Communication Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 125–130
DOI: 10.15439/2024F2586
ISSN 2300-5963 ACSIS, Vol. 41

Sensitivity Analysis in Air Pollution Modeling Supported by High Performance Supercomputers

Tzvetan Ostromsky, Silvi-Maria Gurova
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences (IICT-BAS),
Acad. G. Bonchev Str., Block 25-A,
1113 Sofia, Bulgaria
Email: {ceco,smgurova}@parallel.bas.bg

Meglena Lazarova
Faculty of Applied Mathematics and Informatics, Technical University of Sofia, 8 "St. Kliment Ohridski" Blvd., 1000 Sofia, Bulgaria
Email: { meglena.laz@tu-sofia.bg }

***Abstract*—Environmental modeling (and air pollution modeling in particular) is one of the toughest problems of computational mathematics. All relevant physical, chemical, and photochemical processes in the atmosphere should be taken into account. These are mathematically represented by a complex system of partial differential equations (PDE). To simplify the original PDE system proper splitting procedure is applied. As a result, the initial system is replaced by several simpler submodels, connected with the main physical and chemical processes. Even in the case of a local study of the environment in a relatively small area, the model should be calculated in a large spatial domain, because the pollutants can be moved quickly over long distances, driven by the atmosphere dynamics, especially at high altitudes. One major source of difficulty is the high intensity of the atmospheric processes, which require a small time step to be used to get a stable numerical solution (at least in the chemistry submodel). All this makes the treatment of large-scale air pollution models a heavy computational task that requires efficient numerical algorithms. It has always been a serious challenge for the fastest and most powerful supercomputers of their time. Fortunately, Bulgaria is one of the leading countries in Eastern Europe concerning the supercomputer infrastructure development in recent years.**

***Index Terms*—air pollution modeling, system of PDE, sensitivity analysis, algorithm, supercomputer, scalability.**

I. Introduction

THE IMPORTANCE of environmental security has grown rapidly in the last years of the 21st century. This has become a topical issue in the decision-making process of environmental issues for developed countries around the world. To create reliable scenarios of global environmental changes and their possible consequences, it is necessary to carry out many comprehensive scientific studies on various aspects of the environment and climate. From a scientific point of view, this means carefully analyzing the most important physical and chemical processes during the transport of pollutants, including all possible transformations and interactions with other species [12], in the air over long distances. The realization of this study requires collaboration between scientific experts in the field of environmental modeling, and numerical analysis, as well as the conduct of numerous studies, based on which results are obtained that will be useful to society.

The present work aims to help develop a new reliable and computationally efficient mechanism to investigate the sensitivity of estimated concentration levels for the most important air pollutants, such as nitrogen dioxide (NO2) and especially ozone (O3), to the change of the rates of chemical reactions involved (directly or indirectly) in their formation. Both the actual data estimates and some scenarios with a certain reduction in emissions from land transport in Europe calculated by the Unified Danish Euler Model (UNI-DEM) are used in these sensitivity analysis studies [5, 6, 13].

Sensitivity analysis (SA) [10, 17, 19] is called the study of how uncertainty in the output of a model can be allocated to different sources of uncertainty in the input data, and several sensitivity analysis techniques stand out in the literature [6, 13, 16, 18]. Most existing SA methods rely heavily on assumptions about some special properties related to the behaviour of the model (such as linearity, monotonicity, and additively of the relationship between the input factor and the output of the model). Among the quantitative methods, the methods based on variances are most often used [17]. The main idea of these methods is to estimate how the variance of an input or group of inputs contributes to the variance of the model's output.

The computational tasks involved in processing large-scale air pollution models are enormous. For this reason, it is advisable to simplify the considered large-scale tasks as much as possible, but to strive to preserve the high level of reliability of the obtained results of the considered model. On the one hand, this can be obtained using sensitivity analysis. On the other hand, it is necessary to study the influence of variations of the initial conditions, boundary conditions, and/or chemical rates on the model results. This will assist in creating more correct model simplifications. This kind of analysis can provide the researcher with valuable information about reasonable simplifications and/or identify reliable parameters

and mechanisms to improve. This is because changes in parameters and mechanisms affect model outputs.

