Computers and Mathematics with Applications 70 (2015) 2773–2782

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

Preparing input data for sensitivity analysis of an air pollution model by using high-performance supercomputers and algorithms

Tzvetan Ostromsky [a,*], Ivan Dimov [a], Vassil Alexandrov [b,c], Zahari Zlatev [d]

[a] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev, bl. 25A, 1113 Sofia, Bulgaria

[b] ICREA - Barcelona Supercomputing Centre (BSC-CNS), Carrer Jordi Girona 29, E-08034 Barcelona, Spain

[c] Computational Science, ITESM Monterrey Tech, Mexico

[d] Department of Environmental Science, University of Århus, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark

ARTICLE INFO

Article history:
Available online 19 October 2015

Keywords:
Sensitivity analysis
Air pollution model
Parallel algorithm
Performance
Speed-up
Supercomputer

ABSTRACT

Sensitivity analysis is an important issue in large-scale mathematical modelling. We developed a novel 3-stage method for global sensitivity analysis of the Unified Danish Eulerian Model (UNI-DEM). This is a powerful large-scale air pollution model with an up-to-date high-performance software implementation. There is a number of uncertain internal parameters, especially in the chemistry–emission submodel, which are subject to our quantitative sensitivity study. Efficient Monte Carlo and quasi-Monte Carlo algorithms based on Sobol sequences are used in this study.

A large number of numerical experiments with some special modifications of the model must be carried out in order to collect the necessary input data for the particular sensitivity study. For this purpose we created an efficient high performance implementation SA-DEM, based on the MPI version of the package UNI-DEM. A vast number of numerical experiments were carried out with SA-DEM on an IBM Blue Gene/P, the most powerful parallel supercomputer, at the time of the write-up of this paper, in Bulgaria. Even this powerful machine has some problems with the storage when SA-DEM is to be run with the refined (480×480) version of the mesh. The code was implemented with some enhancements on the IBM MareNostrum III at BSC — Barcelona, the most powerful parallel supercomputer in Spain. This implementation appears to be quite efficient for that challenging computational problem, as our experiments show. Some numerical results and performance analysis of these results will be presented in the paper.

1. Introduction

In a popular definition (due to A. Saltelli [1]), **s**ensitivity **a**nalysis (SA) is the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input. The uncertainties in the model input can be due to various reasons: inaccurate measurements or calculation, approximation, data compression, etc. In order to measure

* Corresponding author.
E-mail addresses: ceco@parallel.bas.bg (T. Ostromsky), ivdimov@bas.bg (I. Dimov), vassil.alexandrov@bsc.es (V. Alexandrov), zz@dmu.dk (Z. Zlatev).
URLs: http://www.parallel.bas.bg/~ceco (T. Ostromsky), http://www.dmu.dk/AtmosphericEnvironment (Z. Zlatev).

http://dx.doi.org/10.1016/j.camwa.2015.07.020
0898-1221/

Table 1
Computational cost of variance-based methods for SA.

Method	Cost (Model runs)	Sensitivities
FAST (1973)	$O(d^2)$	$S_i, \forall i$
Sobol (1993)	$N(2d+2)$	$S_i, TSI(x_i), \forall i$
EFAST (1999)	dN	$S_i, TSI(x_i), \forall i$
Saltelli (2002)	$N(d+2)$	$S_i, \forall i, S_{ij}, \forall l, j, l \neq i$

the specific contribution of the uncertainty in each input parameter, considered to be a potential source, the **s**ensitivity **i**ndices (SIs) have been introduced. Two kinds of sensitivity analysis have been discussed in the literature: local and global. Local SA studies how much some small variations of inputs around a given value can change the value of the output. Global SA takes into account the whole domain of variation changes in the set of input parameters, and apportions the output uncertainty to the uncertainty in the input data.

