Accepted Manuscript

Studying the sensitivity of the pollutants concentrations caused by variations of chemical rates

I. Dimov, R. Georgieva, S. Ivanovska, Tz. Ostromsky, Z. Zlatev

PII: S0377-0427(10)00331-6
DOI: 10.1016/j.cam.2010.05.041
Reference: CAM 7935

To appear in: *Journal of Computational and Applied Mathematics*

Received date: 30 October 2009
Revised date: 21 May 2010

Please cite this article as: I. Dimov, R. Georgieva, S. Ivanovska, T. Ostromsky, Z. Zlatev, Studying the sensitivity of the pollutants concentrations caused by variations of chemical rates, *Journal of Computational and Applied Mathematics* (2010), doi:10.1016/j.cam.2010.05.041

Studying the Sensitivity of the Pollutants Concentrations Caused by Variations of Chemical Rates

I. Dimov [a,b], R. Georgieva [a], S. Ivanovska [a], Tz. Ostromsky [a], Z. Zlatev [c]

[a] *Department of Parallel Algorithms, Institute for Parallel Processing, Bulgarian Academy of Sciences, Acad. G. Bonchev 25 A, 1113 Sofia, Bulgaria*

[b] *Centre for Advanced Computing and Emerging Technologies, School of Systems Engineering, The University of Reading, Whiteknights, P.O. Box 225, Reading, RG6 6AY, UK*

[c] *National Environmental Research Institute, Aarhus University, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark*

Abstract

A systematic procedure for sensitivity analysis of a case study in the area of air pollution modeling has been performed. Contemporary mathematical models should include a large set of chemical and photochemical reactions to be established as a reliable simulation tool. The Unified Danish Eulerian Model is in the focus of our investigation as one of the most advanced large-scale mathematical models that describes adequately all physical and chemical processes.

Variance-based methods are one of the most often used approaches for providing sensitivity analysis. To measure the extent of influence of the variation of the chemical rate constants in the mathematical model over the pollutants concentrations the Sobol' global sensitivity indices are estimated using efficient techniques for small sensitivity indices to avoid a loss of accuracy. Studying of relationships between input parameters and model output as well as internal mechanisms is very useful for a verification and an improvement of the model and also for development of monitoring and control strategies of harmful emissions, for a reliable prediction of the final output of scenarios when the concentration levels of pollutants are exceeded. The proposed procedure can also be applied when other large-scale mathematical models are used.

Key words: Variance-based sensitivity analysis, Sobol' bal sensitivity indices, Air pollution modeling, Multidimensional numerical integration

1 Introduction

Environmental security is rapidly becoming a significant topic of present interest all over the world. It is necessary to carry out many comprehensive scientific studies and to analyze carefully the most important physical and chemical processes during the transport and transformations of air pollutants. An effective performance of such complicated procedures requires a joined research and collaboration between experts in the field of environmental modeling, numerical analysis and scientific computing.

The aim of the present work is to propose a new mechanism for investigation the sensitivity of the calculated concentration levels of important pollutants (like nitrogen dioxide NO_2 and especially ozone O_3) due to variation of rates of the involved chemical reactions in a real-life scenario of air pollution transport over Europe with the Unified Danish Eulerian Model (UNI-DEM).

In investigation of various highly complex engineering, physical, environmental, social, and economic systems it is important to measure relations that describe the effect on the output results when the conditions for the input change.

Sensitivity **a**nalysis (SA) is the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input [21]. Two classes in sensitivity analysis are considered in the existing literature: local SA and global SA. Local SA studies how some small variations of inputs around a given value change the value of the output. Global SA takes into account all the variation range of the inputs, and apportions the output uncertainty to the uncertainty in the input factors.

Several sensitivity analysis techniques are available [21]. Most existing methods for providing SA rely heavily on special assumptions connected to the behavior of the model (such as linearity, monotonicity and additivity of the relationship between input factor and model output). Among quantitative methods, variance-based methods are the most often used [20]. 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.

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. That is why, it is highly desirable to simplify as much as possible the model keeping the needed level of reliability of models' results. A careful sensitivity analysis is needed in order to decide where and how simplifications can be made. On the other hand, it is important to analyze the influence of variations of the initial conditions, the boundary conditions and/or the chemical rates on the model results in order to make right

assumptions about the simplifications which have to be implemented. Such an analysis can give valuable information about the performance of reliable and reasonable simplifications or to identify parameters and mechanisms the accuracy of which should be improved, because the model results are very sensitive to variations of these parameters and mechanisms. Thus, the goal could be

- improving the model,
- increasing the reliability of the results, and
- identifying processes that must be studied more carefully.

The rest of the paper is organised as follows. A description of the used mathematical model is given in Section 2, Subsection 2.1. Here we also describe the approach for evaluating Sobol' global **s**ensitivity **i**ndices (SI) in Subsection 2.3. Subsection 2.4 contains a brief review of two Monte Carlo approaches for small sensitivity indices. Section 3 presents a case study, the proposed scheme for providing sensitivity analysis and some results from numerical experiments. Section 4 contains a discussion about the obtained numerical results. Some concluding remarks are given in Section 5.

2 Mathematical Background

2.1 The Mathematical Model - Unified Danish Eulerian Model

According to the definition, given in [21], sensitivity analysis involves models, model inputs and model outputs.

The focus of our study is in the area of environmental security (air pollution transfer). Contemporary mathematical models of air pollution transport should include a fairly large set of chemical and photochemical reactions to be established as a reliable simulation tool [32]. The investigations and the numerical results that are reported in this paper have been done by using a large-scale mathematical model called the Unified Danish Eulerian Model [6,29,31,33].

The Unified Danish Eulerian Model simulates the transport of air pollutants. It has been developed at the Danish National Environmental Research Institute (http://www2.dmu.dk/AtmosphericEnvironment/DEM/). The space domain of the model contains the whole of Europe, the Mediterranean as well as parts of Asia and Africa. The model gives the possibility to study concentration variations in time of a high number of air pollutants, which is important for environmental protection, agriculture, health care. The mathematical model

takes into account the major physical processes - advection, diffusion, deposition, emissions, and chemical reactions. It must be emphasized that the main types of pollutants - sulphur pollutants, nitrogen pollutants, ammonia-ammonium, ozone, radicals and hydrocarbons can be studied by this model.

UNI-DEM is described mathematically [6,29,31] by the following system of partial differential equations:

$$\begin{aligned}\frac{\partial c_s}{\partial t} = &-\frac{\partial(uc_s)}{\partial x} - \frac{\partial(vc_s)}{\partial y} - \frac{\partial(wc_s)}{\partial z} + \\ &+\frac{\partial}{\partial x}\left(K_x\frac{\partial c_s}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial c_s}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z\frac{\partial c_s}{\partial z}\right) + \\ &+E_s + Q_s(c_1, c_2, \ldots, c_q) - (k_{1s} + k_{2s})c_s, \quad s = 1, 2, \ldots, q.\end{aligned}$$

The number q of equations in this system is equal to the number of chemical species that are studied by the model. The other quantities involved in the model are described below: c_s - concentrations of the chemical species; u, v, w - components of the wind along the coordinate axes; K_x, K_y, K_z - diffusion coefficients; E_s - emissions in the space domain; k_{1s}, k_{2s} - coefficients of dry and wet deposition respectively ($s = 1, \ldots, q$); $Q_s(c_1, c_2, \ldots, c_q)$ - non-linear functions that describe the chemical reactions between species.

