Reliability Engineering and System Safety 107 (2012) 23–28

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Monte Carlo sensitivity analysis of an Eulerian large-scale air pollution model

I. Dimov, R. Georgieva *, Tz. Ostromsky

Department of Parallel Algorithms, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev 25A, 1113 Sofia, Bulgaria

ARTICLE INFO

Article history:
Received 11 December 2010
Received in revised form
22 March 2011
Accepted 28 June 2011
Available online 7 July 2011

Keywords:
Variance-based global sensitivity analysis
Multidimensional numerical integration
Monte Carlo and quasi-Monte Carlo algorithms
Air pollution modeling

ABSTRACT

Variance-based approaches for global sensitivity analysis have been applied and analyzed to study the sensitivity of air pollutant concentrations according to variations of rates of chemical reactions. The Unified Danish Eulerian Model has been used as a mathematical model simulating a remote transport of air pollutants. Various Monte Carlo algorithms for numerical integration have been applied to compute Sobol's global sensitivity indices. A newly developed Monte Carlo algorithm based on Sobol's quasi-random points MCA-MSS has been applied for numerical integration. It has been compared with some existing approaches, namely Sobol's $\Lambda\Pi_\tau$ sequences, an adaptive Monte Carlo algorithm, the plain Monte Carlo algorithm, as well as, eFAST and Sobol's sensitivity approaches both implemented in SIMLAB software. The analysis and numerical results show advantages of MCA-MSS for relatively small sensitivity indices in terms of accuracy and efficiency. Practical guidelines on the estimation of Sobol's global sensitivity indices in the presence of computational difficulties have been provided.

1. Introduction

Environmental security is a very important topic for the modern society. Development of reliable and sustainable mathematical models has a significant role at this area. Specification of the most influential factors (chemical rates, boundary conditions, emission levels) on model outputs using sensitivity analysis techniques already achieves valuable information for an improvement of the model and identification of parameters that must be studied more carefully. It will lead to an increase of reliability and robustness of predictions obtained by large-scale environmental and climate models. On the other hand, sensitivity analysis is a tool useful for all processes where it is important to know which input factors contribute most to output variability [1].

The aim of the present work is

- to study the sensitivity of the concentration levels of important pollutants (like ozone O_3) due to variation of chemical rates applying variance-based techniques for global sensitivity analysis and Monte Carlo approaches for numerical integration,
- to compare the newly developed Monte Carlo algorithm based on Sobol's quasi-random points MCA-MSS [2] with some existing approaches, namely
  - Sobol's $\Lambda\Pi_\tau$ sequences [3],
  - adaptive Monte Carlo algorithm developed in [4,5],
  - Sobol's sensitivity approach implemented in SIMLAB [6],
  - eFAST sensitivity approach carried out via SIMLAB [6], and
  - the plain Monte Carlo algorithm [3,4],
- to show the superior efficiency of the sensitivity analysis methods proposed by the authors in [5],
- to provide practical insights about the case study at hand, and
- to try to provide operation guidelines on the estimation of relatively *small* Sobol's indices in the presence of computational difficulties.

The input data for sensitivity analysis has been obtained during runs of a large-scale mathematical model for remote transport of air pollutants (*Uni*fied *D*anish *E*ulerian *M*odel, UNI-DEM,[1] [7]).

Among quantitative global sensitivity analysis methods, variance-based methods are the most often used [8]. The main idea of these methods is to evaluate how the variance of an input or a group of inputs contributes into the variance of model output. Two of the most often used variance-based methods have been applied—Sobol's approach and *F*ourier *a*mplitude *s*ensitivity *t*est (FAST). The approaches have been implemented using a *M*onte *C*arlo *a*lgorithms (MCA) or SIMLAB software tool for global sensitivity analysis [6]. The results described here can be used for increasing the reliability of the mathematical model results, and identifying input parameters that should be measured more precisely.

* Corresponding author.
E-mail addresses: ivdimov@bas.bg (I. Dimov), rayna@parallel.bas.bg (R. Georgieva), ceco@parallel.bas.bg (Tz. Ostromsky).
URLS: http://parallel.bas.bg/dpa/BG/dimov/index.html (I. Dimov), http://parallel.bas.bg/~rayna/ (R. Georgieva), http://parallel.bas.bg/~ceco/ (Tz. Ostromsky).

[1] UNI-DEM has been developed at the Danish National Environmental Research Institute (http://www2.dmu.dk/AtmosphericEnvironment/DEM/).

0951-8320/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2011.06.007

2. Mathematical background

Consider a scalar model output $u=f(x)$ corresponding to a number of non-correlated model parameters $x=(x_1,x_2,\ldots,x_d)\in U^d\equiv[0;1]^d$ with a joint probability density function $p(x)=p(x_1,\ldots,x_d)$.

The issue about parameter importance can be studied via numerical integration in the terms of *an*alysis *o*f *va*riance (ANOVA). A number of unbiased Monte Carlo estimators for global sensitivity indices have been developed applying ANOVA-decomposition of model function [9–12].

