Reliable Monte Carlo Methods for Multidimensional Sensitivity Analysis

Venelin Todorov, Stoyan Poryazov

Abstract: *Sensitivity analysis is a promising technique for determining the stability, reliability, and efficiency of a mathematical model. Since the basic element in performing this procedure is the calculation of the corresponding numerical indicators, called total sensitivity indices, from a mathematical point of view this task is represented by multidimensional integrals. The total sensitivity index of an input parameter can be calculated with only one integral using the adaptive Monte Carlo method, analogous to the calculation of the first order indices. This makes the applied approach one of the most effective variance reduction based methods in terms of computational efficiency.*

Keywords: *Multidimensional integration, Sensitivity Analysis, Monte Carlo methods, Air pollution modelling.*

MSC*: 65C05, 65U05, 65F10, 65Y20*

ITHEA Keywords*: NUMERICAL ANALYSIS: Applications, SIMULATION AND MODELLING: Applications.*

Introduction

We discuss a systematic approach for sensitivity analysis studies in the area of air pollution modelling. The Unified Danish Eulerian Model (UNI-DEM) Zlatev [1995, 2006] is used in this particular study. Different parts of the large amount of output data, produced by the model, were used in various practical applications, where the reliability of this data should be properly estimated. Another reason to choose this model as a case study here is its sophisticated chemical scheme, where all relevant chemical processes in the atmosphere are accurately represented. We study the sensitivity of concentration variations of some of the most dangerous air pollutants with respect to the anthropogenic emissions levels and with respect to some chemical reactions rates. A special version of UNI-DEM (called SA-DEM) was developed for the purpose of this study. Description of UNI-DEM, SA-DEM and their parallel computer implementations will be given in the next section.

Different efficient stochastic algorithms for multidimensional integration have also been applied on a furter stage of these sensitivity studies. Between them are two adaptive Monte Carlo algorithms, described in more details in the paper. These will be compared with two QMC algorithms, namely Fibonacci lattice rule and Latin hypercube sampling. Fibonacci lattice rule is completely investigated in Wang and Hickernel [2002] and Latin hypercube sampling is described in detail in McKay et al. [1979].

1 Description and implementation of UNI-DEM

UNI-DEM is a powerful large-scale air pollution model for calculation the concentrations of a large number of pollutants and other chemical species in the air along certain time period. Its results can be used in various application areas (environmental protection, agriculture, health care, etc.). The large computational domain covers completely the European region and the Mediterranean.

UNI-DEM is mathematically represented by the following system of partial differential equations (PDE), in which the unknown concentrations c_s of a large number of chemical species (pollutants and other chemically active components) take part. The main physical and chemical processes (advection, diffusion, chemical reactions, emissions and deposition) are represented in that system.

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

where c_s are the concentrations of the chemical species; $u,\ v,\ w$ are the wind components along the coordinate axes; $K_x,\ K_y,\ K_z$ – the diffusion coefficients; E_s – the emissions; $k_{1s},\ k_{2s}$ – dry / wet deposition coefficients; $Q_s(c_1, c_2, \ldots c_q)$ – non-linear functions describing the chemical reactions between species under consideration. The above PDE system is non-linear and stiff. Both non-linearity and stiffness are introduced mainly by the chemical scheme: the condensed CBM-IV (Carbon Bond Mechanism) Zlatev [1995, 2006]. It is quite detailed and accurate, but computationally expensive as well.

For the purpose of efficient numerical treatment, the system (1) is split according to the major physical and chemical processes and the following 3 submodels are formed: ***Advection-diffusion***, ***Chemistry & deposition*** and ***Vertical transport (vertical wind and convection)***.

The following methods are used in the numerical solution of the submodels:

- *Advection-diffusion part:* Finite elements, followed by predictor-corrector schemes with several different correctors.
- *Chemistry-deposition part:* An improved version of the QSSA (Quazi Steady-State Approximation) Zlatev [1995].
- *Vertical transport:* Finite elements, followed by theta-methods.

Spatial and time discretization makes each of the submodels a tuff computational task even for the most advanced supercomputer systems. Efficient parallelization has always been a crucial point in the computer implementation of UNI-DEM. The task became much more challenging with development of the sensitivity analysis version of the code, SA-DEM Zlatev [2006]. It consists of he following three parts:

– A modification of UNI-DEM with ability to modify certain parameters, subject to SA study. By now we have been interested in some chemical rate constants as well as in the input data for the anthropogenic emissions. A small number of input parameters is reserved for this purpose.