Chemical reactions play a significant role in the model. The equations in the model are coupled through the chemical reactions. Moreover, both non-linearity and stiffness of the equations are mainly introduced by the chemistry (see [32]). Thus, the motivation to choose UNI-DEM is that it is one of the models of atmospheric chemistry, where the chemical processes are taken into account in a very accurate way. The chemical scheme used in the model is the well-known condensed CBM-IV (Carbon Bond Mechanism; the scheme was proposed in [8], but some enhancements have been obtained in [29] by adding several reactions for handling the ammonia-ammonium transformations in the atmosphere). It includes 35 pollutants and 71 chemical reactions. The scheme is suitable and adequate to study cases of high concentrations of chemical species. The space domain is discretized in a grid with (96×96) nodes in the two-dimensional version and $(96 \times 96 \times 10)$ nodes in the three-dimensional version of the model. The step of the discretization in the horizontal direction is 50 km and ten non-uniform layers are used in the vertical direction (the height of the layers is gradually increased, being smallest close to the surface).

The evaluation of model results against observations has been done in connection with some practical air pollution studies in various regions of Europe. The results have been presented for Bulgaria [35,36], Denmark [29,34], England [1], Europe [2,4,29–31], Hungary [11,12] and North Sea [9]. UNI-DEM has also been used in some inter-comparisons of European large-scale air pollution models [10,18].

Our main goal is to find out how variations of the input parameters of the model influence the model output. In our studies the chemical rate constants are considered as input parameters and the concentrations of pollutants are output parameters. In this paper, the term "constants" means variables with normal distribution (established experimentally) with mean 1.0.

2.2 Global Sensitivity Indices Concept

When the sensitivity of the concentrations calculated by UNI-DEM (or any other deterministic mathematical model) is studied, it is convenient to introduce some stochastic variables and equations.

It is assumed that the mathematical model can be presented as a model function

$$\mathrm{u} = f(\mathrm{x}), \quad \text{where} \quad \mathrm{x} = (x_1, x_2, \ldots, x_d) \in U^d \equiv [0;1]^d \tag{1}$$

is a vector of input parameters with a joint **p**robability **d**ensity **f**unction (p.d.f.) $p(\mathrm{x}) = p(x_1, \ldots, x_d)$. In general, real problems are characterized by multiple outputs. Here it is assumed that a scalar output is given. It is also assumed that input variables are independent (non-correlated input variables) and the density function $p(\mathrm{x}) = p(x_1, x_2, \ldots, x_d)$ is known, even if x_i are not actually random variables. This implies that the output u is also a random variable, as it is a function of the random vector x, with its own p.d.f.

It is reasonable to introduce an indicator that measures the importance of the influence of a given input parameter onto the output. The main indicator referred to a given input parameter $x_i,\ i = 1, \ldots, d$ (normalised between 0 and 1) is defined as

$$\frac{\mathbf{D}[\mathbf{E}(\mathrm{u}|x_i)]}{\mathbf{D}_\mathrm{u}}, \tag{2}$$

where $\mathbf{D}[\mathbf{E}(\mathrm{u}|x_i)]$ is the variance of the conditional expectation of u with respect to x_i and $\mathbf{D}_\mathrm{u}$ is the total variance according to u. This indicator is named *first-order sensitivity index* by Sobol' [24] or *correlation ratio* by McKay [16]. A brief review of measures of importance used in variance-based methods for sensitivity analysis is given in [5].

The *total sensitivity index* [14] provides a measure of the total effect of a given parameter, including all the possible joint terms between that parameter and all the others. The **t**otal **s**ensitivity **i**ndex (TSI) of input parameter $x_i,\ \ i \in$

$\{1, \ldots, d\}$ is defined in the following way [14,24]:

$$S_{x_i}^{tot} = S_i + \sum_{l_1 \neq i} S_{il_1} + \sum_{l_1, l_2 \neq i, l_1 < l_2} S_{il_1 l_2} + \ldots + S_{il_1 \ldots l_{d-1}}, \tag{3}$$

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 $j^{\text{-th}}$ order sensitivity index (respectively *two-way interactions* for $j = 2$, *three-way interactions* for $j = 3$ and so on) for parameter x_i ($2 \leq j \leq d$). The higher-order terms describe the interaction effects between the unknown input parameters $x_{i_1}, \ldots, x_{i_\nu}$, $\nu \in \{2, \ldots, d\}$ on the output variance. Usually for practical computations the set of input parameters is classified according their TSI [5]: *very important* if $0.8 < S_{x_i}^{tot}$, *important* if $0.5 < S_{x_i}^{tot} < 0.8$, *unimportant* if $0.3 < S_{x_i}^{tot} < 0.5$, and *irrelevant* if $S_{x_i}^{tot} < 0.3$. In subsection 2.3 we will show how sensitivity indices $S_{l_1 \ \ldots \ l_\nu}$ are defined via the variances of conditional expectations $\mathbf{D}_{l_1} = \mathbf{D}[f_{l_1}(x_{l_1})] = \mathbf{D}[\mathbf{E}(\mathrm{u}|x_{l_1})]$, $\mathbf{D}_{l_1 \ \ldots \ l_\nu}, 2 \leq \nu \leq d$ (see, equation (8)). It is often reasonable to assume (see [15,17]) that relatively small subsets of input variables in high-dimensional models have the main impact on the output. The high dimensional sums can be neglected when many practical problems are studied. This means that one can use low-order indices preferably, but should be able to control the contribution of higher-order terms.

2.3 The Sobol' Approach

The Sobol' method is one of the most often used variance-based methods. To our best knowledge the Sobol' sensitivity measure [24] was first published in [23]. An important advantage of this method is that it allows to compute not only the first-order indices, but also indices of a higher-order in a way similar to the computation of the main effects. The total sensitivity index can be calculated with just one Monte Carlo integral per factor.

