Sensitivity Analysis of an Air Pollution Model by Using Quasi-Monte Carlo Algorithms for Multidimensional Numerical Integration

Tzvetan Ostromsky[1(✉)], Ivan Dimov[1], Venelin Todorov[1,2], and Zahari Zlatev[3]

[1] Department of Parallel Algorithms, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences (IICT-BAS), Acad. G. Bonchev 25 A, 1113 Sofia, Bulgaria
{ceco,venelin}@parallel.bas.bg, ivdimov@bas.bg

[2] Department of Information Modelling, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (IMI-BAS), Acad. Georgi Bonchev Street, Block 8, 1113 Sofia, Bulgaria
vtodorov@math.bas.bg

[3] National Centre for Environment and Energy, University of Århus, Frederiksborgvej 399, P.O. Box 358, 4000 Roskilde, Denmark
zz@dmu.dk

Abstract. Sensitivity analysis is a powerful tool for studying and improving the reliability of large and complicated mathematical models. Air pollution and meteorological models are in front places among the examples of such models, with a lot of natural uncertainties in their input data sets and parameters. We present here some results of our global sensitivity study of the Unified Danish Eulerian Model (UNI-DEM). One of the most attractive features of UNI-DEM is its advanced chemical scheme – the Condensed CBM IV, which consider in detail a large number of chemical species and numerous reactions between them.

Four efficient stochastic algorithms (Sobol QMC, Halton QMC, Fibonacci lattice rule and Latin hypercube sampling) have been used and compared by their accuracy in studying the sensitivity of ammonia and ozone concentration results with respect to the emission levels and some chemical reactions rates. The numerical experiments show that the stochastic algorithms under consideration are quite efficient for this purpose, especially for evaluating the contribution of small by value sensitivity indices.

1 Introduction

We discuss a systematic approach for sensitivity analysis studies in the area of air pollution modelling. The Unified Danish Eulerian Model (UNI-DEM) [15,16] 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

G. Nikolov et al. (Eds.): NMA 2018, LNCS 11189, pp. 281–289, 2019.
https://doi.org/10.1007/978-3-030-10692-8_31

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.

Four efficient stochastic algorithms (Sobol QMC, Halton QMC, Fibonacci lattice rule and Latin hypercube sampling) have been applied to sensitivity studies of concentration variations of air pollutants with respect to emission levels and some chemical reactions rates. More information on Sobol QMC algorithm can be found in [1]. For generating Sobol quasirandom sequences we use an adaption of INSOBL and GOSOBL routines, implemented respectively in ACM TOMS Algorithm 647 [7] and ACM TOMS Algorithm 659 [2]. The original code can only compute the "next" element of the sequence. The adapted code allows the user to specify the index of the desired element. The Halton sequence is completely described in [8,9]. Fibonacci lattice rule and Latin hypercube sampling are described in detail in our previous paper [6].

2 Description and Implementation of UNI-DEM and Its Sensitivity Analysis Version

UNI-DEM is a powerful large-scale air pollution model for calculating the concentrations of a large number of pollutants and other chemical species in the air, involved in chemical reactions with the pollutants. Among the most useful output results are the mean values of the pollutants' concentrations for certain time period (day, month, year). Other accumulative functions related to them as well as the peak values, are also calculated. These can be used in various application areas (environmental protection, agriculture, health care, etc.).

UNI-DEM is mathematically represented by the following system (1) of partial differential equations (PDE), in which the unknown concentrations c_s of a number of chemical species in the air (pollutants and other chemically active components) must be calculated. The main physical and chemical processes (advection, diffusion, chemical reactions, emissions and deposition) are represented in that system. It is computated in a large spatial domain (4800×4800 km.), which covers completely the European continent and the Mediterranean. Some typical background concentrations (which are varied both seasonally and diurnally) are used for boundary conditions. The large size of the computational domain and the fact that its west and north boundaries (from where the predominating winds blow) are above the ocean (where the concentrations of most pollutants are, in general, stable and much lower than over the continent) deminishes their effect on the results inside the domain. The I/O data arrays are structured by months, so the output concentrations at the end of an already calculated month are used as initial conditions for the next one. Initially, when there is no such data, calculations begin with a 5-day start-up period with

some background initial concentrations and meteo data from the previous month in order to set up the inititial conditions.

