Journal of Petroleum Science and Engineering xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

Generalized sensitivity analysis study in basin and petroleum system modeling, case study on Piceance Basin, Colorado

Yao Tong[a,*], Tapan Mukerji[a,b]

[a] Department of Energy Resources Engineering, Stanford University, United states
[b] Department of Geophysics, Stanford University, United States

ARTICLE INFO

Keywords:
Sensitivity analysis
Basin modeling
Uncertainty
Distance-based sensitivity analysis
Sobol
Piceance Basin

ABSTRACT

Basin and petroleum system modeling is an interdisciplinary endeavor, and utilizes integrated data to study sedimentary basins. Numerical models are constructed and simulated to quantitatively model the geodynamic processes in sedimentary basins. Often, the modeling process covers large spatial and temporal intervals with many uncertain input parameters. These can be continuous parameters characterized by certain statistical distribution, spatially distributed variables, or discrete parameters such as geological scenarios. Identifying sensitivities from these complex input model parameters and recognizing key uncertainties are crucial and challenging for basin modeling development and applications. The major contribution of this work is to introduce and implement efficient and accurate sensitivity analysis approach for basin and petroleum system modeling discipline.

We investigated two types of sensitivity analysis methods and compared their performance for identifying the impact of uncertain parameters on both spatial and temporal model responses. The first approach utilizes the variance-based Sobol indices to quantify parameter sensitivities, while the second approach is a distance-based sensitivity analysis which utilizes the distance between model responses to determine sensitivities of different parameters. The sensitivity analysis approaches are illustrated through a basin modeling example involving the processes of sediment compaction, source rock maturation and hydrocarbon generation in the Piceance Basin, Colorado, US. Monte Carlo samples of the input uncertainties related to physical properties of the source rock, thermal boundary conditions, and geological setting scenarios are generated. Multiple basin models constructed using these uncertain input parameters are simulated across the geological time span, and time-varying model response (hydrocarbon generation from Lower Cretaceous to present-day) and spatial model response (pressure and porosity distribution at present-day) are obtained. Sensitive parameters that impact these spatio-temporal model responses are then analyzed.

Results show that the distance-based sensitivity analysis approach could achieve similar results with fewer model runs compared to the variance-based Sobol method. Model responses in spatial and temporal domain are impacted by different uncertain input parameters. Subtle relationship between input parameters and model response could also be identified. In particular an unexpected link between chemical kinetics and porosity versus depth behavior was uncovered. The knowledge obtained from sensitivity studies enhance the understanding of the complex geological processes and can benefit the modeling development and forecast capability. Though the sensitivity results are case-specific, the approach and workflow are generally applicable to other basins and earth sciences modeling studies.

1. Introduction

Basin and petroleum system modeling is an interdisciplinary endeavor, and utilizes integrated data to study sedimentary basins. Numerical models are constructed and simulated to quantitatively model the geodynamic processes in sedimentary basins. Often, the modeling process covers large spatial and temporal intervals with many uncertain input parameters. These can be continuous parameters characterized by certain statistical distributions, spatially distributed variables, or discrete parameters such as geological scenarios. Understanding how these complicating input parameters impact the output model responses are critical. Sensitivity analysis tools focus on the study of how the uncertainty in the output of a mathematical model or system can be apportioned to different sources of uncertainty in its inputs, and thus is a powerful tool for uncertainty quantification.

However, in the basin modeling discipline, few systematic sensitiv-

* Corresponding author.

http://dx.doi.org/10.1016/j.petrol.2016.11.019
Received 13 April 2016; Received in revised form 9 November 2016; Accepted 15 November 2016
Available online xxxx

ity analysis studies have been conducted to fully investigate the complex input uncertainties and the associated uncertainties' impact on model responses. Some studies utilized local sensitivity analysis tools such as simple experimental design, or vary-one-factor-a-time, which are not capable of uncovering complex relations between model input and output responses. Some studies utilized global sensitivity method but only narrowed its study focus on one particular model response and one aspect of the sedimentary basin history (Formaggia et al., 2013; Ruffo et al., 2014). Wendebourg (2003) used response surface and experimental design methods to study uncertainty of petroleum generation. Studies in hydrogeology (Wainwright et al., 2014) have conducted sophisticated sensitivity analysis on basin scale problems; however, the model responses of interest are quite different from that of basin and petroleum system modeling interests.

In this study, we aim to investigate some existing generalized sensitivity analysis methods. They are flexible enough to tackle various inputs as well as any model responses in basin and petroleum system modeling. We will apply two global sensitivity analysis methodologies and demonstrate their workflows through a one-dimensional basin model illustrative example. We will compare the performances of these methods and discuss the results and insights from the sensitivity analysis.

2. Methodology

We focus on the use of global sensitivity (as opposed to local or derivative-based sensitivity) which allows us to fully explore the relationship between input uncertainty and model output response uncertainty. Two types of Monte Carlo based global sensitivity analysis methods are investigated with the focus being to illustrate an application of the two established approaches for global sensitivity analysis. The first is the Sobol/Saltelli sensitivity analysis approach, which decomposes and analyzes the variances of the model output response and utilizes the Sobol indices to quantify a parameter's sensitivity. The second approach is a distance-based global sensitivity analysis method (Fenwick et al., 2014; Park et al., 2016) which originates from the regionalized sensitivity analysis approach (Spear and Hornberger, 1980; Spear et al., 1994). This method statistically classifies the model output response into a few clusters, and determines the sensitivities of input data by studying the statistics of parameter variations (e.g., the distance between parameter distributions) between each cluster. McKenna and Smith (2004) applied this method to investigate parameters in object-based simulations of a fluvial system for a ground water flow problem.