The method for global SA applied here is based on a decomposition of an integrable model function f in the d-dimensional factor space into terms of increasing dimensionality:

$$f(\mathrm{x}) = f_0 + \sum_{\nu=1}^{d} \sum_{l_1 < \ldots < l_\nu} f_{l_1 \ldots l_\nu}(x_{l_1}, x_{l_2}, \ldots, x_{l_\nu}), \tag{4}$$

where f_0 is a constant. The total number of summands in equation (4) is 2^d (see [26]) and, in general, this so called high dimensional model representation

[24] is non-unique. But, if each term is chosen to satisfy the following condition

$$\int_0^1 f_{l_1 \dots l_\nu}(x_{l_1}, x_{l_2}, \dots, x_{l_\nu})\, \mathrm{d}x_{l_k} = 0, \quad 1 \le k \le \nu, \quad \nu = 1, \dots, d \tag{5}$$

then (4) is unique. The representation (4) is called ANOVA-representation of the model function $f(\mathrm{x})$ [25]. The functional decomposition of $[0;1]^d$ ANOVA (**an**alysis **o**f **va**riance) has been studied by many authors [3,13,22,28]. Sobol' has proven [23] that the decomposition (4) is unique on the assumption (5) and the functions of the right-hand side can be defined in a unique way by multidimensional integrals [25]:

- $f_0 = \int\limits_{U^d} f(\mathrm{x})\, \mathrm{dx};$
- $f_{l_1}(x_{l_1}) = \int\limits_{U^{d-1}} f(\mathrm{x}) \prod\limits_{k \ne l_1} \mathrm{dx}_k - f_0,\ l_1 \in \{1, 2, \dots, d\};$
- $f_{l_1 l_2}(x_{l_1}, x_{l_2}) = \int\limits_{U^{d-2}} f(\mathrm{x}) \prod\limits_{k \ne l_1, l_2} \mathrm{dx}_k - f_0 - f_{l_1}(x_{l_1}) - f_{l_2}(x_{l_2}), l_1, l_2 \in \{1, \dots, d\}.$

An additional essential property of the terms in the ANOVA-presentation is their mutually orthogonality:

$$\int\limits_{U^d} f_{i_1 \dots i_\mu} f_{j_1 \dots j_\nu}\ \mathrm{dx} = 0, \quad (i_1, \dots, i_\mu) \ne (j_1, \dots, j_\nu), \quad \mu, \nu \in \{1, \dots, d\}.$$

It follows from the assumption that the above subsets of indices differ from one another at least one element and the corresponding integral vanishes for this index due to (5).

The quantities

$$\mathbf{D} = \int\limits_{U^d} f^2(\mathrm{x})\, \mathrm{dx} - f_0^2, \quad \mathbf{D}_{l_1\ \dots\ l_\nu} = \int f^2_{l_1\ \dots\ l_\nu}\, \mathrm{d}x_{l_1} \dots\, \mathrm{d}x_{l_\nu} \tag{6}$$

are called variances (total and partial variances, respectively) and have been obtained after squaring and integrating over U^d the equality (4) on the assumption that $f(\mathrm{x})$ is a square integrable function (thus all terms in (4) are also square integrable functions). Therefore, the total variance of the model output is partitioned into partial variances [23] in the analogous way as the model function, that is the unique ANOVA-decomposition:

$$\mathbf{D} = \sum_{\nu=1}^{d} \sum_{l_1 < \dots < l_\nu} \mathbf{D}_{l_1 \dots l_\nu}. \tag{7}$$

It is obvious that the use of terms of probability theory is based on the following interpretation: in general, the input parameters are random variables distributed in U^d that defines $f_{l_1\ \ldots\ l_\nu}(x_{l_1}, x_{l_2}, \ldots, x_{l_\nu})$ also as random variables with variances (6). For example f_{l_1} is presented by a conditional expectation:

$$f_{l_1}(x_{l_1}) = \mathbf{E}(\mathrm{u}|x_{l_1}) - f_0 \quad \text{and respectively} \quad \mathbf{D}_{l_1} = \mathbf{D}[f_{l_1}(x_{l_1})] = \mathbf{D}[\mathbf{E}(\mathrm{u}|x_{l_1})].$$

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

$$S_{l_1\ \ldots\ l_\nu} = \frac{\mathbf{D}_{l_1\ \ldots\ l_\nu}}{\mathbf{D}}, \quad \nu \in \{1, \ldots, d\} \tag{8}$$

are called Sobol' global sensitivity indices [23,25]. This formula coincides for $\nu = 1$ with (2) and the so defined measures correspond to the main effect of input parameters as well as the interactions effect. Using the definition of these measures as ratios of variances and dividing (7) by $\mathbf{D}$, it is easy to show that the following properties hold for the Sobol' global sensitivity indices: $S_{l_1\ \ldots\ l_\nu} \geq 0$, and $\sum_{\nu=1}^{d} \sum_{l_1<\ldots<l_\nu}^{d} S_{l_1\ \ldots\ l_\nu} = 1$.

Based on the results discussed above it is clear that the mathematical treatment of the problem of providing global sensitivity analysis consists in evaluating total sensitivity indices (3) and in particular Sobol' global sensitivity indices (8) of corresponding order. And that leads to computing of multidimensional integrals: $I = \int_\Omega g(\mathrm{x})p(\mathrm{x})\,\mathrm{dx}$, $\Omega \subset \mathbf{R}^d$, where $g(\mathrm{x})$ is a square integrable function in Ω and $p(\mathrm{x}) \geq 0$ is a probability density function, such that $\int_\Omega p(\mathrm{x})\,\mathrm{dx} = 1$. This means that in general case one needs to compute 2^d integrals of type (6) to obtain $S_{x_i}^{tot}$. As we discussed earlier the basic assumption underlying representation (4) is that the basic features of the model functions (1) describing typical real-life problems can be presented by low-order subsets of input variables [15,17], that are constants, terms of first and second order. Thus, the high-dimensional sums (referred to higher-order interactions effects) in (4) can normally be neglected. Therefore, based on this assumption, one can assume that the dimension of the initial problem can be reduced.

Nevertheless, the calculating of the integrals defined by formulas (6) requires integration of different integrands that is not effective according to the computational cost. The procedure for computing global sensitivity indices measuring effect (main or otherwise) of the input parameters that is overcoming this disadvantage has been proposed by Sobol' [25]. Consider an arbitrary set of m variables $(1 \leq m \leq d-1)$: $\mathrm{y} = (x_{k_1}, \ldots, x_{k_m})$, $1 \leq k_1 < \ldots < k_m \leq d$, and let z be the set of $d-m$ complementary variables. Thus $\mathrm{x} = (\mathrm{y}, \mathrm{z})$. Let $K = (k_1, \ldots, k_m)$.

The variances corresponding to the subsets y and z can be defined as

$$\mathbf{D}_{\mathrm{y}} = \sum_{n=1}^{m} \sum_{(i_1<\ldots<i_n)\in K} \mathbf{D}_{i_1\ \ldots\ i_n}, \qquad \mathbf{D}_{\mathrm{z}} = \sum_{n=1}^{d-m} \sum_{(j_1<\ldots<j_n)\in \bar{K}} \mathbf{D}_{j_1\ \ldots\ j_n}, \tag{9}$$

where the complement of the subset K in the set of all parameter indices is denoted by $\bar{K}$. The first sum in (9) is extended over all subsets $(i_1, \ldots, i_n)$, where all indices $i_1, \ldots, i_n$ belong to K. Then the total variance corresponding to the subset y is $\mathbf{D}_{\mathrm{y}}^{tot} = \mathbf{D} - \mathbf{D}_{\mathrm{z}}$ and it is extended over all subsets $(i_1, \ldots, i_\nu),\ 1 \le \nu \le d$, where at least one $i_l \in K,\ 1 \le l \le \nu$.

