



HAL
open science

Screening method using the derivative-based global sensitivity indices with application to reservoir simulator

Samir Touzani, Daniel Busby

► To cite this version:

Samir Touzani, Daniel Busby. Screening method using the derivative-based global sensitivity indices with application to reservoir simulator. 2013. hal-00789807

HAL Id: hal-00789807

<https://hal.science/hal-00789807>

Preprint submitted on 18 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Screening method using the derivative-based global sensitivity indices with application to reservoir simulator

Samir Touzani^{1*} and Daniel Busby¹

¹ IFP Energies nouvelles, 92852 Rueil-Malmaison, France

e-mail: samirtouzani.phd@gmail.com

*Corresponding author

Résumé — Méthode de criblage basée sur les indices de sensibilité DGSM : Application au simulateur de réservoir — Les simulateurs de réservoir peuvent impliquer un grand nombre de paramètres d'entrées. Cependant, il est possible de réduire cette complexité en ne se focalisant que sur les paramètres influents. La détection de ces paramètres peut être effectuée grâce à l'analyse de sensibilité. Néanmoins, pour des problèmes de grande dimension, les méthodes classiques d'estimation des indices de sensibilité nécessitent un très grand nombre d'évaluation du simulateur de réservoir, ce qui est très coûteux en temps de calcul. Récemment, pour pallier au problème de dimensionnalité, de nouveaux indices de sensibilité ont été définis. Ces indices dit DGSM (pour derivative-based global sensitivity measures) sont basés sur la moyenne des dérivées partielles et ont un lien avec les indices totaux qui sont des indices de sensibilité basés sur la décomposition de la variance. Cet article introduit une version révisée des indices DGSM, ce qui permet de proposer une méthode de criblage efficace et moins coûteuse en temps de calcul que les méthodes classiques. Afin d'apprécier l'efficacité de la méthode proposée, des résultats sur des fonctions analytiques ainsi que sur un cas synthétique de réservoir sont présentées.

Abstract — Screening method using the derivative-based global sensitivity indices with application to reservoir simulator — Reservoir simulator can involve a large number of uncertain input parameters. Sensitivity analysis can help reservoir engineers focusing on the inputs whose uncertainties have an impact on the model output, which allows reducing the complexity of the model. There are several ways to define the sensitivity indices. One of the quantitative definition is the variance based sensitivity indices which can quantify the amount of output uncertainty due to the uncertainty of inputs. However, the classical methods to estimate such sensitivity indices in a high dimensional problem can require a huge number of reservoir model evaluation. Recently, new sensitivity indices based on averaging local derivatives of the model output over the inputs domain have been introduced. These so-called derivative-based global sensitivity measures (DGSM) have been proposed to overcome the problem of dimensionality and are linked to total effect indices which are variance based sensitivity indices. In this work we propose a screening method based on a revised DGSM indices which increase the interpretability in some complex case and have a lower computational cost, as demonstrated by numerical test cases and by an application on a synthetic reservoir test model.

1 INTRODUCTION

Reservoir simulators are complex computer codes that model the physical laws governing the recovery process, and which are mainly modelled by mathematical equations for the three phases flow (oil, gas and water) through porous media. These simulators involve a large number of input parameters. The information gathered on such inputs comes from direct measurements, which are clearly very limited and are marred by considerable uncertainty. Thus, it is important to detect influential inputs, whose uncertainties have an impact on the model output. Once identified, one can reduce the complexity of the model by fixing the non-influential inputs on default values (defined by experts) and focus the attention on the influential inputs.

Sensitivity analysis (SA) is the study of how the variation (uncertainty) in the output of the computer model can be apportioned, qualitatively or quantitatively, to different sources of variation in the input of the model. Put in another way, it is a technique for systematically changing parameters in a model to determine the effects of such changes on the output. The local SA methods refer to the study of the sensitivity at a fixed point in the input domain, typically the simple derivative $\partial Y/\partial x^{(i)}$ of the output Y with respect to a given input $X^{(i)}$ taken at some fixed point \mathbf{x}_0 in the input domain. The global SA (GSA) methods [17–19, 26] refers to the sampling-based methods in which the model is evaluated for combinations of values sampled from the distribution (assumed known) of the inputs. Once the sample is generated, several strategies (including simple input-output scatterplots) can be used to derive global sensitivity measures for the factors.

The variance-based methods are nonlinear with respect to the input parameters, and are based on analysis of variance (ANOVA) decomposition which is the decomposition of the total variance V of output into terms due to individual factors plus terms due to interaction among inputs. Most variance-based methods are quantitative, and in this work we will focus on this class of methods, and more specifically on Sobol's indices.

One of the main issues with variance based methods is computational time. Indeed, a reservoir simulator is often very costly in terms of computational time. Furthermore such variance based methods require generally several thousands simulations that are usually not affordable in common applications. In order to perform SA with a limited number of runs, metamodel methods can then be used. In the latter the simulator input/output relation is approximated using different statistical regression techniques starting from an initial set of carefully chosen training runs. Then, if a reasonably good approximation is obtained, the estimated metamodel is used instead of the complex simulator to compute the sensitivity indices. Metamodel methods have known a quick development in the last decade and different approaches have been suggested in many different scientific

disciplines [2, 3, 10, 16, 20, 27–30]. However, despite significant advances in the area, construction of a sufficiently accurate approximation for high dimensional computer code using relatively low number of model evaluation is problematic.

Screening methods aim at reducing the input dimensionality by identifying the non-influential inputs with a low computational cost in terms of model evaluation. Screening design proposed by Morris [11] is adapted for high dimensional expensive computer model. This method is an one factor at time (OAT) technique that vary one input parameter at a time and measures the impact on the output. Indeed, the method is based on calculating a sensitivity index called elementary effect, which provide a good compromise between accuracy and efficiency. However, even if this method is computationally cheaper than other SA methods, it involves hundreds or thousands (depending on the number of inputs and the complexity of the model) of model evaluation, which is still computationally intensive with realistic reservoir simulators for which each simulation requires several hours or days.

Recently, Sobol and Kucherenko [24 25] have proposed new sensitivity indices based on averaging local derivatives of the model output over the inputs domain. It was shown that the so-called Derivative-based Global Sensitivity Measures (DGSM) can be easily estimated and much faster than the global sensitivity indices. In addition, different methods exist to efficiently compute the derivatives of reservoir simulators, in this case DGSM represent a valid alternative to the Morris method for screening the input parameters.

In this work we propose a revised derivative-based sensitivity index that allows a better convergence of the estimation and increases the interpretability in some complex cases. We propose a screening method based on the defined indices. We then employ the method to perform a screening of a high-dimensional numerical test cases and a reservoir simulator.