– A driver routine that automatically generates a set of tasks to produce the necessary results for a particular SA study. It allows to perform in parallel a large number of runs with common input data (reusing it), producing at once a whole set of values on a regular mesh (used later for calculating the sensitivity indices).

– An additional program for extracting the necessary mean monthly concentrations and computing the normalised ratios (to be analysed further on).

Algorithm

Variance-based methods deliver results that are independent to the models behaviors: linearity, monotonicity and additivity of the relationship between input factor and model output sensitivity measures. The variance-based Sobol' method uses the sensitivity measures (indices) and takes into account interaction effects between inputs. 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 MC integral per factor. The computational cost of estimating all first-order (m = 1) and total sensitivity indices via Sobol' approach is proportional to dN, where N is the sample size and d is the number of input parameters.

The Sobol's method Dimov nad Georgieva [2010]is one of the most often used variance-based methods. It is based on a unique decomposition of the model function into orthogonal terms (summands) of increasing dimension and zero means. Its main advantage is computing in a uniform way not only the first order indices, but also the higher order indices (in quite a similar way as the computation of the main effects). The total sensitivity index can then be calculated with just one Monte Carlo integral per factor.

The Sobol method for global SA, applied here, is based on the so-called *High Dimensional Model Representation (HDMR)* (2) of the model function f (integrable) in the d-dimensional factor spaceGeorgieva [2010]:

$$f(\mathbf{x}) = f_0 + \sum_{s=1}^{d} \sum_{l_1<\ldots<l_s} f_{l_1\ldots l_s}(x_{l_1}, x_{l_2}, \ldots, x_{l_s}) \,, \tag{2}$$

where f_0 is a constant. The representation (2) is not unique. Sobol has proven that under the conditions (3) for the right-hand-side functions

$$\int_0^1 f_{l_1\ldots l_s}(x_{l_1}, x_{l_2}, \ldots, x_{l_s}) \, \mathrm{x}_{l_k} = 0, \quad 1 \le k \le s, \quad s = 1, \ldots, d \tag{3}$$

the decomposition (2) is unique and is called *ANOVA-HDMR* of the model function $f(\mathbf{x})$ (the abbreviation ANOVA comming from *Analysis of Variances*). Moreover, the functions of the right-hand side can be defined in a unique way by multidimensional integrals as follows (see also **?**

).

$$f_0 = \int_{U^d} f(\mathbf{x}) \dot{\mathbf{x}} \, , \tag{4}$$

$$f_{l_1}(x_{l_1}) = \int_{U^{d-1}} f(\mathbf{x}) \prod_{k \neq l_1} \dot{\mathbf{x}}_k - f_0, \; l_1 \in 1, \ldots, d \, , \tag{5}$$

$$\int_{U^d} f_{i_1 \ldots i_\mu} f_{j_1 \ldots j_\nu} \dot{\mathbf{x}} = 0, \;\; (i_1, \ldots, i_\mu) \neq (j_1, \ldots, j_\nu), \;\; \mu, \nu \in \{1, \ldots, d\} \, . \tag{6}$$

Definition: Sobol global sensitivity indices

$$S_{l_1 \, \ldots \, l_\nu} = \frac{\mathbf{D}_{l_1 \, \ldots \, l_\nu}}{\mathbf{D}}, \quad \nu \in \{1, \ldots, d\} \tag{7}$$

are defined as ratios of the partial variances

$$\mathbf{D}_{l_1 \, \ldots \, l_\nu} = \int f^2_{l_1 \, \ldots \, l_\nu} \dot{\mathbf{x}}_{l_1} \ldots \dot{\mathbf{x}}_{l_\nu} \tag{8}$$

over the total variance

$$\mathbf{D} = \int_{U^d} f^2(\mathbf{x}) \dot{\mathbf{x}} - f_0^2, \qquad \mathbf{D} = \sum_{\nu=1}^{d} \sum_{l_1 < \ldots < l_\nu} \mathbf{D}_{l_1 \ldots l_\nu} \, . \tag{9}$$

From equalities (9) the following properties hold:

$$S_{l_1 \, \ldots \, l_s} \geq 0, \qquad \sum_{s=1}^{d} \sum_{l_1 < \ldots < l_s}^{d} S_{l_1 \, \ldots \, l_s} = 1 \, . \tag{10}$$

