Sensitivity Study of a Large-Scale Air Pollution Model by Using Optimized Stochastic Algorithm

Venelin Todorov
Bulgarian Academy of Sciences
Institute of Mathematics and Informatics
ul. G. Bonchev 8, 1113 Sofia, Bulgaria
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria
Email: vtodorov@math.bas.bg,venelin@parallel.bas.bg

Tzvetan Ostromsky
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria
Email: ceco@parallel.bas.bg

Ivan Dimov
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria
Email: ivdimov@bas.bg

Stefka Fidanova
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria
Email: stefka@parallel.bas.bg

Abstract—In this work a systematic procedure for multidimensional sensitivity analysis in the area of air pollution modeling by an optimized latin hypercube sampling has been done. The Unified Danish Eulerian Model (UNI-DEM) is used in our investigation, because this is one of the most advanced large-scale mathematical models that describes adequately all physical and chemical processes. 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. The results obtained with an adaptive approach and latin hypercube sampling has been discussed.

I. Introduction

HERE we discuss a systematic approach for sensitivity analysis studies in the area of air pollution modelling. The Unified Danish Eulerian Model (UNI-DEM) [14], [15] 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 [11], [13]. 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.

Sensitivity studies are nowadays applied to some of the most complicated mathematical models from various intensively developing areas of application ([1], [6]). Different efficient stochastic algorithms for multidimensional integration have also been applied on a further stage of these sensitivity studies. The two Adaptive Monte Carlo algorithms that we are going to used are based on the ideas and results of the importance separation and are completely described in the following works [2], [3], [4], [5].

II. Optimized Latin Hypercube Sampling

Latin Hypercube Sampling is a type of stratified sampling, described for the first time by M.D. McKay in 1979 in [9]. For this standard method we will use the notation LHS. Recently a renewing interest in optimization version of Latin hypercube sampling is described by Budiman Minasny in his works [10]. In the case of integral approximation we must simply divide the domain $[0,1]^d$ into m^d disjoint subdomains, each of volume $\frac{1}{m^d}$ and to sample one point from each of them. Let this sample be $\mathbf{x}_{k,j}$, for dimensions $k = 1, \ldots, m^d$, $j = 1, \ldots, d$. LHS does not require more samples for more dimensions (variables) - it is one of the main advantages of this scheme. Examples of random, stratified and Latin hypercube samplings with 16 points are presented on Figure 1 [8].

Here we propose an optimized version of the Latin Hypercube Sampling which includes a uniformly distributed random point in each interval which improves the accuracy as can be seen from the results below. The description of the optimized latin hypercube sampling (LHSO) algorithm is the following:

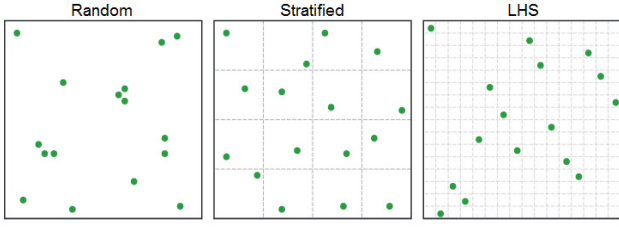1) Divide the distribution of each variable X into n equiprobable intervals.

Fig. 1. Comparison of random, stratified and latin hypercube samplings with 16 points (d = 2, M = 4).

2) For each interval, choose one uniform random number $r \in [0,1]$. Then, let

$$Prob_i = \frac{i-1}{n} + r * \frac{1}{n} \qquad (1)$$

3) Transform the probability into the sampled value x using the inverse of the cumulative distribution function:

$$x = F^{-1}(Prob) \qquad (2)$$

4) The n values obtained for each variable X are paired randomly(e.g. by generating a random permutation of $\{1, 2, \dots, n\}$ for each variable except one) with the n values of the other variables.

III. SENSITIVITY STUDIES WITH RESPECT TO EMISSION LEVELS

The huge output data stream of UNI-DEM contains 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.

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 $[0.5, 1]^4$). 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 I, II, III, respectively. The quantity f_0 is presented by a 4-dimensional integral whereas the rest of the quantities under consideration are presented by an 8-dimensional integrals. In this work for the estimateq quantity we use the notation EQ.

TABLE I
RELATIVE ERROR FOR THE EVALUATION OF $f_0 \approx 0.048$.

# of samples n	ADAPT1 Relative error	ADAPT2 Relative error	LHSO Relative error	LHS Relative error
2^{16}	9.89e-05	4.05e-04	1.45e-05	7.74e-05
2^{18}	3.95e-05	3.83e-06	1.98e-06	3.80e-06
2^{20}	4.99e-05	2.93e-05	6.31e-06	7.16e-06

TABLE II
RELATIVE ERROR FOR THE EVALUATION OF THE TOTAL VARIANCE $\mathbf{D} \approx 0.0002$.

