Journal of Hydrology 528 (2015) 286–300

Contents lists available at ScienceDirect

Journal of Hydrology

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

Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation

Heng Dai [1], Ming Ye *

Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, USA

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form 14 May 2015
Accepted 15 June 2015
Available online 21 June 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Wolfgang Nowak, Associate Editor

Keywords:
Model uncertainty
Model averaging
Scenario uncertainty
Scenario averaging
Variance decomposition
Sparse grid collocation

SUMMARY

Sensitivity analysis is a vital tool in hydrological modeling to identify influential parameters for inverse modeling and uncertainty analysis, and variance-based global sensitivity analysis has gained popularity. However, the conventional global sensitivity indices are defined with consideration of only parametric uncertainty. Based on a hierarchical structure of parameter, model, and scenario uncertainties and on recently developed techniques of model- and scenario-averaging, this study derives new global sensitivity indices for multiple models and multiple scenarios. To reduce computational cost of variance-based global sensitivity analysis, sparse grid collocation method is used to evaluate the mean and variance terms involved in the variance-based global sensitivity analysis. In a simple synthetic case of groundwater flow and reactive transport, it is demonstrated that the global sensitivity indices vary substantially between the four models and three scenarios. Not considering the model and scenario uncertainties, might result in biased identification of important model parameters. This problem is resolved by using the new indices defined for multiple models and/or multiple scenarios. This is particularly true when the sensitivity indices and model/scenario probabilities vary substantially. The sparse grid collocation method dramatically reduces the computational cost, in comparison with the popular quasi-random sampling method. The new framework of global sensitivity analysis is mathematically general, and can be applied to a wide range of hydrologic and environmental problems.

1. Introduction

Sensitivity analysis is a vital tool in hydrological modeling to identify influential parameters for inverse modeling and uncertainty analysis. There has been a growing trend of using global sensitivity analysis, which, in comparison with local sensitivity analysis, considers entire ranges of model parameters and takes into account the interactions between different parameters (van Griensven et al., 2006; Herman et al., 2013; Mishra et al., 2009; Nossent et al., 2011; Pan et al., 2011; Saltelli, 2000; Saltelli et al., 2010; Saltelli and Sobol, 1995; Shi et al., 2014; Song et al., 2015; Yang, 2011). Among various global sensitivity analysis methods, variance-based methods (Sobol', 1993; Saltelli et al., 1999) have gained popularity (Massmann and Holzmann, 2012; van Werkhoven et al., 2008; Wagener et al., 2009; Yang, 2011; Zhang et al., 2013a, 2013b). Different from screening methods (e.g., Morris methods, Morris, 1991), the variance-based methods provide not only ranking of parameter importance but also quantitative sensitivity measures such as global sensitivity indices for different parameters. The quantitative measures have been used recently for model structure diagnosis with respect to model complexity and model structure inadequacy (Rosolem et al., 2012; Gupta et al., 2012). For example, van Werkhoven et al. (2008) used Sobol's variance-based global sensitivity analysis to evaluate whether moderate model complexity is adequate for modeling multiple watersheds. Herman et al. (2013) further extended the global sensitivity analysis to three different models for understanding intermodel differences in dominant model parameters and/or components. These researches have shown promise in revealing contrasting controls across individual models.

This paper presents a new method that uses variance-based global sensitivity analysis beyond individual models but in the model averaging framework. New global sensitivity indices are derived for multiple models to identify influential parameters for not only individual models but also all models on average. This is necessary when model uncertainty exists, because parameter sensitivity varies (sometimes substantially) between models (Herman et al., 2013; Van Werkhoven et al., 2008) and identifying important parameters for a single model may be biased in that the important

* Corresponding author. Tel.: +1 850 644 4587; fax: +1 850 644 0098.
E-mail address: mye@fsu.edu (M. Ye).
[1] Present address: Pacific Northwest National Laboratory, Richland, WA 99352, USA.

http://dx.doi.org/10.1016/j.jhydrol.2015.06.034
0022-1694/© 2015 Elsevier B.V. All rights reserved.

parameters identified for a single model may not be important to the processes that the model intend to simulate. The bias in parameter identification can be reduced by the new global sensitivity indices based on model averaging for addressing model uncertainty. Model uncertainty is caused by multiple plausible interpretations of the system of interest based on available data and knowledge. For the hydrological modeling, model uncertainty is often inevitable, because the open and complex hydrologic systems can be conceptually interpreted and mathematically described in multiple ways (Beven, 2002, 2006; Bredehoeft, 2003, 2005; Neuman, 2003). Model averaging is a popular method for addressing and quantifying model uncertainty by evaluating the weighted average of the simulations of alternative models. The weights are measures of model plausibility, and can be estimated using various methods (Foglia et al., 2013; Lu et al., 2013; Tsai and Elshall, 2013; Ye et al. (2010b); Tsai and Li, 2008; Ye et al. (2008a); Meyer et al., 2007; Ajami et al., 2007; Poeter and Anderson, 2005; Ye et al., 2004; Winter and Nychka, 2010; Wohling and Vrugt, 2008). Schoniger et al. (2014) recently reported a comprehensive review and comparison of the methods used to estimate the Bayesian model averaging weights. Since our new sensitivity indices are based on the concept of model averaging (not specifically Bayesian model averaging), the indices can be evaluated by using the weights estimated via any methods. In addition, the weights can be estimated using expert judgments only or using both expert judgment and observations. The former is equivalent to prior weights and the latter to posterior weights, from a Bayesian viewpoint. In this study, only the prior weights are used, since evaluating posterior weights is beyond the scope of this work.