The procedure for computation of global sensitivity indices is based on the following representation of the variance $\mathbf{D}_{\mathrm{y}}$: $\mathbf{D}_{\mathrm{y}} = \int f(\mathrm{x})\, f(\mathrm{y}, \mathrm{z}')\, \mathrm{dx}\, \mathrm{dz}' - f_0^2$ (see [25]). The last equality allows to construct a Monte Carlo algorithm for evaluating $f_0, \mathbf{D}$ and $\mathbf{D}_{\mathrm{y}}$, where $\xi = (\eta, \zeta)$:

$$\frac{1}{N}\sum_{j=1}^{N} f(\xi_j) \xrightarrow{P} f_0, \qquad \frac{1}{N}\sum_{j=1}^{N} f(\xi_j)\, f(\eta_j, \zeta_j') \xrightarrow{P} \mathbf{D}_{\mathrm{y}} + f_0^2,$$
$$\frac{1}{N}\sum_{j=1}^{N} f^2(\xi_j) \xrightarrow{P} \mathbf{D} + f_0^2, \qquad \frac{1}{N}\sum_{j=1}^{N} f(\xi_j)\, f(\eta_j', \zeta_j) \xrightarrow{P} \mathbf{D}_{\mathrm{z}} + f_0^2.$$

For example, for $m = 1, \mathrm{y} = \{x_{l_1}\}, l_1 \in \{1, \ldots, d\}$ and $\mathrm{z} = \{1, \ldots, d\} \backslash l_1$: $S_{l_1} = S_{(l_1)} = \mathbf{D}_{(l_1)}/\mathbf{D},\ S_{l_1}^{tot} = \mathbf{D}_{l_1}^{tot}/\mathbf{D} = 1 - S_{\mathrm{z}}$.

It is important to estimate the computational cost for computing the sensitivity indices in order to be able to compare this approach with other existing approaches. The computational cost of estimating all first-order ($m = 1$) and total sensitivity indices via the scheme proposed by Sobol' can be defined as $N(2d + 1)$ model function evaluations (N model runs for f_0, dN model runs for the first-order terms, and dN model runs for the total effect terms), where N is the sample size and d is the number of input parameters. It should be noted that the most frequently used variance-based methods as Sobol' method and FAST (Fourier Amplitude Sensitivity Test) (and their improved versions) have a computational cost proportional to dN of estimating all main and total effects of input parameters (see [19]).

The computing of higher-order interactions effect can be performed by an iterative process. For example, $S_{(l_1 l_2)} = \mathbf{D}_{(l_1 l_2)}/\mathbf{D} = S_{l_1} + S_{l_2} + S_{l_1 l_2}$, and $S_{l_1 l_2}$ can be obtained assuming that the corresponding first-order sensitivity indices have been already computed.

2.4 Monte Carlo Approaches for Small Sensitivity Indices

Unfortunately the standard Monte Carlo algorithm for estimating global sensitivity indices, proposed in [23], is spoilt by loss of accuracy when $\mathbf{D}_y << f_0^2$, i.e. in the case of small (in values) sensitivity indices. In Section 3.6 we will discuss this loss of accuracy based on our numerical results presented on Table 1. That is why here we have applied two approaches for evaluating small sensitivity indices - reducing of the mean value (proposed by I.M. Sobol', 1990) and a combined approach (it is a combination of approaches of reducing of the mean value and correlated sampling) suggested in [27]. These approaches are described briefly below.

The concept of the first approach consists of replacement of the original integrand (the mathematical model function) by a function of the following type $\varphi(\mathrm{x}) = f(\mathrm{x}) - c$, where $c \sim f_0$. For numerical experiments we have chosen the constant c to be a Monte Carlo estimate of f_0. Therefore the following formulas hold to estimate the partial and total variances respectively:

$$\mathbf{D}_{\mathrm{y}} = \int \varphi(\mathrm{x})\ \varphi(\mathrm{y},\mathrm{z}')\,\mathrm{dx}\,\mathrm{dz}' - \omega^2, \quad \text{where} \quad \omega = \int \varphi(\mathrm{x})\,\mathrm{dx},$$
$$\mathbf{D} = \int \varphi^2(\mathrm{x})\,\mathrm{dx} - \omega^2, \qquad\qquad \omega = f_0 - c.$$

A new estimator for variances has been proposed in the combined approach:

$$\mathbf{D}_{\mathrm{y}} = \int \varphi(\mathrm{x})\ [\varphi(\mathrm{y},\mathrm{z}')\mathrm{dx}\,\mathrm{dz}' - \varphi(\mathrm{x}')]\,\mathrm{dx}\,\mathrm{dx}',\ \mathbf{D} = \int \varphi(\mathrm{x})[\varphi(\mathrm{x}) - \varphi(\mathrm{x}')]\,\mathrm{dx}\,\mathrm{dx}'.$$

It should be noted that the variance of the second approach is smaller than the variance of the first one under certain conditions specified in the proposition proven in [27]:

Proposition (I. Sobol', E. Myshetskaya, 2007). Denote $\delta = \sup|v(\mathrm{y})|2/\mathbf{D}_{\mathrm{z}}$. If $\delta < 1$ and $S_{\mathrm{z}} > \dfrac{1}{2-\delta}$, then $V_0^{(2)} < V_0^{(1)}$, where $V^{(1)}$ and $V^{(2)}$ refer to the variances of the first and the second approach for small indices respectively and $v(\mathrm{y}) = \int\ g_2(\mathrm{z}')\ g_{12}(\mathrm{y},\mathrm{z}')\,\mathrm{dz}'$.

The used quantities are terms in the ANOVA-like decomposition of the model function:

$$f(\mathrm{x}) = f_0 + g_1(\mathrm{y}) + g_2(\mathrm{z}) + g_{12}(\mathrm{x}), \qquad \text{where}$$
$$g_1(\mathrm{y}) = \int f(\mathrm{x})\,\mathrm{dz} - f_0, \quad g_2(\mathrm{z}) = \int f(\mathrm{x})\,\mathrm{dy} - f_0,$$
$$g_{12}(\mathrm{x}) = f(\mathrm{x}) - \int f(\mathrm{x})\,\mathrm{dy} - \int f(\mathrm{x})\,\mathrm{dz} + f_0.$$

3 Numerical Experiments. A Case Study

3.1 Air Pollution Model

The particular model for our studies is the UNI-DEM, but most of the results can also be applied when other large-scale mathematical models are used. This model is one of the most advanced large-scale mathematical models that describes adequately all physical and chemical processes. A two-dimensional discretization on a 96×96 grid has been selected. All runs were performed for the period 1994-1998. For the sensitivity study a representative summer month has been selected because it is well-known that the concentrations of many chemical species achieve their annual maximum in summer-time. Here it has been chosen to study the sensitivity of the concentrations of one of the most important pollutants - ozone. Sensitivity analysis is applied for the average values of concentrations for this month obtained by using UNI-DEM.