While the classical deterministic methods for numerical integration are effective for sub-integer functions having a relatively small dimension, for high dimensions they become even inapplicable because the number of sub-integral function values required to calculate is growing exponentially. On the other hand, the order of convergence of the adaptive Monte Carlo method for integration is independent of dimension. Therefore, the Monte Carlo approach is an effective apparatus for conducting sensitivity analysis of large-scale systems. Dispersion may increase with increasing dimension, but there are various Monte Carlo techniques to reduce variance - the sampling method and its modifications, the small discrepancy series, the importance partitioning method. The adaptability approach is also a widely used effective approach to improve the convergence order of deterministic and stochastic numerical integration methods. There are different approaches to designing adaptive Monte Carlo algorithms. The adaptive algorithm implemented here does not use any prior information on the smoothness of the subintegral function, but uses posterior variance information Dimov et al. [2003]. The basic idea is to concentrate random points in sub-areas in which the variance is large (in terms of preset accuracy), ie. the approach is based on the recursive division of the area, using a posteriori information about the current division error. The algorithm starts by dividing the intervals along all directions of the M sub-interval, with M being set as an input parameter. For each subdomain, the respective integral and variance are calculated. The resulting variance is

then compared to a predetermined value. The result of the comparison is used to further divide the area and increase the density of random points. Random numbers are used to determine the first and then the next direction to divide. In order to avoid unequal separation at different coordinates, the algorithm is designed so that a coordinate is re-split only after all other coordinates have been selected. The algorithm stops when the standard deviation in all subdivisions obtained after division has reached the predetermined accuracy. Thus, an approximation of the integral with the MC approach is obtained. This algorithm has been used to calculate the relevant sensitivity indices in the study of the effect of chemical reaction rate constants on the concentrations of some pollutants (eg ozone). The computational complexity is proportional to the sample size n and the number of input parameters.

Adaptive strategy Georgieva [2010] is well known method for evaluation of multidimensional integrals, especially when the integrand function has peculiarities and peaks. Adaptive Monte Carlo methods proposed by Lautrup use a "sequence of refinements" of the original area selected so that the computations to be concentrated in subdomains with computational difficulties. There are various adaptive strategies depending on the technique of adaptation. Our adaptive algorithm (simple adaptive Monte Carlo algorithm) does not use any a priori information about the smoothness of the integrand, but it uses a posteriori information for the variance obtained during calculations. The main idea is a concentration of random points in the subregions where the variance is large (in terms of a preliminary given accuracy), i.e. the approach is based on a recursive partitioning of the integration area using a posteriori error information for the current partition. Let p_j and I_{Ω_j} are the following expressions: $p_j = \int_{\Omega_j} p(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and $I_{\Omega_j} = \int_{\Omega_j} f(\mathbf{x})p(\mathbf{x})\,\mathrm{d}\mathbf{x}$. Consider now a random point $\xi^{(j)} \in \Omega_j$ with a density function $p(\mathbf{x})/p_j$. In this case $I_{\Omega_j} = \mathbf{E}\left[\frac{p_j}{N}\sum_{i=1}^{N} f(\xi_i^{(j)})\right] = \mathbf{E}\theta_N$. This adaptive algorithm gives an approximation with an error $\varepsilon \leq c\, N^{-1/2}$, where $c \leq 0.6745\sigma(\theta)$ ($\sigma(\theta)$ is the standard deviation). From the estimation of the error, it can be concluded that, in general, the simple adaptive Monte Carlo algorithm gives an error less than the error of the Plain Monte Carlo algorithm, but the order is the same. The adaptive MC algorithm applied here is described below.

Algorithm

1. **Input data**: *total number of points $N1$, constant $M = 4$(the initial number of subregions taken), constant ε (max value of the variance in each subregion), constant δ (maximal admissible number of subregions), d-dimensionality of the initial region/domain, f - the function of interest.*

   1.1. **Calculate** *the number of points to be taken in each subregion $N = N1/\delta$.*

2. **For** $j = 1,\ M^d$:

   2.1. **Calculate** *the approximation of I_{Ω_j} and the variance $\mathbf{D}_{\Omega_j}$ in subdomain Ω_j based on N independent realizations of random variable θ_N;*

   2.2. **If** $(\mathbf{D}_{\Omega_j} \geq \varepsilon)$ **then**

      2.2.1. **Choose** *the axis direction on which the partition will perform,*

      2.2.2. **Divide** *the current domain into two (G_{j_1}, G_{j_2}) along the chosen direction,*

      2.2.3. **If** *the length of obtained subinterval is less than δ* **then go to** *step 2.2.1* **else** *$j = j_1$ G_{j_1} is the current domain right and* **go to** *step 2.1;*

2.3. Else if $(\mathbf{D}_{\Omega_j} < \varepsilon)$ *but an approximation of* $I_{G_{j_2}}$ *has not been calculated yet*, **then** $j = j_2$ G_{j_2} *is the current domain along the corresponding direction right and* **go to** *step 2.1;*

