Journal of Physics:
Conference Series

PAPER • **OPEN ACCESS**

Development, Computer Implementation, and Sensitivity Analysis of a Numerical Model for Environmental Simulations

To cite this article: Tzvetan Ostromsky *et al* 2025 *J. Phys.: Conf. Ser.* **3145** 012027

Development, Computer Implementation, and Sensitivity Analysis of a Numerical Model for Environmental Simulations

Tzvetan Ostromsky[1*], Silvi-Maria Gurova[1,4], Meglena Lazarova[2,3] and Venelin Todorov[1,4]

[1] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences (IICT-BAS), Acad. G. Bonchev str., bl. 25A, 1113 Sofia, Bulgaria

[2] Department of Mathematical Modelling and Numerical Methods, Faculty of Applied Mathematics and Informatics, Technical University of Sofia, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria

[3] Department of Mathematics, Informatics and Physics, Prof. Dr. Assen Zlatarov State University of Burgas, 1 Prof Yakimov Blvd., Burgas 8000, Bulgaria

[4] Centre of Excellence in Informatics and Information and Communication Technologies, Sofia, Bulgaria

[*]E-mail: **tz.ostromsky@gmail.com**
E-mail: smgurova@parallel.bas.bg , meglena.laz@tu-sofia.bg , venelin@parallel.bas.bg

Abstract. This paper presents the parallel implementation and numerical evaluation of the Danish Eulerian Air Pollution Model (UNI-DEM) and its variant SA-DEM, designed for sensitivity analysis of the model, as well as certain critical parts of it. The model solves a system of partial differential equations describing advection-diffusion transport, chemical reactions, emissions, and deposition using operator separation, robust time-stepping schemes, and domain decomposition for MPI-based execution. Scalability experiments were performed on two supercomputers - IBM MareNostrum III (BSC) and Discoverer (Sofia Tech Park) - using a unified test. (480 × 480 grid, 35 species, Δt = 90 s). The results show almost perfect scaling for the chemical stage and communication-limited scaling for the transport stage, with superlinear speedup observed at low core counts due to cache effects.

A global sensitivity analysis was conducted for 12 kinetic parameters using Sobol dispersion-based indices calculated by several MC/QMC algorithms, such as MSS-1, MSS-2, MSS-2-C, VDC2, FIBO, and OPTL. The results show that the combination of advanced atmospheric modelling, high-throughput computing, and quasi-Monte Carlo techniques provides both computational efficiency and reliable parameter ranking, supporting model downscaling and scenario-based air quality assessment.

1. Introduction and motivation for supercomputer modelling of atmospheric pollution

The first decades of the 21st century have been marked by rapid urbanization, increasing energy consumption, and intensification of transport, leading to an increase in air pollution. The main pollutants are fine particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen oxides (NO_x), sulphur dioxide

(SO_2), ground-level ozone (O_3), volatile organic compounds (VOC), and ammonia (NH_3), which in the atmosphere enter into complex photochemical and heterogeneous reactions and create secondary aerosols and ground-level ozone. The long-range transboundary transport of pollution and the circulation of air masses turn a regional problem into a global one. This has a direct impact on the health of the population, ecosystems, agriculture, and the climate (which in turn has an inverse effect on the spread of pollution). WHO health assessments show that it is chronic inhalation of fine particles and ozone that bears the main burden of premature mortality in Europe and the world [1]. In this context, societies want not only to monitor the concentration of pollutants, but also to predict their value depending on the specific weather conditions and emissions in a given area and possible exceedances.

1.1. Mathematical modelling of transboundary atmospheric pollution

Observations and measurements of pollution by stationary and mobile stations, satellites, etc., provide valuable data, but have certain limitations: spatial and temporal incompleteness, inability to reconstruct the three-dimensional field of chemical processes at each point, as well as inability to predict effects under different scenarios (e.g., changes in emissions, large fire events, sandstorms). Measurements are important, but predictions and scientific analyses require mathematical models that unify the physical and chemical processes in the atmosphere. Simple empirical or statistical models often do not capture the nonlinearity of photochemistry, aerosol fractionation, sedimentation, and secondary aerosol production. This makes the need for global models of atmospheric pollution with high spatial and temporal resolution and a realistic physicochemical scheme critical. These three-dimensional mathematical models, describing air quality in detail, solve a complex system of partial differential equations. It describes the dynamics of pollutant transport (advection and diffusion), chemical reactions, emissions, and dry/wet deposition. Its solution provides an estimate of the concentrations in each cell of the spatial grid covering the studied region. Such a model is the Danish Eulerian Model (DEM)/UNI-DEM [2,3,4,10]. DEM has over 30 years of history and has been used to assess air pollution, health risks, impacts on forestry and agriculture, etc. In its modern parallel implementation, the UNI DEM equations are divided into three submodels: horizontal advection diffusion, chemistry with emissions and deposition, and vertical exchange. Its basic chemical module uses the (condensed) CBM IV mechanism with 35 chemical species and 116 reactions.

