Parallel implementation and one year experiments with the Danish Euleian Model

Tzvetan Ostromsky[1], Ivan Dimov[1], and Zahari Zlatev[2]

[1] Institute for Parallel Processing, Bulgarian Academy of Sciences,
Acad. G. Bonchev str., bl. 25-A, 1113 Sofia, Bulgaria,
E-mail: ceco@parallel.bas.bg, ivdimov@bas.bg,
Web-page: http://www.bas.bg/clpp/
[2] National Environmental Research Institute, Department of Atmospheric Environment, Frederiksborgvej 399 P. O. Box 358, DK-4000 Roskilde, Denmark,
E-mail: zz@dmu.dk
Web-page: http://www.dmu.dk/AtmosphericEnvironment

Abstract. Large scale air pollution models are powerful tools, designed to meet the increasing demand in different environmental studies. The atmosphere is the most dynamic component of the environment, where the pollutants can be moved quickly on far distnce. Therefore the air pollution modeling must be done in a large computational domain. Moreover, all relevant physical, chemical and photochemical processes must be taken into account. In such complex models operator splitting is very often applied in order to achieve sufficient accuracy as well as efficiency of the numerical solution.
The Danish Eulerian Model (DEM) is one of the most advanced such models. Its space domain ($4800 \times 4800\ km$) covers Europe, most of the Mediterian and neighboring parts of Asia and the Atlantic Ocean. Efficient parallelization is crucial for the performance and practical capabilities of this huge computational model. Different splitting schemes, based on the main processes mentioned above, have been implemented and tested with respect to accuracy and performance in the new version of DEM. Some numerical results of these experiments are presented in this paper.

1 INTRODUCTION

The problem for air pollution modelling has been studied for years [8, 15]. An air pollution model is generally described by a system of partial differential equations for calculating the concentrations of a number of chemical species (pollutants and other components of the air that interact with the pollutants) in a large 3-D domain (part of the atmosphere above the studied geographical region). The main physical and chemical processes (horizontal and vertical wind, diffusion, chemical reactions, emissions and deposition) should be adequately represented in the system.

The Danish Eulerian Model (DEM) [1, 10, 14–16] is mathematically represented by the following system of partial differential equations:

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

where

- c_s – the concentrations of the chemical species;
- $u,\ v,\ w$ – the wind components along the coordinate axes;
- $K_x,\ K_y,\ K_z$ – diffusion coefficients;
- E_s – the emissions;
- $k_{1s},\ k_{2s}$ – dry / wet deposition coefficients;
- $Q_s(c_1, c_2, \dots c_q)$ – non-linear functions describing the chemical reactions between species under consideration (Gery et al. (1989)).

2 SPLITTING INTO SUBMODELS

The above rather complex system (1) is split into 3 subsystems / submodels, according to the major physical / chemical processes and the numerical methods applied in their solution.

$$\frac{\partial c_s^{(1)}}{\partial t} = -\frac{\partial(uc_s^{(1)})}{\partial x} - \frac{\partial(vc_s^{(1)})}{\partial y} + \frac{\partial}{\partial x}\left(K_x \frac{\partial c_s^{(1)}}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial c_s^{(1)}}{\partial y}\right) = A_1 c_s^{(1)}(t)$$

horizontal advection & diffusion

$$\frac{\partial c_s^{(2)}}{\partial t} = E_s + Q_s(c_1^{(2)}, c_2^{(2)}, \dots c_q^{(2)}) - (k_{1s} + k_{2s})c_s^{(4)} = A_2 c_s^{(2)}(t)$$

chemistry, emissions & deposition

$$\frac{\partial c_s^{(3)}}{\partial t} = -\frac{\partial(wc_s^{(3)})}{\partial z} + \frac{\partial}{\partial z}\left(K_z \frac{\partial c_s^{(3)}}{\partial z}\right) = A_3 c_s^{(3)}(t)$$

vertical transport

Various splitting schemes have been proposed and analysed in [2–5, 8, 9, 13] The three splitting schemes, discussed in this paper and used in our experiments, are briefly described below.