2.4. Else if $(\mathbf{D}_{\Omega_j} < \varepsilon)$ *but there are subdomains along the other axis directions,* **then go to** *step 2.1;*

2.5. Else *Accumulation in the approximation* I_N *of* I*.*

Computational complexity

Let N be the dimensionality of the problem under consideration. First let's describe briefly the Crude Monte Carlo algorithm. Let ξ be a random point with probability density function $p(x)$. Introducing the random variable

$$\theta = f(\xi)$$

with mathematical expectation equal to the value of the integral I_{G_j}, then

$$E\theta = \int_{G_j} f(x)p(x)dx.$$

Let $\xi_1, \xi_2, \ldots, \xi_N$ be independent realizations of the random point ξ with probability density function $p(x)$ and $\theta_1 = f(\xi_1), \ldots, \theta_N = f(\xi_N)$. Then an approximate value of I_{G_j} is

$$\hat{\theta}_N = \frac{1}{N}\sum_{i=1}^{N}\theta_i.$$

One can easily see that the computational complexity of the Crude Monte Carlo is linear, because in this simple case we have to choose N random points in the domain and every such choice is at the cost of $\mathcal{O}(1)$ operations. One single evaluation of the function in any of these points is also at the cost of $\mathcal{O}(1)$ operations.

In the adaptive Monte Carlo algorithm we are doing the same number of operations as in the Crude Monte Carlo algorithm. For the simple case when we have the two dimensional case ($N = 2$) and on the first step in the optimized adaptive approach we have $M = 4$ subdomains in our optimized Adaptive approach and

$$\hat{\theta}_N = \frac{1}{N_1}\sum_{i=1}^{N_1}\theta_i + \frac{1}{N_2}\sum_{i=1}^{N_2}\theta_i + \frac{1}{N_3}\sum_{i=1}^{N_3}\theta_i + \frac{1}{N_4}\sum_{i=1}^{N_4}\theta_i,$$

where $N_1 + N_2 + N_3 + N_4 = N$, so we have the same number of operations as the Crude Monte Carlo, which computational complexity is linear, to evaluate an approximation of I_{G_j}.

In general case for the adaptive algorithm the computational complexity depends on the integrand. First, let's consider the worst case. We always have a domain(area) with variance greater than the parameter ε and we need to divide this domain to 2^N subdomains. Additionally, for each of these 2^N newly obtained domains, we have to choose N random points into it and every such choice is at the cost of $\mathcal{O}(N)$ operations. We will receive complexity $\mathcal{O}(N.2^N)$.

So we choose only $\mathcal{O}(1)$ subdomains where the variance is greater than the parameter ε and this is independent of N. When we divide the domain on every step adaptiveness is not in all subdomains, but only in $\mathcal{O}(1)$ subdomains. At the beginning we have to choose $\frac{N}{k_0}$ random points. After that when dividing the domain into 2^N subdomains, we choose only $\mathcal{O}(1)$ subdomains, this choice is again independent of N. In these subdomains we choose $\frac{N}{k_1}$ points. On the j^{th} step of the Adaptive approach we choose $\mathcal{O}(1)$ subdomains with $\frac{N}{k_j}$ points. We have that $\sum\limits_{j=0}^{i} \frac{1}{k_j} = 1$. Therefore for the computational complexity we obtain

$$\frac{N}{k_0} + \mathcal{O}(1)\frac{N}{k_1} + \cdots + \mathcal{O}(1)\frac{N}{k_i} =$$

$$= N\mathcal{O}(1)\left(\sum_{j=0}^{i}\frac{1}{k_j}\right) = N\mathcal{O}(1) = \mathcal{O}(N).$$

In this way we can conclude that the computational complexity of the optimized Adaptive algorithm is linear.

Two adaptive approaches ADAPT1 (M=1) and ADAPT2(M=2) will be compared with Fibonacci besed lattice rule (FIBO) and Latin hypercube sampling (LHS).

2 Sensitivity Studies with Respect to Emission Levels

In the huge output data stream of UNI-DEM are the mean monthly concentrations of more than 30 pollutants. We consider 2 of them: *ozone* (O_3) and *ammonia* (NH_3). In particular, we present some results of a sensitivity studiy of the mean monthly concentrations of ammonia.