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

c_s denote 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 and wet deposition coefficients respectively; $Q_s(c_1, c_2, \ldots c_q)$ – 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) [16]. 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)**.

Spatial and time discretization makes each of the submodels a tough 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 [11–13]. It consists of the 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).

Significant improvements of the earlier versions of SA-DEM were made by introducing two additional levels of parallelism: top-level(MPI) and bottom-level(OpenMP). They allow us to use efficiently the computational power of the contemporary cluster supercomputers with multicore nodes. Other important improvement in the data management strategy reduced the number of I/O

Table 1. Time (T) and speed-up **Sp** of SA-DEM (MPI only) on the Spanish supercomputer IBM MareNostrum III at BSC, Barcelona

#CPU	#Nodes	Advection		Chemistry		TOTAL		
		T [s]	**Sp**	T [s]	**Sp**	T [s]	**Sp**	E [%]
10	1	83460	**10**	77273	**10**	171707	**10**	100%
40	3	19448	**43**	16946	**46**	40471	**42**	106%
80	5	9874	**85**	9047	**85**	22261	**77**	96%
160	10	5250	**159**	4562	**169**	12875	**133**	83%
320	20	2895	**288**	2403	**322**	8233	**209**	65%
640	40	1522	**548**	1269	**609**	5387	**319**	50%
960	60	1215	**687**	822	**940**	4075	**421**	44%

operations and pipelined most of them with the computationally intensive stages, reducing significantly the CPU idle time in the parallel MPI processes.

In Table 1 we show some scalability results from experiments with SA-DEM on one of the largest supercomputers in Europe – IBM MareNostrum III (in BSC, Barcelona, Spain). It consists of 3028 nodes IBM dx360 M4 (16 core) with 32 GB RAM per node. It is seen from Table 1 that the chemical stage (the most computationally expensive) scales very well (shows almost linear speed-up in the whole range of experiments). Advection stage scales pretty well in most of the experiments, with understandable slow-down in the highly parallel experiments. It is caused by the significant boundary overlapping of the domain partitioning when approaching the inherent partitioning limitations. In general, SA-DEM performs quite efficiently and show relatively high scalability on such a large supercomputing system.

3 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 study of the mean monthly concentrations of ammonia in Milan.

In this section 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 under consideration is the 4-dimensional hypercube $[0.5, 1]^4$. Polynomials of 2-nd degree have been used as an approximation tool [5]. The input

data have been generated by the improved version of SA-DEM code, specialized for sensitivity studies (see the previous section).

Table 2. Relative error for the evaluation of $f_0 \approx 0.048$.

# samples n	Relative error			
	Sobol	Halton	FIBO	LHS
2^{10}	5.56e−04	3.15e−05	2.09e−04	5.37e−04
2^{12}	1.16e−04	1.14e−04	4.32e−05	2.27e−04
2^{14}	3.14e−05	1.27e−05	2.25e−05	6.28e−05
2^{16}	8.78e−06	8.20e−06	8.70e−06	7.74e−05
2^{18}	1.75e−06	2.40e−06	1.79e−06	3.80e−06
2^{20}	4.97e−07	1.03e−06	4.21e−07	7.16e−06

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 2, 3 and 4, respectively. The quantity f_0 is presented by 4-dimensional integral whereas the rest of quantities under consideration are presented by 2-dimensional integrals following the ideas of *correlated sampling* technique to compute sensitivity measures in a reliable way [10,14].