The rest of the article is structured as follows. The second section provides an overview of the Danish Eulerian Model, encompassing its high-performance parallel code UNI-DEM, and its specialized sensitivity analysis version, SA-DEM. The third section presents numerical outcomes from scalability experiments conducted with SA-DEM on two of Europe's largest supercomputers – IBM MareNostrum III in Barcelona, Spain, and the petascale supercomputer Discoverer in Sofia, Bulgaria, part of the EuroHPC high performance computing network. Finally, concluding remarks from the presented results and some plans for future work are given at the end of the paper.

II. The Danish Eulerian Model (DEM) and its parallel implementation

A. Historical review

The Danish Eulerian Model (DEM), boasting a robust development history spanning over 30 years [2, 3, 14, 15, 21, 22, 23], stands as a formidable tool for assessing large-scale air pollution. Throughout its evolution, it has proven its efficacy in diverse, long-term environmental studies across various domains, including but not limited to environmental protection, human health care, agricultural production, forestry, wildlife, cultural heritage preservation and others. All relevant physical and chemical processes in the atmosphere should be taken into account, which are mathematically represented by a complex PDE system. To simplify it a proper splitting procedure is applied. As a result, the initial system is replaced by several simpler systems (submodels), connected with the main physical and chemical processes. These systems should be calculated in a large spatial domain, as the pollutants migrate quickly over long distances, driven by the atmosphere dynamics, especially at high altitudes. Here they are exposed to temperature, light, and other conditions changes in an extremely wide range, and so does the speed of most chemical reactions. One of the major sources of difficulty is the dynamics of the atmospheric processes, which require a small time step to be used (at least, for the chemistry submodel) to get a stable numerical solution of the corresponding system. All this makes the treatment of large-scale air pollution models a tuff and heavy computational task. It has always been a serious challenge, even for the fastest and most powerful state-of-the-art supercomputers. [7, 23].

B. Description of the Danish Eulerian Model (DEM)

The Danish Eulerian Model (DEM) [3, 20, 21] is mathematically represented by the following system of partial differential equations:

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

in which the parameters' meaning is as follows:

q - number of equations (equal to the number of chemical species),

c_s - concentrations of the chemical species considered,

u, v, w - components of the wind along the coordinate axes,

K_x, K_y, K_z - diffusion coefficients,

E_s - emissions in the space domain,

k_{1s}, k_{2s} - coefficients of dry and wet deposition respectively (s = 1, …, q),

$Q_s(c_1, c_2, \ldots, c_q)$ - non-linear functions that describe the chemical reactions between the species.

C. Splitting into submodels

The above rather complex system is split into three subsystems (submodels), according to the major physical and chemical processes as well as the numerical methods applied in their solution. In particular, these are:

- (i) the horizontal advection and diffusion (2);
- (ii) chemistry, emissions, and deposition (3);
- (iii) vertical exchange (4).

The submodels are described rigorously by the next formulae:

$$(2)\quad \frac{\partial c_s^{[1]}}{\partial t} = -\frac{\partial(uc_s^{[1]})}{\partial x} - \frac{\partial(vc_s^{[1]})}{\partial y} + \frac{\partial}{\partial x}\left(K_x \frac{\partial c_s^{[1]}}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial c_s^{[1]}}{\partial y}\right),$$

$$(3)\quad \frac{\partial c_s^{[2]}}{\partial t} = E_s + Q_s\left(c_1^{[2]}, c_2^{[2]}, \ldots, c_q^{[2]}\right) - (k_{1s} + k_{2s})c_s^{[2]},$$

$$(4)\quad \frac{\partial c_s^{[3]}}{\partial t} = -\frac{\partial(wc_s^{[3]})}{\partial z} + \frac{\partial}{\partial z}\left(K_z \frac{\partial c_s^{[3]}}{\partial z}\right).$$

The discretization of the spatial derivatives in the right-hand sides of the sub-models results in forming three large systems of ordinary differential equations.