Here we present some results of our research on the sensitivity of UNI-DEM output (in particular, the ammonia mean monthly concentrations) with respect to the anthropogenic emissions input variation. The anthropogenic emissions input consists of 4 different components $\mathbf{E} = (\mathbf{E^A}, \mathbf{E^N}, \mathbf{E^S}, \mathbf{E^C})$ as follows:

$\mathbf{E^A}$ – ammonia (NH_3);
$\mathbf{E^N}$ – nitrogen oxides $(NO\ +\ NO_2)$;
$\mathbf{E^S}$ – sulphur dioxide (SO_2);
$\mathbf{E^C}$ – anthropogenic hydrocarbons.

The domain is the 4-dimensional hypercube $[0.5, 1]^4$). Polynomials of 2-nd degree have been used as an approximation tool Georgieva [2010]. The input data have been generated by the improved version SA-DEM code, specialized for sensitivity studies (see the previous section).

The results for relative errors for evaluation of the quantities f_0, total variances and first-order and total sensitivity indices using various stochastic approaches for numerical integration are presented in Tables 1, 2, 3, respectively. The quantity f_0 is presented by 4-dimensional integral whereas the rest of quantities under consideration are presented by 8-dimensional integrals following the ideas of *correlated sampling* technique to compute sensitivity measures in a reliable way Georgieva [2010].

The results in Table 1 show that the algorithms using generalized Fibonacci numbers and LHS simulate the behaviour of the Adaptive Monte Carlo algorithm, but for higher dimensions their

Table 1: Relative error for evaluation of $f_0 \approx 0.048$.

# of samples n	ADAPT1 Relative error	ADAPT2 Relative error	FIBO Relative error	LHS Relative error
2^{10}	1.88e-03	1.03e-03	2.09e-04	5.37e-04
2^{12}	2.05e-04	5.05e-04	4.32e-05	2.27e-04
2^{14}	1.83e-04	1.38e-05	2.25e-05	6.28e-05
2^{16}	9.89e-05	4.05e-04	8.70e-06	7.74e-05
2^{18}	3.95e-05	3.83e-06	1.79e-06	3.80e-06
2^{20}	4.99e-05	2.93e-05	4.21e-07	7.16e-06

Table 2: Relative error for evaluation of the total variance $\mathbf{D} \approx 0.0002$.

# of samples n	ADAPT1 Relative error	ADAPT2 Relative error	FIBO Relative error	LHS Relative error
2^{10}	1.56e-03	4.76e-03	1.63e-01	1.74e-02
2^{12}	2.58e-03	4.28e-04	2.39e-02	1.04e-02
2^{14}	6.03e-04	2.79e-04	2.90e-03	1.04e-02
2^{16}	1.83e-04	5.12e-04	2.65e-04	3.65e-04
2^{18}	5.77e-05	1.21e-04	3.01e-04	1.21e-05
2^{20}	3.42e-05	3.28e-05	1.19e-04	5.96e-05

efficiency decreases. The particular case study confirms the conclusion that these algorithms are suitable and more efficient for smooth functions with relatively low dimensions. From Tables 1 and 2 we can conclude that all stochastic approaches under consideration give reliable relative errors for sufficiently large number of samples. This is not the case for some sensitivity indices, which are very small by absolute value (see Table 3), but fortunately, these are of low importance too. The most efficient in terms of computational complexity is the FIBO algorithm, followed very closely by the LHS algorithm. Adaptive algorithm gives results of the same order as LHS and FIBO, and sometimes even outperforms them – see for example the relative errors for S_1^{tot} in Table 3.

Most influential emissions about ammonia output concentrations are ammonia emissions themselves (about 89% for Milan). The second most influential emissions about ammonia output are sulphur dioxide emissions (about 11%) - see Fig. 1. Pie charts representation of first- and second-order sensitivity indices of the ammonia in Milan on Fig. 1 has been obtained applying the correlated sampling of SobolâĂŹ variance-based approach for multidimensional sensitivity analysis for computing all possible sensitivity measures to study influence of four chosen groups of air pollutant emissions over the concentration of the three important air pollutants.

Table 3: Relative error for estimation of sensitivity indices of the input anthropogenic emissions by using various Monte Carlo and quasi-Monte Carlo approaches ($n = 2^{16} = 65536$).

