WP5: Monte Carlo methods for sensitivity analysis of large mathematical models

1. Major activities and results

Task 5.1. Generation of mesh function for different climate scenarios (pollutants, chemical reactions) using two and three dimensional version of UNI-DEM (Unified Danish Eulerian Model).

By performing a series of test experiments we analyzed and improved the scalability of SA-DEM package for sensitivity analysis of the Danish Eulerian model (with three hierarchical levels of parallelism). It has been implemented and optimized for parallel execution on the IBM BlueGene/P supercomputer installed in Bulgaria. By using it a large number experiments have been done and a large amount of data to be used in sensitivity study of the Danish Eulerian model with respect to some chemical reactions rate coefficients, has been generated. It has been used for development of more efficient Monte Carlo methods for sensitivity analysis of the DEM. The results are presented in detail in [DGOZ_13c, OGZ_12, ODGMZ_13b].

A new set of numerous parallel experiments have been executed to obtain higher density mesh functions, which are further used for calculating the Sobol global sensitivity indices with respect to the chemical reactions rate coefficients. The results have been presented at the 9th International Conference on „Large-Scale Scientific Computations“ (http://parallel.bas.bg/Conferences/SciCom13/), June 3-7, 2013, Sozopol [6].

The improved SA-DEM version for sensitivity analysis of the Danish Eulerian model with three hierarchical levels of parallelism has been implemented and optimized for parallel execution on the IBM MareNostrum III cluster at Barcelona Supercomputing Centre (BSC) in Barcelona, Spain. On the first (highest) level parallelization is between different values of the perturbation coefficient vector. On the second level parallelization is based on decomposition of the computational domain. In both these levels distributed memory parallelization model is applied (by using the MPI library of standard communication routines). On the third (lowest) level the shared memory parallelization model is applied (by using OpenMP standard directives). Its application is limited up to the number of cores per node (16 in this case). IBM MareNostrum III is the most powerful parallel supercomputer in Spain, with 3028 16-core nodes and peak performance about 1 PetaFLOPS. This machine is superior with respect to the IBM BlueGene/P in terms of speed, cores per node and RAM per node. These advantages and especially the larger RAM per node (which turned out to be critical for execution of SA-DEM on the Blue Gene/P) allowed us to run the finest grid version of SA-DEM (480 x 480) in the experiments on MareNostrum III, which caused some problems on IBM BlueGene/P.
Results of numerical experiments on Spanish supercomputer show the parallel efficiency and scalability of the new implementation of the SA-DEM. There have been performed a number of experiments to obtain the necessary data for a sensitivity analysis of the Danish Eulerian model for the six chemical reactions rate coefficients in the chemical submodel. Some of the results are presented in [ODGMZ_13a].

These new results have been presented at the 9th IMACS seminar on Monte Carlo methods, held in Annecy, France in July 15-19, 2013 [8].

<u>Task 5.2.</u> Techniques for approximation of mesh function and error estimation.

A study of various sensitivity analysis approximation tools for presenting the data model UNI-DEM are performed for:

- ammonia, ozone , nitrogen sulphate and nitrate nitrogen to fluctuations in the levels of four major groups of anthropogenic emissions by origin, and
- ozone to variations in rate constants of six chemical reactions.

For the study polynomials of different degrees as individual variables, and the total degree of the polynomial are taken. Numerical analysis of the reliability of the approximations is evaluated by mean square root error. Initially, following only this criterion, all considered polynomials give satisfactory accuracy. Some additional tests made it possible to refine and clarify the reliable apparatus, namely - the comparison with the minimum and maximum values of the data model in the domains of definition.

The results are reported to the Fifth Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences (http://2013.eac4amitans.eu/index.html), Albena, Bulgaria, June 24-29, 2013 [7].

<u>Task 5.3.</u> Sensitivity analysis of the UNI-DEM output parameters according to the emission levels and supporting the decision makers in developing ecological directives.

Main results from the current study were presented in an invited talk at Ninth IMACS Seminar on Monte Carlo Methods, Annecy-le-Vieux, France, July 15–19, 2013 [2].

Sensitivity analysis for ozone concentrations, simulated by UNI-DEM, according to variations of rates of six chemical reactions rate coefficients has been provided. The sensitivity study has been done for the areas of four European cities (Genova, Milan, Manchester, and Edinburgh). It has been carried out by software packages for scientific computing (Matlab and Mathematica). The *correlated sampling approach* proposed by Saltelli and Kucherenko has been applied to compute Sobol sensitivity measures. In comparison with the standard Sobol approach this method provides reliable values for the indices and overcomes the loss of accuracy in the case of small sensitivity indices. The results obtained for the sensitivity indices show: a) the mathematical model UNI-DEM under consideration is additive according to the chosen input parameters – rates of chemical reactions, b) the results obtained during current study are fully consistent with the conclusions about importance of model inputs obtained in the case of smaller number of chemical reactions, c) a new important input parameter (the rate of the time-dependent chemical reaction # 1) is identified, d) insignificant influence of geographical location on the values of the evaluated variables. The results are described in [DGOZ_13a].