In addition to model uncertainty, the new method of variance-based global sensitivity analysis also considers scenario uncertainty by estimating the new global sensitivity indices for a set of scenarios in a scenario averaging framework. Scenario uncertainty is aleatory and an important source of predictive uncertainty. According to IPCC (2000, p.62), "scenarios are images of the future, or alternative futures. They are neither predictions nor forecasts. Rather, each scenario is one alternative image of how the future might unfold. A set of scenarios assists in the understanding of possible future developments of complex systems." Following Meyer et al. (2014), a scenario of hydrologic modeling is defined in this paper as a future state or condition assumed for a system, with the emphasis on those aspects of a scenario that affect the system hydrology. Scenario uncertainty here is similar to the input uncertainty used in surface hydrology (e.g., Kavetski et al., 2006; Vrugt et al., 2008; Renard et al., 2010), but with focus on future states and conditions. While it happens often that the same models are used for different scenarios, different models may be needed for different scenarios, when scenario uncertainty affects model formulation. On the other hand, for the same model, its plausibility may vary under different scenarios, because prior probability may change given that the model is conditioned on scenarios, although scenario uncertainty does not affect model calibration. Based on the concept that model uncertainty depends on scenario uncertainty, Draper et al. (1999) developed a scenario- and model-averaging method to first quantify model uncertainty and then scenario uncertainty. Similarly, Meyer et al. (2007, 2014) developed a hierarchical Bayesian framework to quantify parametric, model, and scenario uncertainty. The hierarchical framework was also used by Rojas et al. (2010) for quantifying scenario and model uncertainties for a system with multiple models and scenarios. It is for the first time that the hierarchical framework is used for global sensitivity analysis to define global sensitivity indices with consideration of multiple models and scenarios. Without loss of generality, the numerical example of this study uses the same models for different scenarios.

As discussed above, instead of considering only parametric uncertainty under a single model and a single scenario, our new variance-based global sensitivity considers the joint effect of parametric, model, and scenario uncertainties on model outputs. This is in a similar spirit of Baroni and Tarantola (2014), who developed a general probabilistic framework to considering all uncertainty sources in global sensitivity analysis. Their framework however does not specify the hierarchical structure from model scenarios to model structures and to model parameters. Since various frameworks of uncertainty quantification have been developed for different purposes of hydrologic modeling (Matott et al., 2009; Renard et al., 2010; Refsgaard et al., 2012; Tartakovsky, 2013), our method of global sensitivity analysis may be extended to the different frameworks to meet the various needs of global sensitivity analysis.

Another focus of this paper is to use computationally efficient methods for global sensitivity analysis, which is well known to be computationally expensive because it requires Monte Carlo (MC) simulation with a large number of model executions (Sobol', 1993; Jansen, 1999; Sobol' et al., 2007; Saltelli et al., 2010). Various methods have been developed to reduce the computational cost. A common practice is to first conduct a Morris analysis to screen out unimportant parameters so that global sensitivity analysis is only conducted for important parameters (e.g., Chu-Agor et al. 2011; Zhao et al., 2011). The quasi-random sampling method developed by Saltelli et al. (2010) is the most popular MC method because the number of model executions needed for Sobol' sensitivity analysis is dramatically reduced in comparison with conventional MC methods. Rakovec et al. (2014) developed a hybrid local–global sensitivity analysis method termed the Distributed Evaluation of Local Sensitivity Analysis (DELSA) to implement Sobol' sensitivity analysis. DELSA is computationally efficient, because it does not use MC methods but uses the results of local sensitivity analysis (i.e., the Jacobian matrix) to approximate the Sobol' variance terms. Another kind of widely used methods is meta-modeling to build cheap-to-compute surrogates or emulators of computationally expensive models so that performing a large number of model executions is computationally affordable (O'Hagan, 2006). The methods of developing surrogates for sensitivity analysis include Taylor series approximation (Hakami et al., 2003), response surface approximation (Helton and Davis, 2003), Fourier series (Saltelli et al., 1999) nonparametric regression (Helton, 1993; Storlie et al., 2009), Kriging (Kleijnen, 2009; Borgonovo et al., 2012; Lamoureux et al., 2014), Gauss process (Rasmussen and Williams, 2006), polynomial chaos expansion (Garcia-Cabrejo and Valocchi, 2014; Formaggia et al., 2013; Oladyshkin et al., 2012; Sudret, 2007), and sparse-grid collocation (Buzzard, 2012; Buzzard and Xiu, 2011). However, the meta-modeling methods may still need a relatively large number of model executions to develop accurate surrogates, and the surrogate development is not always straightforward due to model nonlinearity (Razavi et al., 2012; Zhang et al., 2013a, 2013b). Therefore, there are still urgent needs to develop new computationally efficient methods for performing global sensitivity analysis.

This paper presents a use of the computationally efficient method based on sparse grid collocation (SGC) techniques for variance-based global sensitivity analysis. The SGC techniques were developed for computing multidimensional integration (Smolyak, 1963), and have been shown to be an efficient and effective tool to overcome the curse of dimensionality for high dimensional numerical integration and interpolation (Barthelmann et al., 1999; Bungartz and Griebel, 2004; Gerstner and Griebel, 1998; Xiu and Hesthaven, 2005). The SGC techniques are particularly suitable for variance-based global sensitivity analysis, because the mean and variance needed for the sensitivity analysis are multivariate integrals. Based on quadrature rules (e.g., Gerstner and Griebel, 1998), SGC evaluates mean and variance of a quantity of interest at selected sparse grid points in parameter space. Since the number of