Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 25–28
DOI: 10.15439/2020F108
ISSN 2300-5963 ACSIS, Vol. 23

Optimized Quasi-Monte Carlo methods based on low discrepancy sequences for sensitivity analysis in air pollution modelling

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*—An optimization version of the van der Corput sequence has been used in our sensitivity studies of the model output results for some air pollutants with respect to the emission levels and some chemical reactions rates. Sensitivity analysis of model outputs to variation or natural uncertainties of model inputs is very significant for improving the reliability of these models. Clearly, the progress in the area of air pollution modeling, is closely connected with the progress in reliable algorithms for multidimensional integration.**

I. Introducton

THE main goal of the present work is to develop and investigate efficient stochastic algorithms for multiple numerical integration providing sensitivity analysis (SA) that means evaluating Sobol' sensitivity indices (SIs) [1], [3], [5]. While the classical deterministic grid methods are efficient for low dimensional integrands [2], they become computationally intensive and even impracticable for high dimensions s because the number of required integrand evaluations grows exponentially. In contrast, the convergence rate of the plain **M**onte **C**arlo (MC) integration methods does not depend on the number of dimensions s. That is why the Monte Carlo method is a power tool in sensitivity analysis of large-scale systems [11].

Venelin Todorov is supported by the Bulgarian National Science Fund under Young Scientists Project KP-06-M32/2 - 17.12.2019 "Advanced Stochastic and Deterministic Approaches for Large-Scale Problems of Computational Mathematics" 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. The work is also supported by the Bulgarian National Science Fund under Project DN 12/5-2017 "Efficient Stochastic Methods and Algorithms for Large-Scale Problems".

The Unified Danish Eulerian Model (UNI-DEM)[12], [13] 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. 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. The calculations are done in a large spatial domain, which covers completely the European region and the Mediterranean, for certain time period (meteorological data must be available for it) [2]. In this particular study we use them for two of the most dangerous pollutants: the ozone (O_3) and the ammonia (NH_3). Other accumulative functions related to sone specific applications, maximal values, etc. are also calculated, exported and used in various application areas (environmental protection, agriculture, health care, etc.).

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:

$$\frac{\partial c_s}{\partial t} = -\frac{\partial (uc_s)}{\partial x} - \frac{\partial (vc_s)}{\partial y} - \frac{\partial (wc_s)}{\partial z} + \quad (1)$$

$$+\frac{\partial}{\partial x}\left(K_x \frac{\partial c_s}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial c_s}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial c_s}{\partial z}\right) + \quad (2)$$

$$+E_s + Q_s(c_1, c_2, \ldots, c_q) - (k_{1s} + k_{2s})c_s, \; s = 1, \ldots, q \quad (3)$$

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) [13].

II. The Van der Corput sequence

Let $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(s)})$ for $i = 1,2,\ldots$. The star discrepancy is given by:

$$D_N^* = D_N^*(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \sup_{\Omega \subset E^s} \left| \frac{\#\{\mathbf{x}_n \in \Omega\}}{N} - V(\Omega) \right|, \qquad (4)$$

where $E^s = [0,1)^s$. Let $m = \ldots a_3(m), a_2(m), a_1(m)$ be the representation of m in base b. The radical inverse sequence is given by:

$$m = \sum_{i=0}^{\infty} a_{i+1}(m) b^i, \ \phi_b(m) = \sum_{i=0}^{\infty} a_{i+1}(m) b^{-(i+1)} \qquad (5)$$

and it is fulfilled that

$$D_N^* = O\left(\frac{logN}{N}\right) \qquad (6)$$

The **Van der Corput** sequence is obtained when $b = 2$ [9].

The van der Corput sequence is often used to generate a sequence of points which have a better covering property than pseudorandom points. The van der Corput sequence generates a sequence of points in [0,1] which never repeats. The elements of the van der Corput sequence are strictly less than 1. The van der Corput sequence writes an integer in a given base b, and then its digits are "reflected" about the decimal point. This maps the numbers from 1 to N into a set of numbers in [0,1], which are especially nicely distributed if N is one less than a power of the base. The generation is quite simple. Given an integer I, the expansion of I in base b is generated. Then, essentially, the result R is generated by writing a decimal point followed by the digits of the expansion of I, in reverse order. This decimal value is actually still in base b, so it must be properly interpreted to generate a usable value. In the numerical experiments we will compare the standard van der Corput sequence VDC2 with an optimization variant when we increase the base to $b = 2^4$ - we use the notation VDCO. Such comparison has been made for the first time for the particular model.