The well-known meteorological models MSC-W, CMAQ, and CAMx [5,6,7] have a similar structure and follow a similar modular architecture, but they all face the same difficulty: the need for huge input data (meteorology, emission inventories, boundary conditions) and a small time step, which makes them computationally heavy. For society, these models have great benefits – they allow for "what if" analyses (for example, if we reduce NO_x by 30%, how does ground-level ozone change) and support the development of clean air policies and plans. The disadvantage is that maintaining high spatial resolution and complex chemistry leads to high memory and time costs and requires the use of modern algorithms for area division, load balancing, and optimal time integration schemes.

1.2. Role and advantages of supercomputing

In order for DEM and similar models to be executed in realistic conditions and to perform dozens of iterations for sensitivity analysis or ensemble predictions, it is necessary to use high-performance computing. For a one-year simulation of SA-DEM on a 480×480 grid with 35

chemical components and a step of 90 seconds, it is necessary to run it not once, but dozens and hundreds of times (for example, for sensitivity analysis or ensemble predictions). Parallel versions of the model show that the domain decomposition and MPI-communications allow efficient execution on petascale systems such as Discoverer (EuroHPC, Sofia Tech Park, Bulgaria,[8]) and MareNostrum III (BSC, Barcelona, Spain, [9]), while maintaining the balance between communication and computation. There are several benefits of using parallel version of the model on supercomputers: (i) real-time annual scenarios for hours instead of weeks; (ii) the ability to store large 3D fields and chemical tables in memory; (iii) the ability to run multiple simulated scenarios in parallel, which is key in making management decisions.

In the atmospheric models, uncertainty exists not only in the input data, but also in the rates of chemical reactions. Monte Carlo (MC) and Quasi-Monte Carlo (QMC) methods are used to calculate them. The Sobol series and its modifications, proposed by I.M. Sobol, are quasi-random series with small discrepancy. They fill the integral in multidimensional space more evenly and usually provide faster convergence than classical pseudo-random points, although this effect weakens in high dimensions. MC/QMC allow us to "drill" this parameter space and estimate the Sobol-type sensitivity indices that influence the extent to which the Danish-Eulerian model under consideration [11, 12]. Therefore, several modified MC/QMC algorithms were tested for that purpose - Sobol, MSS 1, MSS 2, MSS 2 S, OPTL, VDC2, and FIBO.

The paper is divided into the following parts: Section 2 presents the mathematical formulation of the Danish Euler model, its decomposition into submodels, discretization and parallel implementation for nowadays state-of-the-art supercomputers. Section 3 presents the numerical results from experiments conducted on two such supercomputers, part of the EuroHPC infrastructure. Concluding remarks and future work plans are presented in Section 4.

2. The Danish Eulerian Model (DEM): formulation, numerical implementation, and sensitivity

The Danish Eulerian Model (DEM) is a large-scale atmospheric chemical transport model for estimating the dispersion, transformation, and deposition of pollutants over Europe. The DEM is modular, integrates detailed meteorology and chemical mechanisms, and is used for air quality assessment, transboundary pollution, and long-term forecasts. In its modern unified parallel implementation (UNI-DEM), the processes are solved on a fixed 2D/3D grid by a system of DMEs including advection, diffusion, chemical and photochemical reactions, emissions, and deposition. The chemical module uses the condensed mechanism CBM-IV (≈ 35 species; 116 reactions), with some stiffness due to the nonlinearity of most chemical reactions. [3, 10].

2.1 Mathematical formulation of DEM

The DEM (Danish Eulerian Model) is described by the system (1) of partial differential equations (PDE) for the pollutant's concentrations c_s , $s=1, \ldots, q$ where q is the number of chemical species (equal to the number of equations), and the other notation is as follows:

- 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 ($s = 1,\ldots, q$),
- $Q_s(c_1, c_2,\ldots, c_q)$ - functions describing the chemical reactions between the species (in general - non-linear).