D. Numerical methods and domain decomposition techniques applied in the submodels

A selection of various numerical methods is used in the numerical solution of the submodels. High stability require-

ments apply to the selected methods, because of the stiffness of the equations to be solved.

In the *advection-diffusion* submodel Finite Elements method (FE) is applied, followed by predictor-corrector schemes with several different correctors.

In the submodel for *chemistry, emissions, and deposition*, an enhanced version of the Quasi Steady-State Approximation (QSSA) is employed [8, 9, 11]. Emphasizing the pivotal role of chemical reactions, the model introduces both nonlinearity and stiffness. This atmospheric chemistry model stands out for its meticulous and precise depiction of chemical processes. The chosen chemical scheme – the condensed CBM-IV (Carbon Bond Mechanism IV), was initially proposed in [8], with subsequent improvements, specifically addressing ammonia-ammonium transformations, documented in [21]. In its basic form (involving 35 chemical species), the model accounts for 116 chemical reactions, of which 69 are time-dependent and 47 are time-independent. This scheme proves to be well-suited and adequate for investigating scenarios involving elevated concentrations of various pollutants.

In the *vertical transport* submodel, the Finite Elements method is generally used, followed by θ-methods. It should be mentioned here that there is an option to switch off the vertical transport submodel, which results in a simplified 2D version of the model.

E. Parallel implementation

The primary tool for parallelization is the MPI standard library. Originally designed as a communication library for distributed memory computers, MPI (Message Passing Interface) has evolved into one of the most popular parallelization tools for application programming, known for its efficiency, portability, and user-friendliness. Its versatility extends to a wide range of parallel systems, including shared-memory computers and clustered systems (where each cluster node operates as a distinct shared-memory machine), ensuring a high level of code portability.

In the context of UNI-DEM, MPI parallelization relies on spatial domain partitioning (see [13, 14, 15] for more detail).

In SA-DEM code, built up on simultaneous execution of several UNI-DEM calls with perturbed internal parameters and/or input data, there is another (coarser) level of parallelism, based on these large potentially parallel tasks. The MPI library allows these to be organized as super-tasks, working on separate communicators.

For a more in-depth explanation of the main computational stages of the model, the numerical methods and techniques engaged in each stage, as well as the parallelization strategy used in the UNI-DEM and SA-DEM codes, one can refer to [1, 4, 14, 15, 21, 23].

III. NUMERICAL RESULTS

This section presents the outcomes of scalability experiments conducted on two of Europe's most powerful general purpose supercomputers, namely the IBM MareNostrum III at the Barcelona Supercomputing Centre (BSC) in UPC - Barcelona, Spain, and the Discoverer system at Sofia Tech Park in the capital of Bulgaria. Both systems are currently parts of the EuroHPC computational infrastructure, as MareNostrum III was upgraded and up-scaled to the more powerful MareNostrum V. Tables I and II below contain summarized results of these experiments. The primary user-defined parameters for the experiments with SA-DEM / UNI-DEM on both machines were configured as follows:

- Discretization grid size: 480 × 480 (two-dimensional fine-resolution grid with step 10 km in both directions);
- Number of chemical species (q): 35;
- Main time-discretization step: 90 seconds (applied to all stages, could be reduced on the chemistry stage, if necessary

TABLE I.
EXECUTION TIME T (IN SECONDS), SPEED-UP (SP), AND PARALLEL EFFICIENCY E (IN %) OF SA-DEM (FINEST GRID) ON THE SPANISH SUPERCOMPUTER IBM MARENOSTRUM III

Time **T** [seconds], speed-up ***(Sp),*** and parallel efficiency **E** [in %]
of SA-DEM (2D MPI version) on the IBM MareNostrum III supercomputer
(480×480) mesh, 35 chem.species, NSIZE = 32

