 'Select' button. Once the list of likelihood functions to be saved is complete, the likelihood file name has to be entered clicking on the corresponding button. Finally, click on the 'Save' button to generate the ASCII file.
3.9 **Filter output variables**

This panel allows the performance of Monte Carlo filtering on the output variables. Main figure commands behave the same way as for the 'Transform likelihood' panel (see Section 3.6.2).

![Filter output variables panel](image)

3.9.1 **Filter figure commands**

*Time point of the output variable*

In the case of time dependent output, this button allows choosing the time point at which filtering has to be performed, e.g. in the case the user would like that the final value of the simulations does not exceed a given threshold. Moreover, it allows the output variable to be filtered at any time points, by setting the time value at "Inf", i.e. the output variable is constrained by lower and upper thresholds for the whole simulation.
Base sample for filtering

This pop-up menu allows choosing the starting sample to be filtered: it can be the original sample or an already filtered sample. In the latter case, the 'Replace' button will be enabled.

3.9.2 The Variable menu

This menu allows choosing the output variable on which filtering has to be performed.
3.10 Save filtered data

This panel allows saving a filtered sample of the Monte Carlo analysis in ASCII files (in SIMLAB format), allowing direct use of the SIMLAB to analyse the filtered data. Filtered samples are firstly picked up using the Samples Menu Item and then the choice must be confirmed clicking on the 'Select' button. Then the names of the sample file, the model output file and the likelihood file have to be entered clicking on the corresponding buttons. Finally, click on the 'Save' button to generate the ASCII files containing only the filtered sample.
3.11 Analyse the posterior distributions of input factors

This panel allows analysing the joint posterior distribution of the model input factors through bootstrap samples and covariance structure. It has two sub-panels: the upper for generating bootstrap samples, the lower for analysing the covariance structure and for viewing bidimensional projections of the joint distributions.

3.11.1 Bootstrap sub-panel

In the left menu, the likelihood measure to use for generating the bootstrap sample must be first selected. Then the user must enter the bootstrap sample size. Click ‘Apply’ to generate the sample. Finally, after editing the new sample name, the sample can be saved. With the ‘Cancel’ button, the program exits without saving the new samples generated, while clicking ‘OK’ the new samples are made permanent.
3.11.2 View sample space sub-panel

Two options are possible: compute correlation structure and view bidimensional projections of the sample space.

Correlation coefficients

Clicking on this button two windows are opened: the first shows bar plots of the correlation coefficients, the second shows a table with the correlation coefficients.
Two sets of correlation coefficients are computed and displayed. The correlation coefficients shown in the upper part of the figures are computed using the sample selected in the Sample List Item. The correlation coefficients shown in the lower part are computed using the Original sample and weighting with the likelihood measure selected in the Likelihood Measure List Item. In the latter case, the following procedure is applied [1].

By normalising likelihood measures we obtain weights such as:

$$\sum_{i=1}^{n} w_i(x) = 1$$  \hspace{1cm} (16)

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

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

$$\hat{E}(x_j) = \sum_{i=1}^{n} x_{ij} w_i(x) = \hat{\mu}_j$$  \hspace{1cm} (17)
\begin{equation}
\hat{V}(x_j) = \sum_{i=1}^{n} x_{ij}^2 w_i(x) - \hat{\mu}_j^2 = \hat{\sigma}_j^2
\end{equation}

and defining the new standardised factors:

\begin{equation}
\tilde{x}_j = \frac{x_j - \hat{\mu}_j}{\hat{\sigma}_j}
\end{equation}

the correlation coefficients can be estimated as:

\begin{equation}
\hat{\rho}_{jk} = \sum_{i=1}^{n} \tilde{x}_{ij} \tilde{x}_{ik} w_i(x) \in (-1,1)
\end{equation}

So by selecting the Original sample for the first type of correlation coefficients and a likelihood measure for the second type, it is possible to view the modifications in the correlation structure induced by conditioning to observed data.

**View projections**

First the user must select a sample in the Sample List Item, then he must select a given parameter \(x_j\) in the Input factor List Item. Clicking on the ‘VIEW’ button, a new window is opened, showing the structure of all the other factors vs. \(x_j\), i.e. showing bidimensional projection of the sample on to the planes \((x_j, x_k)_{k \neq j}\).
3.12 The DEMO files

In the GLUEWIN files downloadable at the IMPACT page of the JRC site there are also demo files usable by GLUEWIN:

Table 2. DEMO files.

<table>
<thead>
<tr>
<th>Sobol'</th>
<th>FAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>demo.m</td>
<td>demo.fac</td>
</tr>
<tr>
<td>model.m</td>
<td>demo.imd</td>
</tr>
<tr>
<td>demo.obs</td>
<td>demo.obs</td>
</tr>
<tr>
<td>demo.out</td>
<td>demo_fast.out</td>
</tr>
<tr>
<td>demo.sam</td>
<td>demo_fast.sam</td>
</tr>
<tr>
<td>series.mat</td>
<td>series_fast.mat</td>
</tr>
</tbody>
</table>

The demo files consider the following time-dependent model with three input factors:

\[
    y(t) = x_1 \frac{\exp(-x_3 t)}{1 + \exp(-x_2 t)}
\]

The support for the three parameters is:

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Distribution</th>
<th>'True' values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>U[0.3-0.7]</td>
<td>0.5</td>
</tr>
<tr>
<td>$x_2$</td>
<td>U[0.2-2]</td>
<td>1</td>
</tr>
<tr>
<td>$x_3$</td>
<td>U[0.05-0.4]</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Two types of samples, and corresponding model outputs, are considered: a Sobol' sample and a FAST sample, created with SIMLAB.

For the Sobol' analysis, the MATLAB files `demo.m` and `model.m` are given, that enable to run the model. Moreover, the sample file (`demo.sam`) the output file (`demo.out`) and the observations file (`demo.obs`) are also given. The pseudo-observed data are obtained running the model with the 'true' values shown in the table above and adding a standard error of 0.005. Such files can be loaded in the 'Measures are given' mode. The file `series.mat` is also given, where the data in binary format are saved. In this case, the file must be loaded in the 'Saved analysis' mode.

For the FAST analysis, the files `demo.fac` (SIMLAB configuration file defining parameter distributions) and `demo.imd` (SIMLAB file for running the demo model 'internally', see SIMLAB manual) are given, for re-running the demo model within SIMLAB. Moreover the following files usable with GLUEWIN are given:
demo_fast.sam (sample file), demo_fast.out (output file), demo.obs (observations file), series_fast.mat (binary file).
4 Concluding remarks

In this report, a software prototype has been presented called GLUEWIN, which embeds Bayesian uncertainty estimation into a Global Sensitivity Analysis. The software has been applied and tested in the case studies of IMPACT, specifically in the co-operations with GKSS (analysis of the WAMPUM model) and with ULANC (analysis of the Transfer Function model) in framework of the Elbe River case study [17, 23].

The software has been made available at the IMPACT project page of the JRC web site, on the address:

http://www.jrc.cec.eu.int/uasa/prj-impact.asp

Staff members of the Applied Statistic sector of JRC also used GLUEWIN as Alfa testers. Thanks are due to R. Girardi, A. Rossi, A. Saltelli, S. Tarantola and U. Callies for numerous comments and suggestions.
5 References

[6] SIMLAB, Ver 1.1, Simulation environment for uncertainty and sensitivity analysis, by Andrea Saltelli, JRC/ POLIS ScaRL.
References