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

$$+ E_s + Q_s\left(c_{1,\ \ldots,\ c_q}\right) - (k_{1s} + k_{2s})c_s \ ; \quad s=1,\ 2,\ \ldots,\ q.$$

2.2. Operator division of DEM

Due to the complexity of system (1), it is transformed by operator-splitting into the three submodels (2) – (4), listed below. The spatial derivatives are discretized (forming large systems of PDEs), and time is integrated with robust computational schemes tailored to each submodel.

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

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

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

The submodels represented in the equations (2) - (4) refer to the following chemical and physical processes: (2) horizontal advection + horizontal diffusion; (3) chemistry + emissions + deposition; and (4) vertical transport exchange (convection + diffusion).

2.3. Discretization of the spatial domain and basic numerical methods applied in the submodels

For the numerical discretization of the spatial domain DEM uses the EMEP grid [13, 14]. Refinements of that grid (up to 480 x 480) are also used in the later versions (UNI-DEM, SA-DEM). The typical computational approaches here are based on finite-volumes or finite-differences with controlled flows. Finite-volumes schemes are preferred because they satisfy the mass conservation law.

The computational approaches applied to the three submodels (2) – (4) are based respectively on:

- Advection–diffusion (2): finite element (FE) + predictor-corrector with selection of correctors according to stability/smoothness.
- Chemistry + deposition (3): improved QSSA (*Quasi-Steady-State Approximation*); the used CBM-IV mechanism (35 species; 116 reactions: 69 time-dependent, 47 independent) describes the transformations in detail and is adequate for high concentrations.
- Vertical exchange (4): again FE in space and θ -methods in time. This module can be switched off (when running 2D scenarios).

2.4. High performance parallelization

For efficiency in the parallel implementation of the model (UNI-DEM), domain decomposition is applied, which naturally supports distributed memory and MPI parallelization. Communication overhead is critical in advection-diffusion due to boundary value exchange; non-blocking MPI communications, careful load balancing, and cache-optimizing settings (e.g.,

CHUNKSIZE/NSIZE) are used to mitigate this. Portability is achieved via MPI library of standard communication tools [15] (for distributed memory parallelization model) and OpenMP-standard directives [16] (for shared memory parallelization model) without the need of major changes between architectures.

2.5. Global Sensitivity Analysis and Evaluation of Sobol' Sensitivty Indices

The goal of global sensitivity analysis is to decompose the variance of the output $u = f(x)$ (concentrations, integrals, etc.) into the shares of the independent inputs x_i (e.g., chemical reaction rates) and their interactions [11].

The ratio $\frac{\mathrm{D}[\mathbf{E}(u \mid x_i)]}{\mathrm{D}_u} \equiv S_i$ is called the first-order sensitivity index (SI) [2], $u = f(\mathbf{x})$ is a model function with independent variables $\mathbf{x} = (x_1, x_2, ..., x_d) \in \mathrm{U}^d \equiv [0,1]^d$, $\mathrm{D}[\mathbf{E}(u \mid x_i)]$ is the variance of u and D_u - the total variance. Total sensitivity index (TSI) is defined as the following sum [12]:

$$S_{x_i}^{tot} = S_i + \sum_{l_1 \neq i} S_{i,l_1} + \sum_{l_1, l_2 \neq i, l_1 < l_2} S_{i,l_1,l_2} + ... + S_{i,l_1,l_2,...l_{d-1}},$$

where $S_{i,l_1,...l_{j-1}}$ denotes the j-th order mutual SI for x_i with respect to

$$x_{l_1}, \ldots, x_{l_j} (1 \leq j \leq d).$$

It is well known [11, 12] that the total and partial variances can be calculated as multidimensional integrals. Thus, the above sensitivity indices can be efficiently evaluated by using MC or QMC integration algorithms (which are, to a great extent, massively parallel and, therefore, computationally efficient). To obtain the necessary data for estimation of Sobol's indices sensitivity analysis, however, multiple DEM runs with different parameters are needed. A special sensitivity analysis version , called SA-DEM, has been developed for that purpose. SA-DEM implements larger parallelism (several UNI-DEM runs in parallel). This allows the distribution of samples across nodes and significantly reduces the "time-to-solution" while maintaining accuracy. At the same time, the "internal" parallelism (domain decomposition) ensures the scalability of a single execution. Thus, the combined scheme takes advantage of petascale systems such as the Spanish IBM MareNostrum III (BSC) and the Bulgaria's largest and most powerful supercomputer – Discoverer. In the next section some results of such experiments are presented and discussed.