Homma and Saltelli discuss in [10] which of the two formulae below gives better estimation of $f_0^2 = \left(\int_{U^d} f(\mathrm{x})\mathrm{dx}\right)^2$ in the expression for total variance and Sobol global sensitivity measures. The first formula is

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

where x and x′ are two independent sample vectors, and the second one is

$$f_0^2 \approx \left\{\frac{1}{n}\sum_{i=1}^{n} f(\mathrm{x}_{i,1},\ldots,\mathrm{x}_{i,d})\right\}^2 \tag{3}$$

In case of estimating sensitivity indices of a fixed order, the first formula (2) is better (as recommended in [10]).

The results in Table 2 show that the algorithms using generalized Fibonacci numbers and LHS simulate the behaviour of Sobol QMCA, but for higher dimensions their efficiency decrease. The particular case study confirms the conclusion that these algorithms are suitable and more efficient for smooth functions with comparatively low dimensions. From Tables 2 and 3 we can conclude that all stochastic approaches under consideration give reliable relative errors for sufficiently large number of samples. The most efficient in terms of computational

Table 3. Relative error for the evaluation of the total variance $\mathbf{D} \approx 0.0002$.

# samples n	Relative error			
	Sobol	Halton	FIBO	LHS
2^{10}	2.28e−03	1.40e−02	1.63e−01	1.74e−02
2^{12}	9.38e−04	7.81e−03	2.39e−02	1.04e−02
2^{14}	1.92e−04	1.77e−03	2.90e−03	1.04e−02
2^{16}	5.86e−05	5.96e−04	2.65e−04	3.65e−04
2^{18}	8.61e−06	1.48e−04	3.01e−04	1.21e−05
2^{20}	1.60e−06	4.77e−05	1.19e−04	5.96e−05

Table 4. Relative error for estimation of sensitivity indices of input parameters using various Monte Carlo and quasi-Monte Carlo approaches ($n \approx 65536$).

Sensit. index	Ref. value	Sobol	Halton	FIBO	LHS
S_1	9e−01	5.78e−06	2.95e−04	3.62e−04	9.79e−03
S_2	2e−04	1.52e−03	3.49e−02	1.74e−01	6.60e−01
S_3	1e−01	4.39e−05	2.30e−03	3.22e−03	8.65e−03
S_4	4e−05	2.87e−03	1.21e−01	4.87e−01	6.70e−01
S_1^{tot}	9e−01	5.19e−06	2.97e−04	4.61e−04	4.31e−04
S_2^{tot}	2e−04	1.36e−04	3.24e−02	3.45e−01	2.94e+01
S_3^{tot}	1e−01	4.65e−05	2.25e−03	1.96e−03	1.10e−02
S_4^{tot}	5e−05	1.57e−03	1.20e−01	5.06e−01	2.41e+02

complexity is the algorithm of Sobol, followed by Halton algorithm. The evaluated sensitivity measures, presented in the tables, are obtained either by multidimensional integrals (total variances) or by ratios of multidimensional integrals (Sobol global sensitivity indices). One can notice also from results in Table 4 that the order of relative error is different for different quantities of interest (see column *Reference value*) for the same sample size. It depends both on the integrand dimension and the magnitude of estimated quantity. The algorithms using generalized Fibonacci numbers and LHS are characterized with unreliable relative errors for small in value sensitivity measures.

4 Sensitivity Studies with Respect to Chemical Reactions Rates

Another part of our research was to 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 [15], namely: ## $1, 3, 7, 22$

(time-dependent) and 27, 28 (time independent). The simplified chemical equations of those reactions are as follows:

[#**1**] $NO_2 + h\nu \Longrightarrow NO + O$;
[#**3**] $O_3 + NO \Longrightarrow NO_2$;
[#**7**] $NO_2 + O_3 \Longrightarrow NO_3$;
[#**22**] $HO_2 + NO \Longrightarrow OH + NO_2$;
[#**27**] $HO_2 + HO_2 \Longrightarrow H_2O_2$;
[#**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 [4]).

