Title

Evaluation of multiple reduced-order models to enhance confidence in global sensitivity analyses

Authors

Zhang, Yingqi
Liu, Yaning
Pau, George
et al.

Publication Date

2016-06-01

DOI

10.1016/j.ijggc.2016.03.003

Peer reviewed

Evaluation of multiple reduced-order models to enhance confidence in global sensitivity analyses

Author links open overlay panel[YingqiZhang](#)[a][YaningLiu](#)[a][GeorgePau](#)[a][SergeyOladyshkin](#)[b][StefanFinsterle](#)[a]
Show more

https://doi.org/10.1016/j.ijggc.2016.03.003Get rights and content

Highlights

- Using multiple reduced order models can significantly reduce the computational cost needed for global sensitivity analysis.
- Validation set provides an evaluation of the performance of each reduced order model (ROM) as a basis for combining them.
- Validation set provides an indication of the confidence in the accuracy of the final estimates from multiple ROMs.
- Using the locally best ROM for combining multiple ROMs is able to screen out ROMs that have a large RMSE locally.

Abstract

Variance-based global sensitivity analysis (e.g., the Sobol' sensitivity index) can be used to identify the important parameters over the entire parameter space. However, one often cannot afford the computational costs of sampling-based approaches in combination with expensive high-fidelity forward models. Reduced-order models (ROM) can substantially accelerate calculation of these sensitivities. However, it is usually difficult to determine what type of ROM should be used and how accurately the ROM represents the high-fidelity model (HFM) results. In this paper, we propose to concurrently use multiple ROMs as a way to assess the robustness of the model-reduction method. Two sets of HFM simulations are needed, one set for building ROMs and the other for validating ROMs. Our goal is to keep the total number of HFM simulations to a minimum. Ideally some of the HFM simulations in the first set can be shared by different ROMs. Based on validation results, the ROMs can be combined with

different schemes. We demonstrate that we can achieve the goal by using four different ROMs and still considerably save computational time compared to using traditional HFM simulation for calculating sensitivity indices. We apply the approach to an example problem of a large-scale geological carbon dioxide storage system, in which the objective is to calculate a sensitivity index to identify important parameters. For this problem, the locally best ROM provides better estimates than the weighted average from all ROMs.

Keywords

Multiple reduced-order models
Global sensitivity analysis
Geological CO_2storage

1. Motivation

This work is motivated by the study of Wainwright et al. (2013), in which they performed a variance-based global sensitivity analyses (GSA) to calculate the Sobol' index (Sobol, 2001, Saltelli, 2002, Sobol et al., 2007) for performance evaluation of a large-scale geological carbon dioxide (CO_2) storage system. The GSA study is part of the National Risk Assessment Partnership (NRAP) project, providing a basis to determine what parameters should be included in the risk assessment framework. The numerical model used to describe this system, referred to as the Kimberlina model, includes twelve discontinuous or continuous (stacked) formations, extending 84 km in the eastern direction and 112 km in the northern direction, and involves the simulation of multi-phase flow processes and related thermodynamic complexities. To conduct variance-based global sensitivity analyses using such a model, the main challenges include: (1) the simulation time is long, i.e., each forward simulation with a high-fidelity model (HFM) takes about 8–12 h on a 12-processor computer; (2) simulations may not be able to finish for given set of modeling parameters, i.e., convergence problems are encountered for sampled parameter combinations that may be numerically unstable; and (3) the number of uncertain parameters is large, leading to the curse of dimensionality problem. To be able to calculate the global sensitivity coefficients, Wainwright et al. (2013) reduced the number of uncertain parameters to five and performed a total of 2100 HFM simulations. This task would be very challenging if the computational resources were not in place. Moreover, one might question whether the sample size is sufficiently large even for the reduced parameter dimension. As a result, we are motivated to propose an approach that substantially reduces computational costs of that

challenging problem, i.e., allows for the calculation of sensitivity indices with about 200 HFM simulations (less than 10% of 2100).

Reduced-order models (ROMs) have been proposed to make such an analysis possible for a computationally intensive forward model (e.g., Blatman and Sudret, 2010, Oladyshkin et al., 2012, Pau et al., 2013a). There are many ROM methods, each with its advantages and disadvantages depending on the properties (e.g., linearity) of the model or function that is being approximated, and the context within which the ROM is employed (e.g., for an accurate forward prediction or for an inverse analysis). It is usually difficult to find one method that is consistently superior to the others. Often times, a good choice of a ROM relies on some prior knowledge of the system response surface (e.g., about its linearity or smoothness). This knowledge may not always exist a priori, which makes applications of ROM methods difficult. In addition, once a ROM method is chosen and applied, the question is how to determine its accuracy, i.e., how well it is able to represent the results that would be obtained with the HFM. In this paper, we propose to use multiple ROMs to address these challenges, while keeping the number of HFM simulations to an affordable level. For our analysis we will require two sets of HFM simulations: one set for ROM construction and one set for ROM validation. We construct four different ROMs for the Kimberlina model. Based on the prediction error from the validation set, a combined model prediction can be obtained for any parameter values in parameter space. Finally we calculate Sobol' sensitivity indices using the predictions from the combined ROM.

The paper has the following structure. We first briefly describe the four ROMs (Section 2) and the Kimberlina model (Section 3). Then we focus on the ROM construction, validation, and combination for prediction (Section 4). Finally, we evaluate the accuracy and efficiency of the proposed method by comparing the results to those obtained by Wainwright et al. (2013), and discuss the confidence in the results (Section 5).

2. Reduced-order modeling methods

ROM refers to a model that is a computationally efficient approximation of an HFM. There are many different ways to approximate a model. In the discipline of subsurface flow and transport simulation, a ROM is usually one of the following:

A model with some simplified or ignored physics, which would have been modeled as accurately as possible in an HFM. This method can be used to study or separate the behaviors caused by various processes. However, replacing a HFM with a simplified

model can inadvertently neglect important process in a complex system that may be important for some combinations of parameters not studied when the simplified model is built.

A model that has a coarser mesh than the HFM but with all the known physics considered. The discrepancies between the coarse model and HFM must however be compensated either through an error model (Kaipio and Somersalo, 2004) or a map between the solutions of the coarse model and HFM (Pau et al., 2014).
A model obtained using projection-based model reduction techniques, which has been widely used in the field of computational fluid dynamics (e.g., Willcox and Peraire, 2002). Last decades, the polynomial chaos expansion (PCE) in which a HFM is projected onto orthogonal or orthonormal polynomial bases (e.g., Xiu and Karniadakis, 2002) has become popular. The intrusive version of it is less practical because it requires an effort in code modification. While mathematically vigorous for approximating partial different equations, these methods are powerful for nonlinear subsurface models.
A response surface model, which establishes an input–output relation between the studied parameters and outputs of interest using either an interpolation technique, a statistical regression method, or a learning method, such as artificial neural network. Model outputs are approximated at un-evaluated parameter points based on existing HFM runs at some sampled parameter points, referred to as snapshots. Examples include linear lookup tables, kriging, non-intrusive polynomial chaos expansion (PCE) (e.g., Oladyshkin et al., 2011, Oladyshkin and Nowak, 2012), Gaussian process regression (GPR) (e.g., Rasmussen and Williams, 2006) and radial basis functions (RBF) (e.g., Regis and Shoemaker, 2004, Regis and Shoemaker, 2005).
Within NRAP, look-up tables, multi-variate adaptive regression spline (MARS) (Harp et al., 2016, Jordan et al., 2015), multi-fidelity models (Bianchi et al., 2016) and ROMs based on field data (Zhang et al., 2016), were developed and used to predict the behavior of various components (i.e., model output for various parameter combinations in the parameter space). Our interest in this work is not to develop a ROM to predict a specific system behavior (e.g., pressure), but to estimate a statistical property that provides an integrative quantification of the model output (e.g., sensitivity). In this paper, we will focus on response surface models since these models are usually non-intrusive, which means one does not have to re-write governing equations. They are relatively easy to implement compared to intrusive methods, and less dependent on the specific problem applied. The methods in the second and third categories are intrusive methods, i.e., the governing equations are changed (for the first category the number of governing

equation is different between a ROM and its HFM). Razavi et al. (2012)provided a comprehensive review on many of these methods.
This study addresses the important issue in practical applications of ROMs that the accuracy of a ROM cannot be easily assessed without comparison to the HFM results obtained during the prediction phase, which are not available in any practical application involving ROMs. We propose to examine the robustness of the approximation by comparing multiple ROMs developed for the same HFM, and to estimate their relative accuracy based on results from a validation set. This approach may also reveal shortcomings or errors in the application of certain ROM methods; they will be discussed as they arise.