An experimental study of a Monte Carlo algorithm based on quasirandom Sobol sequences, a quasi-Monte Carlo algorithm and a quasi-Monte Carlo algorithm with scrambled Sobol sequences has been done. Each of these algorithms lead to reasonably efficient computation of unknown values - values of multidimensional integrals with smooth and non-smooth integrand, and also with the approximate function of the data generated by UNI-DEM (in particular by the integrated version of the model for sensitivity analysis, SA-DEM). The Monte Carlo algorithm based on modified Sobol sequences gives similar results for relative error rate to scrambled quasi-Monte Carlo algorithms. But there are also many cases when this algorithm has a substantial advantage, particularly for the small sensitivity indices. The results are described in [DGOZ_13b].

The influence of emission levels on the concentrations of four important air pollutants (ammonia, ozone, ammonium sulphate and ammonium nitrate) over three European cities (Milan, Manchester, and Edinburgh) with different geographical locations is considered. The emission levels have been simulated by UNI-DEM. A comprehensive study of efficiency of techniques to reduce the variance in

the calculation of small sensitivity indices has been done. Despite the fact that some of the first-order sensitivity indices are small, simplifications of the mathematical model should not be made on this stage since each group of air pollutants under consideration has important influence over some of the chemical species chosen. The results are described in [DGOZ_13c].

Approaches for sensitivity analysis in the case of dependent variables have been studied. An algorithm proposed by Kucherenko, Tarantola and Annoni for the calculation of first-order sensitivity indices and total sensitivity indices has been implemented. The algorithm is a generalization of the standard Sobol approach for independent inputs. It is necessary the distribution of variables to be known as a preliminary information (each input is a random variable). In case of normal distribution (that one can observe in the particular model) the application of the algorithm is directly. The results are presented in the Seventh International Conference on Sensitivity Analysis of Model Output SAMO 2013 (http://www.gdr-mascotnum.fr/mascot13.html/) in Nice, France [3].

Efficient Monte Carlo algorithms for multidimensional integrals have been applied to solve integral equations. The problem for calculation of a functional of the solution of an integral equation is transformed into approximate evaluation of finite number of integrals of increasing dimensionality. The applied algorithms have been developed and studied for multidimensional integrals in our previous studies, while the objective of the present research is the study of their properties for solving Fredholm integral equations of second kind. The algorithms use pseudorandom numbers that are shifted quasirandom Sobol sequences. In the first case, the new pseudorandom point is generated randomly on the sphere with a center the corresponding quasirandom point. In the second case, one more pseudorandom point is generated that is centrally symmetric to the first one. The results for relative error and computational time obtained by these algorithms are compared to results obtained by a known Sobol approach for integral equations. The main assumptions are that the task is solved with a preliminary given accuracy as well as the both kind of approximation errors are balanced (systematic and stochastic). The numerical experiments demonstrate reliability of the results obtained as well as an advantage of the studied algorithms in terms of computational complexity. The results are presented in the Ninth IMACS Seminar on Monte Carlo Methods IMACS MCM 2013 (http://www.lama.univ-savoie.fr/IMACS2013/) in Annecy-le-Vieux, France [4].

***Task 5.4. Computer modeling and studying the effects of the global climate changes and the variations in the amount of the harmful emissions on the pollution levels of the atmosphere in South eastern Europe, and in particular, on the territory of Bulgaria with taking into account the results* obtained in sensitivity analysis of UNI-DEM output results.**

New test experiments have been performed with the parallel code of the Danish Eulerian model for studying long range transport of air pollutants (UNI-DEM) on the supercomputer IBM Blue Gene / P in order to improve the computer model. Another goal of the study was the adjustment of parameters in the model and study of the impact of global climate change on the levels of pollution.

In [ZGD_13a] new results of computer experiments using real meteorological and emission data with the parallel version of UNI-DEM are presented. The aim of the work is to complete the study of the impact of global climate change on the levels of air pollution in the Balkan Peninsula. We have discussed and made conclusions regarding the following important for the modern society issues: (i) the impact of the emissions produced in European countries outside the Balkan region on the concentration levels of hazardous air pollutants in the Balkan countries; (ii) the impact of changes in levels of air pollution on yields in agriculture and the state of the forests; (iii) changes in the number of "bad days" (days on which ozone pollution exceeds the maximum safe threshold) due to climate changes; (iv) the impact of the increase in temperature alone or in combination with other factors over the levels of certain harmful to human health contaminants.