Table 5. Relative error for the evaluation of $f_0 \approx 0.27$.

# samples n	Relative error			
	Sobol	Halton	FIBO	LHS
2^{10}	1.62e−04	1.60e−04	2.08e−03	3.73e−04
2^{12}	4.54e−05	5.55e−05	1.40e−04	2.41e−04
2^{14}	3.59e−06	2.70e−05	3.98e−04	7.53e−05
2^{16}	4.70e−06	1.60e−06	2.61e−04	2.02e−04
2^{18}	5.90e−07	1.02e−06	7.29e−06	2.82e−05
2^{20}	1.36e−07	5.56e−07	4.57e−07	1.04e−05

Table 6. Relative error for the evaluation of the total variance $\mathbf{D} \approx 0.0025$.

# samples n	Relative error			
	Sobol	Halton	FIBO	LHS
2^{10}	5.75e−03	4.86e−02	6.73e+00	1.91e−02
2^{12}	2.43e−03	1.25e−03	5.27e−01	9.99e−02
2^{14}	9.90e−05	1.65e−03	1.02e−01	1.62e−02
2^{16}	5.81e−05	4.34e−04	1.97e−03	3.56e−05
2^{18}	7.71e−06	3.79e−04	4.53e−03	7.78e−03
2^{20}	1.75e−06	3.34e−05	9.33e−03	2.78e−04

The relative errors for evaluation of the quantities f_0, total variances, first-order and total sensitivity indices using various stochastic approaches for numerical integration are presented in Tables 5, 6 and 7 respectively. The quantity f_0 is presented by 6-dimensional integral, whereas the rest of the quantities under consideration are presented by 2-dimensional integrals, following the ideas of *correlated sampling*.

From these tables we can see that Sobol QMCA gives better results than Halton QMCA and the difference is 1–2 orders. 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. More results in favour of this conclusion can be found in [3].

Table 7. Relative error for estimation of sensitivity indices of input parameters using various Monte Carlo and quasi-Monte Carlo approaches ($n \approx 65536$).

Sensist. index	Ref. value	Sobol	Halton	FIBO	LHS
S_1	4e−01	1.83e−04	2.87e−03	3.82e−02	3.04e−02
S_2	3e−01	2.69e−05	3.76e−03	1.03e−02	7.35e−04
S_3	5e−02	1.08e−04	7.27e−03	5.48e−01	2.33e−02
S_4	3e−01	1.37e−04	2.19e−03	1.07e−02	2.47e−02
S_5	4e−07	2.69e−01	3.68e+01	3.40e+03	9.25e+02
S_6	2e−02	2.81e−03	1.30e−02	1.32e+00	3.81e−02
S_1^{tot}	4e−01	1.39e−04	2.79e−03	7.92e−02	2.03e−02
S_2^{tot}	3e−01	4.32e−05	3.26e−03	3.06e−02	1.45e−02
S_3^{tot}	5e−02	1.08e−04	6.43e−03	1.31e+00	1.55e−01
S_4^{tot}	3e−01	3.77e−04	2.11e−03	3.84e−01	1.11e−02
S_5^{tot}	2e−04	1.40e−03	1.38e−02	8.85e+01	1.45e+01
S_6^{tot}	2e−02	1.29e−05	1.04e−02	2.15e+00	9.75e−01
S_{12}	6e−03	6.03e−04	7.92e−03	3.21e+00	8.99e−02
S_{14}	5e−03	2.17e−03	9.12e−03	8.64e+00	2.74e−01
S_{15}	8e−06	9.33e+02	9.36e+02	9.19e+02	9.21e+02
S_{24}	3e−03	4.97e−04	1.83e−02	1.37e+01	7.10e−01
S_{45}	1e−05	1.48e−02	9.08e−01	4.25e+01	1.05e+01

Acknowledgements. The authors would like to thank Rayna Georgieva for her help. This work is supported by the Bulgarian Academy of Sciences through the Program for Career Development of Young Scientists, Grant DFNP-17-88 /28.07.2017; as well as by the Bulgarian NSF under Projects DN 12/4 -2017 and DN 12/5 -2017.