2 GLOBAL SENSITIVITY ANALYSIS AND MORRIS DESIGN

First let consider a mathematical model for a reservoir simulator

$$Y = f(\mathbf{X}) \quad (1)$$

where Y is a scalar output of the computer code, $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$ a unit d -dimensional input vector ($\mathbf{X} \in [0, 1]^d$) which represents the uncertain parameters/factors of the simulator and $f : [0, 1]^d \rightarrow \mathbb{R}$ is a function that models the relationship between the input factors and the output of the computer code.

2.1 Global sensitivity analysis

The main idea from Sobol approach [26] is to decompose the response $Y = f(\mathbf{X})$ into summands of different dimensions via analysis of variance decomposition (ANOVA) defined as

$$f(X) = f_0 + \sum_{j=1}^d f_j(X^{(j)}) + \sum_{j<l} f_{jl}(X^{(j)}, X^{(l)}) + \dots + f_{1,\dots,d}(X^{(1)}, \dots, X^{(d)}) \quad (2)$$

where f_0 is a constant, f_j 's are univariate functions representing the main effects, f_{jl} 's are bivariate functions representing the two way interactions, and so on.

The integrals of every summand of this decomposition over any of its own variables is assumed to be equal to zero, i.e.

$$\int_0^1 f_{j_1,\dots,j_s}(X^{(j_1)}, \dots, X^{(j_s)}) dX^{(j_k)} = 0 \quad (3)$$

where $1 \leq j_1 < \dots < j_s \leq d$, $s = 1, \dots, d$ and $1 \leq k \leq s$. It follows from this property that all the summands in (2) are orthogonal, i.e, if $(i_1, \dots, i_s) \neq (j_1, \dots, j_l)$, then

$$\int_{\Omega^d} f_{i_1,\dots,i_s} f_{j_1,\dots,j_l} d\mathbf{X} = 0 \quad (4)$$

Using the orthogonality, Sobol [26] showed that such decomposition of $f(\mathbf{X})$ is unique and that all the terms in (2) can be evaluated via multidimensional integrals

$$f_0 = E(Y) \quad (5)$$

$$f_j(X^{(j)}) = E(Y|X^{(j)}) - E(Y) \quad (6)$$

$$f_{j,l}(X^{(j)}, X^{(l)}) = E(Y|X^{(j)}, X^{(l)}) - f_j - f_l - E(Y) \quad (7)$$

where $E(Y)$ and $E(Y|X^{(j)})$ are respectively the expectation and the conditional expectation of the output Y . Analogous formulae can be obtained for the higher-order terms. If all the input factors are mutually independent, the ANOVA decomposition is valid for any distribution function of the $X^{(i)}$'s and using this fact, squaring and integrating (2) over $[0, 1]^d$, and by (4), we obtain

$$V = \sum_{j=1}^d V_j + \sum_{1 \leq j < l \leq d} V_{jl} + \dots + V_{1,2,\dots,d} \quad (8)$$

where $V_j = V[E(Y|X^{(j)})]$ is the variance of the conditional expectation that measures the main effect of X_j on Y and $V_{jl} = V[E(Y|X^{(j)}, X^{(l)})] - V_j - V_l$ measures the joint effect of the pair $(X^{(j)}, X^{(l)})$ on Y . The total variance V of Y is defined to be

$$V = E(Y^2) - f_0^2 \quad (9)$$

Variance-based sensitivity indices, also called Sobol indices, are then defined by

$$S_{j_1,\dots,j_s} = \frac{V_{j_1,\dots,j_s}}{V} \quad (10)$$

where $1 \leq j_1 < \dots < j_s \leq d$ and $s = 1, \dots, d$. Thus, $S_j = V_j/V$ is called the first order sensitivity index (or the main effect) for factor $X^{(j)}$, which measures the main effect of $X^{(j)}$ on the output Y , the second order index $S_{jl} = V_{jl}/V$, for $j \neq l$, is called the second order sensitivity index expresses the sensitivity of the model to the interaction between variables $X^{(i)}$ and $X^{(j)}$ on Y and so on for higher orders effects. The decomposition in (8) has the useful property that all sensitivity indices sum up to one.

$$\sum_{j=1}^p S_j + \sum_{1 \leq j < l \leq p} S_{jl} + \dots + S_{1,2,\dots,p} = 1 \quad (11)$$

The total sensitivity index (or total effect) of a given factor is defined as the sum of all the sensitivity indices involving the factor in question.

$$S_{T_i} = \sum_{\#i} \frac{V_l}{V} = \frac{V_{T_i}}{V} \quad (12)$$

where $\#i$ represents all the S_{j_1,\dots,j_s} terms that include the index i . Total effect index of an input $X^{(i)}$ measures the part of output variance explained by all the effects in which it plays a role. Note however that the sum of all S_{T_i} is higher than one because interaction terms are counted several times. It is also important to note that total effect indices can be computed by a single multidimensional integration and do not require computing all high order indices (see Sobol [26]). Then comparing the total effect indices provides information about influential parameters. Indeed, one can suppose that the input is non-influential if its total effect S_{T_i} is less than 0.01.

GSA enables to explain the variability of the output response as a function of the input parameters through the definition of total and partial sensitivity indices. The computation of these indices involves the computation of several multidimensional integrals that are estimated by Monte Carlo method and thus requires huge random samples. For this reason GSA techniques are prohibitive if used directly using the computer code (fluid flow simulator for example).

2.2 Morris's screening method

The screening method introduced by Morris [11] is based on a one factor at time (OAT) experimental design. The points of the Morris design are sampled from a d -dimensional p -level grid, as the range of each input $X^{(i)}$ is divided into p equal level. The impact of varying one input at a time is evaluated by the so-called elementary effect that, for $i = 1, \dots, d$, is defined as

$$d_i^r(\mathbf{X}^r) = \frac{f(\mathbf{X}_i^r) - f(\mathbf{X}^r)}{\Delta} \quad (13)$$

where Δ is a multiple of $1/(p-1)$ with p the number of levels, \mathbf{X}^r is a randomly chosen point in $[0, 1]^d$ such that

$X^{(i)} + \Delta$ is still in $[0, 1]$ and $\mathbf{X}_i^r = X^{(1)}, \dots, X^{(i-1)}, X^{(i)} + \Delta, X^{(i+1)}, \dots, X^{(d)}$. The group of points composed of \mathbf{X}^r and \mathbf{X}_i^r 's are called trajectories. Thus, the Morris design is structured in R random trajectories composed of $R(d+1)$ points. The sensitivity measures proposed by Morris [11] are defined as a statistics of the elementary effect. The first one is the mean $\hat{\mu}_i$

$$\hat{\mu}_i = \frac{1}{R} \sum_{r=1}^R d_i^r \quad (14)$$

which is a measure of the i th input importance. The second statistic is the standard deviation of the elementary effect $\hat{\sigma}_i$

$$\hat{\sigma}_i = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (d_i^r - \hat{\mu}_i)^2} \quad (15)$$

which is a measure of the non-linearity and the involved interactions of the i th input. However, $\hat{\sigma}_i$ does not allow to distinguish between non-linearities and interactions.