Multiple ROM methods have been applied in the field of optimization and been tested for a number of test functions. However, we have not seen them being used and tested for a real application. For example, Muller and Piche (2011)proposed to combine multiple surrogate models via Dempster–Shafer theory. The pignistic probabilities of the models are calculated based on four model characteristics: Correlation coefficients (CC), root mean squared errors (RMSE), maximal absolute errors (MAE) and median absolute deviation (MAD). They found applying model combinations becomes more favorable with increasing number of parameters for the tested functions.
The proposed approach is demonstrated for a variance-based sensitivity analysis (i.e., the evaluation of the Sobol' index), based on the HFM analysis of Wainwright et al. (2013). The Sobol' global sensitivity analysis method provides parameter rankings based on how much uncertainty in the final prediction can be attributed to each input parameter. Usually the large number of HFM evaluations necessary to perform this type of sensitivity analysis becomes a major restriction for its application. In the current study we substantially reduce the computational costs because of ROM methods employed to calculate these sensitivities. Blatman and Sudret (2010) and Oladyshkin et al. (2012) also proposed to use ROM for global sensitivity analysis, but these studies did not address the issue of how to evaluate the estimate of the sensitivity indices if there is no reference available, and therefore how to assess the appropriateness of the selected ROMs in representing the HFM.
We propose to construct different ROMs at the same time for estimating Sobol' indices and obtaining rankings of influential parameters. Then we compare these ROM-predicted indices and associated parameter rankings. A composite result based on the individual ROM results and its variance is also calculated. The goal is to minimize the

total amount of HFM simulations by constructing the different ROMs using shared snapshots.

The following four ROM methods are considered: (1) Gaussian process regression (GPR), which is a Bayesian approach that models the response surface by a mean and a covariance function (Rasmussen and Williams, 2006, Pau et al., 2013a, Pau et al., 2013b); (2) arbitrary polynomial chaos (aPC), which is a stochastic approach in which the model output is represented by a polynomial chaos expansion based on an arbitrary parameter distribution without the need to re-write the governing equations (Oladyshkin and Nowak, 2012); (3) cut high-dimensional model representation (Cut-HDMR) (Rabitz et al., 1999, Li et al., 2001); and (4) random sampling high-dimensional model representation (RS-HDMR) (Li et al., 2002). The two HDMR methods establish input-output relations by multivariate representations, assuming only relatively low-order correlations of the input variables will have an impact on the output, and high-order terms can be ignored (Rabitz et al., 1999, Li et al., 2001).
The reasons to select these four ROMs for calculating the Sobol' index are the following: (1) aPC takes into account the parameter distribution for best estimation of mean stochastic characteristics and the Sobol' index can be directly obtained from the statistical moments calculated by aPC; (2) the two HDMR methods are similar to aPC in that they also use polynomial approximations (Lagrange interpolation for Cut-HDMR and orthogonal polynomial bases for RS-HDMR). However, the differences that exist between the methods may affect the ROM approximation and ultimate results; we investigate these effects; (3) GPR is an approach similar to kriging and can thus be considered an alternative to the other three ROM methods; (4) both GPR and RS-HDMR can be constructed at no extra computational cost (i.e., they can use the same HFM snapshot simulations used to construct aPC and Cut-HDMR).

In the sections below we will briefly describe the basic theories behind the four ROM approaches. The complete formulation and detailed implementation of these ROM methods can be found in the references provided in the previous section.

2.1. Gaussian process regression (GPR)

GPR can be considered a generalized kriging method, except that the (input) variables $\mathbf{p}$ are not limited to spatial variables. The underlying assumption is that the model output $f(\mathbf{p})$ can be characterized by a mean $m(\mathbf{p})$ and a covariance function $k(\mathbf{p},$

p′). The GPR approximations of the model output are represented by $g(\mathbf{p})$. The joint distribution of $f(\mathbf{q})$ and $g(\mathbf{p})$ is then

(1)[f(q)g(p)]∼N(m(q),[K(q,q)K(q,p)K(p,q)K(p,p)]),q∈SN

where $\mathbf{K}_{ij}(\mathbf{q},\mathbf{q}) = k(\mathbf{q}_i,\mathbf{q}_j)$, subscription i and j represent snapshots. $\boldsymbol{S}_N$ represents the sample set containing N snapshots $\{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N\}$ for ROM construction. For any given **p**, the GPR procedure gives the expected value and variance of the approximating function $g(\mathbf{p})$:

(2)E[g(p)]=K(p,q)K(q,q)−1f(q)+m(q),q∈SN

(3)σg2[g(p)]=K(p,p)−K(p,q)K(q,q)−1K(p,q),q∈SN

The use of a variance allows us to estimate the uncertainty due to the ROM approximation. Prior information of the response surface will be helpful in the selection of an appropriate covariance function. In addition, for the squared exponential covariance function we are using, one needs to determine three hyperparameters σ_f^2, l and σ_n^2 in the formulation of a Gaussian process, representing the variance of $f_i(\mathbf{q})$, the characteristic length of the parameters, and the noise variance of $f_i(\mathbf{q})$, where $\mathbf{q}\in\boldsymbol{S}_N$. These hyperparameters are obtained by solving an optimization problem that maximizes the marginal Gaussian likelihood function (the likelihood that the chosen covariance function is a good approximation of the true covariance function). The parameter distribution is not used to build GPR ROM, which means no prior knowledge on such distribution is needed. Since GPR can utilize any arbitrary set of parameters, we use all samples in the training set to construct a ROM based on GPR.

2.2. Arbitrary polynomial chaos (aPC)

The polynomial chaos expansion (PCE) method was introduced by Wiener (1938). The basic idea of PCE is to project the model response surface onto an orthogonal basis in the parameter space, which is an efficient polynomial projection to include nonlinear effects in the stochastic analyses. Among various techniques, we are interested in the non-intrusive approach, which does not require us to manipulate the PDEs of the forward model. The classical PCE method relies on normally distributed parameters to obtain an optimal solution. In recent years, the classical PCE has been extended to generalized polynomial chaos (gPC) (Xiu and Karniadakis, 2002) to take into account a number of theoretical statistical distributions, such as gamma, beta and uniform distributions, and arbitrary polynomial chaos (aPC) to handle parameters with arbitrary distributions (Oladyshkin and Nowak, 2012). If an arbitrary polynomial basis is used (Oladyshkin et al., 2011), the input parameters can have arbitrary distributions. The distribution format can be either discrete, continuous, or discretized continuous, or

specified either through some statistical moments or an analytical expression of the density function.

Based on polynomial chaos theory, the model output can be approximated as

(4)a(p)=∑i=1Nciϕi(p)

Here the function ϕ_i is the multi-variate orthogonal polynomial basis for the vector **p** of input parameters, and c_i are the expansion coefficients that relate model outputs to model inputs. The minimum number of terms N depends on the order of the expansion d (i.e., the polynomial order) and the number of parameters n, as shown in Eq. (5):

(5)N=(d+n)!(d!n!)