Est. quantity	Ref. value	**ADAPT1**	**ADAPT2**	**FIBO**	**LHS**
S_1	9**e**-01	7.67e-04	1.22e-03	3.62e-04	9.79e-03
S_2	2**e**-04	1.47e-03	4.96e-02	1.74e-01	6.60e-01
S_3	1**e**-01	4.11e-03	1.59e-03	3.22e-03	8.65e-03
S_4	4**e**-05	1.04e-01	1.69e-01	4.87e-01	6.70e-01
S_1^{tot}	9**e**-01	4.99e-05	5.36e-05	4.61e-04	4.31e-04
S_2^{tot}	2**e**-04	5.23e-01	5.00e+00	3.45e-01	2.94e+01
S_3^{tot}	1**e**-01	1.15e-02	1.28e-02	1.96e-03	1.10e-02
S_4^{tot}	5**e**-05	1.88e+01	3.43e+01	5.06e+01	2.41e+02

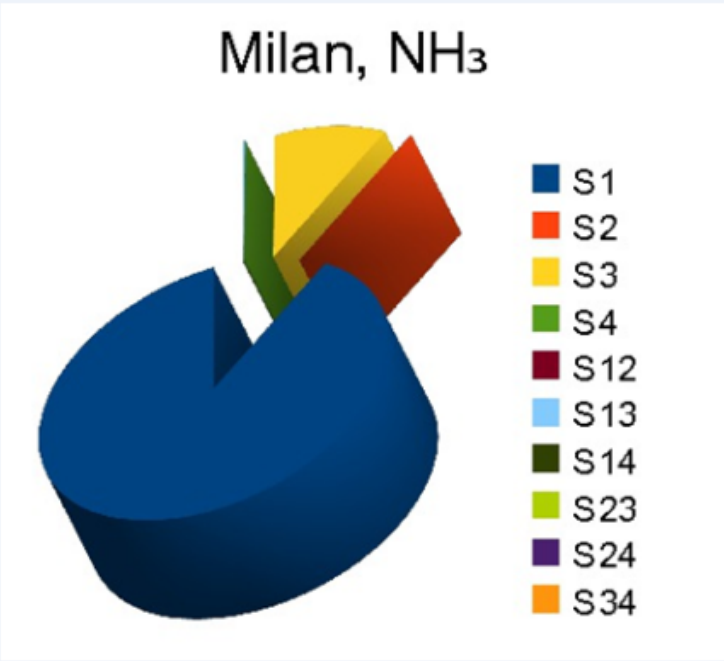

Figure 1: Pie charts representation of first- and second-order sensitivity indices of the ammonia in Milan

Sensitivity Studies with Respect to Chemical Reactions Rates

We will also study the sensitivity of the ozone concentration values in the air over Genova with respect to the rate variation of some chemical reactions of the condensed CBM-IV scheme (Zlatev [1995]), namely: ## $1, 3, 7, 22$ (time-dependent) and $27, 28$ (time independent). The simplified chemical equations of those reactions are as follows:

$[\#\mathbf{1}]\ NO_2 + h\nu \Longrightarrow NO + O;$
$[\#\mathbf{3}]\ O_3 + NO \Longrightarrow NO_2;$
$[\#\mathbf{7}]\ NO_2 + O_3 \Longrightarrow NO_3;$
$[\#\mathbf{22}]\ HO_2 + NO \Longrightarrow OH + NO_2;$
$[\#\mathbf{27}]\ HO_2 + HO_2 \Longrightarrow H_2O_2;$
$[\#\mathbf{28}]\ OH + CO \Longrightarrow HO_2.$

The domain under consideration is the 6-dimensional hypercube $[0.6, 1.4]^6$). Polynomials of second degree have been used for approximation again (see Georgieva [2010]).

Homma and Saltelli discuss in Homma and Saltelli [1996] which is the better estimation of $f_0^2 = \left(\int_{U^d} f(\mathbf{x})\mathrm{d}\mathbf{x}\right)^2$ in the expression for total variance and Sobol global sensitivity measures. In case of estimating sensitivity indices of a fixed order, the formula

$$f_0^2 \approx \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,d})\ f(\mathbf{x}'_{i,1}, \ldots, \mathbf{x}'_{i,d}),$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two independent sample vectors, is better (as recommended in Georgieva [2010]).

Table 4: Relative error for evaluation of $f_0 \approx 0.27$.

# of samples n	**ADAPT1** Relative error	**ADAPT2** Relative error	**FIBO** Relative error	**LHS** Relative error
2^{10}	2.74e-04	3.21e-04	2.08e-03	3.73e-04
2^{12}	9.55e-05	4.43e-05	1.40e-04	2.41e-04
2^{14}	1.20e-04	5.64e-05	3.98e-04	7.53e-05
2^{16}	3.49e-05	3.72e-05	2.61e-04	2.02e-04

The relative errors for evaluation of the quantities f_0, total variances, first and second order sensitivity indices by using various stochastic approaches for numerical integration are presented in Tables 4, 5, 6 respectively. Here the quantity f_0 is presented by a 6-dimensional integral, whereas the total variance and the sensitivity indices are presented by 12-dimensional integrals, following the ideas of *correlated sampling*.

