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

Venelin Todorov  
Institute of Mathematics and Informatics  
ul. G. Bonchev 8, 1113 Sofia, Bulgaria  
Email: vtodorov@math.bas.bg, venelin@parallel.bas.bg

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

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

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

Abstract—In this work a systematic approach 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.

This work is supported by the Bulgarian National Science Fund under Project DN 12/5-2017. Venelin Todorov is supported by the Bulgarian National Science Fund under 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 D01-205/23.11.2018, financed by the Ministry of Education and Science in Bulgaria.

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 use 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 $x_{k,j}$, 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.
2) For each interval, choose one uniform random number \( r \in [0, 1] \). Then, let
\[
Prob_i = \frac{i - 1}{n} + r \cdot \frac{1}{n} \quad (1)
\]
3) Transform the probability into the sampled value \( x \) using the inverse of the cumulative distribution function:
\[
x = F^{-1}(Prob) \quad (2)
\]
4) The \( n \) values obtained for each variable \( X \) are paired randomly (e.g. by generating a random permutation of \( \{1, 2, \ldots, n\} \) 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 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 \( E = (E^A, E^N, E^S, E^C) \) as follows:
\[
\begin{align*}
E^A & \quad \text{ammonia (NH}_3) ; \\
E^N & \quad \text{nitrogen oxides (NO + NO}_2) ; \\
E^S & \quad \text{sulphur dioxide (SO}_2) ; \\
E^C & \quad \text{anthropogenic hydrocarbons.}
\end{align*}
\]

The domain is the 4-dimensional hypercube \([0.5, 1]^4\). The input data have been generated by the improved version SADEM 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.

![Comparison of random, stratified and latin hypercube samplings with 16 points (d = 2, M = 4).](image)

### Table I

<table>
<thead>
<tr>
<th># of samples</th>
<th>ADAPT1</th>
<th>ADAPT2</th>
<th>LHS0</th>
<th>LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
</tr>
<tr>
<td>2(^{16})</td>
<td>9.89e-05</td>
<td>4.05e-04</td>
<td>1.45e-05</td>
<td>7.74e-05</td>
</tr>
<tr>
<td>2(^{18})</td>
<td>3.95e-05</td>
<td>3.83e-06</td>
<td>1.98e-06</td>
<td>3.80e-06</td>
</tr>
<tr>
<td>2(^{20})</td>
<td>4.99e-05</td>
<td>2.93e-05</td>
<td>6.31e-06</td>
<td>7.16e-06</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th># of samples</th>
<th>ADAPT1</th>
<th>ADAPT2</th>
<th>LHS0</th>
<th>LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
</tr>
<tr>
<td>2(^{16})</td>
<td>1.83e-04</td>
<td>5.12e-04</td>
<td>2.17e-04</td>
<td>3.65e-04</td>
</tr>
<tr>
<td>2(^{18})</td>
<td>5.77e-05</td>
<td>4.56e-05</td>
<td>1.21e-05</td>
<td>1.21e-05</td>
</tr>
<tr>
<td>2(^{20})</td>
<td>3.42e-05</td>
<td>3.28e-05</td>
<td>1.87e-05</td>
<td>5.96e-05</td>
</tr>
</tbody>
</table>

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 \( S^{tot}_1 \) 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 \( S^{tot}_1 \)).

### 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:
\[
\begin{align*}
\text{[#1]} \quad & NO_2 + h\nu \rightarrow NO + O; \\
\text{[#3]} \quad & O_3 + NO \rightarrow NO_2; \\
\text{[#7]} \quad & NO_2 + O_3 \rightarrow NO_3;
\end{align*}
\]