Once the orthogonal bases are obtained for arbitrary distributions (Oladyshkin and Nowak, 2012), the remaining problem is to find the N unknown coefficients c_i(i = 1, …, N). This requires at least N equations, which means at least N parameter samples and N HFM simulations are needed. A sampling method named collocation method demands only the number N of HFM simulations and the ROM solutions are exact at the sampling points. The optimal choice of sampling (collocation) points is taken as the roots of the polynomial of one degree higher than the order used in the chaos expansion according to Villadsen and Michelsen (1978).However, if the number of HFM simulations is more than N (the number of equations is more than the number of unknowns), a least-squares method can be used to determine the coefficients.

aPC has mainly been applied to uncertainty quantification because (1) the locations of collocation points are chosen based on the parameter distribution for best estimation of mean stochastic characteristics, and (2) the statistical moments of model outputs as well as the Sobol' index can be calculated analytically. In this work, we will take advantage of this analytical form and use it to test whether the sample size for calculating the Sobol' index in our application is large enough.

2.3. High-dimensional model representation (HDMR)

The HDMR method is based on the assumption that a multivariate function representing a multidimensional physics output can be represented by contributions from independent individual input and correlated contributions from multiple input variables/parameters (Rabitz et al., 1999):

(6)f(p)=f0+∑i=1nfi(pi)+∑1≤i<j≤nnfij(pi,pj)+...+f1,2,...n(p1,p2,...,pn)

Here, **p** is the input vector containing n parameters, and f_0 (i.e., the zeroth-order term) is a constant representing the mean response of f(**p**). The first-order terms (i.e, $f_i(p_i)$) denote the independent contribution to f(**p**) by the ith input variable acting alone, and

higher-order terms reflect correlated contributions from multiple input variables (i.e., the second-order term $f_{ij}(p_i,p_j)$ represents the correlated contribution from variable p_i and p_j, and so on). The last term contains the residual nth order correlated contribution of all input variables. Experience shows that expansion to second order and ignoring higher-order terms can often provide a good approximation (Li et al., 2001, Ziehn and Tomlin, 2008). Notice the expansion order here in HDMR is different than the polynomial order in aPC. For example, the first-order term $f_i(p_i)$ in HDMR may contain higher-order polynomials.Two typical HDMR expansions are Cut-HDMR and random-sampling HDMR (RS-HDMR) depending on the expression of each term on the right-hand side of Eq. (6). In Cut-HDMR, a reference point $p¯=(p¯1,...,p¯n)$ is chosen in the n-dimensional input variable **p** space, and the component functions with respect to reference point $p¯$ have the following forms

(7)$f0=f(p¯)fi(pi)=f(pi,p¯i)−f0fij(pi,pj)=f(pi,pj,p¯ij)−fi(pi)−fj(pj)−f0...$

Here the notation $f(pi,p¯i)=(p¯1,...,p¯i−1,pi,p¯i+1,...,p¯n)$ means that all input variables are at their reference point except pi.

The construction of a Cut-HDMR includes three steps: (1) determine the expansion order and truncate the higher-order terms; (2) select sample points for each input variable to evaluate the function terms using Eq. (7); (3) select an interpolation scheme (e.g., Lagrange interpolation) to approximate the component function values at other points. The Lagrange interpolation scheme results in a polynomial approximation of each component function. If each input variable is sampled at s different values, the total number of HFM simulations N needed for a Cut-HDMR can be calculated using Eq. (8), where d is the expansion order:

(8)$N=∑i=0dn!(n−i)!i!(s−1)i$

In our study, the sample points for each input variable are chosen as the roots of Legendre polynomials mapped to the parameter ranges (Liu, 2013). The Sobol' indices can be approximated by integrating the Cut-HDMR ROM with Gauss–Legendre quadrature with the above-mentioned sample points as abscissas.The second type of the expansion is RS-HDMR. It is also referred to as ANOVA-HDMR (aHDMR). Similar to the statistical tool ANOVA (analysis of the variance), each component function in the expansion is a random quantity uniquely contributing to the overall variance of the output, taking an integral expression:

(9)$f0=∫Knf(p)dpfi(pi)=∫Kn−1f(p)dpi−f0fij(pi,pj)=∫Kn−2f(p)dpij−fi(pi)−fj(pj)−f0...$

Here dpi stands for the product $dp1dp2...dpn$ without dpi. This expansion is mainly used for the purpose of statistical analysis (Rabitz et al., 1999). To reduce the computational burden in approximating integrals, Alis and Rabitz (2001)suggest approximating each

component function by an orthogonal polynomial series expansion. The coefficients of the expansion can then be approximated by selecting a single set of random samples and running the HFM at the selected samples. Similar to aPC, the Sobol' sensitivity indices can be readily extracted from the expansion coefficients, as a result of the orthogonality of the polynomial bases employed. The sample size *N* is typical chosen based on consideration of affordable computational cost and approximation accuracy, i.e., this size *N* is determined by the user.

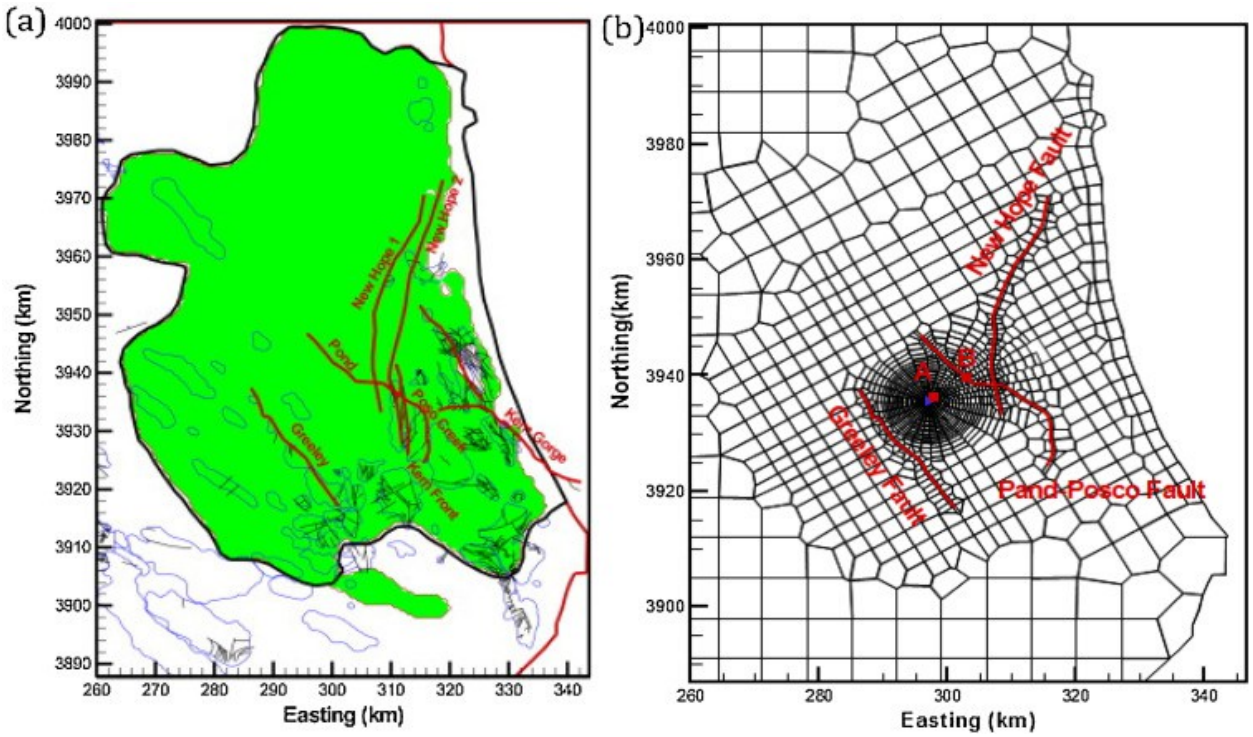
Depending on the details of the implementation of HDMR, both HDMR methods can be polynomial approximations as is aPC. However, the main differences are that (1) the polynomial order for aPC is determined by the user, wheras the polynomial order for Cut-HDMR is determined by *s* (number of samples for each parameter) if the Lagrange interpolation is used, and the polynomial order for RS-HDMR is also determined by the user; (2) the total number of snapshots needed is different, which depends on polynomial order in aPC; the expansion order and *s* in Cut-HDMR. There is no strict requirement for RS-HDMR; and (3) the actual snapshot locations, as well as the choice of the orthogonal bases and expansion terms are different: they are determined by the underlying parameter distribution in aPC; they generally are uniformly distributed in Cut-HDMR; they could be anywhere in the parameter space in RS-HDMR.