Noting that when the model is non-monotonic the elementary effects of opposite signs cancel each other, Campolongo et al. [4] proposed the sensitivity measure $\hat{\mu}_i^*$, which is a revised version of $\hat{\mu}_i$

$$\hat{\mu}_i^* = \frac{1}{R} \sum_{r=1}^R |d_i^r| \quad (16)$$

To identify the non-influential inputs, the sensitivity measures $\hat{\mu}_i^*$ and $\hat{\sigma}_i$ are simultaneously considered.

3 DERIVATIVE-BASED SENSITIVITY ANALYSIS

3.1 Derivative-based global sensitivity measures

First introduced by Sobol and Gresham [21] and then studied in Kucherenko et al. [9], Sobol and Kucherenko [24-25], DGSM are a new sensitivity indices based on averaging local derivatives of the model output over the inputs domain. Assume that $\partial f(\mathbf{X})/\partial x^{(i)}$, for $i = 1, \dots, d$, are square-differentiable. The DGSM indices are defined as

$$\nu_i = \mathbb{E} \left[\left(\frac{\partial f(\mathbf{X})}{\partial x^{(i)}} \right)^2 \right] = \int \left(\frac{\partial f(\mathbf{X})}{\partial x^{(i)}} \right)^2 d\mathbf{x} \quad (17)$$

Thus, calculation of DGSM indices is based on the evaluation of integrals, which is easily performed using a classical Monte Carlo (MC), quasi-Monte Carlo (QMC) or latin hyper-cube (LHS) sampling. The empirical estimator of ν_i is given by

$$\hat{\nu}_i = \frac{1}{n} \sum_{j=1}^n \left(\frac{\partial f(\mathbf{X}_j)}{\partial x^{(i)}} \right)^2 \quad (18)$$

3.1.1 Link between DGSM and GSA

Recently, Sobol and Kucherenko [24] have established the link between the DGSM index ν_i and the total effect index S_{T_i} for input variables following uniform and normal distributions. Here, we assume that $X^{(i)} \sim \mathbb{U}[a_i, b_i]$, for $i = 1, \dots, d$, the link between V_{T_i} and ν_i is defined by the following inequality

$$V_{T_i} \leq \frac{(b_i - a_i)^2}{\pi^2} \nu_i = \nu_i^* \quad (19)$$

Thereby the total effect indices have the following upper bound

$$S_{T_i} \leq \frac{\nu_i^*}{V} = \Upsilon_i \quad (20)$$

where V is the total variance of the model. If $\Upsilon_i \simeq 0$, then $X^{(i)}$ can be considered as non-influent input, which make Υ_i a good candidate for a screening procedure.

3.1.2 Link between DGSM and Morris method

Kucherenko et al. [9] have introduced two derivative-based sensitivity indices that are very similar to the Morris indices μ_i and σ_i . These indices are defined, for $i = 1, \dots, d$, as

$$\bar{M}_i = \int \frac{\partial f(\mathbf{X})}{\partial x^{(i)}} d\mathbf{x} \quad (21)$$

and

$$\bar{\Sigma}_i^2 = \int \left(\frac{\partial f(\mathbf{X})}{\partial x^{(i)}} \right)^2 d\mathbf{x} - \bar{M}_i^2 \quad (22)$$

where \bar{M}_i is equivalent to μ_i and $\bar{\Sigma}_i^2$ equivalent to σ_i^2 . In addition, these indices are more accurate than Morris's indices, which cannot correctly consider effects with characteristic dimensions less than Δ . Indeed, in (13) the elementary effects d_i^r are calculated as finite differences with the increment δ , which has the same order of magnitude as the uncertainty range of inputs. In contrast with the derivative-based indices \bar{M}_i and $\bar{\Sigma}_i^2$, where the elementary effects are substitute by the local derivatives.

We can also note that the DGSM indices ν_i can be defined as

$$\nu_i = \int \left(\frac{\partial f(\mathbf{X})}{\partial x^{(i)}} \right)^2 d\mathbf{x} = \bar{M}_i^2 + \bar{\Sigma}_i^2 \quad (23)$$

3.2 Refining the DGSM index Υ

In the previous section we have shown that the sensitivity index Υ is an upper bound of the total effect index. However, for some complex model Υ can be much larger than the corresponding total effect index. In this case, it is difficult to decide which inputs are influential and which are not. In addition, Υ estimation involves the variance of the model output V . Empirical results (see next section) show that estimation of V requires more model evaluation than estimation

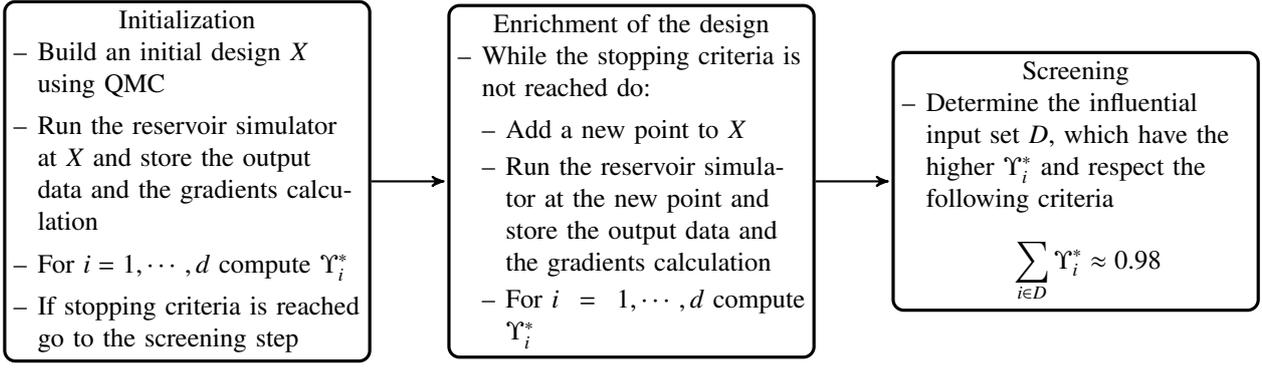


Figure 1: Schematic representation of the screening method

of v .

We propose here a different version of Υ that we call Υ^* , which is defined, for $i = 1, \dots, d$, as

$$\Upsilon_i^* = \frac{v_i^*}{\sum_{j=1}^d v_j^*} \quad (24)$$

This index is a normalized upper bound of V_{T_i} . Indeed, the link between Υ_i^* and V_{T_i} is defined as

$$\frac{V_{T_i}}{\sum_{j=1}^d v_j^*} \leq \Upsilon_i^* \quad (25)$$

In addition, Υ^* has the following useful properties

$$0 \leq \Upsilon_i^* \leq 1 \quad (26)$$

and

$$\sum_{i=1}^d \Upsilon_i^* = 1 \quad (27)$$

The drawback of Υ^* is the loss of the link with the total effect indices. Nevertheless, the use of Υ^* offers a stronger measure to define the non-influential inputs, as we will see in the next sections.