Table 5: Relative error for evaluation of the total variance $\mathbf{D} \approx 0.0025$.

# of samples n	**ADAPT1** Relative error	**ADAPT2** Relative error	**FIBO** Relative error	**LHS** Relative error
2^{10}	9.67e-04	1.18e-03	6.73e+00	1.91e-02
2^{12}	9.10e-04	2.24e-03	5.27e-01	9.99e-02
2^{14}	1.40e-04	1.86e-04	1.02e-01	1.62e-02
2^{16}	3.01e-05	1.48e-04	1.97e-03	3.56e-05

Table 6: Relative error for estimation of sensitivity indices of several chemical reaction rate parameters by using various Monte Carlo and quasi-Monte Carlo approaches ($n = 2^{16} = 65536$).

Est. quantity	Ref. value	**ADAPT1**	**ADAPT2**	**FIBO**	**LHS**
S_1	4**e**-01	1.55e-04	3.48e-04	3.82e-02	3.04e-02
S_2	3**e**-01	4.34e-04	1.58e-04	1.03e-02	7.35e-04
S_3	5**e**-02	3.42e-04	8.09e-05	5.48e-01	2.33e-02
S_4	3**e**-01	4.75e-04	9.04e-04	1.07e-02	2.47e-02
S_5	4**e**-07	1.31e+01	1.07e+01	3.40e+03	9.25e+02
S_6	2**e**-02	1.08e-03	4.54e-04	1.32e+00	3.81e-02
S_{12}	6**e**-03	1.30e-02	7.92e-03	3.21e+00	8.99e-02
S_{14}	5**e**-03	5.30e-03	1.81e-03	8.64e+00	2.74e-01
S_{15}	8**e**-06	9.34e+02	9.34e+02	9.19e+02	9.21e+02
S_{24}	3**e**-03	1.26e-03	7.24e-03	1.37e+01	7.10e-01
S_{45}	1**e**-05	9.93e-02	8.55e-02	4.25e+01	1.05e+01

Quasi-MC lattice rule based on generalized Fibonacci numbers and Latin hypercube sampling produce better results for 6-dimensional integrals in comparison with 12-dimensional integrals. It is clear that with the increasing the dimensionality of the integral Adaptive method produce more accurate results than both FIBO and LHS. Adaptive Monte Carlo algorithm gives better results in case of higher dimensional integrals and lower number of samples. For most of the sensitivity indices Adaptive Monte Carlo algorithm gives more accurate results than FIBO and LHS by at least 2 orders of degree.

The representation of the reference values of the first-order and second-order sensitivity indices of input parameters (for ozone concentrations) are given on Fig.2. One can observe that the second-order sensitivity indices take rather small portions. This fact shows that the UNI-DEM is additive according to the rates of the six chemical reactions studied in this work.

Conclusion

This study focuses on environmental safety. Sensitivity analysis, and in particular the results obtained, play an extremely important two-sided role: to test and improve mathematical models, and, on the other hand, to reliably interpret the numerical results by relevant specialists. Dispersion-based analysis is a very effective tool for in-depth study of the relationship between individual parameters, outputs and internal mechanisms governing the system in question. By identifying the major chemical reactions that affect the behavior of the system, specialists in various fields of application (physics, chemistry) will be able to obtain valuable information about improving the model, which in turn will increase the reliability and sustainability of forecasts. Thus, through a sensitivity analysis, the mathematical model will help to make more accurate estimates of the effects of harmful emissions on human health and agricultural losses. The results obtained show that the stochastic adaptive approach developed is one of the most efficient methods based on reducing the variance in terms of computational efficiency and accuracy.