3. The Kimberlina model

The Kimberlina model (Zhou et al., 2011, Birkholzer et al., 2011, Wainwright et al., 2013) was developed to simulate a hypothetical geological CO_2 storage project in the Southern San Joaquin Basin in California, USA. The geological model was based on actual field data from many oil wells in that region, reflecting realistic geological CO_2 storage reservoir conditions. The numerical model domain shown in Fig. 2 is about 84 km by 112 km in each direction. The storage formation, the Vedder formation, is about 400 m thick and at about 2750 m below the ground surface. The formation dips at an average slope of 7° from a shallow outcrop area located along the eastern model boundary. The Vedder sand is quite permeable, providing sufficient injectivity. The overlying Temblor–Freeman Shale with a thickness of 200 m is considered a suitable caprock for stratigraphic containment of the injected supercritical CO_2. The hypothetical scenario is that CO_2 is to be injected at the center of the model (the center of the spider net grid) at a rate of 5 M ton/yr for 50 years. The 3D numerical mesh of the high-fidelity model (HFM) contains twelve geological formations, represented by about 65,000 elements. The refined mesh in the center of the domain reflects expectation of multiphase processes and strong pressure buildup due to CO_2 injection. The HFM simulation is

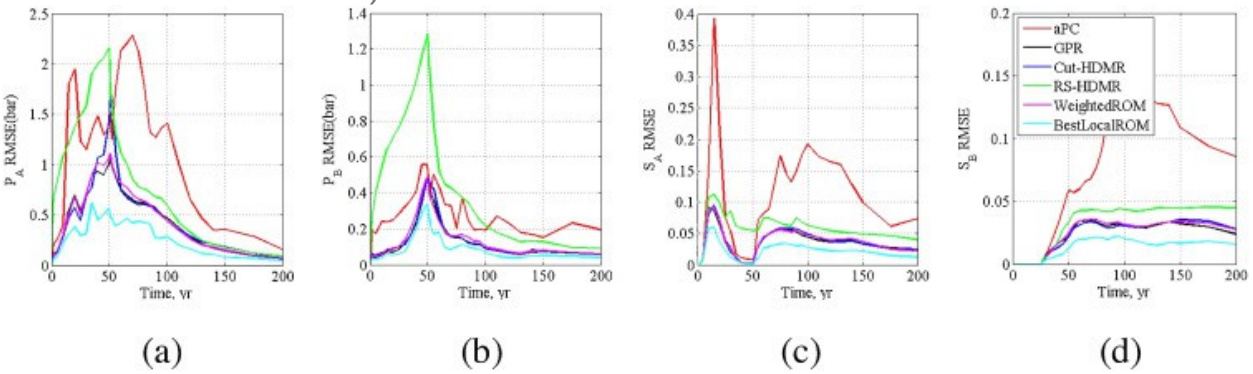
performed using the massively parallel multiphase simulator TOUGH2-MP (Zhang et al., 2008) with the ECO2N module to simulate injection and migration of supercritical CO_2 in the brine reservoir. The ECO2N module describes the thermodynamics and thermophysical properties of H_2O–NaCl–CO_2 mixtures, including phase transitions and dissolutions (Pruess, 2005). The period to be simulated consists of 50 years of injection and 150 years of post-injection. A typical HFM simulation takes about 8–12 h on a 12-processor computer.

In this paper, we use the Kimberlina model of Wainwright et al. (2013) as the high-fidelity model. Examining modeling errors of that model, specifically those caused by the assumption that geological properties are homogeneous within a given layer, is outside the scope of this study. Based on the intended use of the model, different outputs are studied as the performance measure of interest. For example, plume size is likely a relevant metric, so is the overpressure zone above a critical pressure. We consider four outputs of interest: S_A, S_B, P_A and $P_B$representing the CO_2 saturation (S) and pressure increase (P) at two selected locations (Fig. 1b), Point A (near field) and Point B (fault location), which are 1.8 km and 7.3 km updip from the injection location, respectively. They are considered as point-based performance measures of CO_2 plume and pressure behavior due to CO_2 injection. We include both point pressure and saturation outputs because pressure is representative of relatively linear, smooth (over time) behavior, and point saturation, which is representative of a very non-linear (almost binary) output. Since our goal is to study how ROMs' abilities to approximate outputs of different smoothness affect the determination of global sensitivities indices, we omit other outputs that were studied in Wainwright et al. (2013). Five parameters are considered uncertain: reservoir horizontal permeability (Res. k_h); reservoir porosity (Res. φ); reservoir compressibility (Res. $_{\beta p}$); reservoir van Genuchten m parameter (Res. m); and caprock permeability (Caprock k_h). The two permeabilities follow a lognormal distribution, while other parameters are normally distributed. As mentioned earlier, we want to test the accuracy of our approach by comparing results to the ones obtained by Wainwright et al. (2013). The parameter ranges are taken from Wainwright et al. (2013). Wainwright et al. (2013) discussed two variance-based sensitivity measures: The Sobol' sensitivity index to identify the influential parameters, and the total sensitivity index to identify unimportant parameters. The computational costs of the two are similar. We will only discuss the Sobol' sensitivity index.



1. Download high-res image (1MB)
2. Download full-size image

Fig. 1. Plan view (Wainwright et al., 2013) of (a) the Vedder formation (green area) with faults (red lines), and (b) the model domain with numerical grid. In (a), blue polygons show hydrocarbon fields in the region with data used for the development of geologic model and spatial distribution of rock properties. In (b), the red lines delineate the faults that are explicitly represented in the model, the blue point is the injection location, and the red dots (Points A and B) are used for point-based performance measures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



1. Download high-res image (483KB)
2. Download full-size image

Fig. 2. The root mean square error (RMSE) from the comparison to validation set 1 for (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.

The Sobol' sensitivity index S_i, also referred to as the first-order sensitivity index, measures how much uncertainty (variation) in the output $f(\mathbf{p})$ comes from each input parameter p_i. For consistency, we use the same algorithm and the same realizations used by Wainwright et al. (2013) to calculate S_i, which is based on an algorithm developed by Saltelli et al. (2006) and modified by Glen and Isaacs (2012). The number of simulations required by this algorithm is $m(n + 2)$, where $m$is the number of the randomly generated sets of parameters, and n is the number of parameters. Wainwright et al. (2013) set $m = 300$ and $n = 5$, resulting in 2100 HFM simulations. The question remains if $m = 300$ is sufficient since no convergence study (i.e., calculating the sensitivity index with an increasing number of realizations until the results are stable) was performed.

4. ROM-based analysis

4.1. ROM construction

Our goal of using a ROM to calculate the Sobol' sensitivity index S_i is to reduce the computational cost by at least a factor of 10, which means the number of HFM simulations for all the ROMs should be less than 210. We make the following decisions:

1. We use a second-order aPC, which needs 21 HFM simulations according to Eq. (5). The work of Oladyshkin et al., 2011, Oladyshkin et al., 2012indicates that sufficiently accurate results can be obtained with a second-order aPC.
2. We use the second-order expansion of the Cut-HDMR method since it is typically sufficient for most problems (Rabitz et al., 1999, Li et al., 2001). In addition, $s = 3$ samples (mapped Legendre polynomial roots) is used for each parameter to limit the number of HFM simulations to 51.
3. The total number of HFM runs needed to construct aPC and Cut-HDMR is thus 72. Results from these 72 HFM simulations are then used to construct both GPR and RS-HDMR, i.e., these two ROMs are built at no extra HFM simulation costs.
4. GPR and RS-HDMR are constructed based on the same 72 HFM snapshots.