2.1. Sobol sensitivity analysis

The Sobol method (Saltelli et al., 2008; Sobol, 2001) is a variance-based sensitivity analysis approach. It decomposes the variance of a certain model output (response) into portions that are contributions from variation of different input parameters. In this way, the corresponding impact of the input data to output model response are quantified from the analysis of these variances. The detailed workflow and methodology are summarized in the section below.

Given any model that may be viewed as $Y = f(X)$, where X is a vector of k uncertain model inputs $\{X_1, X_2, X_3, \ldots, X_k\}$, and Y is a univariate model output, the variance of Y can be decomposed into terms attributable to each input parameter, as well as the interactions between the parameters. The Sobol index S_i (first-order sensitivity index) is defined as: $S_i = Var(E[Y|X_i])/Var(Y)$. It measures the effect of varying X_i alone and quantifies the contribution of each X_i to the total variability of Y. Another index S_{ti} (total-effect sensitivity index) is defined as: $S_{ti} = 1 - Var(E[YX_{\sim i}])/Var(Y)$, where the $X_{\sim i}$ notation indicates the set of all variables except X_i, and S_{ti} measures the total effect of X_i including all variance caused by interactions. Note that $S_{ti} = 1 - S_{\sim i}$ (Sudret, 2008) where $S_{\sim i}$ is the sum of sensitivities from all parameters except the parameter i. The original Sobol methods and the later extension studies (Glen and Isaacs, 2012; Saltelli et al., 2010, 2008; Sobol, 2001; Wainwright et al., 2014) provide different means of estimating S_i and S_{ti}. In this study, we adopt one of the most recent algorithm developed by Wainwright et al. (2014) to estimate Sobol Indices using the calculation shown below:

- Generate two sets of sample matrices A and B, each of which is an $n \times k$ matrix, where n is the number of Monte Carlo sample models, k is the number of uncertain parameters to be investigated;
- From A and B, create matrix $C_i (i = 1,2,\ldots,k)$ such that the i-th column equals to the i-th column of matrix A. The other columns of C_i are equal to matrix B;
- Simulate models with input parameter sets given by matrix A, B and C, this requires $n^*(k + 2)$ model runs;
- The simulation results of models from parameter sets A, B, C are n-dimensional vectors: $\{a_m\}$, $\{b_m\}$, $\{C_{i,m}\} (m = 1,2,\ldots,n)$
- S_i and S_{ti} can be calculated follow the equations developed in (Wainwright et al., 2014) using the following relationship:

$$S_i = 1 - \frac{1}{\sigma_y^2} \frac{1}{2(n-1)} \sum_{m=1}^{n} (a_m - c_{i,m})^2$$

$$S_{ti} = \frac{1}{\sigma_y^2} \frac{1}{2(n-1)} \sum_{m=1}^{n} (c_{i,m} - b_m)^2$$

Where σ_y is the variance of model response Y

2.2. Distance-based global sensitivity analysis

One limitation of the Sobol method is the expensive computational cost when the number of parameters increases. As shown in Section 2.1, with n sets of k-dimensional parameter vectors of Monte Carlo sampling, this method requires $n(k + 2)$ model runs. For every increase of one parameter, the total number of model runs increases by n. In order to achieve a better computational efficiency, an alternative approach originating from the regionalized sensitivity analysis (RSA) is investigated. The key concept of this method is to classify or cluster the output model responses and compare the statistical distribution of the input parameters between different class/cluster in order to determine the parameter sensitivities (Spear et al., 1994). This approach was extended by later authors (Bastidas et al., 1999; Fenwick et al., 2014; Pappenberger et al., 2008). In the most recent work from Fenwick et al. (2014), they conducted comprehensive investigation of sensitivities from different type of input data as well as their interactions, and show practical applications in petroleum reservoir modeling. The distance-based generalized sensitivity analysis (DGSA) method is capable of tackling various model responses as well. Unlike the Sobol method, in DGSA methods, one may choose univariate scalar or a vector of model output response and conduct the similar workflow to investigate sensitivities from different perspectives. Considering these advantages, we implement this methodology in this study following the workflow of Fenwick et al., 2014, summarized as below:

1. Generate input data sample matrices which is an $n \times k$ matrix, where n is the number of Monte Carlo samples and k is the number of uncertain parameters to be investigated;
2. Simulate models with input parameter sets given by the input data matrix;
3. Compute the pairwise distance matrix for each model's response; with n number of models, the distance matrix size is $n \times n$. In this study, Euclidean distance is employed to construct the distance matrix (Scheidt and Caers, 2009); the Euclidean distance between two vector valued model responses, $\mathbf{p}$=(p_1,p_2,...,p_k), and $\mathbf{q}$ =(q_1,q_2,