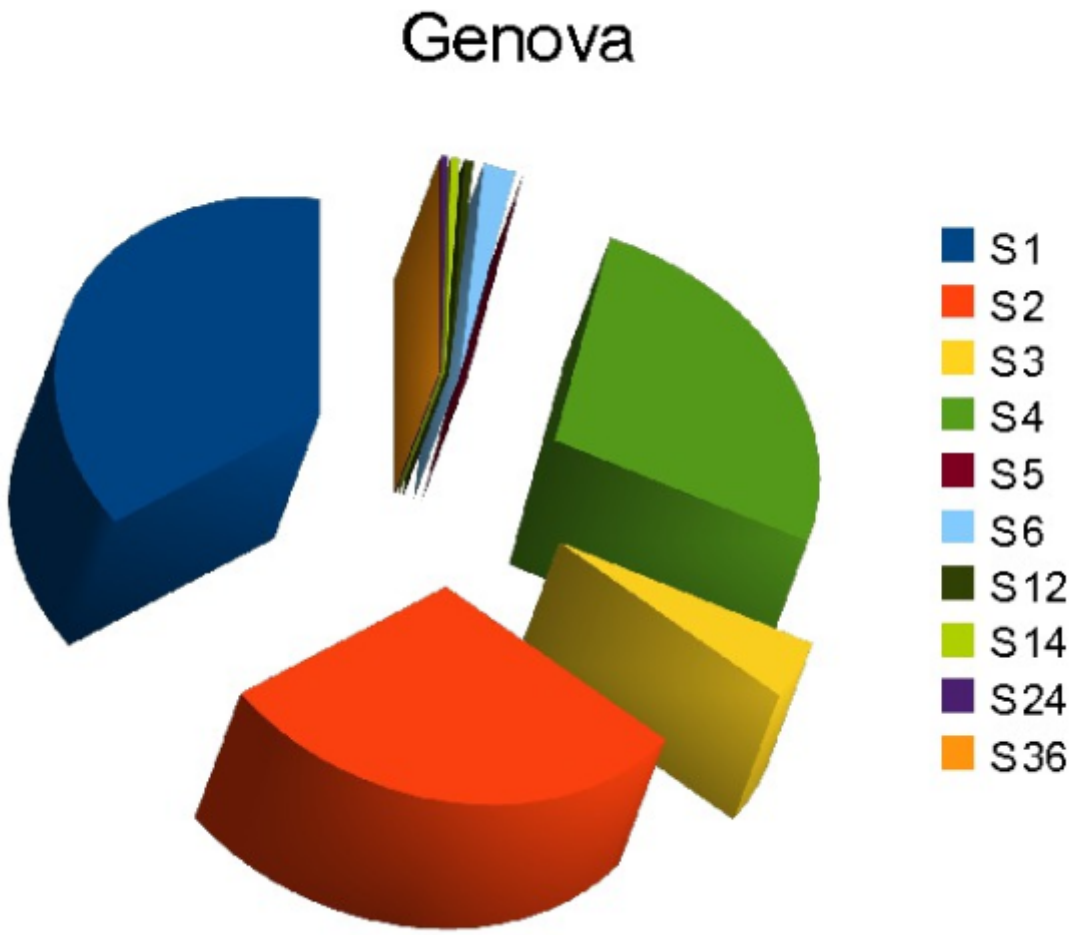

Figure 2: Pie charts representation of first- and second-order sensitivity indices of the ozone in Genova

Acknowledgements

Venelin Todorov is supported by the National Scientific Program "Young scientists and Postdoctoral candidates" 2020-2021 and by the National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICT in SES)", contract No DO1-205/23.11.2018, financed by the Ministry of Education and Science in Bulgaria.

Bibliography

I. Dimov, R. Georgieva (2010). Monte Carlo algorithms for evaluating Sobol sensitivity indices. Mathe- matics and Computers in Simulation, 81 (3), 506-514.

Dimov I., Karaivanova A., Georgieva R., Ivanovska S. Parallel Importance Separation and Adaptive Monte Carlo Algorithms for Multiple Integrals, Springer Lecture Notes in Computer Science, 2542, Springer-Verlag, Berlin, Heidelberg, 2003, New York: 99–107.

R. Georgieva, PhD thesis, Sofia, 2010.

McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2), 239âĂŞ45 (1979)

T. Homma, A. Saltelli, Importance Measures in Global Sensitivity Analysis of Nonlinear Models, Reliability Engineering and System Safety 52, 1996, 1–17.

S. Poryazov. A suitable unit of sensitivity in telecommunications. TELECOM 2011, 13-14.10.2011, Sofia, ISSN 1314-2690, p. 165–172.

Z. Zlatev, Computer treatment of large air pollution models, 1995.

Y. Wang and F. J. Hickernell, *An historical overview of lattice point sets*, 2002

Z. Zlatev, I. T. Dimov, Computational and Numerical Challenges in Environmental Modelling, Elsevier, Amsterdam, 2006.

Authors' Information

Venelin Todorov *- Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Information Modeling Department, Acad. Georgi Bonchev Str., Block 8, Sofia 1113, Bulgaria;*
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences,
Department of Parallel Algorithms, Acad. Georgi Bonchev Str., Block 25A, Sofia 1113, Bulgaria;
e-mail: vtodorov@math.bas.bg,venelin@parallel.bas.bg
Major Fields of Scientific Research: Monte Carlo methods, Sensitivity Analysis Multi-dimensional integrals, Digital nets, Lattice sequences