Environmental Modelling & Software 64 (2015) 230–239

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects

Y.P. Khare*, R. Muñoz-Carpena, R.W. Rooney, C.J. Martinez

Department of Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, PO Box 110570, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 29 January 2014
Received in revised form
7 November 2014
Accepted 14 November 2014
Available online 27 December 2014

Keywords:
Trajectory sampling
Elementary effect method
Sensitivity analysis
Parameter screening efficiency

ABSTRACT

Environmental models are inherently complex and often characterized by high dimensionality. The method of elementary effects (EE) is one of the most widely used parameter screening technique implemented to reduce burden on computational resources required for thorough model evaluation. Due to issues like inefficient screening and excessive sampling time, the development of more effective EE sampling strategies has been a recent research focus. This paper presents a new sampling strategy - Sampling for Uniformity (SU) – based on the principles of meeting close-to-theoretical parameter distributions and maximizing trajectory spread. The performance of the SU relative to existing strategies was evaluated using a number of criteria including generated parameter distributions' uniformity, time efficiency, trajectory spread, and screening efficiency. The SU performed better than some trajectory-based benchmark strategies across the evaluation criteria, underlining the effectiveness of multi-criteria based sampling and the need to focus future efforts on exploring other combinations of sampling criteria.

1. Introduction

Environmental models are increasingly used in a regulatory decision making framework to tackle a variety of complex problems (e.g. Bennett et al., 2013; Jakeman et al., 2006; Warmink et al., 2010). Both the European Commission (EC, 2009) and the United States Environmental Protection Agency (EPA, 2009) have emphasized the importance of risk assessment and have provided the guidelines on assessing and reducing the risks in environmental model applications. Sensitivity analysis is one of the fundamental tools used for model evaluation (Liu et al., 2008; Refsgaard et al., 2007; Saltelli et al., 2000, 2008). Today, several sensitivity analysis methods are available ranging from derivative-based local approaches to more rigorous variance-based global sensitivity analysis (GSA) methods such as the Fourier Amplitude Sensitivity Test (FAST) and the method of Sobol' (Cacuci and Ionescu-Bujor, 2004; Cukier et al., 1978; Helton and Davis, 2002; Saltelli et al., 2000, 2005; Sobol', 1993). The merits of GSA methods over local methods are well established (Saltelli et al., 2004, 2008). However, variance-based GSA methods become time consuming with an increase in the number of parameters to the point of computational infeasibility (Kucherenko et al., 2009; Saltelli et al., 2005). This is often the case with many environmental and ecological models as they are typically characterized by tens to hundreds of parameters where the consideration of spatial variability further increases the model dimensionality, ultimately requiring a large number of model runs and computational resources (Ciric et al., 2012; Foglia et al., 2009; Herman et al., 2013a,b; Jawitz et al., 2008; Makler-Pick et al., 2011; Moreau et al., 2013; van Griensven et al., 2006). Thus parameter screening, the initial separation of important model parameters from unimportant ones, becomes essential before applying rigorous variance-based GSA methods.

Various methods such as first and higher order derivatives, one at a time (OAT) approaches, the method of elementary effects (EEs) (Morris, 1991), the systematic fractional replicate design (Cotter, 1979), the iterated fractional factorial design (Andres and Hajas, 1993), and the sequential bifurcation design (Bettonvil, 1990; Bettonvil and Kleijnen, 1997) have been developed for low computational cost parameter screening exercises. Among these, the method of EEs (and its variants) is the most widely used screening method in environmental modeling studies and has been

* Corresponding author. Tel.: +1 352 392 1864x290; fax: +1 352 392 4092.
E-mail addresses: khareyogesh1@ufl.edu (Y.P. Khare), carpena@ufl.edu (R. Muñoz-Carpena), rrooney@ufl.edu (R.W. Rooney), chrisjm@ufl.edu (C.J. Martinez).

http://dx.doi.org/10.1016/j.envsoft.2014.11.013
1364-8152/

recommended as a part of a modern statistical framework for GSA (e.g. Cariboni et al., 2007; Campolongo et al., 1999; Muñoz-Carpena et al., 2007; Tong and Graziani, 2007; Yang et al., 2012; Zhan et al., 2013). However, the original method of EEs i.e. Morris (1991) method was found to have drawbacks regarding parameter sampling and the calculation of sensitivity measures sometimes leading to unreliable parameter screening (Campolongo et al., 2007).

Santner et al. (2003) categorized parameter sampling schemes for computer experiments into three types: (1) Latin Hypercube Sampling (LHS), (2) distance measure criteria, and (3) uniformity. Each type has advantages and disadvantages (e.g. Fang et al., 2000; Johnson et al., 1990; Santiago et al., 2012a,b; Wiens, 1991). However, a sampling strategy belonging to only one of the above three types is not totally satisfactory for sampling designs (Santner et al., 2003). They also suggested that care should be taken when combining design types to avoid computational infeasibility. It should be noted that most of the computer experiment sampling strategies generate parameter samples in a unit hyperspace which are then transformed to user-defined distributions.