The chemical reactions taking place during air pollution transport are among the most important processes. This is why we decided to concentrate our attention to the chemical reactions.

3.2 Sensitivity Tests with UNI-DEM Performed in a Previous Study

The idea of performed sensitivity tests with UNI-DEM that is proposed and discussed in [6,7] is based on the computation of standard deviations and skewness of the pollutants concentrations under consideration, for example ozone and nitrogen dioxide. The experiments were performed with a sequence of 100 normally distributed random values of the constant rate of the reaction $O_3 + NO \Longrightarrow NO_2$ using the two-dimensional version of UNI-DEM.

As a result of these experiments some major conclusions have been drawn about

- the relationship between emissions amount and pollutants concentrations,
- the distributions of the standard deviations of pollutants concentrations and the influence of the variety of the variance of the sequence of normally distributed random values of the rate constant of the above chemical reaction on the standard deviations of pollutants concentrations, and
- the sensitivity of the dependence of pollutants concentrations on the variety of the chemical rate constants.

3.3 Using a Box-model to Obtain Initial Guesses about the Sensitivity

The computer treatment of a large-scale air pollution model is an extremely expensive process. Furthermore, the treatment of the chemical reactions is the most time-consuming part of any air pollution model. Therefore, it is necessary to find some simple way for obtaining an initial evaluation of the sensitivity of the concentrations of the different species to variations of the rates of the chemical reactions. The so-called "box-model" can conveniently be used in the solution of this sub-task [31]. Consider the system of partial differential equations by which UNI-DEM is described mathematically (see Subsection 2.1). Remove all terms excluding the emission and chemical terms. Since these terms do not depend on the spatial variables, a system of ordinary differential equations $dg_{s,i,j,k}/dt = E_{s,i,j,k} + Q_{s,i,j,k}(g_{1,i,j,k}, g_{2,i,j,k}, \ldots, g_{q,i,j,k})$, where $g_{s,i,j,k}(t)$ is the value of the concentration c_s at an arbitrary grid-point (x_i, y_j, z_k) in the space domain at time t has to be considered. This means that, roughly speaking, the model is considered at a given grid-point, which also explains the use of the term box-model. While the computer treatment of the whole UNI-DEM leads to the solution of huge systems of ordinary differential equations (containing millions of equations) during many time-steps, the box-model is a rather small system, which does not cause computational difficulties. It was possible to run this system in a large number of sensitivity tests where the chemical rates of all chemical reactions were cyclicly varied by multiplying them by a factor α, where $\alpha \in [0.1, 2.0]$. The resulting simple optimization procedure was used to determine the chemical rates for which the overall results (including all concentrations) are most sensitive. By using this procedure, it was possible to find out that the results seem to be most sensitive to the variation of the 3-rd, 22-nd and 28-th chemical reactions from the list of reactions of the condensed CBM IV scheme ([29]). The simplified chemical equations of these reactions are as follows:

$\#3 \quad O_3 + NO \Longrightarrow NO_2$

$\#22 \quad HO_2 + NO \Longrightarrow OH + NO_2$

$\#28 \quad OH + CO \Longrightarrow HO_2$

Note that the ozone does not necessarily participate in all these reactions. Important precursors of ozone participate instead.

3.4 First Stage of Computations based on UNI-DEM

This stage of computations consists of generation of input data for providing sensitivity analysis. On the other hand, the first computational stage repre-

sents an output of the UNI-DEM. The model runs have been done for the chemical rates variations with a fixed set of perturbation factors $\alpha = \{\alpha_i\}, i = 1, \ldots, d$, applied to the constants of a corresponding subset of d chemical reactions, selected for these experiments among the total number of reactions in the target (CBM IV) chemical scheme used in the model.

The generated data is ratios of the following type

$$r_s(\alpha) = \frac{c_s^{\alpha}(a_s^{i_{max}}, b_s^{j_{max}})}{c_s^{max}}, \qquad \alpha_i \in \{0.1, 0.2, \ldots, 2.0\},$$

where the lower index s corresponds to the chemical pollutant, $s = 1, \ldots, 35$. The denominator $c_s^{max} = c_s^{max}(a_s^{i_{max}}, b_s^{j_{max}})$ is the maximum mean value of the concentration of chemical species s (in the numerical experiments - for July 1998) obtained for $\alpha = (1, \ldots, 1)$, i.e. without any perturbations, $a_s^{i_{max}}$ and $b_s^{j_{max}}$ are the coordinates of the point, where this maximum has been reached, and i_{max}, j_{max} are the mesh indices of this point. The nominator represents the values of the concentrations of the corresponding pollutant for a given set of values of the perturbation parameters $\alpha_i \in \{0.1, \ldots, 2.0\}$, computed at the point $(a_s^{i_{max}}, b_s^{j_{max}})$. Thus the input data that would be analyzed by the applied and investigated here sensitivity analysis tool is a set of pollutant concentrations normalized according to the maximum mean value of the concentration of the corresponding chemical species.

The numerical results on the first stage has been obtained on a SunFire E25000 supercomputer at the Technical University of Denmark (DTU). Long series of scenarios where (a) the meteorological conditions were fixed while the emissions were varied and (b) the emissions were kept constant, while meteorological conditions for different years are used were run. Based on those numerical results a number of plots similar to one presented on Figure 1 were analysed. The results indicate that if the emission changes are considerably large, then the effect of these changes is greater than the effect of the inter-annual meteorological conditions. The point (near the border of Germany, France and Switzerland) where the average ozone concentrations in July 1998 was maximal is taken and used for all five years.

We should mention here that the reliability of the model is an important issue. The reliability of the results obtained by UNI-DEM was checked by comparing them with observations produced by a large number of stations located over Europe, as well as, by running the "rotation test" (see, [31]). The results show that the accuracy of the numerical simulations is good enough for sensitivity analysis studies.

Figure 1 shows the distribution of ozone concentrations over Europe for July 1998. Similar plots are drawn for other important pollutants like sulfur dioxide, nitrogen dioxide, ozone, peroxy radicals and nitrogen oxide. The results

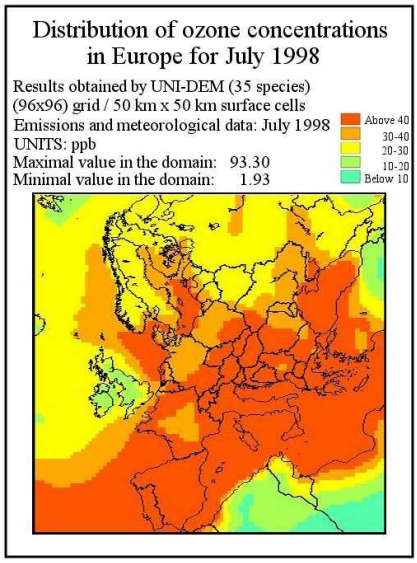

Fig. 1. Distribution of ozone concentrations in Europe for July 1998.