4 DGSM-BASED SCREENING METHOD

As shown in section 3, estimating Υ^* for each input allows one to detect the influential inputs in the model. Indeed, one can state that if $\hat{\Upsilon}_i^* \leq 0.01$, the corresponding input $X^{(i)}$ can be defined as non-influential one. However, such criteria may be too strong in the case of very high dimensional model (typically more than 100 input parameters), the fact is that the sum of effects due to inputs with small sensitivity indices may be significant on the output model variance. Because of that, and by using the property (27) of the Υ^* measures, it is more robust to state that the influential inputs

are the set D of d^* inputs whose the Υ_i^* are the most high and which respect the following criteria

$$\sum_{i \in D} \Upsilon_i^* \approx 0.98 \quad (28)$$

Given that the reservoir simulator evaluation may be computationally demanding, it is important to use a sequential strategy to build the design of point by reusing at each step the already evaluated points. The use of the so-called QMC Sobol sequence [22, 23] is an efficient way to build a sequential design. Our choice is motivated by two main properties of the QMC Sobol sequences. First, this technique is based on the generation of deterministic quasi-random sequences with a good space-filling property of the unit hypercube, in other words the input domain are well covered for fairly small sets. Second, the points of the Sobol sequence are independent. That is, by enriching the design sequentially, one keeps the space filling properties of the Sobol sequence. Since a sequential method computes successive estimation of the Υ_i^* indices, a practical test is needed to determine when to stop the iteration. In this work we propose to use the following error criteria

$$err_l = \frac{1/10 \sum_{k=1}^{10} \|\Upsilon_{(l)}^* - \Upsilon_{(l-k)}^*\|}{\|\Upsilon_{(l)}^*\|} \quad (29)$$

where vectors $\Upsilon_{(l)}^* = (\Upsilon_1^*, \dots, \Upsilon_d^*)$ are the l th estimation of Υ_i^* indices and $\|\cdot\|$ is the Euclidean norm. Thus, we define the stopping criteria as $err \leq 0.05$. A schematic representation of the entire screening method is shown in Figure 1. Note that one can use the same stopping criteria (29) to estimate Υ_i indices.

5 NUMERICAL TESTS

In this section, two numerical test cases are used to demonstrate the estimation performance of the DGSM index Υ^*

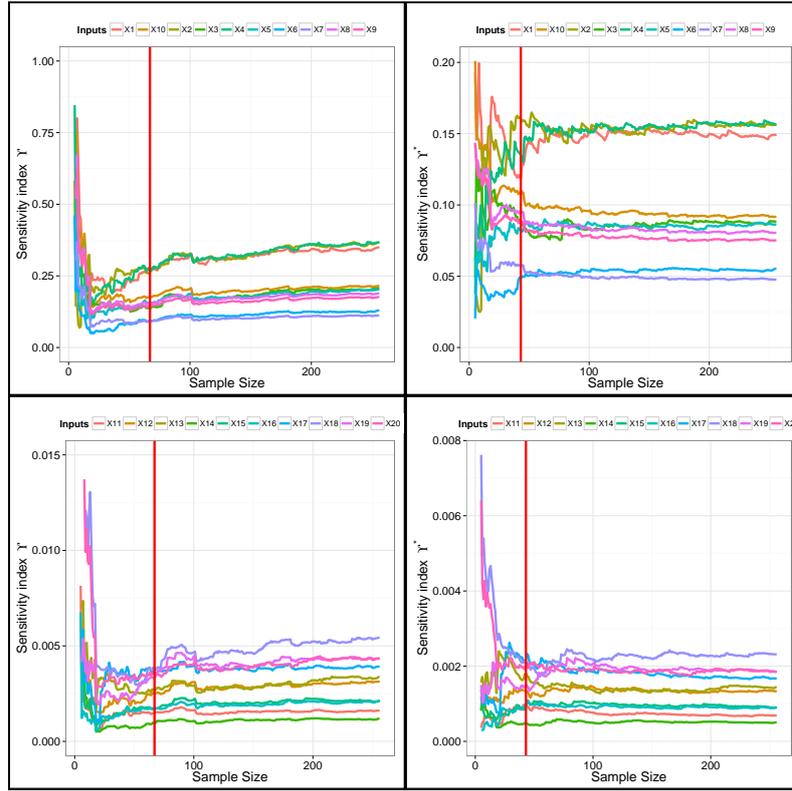


Figure 2: Convergence of the Υ_i and Υ_i^* indices estimates versus the sample size for the Morris function

and the accuracy of the proposed screening method to detect the influential inputs. Adopting the QMC sampling method, each input parameter $X^{(i)}$ is uniformly distributed in $[0, 1]$.

5.1 The test case of Morris

The test function proposed by Morris [11] contains 20 input parameters and defined as follow

$$y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{i,j} w_i w_j + \sum_{i<j<l}^{20} \beta_{i,j,l} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{i,j,l,s} w_i w_j w_l w_s$$

where $w_i = 2 \times (X^{(i)} - 1/2)$ except for $i=3,5$, and 7 , where $w_i = 2 \times (1.1X^{(i)} / (X^{(i)} + 1) - 1/2)$. Coefficients with relatively large values are

$$\begin{cases} \beta_i = 20 & \text{for } i = 1, \dots, 10 \\ \beta_{i,j} = -15 & \text{for } i, j = 1, \dots, 6 \\ \beta_{i,j,l} = -10 & \text{for } i, j, l = 1, \dots, 5 \\ \beta_{i,j,l,s} = 5 & \text{for } i, j, l, s = 1, \dots, 4 \end{cases}$$

the remaining β_i and $\beta_{i,j}$ are independently generated from a standard normal distribution. The remaining $\beta_{i,j,l}$ and $\beta_{i,j,l,s}$

are sets to zero.

In Figure 2 one can see the results of computing Υ_i and Υ_i^* sequentially with a QMC design ranging from $n = 5$ to $n = 256$. In addition, the sample size when the stopping criteria of the proposed screening method is reached is represented by the red vertical line. Thereby, we can notice that Υ_i^* converge faster than Υ_i . Indeed, the stopping criteria is reached at $n = 43$ for Υ_i^* and at $n = 67$ for Υ_i estimations. In Table 1, for the inputs selected by the proposed screening method, the values of the total effect indices (obtained by the so-called extended-FAST method [18] using a sample of size $N = 3.5 \times 10^4$) as well as the values of Υ_i and Υ_i^* (computed at the stopping criteria sample size $n = 67$ and $n = 43$) are reported. Notice that the values of the indices $S_{T_{11}}, \dots, S_{T_{20}}$ are smaller than 0.005 and then the corresponding inputs are considered to be non-influential.