References

1. Antonov, I., Saleev, V.: An economic method of computing LP_τ-sequences. USSR Comput. Math. Phy. **19**, 252–256 (1979)
2. Bratley, P., Fox, B.: Algorithm 659: implementing Sobol's quasirandom sequence generator. ACM Trans. Math. Softw. **14**(1), 88–100 (1988)
3. Dimitrov, S., Dimov, I., Todorov, V.: Latin hypercube sampling and fibonacci based lattice method comparison for computation of multidimensional integrals. In: Dimov, I., Faragó, I., Vulkov, L. (eds.) NAA 2016. Lecture Notes in Computer Science, vol. 10187, pp. 296–303. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57099-0_32
4. Dimov, I., Georgieva, R., Ostromsky, Tz, Zlatev, Z.: Variance-based sensitivity analysis of the unified danish eulerian model according to variations of chemical rates. In: Dimov, I.T., Faragó, I., Vulkov, L. (eds.) NAA 2012. LNCS, vol. 8236, pp. 247–254. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41515-9_26

5. Dimov, I.T., Georgieva, R., Ostromsky, T., Zlatev, Z.: Sensitivity studies of pollutant concentrations calculated by UNI-DEM with respect to the input emissions. Centr. Eur. J. Math. **11**(8), 1531–1545 (2013). https://doi.org/10.2478/s11533-013-0256-2. Numerical Methods for Large Scale Scientific Computing
6. Dimov, I.T., Georgieva, R., Todorov, V., Ostromsky, Tz.: Efficient stochastic approaches for sensitivity studies of an Eulerian large-scale air pollution model. In: AIP Conference Proceedings, vol. 1895, no. 1, p. 050004 (2017). https://doi.org/10.1063/1.5007376
7. Fox, B.: Algorithm 647: implementation and relative efficiency of quasirandom sequence generators. ACM Trans. Math. Softw. **12**(4), 362–376 (1986)
8. Halton, J.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik **2**, 84–90 (1960)
9. Halton, J., Smith, G.B.: Algorithm 247: radical-inverse quasi-random point sequence. Commun. ACM **7**, 701–702 (1964)
10. Homma, T., Saltelli, A.: Importance measures in global sensitivity analysis of nonlinear models. Reliab. Eng. Syst. Saf. **52**, 1–17 (1996)
11. Ostromsky, T.z., Dimov, I.T., Georgieva, R., Zlatev, Z.: Air pollution modelling, sensitivity analysis and parallel implementation. Int. J. Environ. Pollut. **46**(1/2), 83–96 (2011)
12. Ostromsky, T.z., Dimov, I., Georgieva, R., Zlatev, Z.: Parallel computation of sensitivity analysis data for the Danish Eulerian model. In: Lirkov, I., Margenov, S., Waśniewski, J. (eds.) LSSC 2011. LNCS, vol. 7116, pp. 307–315. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29843-1_35
13. Ostromsky, T.z., Dimov, I.T., Marinov, P., Georgieva, R., Zlatev, Z.: Advanced sensitivity analysis of the Danish Eulerian Model in parallel and grid environment. In: Proceedings of 3rd International Conference on AIP, AMiTaNS 2011, vol. 1404, pp. 225–232 (2011)
14. Sobol, I.M., Tarantola, S., Gatelli, D., Kucherenko, S., Mauntz, W.: Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliab. Eng. Syst. Saf. **92**, 957–960 (2001)
15. Zlatev, Z.: Computer Treatment of Large Air Pollution Models. KLUWER Academic Publishers, Dorsrecht-Boston-London (1995)
16. Zlatev, Z., Dimov, I.T.: Computational and Numerical Challenges in Environmental Modelling. Elsevier, Amsterdam (2006)