Table 1 summarizes the original and recent developments in parameter sampling and sensitivity measures for the EEs method. Theoretical basis of these strategies vary considerably. Each successive work has sought to achieve a more effective sampling strategy, i.e. 'effective exploration' of the parameter hyperspace at 'low computational cost'. Low computational cost implies the selection of fewer trajectories for reliable screening as well as the time required for sample generation. Morris (1991) discussed advantages of trajectory based designs over LHS when one needs to restrict the number of model runs. The stratified sampling used by Morris (1991) from a definite number of levels (p) did not employ a distance measure or distribution criteria and can be regarded as a random sampling over the gridded parameter space. This sampling strategy by Morris will be referred as MM hereon in this article. While LHS may not be efficient, some of its variants have attractive properties such as orthogonality, which can be useful for screening methods (Santner et al., 2003). van Griensven et al. (2006) presented a LHS based sampling for the EE method while a simplex-based sampling strategy was introduced by Pujol (2009). The strategy of optimized trajectories (OT) (Campolongo et al., 2007) and the modified optimized trajectories (MOT) (Ruano et al., 2012) based their sampling on maximizing Euclidean distance (ED) between trajectories. Note that although ED is not the only possible distance criteria, it is the only one explored so far for sample optimization (Campolongo et al., 2007; Santner et al., 2003). The cell-based trajectory sampling (Saltelli et al., 2009) uses a rather complicated parameter sampling scheme. Given an individual input parameter X_i, EEs are calculated by tracking and combining model values for sample points which differ from each other in all parameter coordinates except the X_i^{th}. This method, originally designed for models with strong interactive effects, was found to be inferior to some other methods in identifying important parameters based on total sensitivity (Campolongo et al., 2011). The radial sampling strategy proposed by Campolongo et al. (2011) was designed as an extension of the basic OAT approach and was found to perform the best for most of the test functions in their study. Sobol' quasi-random sequences (Sobol', 1976) were used in radial sampling for the sample design. However, Sobol' sequences are inefficient when the number of parameters is high (Saltelli et al., 2000) and repeated use of the trajectory base points makes success of this strategy dependent on specific points (see Herman et al., 2013a).

It is noteworthy that none of the screening sampling strategies discussed above were based on the principle of reproducing uniform parameter distributions closely. A uniform parameter distribution is indicated in this context by the probability density function of an individual parameter values from all trajectories. However, only two of the above strategies (Campolongo et al., 2007; Ruano et al., 2012) discuss the importance of matching the distributions of generated parameter samples. One of the possible reasons that uniform distribution sampling has not been previously considered could be the associated run time efficiency, as mentioned in Santner el al. (2003). In addition, none of the previous studies, with the notable exception of Ruano et al. (2012), considered the issue of run time, which is an important feature from the application point of view. The MOT method was developed by Ruano et al. (2012) in response to the large computational time requirement for sample generation using the OT, especially for high dimensional models.

In spite of the availability of an array of sampling strategies for EE-based parameter screening, an efficient sampling strategy which is suitable across multiple criteria for all types of models still eludes the modeling community. The objective of this research was to develop a new screening sampling strategy by combining two parameter sampling criteria, so that it (i) can be used for a wide range of models, (ii) is fast enough from a practical point of view, and (iii) produces parameter distributions that closely resemble the

Table 1
Elementary effect methods and sampling strategies for parameter screening sensitivity analysis.

Sr.	Method and authors	Availability	Sensitivity measures[a]	Improvements
1	Method of elementary effects or Morris Method (MM) (Morris, 1991)	(1) As a part of SimLab v2.2 (2) Matlab (http://ipsc.jrc.ec.europa.eu/?id=756)	μ_i and σ_i	Sampling strategy[b], sensitivity measures
2	Method of Optimized Trajectories (OT) (Campolongo et al., 2007)	Matlab (http://ipsc.jrc.ec.europa.eu/?id=756)	μ_i^* and σ_i	Sampling strategy, sensitivity measures
3	Combined LHS and one at a time sampling (van Griensven et al., 2006)	NA[c]	μ_i^*[d]	Sampling strategy
4	Simplex based sampling (Pujol, 2009)	R (http://cran.r-project.org/web/packages/sensitivity/)	μ_i^* and σ_i	Sampling strategy
5	Cell based trajectory sampling (Saltelli et al., 2009)	NA[c]	μ_i^* and σ_i	Sampling strategy
6	Radial sampling (Campolongo et al., 2011)	NA	μ_i^* and σ_i	Sampling strategy
7	Modified Optimized Trajectory Method (MOT) (Ruano et al., 2012)	Matlab (personal communication with authors)	μ_i^* and σ_i	Sampling strategy
8	Sampling for Uniformity (SU) (This article)	Matlab (http://abe.ufl.edu/carpena/software/SUMorris.shtml)	μ_i^* and σ_i	Sampling strategy

[a] μ_i = mean of distribution of elementary effects (EE) for ith parameter, μ_i^*: mean of distribution of absolute values of elementary effects (EE) for ith parameter, σ_i: standard deviation of distribution of elementary effects (EE) for ith parameter.
[b] Original development.
[c] Not freely available.
[d] van Griensven et al. (2006) use sensitivity measure similar to μ_i^*.