The lattice sequence that we are going to use in our experimental study, namely Fibonacci based lattice rule FIBO and polynomial lattice sequence are described in [6], [7], [10].

III. Sensitivity Studies with Respect to Emission Levels

In this section we will present the results for the sensitivity of UNI-DEM output (in particular, the ammonia mean monthly concentrations) with respect to the anthropogenic emissions input data variation are shown and discussed in this section. The anthropogenic emissions input consists of 4 different components

$\mathbf{E^A}$ – 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 second degree have been used as an approximation tool. The input data have been generated by the improved SA-DEM version, developed for the purpose of our sensitivity studies (see the previous section).

Results of the relative error estimation for the quantities f_0, the total variance $\mathbf{D}$, first-order (S_i) and total (S_i^{tot}) sensitivity indices are given in Tables I, II, III, respectively. f_0 is presented by a 4-dimensional integral, while the rest of the above quantities are presented by 8-dimensional integrals, following the ideas of *correlated sampling* technique to compute sensitivity measures in a reliable way [8]. The four different stochastic approaches used for numerical integration are presented in separate columns of the tables.

A study of the computational efficiency of the stochastic algorithms under consideration for evaluating sensitivity measures presented by multidimensional integrals (total variance) or a ratio of multidimensional integrals (Sobol global sensitivity indices) have been made. The results show that the computational efficiency of the algorithms depends on integrand dimension and magnitude of estimated quantity. The order of relative error is different for different quantities of interest (see column *Reference value*) for the same sample size.

When $n = 2^{16}$ it can be seen that the modified lattice rule LATM produce better results than the lattice rule based on Fibonacci generalized numbers FIBO - see Table III. The optimized van der Corput sequence produce worse results than both of the lattice sequences FIBO and LATM. It can be seen that the van der Corput sequence with base $b = 2$ gives worse results by at least 1-2 orders to the optimized van der Corput sequence with base $b = 2^4$ VDCO.

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%).

IV. Sensitivity Studies with Respect to Chemical Reactions Rates

In this section we will 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 ([12]), namely: # $1,3,7,22$ (time-dependent)

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

# of samples n	**VDC2** Relative error	**VDCO** Relative error	**FIBO** Relative error	**LATM** Relative error
2^{10}	3.72e-04	1.64e-05	2.09e-04	8.46e-04
2^{12}	4.39e-04	5.11e-05	4.32e-05	1.79e-04
2^{14}	2.15e-03	4.45e-05	2.25e-05	2.62e-06
2^{16}	3.11e-04	5.56e-06	8.70e-06	4.14e-07
2^{18}	7.66e-05	1.03e-06	1.79e-06	1.17e-06
2^{20}	1.48e-04	3.56e-07	4.21e-07	1.15e-06

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

# of samples n	**VDC2** Relative error	**VDCO** Relative error	**FIBO** Relative error	**LATM** Relative error
2^{10}	9.21e-02	1.78e-02	1.63e-01	1.54e-02
2^{12}	6.37e-02	4.11e-03	2.39e-02	3.67e-03
2^{14}	4.26e-02	1.34e-03	2.90e-03	1.49e-03
2^{16}	2.22e-03	3.19e-04	2.65e-04	1.61e-03
2^{18}	2.22e-03	1.13e-04	3.01e-04	1.48e-03
2^{20}	7.58e-03	5.76e-05	1.19e-04	1.46e-03

and # $27, 28$ (time independent). The simplified chemical equations of those reactions are:

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

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

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

Homma and Saltelli discuss in [4] which is the better estimation of

$$f_0^2 = \left(\int_{U^d} f(\mathbf{x}) \mathrm{d}\mathbf{x} \right)^2 \tag{7}$$

in the expression for total variance and Sobol global sensitivity measures. The first formula is

$$f_0^2 \approx \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,d}) \; f(\mathbf{x}'_{i,1}, \ldots, \mathbf{x}'_{i,d}) \tag{8}$$

and the second one is

$$f_0^2 \approx \left\{ \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,d}) \right\}^2 \tag{9}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two independent sample vectors. In case of estimating sensitivity indices of a fixed order, formula (8) is better (as recommended in [4]), here we use it too.