2.1 Sequential splitting scheme, used by default in UNI-DEM

$$\left.\begin{array}{c} \frac{dc_s^{(1)}(t)}{dt} = A_1 c_s^{(1)}(t), \quad t \in ((k-1)\tau, k\tau] \\ c_s^{(1)}((k-1)\tau) = c_s^{(2)}((k-1)\tau) \end{array}\right\} \tag{2}$$

$$\left.\begin{array}{c} \frac{dc_s^{(2)}(t)}{dt} = A_2 c_s^{(2)}(t), \quad t \in ((k-1)\tau, k\tau] \\ c_s^{(2)}((k-1)\tau) = c_s^{(1)}(k\tau) \end{array}\right\} \tag{3}$$

$s = 1, 2, \ldots, q($ q – the number of chemical species).

Equations (2)–(3) describe the sequential splitting scheme in the 2-D case (without the vertical transport). The splitting error of this scheme, used by default in UNI-DEM, is $\mathcal{O}(\tau)$, where τ is the time step.

2.2 Marchuck - Strang splitting scheme

$$\left.\begin{array}{c} \frac{dc_s^{(1)}(t)}{dt} = A_1 c_s^{(1)}(t), \quad t \in ((k-1)\tau, (k-\frac{1}{2})\tau] \\ c_s^{(1)}((k-1)\tau) = \hat{c}^{(1)}((k-1)\tau) \end{array}\right\} \tag{4}$$

$$\left.\begin{array}{c} \frac{dc_s^{(2)}(t)}{dt} = A_2 c_s^{(2)}(t), \quad t \in ((k-1)\tau, k\tau] \\ c_s^{(2)}((k-1)\tau) = w_k^{(1)}((k-\frac{1}{2})\tau) \end{array}\right\} \tag{5}$$

$$\left.\begin{array}{c} \frac{d\hat{c}_s^{(1)}(t)}{dt} = A_1 \hat{c}_s^{(1)}(t), \quad t \in ((k-\frac{1}{2})\tau, k\tau] \\ \hat{c}_s^{(1)}((k-\frac{1}{2})\tau) = c_s^{(2)}(k\tau) \end{array}\right\} \tag{6}$$

$s = 1, 2, \ldots, q($ q – the number of chemical species).

Equations (4)–(6) describe the symmetric splitting scheme (due to Marchuck and Strang) in the 2-D case (without the vertical transport). This scheme has higher order of accuracy, $\mathcal{O}(\tau^2)$, where τ is the time step.

2.3 Weighted sequential splitting scheme

The sequential splitting scheme (2)–(3) is applied twice on each step with reverse order of the two operators A_1 and A_2. The average of the two results for $c_s((k-1)\tau)$ is taken as initial value for calculations of the $c_s(k\tau)$ on the next step. This scheme has also second order of accuracy, $\mathcal{O}(\tau^2)$.

3 PARALLELIZATION STRATEGY AND NUMERICAL METHODS, USED IN THE SOLUTION OF THE SUBMODELS

Although the splitting is a crucial step in the efficient numerical treatment of the model, after discretization of the large computational domain each submodel becomes itself a huge computational task. In addition, the dynamics of the chemical and photo-chemical processes requires using of small time-step to keep stability of the computations. Large parallel supercomputers must be used in order to meet the high speed and storage requirements. Moreover, development and impementation of efficient parallel algorithms is very important for improving the practical capabilities and the performance of DEM. That topic has been discussed in more detail in [10–12, 15, 16].

Distributed memory parallelization model via MPI [6] is used in the current UNI-DEM version considered in this work. For maximum portability only standard MPI routines are used in the UNI-DEM code. Parallelization is based on domain decomposition of the horizontal grid, which implies certain restrictions on the number of MPI tasks and requires communication on each time step. Improving the data locality for more efficient cache utilization is achieved by using *chunks* to group properly the small tasks in the chemistry-deposition and vertical exchange stages.

Additional **pre-processing** and **post-processing** stages are needed for scattering the input data and gathering the results. These are cheap, but their relative weight grows up with increasing the number of MPI tasks, affecting the total speed-up and scalability.

The numerical methods, used in the solution of the submodels, are given below.

- **Advection-diffusion part:** Finite elements, followed by predictor-corrector schemes with several different correctors. The native parallel tasks in this case are the calculations for a given pollutant in a given layer. There are enough parallel tasks, sometimes too big to fit into the cache (depends on the subdomain size).
- **Chemistry-deposition part:** An improved version of the QSSA (Quazi Steady-State Approximation) (Hesstvedt et al. - [7]). The native parallel tasks here are the calculations in a single grid-point. These small tasks are grouped in *chunks* for efficient cache utilization.
- **Vertical transport:** Finite elements, followed by θ-methods. The native parallel tasks here are the calculations along each vertical grid-line. The number of these tasks is large, while the tasks are relatively small. They also can be grouped in *chunks*, like those in the chemical stage.