3. Numerical Experiments and Discussion

The numerical experiments presented in this section were performed to evaluate the scalability and computational efficiency of the parallel version of the SA-DEM model when running on two high-performance petascale computing infrastructures: IBM Mare Nostrum III (owned by Barcelona Supercomputing Center, Spain) and Discoverer (owned by Petascale Bulgaria consortium and hosted at Sofia Tech Park). Both supercomputers are integral components of the EuroHPC Joint Undertaking, which aims to strengthen Europe's

supercomputing capabilities. Notably, MareNostrum III has since been upgraded to the more powerful MareNostrum V, a next-generation machine with significantly enhanced processing power and memory capabilities.

MareNostrum III was ranked among the top 50 supercomputers in the world during its peak. It featured over 3,000 compute nodes, each equipped with dual Intel Sandy Bridge processors, totaling more than 48,000 cores. It offered 1.1 petaflops of peak performance, supported by a high-speed interconnect and a large shared file system, making it well-suited for large-scale scientific and environmental simulations like those performed in this study.

Discoverer, inaugurated in 2021 and hosted at Sofia Tech Park, is the most powerful supercomputer in Southeast Europe and one of the key EuroHPC systems. Built by Atos using its BullSequana XH2000 platform, Discoverer delivers over 4.5 petaflops of peak performance and includes more than 250,000 cores. It supports a wide range of scientific, industrial, and public-sector applications, from weather forecasting to artificial intelligence, and provides access to high performance computing resources for researchers across Europe.

Both platforms allow running the model in parallel by using its 2-level MPI parallelization structure, based on spatial domain decomposition. Since horizontal advection-diffusion is the most communication-intensive module, and chemistry is the most computationally intensive, the selected test cases allow observing both the effect of network latency/throughput and the influence of CPU performance when increasing the number of processes, with the results presented in terms of speedup S(p) and parallel efficiency $E(p)$.

The results of the experiments on the two computing systems are presented in Figure 1 and Figure 2 respectively. In order to obtain comparable results, all the experiments were carried out with identical model configuration set: grid size 480×480 (refined EMEP grid with a step 10 km), 35 chemical species, step Δt=90sec, NSIZE = 32. The obtained plots demonstrate the load balance and efficiency of the SA-DEM parallelization, shows the limits of SA-DEM efficiency and allows us to identify the dominant sources of delay.

3.1. Scalability Results on the IBM MareNostrum III Supercomputer

This subsection examines the scalability of the parallel implementation of SA-DEM on the IBM MareNostrum III. The initial number of parallel tasks is chosen to guarantee a sufficiently large local subproblem (a set of 10 MPI tasks is the starting point), trying to avoid overflow in the distributed memory.