5.

The next step is to validate the four ROMs.

4.2. ROM validation

The goal of the validation set is to investigate how well each ROM can estimate the output of interest over the entire parameter range. Therefore, the parameter samples should be as uniformly distributed as possible over the entire range. For this purpose, we generate a Sobol' sequence (Sobol, 1967) containing 100 HFM realizations, with the first 50 realizations forming validation set 1, and all 100 realizations forming validation set 2. The Sobol' sequence provides the best uniform distribution in the multi-dimensional unit hypercube for a given number of realizations. Comparing errors from the two validation sets can indicate if a sample size of 50 is sufficient for validation purposes.

We will use the root mean square error (RMSE) at different times to compare the ROM predictions relatively the HFM (set 2 is shown in Fig. 2). The RMSEs of the ROMs have similar temporal characteristics. Figure 4 of Wainwright et al. (2013)shows the time evolution of pressure and saturation at points A and B for their reference parameter set. Figure 5 of Wainwright et al. (2013) is a set of Monte Carlo simulation results showing the time evolution of the pressure and saturation at points A and B, which can be used to understand these RMSEs. For pressure estimation, the error tends to be highest around year 50, corresponding to the highest pressure build-up at the end of injection. For S_A predictions, two main increases in RMSEs from the ROMs correspond to a sudden saturation increase at early time between 20 and 30 years, and a decrease around 70 years in most realizations. The actual location of relatively large errors varies for different ROMs, because the time when S_A increases or decreases depends on the ROM's accuracy with which it represents the system state at the particular realization of the validation parameter set. This behavior can be attributed to the fact that in general ROMs cannot capture sudden output changes very well, as is the case for S_A. However, the RMSE curve for S_B predictions is smoother than for $S_A$predictions due to the fact that Point B is much farther away. For all the ROMs, the largest error over time is higher at Point A than at Point B. This is because Point A is closer to the injection point, which leads to a larger magnitude in both pressure and saturation changes over the simulation period.

Amongst the four different ROMs, GPR and Cut-HDMR have the smallest RMSE, and their predictions are almost indistinguishable. Surprisingly, RS-HDMR performs poorly for pressure and reasonably well for saturation. This indicates that RS-HDMR has a

unique advantage in situations where the response surface is not smooth. In this analysis, aPC performs poorly compared to GPR and Cut-HDMR. In particular, aPC has large errors in saturation at a few (temporal) locations. However, the large error is likely due to a mismatch between the distribution of the training sample set and the validation sample set. Recall that the HFM simulations used to build aPC are selected based on the parameter distribution, i.e., more samples are taken from the region with higher probability density, compared to the validation set, which is sampled from a uniform distribution. This leads to relatively large errors from realizations in the validation set that are far outside the snapshot range (recall that only 21 snapshots were taken in a five-dimensional parameter space to construct aPC).

If our final goal were to compare the four ROMs, the comparison would not have been fair because the number of HFM runs were different for different ROMs. However, such a comparison is just an intermediate step for validating ROMs and it does not serve to select the best ROM. For aPC and Cut-HDMR, there is an optimal way of choosing the samples used to construct the ROM; it depends on the number of parameters and polynomial orders, resulting in a different number of HFM runs used to construct these ROMs. On the other hand, GPR and RS-HDMR can handle any given set of samples, allowing us to just reuse the samples used by aPC and Cut-HDMR. Thus, we optimally created 4 different ROMs using the minimal total number of HFM runs.

There is not much difference in the RMSE obtained for validation sets 1 and 2, indicating that a Sobol' sequence of 50 realizations is sufficient for validation purposes involving five parameters. Nonetheless, for combining individual ROMs (see next section), we use the results from validation set 2, which contains 100 realizations.

4.3. ROM combination

Even though on average GPR and Cut-HDMR estimates have smaller errors than RS-HDMR and aPC estimates, each ROM has the potential to provide better estimates than its competitors depending on (1) the output of interest; (2) the prediction time; (3) the sampled parameter set. As a result, we decide to investigate two methods to estimate predictions using a combination of all four ROMs: (1) use a weighted average to combine the predictions from all four ROMs with weights computed from (10)wi=1RMSEi∑i=141RMSEi
where w_i is the weight (a function of output of interest and time), and i is an index indicating estimates from aPC, GPR, Cut-HDMR and RS-HDMR; and (2) use the ROM

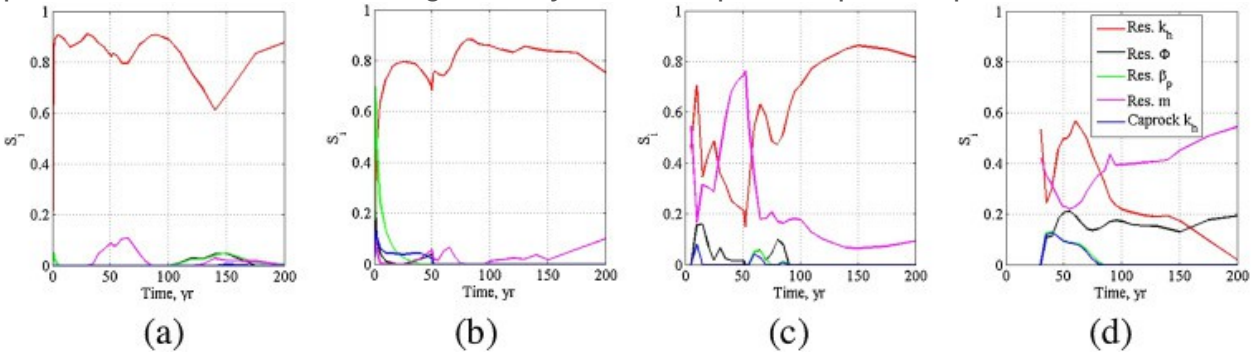
estimate that gives the best local estimate, i.e., for each prediction, find the closest realization (in parameter space) from the validation set, check which ROM produces the least error for an output of interest at a particular time, then use that ROM for this particular prediction of the output of interest at this time.

Combining ROMs using the weighted method requires a total number of weights that is equal to the product of the number of outputs (i.e., P_A, P_B, S_A, S_B) times the number of prediction times, i.e., in our study 4 × 32 = 128.

Fig. 2 also shows the RMSE for the prediction of 100 realizations in the validation set using both the weighted average ROMs and the locally best ROM. As expected, the error from the locally best ROM is smallest compared to individual ROM predictions and the weighted average ROM predictions. For this particular problem, we have also tested different weighting schemes for calculating w_i when computing the RMSE of the weighted average ROM and found the results insensitive to the weighting scheme used. This is because both GPR and Cut-HDMR have similar RMSEs that are either significantly smaller than or the same as those of other ROMs, and thus similar larger weights or same weights (compared to other ROMs) no matter what schemes are used.