The relative error estimation for the quantities f_0, the total variance $\mathbf{D}$ and some sensitivity indices are given in Tables IV, V, VI respectively. The four different stochastic approaches used for numerical integration are presented in separate columns of the tables. The quantity f_0 is presented by 6-dimensional integral, while the rest are presented by 12-dimensional integrals.

When $n = 2^{16}$ it can be seen that the van der Corput sequences produce similar results to the modified lattice rule LATM and Fibonacci generalized numbers FIBO - see Table VI. When $n = 2^{16}$ the optimized van der Corput sequence VDCO is far better than LATM and FIBO for same of the sensitivity indices - see the value for S_{45} and S_{12} in Table VI. Overall LATM produce better results than the Fibonacci based lattice rule by at least one order - see Table VI. None of the 4 methods estimates S_{15} and S_5 reliably, which has extremely small reference values. This natural "size effect" does not destroy the accuracy of the corresponding total sensitivity indices (which are much larger, so the influence of S_5 and S_{15} is negligible).

From these tables we can see the optimized van der Corput sequence gives better results than both the lattice sequences and the original van der Corput sequence with increasing the dimensionality of the integral. The results obtained by the optimized van der Corput sequence are better for 12 dimensional integrals compared with the results for 8 dimensional integrals. To summarize, Tthe algorithm has been successfully applied to compute global Sobol sensitivity measures corresponding to the influence of several input parameters (six chemical reactions rates and four different groups of pollutants) on the concentrations of important air pollutants.

TABLE III
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS USING VARIOUS MONTE CARLO AND QUASI-MONTE CARLO APPROACHES ($n \approx 65536$).

Est. qnt.	Ref. val.	**VDC2**	**VDCO**	**FIBO**	**LATM**
S_1	9e-01	3.13e-02	4.56e-04	3.62e-04	7.27e-04
S_2	2e-04	1.28e+00	3.34e-02	1.74e-01	2.76e-02
S_3	1e-01	9.13e-02	2.22e-03	3.22e-03	4.24e-03
S_4	4e-05	8.30e-01	3.45e-02	4.87e-01	1.65e-02
S_1^{tot}	9e-01	7.54e-03	1.16e-04	4.61e-04	5.14e-04
S_2^{tot}	2e-04	4.69e+01	1.24e-01	3.45e-01	2.21e-01
S_3^{tot}	1e-01	4.14e-02	1.10e-03	1.96e-03	6.41e-03
S_4^{tot}	5e-05	5.54e+02	1.60e-01	5.06e-01	1.60e-01

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

# of samples n	**VDC2** Relative error	**VDCO** Relative error	**FIBO** Relative error	**LATM** Relative error
2^{10}	1.25e-02	8.13e-04	2.08e-03	7.12e-03
2^{12}	1.70e-03	2.12e-04	1.40e-04	1.80e-03
2^{14}	3.56e-03	6.15e-05	3.98e-04	4.04e-05
2^{16}	8.66e-04	7.09e-05	2.61e-04	9.91e-06
2^{18}	4.63e-04	2.33e-06	7.29e-06	7.24e-06
2^{20}	7.85e-05	3.51e-07	4.57e-07	7.04e-06

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

# of samples n	**VDC2** Relative error	**VDCO** Relative error	**FIBO** Relative error	**LATM** Relative error
2^{10}	2.23e-02	1.10e-02	6.73e+00	3.11e-02
2^{12}	2.04e-01	8.45e-03	5.27e-01	8.76e-02
2^{14}	3.94e-02	6.44e-04	1.02e-01	7.54e-04
2^{16}	1.05e-03	2.71e-04	1.97e-03	9.13e-04
2^{18}	1.27e-02	1.01e-04	4.53e-03	2.22e-03
2^{20}	1.53e-02	9.56e-05	9.33e-03	2.22e-03

TABLE VI
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS USING VARIOUS QMC APPROACHES ($n \approx 65536$).