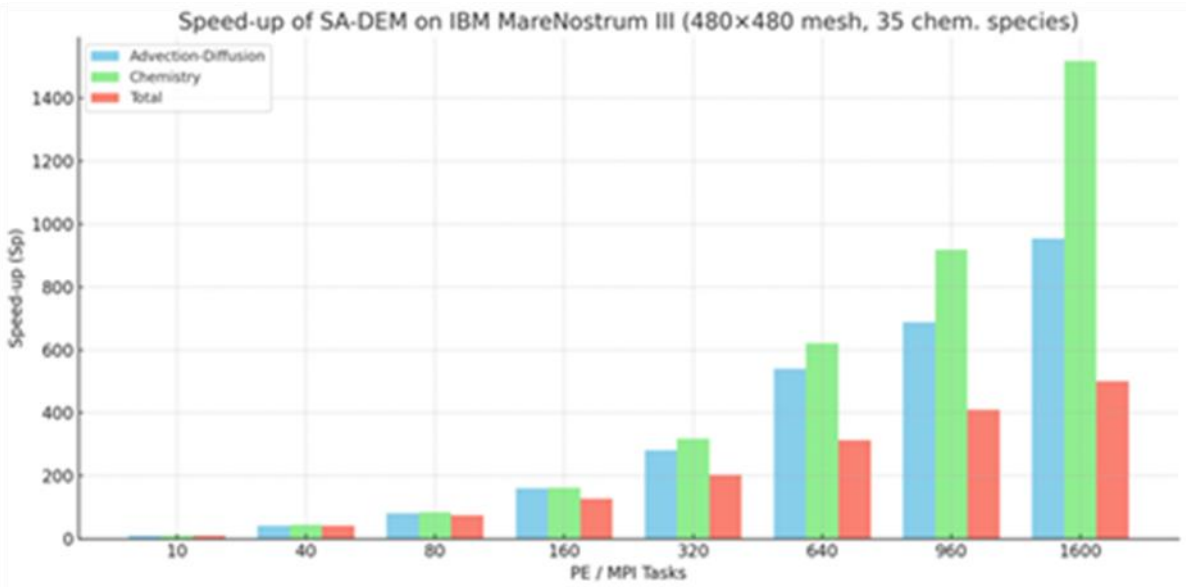

Figure 1. Speed-up (Sp) of Advection–Diffusion (blue), Chemistry (green), and the total (red) for the 2D SA-DEM model on the finest grid, computed on IBM MareNostrum III.

Figure 1 shows the typical behavior of the three sub-computational stages and the overall code. The chemical stage demonstrates almost perfect scaling over a wide range of processes, as it does not need boundary value exchanges and is computationally dominated. Advection-diffusion scales well to intermediate sizes, but enters a regime of diminishing returns earlier due to communications between neighboring sub-areas (boundary cell exchanges). The overall speed-up in the low range of processes often exceeds the linear expectation (superlinear effect), which is explained by better cache-locality and reduced main memory misses when the working set falls into the cache. With increasing processes, the efficiency starts to decline: for chemistry, later and more smoothly; for advection-diffusion, earlier due to an increasing relative share of communications and synchronizations. The balance between the two stages also determines the behavior of the total Sp and $E(p)$: at moderate scales the growth is close to ideal, and after the "tipping point" the communication overhead dominates. This outlines the practical limit of scaling SA-DEM onto MareNostrum III for the given dimensionality and resolution.

3.2. Scalability Results on the Discoverer Supercomputer

The Discoverer runs follow the same experimental setup (480×480 grid, 35 species, Δt = 90sec, NSIZE = 32), for the purpose to get comparable with the MareNostrum III results. The Bull Sequana XH2000 architecture with modern CPUs and high-performance networking provides a higher peak and aggregate core count, which is beneficial for greater parallelism (many simultaneous subsamples) and sensitivity (see Section 3.5).

