LBNL-1005838

Global Sensitivity and Data-Worth Analyses in iTOUGH2

User's Guide

Haruko Murakami Wainwright
Stefan Finsterle

Earth Sciences Division
Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720

July 2016

This work was supported, in part, by the U.S. Dept. of Energy under Contract No. DE-AC02-05CH11231.

DISCLAIMER

Table of Contents

1. INTRODUCTION **1**

2. METHODS **3**

2.1 Local Sensitivity Analysis **3**

2.2 Morris Global Sensitivity Analysis **6**

2.3 Sobol' Global Sensitivity Analysis **8**

2.4 Data-Worth Analysis **10**

3. INPUT AND OUTPUT FORMATS **14**

3.1 Input Formats **14**

3.1.1 Local Sensitivity Analysis 14

3.1.2 Morris Global Sensitivity Analysis 16

3.1.3 Sobol' Global Sensitivity Analysis 16

3.1.4 Data-Worth Analysis 17

3.2 Output Formats **19**

3.2.1 Local Sensitivity Analysis 19

3.2.2 Morris Global Sensitivity Analysis 19

3.2.3 Sobol' Global Sensitivity Analysis 19

3.2.4 Data-Worth Analysis 20

4. SAMPLE PROBLEMS **21**

4.1.1 Sensitivity Analyses 21

4.1.2 Data-Worth Analysis 28

5. CONCLUDING REMARKS **39**

Acknowledgments **39**

References **40**

List of Figures

Figure 1: iTOUGH2 input file for local sensitivity analysis. 15

Figure 2: iTOUGH2 block > COMPUTATION for the Morris One-At-A-Time global sensitivity analysis. 16

Figure 3: iTOUGH2 block > COMPUTATION for the Sobol' global sensitivity analysis. 16

Figure 4: Schematic of gas-pressure-pulse-decay apparatus. 21

Figure 5: Local, scaled sensitivity coefficients as a function of time for upstream and downstream reservoir pressures observable during a gas-pressure-pulse-decay experiment. 22

Figure 6: Time evolution of sensitivity index: (a) Mean EE, (b) Mean |EE| (c) Sobol' index, and (d) Sobol' total sensitivity index. In (a)-(d), the thin lines represent the 95% confidence intervals. 24

Figure 7: Mean EE vs. SD from the Morris method. The circle, square and triangle on each curve represent 100, 748 and 7150 seconds, respectively. The end of each line corresponds to 6.705×10^4 seconds. 26

Figure 8: The interaction effect (the difference between the Sobol' index and the total sensitivity index) as a function of Sobol' index. The circle, square and triangle on each curve represent 100, 748.48 and 7150.37 seconds, respectively. The end of each line corresponds to 6.705×10^4 seconds. 27

Figure 9: Five-spot well pattern with grid for modeling a 1/8 symmetry domain. 28

Figure 10: Excerpt of TOUGH2-EOS1 input file for simulating five-spot geothermal reservoir operation. 29

Figure 11: Excerpt of iTOUGH2 input file, showing > PARAMETER block with potentially influential parameters. 33

Figure 12: Excerpt of iTOUGH2 input file, showing > OBSERVATION block with potential calibration data sets. 34

Figure 13: Excerpt of iTOUGH2 input file, showing > COMPUTATION block, which initiates the data-worth analysis. 35

Figure 14: Excerpt of iTOUGH2 output file, showing results of residual and data-worth analysis. 37

Figure 15: Excerpt of iTOUGH2 output file, showing summary statistics of data-worth analysis. 38

1. INTRODUCTION

iTOUGH2 (Finsterle, 2010) has been developed as an inverse modeling and parameter estimation (PE) tool for various modules of the non-isothermal multiphase flow and transport simulator TOUGH2 (Pruess et al., 2012; Falta et al., 1995; Pruess and Battistelli, 2002; Oldenburg et al., 2004; Pruess, 2005; Finsterle and Kowalsky, 2007; Pruess, 2011; Doughty, 2013). As analysis capabilities have been added, iTOUGH2 has become an integrated framework for hydrogeological modeling under uncertainty – from test design to site characterization to prediction – including parameter estimation (PE), uncertainty analysis (UA), and sensitivity analysis (SA). Moreover, all iTOUGH2 analysis tools can be applied to any stand-alone, text-based simulator through the use of the PEST interface (Finsterle, 2010; Finsterle and Zhang, 2011).

A considerable fraction of numerical simulation runs are performed in the broad context of a sensitivity analysis. Modeling results are obtained for different conceptual models, scenarios, forcing terms, and parameter values and are compared to each other to examine the impact of these factors on model predictions. The motivations for performing such sensitivity analyses are as varied as the purposes for developing a numerical model in the first place, which range from improving fundamental process understanding to site characterization to predictive modeling to performance assessment and decision support. Depending on these ultimate modeling goals, sensitivity analyses are used to (1) uncover errors in mechanistic models, (2) build parsimonious data-driven models, (3) identify key processes in diagnostic models, (4) determine influential factors in prognostic models, and (5) defend the robustness of a particular model. More specifically, sensitivity analyses can help to (6) identify which (uncertain) parameters have the greatest effect on model predictions and prediction uncertainties, and consequently (7) which properties need to be estimated with high accuracy. This allows one to (8) establish research priorities, and to (9) rank, screen, and thus reduce the number of parameters to be varied or estimated, lowering the computational burden of the subsequent parameter estimation and uncertainty analyses. Complementary to the measures of parameter influence, sensitivity indices may be used to (10) identify which observations are likely to contain useful information about the relevant parameters to be estimated, and (generally referred to as a value-of-information or data-worth analysis) (11) evaluates an observation's potential to reduce the uncertainty of the parameters to be estimated by inverse modeling, or (12) an observation's potential to reduce prediction uncertainty through a more precise estimation of influential parameters.

In general, a sensitivity analysis examines the relation between a parameter that is an input to the numerical model, and an observable variable that is an output of the numerical model. The input parameter can be a parameter to be estimated by inverse modeling, an uncertain parameter used in a predictive model, or an operational or design variable. Correspondingly, the output variable can be the observation at a calibration point or the value of an objective function, a predictive variable, or a cost function or other performance measure. It is therefore essential to clearly identify and report the objective of any sensitivity analysis for its results to be interpreted correctly, specifically when parameters are ranked according to their importance, or a data points are valued according to their worth, where worth is a measure of how much the estimation or prediction uncertainty can be reduced by collecting a given data point or data set.

In this manual, we declare observations to be *sensitive* or *insensitive* with respect to a change in a parameter, whereas parameters are *influential* or *non-influential* regarding the declared objective to be calculated by the numerical model.

In contrast to the derivative-based, local sensitivity method (Cacuci, 2003), global sensitivity analysis (GSA) methods, such as the Morris and Sobol' methods (Morris, 1991; Sobol', 2001; Saltelli et al., 2008), explore the parameter space so that they provide robust sensitivity measures in the presence of nonlinearity and interactions among parameters. The Morris one-at-a-time (OAT) method is a computationally frugal method that changes one parameter at a time from randomly generated reference parameter sets, and computes the difference in outputs. The Sobol' method provides the variance-based sensitivity indices that quantify the relative contribution of each parameter to the uncertainty in outputs.

This manual explains the use of local sensitivity analysis, the global Morris OAT and Sobol' methods, and a related data-worth analysis as implemented in iTOUGH2. In addition to input specification and output formats, it includes some examples to show how to interpret results. Wainwright et al. (2013; 2014) demonstrated that GSA could be used not just to identify important parameters but also to identify nonlinearity and interactions among parameters. This manual shows additional plots that can be created from the iTOUGH2 outputs to visualize such information.

2. METHODS

In this section, we introduce the three sensitivity methods implemented in iTOUGH2:

(1) local sensitivity coefficients,

(2) the Morris OAT method, and

(3) the Sobol' variance-based method, as well as

(4) data-worth analysis.

Although the local SA is included in the original iTOUGH2 manual (Finsterle, 2010), it is described here for completeness and to highlight the difference to the global methods. More details of GSA are documented in Morris (1991), Sobol' (2001), Saltelli et al. (2008) and Wainwright et al. (2013, 2014).

We denote a set of parameters by $\mathbf{p} = [p_1, p_2, \ldots, p_n]$ and a set of model outputs $\mathbf{z} = [z_1, z_2, \ldots, z_m]$, which are a function of $\mathbf{p}$, i.e., $\mathbf{z} = f(\mathbf{p})$, where n is the number of parameters, m is the number of output variables of interest, and f represents any hydrological or other type of forward model.

2.1 Local Sensitivity Analysis

The local sensitivity coefficient is defined as a partial derivative, i.e., the change of an output variable caused by a unit change in each parameter from the reference value.

$$S_{ij} = \left. \frac{\partial z_i}{\partial p_j} \right|_{\mathbf{p}^*} \tag{1}$$

While there are several methods to calculate sensitivity coefficients (e.g., analytical, adjoint-state method, automatic differentiation, perturbation method), iTOUGH2 evaluates the derivative by changing each parameter i by a small increment δp_i from its reference value p_i^* and computes the difference of the output. If approximating the partial derivatives using a first-order finite-difference approximation, the total number of forward simulations required to calculate n column vectors of length m, each holding the sensitivity coefficients of all outputs with respect to a parameter, is $(n+1)$, i.e., the reference-case simulation plus the n simulations for small increments in the n parameters.

$$S_{ij} \approx \frac{z_i(p_1^*, \ldots, p_j^* + \delta p_j, \ldots, p_n^*) - z_i(p_1^*, \ldots, p_n^*)}{\delta p_j} \tag{2}$$

If centered finite differences are used, the number of simulations increases to $(2n+1)$:

$$S_{ij} \approx \frac{z_i(p_1^*,\ldots,p_j^*+\delta p_j,\ldots,p_n^*) - z_i(p_1^*,\ldots,p_j^*-\delta p_j,\ldots,p_n^*)}{2\delta p_j} \tag{3}$$

Because the units of the partial derivatives are the units of the model output over the units of the parameter, they cannot be readily compared to each other if we are concerned with the sensitivity of different model outputs with respect to parameters of different types. We therefore introduce a scaled, dimensionless sensitivity index, which is defined as

$$\overline{S}_{ij} = S_{ij} \cdot \frac{\sigma_{p_j}}{\sigma_{z_i}} = \frac{\sigma_{p_j}}{\sigma_{z_i}} \cdot \left.\frac{\partial z_i}{\partial p_j}\right|_{\mathbf{p}^*} \tag{4}$$

where σ_p is the parameter-scaling factor, and σ_z is the output- or observation-scaling factor, also referred to as parameter variation or standard deviation, and observation standard deviation.

The choice of these scaling factors appears somewhat subjective, and alternative definitions of the scaled sensitivity coefficients have been proposed in the literature, specifically $\sigma_p = p$ (e.g., Hill and Tiedeman, 2007) and $\sigma_z = z$, or various variations thereof. While all of these alternatives result in non-dimensional scaled sensitivity coefficients, we argue that a careful choice of these scaling factors enhances the interpretative power of the resulting composite sensitivity measures.

For example, σ_p is probably best thought of as the amount by which the parameter would be changed in a conventional sensitivity analysis, where the parameter is perturbed from its base-case value by an amount considered "reasonable" to examine its impact on the model output. (If different perturbations would be chosen depending on whether the parameter is increased or decreased from its reference value, a parameter transformation is probably in order.) It can also be viewed as the standard deviation or range of the parameter that represents the parameter variability or its uncertainty. If scaled sensitivity coefficients are used as measures of relative parameter influence, σ_p reflects the leverage of this parameter's uncertainty on model predictions and thus indicates the potential need to obtain better estimates through independent measurements or inverse modeling. Accurately known parameters with a sufficiently low σ_p value are thus correctly identified as non-influential in the sense that they do not deserve our prime attention when designing an experiment. If sensitivity coefficients are used as rudimentary data-worth measures, σ_p reflects the target estimation uncertainty; the more accurately a parameter needs to be estimated, the comparatively less value does a certain data point with a given measurement uncertainty (σ_z) have when used as a calibration point in inverse modeling. Finally, if a sensitivity coefficient is used for a first-order uncertainty propagation analysis, σ_p reflects our current state of knowledge about this parameter, and implies our presumption that the model is approximately linear within the range defined by σ_p.

Understanding the role that σ_p plays in the evaluation of scaled sensitivity coefficients is essential, and the interpretations given here may help obtain more meaningful results than the indiscriminate choice $\sigma_p = p$. The fact that σ_p has specific meaning depending on the context within which the sensitivity coefficients are used also allows the modeler to examine the impact and robustness of these analyses with respect to changes in their objectives or other conditions. For example, should independent information about a parameter become available, this will impact data worth of potential data sets. If this data worth were examined by scaled sensitivity

analysis alone, the choice $\sigma_p = p$ would miss that important influence, while an appropriate reduction in σ_p will capture this often significant influence. Of course, the effects of prior information are properly included in a formal data-worth analysis (as the one presented below).

In addition to the interpretations outlined above, the main purpose of the scaling of the partial derivatives (Eq. (1)) by the ratio σ_p / σ_z is to make the sensitivity measures dimensionless, thus allowing one to compare parameters and observations of different types, units, and scales. Moreover, it enables evaluation of integral measures, e.g., the sum of the absolute values of scale sensitivity coefficients for a given parameter, a given observation type, time, location, or data set. These composite indices provide useful information about relative parameter influence or output sensitivity. If the composite measure is a sum of rows of the sensitivity matrix along a column, it refers to a parameter's influence on model output; if it is the sum of columns over one or multiple rows, it refers to the sensitivity or information content of an observation, a data set, or a data type.

It should be noted, however, that these composite sensitivity measures do not account for linear dependencies and statistical correlations among closely spaced parameters and observations; they may thus be misleading. Synthetic inversions or formal data-worth analyses overcome this limitation.