Our primary motivation in this work is to search for efficient numerical algorithms for computing the global sensitivity indices of certain large-scale mathematical models like the **Uni**fied **D**anish **E**ulerian **M**odel (UNI-DEM) [2]. Such large-scale models are often described by systems of partial differential equations (the number of equations being equal to the number of chemical species studied by the model). It is not uncommon that each of these systems contains several millions of equations. This means that the computational tasks arising in the treatment of large-scale air pollution models are enormous, and great difficulties arise even when modern high-performance computers are used. Therefore, it is highly desirable to simplify as much as possible the model. A careful sensitivity analysis is needed in order to decide where and how simplifications can be made. On the other hand, air pollution modellers might ask the extent to which their results depend on assumptions of initial conditions, boundary conditions, or chemical reaction rate constants. This analysis can give valuable information about how precise the model output is and identify which variables should be investigated more closely if the uncertainty is unacceptably high. The goals may be to rank the importance of input variables, improve precision, screening, and decision making.

There are several alternative sensitivity analysis techniques. Generally, some methods for providing SA have been developed under special assumptions on the behaviour of the model (such as linearity, monotonicity and additivity of the relationship between input factor and model output).

In the local approaches the sensitivity of each unknown input value is computed by keeping the other parameter fixed and only varying the certain input parameter in a local area around its nominal value. These can be interpreted as **o**ne-**a**t-a-**t**ime experiments (OAT). One drawback of OAT techniques is that it is not possible to compute the effects which are caused by the interactions between the unknown input parameters.

Screening methods [3] are meant to deal with models containing hundreds of input factors, or with very computationally expensive models. They are economical from a computational point of view, but as a drawback, they tend to provide qualitative sensitivity measures, i.e. they rank the input factors in order of importance, but do not quantify how much a given factor is more important than another.

Among quantitative methods, variance-based methods are the most often used [4]. 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 an output variable. Variance-based methods deliver global, quantitative and model-independent sensitivity measures. A general sensitivity concept, namely the variance-based sensitivity analysis using a Monte Carlo technique, has been used in [5]. This concept is sampling-based, that is why a Monte Carlo simulation is applied. The techniques based on Monte-Carlo methods require a lot of simulations. The uncertain input parameters are modelled by random variables and characterized by their probabilistic density functions. The variance-based analysis focuses on the following questions: *"Which of the input variables variances influences the model output variance at most?"* and *"Which of the input variables has to be known more accurately to reduce the output variance?"*

Two alternative ways to calculate sensitivity indices are currently adopted: the method proposed by Sobol [6], and the extended version of the **F**ourier **A**mplitude **S**ensitivity **T**est (FAST) [7]. These methods do not need model to be linear and additive. In the FAST method the variance of the model output (a d-dimensional integral) is re-written as a one-dimensional integral with respect to a scalar variable s, by transforming each input variable x_i to be of the form $x_i = G_i(sin(\omega_i s))$, for an appropriate set of transformations G_i and integer frequencies ω_i (for more detail see [7]).

The Sobol indices are superior with respect to FAST in computation of the higher interaction terms (which is done in a way similar to the computation of the main effects). In the Sobol measure each effect (main or otherwise) is computed by evaluating just one multidimensional integral, which means that the total sensitivity index can be computed with just one Monte Carlo integral per factor.

Table 1 presents the computational cost of the most widely used variance-based methods: FAST (Fourier Amplitude Sensitivity Test) [8], the Sobol method [6], Extended FAST (EFAST) [7], as well as Saltelli's improvement of the Sobol method [9].

This work is aimed to evaluate Sobol SIs, which is connected with multiple numerical integration. While the classical deterministic grid methods are efficient for low dimensional integrands [10], their computational price grows fast and they become impractical for high dimensions d because the number of required integrand evaluations grows exponentially with d. In contrast, the convergence rate of the plain **M**onte **C**arlo (MC) integration method [11] does not depend on the number of dimensions d. That fact makes the Monte Carlo method a very useful tool in sensitivity analysis of large-scale models.