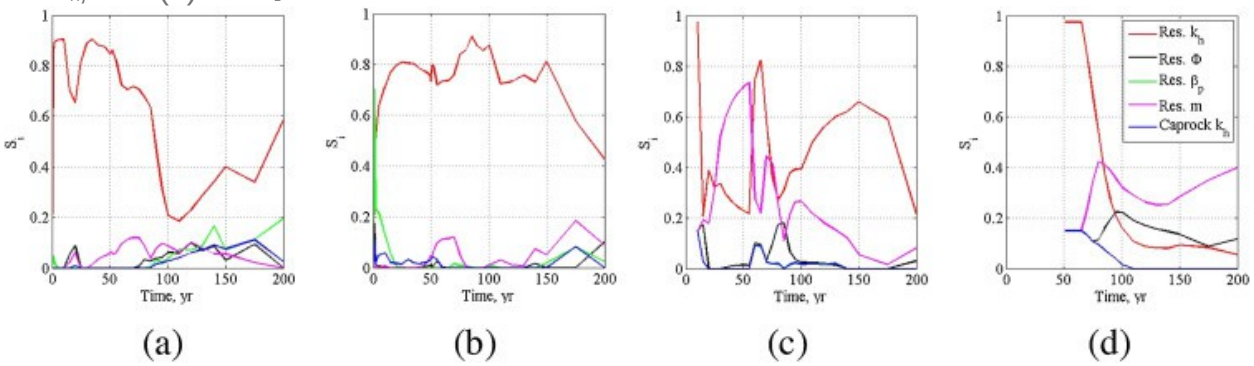
4.4. ROM-based Sobol' sensitivity index

The estimated S_i from each individual ROM is plotted in Fig. 3, Fig. 4, Fig. 5, Fig. 6. For comparison purpose, Fig. 9 shows the Sobol' indices calculated based on the 2100 HFM simulation results from Wainwright et al. (2013). Because the pressure propagates much faster than the migration of the CO_2, the pressure at Points **A** and **B** is sensitive to changes in the parameters almost immediately after CO_2 injection starts, while it takes some time for the sensitivity of saturation to appear at the more distant Point **B**. Pressure is primarily sensitive to only the permeability k_h over the entire simulation period, while saturation has significantly more complex temporal dependence on time.



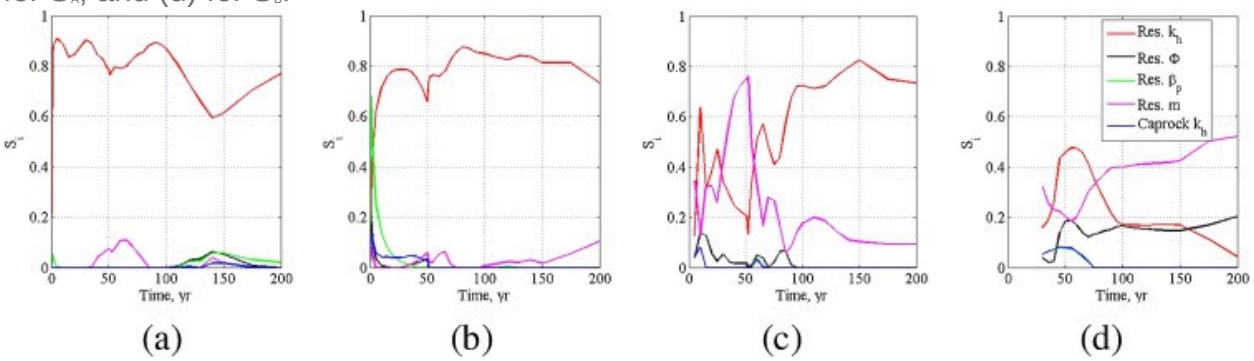
1. Download high-res image (401KB)
2. Download full-size image

Fig. 3. The Sobol' sensitivity index S_i calculated using GRP (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.



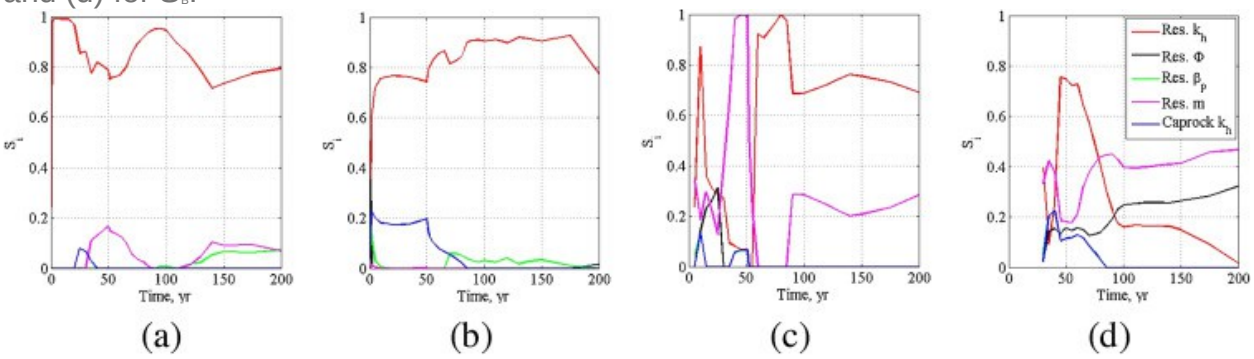
1. Download high-res image (428KB)
2. Download full-size image

Fig. 4. The Sobol' sensitivity index S_i calculated using the aPC (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.



1. Download high-res image (377KB)
2. Download full-size image

Fig. 5. The Sobol' index S_i calculated using the Cut-HDMR (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.



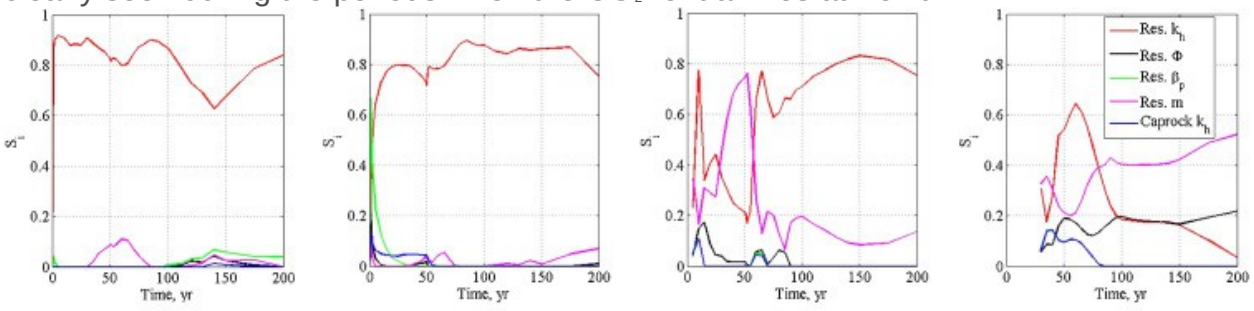
1. Download high-res image (387KB)
2. Download full-size image

Fig. 6. The Sobol' index S_i calculated using the RS-HDMR (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.

Compared to Fig. 9, all ROMs are able to identify the influential parameters correctly at low computational costs. This suggests ROMs with larger RMSE, i.e., aPC and RS-HDMR, are useful in statistical applications, such as uncertainty quantification; conversely, they may not be applicable in areas that need accurate individual predictions, such as in inversions or other optimization problems. Large errors for isolated parameter sets will only have a small effect on the overall statistical results evaluated over the entire parameter space.

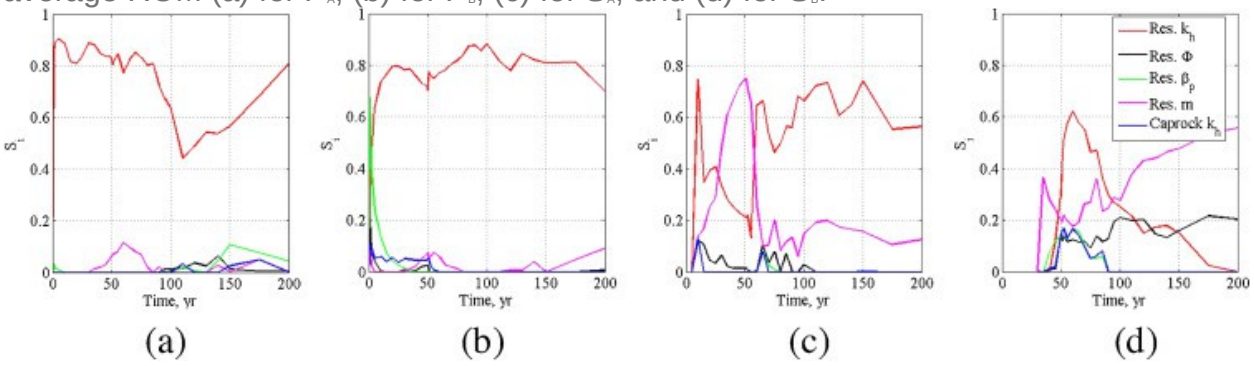
However, ROMs with smaller RMSE, as determined from the previous section, are able to capture the temporal changes in S_i more accurately. GPR and Cut-HDMR both result in very similar S_i estimates compared to those determined using HFM. For aPC, S_i starts to be non-zero at about 50 years, while the other three ROMs calculate non-zero sensitivities after approximately 30 years. The later appearance of S_i is due to the fact that none of the 21 training HFMs used to build the aPC ROM results in a CO_2 plume arrival earlier than 50 years. However, Fig. 6shows that despite the large RMSE of the pressure predicted by RS-HDMR, the negative effects on the computed S_i is small and limited to parameters that are non-influential. Thus, while ROMs with small RMSE generally estimate S_i more accurately, ROMs with larger RMSE can still lead to a relatively good estimate of S_i.