2.1. Sobol's approach

In Sobol's approach [11] the variance of the square integrable model function is decomposed into terms of increasing dimension:

$$f(x)=f_0+\sum_{\nu=1}^{d}\sum_{l_1<\cdots<l_\nu} f_{l_1\ldots l_\nu}(x_{l_1},x_{l_2},\ldots,x_{l_\nu}),\quad f_0=\int f(x)\,dx, \tag{1}$$

where f_0 is a constant. The representation (1) is unique (called ANOVA—representation of the model function $f(x)$ [11]) if $\int_0^1 f_{l_1\ldots l_\nu}(x_{l_1},x_{l_2},\ldots,x_{l_\nu})\,dx_{l_k}=0$, $1\le k\le\nu$, $\nu=1,\ldots,d$. The quantities $\mathbf{D}=\int_{U^d}f^2(x)\,dx-f_0^2$, $\mathbf{D}_{l_1\ \ldots\ l_\nu}=\int f^2_{l_1\ \ldots\ l_\nu}\,dx_{l_1}\ldots dx_{l_\nu}$ are called variances (total and partial variances, respectively). Therefore, the total variance of the model output is partitioned into partial variances [13] in the analogous way as the model function, that is the ANOVA-decomposition: $\mathbf{D}=\sum_{\nu=1}^{d}\sum_{l_1<\cdots<l_\nu}\mathbf{D}_{l_1\ldots l_\nu}$.

Based on the above assumptions about the model function and the output variance, the following quantities:

$$S_{l_1\ \ldots\ l_\nu}=\frac{\mathbf{D}_{l_1\ \ldots\ l_\nu}}{\mathbf{D}},\quad \nu\in\{1,\ldots,d\} \tag{2}$$

are called Sobol's global sensitivity indices [11,13].

The *t*otal *s*ensitivity *i*ndex (TSI) of input parameter x_i, $i\in\{1,\ldots,d\}$, is defined in the following way [13]: $S_i^{tot}=S_i+\sum_{l_1\ne i}S_{il_1}+\sum_{l_1,l_2\ne i,l_1<l_2}S_{il_1l_2}+\cdots+S_{il_1\ldots l_{d-1}}$, where S_i is called *the main effect (first-order sensitivity index)* of x_i and $S_{il_1\ldots l_{j-1}}$ is the jth-order sensitivity index (respectively *two-way interactions* for $j=2$, and so on) for parameter x_i $(2\le j\le d)$. The mathematical treatment of the problem of providing global sensitivity analysis consists in evaluating total sensitivity indices and in particular Sobol's global sensitivity indices (2) of corresponding order that leads to computing of multidimensional integrals I_N.

A numerical approach for evaluating *small* sensitivity indices that combines reduction of the mean value and correlated sampling suggested in [12] has been applied. Analyzing the structure of the error of computing *small* sensitivity indices one can see that if $\mathbf{D}_{l_1\ldots l_\nu}\ll f_0^2$, then the method is spoilt by loss of accuracy. In a non-linear large-scale model like UNI-DEM it happens that even small by value indices may be important to be estimated. That is why we pay a special attention to *small* sensitivity indices.

2.2. Fourier amplitude sensitivity test

The method is based on Fourier transformation of uncertain model input parameters into a frequency domain [14]. It converts a multidimensional integral over all the uncertain parameters to a one-dimensional integral defined by a set of parametric equations, i.e. the idea is to explore the input space along a curve.

An *e*xtension of the *FAST* (or eFAST) method [15] was developed to estimate the total effects. Decomposition of variance in eFAST works by varying different parameters at different frequencies, encoding the identity of parameters in the frequency of their variation. Fourier analysis then measures the strength of each parameter frequency in the model output.

2.3. Adaptive Monte Carlo algorithm

The aim of this algorithm is to increase the reliability for evaluating Sobol's global sensitivity indices by reducing the variance of the corresponding Monte Carlo estimator. In particular, a technique for numerical integration of functions with local computational difficulties has been applied. This is important since the integrand has *computational irregularities* in a number of practical applications. There are various adaptive Monte Carlo algorithms depending on the strategy of adaptation [4,16,17]. The adaptive Monte Carlo algorithm applied here uses a posteriori information about the variance on the corresponding subdomains. A schematic description of the algorithm is presented in Fig. 1. The idea of the algorithm consists of the following [5]: the domain of integration is separated initially into subdomains with identical volume. The subdomain with the largest standard

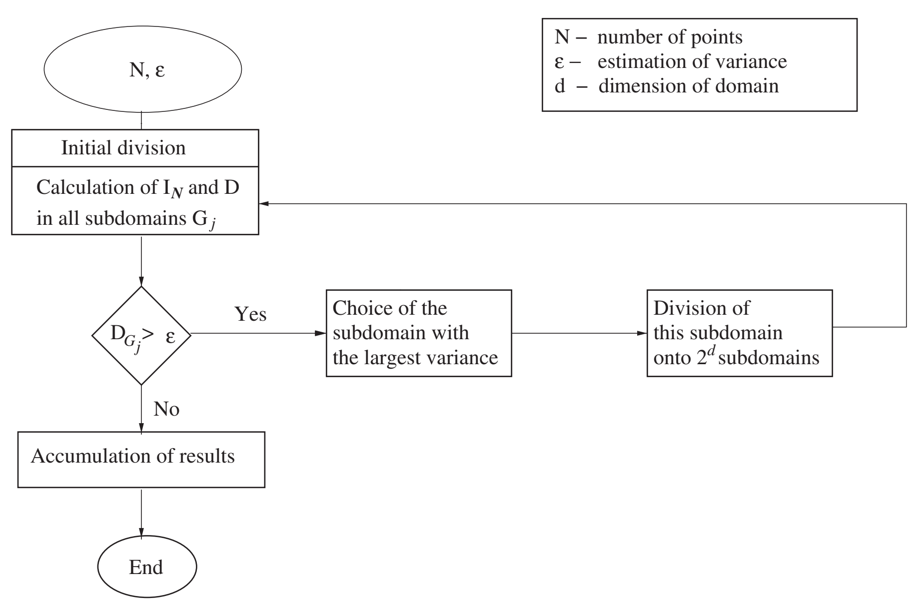

Fig. 1. The concept of the adaptive Monte Carlo algorithm.