In addition to the local sensitivity coefficients of the system response with respect to the parameters (Eq. (1)), iTOUGH2 also calculates the change in the objective function Φ as parameter j is perturbed by δp_j:

$$S_j^{OF} = \left|\Delta\Phi\right\|_{\mathbf{p}^*} = \left|\Phi[f(p_1^*,\ldots,p_j^*+\delta p_j,\ldots,p_n^*)] - \Phi[f(p_1^*,\ldots,p_n^*)]\right| \tag{5}$$

This local sensitivity measure can be used to select the parameters that most likely help improve the match to measured data. While the objective function Φ is often chosen as the sum of the squared weighted residuals (where the weight is the inverse of σ_z), iTOUGH2 provides the user with alternative functions (referred to as "robust estimators").

Note that the sensitivity analysis of the system response, Eq. (1), refers to the model output only and thus does not make use of any measured data; on the other hand, Eq. (5) refers to the objective function, which includes the residuals and thus the measured data.

Once the sensitivity coefficients are calculated and stored in the Jacobian matrix **J**, a simple linear uncertainty analysis can be performed to obtain the covariance matrix of the estimated parameters, $\mathbf{C}_{pp}$:

$$\mathbf{C}_{pp} = s_0^2 \left(\mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{J}\right)^{-1} \tag{6}$$

Here, **J** is the $n \times m$ Jacobian matrix, holding the sensitivity coefficients S_{ij}; $\mathbf{C}_{zz}$ is the $m \times m$ observation covariance matrix, containing the variances σ_z^2 on its diagonal, and s_0^2 is the estimated error variance. If no measured data are available—which is often the case in a sensitivity analysis—s_0^2 is defined as 1.0; if measured data are available, s_0^2 is given by

$$s_0^2 = \frac{\mathbf{r}^T \mathbf{C}_{zz}^{-1} \mathbf{r}}{m-n} \tag{7}$$

Here, **r** is the residual vector with elements $r_i = (z_i^* - z_i(\mathbf{p}))$, where z_i^* is the measured data point that corresponds to the model-calculated, observable variable z_i. Finally, a linear uncertainty propagation analysis yields the covariance matrix of the model prediction:

$$\mathbf{C}_{\hat{z}\hat{z}} = \hat{\mathbf{J}} \mathbf{C}_{pp} \hat{\mathbf{J}}^T \tag{8}$$

In Eq. (8), the Jacobian matrix $\hat{\mathbf{J}}$ holds sensitivity coefficients of the prediction of interest with respect to the parameters **p**, whose uncertainty is described by $\mathbf{C}_{pp}$.

Based on the sensitivity matrix, iTOUGH2 also performs an analysis in which the relative contribution of a parameter to the prediction uncertainty is evaluated. First, prediction uncertainty is evaluated as a function of parameter uncertainty based on a linearity and normality assumption using Eq. (8). The parameter covariance matrix $\mathbf{C}_{pp}$ is either (Application Mode A) user-specified (in the case of a first-order-second-moment (FOSM) error propagation analysis), or (Application Mode B) calculated using Eq. (6) (in the case of sensitivity analysis, data-worth analysis, or an inversion). The contribution of a parameter to prediction uncertainty is the examined in two ways. In Option (i), each parameter—one at a time—is assumed to be perfectly known, i.e., it is removed from the uncertainty propagation analysis. The covariance matrix resulting from fixing the *j*-th parameter is denoted by $\mathbf{C}_{\hat{z}\hat{z},j}$. An $m \times n$ matrix is then constructed with elements

$$PCPU_{ij} = \left(1 - \frac{\sigma^2_{\hat{z}_{i,j}}}{\sigma^2_{\hat{z}_i}}\right) \times 100\% \tag{9}$$

In this way, the relative contribution in percent of each parameter to prediction uncertainty can be assessed. It is important to realize that in the context of FOSM analysis (Application Mode A), the relative contributions for a given prediction add up to 100%. However, this is not the case in Application Mode B, because having perfect knowledge of one parameter also reduces the (notional) estimation uncertainty of all other parameters that are correlated to the parameter being fixed, thus increasing this parameter's influence on prediction uncertainty.

In Option (ii), all parameters except parameter *j* are assumed to be perfectly known, and the process is repeated.

2.2 Morris Global Sensitivity Analysis

The Morris one-at-a-time (OAT) method is a global SA method developed by Morris (1991). It can be considered an extension of a local SA, since the Morris method randomly generates sets of reference values from the entire parameter range, and computes the difference of output caused by a fixed parameter change. The normalized parameter range is partitioned into $(k-1)$ equally-sized intervals so that each normalized parameter takes values from the set $\{0, 1/(k-1), 2/(k-1), \ldots, 1\}$. A fixed normalized increment is calculated as $\Delta = k/\{2(k-1)\}$. A random

reference point ξ is chosen from the set $\{0, 1/(k-1), 2/(k-1), \ldots, 1-\Delta\}$, and the increment Δ is added to each parameter in random order. The increments and parameters are back-transformed, i.e., $p_i = p_{i,\min} + \xi(p_{i,\max} - p_{i,\min})$ and $\Delta_i = \Delta(p_{i,\max} - p_{i,\min})$, and two variations of the elementary effect (EE) for parameter p_i are calculated as:

$$EE_i = \left(\frac{1}{\sigma_z} \cdot \frac{f(p_1^*, \ldots, p_i^* + \Delta_i, \ldots, p_n^*) - f(p_1^*, \ldots, p_n^*)}{\Delta_i} \right) \tag{9}$$

$$EE_i = \left(\frac{\Phi[f(p_1^*, \ldots, p_i^* + \Delta_i, \ldots, p_n^*)] - \Phi[f(p_1^*, \ldots, p_n^*)]}{\Delta_i} \right) \tag{10}$$

where **p*** is the randomly selected parameter set, and σ_z is the output-scaling factor. Eq. (9) is the original formulation. However, as the sign of a sensitivity coefficient is often irrelevant for the evaluation of parameter influence, taking the mean of the absolute values of EE can be used to represent parameter sensitivity (see below). In iTOUGH2, the elementary effect is also evaluated with respect to the objective function Φ; Eq. (10) thus includes the measured data and evaluates a parameter's influence on changing the goodness-of-fit criterion rather than the influence on a composite measure of the model output, as does Eq. (9).

Although the original sensitivity method proposed by Morris (1991) scales to EE values only by the parameter range, it is possible to scale EE by an output-weighting factor, as we discussed for the local sensitivity method.

By conducting simulations over multiple "paths" (i.e., multiple sets of reference parameter values and multiple, random orders of changing each parameter), an ensemble of EEs is obtained for each parameter. The number of required runs is $r(n + 1)$, where r is the number of paths. Using the ensemble of EEs, we can compute three summary statistics: the mean of EE according to Eq. (9), the mean of absolute EEs (mean |EE|), and the standard deviation (SD) of EEs. The mean EE and mean |EE| can be regarded as a global sensitivity index, since they represent the average effect of each parameter over the parameter space. Wainwright et al. (2014) discussed that the sign (i.e., positive or negative) of mean EEs is useful for system understanding. In addition, Wainwright et al. (2014) showed that, since the mean |EE| and Sobol' total sensitivity index are both computed by perturbing one parameter at a time (see Section 2.3), the mean |EE| provides information similar to the total sensitivity index with less computational cost, and hence the mean |EE| can be used to identify non-influential factors (Saltelli et al., 2008). The standard deviation of EE is used to identify nonlinear or interaction effects as well as to compute the standard error of the mean (SEM), i.e., SEM = SD/$r^{0.5}$ (Morris, 1991). The SEM represents the uncertainty in estimating the sensitivity index.

The results of a Morris global sensitivity analysis can be visualized by plotting the mean elementary effect EE against the standard deviation SD. Parameters with a relatively high absolute value of EE can be considered—on average, i.e., evaluated over the entire parameter range—more influential than parameters with |EE| closer to zero. Moreover, parameters with a small SD value enter the model f in an approximately linear fashion; parameters with SD values

significantly larger than EE exhibit effects from nonlinearity and/or parameter interaction. SEM can be used to judge whether SD is indeed significantly greater than |EE|.

Results from a Morris analysis contain global sensitivity information even if only a small number of paths can be evaluated. This frugality makes the Morris method a very useful GSA tool also for computationally expensive high-fidelity models. Specifically, EE can be interpreted as an average parameter influence over the chosen parameter range, and the variability of the elementary effect provides insights into the models degree of nonlinearity and parameter interaction. As these are only relative and qualitative sensitivity measures, there is no stringent requirement to evaluate a large number of paths to achieve statistical convergence; moreover, the SEM value provides some indication of the accuracy of the estimated elementary effect.

It is important to realize that the results of a Morris analysis depend relatively strongly on the parameter range specified. Conversely, the number of partitions, k, is relatively insignificant ($4 \le k \le 12$ is an appropriate range; note that P has to be an even number). Finally, as mentioned above, the number of paths, r, can be chosen pragmatically based on the available computational resources.

2.3 Sobol' Global Sensitivity Analysis

While the local and Morris sensitivity methods are difference-based, the Saltelli method is variance-based. Here we define the random variable Z_j and the random $\mathbf{P} = [P_1, P_2, P_3, \ldots, P_n]$ for the system response and the parameters, respectively. The sampled response and parameters are z_j and $\mathbf{p}$.

The Sobol' global sensitivity analysis evaluates two conditional variances as sensitivity indices. The first-order sensitivity index (referred to as the Sobol' index) is defined by

$$S_{ij} = \frac{V[E[Z_j \mid P_i]]}{V[Z_j]} \tag{11}$$

where $\mathbf{E}[\bullet]$ and $\mathbf{V}[\bullet]$ represent mean and variance, respectively. S_{ij} quantifies the first-order effect, i.e., the relative contribution of P_i to the uncertainty of Z_j. As it measures the variability of the output with respect to an individual parameter, it excludes the interaction effect with other parameters. It Sobol' index is used to identify influential parameters. The total sensitivity index is defined as

$$S_{tij} = 1 - \frac{V_{-i}[E[Z_j \mid P_{-i}]]}{V[Z_j]} \tag{12}$$

where $\mathbf{E}[Z_j|P_{-i}]$ represents the mean of Z_j conditioned on all the parameters but P_i. S_{tij} accounts for the total effect of P_i including interaction effects, and is used to identify parameters with negligible effects and parameters that can be fixed. The interaction effect is the effect of each parameter depending on other parameters.

We compute the sensitivity index S_i using an algorithm developed by Saltelli et al. (2003) and modified by Glen and Issacs (2012). It starts from generating two sets of sample matrices **A** and **B**, each of which is an *r*-by-*n* matrix containing *r* sets of *n*-dimensional parameter vectors from Monte-Carlo (MC) sampling. Any underlying parameter distributions can be used. From **A** and **B**, we create matrices $\mathbf{C}_i$ (i=1,…,n) such that the i-th column of $\mathbf{C}_i$ is the same as the i-th column of **A** (i.e., $\mathbf{C}_{i,(l,\,i)} = \mathrm{A}_{(l,\,i)}$ for l=1,2,….r), and the other columns of $\mathbf{C}_i$ are the same as **B** (i.e., $\mathbf{C}_{i,(l,\,h)} = \mathbf{B}_{(l,\,h)}$ for l=1,2,….r and $h \neq i$). The simulation results from the parameter sets **A**, **B** and $\mathbf{C}_i$ are *r*-dimensional vectors: $\mathbf{a}_j$, $\mathbf{b}_j$ and $\mathbf{c}_{ij}$ (for each output Z_j), respectively (i.e., *r* realizations of simulations for each vector). The number of required simulations is $r(n+2)$; the proportionality to the number of parameters is similar to the Morris method. Following Glen and Issacs (2012), the Saltelli sensitivity index is computed as a correlation coefficient between $\mathbf{a}_j$ and $\mathbf{c}_{ij}$:

$$S_{ij} = \frac{1}{\sigma_{Zj}^2}\frac{1}{r-1}\sum_{l=1}^{r}(a_{j,l}-\mu_{Zj})(c_{ij,l}-\mu_{Zj}), \quad (13)$$

where μ_{Zj} is the overall mean of Z_j and $\sigma_{Zj}{}^2$ is the overall variance of Z_j. Although Glen and Issacs (2012) did not comment on this, Eq. (13) provides a more intuitive way to understand S_{ij}. The parameter sets $\mathbf{C}_i$ and **A** share the same values only for the i-th parameter. If the i-th parameter is more influential than the other parameters, the i-th parameter determines the results so that the simulation results $\boldsymbol{c}_{ij}$ and $\boldsymbol{a}_j$ should be similar and hence have higher correlation.

Similar to the Saltelli sensitivity index, the total sensitivity index can be computed as:

$$S_{tij} = 1 - \frac{1}{\sigma_{Zj}^2}\frac{1}{r-1}\sum_{l=1}^{r}(c_{ij,l}-\mu_{Zj})(b_{j,l}-\mu_{Zj}). \quad (14)$$

Using the covariance-semivariogram relationship, we can re-write the total sensitivity index as:

$$S_{tij} = \frac{1}{\sigma_{Zj}^2}\frac{1}{2(r-1)}\sum_{l=1}^{r}(c_{ij,l}-b_{j,l})^2. \quad (15)$$

Since $\mathbf{C}_i$ and **B** have the same parameter values except for the i-th parameter, $(\boldsymbol{c}_{ij} - \boldsymbol{b}_j)$ is equivalent to taking a difference in z_j when perturbing the i-th parameter with the other parameters fixed. When we perturb the i-th parameter, the change in Z_j includes all the effects associated with the i-th parameter. S_{tij}, therefore, includes the interaction effects associated with the i-th parameter. This procedure is the same as the Morris and local sensitivity methods, except that the difference in the output is not divided by the parameter difference (Δ). This similarity would be the reason why Campolongo et al. (2007) observed the mean |EE| being a good proxy for S_{tij}.

Due to its high computational expense, it is important to account the uncertainty in S_{ij} and S_{tij} caused by a given number of simulations, and hence to compute the confidence interval of S_{ij} and S_{tij} (Wainwright et al., 2014). Interpreting the sensitivity indices as correlations coefficients allows us to compute the confidence interval of S_{ij} and S_{ti} (Fisher, 1921) and to compare the sensitivity indices of different parameters with a limited number of simulations. The 95%

confidence interval of S_{ij} is given as $\tanh\{\text{arctanh}(S_{ij}) \pm 1.96\text{SE}\}$, where SE is the standard error given by $\text{SE} = (r-3)^{-0.5}$. Similarly, the 95% confidence interval of S_{tij} is given as $1-\tanh\{\text{artanh}(S_{tij}) \pm 1.96\text{SE}\}$. The standard error and confidence interval shrink as r increases. Having a confidence interval is useful to determine the sufficient number of simulation runs r such that the ranking of two parameters can be delineated when the confidence intervals become small and are no longer overlapping.