Est. qnt.	Ref. val.	**VDC2**	**VDCO**	**FIBO**	**LATM**
S_1	4e-01	1.07e-01	1.92e-02	3.82e-02	1.50e-02
S_2	3e-01	5.08e-02	7.45e-03	1.03e-02	2.14e-02
S_3	5e-02	6.37e-04	5.22e-04	5.48e-01	8.28e-02
S_4	3e-01	7.39e-02	4.71e-03	1.07e-02	6.81e-03
S_5	4e-07	7.26e+02	9.00e+01	3.40e+03	2.07e+03
S_6	2e-02	4.10e-01	5.56e-03	1.32e+00	1.19e-02
S_1^{tot}	4e-01	8.89e-02	6.56e-02	7.92e-02	1.07e-02
S_2^{tot}	3e-01	1.06e-02	1.41e-02	3.06e-02	2.28e-02
S_3^{tot}	5e-02	1.25e-01	5.13e-02	1.31e+00	4.92e-02
S_4^{tot}	3e-01	1.51e-01	8.23e-03	3.84e-01	1.93e-02
S_5^{tot}	2e-04	3.45e+02	3.02e+00	8.85e+01	6.78e+00
S_6^{tot}	2e-02	1.63e+00	4.87e-02	2.15e+00	7.63e-02
S_{12}	6e-03	9.16e-01	1.34e-01	3.21e+00	2.21e-01
S_{14}	5e-03	1.91e-01	1.23e-01	8.64e+00	1.31e+00
S_{15}	8e-06	8.45e+02	1.72e+02	9.19e+02	9.62e+02
S_{24}	3e-03	3.25e-01	1.88e-02	1.37e+01	5.63e-01
S_{45}	1e-05	1.41e-01	4.23e-02	4.25e+01	3.87e+01

V. CONCLUSION

The computational efficiency (in terms of relative error and computational time) of several stochastic algorithms for multidimensional numerical integration has been studied to analyze the sensitivity of UNI-DEM model output to variation of input emissions of the anthropogenic pollutants and of rates of several chemical reactions. The algorithms have been successfully applied to compute global Sobol sensitivity measures corresponding to the influence of several input parameters on the concentrations of important air pollutants. The study has been done for the areas of several European cities with different geographical locations. The novelty of the proposed approaches is that the van der Corput sequence and its optimized version have been applied for the first time to sensitivity studies of the particular air pollution model. This is also the first time the Fibonacci based lattice rule and the modified lattice sequence based on transformation function are compared with the van der Corput sequence and its optimized version to the problem under consideration. The numerical tests show that the presented optimized stochastic approach is the most efficient for the multidimensional integrals under consideration and especially for computing small by value sensitivity indices.

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. Dimov, R. Georgieva, Monte Carlo algorithms for evaluating Sobol' sensitivity indices, Mathematics and Computers in Simulation, Volume 81, Issue 3, 2010, Pages 506-514, ISSN 0378-4754, https://doi.org/10.1016/j.matcom.2009.09.005.

[3] 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

[4] 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.

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

[6] Sloan I.H. and Kachoyan P.J., Lattice methods for multiple integration: Theory, error analysis and examples, SIAM J. Numer. Anal. 24, pp. 116–128, 1987, https://doi.org/10.1137/0724010.

[7] I.H. Sloan and S. Joe, *Lattice methods for multiple Integration*, (Oxford University Press 1994).

[8] 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.

[9] van der Corput, J.G. (1935), "Verteilungsfunktionen (Erste Mitteilung)" (PDF), Proceedings of the Koninklijke Akademie van Wetenschappen te Amsterdam (in German), 38: 813–821, Zbl 0012.34705

[10] Wang Y., Hickernell F.J. (2002) An Historical Overview of Lattice Point Sets. In: Fang KT., Niederreiter H., Hickernell F.J. (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-56046-0_10.

[11] 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.

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

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