# of samples n	ADAPT1 Relative error	ADAPT2 Relative error	LHSO Relative error	LHS Relative error
2^{16}	1.83e-04	5.12e-04	2.17e-04	3.65e-04
2^{18}	5.77e-05	1.21e-04	4.56e-05	1.21e-05
2^{20}	3.42e-05	3.28e-05	1.87e-05	5.96e-05

The results show that the algorithms using LHSO simulate the behaviour of the Adaptive Monte Carlo algorithm, and for some small in value sensitivity indices as S_4^{tot} LHSO gives better results than the adaptive approach. Adaptive algorithm gives results of the same order as LHS, and sometimes even outperforms it (see for example the relative errors for S_1^{tot}).

IV. 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 of variation of some chemical reactions of the condensed CBM-IV scheme ([14]), namely: # $1, 3, 7, 22$ (time-dependent) and # $27, 28$ (time independent). The simplified chemical equations of those reactions are as follows:

$$[\#1] \quad NO_2 + h\nu \Longrightarrow NO + O;$$
$$[\#3] \quad O_3 + NO \Longrightarrow NO_2;$$
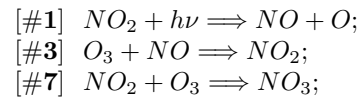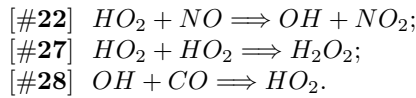$$[\#7] \quad NO_2 + O_3 \Longrightarrow NO_3;$$

TABLE III
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF THE INPUT ANTHROPOGENIC EMISSIONS BY USING VARIOUS MONTE CARLO APPROACHES ($n = 2^{16} = 65536$).

EQ	Ref. value	ADAPT1	ADAPT2	LHSO	LHS
S_1	9e-01	7.67e-04	1.22e-03	1.16e-04	9.79e-03
S_2	2e-04	1.47e-03	4.96e-02	7.10e-03	6.60e-01
S_3	1e-01	4.11e-03	1.59e-03	1.19e-03	8.65e-03
S_4	4e-05	1.04e-01	1.69e-01	6.10e-02	6.70e-01
S_1^{tot}	9e-01	4.99e-05	5.36e-05	7.31e-05	4.31e-04
S_2^{tot}	2e-04	5.23e-01	5.00e+00	1.07e-01	2.94e+01
S_3^{tot}	1e-01	1.15e-02	1.28e-02	5.45e-03	1.10e-02
S_4^{tot}	5e-05	1.88e+01	3.43e+01	7.54e-01	2.41e+02

$$[\#22] \quad HO_2 + NO \Longrightarrow OH + NO_2;$$
$$[\#27] \quad HO_2 + HO_2 \Longrightarrow H_2O_2;$$
$$[\#28] \quad OH + CO \Longrightarrow HO_2.$$

The domain under consideration is the 6-dimensional hypercube $[0.6, 1.4]^6$).

Homma and Saltelli discuss in [7] which is the better estimation of $f_0^2 = \left(\int_{U^d} f(\mathbf{x}) d\mathbf{x} \right)^2$ 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 [7], [12]).

TABLE IV
RELATIVE ERROR FOR EVALUATION OF $f_0 \approx 0.27$.

# of samples n	ADAPT1 Relative error	ADAPT2 Relative error	LHSR Relative error	LHS Relative error
2^{16}	3.49e-05	3.72e-05	1.12e-05	2.02e-04
2^{18}	5.90e-07	1.02e-06	6.14e-06	2.82e-05
2^{20}	1.36e-07	5.56e-07	8.71e-07	1.04e-05

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 IV, V, VI 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 V
RELATIVE ERROR FOR EVALUATION OF THE TOTAL VARIANCE
$\mathbf{D} \approx 0.0025$.

# of samples n	ADAPT1 Relative error	ADAPT2 Relative error	LHSO Relative error	LHS Relative error
2^{16}	3.01e-05	1.48e-04	8.23e-05	3.56e-05
2^{18}	7.71e-06	3.79e-04	6.12e-05	7.78e-03
2^{20}	1.75e-06	3.34e-05	5.21e-05	2.78e-04

TABLE VI
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF SEVERAL
CHEMICAL REACTION RATE PARAMETERS BY USING VARIOUS MONTE
CARLO AND QUASI-MONTE CARLO APPROACHES ($n = 2^{16} = 65536$).

EQ	Ref. value	ADAPT1	ADAPT2	LHSO	LHS
S_1	4e-01	1.55e-04	3.48e-04	7.53e-05	3.04e-02
S_2	3e-01	4.34e-04	1.58e-04	5.19e-04	7.35e-04
S_3	5e-02	3.42e-04	8.09e-05	7.24e-04	2.33e-02
S_4	3e-01	4.75e-04	9.04e-04	4.31e-04	2.47e-02
S_5	4e-07	1.31e+01	1.07e+01	2.37e+01	9.25e+02
S_6	2e-02	1.08e-03	4.54e-04	5.21e-04	3.81e-02
S_{12}	6e-03	1.30e-02	7.92e-03	7.70e-03	8.99e-02
S_{14}	5e-03	5.30e-03	1.81e-03	5.11e-03	2.74e-01
S_{15}	8e-06	9.34e+02	9.34e+02	9.06e+02	9.21e+02
S_{24}	3e-03	1.26e-03	7.24e-03	4.21e-03	7.10e-01
S_{45}	1e-05	9.93e-02	8.55e-02	1.14e-01	1.05e+01