2.4 Data-Worth Analysis

A data-worth analysis complements a sensitivity analysis in that it specifically identifies the contribution each (potential or existing) data point makes to the solution of an inverse problem and a subsequent predictive simulation. The theory, a demonstration case, and interpretation of data-worth measures are discussed in *Finsterle* [2015]. The basic idea behind the proposed approach is to examine how the addition of potential data (or removal of existing data) reduces (or increases) the uncertainty in select predictions made by a model that is to be calibrated against these data. The arguments for using prediction uncertainty as the ultimate criterion for evaluating data worth was presented above, even though the analysis could be pushed further to the level where monetary values are assigned to each data point or data set within a risk-cost-benefit decision framework.

Including calibration and prediction phases in a single data-worth analysis has considerable advantages, as it automatically identifies data that contain information about those parameters that are most influential on the predictions of interest. However, data-worth analysis can also be limited to the calibration step (i.e., by only evaluating the uncertainty in the estimated parameters), or on the prediction step (i.e., by only evaluating the contribution of individual parameter's prior information on prediction uncertainty).

The data-worth analysis method described here makes use of the Jacobian matrix evaluated at a reference parameter point or the best-estimate parameter set after an inversion. It is therefore a local analysis that furthermore relies on the linearity and normality assumptions that underlie Eqs. (6) and (8).

1. The method consists of the following steps: Select observable variables to be calculated by the model. Categorize them into three groups:

    a. *Actual observations*: Observations that actually exist or are expected to be collected, and that will be used as calibration points. This will be referred to as the reference data set.

    b. *Potential observations*: Observations that could be collected for use in model calibration, should the data-worth analysis prove their value.

    c. *Predictions*: The target output of interest to be calculated by the prediction model. Examples include the energy produced from a geothermal well after 30 years of exploitation; the maximum contaminant concentration at a drinking water well over a regulatory compliance period; the subsidence at a critical infrastructure due to

groundwater pumping; and the risk of induced seismicity caused by reservoir stimulation.

Prior information about the parameters of Step 1 as well as Tikhonov regularization terms can be included as actual or potential observations.

2. Select parameters that potentially influence the prediction of interest. They may be provided as prior information or estimated by inverse modeling.

3. Develop a calibration model that simulates actual and potential observations.

4. Develop a prediction model that simulates target output of interest.

5. Evaluate the sensitivity coefficients (Eq. 1) of all observations defined in Step 2 with respect to all parameters selected in Step 1. The resulting Jacobian matrix has three submatrices, each consisting of the rows that are assigned to one of the three observation groups.

6. Using the set of (presumably or actually) existing calibration points (but omitting potential observations), evaluate the covariance matrix of the estimated parameters, $\mathbf{C}_{pp}$, using Eq. (6).

7. Propagate the uncertainty in the estimated parameters, $\mathbf{C}_{pp}$, to uncertainties of the predictions, $\mathbf{C}_{\hat{z}\hat{z}}$, using Eq. (8).

8. Remove (-) one actual calibration point (or one actual calibration data set), labeled *k*, or add (+) one potential observation point (or potential data set) and re-evaluate the covariance matrix of the estimated parameters, $\mathbf{C}_{pp,\pm k}$.

9. Evaluate the covariance matrix of the model predictions, $\mathbf{C}_{\hat{z}\hat{z},\pm k}$.

10. Scale matrices $\mathbf{C}_{\hat{z}\hat{z}}$ and $\mathbf{C}_{\hat{z}\hat{z},\pm i}$ by the acceptable prediction uncertainty, i.e., $\bar{c}_{ij} = c_{ij} \big/ (\sigma_{z_i} \cdot \sigma_{z_j})$; the scaled matrices are designated by $\bar{\mathbf{C}}_{\hat{z}\hat{z}}$ and $\bar{\mathbf{C}}_{\hat{z}\hat{z},\pm k}$, respectively. (Note that the diagonal elements of $\mathbf{C}_{\hat{z}\hat{z},\pm i}$ indicate whether the attained prediction uncertainty is indeed acceptable.)

11. Evaluate the data worth as a measure of the relative increase in the prediction uncertainty caused by the removal of existing data, or the relative decrease in the prediction uncertainty caused by adding a potential data using one of the following metrics:

Metric 1:
$$\omega_{\pm k} = \begin{cases} 1 - \dfrac{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z},+k})}{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z}})} & \text{when adding potential data} \\ 1 - \dfrac{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z}})}{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z},-k})} & \text{when removing existing data} \end{cases} \tag{16a}$$

Metric 2: $$\omega_{\pm k} = \begin{cases} \dfrac{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z}})}{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z},+k})} - 1 & \text{when adding potential data} \\ \dfrac{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z},-k})}{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z}})} - 1 & \text{when removing existing data} \end{cases} \tag{16b}$$

Metric 3: $$\omega_{\pm k} = 1 - \frac{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z},\pm k})}{\mathrm{tr}(\bar{\mathbf{C}}_{\hat{z}\hat{z}})} \tag{16c}$$

Metric 1 produces data-worth measures between 0 and 1, regardless of whether existing data are removed or potential data added. Metric 2 uses the augmented data set as reference when adding information, and the existing data set as reference when removing information; it yields ω values between 0 and ∞. Metric 3 consistently uses the existing data as reference; it produces ω values between -∞ and 1, where positive values indicate the worth of adding potential data, and negative values that of removing existing information. In all metrics, a value of 0 indicates that collecting data point or data set k has no benefits.

If the data-worth analysis is to be restricted to the calibration phase (i.e., if the goal is to evaluate the worth of information for solving a particular inverse problem), Step 9 is skipped, and Steps 10 and 11 are performed on the estimation covariance matrices $\mathbf{C}_{pp}$ and $\mathbf{C}_{pp,\pm k}$, using the acceptable parameter uncertainty σ_p instead of σ_z. (The diagonal elements of $\mathbf{C}_{pp,\pm k}$ indicate whether the attained estimation uncertainty is indeed acceptable.) Conversely, if the data-worth analysis is to be limited to the prediction phase, Step 6 is skipped and replaced by a covariance matrix that reflects the uncertainty of added or removed prior information. The scaling in Step 10 is necessary as the covariance matrices are not dimensionless, but data worth is evaluated based on a composite measure, i.e., the trace of these matrices (or, alternatively, the determinant). Moreover, the scaling allows the modeler to appropriately weigh the data-worth criteria, reflecting the specific objectives of the study. Data-worth analyses performed with different relative weights assigned to predictions could be used to highlight the trade off between competing prediction targets (similar to a Pareto front in the context of multi-objective optimization).

Unlike a sensitivity analysis, the data-worth analysis properly accounts for redundancy in the information content of closely spaced data points, as well as for the lack of parameter identifiability due to correlations. A data point that contains complementary information about otherwise strongly correlated parameters has higher data worth than a data point with sensitivity coefficients that may be higher, but are similar for concurrently estimated parameters and similar for neighboring data points. If such data redundancy is prevalent, the sum of individual worth values for data that belong to a certain data set is smaller than the worth of adding the entire data set. Relying on the results of a sensitivity analysis to identify observations suitable for model calibration or the development of a robust prediction model may thus be misleading.

Data worth depends not only on the reference parameter set, but also on the amount and quality of all the other data points presumed available for model calibration, including prior information.

A data-worth analysis should thus be repeated for different reference parameter sets and different reference data sets with varying assumptions about the error structure of the residuals.

In summary, the data-worth value $\omega_{\pm k}$ can be interpreted as the contribution of, respectively, adding (or removing) potential (or existing) data points (or data sets) to the reduction (or increase) in overall estimation uncertainty, which is measured by the trace of the scaled parameter or prediction covariance matrix (see Figure 1). Starting with reference data, the uncertainty analysis (Eqs. (6) and (8)) determines whether the estimation or prediction uncertainties are sufficiently low, i.e., acceptable for the decision maker. If so, the data-worth analysis indicates which existing data could be removed to arrive at a cheaper design with minimal impact on the quality of the estimated parameters and without substantially increasing prediction uncertainty. If uncertainties are unacceptably high, the data-worth analysis suggests which potential data could be added to the reference data set to effectively reduce the estimation and prediction uncertainty. A few iterations of this process—conducted prior to actual data collection—is likely to yield a testing or monitoring design that is robust and effective in reaching the calibration and modeling goals.

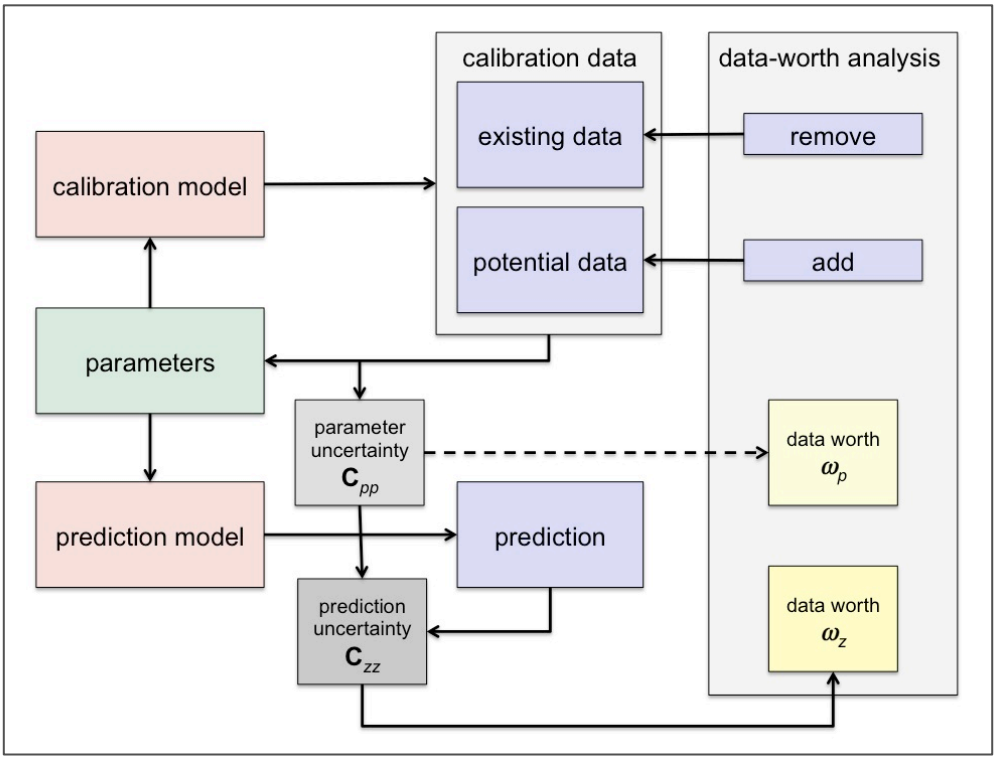

Figure 1: Elements of data-worth analysis. Influential parameters to be determined from calibration data are shared with a prediction model to determine how the removal of existing or addition of potential data impacts the estimation or prediction uncertainty. Data worth may also be calculated based on parameter uncertainty (dashed line).

3. INPUT AND OUTPUT FORMATS

3.1 Input Formats

The input formats are similar to the ones for PE, UA and other iTOUGH2 analysis methods, which allows for a seamless transition from GSA to PE, UA or other analyses using iTOUGH2. The `> PARAMETER` block includes the parameters of interest, their transformation, reference value, uncertainty ranges and/or distributions, and other details (see fourth-level commands). The `> OBSERVATION` block is exactly the same as for other applications, and includes the outputs or performance measures of interest, transformation, standard deviation, and other details (see fourth-level commands). Note that SA generally does not require measured data, as it only refers to the simulation output; this can conveniently be indicated by the use of command `>>>> NO DATA`.

3.1.1 Local Sensitivity Analysis

The local sensitivity analysis requires information about the parameter's reference value, the scaling coefficients σ_p and σ_z, the perturbation factor δ, and the method used to numerically calculate the derivatives.

The reference parameter values, i.e., the local point **p*** in the parameter space for which the derivative-based local sensitivity analysis is performed, is given by the values in the TOUGH2 input file, commands `>>>> PRIOR`, `>>>> GUESS` or `>> GUESS`, each overwriting the previous definition.

The scaling factor σ_p is given using command `>>>> VARIATION` or `>>>> DEVIATION` within block `> PARAMETER`. Command `>>>> VARIATION` is specifically used for scaling sensitivity coefficients, while command `>>>> DEVIATION` scales the sensitivity coefficients and at the same time weighs the difference between the prior information and the parameter value as estimated during an inversion. The scaling factor σ_z is given using command `>>>> DEVIATION` (or related commands) within block `> OBSERVATION`.

The perturbation factor δ applied to all parameters is given by command `>>> PERTURB`; it can be overwritten for individual parameters by command `>>>> PERTURB`. If a negative perturbation value is given, the parameter will be changed by the specified value rather than a percentage of the parameter value. Specifying a negative perturbation is required if the parameter value is zero.

Figure **2** shows as an example an iTOUGH2 input file that invokes a local sensitivity analysis in which the influence of porosity and residual gas saturation on the amount of entrapped gas and the spreading of the gas plume is evaluated over a 30 year period of CO_2 injection.

```
> PARAMETER
  >> POROSITY
     >>> MATERIAL      : RESER
         >>>> VARIATION:  0.05 (= sigma_p)
         <<<<
     <<<

  >> RELATIVE permeability
     >>> MATERIAL      : DEFAU
         >>>> INDEX     :     2
         >>>> GUESS     :   0.0
         >>>> PERTURB   : -0.01
         >>>> VARIATION:  0.10 (= sigma_p)
         <<<<
     <<<
  <<

> OBSERVATION
  >> TIME: 30 EQUALLY spaced YEARS
       1.0 30.0

  >> TOTAL TRAPPED GAS VOLUME
     >>> MATERIAL: RESER
         >>>> NO DATA
         >>>> DEVIATION: 1000.0 m^3 (= sigma_z)
         <<<<
     <<<

  >> SECOND MOMENT of GAS plume in X-direction
     >>> entire MODEL
         >>>> NO DATA
         >>>> DEVIATION: 2500.0 m^2 (= sigma_z)
         <<<<
      <<<
  <<

> COMPUTATION
  >> OPTION
     >>> local SENSITIVITY ANALYSIS
     <<<
  >> JACOBIAN
     >>> PETURB parameters by : 5 %
     >>> use CENTERED finite differences
     <<<
  <<
<
```

Figure 2: iTOUGH2 input file for local sensitivity analysis.