It can be seen that for this test case both indices (Υ_i and Υ_i^*) are able to identify the influential inputs correctly at the stopping criteria (29). Furthermore, even if at $n = 67$ the indices Υ_i are under estimated they are almost for all inputs greater but close to the total effect indices. To conclude on this numerical test, we can say that for the Morris function the developed screening method based on Υ_i^* is efficient and not consuming in terms of model evaluations. However, estimating Υ_i indices provides more information since it is

very close as upper bound to the total effect indices but the drawback is the additional model evaluation cost.

Input	Total effect	$\Upsilon_i(n=67)$	$\Upsilon_i^*(n=43)$
$X^{(1)}$	0.240	0.249	0.114
$X^{(2)}$	0.241	0.244	0.164
$X^{(3)}$	0.095	0.127	0.088
$X^{(4)}$	0.245	0.252	0.137
$X^{(5)}$	0.098	0.137	0.083
$X^{(6)}$	0.082	0.081	0.048
$X^{(7)}$	0.050	0.078	0.053
$X^{(8)}$	0.105	0.133	0.091
$X^{(9)}$	0.099	0.123	0.084
$X^{(10)}$	0.106	0.173	0.118

TABLE 1: Estimated sensitivity indices for the Morris function

5.2 The g-Sobol function

Consider the g-Sobol function, which is strongly nonlinear and is described by a non-monotonic relationship. Because of its complexity and the availability of analytical sensitivity indices, this function is a well-known test case in the studies of GSA. Let define the g-Sobol function for 200 input parameters as follow

$$g_{\text{Sobol}}(X^{(1)}, \dots, X^{(8)}) = \prod_{k=1}^8 g_k(X^{(k)})$$

with

$$g_k(X^{(k)}) = \frac{|4X^{(k)} - 2| + a_k}{1 + a_k}$$

where $\{a_1, \dots, a_{20}\} = \{0, 0.25, 0.5, 0.75, 1, 2, \dots, 16\}$ and $\{a_{21}, \dots, a_{200}\} = 99$. The contribution of each input $X^{(k)}$ to the variability of the model output is represented by the weighting coefficient a_k . The lower this coefficient a_k , the more significant the variable $X^{(k)}$.

The analytical values of Sobol's indices are given by

$$V_j = \frac{1}{3(1 + a_j)^2}, \quad V = \prod_{k=1}^d (V_k + 1) - 1,$$

$$S_{j_1, \dots, j_s} = \frac{1}{V} \prod_{k=1}^s V_k$$

where $1 \leq j_1 < \dots < j_s \leq d$ and $s = 1, \dots, d$. The analytical values of the total effect indices are shown in table (2). Figure 3 shows the sequential estimation of Υ_i^* and Υ_i indices with a QMC design ranging from $n = 5$ to $n = 256$. As for the previous test example the estimation of Υ_i^* indices

converge faster than those Υ_i . In this test case the stopping criteria (29) is reached at $n = 41$ for Υ_i^* and $n = 59$ for Υ_i . The analytical values of the total effect indices as well as the values of Υ_i and Υ_i^* (estimated at the sample size $n = 59$ and $n = 41$) for the inputs which are identified as influential by the screening method are reported on Table 2. Notice that the values of the indices $S_{T_{12}, \dots, T_{200}}$ are smaller than 0.007 and then the corresponding inputs are considered to be non-influential. One can see that the informations provided by the indices $\Upsilon_1, \dots, \Upsilon_4$ are difficult to interpret. Indeed, for the g-Sobol function the values of Υ_i 's provide only a qualitative informations, because for some inputs $\Upsilon_i > 1$ which is higher than the maximal value for the total effect indices. These results may due to the model non linearity with respect to the inputs. On the other side, despite the non-linearity and non-monotonicity of the model the Υ_i^* measures performs very well in terms of quantitative interpretability.

The above two numerical tests, show us that both DGSM indices are adapted to identify the non-influential inputs. Moreover, the (27) property of Υ_i^* indices allows one to use an automatic screening method regardless the complexity of the studied model.

Input	Total effect	Υ_i (n=59)	Υ_i^* (n=41)
$X^{(1)}$	0.396	3.671	0.338
$X^{(2)}$	0.279	3.111	0.266
$X^{(3)}$	0.205	1.781	0.144
$X^{(4)}$	0.156	1.383	0.088
$X^{(5)}$	0.122	0.911	0.072
$X^{(6)}$	0.057	0.397	0.031
$X^{(7)}$	0.032	0.234	0.017
$X^{(8)}$	0.021	0.169	0.011
$X^{(9)}$	0.015	0.106	0.008
$X^{(10)}$	0.011	0.078	0.005
$X^{(11)}$	0.009	0.061	0.004

TABLE 2: Analytical total effect indices and estimated DGSM indices for the g-Sobol function

6 RESERVOIR FORECASTING APPLICATION

In this section, the proposed screening method is applied on a reservoir simulator. As the goal here is to apply the method on a high dimensional case, we choose to use horizontal and vertical permeability as inputs parameters. However, since the number of grid blocks in the considered reservoir simulation model is large, we applied the most basic parametrization technique which is the zonation to reduce the dimension of the problem. This technique consists in dividing the reservoir into relatively small number of zones (subregions)