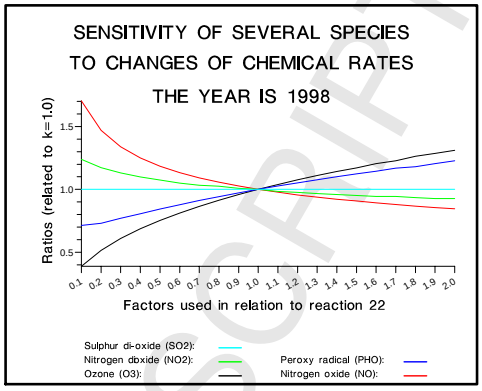

Fig. 2. Sensitivity of several species to changes of chemical rates (1998).

presented on these plots allow us to identify the computational mesh-points where each pollutant reaches its maximum.

We study numerically how the chemical rate constants (considered as random variables) influence the output results. As an example we show how the rate constant of the 22^{nd} chemical reaction of CBM-4 scheme influences the concentrations of sulfur dioxide, nitrogen dioxide, ozone, peroxy radicals and nitrogen oxide for a known scenario taken for July 1998 (see Figure 2). The results are for the mesh point of the computational domain where the concentrations reach their maximum. Figure 2 illustrates the fact that the influence of this particular rate constant (for the 22^{nd} reaction) to nitrogen dioxide, ozone, peroxy radicals and nitrogen oxide is significant, but at the same time the influence to sulfur dioxide is practically negligible.

In order to check the substantiability of the results we run the model for a period of 5 years - from 1994 to 1998 (see Figure 3). We can see how the situation changes for different years with different meteorological conditions. A number of plots were prepared for different pollutants and Figure 3 is just an example illustrating how the same 22^{nd} chemical reaction of CBM-4 scheme influences the concentrations of ozone. One can see that the influence does not change a lot from one year to another. In this way the tendency of concentration change depending on the variation of the perturbation factor remains the same (or very similar) for different years (see Figure 3).

It is studied numerically how various chemical rate reactions influence air-pollution concentrations. An example is shown on Figure 4. Analyzing pre-

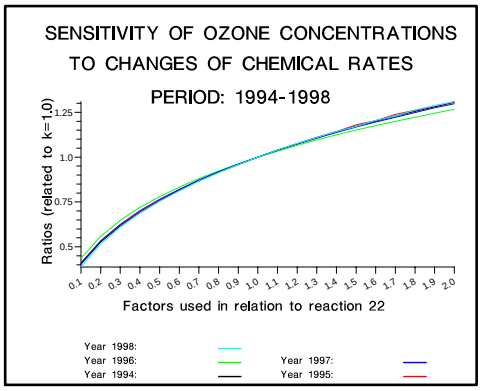

Fig. 3. Sensitivity of ozone concentrations to changes of chemical rates.

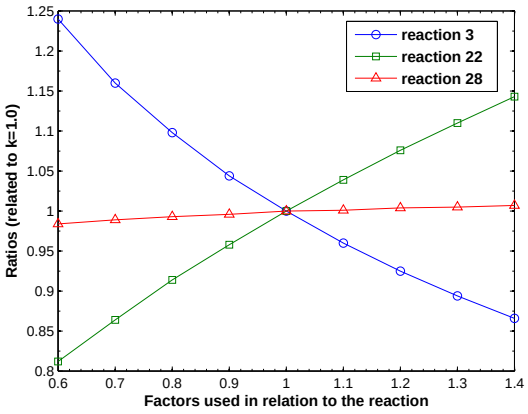

Fig. 4. Sensitivity of ozone concentrations to changes of chemical rates (July 1998).

sented results for reactions ## 3, 22 and 28 of CBM-4 scheme one can conclude that the influence of reactions ## 3 and 22 on ozone concentrations is significant. At the same time the influence of reaction # 28 is relatively low.

3.5 Second Stage of Computations with UNI-DEM

The second stage of computations consists of two steps: (i) Approximation, and (ii) Computing of Sobol' global sensitivity indices.

As a result of computations with the use of UNI-DEM we obtain tables of the values of the model function. These values represent the quotient between the values of the concentration of the corresponding pollutant for a given set of values of the perturbation parameters $\alpha_i \in \{0.1, \ldots, 2.0\}$, computed at the point where the maximum average value of the concentration has been reached, and this average value for $\alpha = (1, \ldots, 1)$. Since the application of the sensitivity analysis method supposes that the model is given with a function (1), the first step is to use approximation in order to produce a function from the table of values.

This stage is an important link between the generation of experimental data and the mathematical technology for sensitivity analysis. The precise approximation of data is crucial for the overall reliability of the obtained sensitivity indices. That is why the investigation and determination of an applicable instruments for approximation of the table function is important part of the work.

As an initial step we use polynomials of third and forth degree as an approximation tool, where $p_s(\mathrm{x})$ is the polynomial that approximates the mesh

function given in the table that corresponds to the s-th chemical specie:

$$p_s(\mathrm{x}) = \sum_{j=0}^{k} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_d = 0 \\ \nu_1+\ldots+\nu_d = j}}^{k} a_{\nu_1\ldots\nu_d}\ x_1^{\nu_1} x_2^{\nu_2}\ldots x_d^{\nu_d}, \qquad k = 3, 4.$$

The squared 2-vector norm $\| p_s - r_s \|_2^2 = \sum_{l=1}^{n} [p_s(\mathrm{x}_l) - r_s(\mathrm{x}_l)]^2,\ \mathrm{x}_l \in [0.1; 2.0]^3$ in the case of a polynomial of 4-th degree in three variables is $\| p_s - r_s \|_2^2 = 0.016$ for $\mathrm{x}_l \in [0.1; 2.0]^3$ and $\| p_s - r_s \|_2^2 = 0.00005$ for $\mathrm{x}_l \in [0.6; 1.4]^3$ in our numerical experiments. The obtained results show that the presented norm is more influenced by the domain than by the degree of the polynomial. This feature is one of the preconditions for the specification of the domain of perturbation factors.

3.6 Analysis of the Results

Since three chemical reactions have been chosen as the most important for the distribution of pollutants concentrations (for example ozone), the domain of integration is a cube: $\Omega = [0.6; 1.4]^3$. This interval has been specified according to the variation of the perturbation factors and the regions where the mathematical model function (an approximation function of 'real' experimental data obtained using UNI-DEM) has only positive values (it is a natural requirement because it is interpreted physically as a pollutant concentration). Actually, it has been established experimentally that a chemical rate varies with a normal distribution with mean 1.0. It should be emphasized also that the law of conservation of mass (matter) of chemical reactions may be broken for larger intervals.

The results from the numerical experiments are presented in Table 1. The following notation is used: g_0 is the integral over the integrand $g(x)$; c is a constant obtained as a Monte Carlo estimate of f_0. Polynomials of 3-rd (20 unknown coefficients) and 4-th (35 unknown coefficients) degree are used for data approximation.

According to Table 1 the results for total sensitivity indices obtained using the combined approach for small indices are the most reliable.