3.1.2 Morris Global Sensitivity Analysis

Unlike for the local sensitivity analysis of Section 3.1.1, global sensitivity analyses require that the bounds of the *n*-dimensional parameter hypercube be defined. The global sensitivity measures will be suitable averages of the sensitivities within this parameter hypercube. In iTOUGH2, the admissible parameter range is defined for each parameter using command `>>>> RANGE`. Other than this additional requirement, parameters are defined in the usual way using subcommands of iTOUGH2 block `> PARAMETER`, as shown in Figure **2**.

Figure **3** shows the `> COMPUTATION` block that invokes a Morris global sensitivity analysis. Note that the number of partitions *k* has to be an even number (Morris, 1991). If a simulation fails, the affected path may be discarded and a new parameter set generated by resampling.

```
> COMPUTATION
  >> OPTION
     >>> SENSITIVITY ANALYSIS          : MORRIS OAT
         >>>> number of PATHS, r       : 10
         >>>> number of PARTITIONS, k  : 6
         >>>> RESAMPLE up to           : 10 failed trajectories
         <<<<
     <<<
  <<
```

Figure 3: iTOUGH2 block `> COMPUTATION` for the Morris One-At-A-Time global sensitivity analysis.

3.1.3 Sobol' Global Sensitivity Analysis

For the Sobol' method, the `> PARAMETER` block is the same as the one for Monte Carlo simulations in iTOUGH2. The user needs to specify the parameter ranges, and the parameter transformation, e.g., `>>>> LOGARITHM`. In the `> COMPUTATION` block (Figure **4**), the user needs to select keyword `SALTELLI` or `SOBOL` together with command `>>> SENSITIVITY ANALYSIS`, and specify the number of realizations (*r*) and the seed number to initiate the generation of random numbers.

```
> COMPUTATION
  >> OPTION
     >>> SENSITIVITY ANALYSIS       : SALTELLI
         >>>> number of SAMPLES     : 1000
         >>>> SEED number           : 7777
         <<<<
     <<<
  <<
```

Figure 4: iTOUGH2 block `> COMPUTATION` for the Sobol' global sensitivity analysis.

3.1.4 Data-Worth Analysis

A data-worth analysis is invoked by command `>>> DATA-WORTH ANALYSIS`, which is in the `>> OPTION` section of the `> COMPUTATION` block. By default, the data worth is evaluated for each calibration point. If keyword `SET` is added to the command line, data worth is instead evaluated for an entire data set (a data set is defined by each third-level command in block `> OBSERVATION`, and includes all observations contained within an individual `>>>> DATA` block). Parameters are defined analogous to a local sensitivity analysis (see Section 3.1.1 and Figure **2**).

As discussed in Section 2.4, a data-worth analysis can be performed in two ways:

- *Option A*: Evaluate the worth of calibration data with respect to their ability to reduce the uncertainties of estimated parameters (see $\mathbf{C}_{pp}$ in Eq. (6))
- *Option B*: Evaluate the worth of calibration data with respect to their ability to reduce prediction uncertainties (see $\mathbf{C}_{zz}$, Eq. (8))

If *Option A* is chosen, the `> OBSERVATION` block contains all observations potentially (or actually) used for estimating the parameters defined in block `> PARAMETER`. Observations are thus defined as those in a local sensitivity analysis (see Figure **2**).

If *Option B* is chosen, the user must first develop a simulation that consists of two sub-models. The first model (referred to as the calibration model) simulates the processes and conditions during data collection; the second model (referred to as the prediction model) simulates a future system behavior, whose prediction uncertainty shall be minimized through the collection of high-worth calibration data. The calibration and prediction models are preferably very similar (as the parameters estimated by inverse modeling are always scale-, process-, and model-related, which means they should be used in a model with consistent structure). Conversely, they may be very different (e.g., if formation properties are determined through inversion of data collected in a laboratory experiment of a core sample, and then used in a prediction model of field-scale reservoir behavior).

There are different ways these two models could be implemented:

- A single model represents both the calibration and prediction phases. This is certainly the easiest and preferred scenario, as the parameters estimated during the calibration phase are used in a prediction model that has a similar scale and model structure. An example would be a geothermal reservoir model, in which injection, observation and production data during the first 5 years of operation are dedicated for model calibration, and the same model is then used to predict reservoir performance for another 25 years. The data-worth analysis will determine the value of individual calibration data points during the first 5 years with respect to their ability to reduce the uncertainty in the predicted geothermal productivity after 30 years. This approach will be further illustrated in Section 4.1.2.
- The calibration and prediction models are two separate models that are combined into a single model and run concurrently. First, two separate meshes are generated (with unique element names). These two meshes are then combined into a single mesh, which

is defined by a single ELEME and a single CONNE block. The two models must share the same materials so the parameters estimated using the calibration model are also used in the prediction model. Sample Problem 2, Part 2 (also used to demonstrate the Morris and Sobol' global sensitivity analyses, see Section 4.1.1) uses this approach to concurrently analyze a single parameter set using observations that refer to more than on model.

- The two separate calibration and prediction models are run sequentially. However, this approach requires that at least one of the two models be executed as an external model, which needs to be linked to iTOUGH2 using the PEST protocol (Finsterle, 2010). Again, the two models must share the same parameter values. If only one of the models is run externally, whereas the other model is fully integrated into iTOUGH2, each parameter must be defined twice (once as a regular, internal parameter, and once as an external PEST parameter) and linked to one another using command >>>> TIED.

In the > OBSERVATION block, the user must indicate which observations refer to potential (or actual) calibration data, and which observations refer to target model predictions. Recall that the data-worth analysis is performed only for calibration data based on their ability to reduce prediction uncertainty. Actually or presumably available calibration data are defined analogous to those in a regular sensitivity analysis or inversion (see, for example, Figure **2**); these data constitute the reference data set and will be removed during the data-worth analysis. Potentially measured data are indicated by command >>>> POTENTIAL; these data will be added during the data-worth analysis. The observations that refer to output from the prediction model are indicated by command >>>> PREDICTION or command. These observations provide the criteria based on which the value of actual or potential calibration data is calculated.

If no prediction data are provided, *Option A* is automatically invoked; if one or more prediction data points are provided, *Option B* is invoked, unless keyword PARAMETER is added to the command line (e.g., >>>> DATA-WORTH ANALYSIS based on PARAMETER covariance matrix Cpp).

The command line also accepts keyword DETERMINANT, in which case data worth is evaluated using the determinant (rather than the trace) of the covariance matrix (see Eq. (6) or (8)).

A data-worth analysis can be explicitly requested as described above. Moreover, a data-worth analysis is also performed automatically after a local sensitivity analysis and an inversion that uses a derivative-based minimization algorithm. (Using command >>> SENSITIVITY ANALYSIS or command >>> DATA-WORTH ANALYSIS thus yields the same output.) However, since the repeated evaluation of the estimation (and prediction) covariance matrices (see Step 8 in Section 2.4) may become computationally expensive, the data-worth analysis is performed automatically only if the number of calibration points is smaller than 500. If more than 500 calibration points are given, the data-worth analysis is performed for each data set (rather than each data point), reducing computational cost. Keywords with command >>> DATA-WORTH ANALYSIS, if given as a subcommand of >> OUTPUT, provide additional control. The data-worth analysis after a local sensitivity analysis or inversion can be suppressed using command >>> OMIT DATA-WORTH ANALYSIS. Keywords PARAMETER, SET, and DETERMINANT are also available in

the `>> OUTPUT` block version of command `>>> DATA-WORTH ANALYSIS`; they have the same effect as described above. An example of a data-worth analysis is given in Section 4.1.2.

3.2 Output Formats

3.2.1 Local Sensitivity Analysis

The iTOUGH2 output file contains the $n \times m$ scaled sensitivity matrix after a sensitivity analysis has been performed specifically, but also after any inversion that uses a derivative-based minimization algorithm. The unscaled sensitivity (or Jacobian) matrix can be requested using command `>>> SENSITIVITY` in the `>> OUTPUT` block. In addition, the following composite sensitivity measures are evaluated:

- Sum of absolute scaled sensitivity coefficients for each parameter
- Sum of absolute scaled sensitivity coefficients for each observation
- Sum of absolute scaled sensitivity coefficients for each data set and each parameter
- Sum of absolute scaled sensitivity coefficients for each data set and all parameters
- Sensitivity of objective function with respect to each parameter (Eq. (5))
- Relative contribution of parameter uncertainty to prediction uncertainty, assuming the parameter is fixed
- Relative contribution of parameter uncertainty to prediction uncertainty, assuming that all except on of the parameters are fixed

In addition, a linear uncertainty analysis is performed to obtain the covariance matrix of the estimated parameters (see Eq. (6)) as well the uncertainty of each model prediction (see Eq. (8)). A data-worth analysis is also performed. If no data are provided, these analyses are based on an estimated error variance of 1.0, i.e., they assume that the data were matched to a level that is consistent with the *a priori* defined error variances specified for each data point.

3.2.2 Morris Global Sensitivity Analysis

iTOUGH2 computes the mean EE, mean |EE| and SD of EE for all the outputs of interest and for each parameter. In addition, it provides those three values for the mean of all the outputs (heading: `SYSTEM STATE`) and for the objective function S (heading: `OBJECTIVE FUNCTION`) specified in the same manner as other applications in iTOUGH2. Such composite indices are useful when the number of output values is large (e.g., time series).

3.2.3 Sobol' Global Sensitivity Analysis

In the iTOUGH2 output file, the sensitivity and total sensitivity indices are provided in a matrix form for all the outputs and for each parameter. The indices are also calculated for the system state and objective function.

3.2.4 Data-Worth Analysis

The results of the data-worth analysis are printed as the last column in the `RESIDUAL ANALYSIS` section of the iTOUGH2 output file. Only observations that serve as actual or potential calibration points show a data-worth value (labeled `DWi`); observations representing model predictions show the text "`predict.`" in the last column.

`DWi` is the data worth of the corresponding observation as a percentage of the total worth of all calibration points. The absolute data-worth value can be calculated by multiplying `DWi` with the scaling factor, which is reported in the legend entry for `DWi`, located at the top of the residual analysis section of the output file. The legend also indicates (i) whether the data-worth analysis was performed for individual observation points or entire data sets, (ii) whether $\mathbf{C}_{pp}$ or $\mathbf{C}_{\hat{\varepsilon}\hat{\varepsilon}}$ was used as the target criterion, and (iii) whether the trace or determinant of these matrices was evaluated. Finally, the data worth of entire data sets and data types is reported in the `DWA` columns of the summary tables at the end of the residual analysis section.

An example of a data-worth analysis is given in Section 4.1.2.

4. SAMPLE PROBLEMS

4.1.1 Sensitivity Analyses

To demonstrate the iTOUGH2-GSA module, we use the gas-pressure-pulse-decay experiment, which is Sample Problem # 2 documented in *iTOUGH2 Sample Problems* (Finsterle, 2010). The experiment description is repeated here for completeness. The experiments were conducted using a specially designed permeameter with small gas reservoirs. A schematic of the experimental apparatus is shown in Figure **5**. To conduct a test, the upstream reservoir is rapidly pressurized to a value about 300 kPa above the initial pressure of the system using nitrogen gas. Gas starts to flow through the dry sample, and the pressures in both the upstream and downstream reservoirs are monitored as they equilibrate with time.

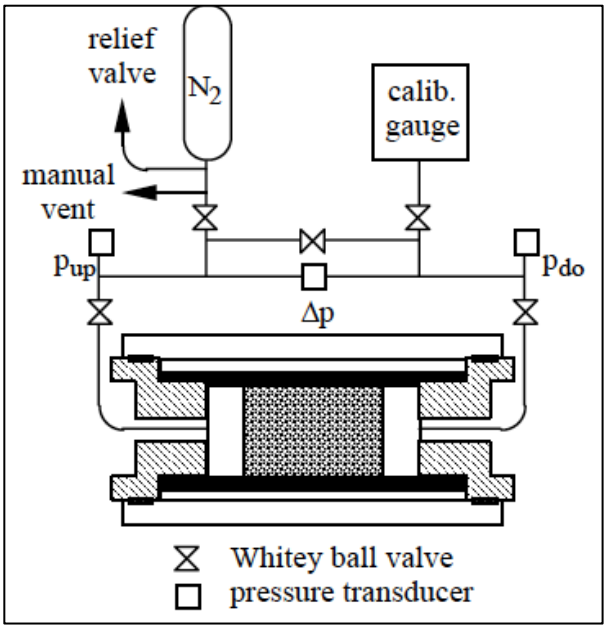

Figure 5: Schematic of gas-pressure-pulse-decay apparatus.

In porous media with very low permeability and porosity, gas mass flow F [$kg \cdot s^{-1} \cdot m^{-2}$] may be enhanced as a result of slip flow known as the Klinkenberg effect.

$$\mathbf{F} = -k\left(1+\frac{b}{P_g}\right)\frac{\rho_g}{\mu_g}\nabla P_g \tag{19}$$

Here, k is the absolute permeability, ρ_g is the gas density, μ_g is the gas dynamic viscosity, and P_g is the gas pressure. The term in parentheses accounts for enhanced gas slip flow, which occurs when the mean free path of the molecules is large relative to the characteristic dimension of the pores. Slip flow is important at low pressures and in small pores, when a significant fraction of molecular collision is with the pore wall rather than with other gas molecules.