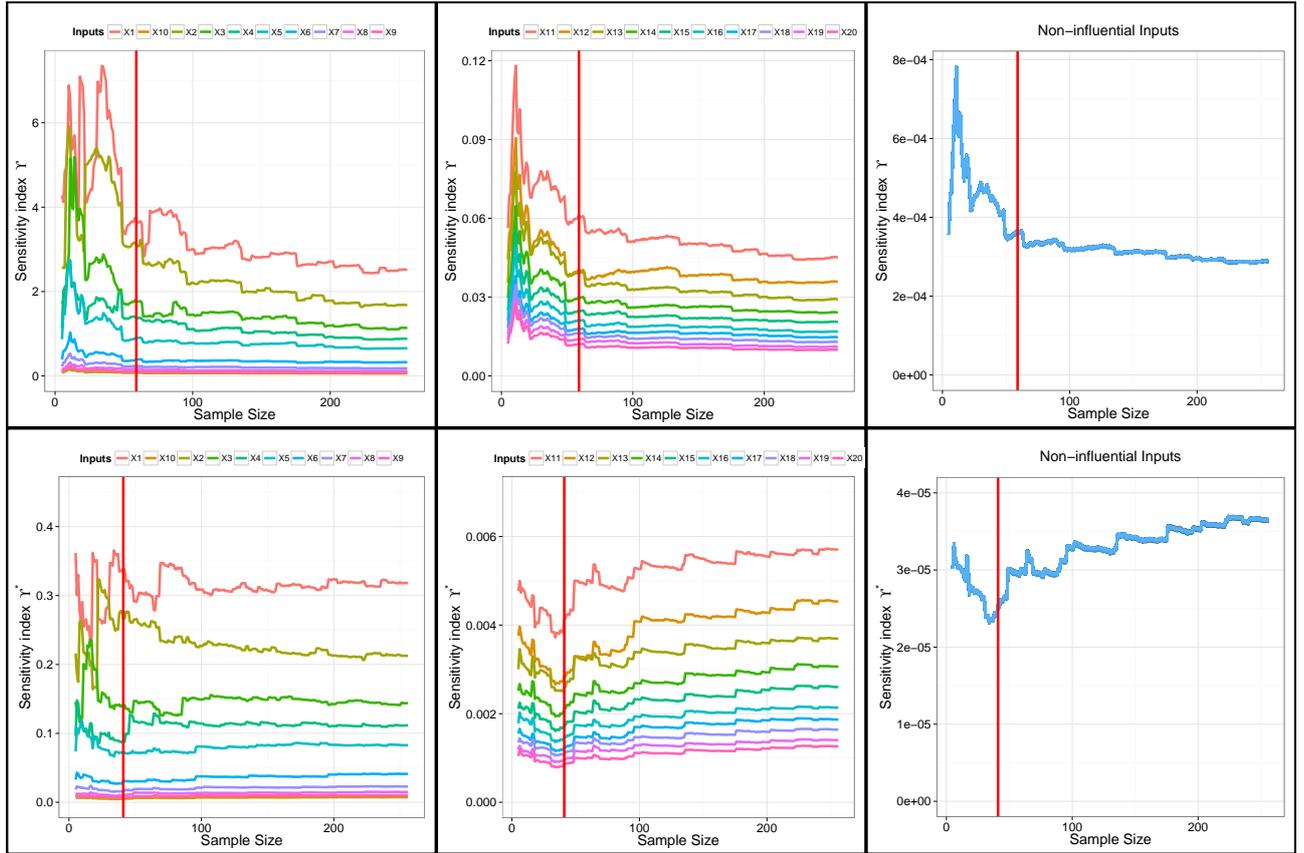


Figure 3: Convergence of the Υ_i and Υ_i^* indices estimates versus the sample size for the g-Sobol function

and assume that each zone is homogeneous. In other words, one fixes the permeability (horizontal or vertical) over all the grid blocks of the considered zone.

6.1 Reservoir model description

The PUNQS case is a synthetic reservoir model taken from a real field located in the North Sea. The PUNQS test case, which is qualified as a small-size model, is frequently used as a benchmark reservoir engineering model for uncertainty analysis and for history-matching studies Floris et al. [7]. The geological model contains $19 \times 28 \times 5$ grid blocks, 1761 of which are active. The reservoir is surrounded by a strong aquifer in the North and the West, and is bounded to the East and South by a fault (see Figure 4). A small gas cap is located in the centre of the dome shaped structure. The geological model consists of five independent layers, where the porosity distribution in each layer was modelled by geo-statistical simulation. The layers 1, 3, 4 and 5 are assumed to be of good quality, while the layer 2 is of poorer quality. The field contains six production wells located around the gas-oil contact. Due to the strong aquifer, no injection wells

are required. For more detailed description on the PUNQS model, see PUNQS [14].

As input parameters, we considered the horizontal and vertical permeability of 60 zones (12 for each layer). Thus, we have a model of 180 inputs, which are supposed independent and defined as follow

- $Z1, \dots, Z60$: horizontal permeability in X direction
- $Z61, \dots, Z120$: horizontal permeability in Y direction
- $Z121, \dots, Z180$: vertical permeability

The values of the permeability at each zone are distributed uniformly over $[PZ_i - 0.2PZ_i, PZ_i + 0.2PZ_i]$, where PZ_i is the arithmetic mean of the permeability values of the grid blocks which compose the i th zone.

The analyzed output is the production watercut data (the proportion of water in the produced oil) after 20 years of production of the well 5 which the perforation location correspond to the grid blocks [17; 11; 3 : 4], where the notation 3 : 4 means that there is a perforation at layer 3 and at layer 4. The reservoir test model was run using the PumaFlowTM [13] simulator, which allows one to compute gradients using gradient simulator method [1, 8] with an additional $\approx 33\%$

of the simulation time per each calculated gradient.

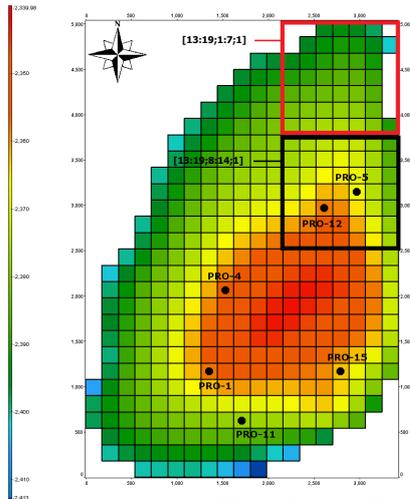


Figure 4: Top structure map of the PUNQS reservoir field (layer 1), the red square corresponds to the area of the grids position [13 : 19; 1 : 7; 1] and the blue square corresponds to the area [13 : 19; 8 : 14; 1].

6.2 Screening

For this reservoir model, we first performed a convergence study of the DGSM indices. Υ_i and Υ_i^* are computed sequentially with a QMC design ranging from $n = 5$ to $n = 200$. From Figure 5 one can see that Υ_i^* indices converge much faster than Υ_i indices. The stopping criteria (29) was reached at $n = 18$ for Υ_i^* and at $n = 42$ for Υ_i indices estimation.

The screening method identifies 13 zones as influential. The DGSM sensitivity measures of these 14 parameters are reported in Table 3. We can see that these zones correspond to the region where the studied well is located and the closest north region. This result demonstrates the relevance of application on a reservoir simulator of the developed screening methodology.