PE / MPI tasks	nodes	STAGE						TOTAL		
		Advection-diffusion			Chemistry					
		T [s]	(*Sp*)	*E* [%]	T [s]	(*Sp*)	*E* [%]	T [s]	(*Sp*)	*E* [%]
10	1	82927	(*10*)	100 %	74835	(*10*)	100 %	168847	(*10*)	100 %
40	3	19578	(*42*)	108 %	16624	(*45*)	113 %	40282	(*42*)	105 %
80	5	10178	(*81*)	103 %	8953	(*84*)	104 %	22471	(*75*)	94 %
160	10	5188	(*160*)	98 %	4632	(*162*)	101 %	13273	(*127*)	80 %
320	20	2945	(*282*)	88 %	2352	(*318*)	99 %	8331	(*203*)	63 %
640	40	1537	(*540*)	84 %	1204	(*622*)	97 %	5387	(*313*)	49 %
960	60	1206	(*688*)	72 %	815	(*918*)	96 %	4118	(*410*)	43 %
1600	100	869	(*954*)	60 %	493	(*1517*)	95 %	3373	(*501*)	31 %

in order to ensure numerical stability, as well as for more accurate calculations of the extreme values of concentrations for some unstable species, involved in very quick chemical reactions);

- Chunk size parameter NSIZE (used for improving data locality of the parallel computations and better utilization of the fastest cache memory): 32;
- Period of simulations: one year.

A. Numerical results from scalability experiments with SA-DEM by using the IBM MareNostrum III supercomputer at BSC

In Table I results of running SA-DEM on the most powerful Spanish supercomputer – IBM MareNostrum III at BSC (located in the Technical University of Catalonia (UPC) - Barcelona) are presented. The main technical parameters of this system are described on its webpage [24]. The size of the test problem makes it impractical to use less than 10 parallel tasks in the simulations. That's why the table starts with experiment with 10 MPI tasks on one node (capable of executing up to 16 tasks in parallel on its 16 cores), assuming also a speed-up (Sp) of 10 in this case and measuring the efficiency E with respect to this initial experiment. We should mention also that not always the number of calculations determines the computing time, the amount of data transfers between different levels storage must also be taken into account. The limited size of cache (the fastest access storage) makes it more efficient in processing smaller tasks (their size is inversely proportional to their number), which explains the super-linear speed-up for the few initial rows of the table (the same applies, as well, to Table II).

B. Numerical results from scalability experiments with SA-DEM and UNI-DEM by using the largest Bulgarian supercomputer Discoverer

The outcomes from scalability experiments utilizing SA-DEM / UNI-DEM on Bulgaria's largest supercomputer (Discoverer), are reported in this section (Table 2 and Table 3). The machine was installed three years ago at Sofia Tech Park by Atos company. In November 2023, the new ranking of the top 500 supercomputers in the world was released, where Discover took place 166 [27]. This supercomputer is a crucial component of a novel network comprising eight new high-performance supercomputers across the European Union, established and supervised by the European High-Performance Computing Joint Undertaking (EuroHPC JU).

IV. CONCLUSIONS

Analyzing the results from the numerical experiments presented in the previous section the following insights were made:

- The parallel MPI implementation of SA-DEM demonstrates excellent balance, portability, and efficiency on some of leading Europe's supercomputers (including Bulgaria's most powerful supercomputer Discoverer), integrated into the EuroHPC computing infrastructure.
- The efficiency and speed-up are higher in the computationally-intensive stages. In particular, the chemistry-deposition stage (which does not need any communication

TABLE II.
EXECUTION TIME **T** (IN *SECONDS)*, SPEED-UP ***(SP),*** AND PARALLEL EFFICIENCY **E** (*IN* %) OF EXPERIMENTS WITH SA-DEM (2D FINE GRID VERSION) ON THE BULGARIAN PETASCALE SUPERCOMPUTER DISCOVERER (PART OF THE EUROHPC INFRASTRUCTURE)