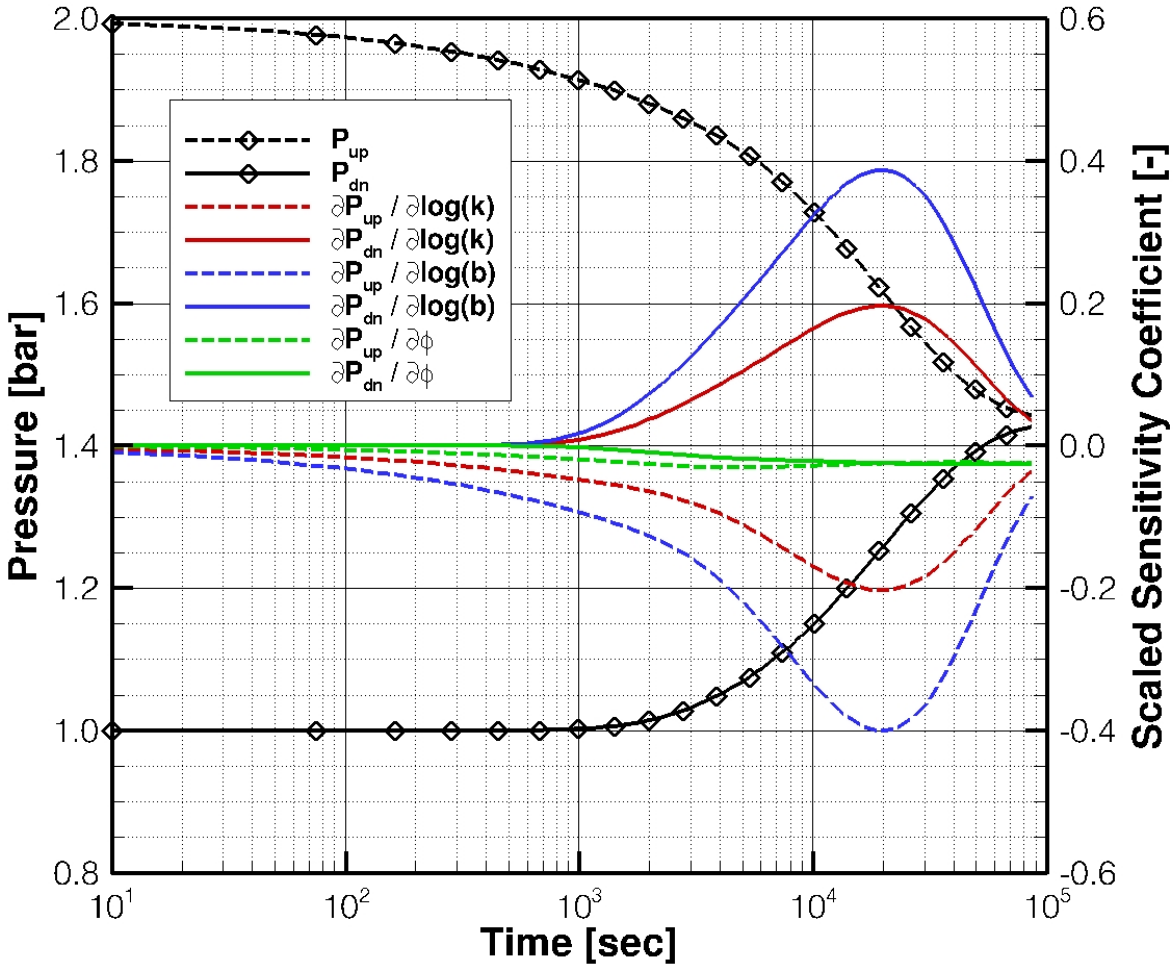

Figure 6: Pressure (symbols) in upstream and downstream reservoirs during gas-pressure-pulse-decay experiment and scaled sensitivity coefficients with respect to log(k), log(b), and ϕ as a function of time.

Figure **6** shows the scaled sensitivity coefficients of the upstream and downstream reservoir pressures with respect to the three parameters as a function of time. The sensitivity coefficients with respect to porosity are negative because an increase in porosity leads to a decrease in pressure in both the upstream and downstream reservoirs. The porosity's influence increases with time and reaches a constant, non-zero value at steady state, when the pressure in the upstream and downstream reservoirs equilibrate slightly below half of the imposed pressure increase (of 300 kPa), because some of the injected gas is needed to compress the gas in the sample pore space. The steady-state pressures are thus sufficient to independently determine porosity.

Conversely, upstream/downstream reservoir pressures have negative/positive sensitivity coefficients, respectively, with respect to permeability and the Klinkenberg factor, as in increase

in gas mobility leads to a decrease/increase in the upstream/downstream reservoir pressure. Moreover, log(k) and log(b) have zero (or a small) influence at early times (when the gas has not yet reached the downstream reservoir). The influences of these two parameters reaches their maxima when the reservoir pressures change most, and approach again zero as gas flow ceases near steady-state conditions. It is important to note that the sensitivity coefficients for log(k) and log(b) are very similar, i.e., they are nearly linearly dependent of each other, which can be expected by inspecting Eq. (19). As a result, the two parameters are highly correlated and thus cannot be determined independently using a single gas-pressure-pulse-decay experiment. This is clearly reflected in a correlation coefficient of the estimation covariance matrix that is very close to -1.0, i.e., an unit increase in one of the two parameters can be almost completely compensated by a unit decrease in the other parameter. Such correlation coefficients are not directly obtained from a local sensitivity analysis, but require an uncertainty analysis. The issue of strong parameter correlations can be resolved, and all three parameters can be estimated accurately with an appropriate change in the experimental design, as discussed in detail in Finsterle and Persoff (1997).

The composite scaled sensitivity measures indicate that (1) the downstream pressure at 0.5 days is the most sensitive observation to changes in all three parameters, (2) the time series of all upstream pressures contains overall more information about the three parameters than the downstream pressures, and (3) the parameters log(k) and log(b) are almost equally influential and significantly more influential on pressure observations than porosity. Moreover, from the covariance matrix of the estimated parameters it can be deduced that (4) the estimation uncertainty of log(k) and log(b) would be extremely large as a result of (5) a very strong negative correlation between these two parameters, leading to (6) very low parameter independence as measured by the ratio between the conditional and marginal standard deviations, whereas (7) porosity could be estimated relatively independently with very low estimation uncertainty. The error propagation analysis indicates that (8) the uncertainty in the predicted reservoir pressures due to uncertainty in the estimated parameters would be less that 400 Pa, and (9) all observations are very well controlled by other observations, indicating some redundancy in the information content of neighboring data points. Finally, the data-worth analysis suggests that (10) the last observations in the upstream and downstream reservoirs have the highest potential to reduce the trace of the estimation covariance matrix, and thus possess the highest data worth.

The local sensitivity analysis is based on reference parameter values, which are uncertain as they rely on prior information that does not contain the additional data obtained from the experiment. The results of the local sensitivity analysis may be misleading if the reference parameters are different from the true values, and if the model is highly nonlinear. In GSA, on the other hand, we can compute the sensitivity over the parameter range so that the sensitivity indices are more robust and more representative with respect to the defined parameter ranges or distributions.

The iTOUGH2 input files are *samMOATi* and *samSOBOLi* included in the Sample directory for the Morris and Sobol' methods, respectively. They are quite similar to the original sample problem (*sam2pli*). The difference is that (1) the parameter range is specified for the Morris method (`>>>> RANGE` in the `> PARAMTER` block), (2) the sensitivity methods and the number of paths or samples is specified in the `> COMPUTATION` block (see Figure **3** and Figure **4**). Among the three parts of the experiments, we use Part 1 (inversion of a single gas-pressure-pulse-decay experiment) and Part 3 (concurrent inversion of three gas-pressure-pulse-decay

experiments performed on three pressure levels; see report *iTOUGH2 Sample Problems* and Finsterle and Persoff (2007)). We evaluate the sensitivity of the pressure in the downstream reservoir (p_{do} in Figure **5**).

Figure 7 shows the time evolution of the sensitivity indices from Morris and Sobol' methods in the Part-1 experiment. The number of simulations is 40 (r = 10, n = 3), and 5,000 (n = 3, r = 1,000) for the Morris and Saltelli methods, respectively. The confidence intervals are shown so that we can take into account the limited number of simulations when we compare the importance of the parameters. For the Morris method, the output scaling factor σ_z is 1000 Pa, which is the same value as used for the local sensitivity method.

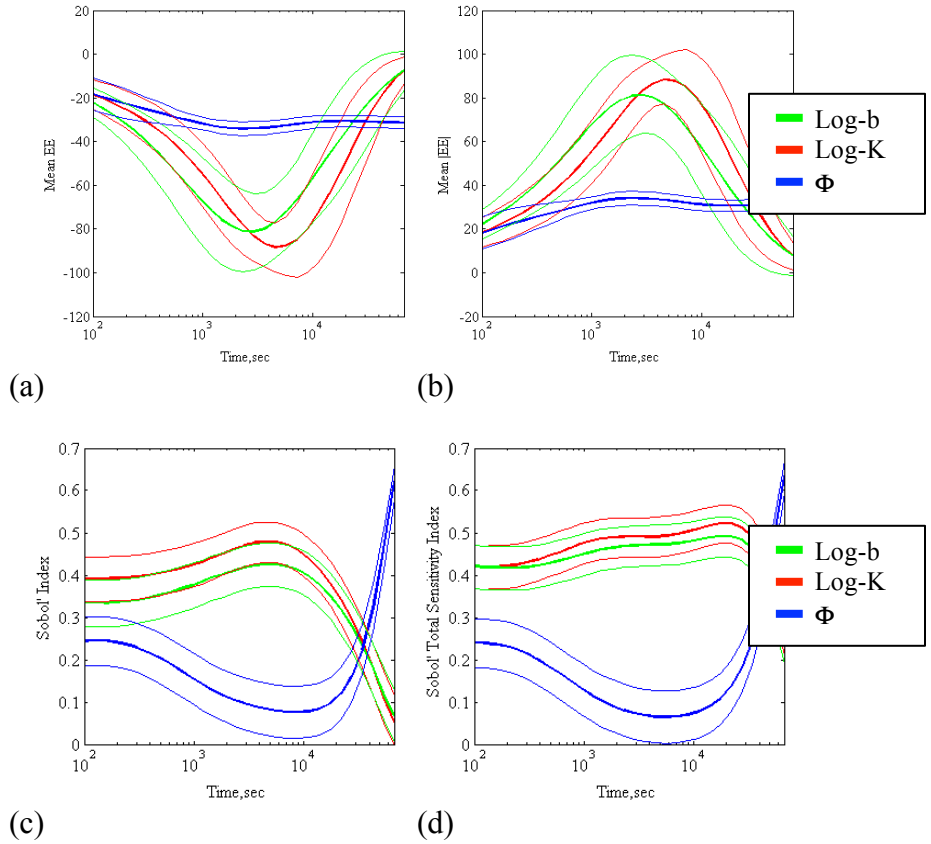

Figure 7: Time evolution of sensitivity index: (a) Mean EE, (b) Mean |EE| (c) Sobol' index, and (d) Sobol' total sensitivity index. In (a)-(d), the thin lines represent the 95% confidence intervals.

At each time point, we may define the importance ranking. Figure 7 shows that the important parameters change with time. In Figure 7a, Mean EE provides the sign of the sensitivity, which helps understand the physics of the problem. The Mean-EE values are all negative, which means that increasing any of these three parameters decreases the downstream pressure. In Figure 7 b, Mean |EE| suggests that the permeability and Klinkenberg factors have a large impact around 10^3-10^4 seconds, while the porosity becomes more important after 5×10^4 seconds. Although the number of paths and number of simulations are not large (r = 10), the confidence intervals are separated well enough to confirm the relative importance (such as the importance of porosity in a

later time). This is the same finding as obtained with the local sensitivity method, although, with GSA, we can account for the nonlinearity and/or interaction effects, since we compute the average of the effects over the prescribed parameter range. At the end of the experiment, the porosity has significantly higher influence than the other two parameters. The physical explanation for this change in influence has been given above for the discussion of the local sensitivity analysis results. Based on this result, we can capture the intervals in which the three parameters are influential, and this experiment time is large enough to estimate the porosity.
In Figure 7c and d, the Sobol' indices provide the relative contribution of each parameter to the uncertainty of the output. The Sobol' indices have better quantitative interpretation. From Figure 7c, for example, we may conclude that both the permeability and Klinkenberg factor account for approximately 40% of the pressure variance between 10^3 and 10^4 seconds, while the porosity accounts for only approximately 10%. Since the indices are normalized by the variance at each time step, the ranking of parameter importance is more easily recognized than Mean |EE|, although the effect of each parameter at different times cannot be compared. The patterns of the three parameters are similar in Figure 7c and d, although the total sensitivity index (Figure 7d) is higher for the permeability and Klinkenberg factor. The higher total sensitivity index compared to the Sobol' index suggests that the permeability and Klinkenberg factor has an interaction effect. Such an interaction effect is evident from Eq. (19), but this is noted here to explain the use of GSA to interpret interaction effects. Such interaction effects are often not evident, especially for large-scale models involving complex processes.

Figure 8 shows a cross-plot between the mean and SD of EE, following Morris (1991). Each curve corresponds to the time evolution for a parameter's sensitivity index. The two black lines represent Mean EE = ±2SEM. These lines can be drawn by substituting SEM by $SD/r^{0.5}$, following the definition of SEM. Since SEM decreases with increasing number of paths (r), the slope of the line increases with the number of paths. Although the number of paths ($r = 10$) is small, the SD is small, and all the parameters are below the black lines, indicating that their non-zero impact is statistically significant. All the parameters have a non-zero value of SD, indicating that they have nonlinearity and/or interaction effects. The ratio between the mean and SD of EE is larger for the permeability and Klinkenberg factor, which suggests the larger nonlinearity and/or interaction effects of these two parameters.