To corroborate these screening results we build a meta-model using a standard implementation of Gaussian process method (GP). The GP code used here is a commercial version implemented in the CougarFlowTM software [5]. For more detail on the technical aspect of the used GP we refer to section 3 of Busby et al. [3]. The GP meta-model is build using the results obtained on the QMC design of the size $n = 200$. However, instead of using the full design of 180 inputs we just selected the 13 inputs identified as influential by the screening method. Thus, rather than building meta-model that approximate a reservoir model of 180 inputs we build a meta-model \hat{f} which involves only the parameters

that supposed to be influential on the output. To assess the prediction accuracy of the meta-model we performed an extra 100 random evaluations of PUNQS simulator (with 180 inputs) and compare the simulator results to the meta-model ones. The measure of the accuracy is given by the Q_2 criteria defined by

$$Q_2 = 1 - \frac{\sum_{i=1}^{n_{test}} (y_i - \hat{f}(\mathbf{x}_i))^2}{\sum_{i=1}^{n_{test}} (y_i - \bar{y})^2}, \text{ with } n_{test} = 100 \quad (30)$$

where y_i denotes the i th simulator evaluation on test set, \bar{y} is their empirical mean and $\hat{f}(\mathbf{x}_i)$ is the predicted value at the design point $\mathbf{x}_i = (x_i^{(1)}, \dots, x_i^{(180)})$. The empirical Q_2 criteria of the considered meta-model \hat{f} is equal to 0.94, which means that the meta-model explain 94% of the output variance. Thus the obtained meta-model is sufficiently accurate to perform a global sensitivity analysis. In the second column of Table 3 the reported total effect indices was computed through the meta-model \hat{f} and using extended-FAST method. From Table 3 one can say that the developed screening method permits to detect the most important inputs. Concerning the inputs Z106 whose total effect index (estimated using the meta-model) is equal to zero, we can suspect that the influence of this parameter has been underestimated by the considered meta-model. We can notice that for this reservoir model the values of Υ_i and Υ_i^* indices have the same magnitude for all selected inputs at the stopping criteria and the values of Υ_i are smaller than the estimated total effect which is due to the underestimation of Υ_i indices at $n = 42$.

Input	Total effect	Υ_i (n=42)	Υ_i^* (n=18)	Grid blocks position
Z58	0.625	0.528	0.542	[13:19;8:14;5]
Z118	0.133	0.158	0.176	[13:19;8:14;5]
Z117	0.066	0.059	0.051	[13:19;1:7;5]
Z10	0.058	0.055	0.058	[13:19;8:14;1]
Z34	0.035	0.045	0.047	[13:19;8:14;3]
Z93	0.022	0.022	0.022	[13:19;1:7;3]
Z94	0.014	0.019	0.021	[13:19;8:14;3]
Z46	0.021	0.018	0.021	[13:19;8:14;4]
Z70	0.014	0.013	0.015	[13:19;8:14;1]
Z105	0.005	0.012	0.012	[13:19;1:7;4]
Z69	0.009	0.005	0.004	[13:19;1:7;1]
Z106	0	0	0.004	[13:19;8:14;4]
Z166	0.009	0.004	0.004	[13:19;8:14;4]

TABLE 3: Estimated sensitivity indices for the production watercut output after 20 years of production of the well 5

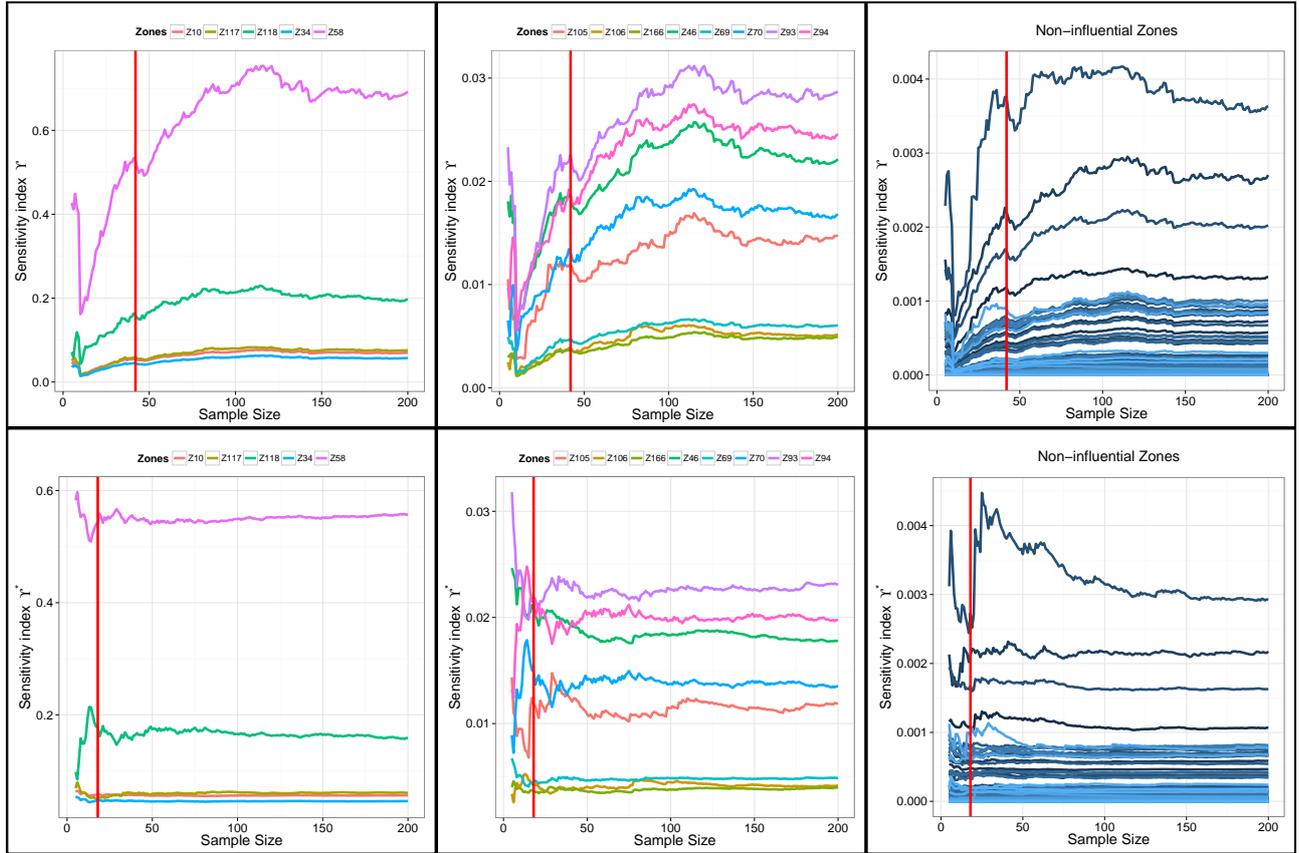


Figure 5: Convergence of the Υ_i and Υ_i^* indices estimates versus the sample size for the production watercut output after 20 years of production of the well 5. Graphics of the third line correspond to the Convergence of the Υ_i and Υ_i^* indices estimates of the non-influential inputs

7 CONCLUSIONS AND DISCUSSIONS

In this work, we presented a new sequential screening method which is based on DGSM indices. We defined a new DGSM index Υ_i^* in order to have a stronger quantitative measure to define the non-influential inputs. We also used the QMC Sobol sequence sampling method, which allows an intelligent sequential estimation of the DGSM indices in order to reduce the number of model evaluation. We empirically showed, by applying on two analytical models and a reservoir synthetic test case, that the proposed screening method is efficient to detect the non-influential inputs for an acceptable computational cost.