We analyze how variations of the input parameters influence on the model output, where we consider the chemical rate constants as input parameters and concentrations of pollutants (more exactly, normalised according to the maximum mean value for July 1998 of the concentration of the corresponding

Table 1
Total sensitivity indices of input parameters obtained using different approaches of sensitivity analysis.

approach / estimated quantity	Standard (Sobol')		Approaches for small indices	
			red. of the m.v.	combined
	$x \in [0.1; 2.0]$	$x \in [0.6; 1.4]$	$x \in [0.6; 1.4]$	$x \in [0.6; 1.4]$
integrand $g(x)$	$f(x)$	$f(x)$	$f(x) - c$	$f(x) - c$
c	-	-	0.51737	0.51737
g_0	0.51520	0.51634	0.25145	0.25145
$\mathbf{D}$	0.26181	0.26446	0.07061	0.00530
S_1	0.26386	0.26530	0.27354	0.52979
S_2	0.26447	0.26359	0.26713	0.46142
S_3	0.25348	0.25209	0.22406	0.00222
$\sum_{i=1}^{3} S_i$	0.78182	0.78097	0.76474	0.99342
S_{12}	0.06885	0.06941	0.07994	0.00628
S_{13}	0.06598	0.06634	0.06845	0.00009
S_{23}	0.06613	0.06592	0.06686	0.00021
$\sum_{i,j=1, i\leq j}^{3} S_{ij}$	0.20096	0.20167	0.21525	0.00658
S_{123}	0.01722	0.01736	0.02001	0.000003
$S_{x_1}^{tot}$	0.41592	0.41841	0.44195	0.53615
$S_{x_2}^{tot}$	0.41667	0.41627	0.43395	0.46791
$S_{x_3}^{tot}$	0.40281	0.40170	0.37938	0.00252

chemical species) as output parameters. All sensitivity indices (first-, second- and third-order as well as total effects) obtained using standard Sobol' approach for computing sensitivity indices and both presented approaches for small indices are given in the Table 1. The domain of integration is $[0.6; 1.4]^3$ for all computations, but as a first case the data in the whole domain has been used to determine unknown coefficients of the approximation function and as a second case the data only in the subdomain $[0.6; 1.4]^3$. One can see that the results in both cases are very similar. Thus, we need only the data in $[0.6; 1.4]^3$ for providing a reliable sensitivity analysis using the presented scheme.

Obviously the results shown in the Table 1 and referred to the standard Sobol' approach show that $\mathbf{D}_y = \mathbf{D}S_y << f_0^2$ in this particular case. This means that the improved approaches that overcome the loss of accuracy for small indices should be used. On the other hand, the results obtained with the approach of

Table 2
First-order and total sensitivity indices of input parameters obtained using combined approaches and two approaches implemented in R package.

estimated quantity	$x \in [0.1; 2.0]^3$			$x \in [0.6; 1.4]^3$		
	combined approach	R Package		combined approach	R Package	
		Sobol', Saltelli	FAST		Sobol', Saltelli	FAST
S_1	0.48262	0.46645	0.47933	0.52979	0.53029	0.52783
S_2	0.51080	0.54567	0.50928	0.46142	0.47884	0.46034
S_3	0.00104	0.00288	0.00101	0.00222	0.00254	0.00221
$S_{x_1}^{tot}$	0.48807	0.48641	0.48800	0.53615	0.53760	0.53609
$S_{x_2}^{tot}$	0.51592	0.50845	0.51604	0.46791	0.45089	0.46799
$S_{x_3}^{tot}$	0.00157	-0.00043	0.00308	0.00252	0.00180	0.00365

reducing the mean value suggest that the estimation especially of very small indices should be performed using the combined approach. It confirms the priority of the combined approach according to variance of the corresponding estimator over the first approach (see Proposition from Subsection 2.4). The results for total sensitivity indices obtained using the combined approach for small indices are the most reliable - the values of total effects are fully consistent with the expected tendencies according to Figure 4.

A comparison of results obtained using the present scheme for sensitivity analysis (applying the combined approach for small sensitivity indices) and results obtained using the available software tool for sensitivity analysis SA - R language (or R Package) and environment for statistical computing (http://www.r-project.org/) is given in Table 2. One can see that the results are close with an exception for $S_{x_3}^{tot}$ obtained with the R package using Sobol'& Saltelli approach, where we found a negative value for $S_{x_3}^{tot}$ which is not acceptable. The reason that one would prefer our approach is that we are able to control the accuracy at each stage of the computations, i.e. at the stage of

- approximation of the mesh function by changing the polynomial degree and
- computing total sensitivity indices by applying the refined technique suitable for computing *small* indices.

4 Applicability of the Results

Sensitivity analysis, and in particular the results achieved have an important twofold role: for mathematical models verification and/or improvement,

and/or on the other hand, for a reliable interpretation of experts of main effect, interaction and higher-order interaction effect of input parameters on model output. Variance-based analysis is an useful tool for an advanced investigation of relationships between model parameters, output results and internal mechanisms regulating the system under consideration. Specifying the most important chemical reactions for the model output the specialists from various applied fields (chemistry, physics) may obtain valuable information for an improvement of the model and thus it will lead to an increase of reliability and robustness of predictions. In this way the mathematical models will become able to predict better the effects of high pollution levels (a) on human health and (b) on losses of crops in agriculture using sensitivity analysis.

Our numerical results show that the standard Sobol' approach and the reduction of the mean value are applicable in cases when the sensitivity indices are not very small. In most cases the standard approach gives reliable results if $S_{x_i} \geq 0.35$. If $S_{x_i} < 0.35$, then the more complicated combined approach should be used. Our advice to people dealing with sensitivity analysis would be to apply the standard approach if there is a priori information that the values of S_{x_i} are larger than 0.35. If such an information doesn't exist we would recommend to use the standard Sobol' approach and if the computed value is less than 0.4 to perform computations again with the combined approach.

5 Concluding Remarks

A systematic scheme for providing sensitivity analysis to a mathematical model of atmospheric chemistry (UNI-DEM) has been used to analyse the sensitivity of concentrations of some important air pollutants to chosen chemical rate reactions.

We have demonstrated that the important advantage of the method we use is the possibility to compute not only the first-order indices, but also indices of a higher-order. Furthermore, the total sensitivity index can be calculated with just one Monte Carlo integral per factor in a way similar to the computation of the first-order indices. It makes this approach one of the most efficient variance-based methods from the point of view of its computational cost (for estimating all first-order and total sensitivity indices). The computational cost is proportional to the sample size and the number of input parameters.

We also have shown that for some considerations *small* sensitivity indices are important. To be able to get relevant estimates of *small* indices one needs to apply a special combined technique which includes a variance reduction method and correlated sampling.

As a future research in this area we plan

- to consider other approximation tools (especially approximation by cubic B-splines looks promising);
- to perform computations with the refined 3D version of UNI-DEM;
- to study model sensitivity on emissions levels and boundary conditions.