We can conclude that all stochastic approaches under consideration give reliable relative errors for sufficiently large number of samples. The adaptive MC algorithm outperforms LHS in the case of small number of samples, higher dimensions and small by value sensitivity indices. The algorithms using generalized LHSO simulate the behaviour of Adaptive Monte Carlo algorithm, and even for higher dimensions sometimes gives better results than the to adaptive approach by at least one order. The most efficient in terms of computational complexity is the LHSO algorithm, followed very closely by the LHS algorithm. The algorithm LHS is characterized with unreliable relative errors for small in value sensitivity measures. MC approach based on optimized Latin hypercube sampling produce significantly better results for 6-dimensional and 12-dimensional integrals in comparison with the standard latin hypercube sampling LHS.

V. CONCLUSION

A comprehensive experimental study of Monte Carlo algorithm based on optimized latin hypercube sampling and adaptive approach for multidimensional numerical integration has been done. Such comparison has been made for the first time and this motivates the present study. The algorithms have been successfully applied to compute global Sobol sensitivity measures corresponding to the six chemical reactions rates and four different groups of pollutants. The numerical tests show that the optimized stochastic approach gives comparable results to the adaptive approach and better than the standard latin hypercube sampling and this method is very efficient for multidimensional numerical integration.

REFERENCES

[1] G. Dimitriu: Global Sensitivity Analysis for a Chronic Myelogenous Leukemia Model: Proc. 9th International Conference NMA'2018, Borovets, Bulgaria, August 20-24, 2018, LNCS 11189, Springer, Jan 2019. DOI: 10.1007/978-3-030-10692-8_42

[2] I.T. Dimov, R. Georgieva. Monte Carlo Method for Numerical Integration based on Sobol' Sequences. Numerical Methods and Applications (I. Dimov, S. Dimova, N. Kolkovska - Eds.), LNCS 6046, Springer, 2011, 50–59, https://doi.org/10.1007/978-3-642-18466-6_5.

[3] I. T. Dimov, R. Georgieva. Multidimensional Sensitivity Analysis of Large-scale Mathematical Models. O.P. Iliev et al. (eds.), Numerical Solution of Partial Differential Equations: Theory, Algorithms, and Their Applications, Springer Proceedings in Mathematics & Statistics 45, Springer Science+Business Media, New York, 2013, 137–156, https://doi.org/10.1007/978-1-4614-7172-1_8.

[4] Dimov I., Karaivanova A., Error analysis of an adaptive Monte Carlo method for numerical integration, *Mathematics and Computers in simulation* 47 pp. 201–213, 1998, https://doi.org/10.1016/S0378-4754(98)00103-7.

[5] Dimov I., Karaivanova A., Georgieva R., Ivanovska S. (2003) Parallel Importance Separation and Adaptive Monte Carlo Algorithms for Multiple Integrals, Springer Lecture Notes in Computer Science, 2542, Springer-Verlag, Berlin, Heidelberg, New York: 99–107, https://doi.org/10.1007/3-540-36487-0_10.

[6] H. Hamdad, Ch. Pézerat, B. Gauvreau, Ch. Locqueteau, Y. Denoual, Sensitivity analysis and propagation of uncertainty for the simulation of vehicle pass-by noise, Applied Acoustics Vol. 149, Elsevier, pp. 85-98 (June 2019). DOI: 10.1016/j.apacoust.2019.01.026

[7] T. Homma, A. Saltelli, Importance Measures in Global Sensitivity Analysis of Nonlinear Models, Reliability Engineering and System Safety 52, 1996, 1–17, https://doi.org/10.1016/0951-8320(96)00002-6.

[8] Jarosz, W.: Efficient Monte Carlo Methods for Light Transport in Scattering Media, PhD dissertation, UCSD, (2008).

[9] 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), 239–45 (1979), https://www.jstor.org/stable/1268522?seq=1.

[10] Minasny B., McBratney B.: A conditioned Latin hypercube method for sampling in the presence of ancillary information, Journal Computers and Geosciences archive, Volume 32 Issue 9, November, 2006, Pages 1378-1388, https://doi.org/10.1016/j.cageo.2005.12.009.

[11] S. Poryazov, E. Saranova, and I. Ganchev. Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems. Autonomous Control for a Reliable Internet of Services. Springer, Cham, 2018, 151-181, https://doi.org/10.1007/978-3-319-90415-3_7.

[12] Sobol I.M., Tarantola S., Gatelli D., Kucherenko S., Mauntz W., Estimating the approximation error when fixing unessential factors in global sensitivity analysis, Reliability Engineering & System Safety, 2007, 92, 957–960, https://doi.org/10.1016/j.ress.2006.07.001.

[13] Zheleva, I., Georgiev, I., Filipova, M., & Menseidov, D. (2017, October). Mathematical modeling of the heat transfer during pyrolysis process used for end-of-life tires treatment. In AIP Conference Proceedings (Vol. 1895, No. 1, p. 030008). AIP Publishing LLC, https://doi.org/10.1063/1.5007367.

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

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