Time T [seconds], speed-up (Sp), and parallel efficiency E [in %] of SA-DEM (2D MPI version) on DISCOVERER supercomputer (480×480) mesh, 35 chem.species, NSIZE = 32

PE / MPI tasks	nodes	STAGE						TOTAL		
		Advection-diffusion			Chemistry					
		T [s]	*(Sp)*	*E* [%]	T [s]	(*Sp*)	*E* [%]	T [s]	(*Sp*)	*E* [%]
10	1	73028	*(10.0)*	100 %	64817	*(10.0)*	100 %	148225	*(10)*	100%
20	2	36842	*(19.8)*	99 %	30296	*(21.4)*	107 %	72341	*(20.5)*	102%
40	3	18458	*(39.6)*	99 %	15465	*(41.9)*	105%	37949	*(39.1)*	98%
80	5	9634	*(75.8)*	95 %	7981	*(81.2)*	102 %	20636	*(71.8)*	90%
120	10	4807	*(152)*	95 %	3925	*(165)*	103 %	11795	*(126)*	79%
320	20	2525	*(289)*	90 %	2037	*(318)*	99 %	6861	*(216)*	68%
640	40	1290	*(566)*	88 %	1034	*(627)*	98 %	4810	*(308)*	48%
960	60	908	*(805)*	84 %	697	*(930)*	97 %	3363	*(441)*	46%
1600	100	764	*(956)*	60 %	445	*(1457)*	91 %	2781	*(533)*	33%

between the tasks) has almost linear overall speed-up, even super-linear concerning the number of MPI tasks (a cache-size effect due to decreasing size of processed data per task with increasing the number of MPI tasks).

• The advection-diffusion stage scales pretty well too, taking into account that there is some unavoidable computational overhead due to overlapping subdomain boundaries of the partitioning. In the chemically-intensive stages, the hybrid MPI-OpenMP code leverages a reduction in time, thanks to the involvement of multiple threads with the OpenMP lower level of parallelism activated at the core level within a node. However, this optimization has minimal impact on other stages.

• To maintain comparability, the size of chunks at the chemistry-deposition stage (NSIZE=32) has been consistent across all experiments. Although this might not be optimal in every scenario, particularly when the number of MPI tasks is high, using a lower value for NSIZE can potentially yield better results.

• As computations span a substantial number of nodes, communications between MPI processes on different nodes become costly, resulting in some speed-up degradation and reduced overall efficiency. To address this, efforts should focus on optimizing I/O processes to mitigate this negative impact, forming part of our future tasks.

ACKNOWLEDGMENT

This research of the authors was partially supported by the following projects:

The work of Tzvetan Ostromsky has been supported by the Bulgarian National Science Fund via the project "Efficient methods for modeling, optimization and decision making" (KP-06-N52/5) as well as by the Ministry of Education and Science (MES) of Bulgaria (Agreement № Д01-27/06.02.2024)**.**

The work of Silvi-Maria Gurova was carried out using the infrastructure purchased under the National Roadmap for RI, financially coordinated by the MES of the Republic of Bulgaria (Grant No. DO1-325/01.12.2023) and supported also by Grant No 30 from CAF America.

The work of Meglena Lazarova was supported by the Bulgarian National Science Fund under Project KP-06-M62/1 "Numerical deterministic, stochastic, machine and deep learning methods with applications in computational, quantitative, algorithmic finance, biomathematics, ecology and algebra" from 2022.

The European High-Performance Computing Joint Undertaking (Euro-HPC JU) is kindly acknowledged for granting us access and computer time on the DISCOVERER supercomputer.

REFERENCES

[1] Alexandrov V. and et al., Numerical integration of chemical ODE problems arising in air pollution models. Env. Modeling and Assessment **2,** 1997, 365–377.

[2] Christensen J., The Danish Eulerian Hemispheric Model. In: Gryning SE. Schiermeier F.A. (eds.) Air Pollution Modeling and Its Application XI NATO Challenges of Modern Society 21 Springer Boston MA, 1996.