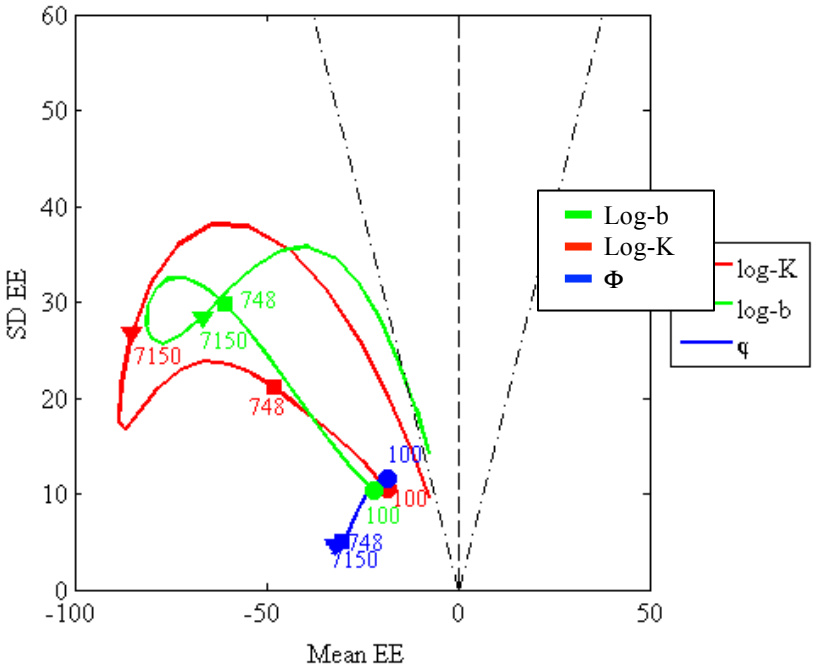

Figure 8: Mean EE vs. SD from the Morris method. The circle, square and triangle on each curve represent 100, 748 and 7150 seconds, respectively. The end of each line corresponds to 6.705×10^4 seconds.

Figure 9 shows the difference between the Sobol' index and the total sensitivity index as a function of the Sobol' index, comparing the first-order effect and the interaction effects. All the parameters show interaction effects, since the difference is larger than zero. The permeability and Klinkenberg factor have a particularly large difference relative to their Sobol' indices, suggesting that they have a large interaction effect compared to the first-order effect. Comparing Figure 8 and Figure 9 allows us to separate interaction from nonlinearity effects, since Morris's SD of EE includes both, but the difference between the Sobol' indix and total sensitivity index represents only the interaction effects. Although Eq. (19) suggests this flow process includes both nonlinearity and interactions (between permeability and Klinkenberg factor), SD of EE of the permeability and Klinkenberg are caused mainly by the interaction effects.

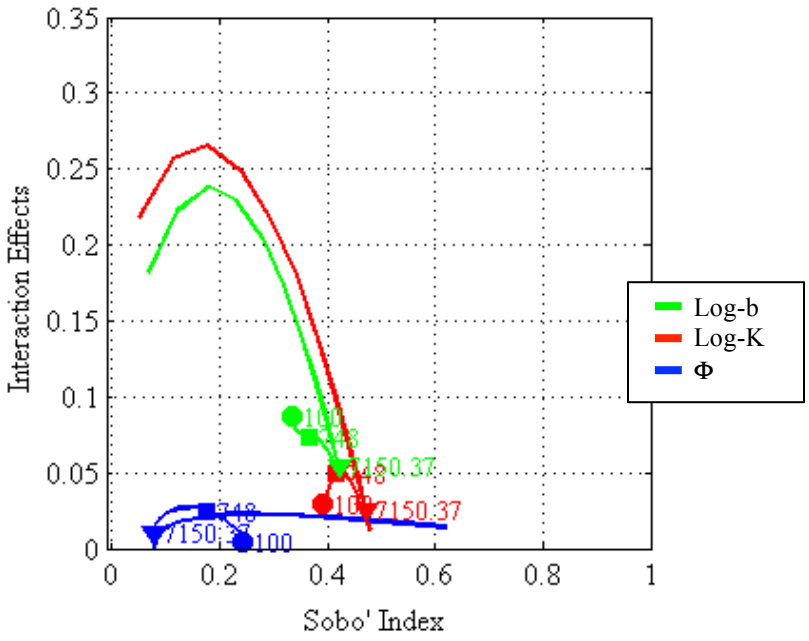

Figure 9: The interaction effect (the difference between the Sobol' index and the total sensitivity index) as a function of Sobol' index. The circle, square and triangle on each curve represent 100, 748.48 and 7150.37 seconds, respectively. The end of each line corresponds to 6.705×10^4 seconds.

4.1.2 Data-Worth Analysis

The following example demonstrates how a data-worth analysis is set up in iTOUGH2, and how some of the results of the analysis are interpreted. The test case is based on a modified version of the five-spot geothermal injection/production problem described as Problem No. 4 in the *TOUGH2 User's Guide* (Pruess et al., 2012) as well as Problem No. 3 in *iTOUGH2 Sample Problems* (Finsterle, 2010). The problem considers a large well field with wells arranged in a five-spot configuration (Figure 10). Because of symmetry, only 1/8 of the basic pattern needs to be modeled, with an injection and production well located in the corners of the model, pumping at a constant rate of 30 kg/s for 30 years. For simplicity, the geothermal reservoir, which has a uniform initial temperature of 300°C and pressure of 85.93 bars, is represented by a single layer, discretized into 36 primary elements. Each of these elements is partitioned into five overlapping elements using the MINC methodology to represent the fracture network as a continuum, which is embedded in multiple matrix continua of increasing average distance from the fractures. As a modification to the original formulation, the injected water is considered to contain a conservative tracer. The simulations are performed using the equation-of-state (EOS) module 1 (Pruess et al., 2012), which handles nonisothermal, two-phase flow of two water components. An excerpt of the TOUGH2-EOS1 input file is shown in Figure 11.

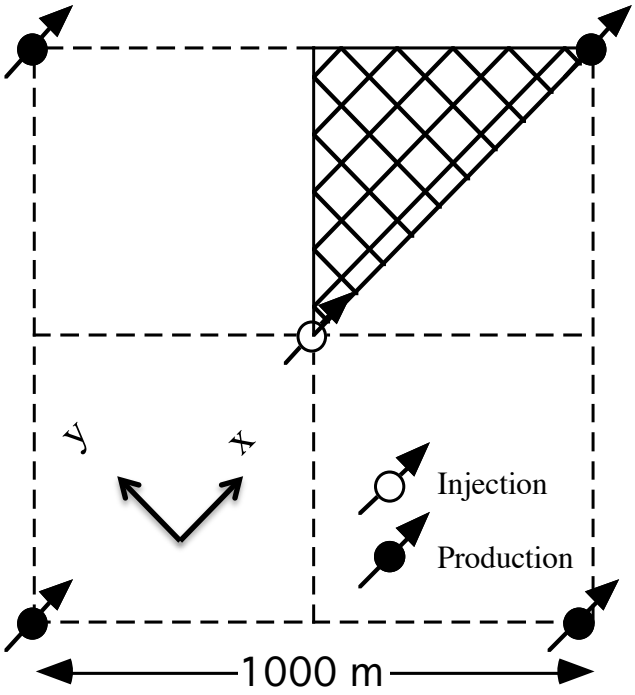

Figure 10: Five-spot well pattern with grid for modeling a 1/8 symmetry domain.

```
Five-spot geothermal reservoir model (cf. SPE-18426), for data-worth analysis
ROCKS----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
POMED          2650.       .01    6.E-15    6.E-15    6.E-15       2.1     1000.
FRACT          2650.       .50    6.E-15    6.E-15    6.E-15       2.1     1000.
MATRX          2650.    1.E-10    0.E-15    0.E-15    0.E-15       2.1     1000.

START----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
----*----1 MOP: 123456789*123456789*1234 ---*----5----*----6----*----7----*----8
PARAM----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
   2 999     99910000000000000 4 0  6
 -1000.00                1.E5 3.15576E7 KA 1
     1.E-5                                              1.E-8
                300.                0.01                1.E-4
MULTI----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
    2    3    2    6
RPCAP----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
    3     .30       .05
    1             0.        0.        1.
GENER----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
 AA 1INJ 1                         COM2       3.75     5.0E5
 KA 1PRO 1                         MASS      -3.75

ELEME----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
 AA 1          POMED0.1906E+060.1250E+04          0.        0.        0.1525E+03
 BA 1          POMED0.7625E+060.5000E+04          0.7071E+020.        0.1525E+03
 CA 1          POMED0.7625E+060.5000E+04          0.1414E+030.        0.1525E+03
 DA 1          POMED0.7625E+060.5000E+04          0.2121E+030.        0.1525E+03
 EA 1          POMED0.7625E+060.5000E+04          0.2828E+030.        0.1525E+03
 FA 1          POMED0.7625E+060.5000E+04          0.3536E+030.        0.1525E+03
 GA 1          POMED0.7625E+060.5000E+04          0.4243E+030.        0.1525E+03
 HA 1          POMED0.7625E+060.5000E+04          0.4950E+030.        0.1525E+03
 IA 1          POMED0.7625E+060.5000E+04          0.5657E+030.        0.1525E+03
 JA 1          POMED0.7625E+060.5000E+04          0.6364E+030.        0.1525E+03
 KA 1          POMED0.1906E+060.1250E+04          0.7071E+030.        0.1525E+03
 BB 1          POMED0.7625E+060.5000E+04          0.7071E+020.7071E+020.1525E+03
 CB 1          POMED0.1525E+070.1000E+05          0.1414E+030.7071E+020.1525E+03
...
 GE 1          POMED0.7625E+060.5000E+04          0.4243E+030.2828E+030.1525E+03
 FF 1          POMED0.3812E+060.2500E+04          0.3536E+030.3536E+030.1525E+03
HTX00          POMED        0.

CONNE----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
 AA 1 BA 1                  10.3536E+020.3536E+020.1078E+05
 BA 1 CA 1                  10.3536E+020.3536E+020.1078E+05
 BA 1 BB 1                  20.3536E+020.3536E+020.2157E+05
 CA 1 DA 1                  10.3536E+020.3536E+020.1078E+05
...
 FE 1 GE 1                  10.3536E+020.3536E+020.2157E+05
 FE 1 FF 1                  20.3536E+020.3536E+020.2157E+05

INCON----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8

MESHMAKER1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
MINC
PART THRED     DFLT
  5  4OUT       100.     100.0     100.0
       .02       .08       .20       .35

ENDCY----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
```

Figure 11: Excerpt of TOUGH2-EOS1 input file for simulating five-spot geothermal reservoir operation.

For the data-worth analysis, we consider a scenario in which a model is to be developed for the prediction of two specific performance measures, namely (1) the enthalpy of the produced fluid mixture, and (2) the temperature in the center of the matrix near the edge of the reservoir; both measures are evaluated after 30 years of exploitation. They may be considered relevant to evaluate the field's long-term productivity and sustainability. Production enthalpy and reservoir temperature shall be predicted as accurately as possible using a model that is calibrated against data to be collected during the first five years of operation. A data-worth analysis is performed to examine which data set among an assembly of planned and potential data sets helps most reaching the objective of obtaining reliable long-term predictions of enthalpy and temperature, and which data set could be omitted without considerable loss of model predictability.

The data-worth analysis as formulated here is of *Option B* (see Section 3.1.4), i.e., the criterion used to compute data worth is prediction uncertainty (rather than uncertainty of estimated parameters). Nevertheless, the prediction uncertainty is linked to the accuracy with which the model parameters can be estimated using data from the first five years of production. This means that a set of influential parameters needs to be identified that affect the model predictions of interest. This identification can be done using one of the local or global sensitivity analysis options described in Sections 2.1–2.3. However, here we do not perform a separate sensitivity analysis for the purpose of selecting influential parameters. Instead, this step is integrated in the data-worth analysis, which highlights the importance of (1) accounting for redundancies in the information content of data sets, and (2) specifying the overall objectives of data collection in support of modeling.

Once the set of potentially influential parameters has been selected, data sets (type, location, sampling frequency, measurement accuracy) that are likely to contain information about these parameters have to be identified. These data sets are further designated as (1) existing data sets (or data sets that are planned to be collected and available for model calibration), and (2) potential data sets (i.e., data sets that are proposed to be collected if the data-worth analysis confirms that their contributions to the modeling goals are substantial). It is primarily the value of these potential calibration data sets what will be examined in the data-worth analysis; however, the analysis also examines the loss of prediction accuracy should one of the planned (or actually existing) data sets be removed. Implementation of these three steps (i.e., (1) parameter definition, (2) data set definition, and (3) data-worth analysis) in iTOUGH2 are discussed in detail next; the corresponding iTOUGH2 input file is *samDWi*.

Figure 12 shows the `> PARAMETER` block of the iTOUGH2 input file *samDWAi*. Seven potentially influential parameters are selected, with reference values (indicated in brackets) as specified in the TOUGH2 input file (Figure 11): (1) $\log(k_X)$ [-14.2 $\log(m^2)$], (2) $\log(k_Y)$ [-14.2 $\log(m^2)$], (3) porosity of the fracture continuum [50%], (4) residual liquid saturation of the fracture continuum [30%], (5) thermal conductivity [2.1 W/m/°C], (6) fracture spacing [100 m], and (7) initial reservoir temperature [300°C].

Commands `>>>> VARIATION` and `>>>> DEVIATION` are used to specify the scaling factor of the parameters, σ_p; the latter command furthermore indicates that the initial `>>>> GUESS` should be treated as prior information. It is important to realize that the (somewhat subjective) choice of σ_p has no impact on an *Option B* data-worth analysis; however,

it affects the composite sensitivity measures of the local sensitivity analysis as well as an *Option A* data-worth analysis.

Figure 13 shows the `> OBSERVATION` block in which planned and potential data sets for model calibration as well as the prediction variables are defined. First, 61 points in time are defined, assuming that a data point will be collected every month for the first five years of exploitation and used for model calibration. An additional point is defined at 30 years, at which time the calculated production enthalpy and reservoir temperature will be extracted from the model results and analyzed for prediction uncertainty.

Next, the five calibration data sets are defined, which are (1) the injection pressure, (2) the temperature of water flowing in a facture at $X = 354$ m and $Y = 0$ m, i.e., midway between the injection well and production well, (3) the temperature of the rock matrix (volume average of matrix continua) at $X = 141$ m and $Y = 141$ m, (4) the production enthalpy, and (5) the tracer concentration in the production well.