In [ZDFG_13] we summarize in the multidimensional aspect of some results obtained in connection with the modeling of processes of advection in the Danish Eulerian model for long-range transport of air pollutants. It is shown that a high degree of accuracy can be achieved in the multidimensional

case, when the Crank-Nicolson numerical scheme is combined with the Richardson extrapolation. Applications of the techniques developed go beyond the task of transfer of pollutants and have applications in solving problems with large dimensionality in other branches of science and engineering.

In [ZG_13] we present and discuss computer experiments performed to investigate the use of pre-conditioners built on the basis of the approximate LU factorization using all the advantages of the intrinsic properties of the coefficient matrix systems, including their scarcity. Certain important conclusions about efficiency of various computational algorithms and selection criteria for the end of the calculations.

The paper [ZGD_13b] presents the results of the research areas of absolute stability for solving ordinary differential equations by the methods of Runge - Kutta of order from one to four, including combination with the Richardson extrapolation. This leads to the possibility of larger time steps in the computer implementation of the algorithms and thus significantly reduce the time to implement the relevant computer codes .

In [ZGD_13c] results of computer experiments using real meteorological and emission data with the parallel version of UNI-DEM are presented. The goal of the article is to study the effects of systematic changes of pollution concentrations caused by human activities. This is done in a variety of computational areas that cover mainly the territory of Europe. It was found that while for the sulphate pollution concentrations there is a direct and almost linear relationship between their changes and the pollution produced by human activities, it is at most harmful contaminants that is not so . It is shown that although the modification of contaminants from human activities is the same for the whole computational domain variation in quantity of the same pollutants varies from one to another sub-domain of the model.

In [ZGD_13d] results produced by computer simulations to study the parallel performance of the Danish Eulerian model for the study of long-distance transfer of pollutants in the air (UNI-DEM) are presented. The influence of the uncertainty of input data (meteorological data , emission data and the rates of chemical reactions that are implemented in the model) on the model output is studied . In this work we examined the organization of parallel processing and scalability of the algorithms used.

2. Publications with acknowledgments to the project DCVP 02/1

(i) published

[DGOZ_12]

I. T. Dimov, R. Georgieva, Tz. Ostromsky, Z. Zlatev, *Variance-based Sensitivity Analysis of the Unified Danish Eulerian Model According to Variations of Chemical Rates*. - In: I. Dimov, I. Faragó, and L. Vulkov (Eds.) Proceedings of NAA 2012, LNCS 8236, Springer, 2013, 247 – 254. ISSN: 0302-9743. SJR (2012):0.517.

[DGOZ_13b].

I. T. Dimov, R. Georgieva, Tz. Ostromsky, Z. Zlatev, *Advanced Algorithms for Multidimensional Sensitivity Studies of Large-scale Air Pollution Models based on Sobol Sequences*. Special issue of Computers and Mathematics with Applications 65 (3), „Efficient Numerical Methods for Scientific Applications“. Elsevier, 2013, 338 - 351. ISSN: 0898–1221. IF (2012): 2.069. 5-year IF: 1.894.

[DGOZ_13c].

I. T. Dimov, R. Georgieva, Tz. Ostromsky, Z. Zlatev. *Sensitivity Studies of Pollutant Concentrations Calculated by UNI-DEM with Respect to the Input Emissions*. Central European Journal of Mathematics, "Numerical Methods for Large Scale Scientific Computing" 11 (8), 2013, 1531 – 1545. Doi: 10.2478/s11533-013-0256-2. IF (2012): 0.405.

[ODGMZ_13b]

Tz. Ostromsky, I. T. Dimov, R. Georgieva, P. Marinov, Z. Zlatev. High Performance Computing of Data for a New Sensitivity Analysis Algorithm, Applied in an Air Pollution Model. - In: I. Dimov, I. Faragó, and L. Vulkov (Eds.) Proceedings of NAA 2012, LNCS 8236, Springer, 2013, 428–436. ISSN: 0302-9743. SJR (2012): 0.517.

[OGZ_12]

Tz. Ostromsky, K. Georgiev, Z. Zlatev, An Efficient Highly Parallel Implementation of a Large Air Pollution Model on an IBM Blue Gene Supercomputer. In: AIP Conf. Proc. 1487, 2012, 135 - 142. ISSN: 0094-243X. ISBN: 978-0-7354-1099-2.

[ZGD_13a]

Z. Zlatev, K. Georgiev, I. Dimov, *Influence of Climatic Changes on Pollution Levels in the Balkan Peninsula*, Computers and Mathematics with Applications, **65**, 2013, 544–562, Elsevier, ISSN 0898–1221, IF (2012): 2.069. 5-year IF: 1.894.