[3] Christensen J., The Danish Eulerian hemispheric model — a three-dimensional air pollution model used for the arctic. Atmospheric Environment, **31(24),** 1997**,** 4169-4191.

[4] Dimov I., Monte Carlo Methods For Applied Scientists. World Scientific, 2007.

[5] Dimov I. et al., A Super-Convergent Stochastic Method Based on the Sobol Sequence for Multidimensional Sensitivity Analysis in Environmental Protection. *Axioms*, 2023, **12(2):**146.

[6] Dimov I. et al., Studying the sensitivity of pollutants' concentrations caused by variations of chemical rates. J. Comput. Appl. Math. **235,** 2010**,** 391–402.

[7] Dimov I. et al., Computational challenges in the numerical treatment of large air pollution models. Ecological Modelling **179**, 2004, 187–203.

[8] Gery M. et al., A photochemical kinetics mechanism for urban and regional scale computer modelling. J. Geophys Res. **94(D10),** 1989, 12925–12956.

[9] Hesstvedt E. et al., Quasi-steady-state approximations in air pollution modeling: comparison of two numerical schemes for oxidant prediction. Int. Journal of Chemical Kinetics **10**, 1978, 971–994.

[10] Homma T., Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering and System Safety **52**, 1996, 1–17.

[11] Hov Ø. et al., Comparison of numerical techniques for use in air pollution models with non-linear chemical reactions. Atmospheric Environment **23**, 1988, 967–983.

[12] Hvidtfeldt U. A. et al., Evaluation of the Danish AirGIS air pollution modeling system against measured concentrations of PM2.5, PM10, and black carbon. Environmental Epidemiology **2(2**), 2018, pe014.

[13] Ostromsky Tz. et al., Air pollution modeling, sensitivity analysis, and parallel implementation. Int. Journal of Environment and Pollution **46 (1-2),** 2011, 83–96.

[14] Ostromsky Tz. and Zlatev Z., Parallel Implementation of a Large-scale 3-D Air Pollution Model. Large Scale Scientific Computing (S. Margenov, J. Wasniewski, P. Yalamov, Eds.) Spinger **2179**, 2001, 309–316.

[15] Ostromsky Tz. and Zlatev Z., Flexible Two-level Parallel Implementations of a Large Air Pollution Model. Numerical Methods and Applications (I.Dimov, I.Lirkov, S. Margenov, Z. Zlatev eds.) LNCS **2542**, 2002, 545–554.

[16] Ostromsky Tz. et al., Advanced Sensitivity Analysis of the Danish Eulerian Model in Parallel and Grid Environment. AIP Conf. Proc**. 1404**, 2011, 225–232.

[17] Saltelli A. et al., Sensitivity Analysis. John Wiley & Sons publishers, Probability and Statistics series, 2000.

[18] Saltelli A. et al., Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Halsted Press New York, 2004.

[19] Sobol I., Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates Mathematics and Computers in Simulation, **55 (1-3**), 2001, 271–280.

[20] WEB-site of the Danish Eulerian Model: https://www2.dmu.dk/AtmosphericEnvironment/DEM/ (Last accessed December 2023).

[21] Zlatev Z., Computer treatment of large air pollution models Kluwer, 1995.

[22] Zahari Z., Comprehensive Air Pollution Studies with the Unified Danish Eulerian Model, 2003, 1125-1137.

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

[24] Technical documentation of MareNostrum III: https://www.bsc.es/marenostrum/marenostrum/mn3 (Last accessed April 2024)

[25] The official web page of the supercomputer Discoverer: https://sofiatech.bg/en/petascale-supercomputer/ (Last accessed April 2024)

[26] Discoverer HPC Doc: https://docs.discoverer.bg/index.html (Last accessed April 2024)

[27] Top500 list of supercomputers – November 2023: https://www.top500.org/system/179948/ (Last accessed April 2024)