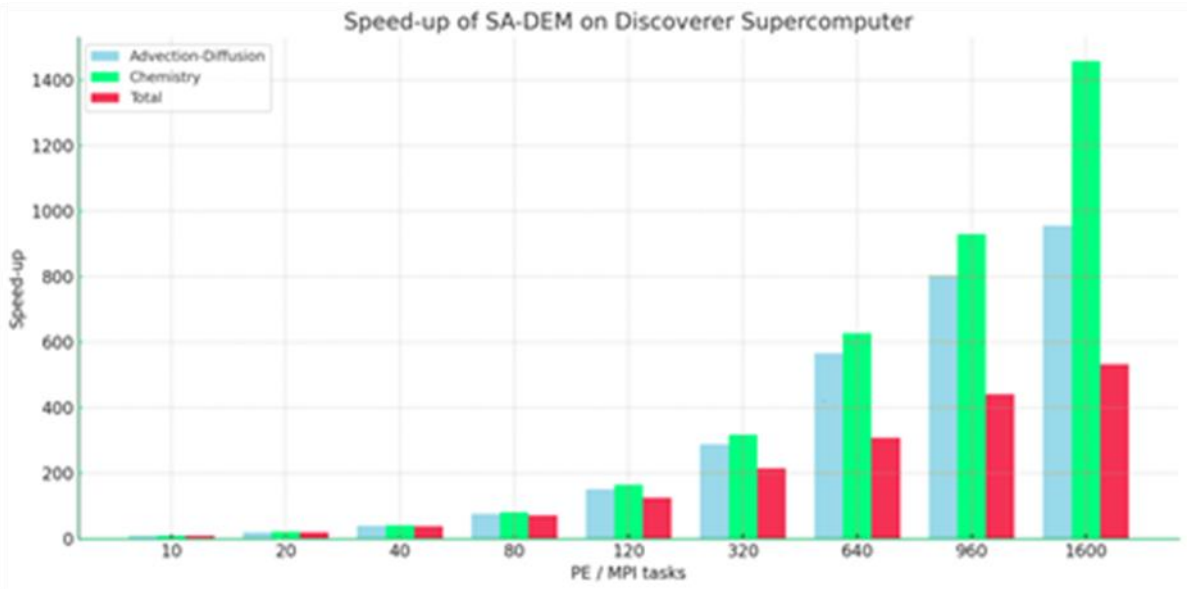

Figure 2. Speed-up (Sp) of Advection–Diffusion (blue), Chemistry (green), and the total (red) for the 2D SA-DEM model on the finest grid, computed on the Discoverer supercomputer (Sofia Tech Park).

Figure 2 shows that the chemical stage is again the "perfect scaler", maintaining high $E(p)$ up to larger scales compared to MareNostrum III. This corresponds to higher computational efficiency and/or lower latency of interprocess operations at the same "hard" domain boundaries. Similar dynamics are observed for advection-diffusion: good efficiency up to an intermediate number of processes and earlier "saturation" at very large configurations, caused by an increasing share of communication steps (exchange at boundaries, barriers, etc.). In the low to medium range, Sp can again be superlinear (cache effect), but the plateau is reached at a higher number of processes, which suggests a deeper cache hierarchy/better memory and network throughput. The overall speed-up and efficiency of the entire model lags closely behind chemistry (compute-dominant) up to the point where the communications of the transport stage start to dictate the pace. The practical conclusion is that Discoverer allows to "push" the limit of useful scaling - especially valuable for ensemble and sensitivity programs.

3.3. Comparative discussion: MareNostrum vs. Discoverer

The chemistry module scales almost perfectly on both platforms across the entire spectrum of experiments performed, while advection-diffusion slows down towards the end of the table (due to fragmentation and increased intensity of boundary communications). On MareNostrum III, the "tipping point" occurs at a smaller number of tasks, while Discoverer maintains high efficiency at a larger scale, which is also reflected in a higher overall Sp. On both systems, a superlinear effect in the low range is reported due to better cache exploitation; the effect persists longer on Discoverer. In summary, the behaviour is qualitatively similar, but quantitatively, Discoverer extends the "zone of most efficient scaling", useful for multi-stage parallelized algorithms.

3.4. Sensitivity analysis and estimation of Sobol indices

The aim is to estimate global sensitivity indices (in particular, Sobol) for SA-DEM outputs against chemical reaction rates. The mathematical formulation (definitions of first-order S_i, total S_i^{tot}). indices, and ANOVA-decomposition) remains unchanged. The estimation is reduced to multidimensional integrals (in this case, 12-dimensional) on $[0,1]^d$, where d is the number of parameters. Several families of MC/QMC schemes are compared: Sobol (small discrepancy series), MSS-1, MSS-2, MSS-2-S (modified MC on Sobol construction), VDC2 (van der Corput, base 2), FIBO (Fibonacci-type constructions), and OPTL (optimal lattice rule). Table 1 gives relative errors for multiple indices (e.g. $S_1, S_2, S_3, S_4, S_5, S_6$- total indices, etc.) for a fixed number of samples (e.g. n=65536).

(1) General pattern of behavior: Two lines stand out as the most reliable "by default": Sobol and MSS-2. For both first-order and total indices, they give systematically low relative errors, with Sobol often being the best for a large set of indices (e.g. S_1, S_2, S_6- some total), and MSS-2 winning for others (e.g., S_2, S_5, S_6, $S_4^{tot}, S_5^{tot}, S_6^{tot}$). This suggests that the Sobol series provides very good coverage of the parameter space and fast convergence, while MSS-2-type modifications are more robust where the estimated quantity is more "difficult" (e.g., with stronger nonlinearities/interactions).

(2) Competing variants: OPTL (optimal lattice rule) and MSS-1 often give errors close to MSS-2 and Sobol, and in some indices they are even the best (e.g. S_3 for OPTL, S_4, and S_1^{tot} for MSS-1). This is in line with the theory: with well-constructed lattices (optimal generator vectors) or with well-chosen sample modifications, lattice rules and hybrid MCs can achieve low errors, especially at moderate effective dimensions.