(ii) accepted

[ODGMZ_13a]

Tz. Ostromsky, I. T. Dimov, R. Georgieva, P. Marinov, Z. Zlatev, *Sensitivity Study of a Large-Scale Air Pollution Model by Using High-Performance Computations and Monte Carlo Algorithms*. AIP Conference Proceedings, ISSN: 0094-243X.

[ZDFG_13]

Z. Zlatev, I. Dimov, I. Farago, K. Georgiev, A. Havasi, Tz. Ostromsky, *Application of Richardson Extrapolation with the Crank–Nicolson Scheme for Multi–dimensional Advection.*

[ZG_13]

Z. Zlatev, K. Georgiev, *Applying Approximate LU-factorizations as Preconditioners in Eight Iterative Methods for Solving Systems of Linear Algebraic Equations*, Central European Journal of Mathematics, Springer, IF: 0581.

[ZGD_13c]

Z. Zlatev, K. Georgiev, I. Dimov, *Sensitivity of European Pollution Levels to Changes of Human-Made Emissions*, глава от електронна книга, озаглавена „Advanced Numerical Methods for Complex Environmental Models: Needs and Availability".

[ZGD_13d]

Z. Zlatev, K. Georgiev, I. Dimov. *Parallel Computations in a Large-Scale Air Pollution Model*, глава от електронна книга, озаглавена „*Advanced Numerical Methods for Complex Environmental Models: Needs and Availability"*.

(iii) submitted

[ZGD_13b**]**

Z. Zlatev, K. Georgiev, I. Dimov, *Absolute Stability Properties of the Richardson Extrapolation Combined with Explicit Runge-Kutta Methods*.

3. Presentations and talks

[1] I. T. Dimov, Sensitivity analysis for large-scale problems, Workshop on Supercomputer Applications, 31 март – 2 април 2013 г., Трявна.

[2] I. T. Dimov, *Efficient Monte Carlo Algorithms with Applications to Sensitivity Analysis*, Ninth IMACS Seminar on Monte Carlo Methods, July 15 - July 19, 2013, Annecy-le-Vieux, France.

[3] I. T. Dimov, R. Georgieva, *Global Sensitivity Study of Compact Models in Nanodevices with Correlated Inputs*, 7th International Conference on Sensitivity Analysis of Model Output, Nice, France, July 1-4, 2013.

[4] I. T. Dimov, R. Georgieva, *Advanced Monte Carlo Algorithms for Solving Integral Equations*, Ninth IMACS Seminar on Monte Carlo Methods, July 15 - July 19, 2013, Annecy-le-Vieux, France.

[5] I. T. Dimov, R. Georgieva, P. Marinov, Tz. Ostromsky, Z. Zlatev, *Sensitivity Studies of a Large-Scale Air Pollution Model: Approximation Techniques and Monte Carlo Algorithms*, 9th International Conference on „Large-Scale Scientific Computations", June 3-7, 2013, Sozopol.

[6] Tz. Ostromsky, I. T. Dimov, R. Georgieva, P. Marinov, Z. Zlatev, *Sensitivity Study of a Large-Scale Air Pollution Model with Respect to Various Input Data Sets, Based on High-Performance Computations*, 9th International Conference on „Large-Scale Scientific Computations", June 3-7, 2013, Sozopol.

[7] Tz. Ostromsky, I. T. Dimov, R. Georgieva, P. Marinov, Z. Zlatev, *Sensitivity Study of a Large-Scale Air Pollution Model by Using High-Performance Computations and Monte Carlo Algorithms*, Fifth Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences, Albena, Bulgaria, June 24-29, 2013.

[8] Tz. Ostroms*ky, I. T. Dimov, R.* Georgieva, P. Marinov, Z. Zlatev, *A Monte Carlo Method for Sensitivity analysis of an Air Pollution Model - Implementation, Performance and Results*, July 15 - July 19, 2013, Annecy-le-Vieux, France.

4. Others

[1] Technical and organizational activities:

a) participation in organization of the workshop on „Supercomputer Applications“ in the framework of SuperCA++, 31.03 – 02.04.2013, Tryavna;

б) participation in organization of the first Europe – China HPC Conference on “Partnership for supercomputing applications”, April 08 – 10, 2013, Grand Hotel Sofia, Sofia Bulgaria.

[2] Organizational and financial activities:

a) participation in regular meetings of the operational group for organization of the work on WP1, WP4 and WP5 in IICT-BAS (S. Margenov, E. Atanassov, I. Dimov, K. Georgiev).

[3] Additional activities for dissemination and popularization of the obtained results in the frame of:

a) Conference on Large Scale Scientific Computing, Sozopol, June 3-7, 2013 (organized by IICT – BAS);
b) Ninth IMACS Seminar on Monte Carlo Methods, July 15 - July 19, 2013, Annecy-le-Vieux, France.