Fig. 7, Fig. 8 are the Sobol' indices calculated from the 2100 predictions obtained using the weighted average ROM and locally best ROM, respectively. Since GPR and Cut-HDMR have smaller RMSE, and therefore larger weights in the weighted average ROM, the S_i estimates using the weighted average ROM are expected to be similar to the ones from GPR and Cut-HDMR. This is confirmed by Figs. 3, 5 and 7. The locally best ROM best approximates the temporal variation of $S_i$obtained by the HFM. This can be clearly seen during the periods when the CO_2front arrives at Point **B**.



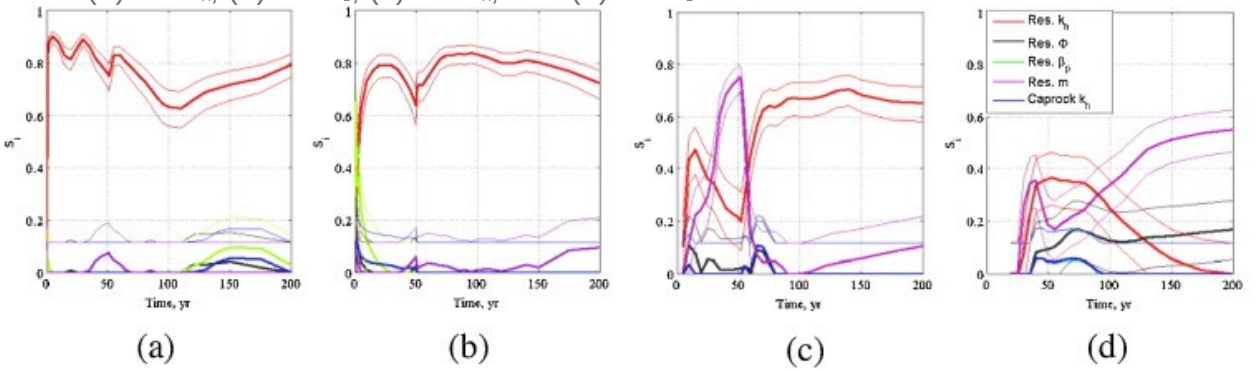
1. Download high-res image (384KB)
2. Download full-size image

Fig. 7. The Sobol' sensitivity index S_i calculated using the predictions from the weighted average ROM (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.



1. Download high-res image (397KB)
2. Download full-size image

Fig. 8. The Sobol' sensitivity index S_i calculated using the predictions from the best local ROM (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B.



1. Download high-res image (445KB)
2. Download full-size image

Fig. 9. The Sobol' sensitivity index S_i calculated using the HFM (Wainwright et al., 2013) (a) for P_A; (b) for P_B; (c) for S_A; and (d) for S_B. The thick lines are the estimated sensitivity lines, and the thin lines represent their confidence intervals.

In addition to sample-estimated values, aPC has the advantage that the analytical form of Sobol's indices can be obtained easily (Oladyshkin et al., 2012). We compare the Sobol's indices directly from the aPC analytical form and those from the analysis of aPC prediction results of 2100 (m = 300) realizations. They are similar (the same main influential parameters are identified) with negligible differences, suggesting that m = 300 is sufficient.

5. Summary and conclusion

Although the Sobol' indices of Wainwright et al. (2013) can be used to verify the proposed method in our example, such "true" results do not exist in real applications.

We have showed that ROMs with small RMSE will typically lead to better estimates of S_i, although the converse is not necessarily true. However, small RMSE for even a small validation sample set increases our confidence in the S_i computed by a ROM. Based on this observation and Fig. 2, it is clear that results from the locally best ROM (Fig. 8) are considered to be the best estimates. The locally best ROM also provides better estimates than each individual ROM, as well as the weighted average ROM. With the locally best ROM, the total computational effort is associated with running 21 + 72 + 100 = 193 HFM realizations, compared to 2100 HFM simulations needed by Wainwright et al. (2013); we achieve 90% reduction in computational cost.

This work was motivated by the difficulty in performing global sensitivity analyses using time-consuming HFM simulations, and the difficulty in evaluating ROM performance if it were used for such an analysis. We propose to use multiple ROMs to approximate the HFM simulation results while limiting the total number of HFM simulations needed to build those ROMs. We reach the following conclusions for the proposed method:

1. The computational effort is significantly reduced (in our case, it is reduced by more than 90%) while obtaining sufficiently accurate estimates of the Sobol sensitivity indices. The HFM simulations include two sets: one set for building ROMs, and one set for ROM validation. The choice that two of the four ROMs can be created using the same HFM simulations already available from the construction of the other two ROMs helps reduce the number of HFM simulations needed for multiple ROMs.

2. The use of a validation set provides (1) an evaluation of the performance of each ROM, and therefore a basis for developing a combined ROM that is either a weighted average of all four ROMs or the locally best ROM; and (2) an indication of the confidence one might have in the accuracy of the final estimates.

3. The efficiency of the combined ROM (either weighted average ROM or the locally best ROM) allows one to conduct a convergence study of the statistical estimates or global sensitivity analysis, which further increases our confidence in these estimates.

4.

Compared to the weighted average ROM, using the locally best ROM to combine ROMs has the advantage to weigh in the ROMs that perform well at some point in the parameter space, but screen out ROMs that have a large RMSE at this point.

5.

The proposed method provides a powerful tool in uncertainty quantifications for large models within a risk assessment framework.

Acknowledgments

The authors wish to acknowledge Haruko Wainwright for sharing her data and sensitivity analysis algorithm with us. This work was conducted as part of National Risk Assessment Partnership (NRAP) effort, supported by the Assistant Secretary for Fossil Energy, Office of Sequestration, Hydrogen, and Clean Coal Fuels, of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231.

References

Alis and Rabitz, 2001

O.F. Alis, H. Rabitz**Efficient implementation of high dimensional model representations**

J. Math. Chem., 29 (2) (2001), pp. 127-142

CrossRefView Record in Scopus

Bianchi et al., 2016

M. Bianchi, L. Zheng, J.T. Birkholzer**Combining multiple lower-fidelity models for emulating complex model responses for CCS environmental risk assessment**

Int. J. Greenh. Gas Control, 46 (2016), 10.1016/j.ijggc.2016.01.009

Birkholzer et al., 2011

J.T. Birkholzer, Q. Zhou, A. Cortis, S. Finsterle**A sensitivity study on regional pressure buildup from large-scale CO_2 storage projects**

Energy Procedia, 4 (2011), pp. 4371-4378

ArticleDownload PDFView Record in Scopus

Blatman and Sudret, 2010

G. Blatman, B. Sudret**Efficient computation of global sensitivity indices using sparse polynomial chaos expansions**

Reliab. Eng. Sys. Saf., 95 (2010), 10.1016/j.ress.2010.06.015

Glen and Isaacs, 2012

G. Glen, K. Isaacs**Estimating Sobol' sensitivity indices using correlations**

Environ. Modell. Software, 37 (2012), pp. 157-166, 10.1016/j.envsoft.2012.03.014

ArticleDownload PDFView Record in Scopus

Harp et al., 2016

D.R. Harp, R.J. Pawar, W.J. Carey, C.W. Gable**Reduced order models of transient CO_2 and brine leakage along abandoned wellbores from geologic carbon sequestration reservoirs**
Int. J. Greenh. Gas Control, 45 (2016), pp. 150-162, 10.1016/j.ijggc.2015.12.001
ArticleDownload PDFView Record in Scopus