(3) Weaker performers: VDC2 and FIBO perform significantly worse, especially for small indices (where high sensitivity of the integrator to "fine" effects is needed). In Table 1, this is seen as errors larger by one/several orders of magnitude. The reason is that basic one-dimensional constructive series and simple Fibonacci-type compositions do not provide a sufficiently uniform distribution in higher dimensions and/or are more susceptible to aliasing in subspaces, especially when the function has "sharp" characteristics (hard chemistry, nonlinearities).

(4) Small vs. large indices: At large index values (dominant parameters), most schemes perform well. The differences between Sobol, MSS-2, OPTL, and MSS-1 are minimal. At small indices (weak contributions/interactions) the picture changes: Sobol and MSS-2 retain low error; OPTL/MSS-1 behave decently, but not always; VDC2/FIBO degrade significantly. For applied purposes (ranking significant reactions and eliminating unimportant ones) this difference is critical: if the algorithm overestimates the error at small indices, there is a risk of "missing" subtle but important interactions.

(5) Practical recommendation: For SA-DEM and similar atmospheric models, Sobol and MSS-2 are the "first choice", OPTL/MSS-1 are good alternatives, especially when there are indications of moderate effective dimensionality, and VDC2/FIBO are more suitable for exploratory/fast studies, but not for fine-tuning of small indices. From the point of view of computational resources, all these schemes are perfectly parallelizable, which fits naturally into SA-DEM on cluster systems: samples are distributed across nodes/communicators, and inside each execution the classical domain decomposition operates.

Table 1. Relative error using various Monte Carlo approaches (n = 65536)

EQ	Ref. value	Sobol	MSS-1	MSS-2	MSS-2-S	VDC2	FIBO	OPTL
S_1	4e-1	1e-4	4e-4	2e-4	2e-2	1e-1	4e-2	2e-3
S_2	3e-1	3e-5	2e-4	3e-4	6e-2	5e-2	1e-2	3e-4
S_3	5e-2	2e-4	2e-3	9e-4	8e-2	6e-4	5e-1	9e-5
S_4	3e-1	3e-4	2e-5	2e-4	4e-3	7e-2	1e-2	1e-3
S_5	4e-7	3e-1	7e+0	7e-2	2e+2	7e+2	3e+3	9e-1
S_6	2e-2	3e-4	1e-3	3e-4	4e-2	4e-1	1e+0	6e-3
S_1^{tot}	4e-1	1e-4	4e-5	2e-4	5e-2	9e-2	8e-2	5e-4
S_2^{tot}	3e-1	4e-5	5e-4	2e-4	3e-2	1e-2	3e-2	1e-3
S_3^{tot}	5e-2	3e-4	2e-3	8e-4	4e-2	1e-1	1e+0	2e-3
S_4^{tot}	3e-1	2e-4	5e-4	2e-4	4e-2	1e-1	4e-1	1e-3
S_5^{tot}	2e-4	7e-3	1e-2	4e-3	1e+0	3e+2	9e+1	3e-1
S_6^{tot}	2e-2	4e-4	1e-3	3e-4	4e-2	1e+0	2e+0	2e-3
S_{12}	6e-3	2e-4	5e-3	1e-3	7e-1	9e-1	3e+0	5e-2
S_{14}	5e-3	2e-3	2e-2	2e-3	1e+0	2e-1	8e+0	9e-2
S_{15}	8e-6	3e-2	1e-1	5e-2	5e+0	8e+2	1e+1	5e-1
S_{24}	3e-3	1e-3	2e-2	6e-3	1e+0	3e-1	9e+2	5e-2
S_{45}	1e-5	4e-2	1e-1	2e-2	4e+0	1e-1	4e+1	3e-2

4. Concluding remarks

This work presents the development and parallel implementation of the Danish Eulerian Model (DEM/UNI-DEM) and its sensitivity analysis variant SA-DEM, including the modules for horizontal advection-diffusion, chemistry with emission and deposition, and vertical exchange. The model is formulated as a system of coupled DDEs and is solved with operator-splitting, stable time schemes, and domain decomposition for MPI parallelization; the chemical module uses the condensed mechanism CBM-IV (35 species, 116 reactions). Our scalability experiments on two petascale supercomputers - IBM MareNostrum III (BSC) and Discoverer (Sofia Tech Park) – confirm, that the chemical stage is an almost ideal scaler (without interdomain communications). The scalability of the advection-diffusion stage is limited by the nessesity of exchanging boundary values between subdomains. A superlinear effect is often observed in experiments with small number of tasks (cache-size effect), and the useful scaling limit is higher on Discoverer. For the sake of comparability, an uniform configuration has been used in all experiments (grid size 480×480, 35 species, Δt = 90 s, NSIZE=32). In the analysis of global sensitivity (Sobol indices) versus chemical reaction rates, the estimates of 12-dimensional integrals show a clear advantage of Sobol' and MSS-2 (lowest relative error for multiple indices), close results with OPTL/MSS-1, and significantly weaker performance of VDC2/FIBO, especially for small indices. Parallel execution of MC/QMC samples in SA-DEM allows for practical time-to-result, which makes large-scale sensitivity campaigns on petascale supercomputers possible.