Of these five data sets, only the injection pressure (Set 1) and production enthalpy (Set 4) are planned to be measured, i.e., they are expected to be existing data sets. These two data sets thus constitute the reference data available for model calibration. The other data sets are potential data sets (indicated by command `>>>> POTENTIAL`), i.e., the data-worth analysis examines how they might help reduce prediction uncertainty if they were added to the reference data sets; they are potentially available for model calibration. To put the value of potential data sets in perspective, the analysis also evaluates the increase in prediction uncertainty should one of the presumably existing data sets (i.e., injection pressure and production enthalpy) be removed from the reference set of calibration points. Note that in a full value of information analysis, the benefits of having additional data sets available for model calibration will have to be compared to the costs incurred by collecting them. In this example, pressure and enthalpy data are expected to be collected routinely as part of reservoir operation (they were thus selected as "existing" data sets), whereas collecting temperature and concentration data is likely to be costly, as it may require that observation wells be drilled and a tracer test be performed (they were thus selected as "potential" data sets).

A `>>>> TIME WINDOW` is specified for each of these data sets so that calibration would only occur for the first five years (omitting the point at 30 years). The expected average residual after calibration, σ_z, is specified for each data set using command `>>>> DEVIATION`. This scaling factor affects the results of the data-worth analysis (as it does those of a composite sensitivity analysis) as it reflects the quality of the data collected. The command `>> NO DATA` clearly indicates that a data-worth analysis can (and should) be performed *before* data are actually collected. It also means that no inversion will be performed; instead, the analysis assumes that an inversion would result in the reference parameter values, and that the final residuals will be statistically consistent with the expected errors as specified in command `>>>> DEVIATION`. Note that a data-worth analysis can (and generally will) be performed also if actual data are available, and an actual inversion is being performed.

The next two blocks (i.e., commands `>> TEMPERATURE` and `>> ENTHALPY`) define the predictions to be used as criteria for the data-worth analysis. The calibration data set that contributes most to a reduction in the overall prediction uncertainty will be assigned the highest data worth. The two model outputs of interest are (1) the temperature in the middle of a matrix block at $X = 212$ m and $Y = 71$ m, and (2) the production enthalpy. These observations should

only be analyzed after 30 years of exploitation, which is accomplished by setting the appropriate `>>>> TIME WINDOW`. Here, the command `>>>> DEVIATION` is used to set a relative weight between the two target predictions. While reaching a prediction uncertainty lower than the indicated level cannot be guaranteed (iTOUGH2 calculates the attainable prediction uncertainty assuming all calibration data sets are actually collected), the specified σ_z indicate our relative interest in a predicted model output. The higher the σ_z, the less will the data-worth analysis be valuing calibration sets that help reduce the uncertainty in the corresponding prediction. Finally, we indicate that these data sets are specified as data-worth analysis criteria by adding the keyword `PREDICTION` to command `>>>> NO DATA`.

Finally, the data-worth analysis is initiated in block `> COMPUTATION` (Figure 14). The first analysis will consider the worth of colleting entire data sets, with prediction uncertainty as the criterion. Alternative analyses are indicated, but deactivated by the #-sign in the first column.

The data-worth analysis requires running the calibration and prediction model (n+1)=8 times to calculate the Jacobian matrices **J** and $\hat{\mathbf{J}}$, respectively. In this particular example, the calibration and prediction phases are simulated using the same model; the calibration phase consists of the injection-production operation during the first five years, whereas the prediction phase consists of the subsequent 25 years of exploitation. The numerically evaluated partial derivatives of the first five data sets (see Figure 13) with respect to the seven adjustable parameters (see Figure 12) define **J**, whereas those of the last two observation blocks (temperature and enthalpy at 30 years) with respect to the same seven parameters define $\hat{\mathbf{J}}$.

The composite sensitivity measures of the scaled sensitivity matrix $\overline{S}_{ij} = J_{ij} \cdot (\sigma_{p_i} / \sigma_{z_j})$ indicate (1) that initial reservoir temperature and the logarithm of the absolute permeability of the fracture continuum in X direction are the most influential parameters, and (2) that measurements of injection pressure and temperature are the most sensitive observations. This is consistent with the intuitive understanding that pressure is very sensitive to changes in reservoir temperature and permeability, and consequently that accurate measurements of injection pressure and temperature at an observation well contain the most information about key parameters. (It should be noted that in a single-component system, pressure and temperature are largely redundant observations, as under two-phase liquid-steam conditions, the pressure is equal to the saturated vapor pressure, which is a function of temperature.)

Analyzing the contribution of each parameter to prediction uncertainty suggests that, considering all existing and potential observations, permeability and initial reservoir temperature contribute most to prediction uncertainty, confirming the parameter ranking from the composite sensitivity measures. Looking at the PCPU value (see Eq. (9)) for the temperature and enthalpy after 30 years shows that fracture spacing and thermal conductivity (while considered overall the least influential parameters) contribute significantly to the uncertainties of the two predictions that are of particular interest for this study.

```
> PARAMETER
  >> ABSOLUTE permeability
     >>> ROCKS: FRACT
         >>>> ANNOTATION: Fract. perm. X
         >>>> LOGARITHM
         >>>> INDEX     : 1
         >>>> VARIATION : 0.25
         <<<<
     >>> ROCKS: FRACT
         >>>> ANNOTATION: Fract. perm. Y
         >>>> LOGARITHM
         >>>> INDEX     : 2
         >>>> VARIATION : 0.25
         <<<<
     <<<
  >> POROSITY
     >>> MATERIAL: FRACT
         >>>> ANNOTATION: Fract. porosity
         >>>> VALUE
         >>>> VARIATION : 0.20
         <<<<
     <<<
  >> RELATIVE permeability
     >>> MATERIAL: FRACT
         >>>> ANNOTATION: Slr
         >>>> VALUE
         >>>> PARAMETER : 1
         >>>> VARIATION : 0.05
         <<<<
     <<<
  >> WET thermal CONDUCTIVITY
     >>> MATERIAL: MATRX FRACT
         >>>> ANNOTATION: Therm. cond.
         >>>> VALUE
         >>>> DEVIATION : 0.5
         <<<<
     <<<
  >> MINC
     >>> MODEL
         >>>> ANNOTATION: Fract. spacing
         >>>> VALUE
         >>>> PARAMETER : 1 2 3
         >>>> DEVIATION :  20.0
         >>>> PERTURB   :  -5.0
         <<<<
     <<<
  >> INITIAL
     >>> MATERIAL: DEFAU
         >>>> ANNOTATION: Initial temp.
         >>>> VALUE
         >>>> INDEX     : 1
         >>>> DEVIATION :   2.0
         <<<<
     <<<
  <<
```

Figure 12: Excerpt of iTOUGH2 input file, showing `> PARAMETER` block with potentially influential parameters.

```
> OBSERVATION
  >> TIMES: 61 EQUALLY spaced in YEARS (calibration points)
     0.0 5.0
  >> TIME :  1                in YEARS (prediction points)
     30.0
Presumably existing and potential calibration data sets
  >> PRESSURE
     >>> ELEMENT: _AA_1
         >>>> ANNOTATION  : Inject. pres.
         >>>> FACTOR      : 1.E5 (bar - Pa)
         >>>> NO DATA right now, but expected to be available
         >>>> DEVIATION   :  2.0 bar
         >>>> TIME WINDOW : -1.0  6.0 YEARS
         <<<<
     <<<
  >> TEMPERATURE
     >>> ELEMENT: _FA_1
         >>>> ANNOTATION  : Fract. T center
         >>>> POTENTIALLY available if confirmed valuable
         >>>> DEVIATION   :  2.0
         >>>> TIME WINDOW : -1.0  6.0 YEARS
         <<<<
     >>> ELEMENT: 2CC_1 +3 [1-1]
         >>>> ANNOTATION  : Matrix T off-center
         >>>> POTENTIALLY available if confirmed valuable
         >>>> DEVIATION   :  2.0
         >>>> TIME WINDOW : -1.0  6.0 YEARS
         <<<<
     <<<
  >> ENTHALPY
     >>> SINK: PRO_1
         >>>> ANNOTATION  : Enthalpy
         >>>> NO DATA right now, but expected to be available
         >>>> conversion FACTOR: 1000.0  (kJ/kg - J/kg)
         >>>> DEVIATION   : 100.0 kJ/kg
         >>>> TIME WINDOW :  -1.0  6.0 YEARS
         <<<<
     <<<
  >> MASS FRACTION
     >>> ELEMENT: _KA_1
         >>>> ANNOTATION  : Cl conc.
         >>>> POTENTIALLY available if confirmed valuable
         >>>> LIQUID PHASE
         >>>> COMPONENT   : 2
         >>>> DEVIATION   :  0.1
         >>>> TIME WINDOW : -1.0  6.0 YEARS
         <<<<
     <<<
Predictions for data-worth analysis
  >> TEMPERATURE
     >>> ELEMENT: 5DB_1
         >>>> ANNOTATION  : Temp. in 30 yrs
         >>>> PREDICTION for DWA
         >>>> DEVIATION   :  2.0  C
         >>>> TIME WINDOW :  29.0 31.0 YEARS
         <<<<
     <<<
  >> ENTHALPY
     >>> SINK: PRO_1
         >>>> ANNOTATION  : Enth. in 30 yrs
         >>>> conversion FACTOR: 1000.0  (kJ/kg - J/kg)
         >>>> PREDICTION for DWA
         >>>> DEVIATION   : 100.0 kJ/kg
         >>>> TIME WINDOW :  29.0 31.0 YEARS
         <<<<
     <<<
```

Figure 13: Excerpt of iTOUGH2 input file, showing `> OBSERVATION` block with potential calibration data sets.

```
> COMPUTATION

  >> OPTION
 OBS >>> DATA-WORTH ANALYSIS for data SETS   based on PREDICTION uncertainty
#OBP >>> DATA-WORTH ANALYSIS for data points based on PREDICTION uncertainty
#OAS >>> DATA-WORTH ANALYSIS for data SETS   based on PARAMETER  uncertainty
#OAP >>> DATA-WORTH ANALYSIS for data points based on PARAMETER  uncertainty
     <<<

  >> OUTPUT
     >>> YEARS
     <<<
  <<
<
```

Figure 14: Excerpt of iTOUGH2 input file, showing `> COMPUTATION` block, which initiates the data-worth analysis.

Figure **15** shows the section of the iTOUGH2 output file that is relevant to data-worth analysis. After the header "`RESIDUAL ANALYSIS`", the explanation about variable `DWi` states (i) that Metric 1, i.e., Eq. (16a), is used to evaluate data worth, (ii) that the data-worth value is calculated based on prediction uncertainty (i.e., matrix $\mathbf{C}_{\underset{\sim}{zz}}$, rather than estimation uncertainty, matrix $\mathbf{C}_{pp}$), (iii) that the overall uncertainty is measured by the trace (rather than the determinant) of the covariance matrix, and (iv) that the analysis is performed by removing entire data sets (rather than individual data points).

Next, information about each observation is given, with `DWi` in the final column, followed by a + or – sign, indicating whether the corresponding observation was added to or removed from the reference data set. Because the data-worth analysis was performed for entire data sets, each data point belonging to a data set has the same `DWi` value (which is the data set's worth divided by the number of calibration points in that set).

This analysis suggests that measuring enthalpy in the production well during the first five years of production is valuable; removing that data set would considerably increase the uncertainty with which reservoir temperature and future production enthalpy could be predicted. The high data worth of enthalpy is not surprising given that enthalpy is the same data type as the prediction of interest. Removing measurements of injection pressure is less detrimental.

Of the three potential data sets to be added to the reference data sets, measuring temperature along the likely flow path between the injector and producer as well as chloride concentrations appears to be most beneficial. This is likely because the arrival time of the cooling front and in particular tracer breakthrough data are the only data that contain useful information about porosity, which in turn affects production enthalpy.

It also appears beneficial to have some quantitative prior information about thermal conductivity, as this value cannot be well determined during the first year of production, but it has a relatively high impact on the long-term predictions; a similar argument can be made for fracture spacing. Conversely, an independent prior estimate of initial reservoir temperature is of little value, mainly because it is well constrained by the available calibration data.

The final two observations (No. 313 and 314) do not have a `DWi` value, because these are not calibration points, but the predictions whose uncertainty is used as the criterion for evaluating data worth. Note that the data-worth value would be different if the analysis were performed for individual data points, because the degree of redundancy and reference data set are different (remove `#OBP` (see Figure 14) to run an analysis using *O*ption *B* for individual data *P*oints).

The final two tables shown in Figure **16** contain the relative data worth of individual data sets and data types, respectively. The tables show that measuring production enthalpy (as is planned) is essential for calibrating the model, and that obtaining reservoir temperatures would further increases the accuracy of the predicted state after 30 years of production. This is intuitively reasonable as these two calibration data sets are of the same type as the target prediction. Conversely, measuring the injection pressure is the least valuable observation despite its high information content for estimating permeability. Recall that permeability has only a limited impact on the target variables (production enthalpy and temperature), mainly because fluid is injected and produced at prescribed mass flow rates, i.e., the amount of fluid flowing through the reservoir and related heat mining efficiency does not strongly depend on permeability. Pressure measurements are thus not of primary importance for this particular operational scenario.

In general, the absence or availability of prior information about the parameters is of limited significance given the information contained in entire data sets, and given the assumed standard deviation with which such prior information could be provided. The data-worth value for prior information may be high if the sensitivity coefficients of the calibration points (matrix **J)** with respect to this parameter are small, but the sensitivity coefficients of the model predictions (matrix $\hat{\mathbf{J}}$) are high. This indicates that the calibration data sets do not contain information about a parameter that is important for the prediction, and that it is therefore essential to obtain an independent estimate of this influential parameter.

The results of the data-worth analysis would be considerably different if the objective were to minimize the estimation uncertainty (i.e., the result of the calibration phase only), without considering the purpose for which the calibrated model will be used. Such an analysis can be conducted by removing `#OAS` or `#OAP` (see Figure 14), and by deleting the second `>> TIMES` block, which defines the (now obsolete) prediction time of 30 years (see Figure 13). In such an analysis, the relative importance of enthalpy is somewhat reduced while tracer measurements would be considered a more valuable calibration data set, because it is the only observation type that helps constrain fracture porosity.