### Table III

<table>
<thead>
<tr>
<th>EQ</th>
<th>Ref. value</th>
<th>ADAPT1</th>
<th>ADAPT2</th>
<th>LHS0</th>
<th>LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_1</td>
<td>9e-01</td>
<td>7.67e-04</td>
<td>1.22e-03</td>
<td>1.16e-04</td>
<td>9.79e-03</td>
</tr>
<tr>
<td>S_2</td>
<td>2e-04</td>
<td>1.47e-03</td>
<td>4.96e-02</td>
<td>7.10e-03</td>
<td>6.60e-01</td>
</tr>
<tr>
<td>S_3</td>
<td>1e-01</td>
<td>4.11e-03</td>
<td>1.59e-03</td>
<td>1.19e-03</td>
<td>8.65e-03</td>
</tr>
<tr>
<td>S_4</td>
<td>4e-05</td>
<td>1.04e-01</td>
<td>1.69e-01</td>
<td>6.10e-02</td>
<td>6.70e-01</td>
</tr>
<tr>
<td>S_{1\text{tot}}</td>
<td>9e-01</td>
<td>4.99e-05</td>
<td>5.36e-05</td>
<td>7.31e-05</td>
<td>4.31e-04</td>
</tr>
<tr>
<td>S_{2\text{tot}}</td>
<td>2e-04</td>
<td>5.23e-01</td>
<td>5.00e+00</td>
<td>1.07e-01</td>
<td>2.94e+01</td>
</tr>
<tr>
<td>S_{3\text{tot}}</td>
<td>1e-01</td>
<td>1.15e-02</td>
<td>1.28e-02</td>
<td>5.45e-03</td>
<td>1.10e-02</td>
</tr>
<tr>
<td>S_{4\text{tot}}</td>
<td>5e-05</td>
<td>1.88e+01</td>
<td>3.43e+01</td>
<td>7.54e-01</td>
<td>2.41e+02</td>
</tr>
</tbody>
</table>

The relative errors for evaluation of the total variance \( D \approx 0.0002 \).
The domain under consideration is the 6-dimensional hypercube $[0,1]^6$). Homma and Saltelli discuss in [7] which is the better estimation of $f_0^2 = \left( \int_{x_0} f(x) dx \right)$ in the expression for total variance and Sobol' global sensitivity measures. In case of estimating sensitivity indices of a fixed order, the formula $f_0^2 \approx \frac{1}{n} \sum_{i=1}^{n} f(x_i,1,\ldots,x_{i,d}) f(x_i',1,\ldots,x_{i,d})$, where $x$ and $x'$ are two independent sample vectors, is better (as recommended in [7], [12]).

### TABLE IV

<table>
<thead>
<tr>
<th># of samples</th>
<th>ADAPT1</th>
<th>ADAPT2</th>
<th>LHSO</th>
<th>LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
</tr>
<tr>
<td>$2^{16}$</td>
<td>3.49e-05</td>
<td>3.72e-05</td>
<td>1.12e-05</td>
<td>2.02e-04</td>
</tr>
<tr>
<td>$2^{18}$</td>
<td>5.90e-07</td>
<td>1.02e-06</td>
<td>6.14e-06</td>
<td>2.82e-05</td>
</tr>
<tr>
<td>$2^{20}$</td>
<td>1.36e-07</td>
<td>5.56e-07</td>
<td>8.71e-07</td>
<td>1.04e-05</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th># of samples</th>
<th>ADAPT1</th>
<th>ADAPT2</th>
<th>LHSO</th>
<th>LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
<td>Relative error</td>
</tr>
<tr>
<td>$2^{16}$</td>
<td>3.01e-05</td>
<td>4.86e-04</td>
<td>8.23e-05</td>
<td>3.56e-05</td>
</tr>
<tr>
<td>$2^{18}$</td>
<td>7.71e-06</td>
<td>3.79e-04</td>
<td>6.12e-05</td>
<td>7.78e-03</td>
</tr>
<tr>
<td>$2^{20}$</td>
<td>1.75e-06</td>
<td>3.34e-05</td>
<td>5.21e-05</td>
<td>2.78e-04</td>
</tr>
</tbody>
</table>

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 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 test 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