In conclusion, the combination of robust parallel model, HPC, and proper MC/QMC integration algorithm can be applied together to reduce simulations' time, to increase the reliability of model results, and to estimate the influence of certain inaccuracy in the input data or parameters of the chemical scheme to the model output. Furthermore, this work aids in

ranking “by importance“ the chemical reactions, guiding for “safe“ model simplifications, and supports reasonable “what-if” scenarios for clean air policies. Future work could include optimizing I/O and network communications across multiple nodes, exploring other chemical mechanisms, and extending ensemble predictions and training applications.

Acknowledgments

The authors of this article were partially supported by the following projects:

- The work of Tzvetan Ostromsky was supported by the Bulgarian National Science Fund under the project "Integrated approach to creating digital twins of archaeological immovable monuments using innovative technologies”, contract KP-06-N82/1.
- The work of Silvi-Maria Gurova was supported by the National Geoinformation Centre (part of the National Roadmap for Research Infrastructures) through grant No. D01-93/25.106.2025, funded by the Ministry of Education and Science of the Republic of Bulgaria. Both authors (S.-M. Gurova and Venelin Todorov) were partially supported by the Centre of Excellence in Informatics and ICT under the Grant No BG16RFPR002-1.014-0018-C01, financed by the Research, Innovation and Digitalisation for Smart Transformation Programme 2021-2027 and co-financed by the European Union.
- The work of Meglena Lazarova is supported by the “Scientific Research” Fund of Ruse University under Project No. 2025 – FNSE – 02 “Research of Natural and Anthropogenic Phenomena”.

The European High Performance Computing Joint Undertaking (EuroHPC) is acknowledged too for giving us opportunity to use the supercomputer Discoverer.

References

[1] Ten Threats to Global Health in 2019. Available online: https://www.who.int/news-room/spotlight/ten-threats-to-global- 609 health-in-2019 (accessed on 3 October 2025)
[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] Zlatev Z., Computer treatment of large air pollution models, Kluwer, 1995.
[5] Gao Z., Zhou X., A review of the CAMx, CMAQ, WRF-Chem and NAQPMS models: Application, evaluation and uncertainty factors, Environmental Pollution, 343, 2024, 123183.
[6] Ivanov V., Gadzhev G., Ganev K., Downscaling Effect of Dry and Wet Deposition for Nitrogen and Sulphur compounds in Comprising CMAQ and EMEP models over Urban Area, Air Pollution Modeling Its Application XXIX, Springer, 129-134, 2024.
[7] Gadzhev, G., Ivanov, V. (2021). Modelling of the Seasonal Sulphur and Nitrogen Depositions over the Balkan Peninsula by CMAQ and EMEP-MSC-W, Studies in Systems, Decision and Control, Springer Cham. , 361,171-183, 2021.
[8] Discoverer HPC Doc, Available online: https://docs.discoverer.bg/index.html (accessed on 3 October 2025)
[9] Technical documentation of MareNostrum III: Available online: https://www.bsc.es/marenostrum/marenostrum/mn3 (accessed on 3 October 2025)
[10] Website of the Danish Eulerian Model: Available online: https://www2.dmu.dk/AtmosphericEnvironment/DEM (accessed on 3 October 2025)
[11] Saltelli A. et al., Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Halsted Press New York, 2004.
[12] 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.
[13] Amann M. et al., Global scale Modelling within EMEP: Progress report, 2021
[14] The EMEP grid documentation: Available online: https://www.emep.int/mscw/emep_grid.html (accessed on 3 October 2025)

[15] Gropp W., Lusk E., Skjellum A., Using MPI: Portable Parallel Programming with the Message-Passing Interface, MIT Press, 2014.
[16] Quinn M.J., Parallel Programming in C with MPI and OpenMP, McGraw-Hill Education, 2003.