Computing DGSM indices requires model gradients estimation. A classical way to compute the derivatives is to use the finite-difference approximation method. However, this method suffer from the fact that the required number of model evaluation is equal to $n(d+1)$, where d is the number of inputs and n the number of points where derivatives are estimated. Since a reservoir simulator evaluation is gener-

ally time consuming, the finite-difference method is infeasible for models with a high number of inputs (roughly more than 20).

Therefore, it is clear that the ability to calculate derivatives efficiently is important to estimating DGSM indices within an acceptable computational cost. In the framework of reservoir simulation different methods have been developed for more or less computationally efficient gradients calculation. In this paper we utilized a reservoir simulator, which allows to compute gradients using the direct method, or also called in reservoir engineering the gradient simulator [1, 8]. This method is based on the solution of the governing analytical finite difference equations of flow, which automatically calculates the gradients during the simulation with an additional $\approx 33\%$ of the simulation time per each calculated gradient. Thus the required number of model evaluation to estimate DGSM indices is $\approx n(d+1)/3$. However, the most efficient method to calculate the gradient of a functional with respect to the reservoir simulator parameters is the adjoint state method when this functional depends on those models

parameters through state variables, which are the solution of the differential equation that define the problem. The advantage of this method comparing to the gradient simulator method is that it consists of the computation of one unique extra linear system and the computation of the gradient with respect to the model parameters is equivalent to one evaluation of the simulator. In other words, it means that the computational cost of gradient calculation is independent of the number of model parameters. So the required number of model evaluation to estimate DGSM indices is equal to $2n$. For more details on the mathematical aspect of the adjoint state method and its applications in reservoir simulation we refer to [6, 12, 15].

In addition to further testing on reservoir models, using a reservoir simulator which allows to compute gradients using adjoint state method is a topic of futur work.

ACKNOWLEDGEMENTS

This work was partially supported by French National Research Agency (ANR) through COSINUS program (Project COSTA BRAVA noANR-09-COSI-015). We wish to Frédéric Roggero, Bertrand Iooss and Frédéric Delbos for useful suggestions and discussions.

REFERENCES

- 1 F. Anterion, R. Eymard, and B. Karcher. Use of parameter gradients for reservoir history matching. *Proceeding of SPE Symposium on Reservoir Simulation, 6-8 February 1989*, (SPE 18433), 1989.
- 2 G. Blatman and B. Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering and System Safety*, 95:1216–1229, 2010.
- 3 D. Busby, C. L. Farmer, and A. Iske. Hierarchical nonlinear approximation for experimental design and statistical data fitting. *SIAM Journal on Scientific Computing*, 29(1):49–69, 2007.
- 4 F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22:1509–1518, 2007.
- 5 CougarFlow. IFP Energies nouvelles uncertainty analysis software. 2011. URL <http://www.openflowsuite.com/software/cougarflow>.
- 6 D. Eydinov, S. Aanonsen, J. Haukas, and I. Aavatsmark. A method for automatic history matching of a compositional reservoir simulator with multipoint flux approximation. *Computational Geosciences*, 12 (2):209–225, 2008.
- 7 F.J.T Floris, M.D. Bush, M. Cuypers, F. Roggero, and A.R. Syversveen. Methods for quantifying the uncertainty of production forecasts- a comparative study. *Petroleum Geoscience*, 7:S87–S96, 2001.
- 8 J.E. Killough, Y. Sharma, A. Dupuy, R. Bissell, and J. Wallis. A multiple right hand side iterative solver for history matching. *Proceeding of SPE Symposium on Reservoir Simulation, 12-15 February 1995*, (SPE 29119), 1995.
- 9 S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, and N. Shah. Monte carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering and System Safety*, 94:1135–1148, 2009.
- 10 A. Marrel, B. Iooss, F. Van Dorpe, and E. Volkova. An efficient methodology for modeling complex computer codes with gaussian processes. *Computational Statistics and Data Analysis*, 52:4731–4744, 2008.
- 11 M.D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–174, 1991.
- 12 R-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167 (2):495–503, 2006.
- 13 PumaFlow. IFP Energies nouvelles reservoir simulator. 2011. URL <http://www.openflowsuite.com/software/pumaflow>.
- 14 PUNQS. Production forecasting with uncertainty quantification. website. 1996. URL <http://www.fault-analysis-group.ucd.ie/Projects/PUNQ.html>.
- 15 B. Rama Rao and S. Mishra. Adjoint sensitivity analysis for mathematical models of coupled nonlinear physical processes. *Proceeding of Proceedings of the ModelCARE 96 Conference September 1996*, 1996.
- 16 J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical science*, 4: 409–435, 1989.
- 17 A. Saltelli and I. Sobol. About the use of rank transformation in sensitivity of model output. *Reliability Engineering and System Safety*, 50:225–239, 1995.
- 18 A. Saltelli and I. Sobol. A quantitative, model independent method for global sensitivity analysis of model output. *Reliability Engineering and System Safety*, 41(1):39–56, 1999.
- 19 A. Saltelli, K. Chan, and M. Scott. *Sensitivity analysis*. Wiley, 2000.
- 20 T. J. Santner, B. J. Williams, and W. I. Notz. *The design and analysis of computer experiments*. Springer, 2003.
- 21 I. Sobol and A. Gresham. On an alternative global sensitivity estimators. *Proceeding of SAMO 1995*, pages 40–42, 1995.
- 22 I. M. Sobol. Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, 16 (5):236–242, 1976.
- 23 I. M. Sobol and S. Kucherenko. On quasi-monte carlo integrations. *Mathematics and Computers in Simulation*, 47:103–112, 1998.
- 24 I. M. Sobol and S. Kucherenko. Derivative based global sensitivity measures and the link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79:3009–3017, 2009.
- 25 I. M. Sobol and S. Kucherenko. A new derivative based importance criterion for groups of variables and its link with global sensitivity indices. *Computer Physics Communications*, 181 (7):1212–1217, 2010.
- 26 I.M Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414, 1993.
- 27 C. B. Storlie and J. C. Helton. Multiple predictor smoothing methods for sensitivity analysis: Description of techniques. *Reliability engineering and system safety*, 93:28–54, 2008.
- 28 C. B. Storlie and J. C. Helton. Multiple predictor smoothing methods for sensitivity analysis: examples results. *Reliability Engineering and System Safety*, 93:57–77, 2008.

- 29 C. B. Storlie, L. P. Swiler, J. C. Helton, and C. J. Sallaberry. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability engineering and systems safety*, 94, (11): 1735–1763, 2009.
- 30 S. Touzani and D. Busby. Smoothing spline analysis of variance approach for global sensitivity analysis of computer codes. *Reliability Engineering and System Safety*, 112:67 – 81, 2013.