Jordan et al., 2015

A.B. Jordan, P.H. Stauffer, D.R. Harp, W.J. Carey, R.J. Pawar**A response surface model to predict CO_2 and brine leakage along cemented wellbores**
Int. J. Greenh. Gas Control, 33 (2015), pp. 27-39, 10.1016/j.ijggc.2014.12.002
ArticleDownload PDFView Record in Scopus

Kaipio and Somersalo, 2004

J.P. Kaipio, E. Somersalo**Computational and statistical methods for inverse problems**
Applied Mathematical Sciences, Springer, Berlin (2004)

Li et al., 2001

G.Y. Li, C. Rosenthal, H. Rabitz**High dimensional model representations**
J. Phys. Chem. A, 105 (33) (2001), pp. 7765-7777
CrossRefView Record in Scopus

Li et al., 2002

G.Y. Li, S.W. Wang, H. Rabitz**Practical approaches to construct RS-HDMR component functions**
J. Phys. Chem. A, 106 (37) (2002), pp. 8721-8733
CrossRefView Record in Scopus

Liu, 2013

Y. Liu**Non-intrusive methods for probabilistic uncertainty quantification and global sensitivity analysis in nonlinear stochastic phenomena**
PhD Thesis, Florida State University, USA (2013)

Muller and Piche, 2011

J. Muller, R. Piche**Mixture surrogate models based on Dempster–Shafer theory for global optimization problems**
J. Global Optim., 51 (1) (2011), pp. 79-104
CrossRefView Record in Scopus

Oladyshkin and Nowak, 2012

S. Oladyshkin, W. Nowak**Date-driven uncertainty quantification using the arbitrary polynomial chaos expansion**
Reliab. Eng. Syst. Saf., 106 (2012), pp. 179-190
ArticleDownload PDFView Record in Scopus

Oladyshkin_et_al., 2011

S. Oladyshkin, H. Class, R. Helmig, W. Nowak**An integrative approach to robust design and probabilistic risk assessment for CO_2storage in geological formations**
Comput. Geosci. (2011), 10.1007/s10596-011-9224-8

Oladyshkin et al., 2012

S. Oladyshkin, F.P.J. de Barros, W. Nowak**Global sensitivity analysis: a flexible and efficient framework with an example from stochastic hydrogeology**
Adv. Water Resour., 37 (2012), pp. 10-22, 10.1016/j.advwatres.2011.11.001
ArticleDownload PDFView Record in Scopus

Pau et al., 2013a

G.S.H. Pau, Y. Zhang, S. Finsterle**Reduced order models for many-query subsurface flow applications**
Comput. Geosci. (2013), 10.1007/s10596-013-9349-z

Pau et al., 2013b

G.S.H. Pau, Y. Zhang, S. Finsterle, H. Wainwright, J. Birkholzer**Reduced order modeling in iTOUGH2**
Comput. Geosci. (2013), 10.1016/j.cageo.2013.08.008

Pau et al., 2014

G.S.H. Pau, G. Bisht, W.J. Riley**A reduced-order modeling approach to represent subgrid-scale hydrological dynamics for land-surface simulations: application in a polygonal tundra landscape**
Geosci. Model Dev., 7 (2014), pp. 2091-2105
CrossRefView Record in Scopus

Pruess, 2005

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

Rabitz et al., 199

H. Rabitz, O.F. Alis, J. Shorter, K. Shim**Efficient input–output model representations**
Comput. Phys. Commun., 117 (1–2) (1999), pp. 11-20
ArticleDownload PDFView Record in Scopus

Rasmussen and

C.E. Rasmussen, C.K.I. Williams**Gaussian Processes for Machine Learning**
The MIT Press (2006)
ISBN 0-262-18253-X
http://www.gaussianprocess.org/gpml/

Razavi et al., 20

S. Razavi, B.A. Tolson, D.H. Burn**Review of surrogate modeling in water resources**
Water Resour. Res., 48 (2012), 10.1029/2011WR011527
W07401

Regis and Shoer

R.G. Regis, C.A. Shoemaker**Local function approximation in evolutionary algorithms for the optimization of costly functions**
IEEE Trans. Evol. Comput., 8 (5) (2004)

Regis and Shoer

R.G. Regis, C.A. Shoemaker**Constrained global optimization of expensive black box functions using radial basis functions**

J. Global Optim., 31 (2005), pp. 153-171
CrossRefView Record in Scopus

Saltelli, 2002

A. Saltelli**Making best use of model evaluations to compute sensitivity indices**
Comput. Phys. Commun., 145 (2) (2002), pp. 280-297
ArticleDownload PDFView Record in Scopus

Saltelli et al., 200

A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo**European commission joint research centre of Ispra (I). Sensitivity analysis practices: strategies for model-based inference**
Reliab. Eng. Syst. Saf., 91 (10–11) (2006), pp. 1109-1125, 10.1016/j.ress.2005.11.014
ArticleDownload PDFView Record in Scopus

Sobol, 1967

I.M. Sobol**Distribution of points in a cube and approximate evaluation of integrals**
U.S.S.R Comput. Maths. Math. Phys., 7 (1967), pp. 86-112
ArticleDownload PDFView Record in Scopus

Sobol, 2001

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

Sobol et al., 200

I.M. Sobol, S. Tarantola, D. Gatelli, S.S. Kucherenko, W. Mauntz**Estimating the approximation error when fixing unessential factors in global sensitivity analysis**
Reliab. Eng. Syst. Saf., 92 (7) (2007), pp. 957-960
ArticleDownload PDFView Record in Scopus

Villadsen and Mi

J. Villadsen, M.L. Michelsen**Solution of Differential Equation Models by Polynomial Approximation**
Prentice-Hall (1978)

Wainwright et al.

H.M. Wainwright, S. Finsterle, Q. Zhou, J.T. Birkholzer**Modeling the performance of large-scale CO_2 storage systems: a comparison of different sensitivity analysis methods**
Int. J. Greenh. Gas Control, 17 (2013), pp. 189-205
ArticleDownload PDFView Record in Scopus

Wiener, 1938

N. Wiener**The homogeneous chaos**
Am. J. Math., 60 (1938), pp. 897-936
CrossRefView Record in Scopus

Willcox and Pera

K. Willcox, J. Peraire**Balance model reduction via the proper orthogonal decomposition**
AIAA J., 40 (11) (2002)

Xiu and Karniada

D. Xiu, G.E. Karniadakis**The Wiener–Askey polynomial chaos for stochastic differential equations**
SIAM J. Sci. Comput., 24 (2002), pp. 619-644
CrossRefView Record in Scopus

Zhang et al., 200

K. Zhang, Y.S. Wu, K. Pruess**User's Guide for TOUGH2-MP—A Massively Parallel Version of the TOUGH2Code, Report LBNL-315E**
Lawrence Berkeley National Laboratory, Berkeley, CA, USA (2008)

Zhang et al., 201

Y. Zhang, C.M. Oldenburg, L. Pan**Fast estimation of dense gas dispersion from multiple continuous CO_2 surface leakage sources for risk assessment**
Int. J. Greenh. Gas Control (2016)
in press

Zhou et al., 2011

Q. Zhou, J.T. Birkholzer, J.L. Wagoner**Modeling the potential impact of geologic carbon sequestration in the southern San Joaquin basin, California**
The Ninth Annual Carbon Capture & Sequestration, Pittsburgh, PA (2011)

Ziehn and Tomlin

T. Ziehn, A.S. Tomlin**Global sensitivity analysis of a 3-dimensional street canyon model—part I: the development of high dimensional model representations**
Atmos. Environ., 42 (8) (2008), pp. 1857-1873
ArticleDownload PDFView Record in Scopus