6 Acknowledgement

This work is supported by NATO under project NATO CLG-982641 and by NSF of Bulgaria under Grants DO 02-115/2008, DTK 02/44/2009 and DO 02-215/2008.

References

[1] S. Abdalmogith, R. Harrison, Z. Zlatev, Intercomparison of inorganic aerosol concentrations in the UK with predictions of the Danish Eulerian Model, *Journal of Atmospheric Chemistry* **54** (2006), 43–66.

[2] C. Ambelas Skjøth, A. Bastrup-Birk, J. Brandt, Z. Zlatev, Studying variations of pollution levels in a given region of Europe during a long time-period, *Systems Analysis Modelling Simulation* **37** (2000), 297–311.

[3] G. Archer, A. Saltelli, I. Sobol', Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *J. Stat. Comput. Simul.* **58** (1997), 99–120.

[4] A. Bastrup-Birk, J. Brandt, I. Uria, Z. Zlatev, Studying cumulative ozone exposures in Europe during a 7-year period, *Journal of Geophysical Research* **102** (1997), 23917–23035.

[5] K. Chan, A. Saltelli, S. Tarantola, Sensitivity analysis of model output: variance-based methods make the difference, In: S. Andradottir, K.J. Healy, D.H. Withers, and B.L. Nelson, eds., *Proceedings of the 1997 Winter Simulation Conference* (1997), 261–268.

[6] I. T. Dimov, *Monte Carlo Methods For Applied Scientists*, World Scientific (2007).

[7] I. T. Dimov, Z. Zlatev, Testing the sensitivity of air pollution levels to variations of some chemical rate constants, *Notes on Numerical Fluid Mechanics* **62** (1997), 167–175.

[8] M. Gery, G. Whitten, J. Killus, M. Dodge. A photochemical kinetics mechanism for urban and regional scale computer modelling, *J. Geophys Res.* **94** (D10) (1989), 12925–12956.

[9] R. Harrison, Z. Zlatev, C. Ottley, A comparison of the predictions of an Eulerian atmospheric transport chemistry model with experimental measurements over the North Sea, *Atmospheric Environment* **28** (1994), 497–516.

[10] H. Hass, M. van Loon, C. Kessler, R. Stern, J. Mathijsen, F. Sauter, Z. Zlatev, J. Langner, V. Foltescu, M. Schaap, Aerosol modelling: Results and

intercomparison from European regional-scale modelling systems, GSFNational Research Center for Environment and Health, International Scientific Secretariat (ISS), EUROTRAC-2, Munich (2004).

[11] Á. Havasi, L. Bozó, Z. Zlatev, Model simulation on transboundary contribution to the atmospheric sulfur concentration and deposition in Hungary, *Időjárás* **105** (2001), 135–144.

[12] Á. Havasi, Z. Zlatev, Trends of Hungarian air pollution levels on a long time-scale, *Atmospheric Environment* **36** (2002), 4145–4156.

[13] W. Hoeffding, A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.* **19** (3) (1948), 293–325.

[14] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering and System Safety* **52** (1996), 1–17.

[15] J. Jacques, C. Lavergne, N. Devictor, Sensitivity analysis in presence of model uncertainty and correlated inputs, *Reliability Engineering and System Safety* **91** (2006), 1126–1134.

[16] M. McKay, Evaluating prediction uncertainty, *Technical Report NUREG/CR-6311*, US Nuclear Regulatory Commission and Los Alamos National Laboratory (1995).

[17] H. Rabitz, O. Alis, J. Shorter, K. Shim, Efficient input-output model representations, *Computer Phys. Comm.* **117** (1999), 11–20.

[18] M. Roemer, M. Beekman, R. Bergsröm, G. Boersen, H. Feldmann, F. Flatøy, C. Honore, J. Langner, J. Jonson, J. Mathijsen, M. Memmesheimer, D. Simpson, P. Smeets, S. Solberg, D. Stevenson, P. Zandveld, Z. Zlatev, Ozone trends according to ten dispersion models, GSF National Research Center for Environment and Health, International Scientific Secretariat (ISS), EUROTRAC-2, Munich, (2004).

[19] A. Saltelli, Making best use of model valuations to compute sensitivity indices, *Computer Physics Communications* **145** (2002), 280–297.

[20] A. Saltelli, K. Chan, M. Scott, *Sensitivity Analysis*, John Wiley and Sons publishers, Probability and Statistics series (2000).

[21] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Halsted Press (New York, 2004).

[22] I. Sobol', *Multidimensional quadrature formulas and Haar functions*, (Nauka, Moscow, 1969) (in Russian).

[23] I. M. Sobol', Sensitivity estimates for nonlinear mathematical models, *Matem. Modelirovanie* **2** (1) (1990), 112–118.

[24] I. M. Sobol', Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment* **1** (1993), 407–414.

[25] I. M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation*, **55** (1-3) (2001) 271–280.

[26] I. M. Sobol', Theorem and examples on high dimensional model representation, *Reliability Engineering and System Safety* **79** (2003), 187–193.

[27] I. Sobol', E. Myshetskaya, Monte Carlo estimators for small sensitivity indices, *Monte Carlo Methods and Applications* **13** (5-6) (2007), 455–465.

[28] A. Takemura, Tensor analysis of ANOVA decomposition, *J. Amer. Statist.*

Assoc. **78** (1983), 894–900.

[29] Z. Zlatev, *Computer treatment of large air pollution models*, KLUWER Academic Publishers (Dorsrecht-Boston-London, 1995).

[30] Z. Zlatev, Massive data sets issues in air pollution modelling, *In: Abello J, Pardalos PM, Resende MGC (eds.), Handbook on Massive Data Sets in Science and Engineering*, Kluwer Academic Press, Dordrecht-Boston-London (2002), 1169–1220.

[31] Z. Zlatev, I. T. Dimov, *Computational and Numerical Challenges in Environmental Modelling* (Elsevier, Amsterdam, 2006).

[32] Z. Zlatev, I. T. Dimov, K. Georgiev, Modeling the long-range transport of air pollutants, *IEEE Computational Science and Engineering*, **1** (3) (1994), 45–52.

[33] Z. Zlatev, I. T. Dimov, K. Georgiev, Three-dimensional version of the Danish Eulerian Model, *Zeitschrift für Angewandte Mathematik und Mechanik*, **76** (1996) S4, 473-476.

[34] Z. Zlatev, I. T. Dimov, Tz. Ostromsky, G. Geernaert, I. Tzvetanov, A. Bastrup-Birk, Calculating losses of crops in Denmark caused by high ozone levels, *Environmental Modelling and Assessment*, **6** (2001), 35–55.

[35] Z. Zlatev, D. Syrakov, A fine resolution modelling study of pollution levels in Bulgaria. Part1: SO2 and NO2 pollution, *International Journal of Environment and Pollution*, **22** (1-2) (2004), 186–202.

[36] Z. Zlatev, D. Syrakov, A fine resolution modelling study of pollution levels in Bulgaria. Part 2: High ozone levels, *International Journal of Environment and Pollution*, **22** (1-2) (2004), 203-222.