Finally, the uncertainty of the predicted temperature and enthalpy is estimated to be 4.5ºC and 240 kJ/kg, respectively (see Figure **15**), assuming that only data from the reference set are available for model calibration. If these prediction uncertainties are not acceptable, the data-worth analysis suggests to add temperature measurements rather than to perform a tracer test.

****

RESIDUAL ANALYSIS

****

RESIDUAL : Measured - computed
C.O.F. : Relative contribution to final objective function
STD. DEV.: A posteriori standard deviation of calculated system response
Yi : Local reliability or influence; observations with Yi < 0.25 are poorly controlled
Wi : Studentized residual; if abs(Wi) > u(0.95) = 1.96 observation is potential outlier
DWi : Data-Worth Analysis: Worth of potential and/or existing calibration data [-]
Metric No. 1: 1-f(C+)/f(C) or 1-f(C)/f(C-), where f(C) = trace(Czz)
Changes in covariance matrix Czz are caused by adding (+) potential or removing (-) existing calibration data sets

#	OBSERVATION	AT TIME [a]	MEASURED	COMPUTED	RESIDUAL	WEIGHT	C.O.F [%]	STD. DEV.	Yi	Wi	DWi +/-
1	Fract. perm. X		-1.42218E+01	-1.42000E+01	-2.18487E-02	1.000E-50	0.000	5.634E-02			0.014 +
2	Fract. perm. Y		-1.42218E+01	-1.42000E+01	-2.18487E-02	1.000E-50	0.000	3.576E-01			0.363 +
3	Fract. porosity		5.00000E-01	5.00000E-01	0.00000E+00	1.000E-50	0.000	5.665E-01			0.079 +
4	Slr		3.00000E-01	3.00000E-01	0.00000E+00	1.000E-50	0.000	7.459E-02			0.065 +
5	Therm. cond.		2.10000E+00	2.10000E+00	0.00000E+00	2.000E+00	0.000	4.631E-01			0.538 -
6	Fract. spacing		1.00000E+02	1.05000E+02	-5.00000E+00	4.000E-02	0.000	1.416E+01			0.116 -
7	Initial temp.		3.00000E+02	3.00000E+02	0.00000E+00	5.000E-01	0.000	1.223E+00			0.001 -
8	Inject. pres.	0.00000E+00	0.00000E+00	8.56477E+06	-8.56477E+06	5.000E-06	0.064	1.50690E+05	0.00	0.00	0.308 -
13	Inject. pres.	8.33333E-02	0.00000E+00	9.98088E+06	-9.98088E+06	5.000E-06	0.087	9.65389E+04	0.00	0.00	0.308 -
...											
308	Inject. pres.	5.00000E+00	0.00000E+00	1.02458E+07	-1.02458E+07	5.000E-06	0.091	3.22777E+04	0.00	0.00	0.308 -
9	Fract. T center	0.00000E+00	0.00000E+00	3.00000E+02	-3.00000E+02	5.000E-01	0.782	1.22335E+00	0.00	0.00	0.498 +
14	Fract. T center	8.33333E-02	0.00000E+00	3.00036E+02	-3.00036E+02	5.000E-01	0.782	1.22751E+00	0.00	0.00	0.498 +
...											
309	Fract. T center	5.00000E+00	0.00000E+00	2.97439E+02	-2.97439E+02	5.000E-01	0.769	2.25178E+00	0.00	0.00	0.498 +
10	Matrix T off-ce	0.00000E+00	0.00000E+00	3.00000E+02	-3.00000E+02	5.000E-01	0.782	1.22335E+00	0.00	0.00	0.325 +
15	Matrix T off-ce	8.33333E-02	0.00000E+00	3.00011E+02	-3.00011E+02	5.000E-01	0.782	1.22465E+00	0.00	0.00	0.325 +
...											
310	Matrix T off-ce	5.00000E+00	0.00000E+00	2.91076E+02	-2.91076E+02	5.000E-01	0.736	2.44647E+00	0.00	0.00	0.325 +
11	Enthalpy	0.00000E+00	0.00000E+00	1.34488E+06	-1.34488E+06	1.000E-05	0.006	6.90726E+03	0.00	0.00	0.884 -
16	Enthalpy	8.33333E-02	0.00000E+00	1.50075E+06	-1.50075E+06	1.000E-05	0.008	8.34776E+04	0.00	0.00	0.884 -
...											
311	Enthalpy	5.00000E+00	0.00000E+00	1.38545E+06	-1.38545E+06	1.000E-05	0.007	1.21766E+04	0.00	0.00	0.884 -
12	Cl conc.	0.00000E+00	0.00000E+00	7.68313E-23	-7.68313E-23	1.000E+01	0.000	4.28222E-21	0.00	0.00	0.558 +
17	Cl conc.	8.33333E-02	0.00000E+00	1.30151E-07	-1.30151E-07	1.000E+01	0.000	1.22405E-06	0.00	0.00	0.558 +
...											
312	Cl conc.	5.00000E+00	0.00000E+00	9.57731E-01	-9.57731E-01	1.000E+01	0.003	1.92311E-01	0.00	0.00	0.558 +
313	Temp. in 30 yrs	3.00000E+01	0.00000E+00	2.45271E+02	-2.45271E+02	5.000E-51	0.000	4.52151E+00	0.00	0.00	predict
314	Enth. in 30 yrs	3.00000E+01	0.00000E+00	1.33094E+06	-1.33094E+06	1.000E-55	0.000	2.42406E+05	0.00	0.00	predict

```
MEAN      : Mean of residuals
MEDIAN    : Median of residuals
STD. DEV. : Root mean squared deviation of residuals from mean
AVE. DEV. : Mean absolute deviation of residuals from mean
SKEWNESS  : Degree of asymmetry of residuals around mean
KURTOSIS  : Relative peakedness of distribution
BCT       : Box-Cox transformed data
M/S       : Ratio of mean and standard deviation
DWA       : Relative Data Worth [%]
COF       : Contribution to final objective function [%]
```

DATASET		DATAPOINTS	MEAN	MEDIAN	STD. DEV.	AVE. DEV.	SKEWNESS	KURTOSIS	M/S	DWA	COF
PRIOR INFORMATION		3								0.74	0.00
Inject. pres.	[Pa]	61	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0.000	0.000	0.00	11.86 -	0.00
Fract. T center	[C]	61	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0.000	0.000	0.00	19.21 +	0.00
Matrix T off-ce	[C]	61	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0.000	0.000	0.00	12.55 +	0.00
Enthalpy	[W]	61	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0.000	0.000	0.00	34.11 -	0.00
Cl conc.	[kg/kg]	61	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0.000	0.000	0.00	21.52 +	0.00
ALL RESIDUALS	WEIGHTED	314	-7.152E+01	-5.104E+01	6.374E+01	5.560E+01	-0.288	-1.750	1.12	100.00	100.00

DATATYPE		DATAPOINTS	MEAN	MEDIAN	STD. DEV.	AVE. DEV.	SKEWNESS	KURTOSIS	M/S	DWA	COF
PRIOR INFORMATION		3								0.74	0.00
PRESSURE	[Pa]	61	-1.017E+07	-1.022E+07	2.163E+05	6.433E+04	6.752	47.214	47.04	11.86	0.00
MASS FRACTION	[kg/kg]	61	-5.507E-01	-6.683E-01	3.571E-01	3.108E-01	0.401	-1.476	1.54	21.52	0.00
TEMPERATURE	[C]	123	-2.974E+02	-2.987E+02	5.321E+00	2.173E+00	7.712	71.979	55.89	31.76	0.00
ENTHALPY	[W]	62	-1.396E+06	-1.394E+06	2.000E+04	1.104E+04	-1.670	11.980	69.81	34.11	0.00

Figure 16: Excerpt of iTOUGH2 output file, showing summary statistics of data-worth analysis.

5. CONCLUDING REMARKS

In addition to its inverse modeling capabilities, iTOUGH2 includes options for performing sensitivity and data-worth analyses, which improve system understanding and help design experiments and monitoring systems, so the data collected contain the information needed t calibrate the model.

A conventional, local sensitivity analysis can be performed as a pre-calibration task. Howev sensitivity coefficients and composite sensitivity measures are also evaluated at the end of a inversion that uses a derivative-based minimization algorithm (specifically the Levenberg-Marquardt method).

Two global sensitivity analysis methods are implemented. The Morris one-at-a-time metho be considered an approach of great practical value, as it identifies influential parameters as as the impact of nonlinearity and/or interaction effects with a limited number of simulation In addition, by having confidence intervals of the indices, it is possible to compare the importance of parameters with a limited number of simulation runs.

The Sobol' method yields a statistically more quantitative global sensitivity index in the co of UQ, at the expense of having to evaluate significantly more simulation runs. Similarity o total sensitivity index and the Morris mean |EE| suggests that mean |EE| would be sufficient could be used instead of the total sensitivity index, requiring fewer simulations. An advanta the Sobol' method is that having two indices – Sobol' index and total sensitivity index – all us to identify the presence and magnitude of interactions effects.

Finally, iTOUGH2 can be used to perform a data-worth analysis, which is based on local sensitivity analyses of a calibration and prediction model. It provides insights into the relati value of data points or entire data sets for the purpose of model calibration and related mod predictions.

Sensitivity and data-worth analysis are essential tools for the development of more reliable more defensible numerical models. They also help improve the design and analysis of labor experiments, field tests, and monitoring systems, making data collection more cost effectiv

Acknowledgments

This work was supported, in part, by the TOUGH Development Grant and the U.S. Dept. o Energy under Contract No. DE-AC02-05CH11231.

References

Cacuci, D.G., 2003, *Sensitivity and Uncertainty Analysis, Vol. 1: Theory*, Boca Raton, FL: Chapman and Hall/CRC Press.

Doughty, C., *User's Guide for Hysteretic Capillary Pressure and Relative Permeability Functions in TOUGH2*, Report LBNL-6533E, Lawrence Berkeley National Laboratory, Berkeley, Calif., March 2013.

Falta, R. W., K. Pruess, S. Finsterle, and A. Battistelli, *T2VOC User's Guide*, Report LBL-36400, Lawrence Berkeley Laboratory, Berkeley, Calif., March 1995.

Finsterle, S., Practical notes on local data-worth analysis, *Water Resour. Res.*, *51*(12), 9904–9924, doi: 10.1002/2015WR017445, 2015.

Finsterle, S., and P. Persoff, Determining permeability of tight rock samples using inverse modeling, *Water Resour. Res.*, *33* (8), 1803–1811, 1997.

Finsterle, S., and M.B. Kowalsky, *iTOUGH2-GSLIB User's Guide*, Report LBNL/PUB-3191, Lawrence Berkeley National Laboratory, Berkeley, Calif., June 2007.

Finsterle, S., *iTOUGH2 User's Guide*, Report LBNL-40040, Lawrence Berkeley National Laboratory, Berkeley, Calif., 2010.

Finsterle, S., *iTOUGH2 Universal Optimization Using the PEST Protocol — User's Guide*, Report LBNL-3698E, Lawrence Berkeley National Laboratory, Berkeley, Calif., July 2010.

Finsterle, S., and Y. Zhang, Solving iTOUGH2 simulation and optimization problems using the PEST protocol, *Environmental Modelling and Software*, *26*, 959–968, doi:10.1016/j.envsoft.2011.02.008, 2011.

Fisher, R. A., On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32, 1921.

Foglia, L., M.C. Hill, S.W. Mehl, and P. Burlando, Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, *Water Resour. Res.*, *45*, W06427, doi:10.1029/2008WR007255, 2009.

Glen, G. and K. Issacs, Estimating Sobol sensitivity indices using correlations, *Environ. Modell. Softw.*, 37, 157-166, 10.1016/j.envsoft.2012.03.014, 2012.

Morris, M.D., Factorial Sampling Plans for Preliminary Computational Experiments, *Technometrics*, 33(2):161-174, 1991.

Oldenburg, C.M., G.J. Moridis, N. Spycher, and K. Pruess, *EOS7C Version 1.0: TOUGH2 Module for Carbon Dioxide or Nitrogen in Natural Gas (Methane) Reservoirs*, Report LBNL-56589, Lawrence Berkeley National Laboratory, Berkeley, Calif., 2004.

Pruess, K. and A. Battistelli, *TMVOC, A Numerical Simulator for Three-Phase Non-isothermal Flows of Multicomponent Hydrocarbon Mixtures in Saturated-Unsaturated Heterogeneous Media*, Report LBNL-49375, Lawrence Berkeley National Laboratory, Berkeley, Calif., April 2002.

Pruess, K., *ECO2N: A TOUGH2 Fluid Property Module for Mixtures of Water, NaCl, and CO_2*, Report LBNL-57952, Lawrence Berkeley National Laboratory, Berkeley, Calif., 2005.

Pruess, K., *ECO2M: A TOUGH2 Fluid Property Module for Mixtures of Water, NaCl, and CO_2, Including Super- and Sub-Critical Conditions, and Phase Change Between Liquid and Gaseous CO_2*, Report LBNL-4590E, Lawrence Berkeley National Laboratory, Berkeley, Calif., 2011.

Pruess, K., C. Oldenburg, and G. Moridis, *TOUGH2 User's Guide, Version 2.1*, Report LBNL-43134, Lawrence Berkeley National Laboratory, Berkeley, Calif., 2012.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, West Sussex, England, 2008.

Sobol, I.M., Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, 55(1-3), 271-280, 2001.

Wainwright, H. M., S. Finsterle, Q. Zhou, Q., and J.T. Birkholzer, Modeling the performance of large-scale CO_2 storage systems: A comparison of different sensitivity analysis methods. *International Journal of Greenhouse Gas Control*, 17, 189-205, 2013.

Wainwright, H. M., S. Finsterle, Y. Jung, Q. Zhou, and J.T. Birkholzer, Making sense of global sensitivity analyses. *Computers & Geosciences*, 65, 84-94, 2014.