4 NUMERICAL RESULTS

The experiments, presented in this section, are performed on the SUN Sunfire 6800 parallel machine (24 CPU UltraSparc-III / 750 MHz), located at the Danish

Technical University (DTU) in Lyngby, Denmark. All experiments are for a time period of 1 year (1998). The number of chemical species considered by the chemical submodel in the current version is $q = 35$. Chunk size equal to 48 is used in the experiments, which seems to be optimal for this problem and the target machine.

4.1 Accuracy results

These results are obtained by experiments with real data (meteorological data sets, collected by the Norwegian Meteorological Institute; as well as emmision data) for 1998. The results are compared to the records of several stationary measuring stations throughout Europe, wich have enough measurements. It is not possible to extract only the splitting error in these experiments. Other errors, which are present in the results, are as follows:

- Input data error;
- Error of the numerical methods, used in the different submodels;
- Spatial discretization error (depending on the grid step);
- Computational (rounding) error;

and so on. In addition, the error of the measurement instruments is also included. Such accuracy results for three major pollutants: nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and amonia (NH_3 + NH_4); are given in Tables 1 – 3 respectively.

UNI-DEM results for NO_2 on ($96 \times 96 \times 1$) grid, in comparison with the measurements of 35 stations in Europe										
Month	Concentration [mol/l]				‖Discrepancy %‖			Correlation		
	Obs.	**Seq.**	**Weig.**	**M.-S.**	**Seq.**	**Weig.**	**M.-S.**	**Seq.**	**Weig.**	**M.-S.**
January	2.79	3.20	3.21	2.35	13	13	19	0.67	0.66	0.62
February	3.24	4.96	4.97	3.83	35	35	15	0.61	0.61	0.59
March	2.18	1.83	1.84	1.38	19	19	58	0.69	0.69	0.75
April	1.98	1.76	1.76	1.39	12	12	42	0.59	0.59	0.49
May	1.81	2.05	2.05	1.66	12	12	9	0.64	0.64	0.66
June	1.58	2.33	2.34	1.95	32	32	19	0.56	0.56	0.47
July	1.57	1.86	1.86	1.56	15	16	1	0.50	0.50	0.49
August	1.70	1.99	1.99	1.81	15	15	6	0.59	0.59	0.57
September	1.84	2.42	2.42	2.11	24	24	13	0.65	0.65	0.56
October	1.96	2.38	2.39	1.85	18	18	6	0.57	0.57	0.53
November	3.10	3.97	3.97	3.04	22	22	2	0.65	0.65	0.70
December	3.33	4.10	4.11	3.10	19	19	8	0.70	0.69	0.71
mean	2.26	2.74	2.74	2.17	17	17	4	0.69	0.69	0.68

Table 1. NO_2 results (concentrations, discrepancy and correlation factor) for 1998.

The following abbreviations are used in the column headings of these tables:
Obs. - observed by the measurement stations;
Seq. - computed by DEM with sequenial splitting;
Weig. - computed by DEM with weighted sequenial splitting;
M.-S. - computed by DEM with Marchuck - Strang splitting.

UNI-DEM results for SO_2 on $(96 \times 96 \times 1)$ grid, in comparison with the measurements of 35 stations in Europe										
Month	Concentration [mol/l]				‖Discrepancy %‖			Correlation		
	Obs.	**Seq.**	**Weig.**	**M.-S.**	**Seq.**	**Weig.**	**M.-S.**	**Seq.**	**Weig.**	**M.-S.**
January	1.36	2.70	2.72	2.49	50	50	46	0.83	0.83	0.83
February	1.41	2.02	2.04	1.91	30	31	26	0.61	0.61	0.69
March	1.11	1.49	1.51	1.41	26	26	22	0.64	0.64	0.64
April	0.80	1.06	1.07	1.01	25	25	21	0.60	0.60	0.64
May	0.73	1.13	1.14	1.12	35	36	35	0.71	0.71	0.72
June	0.51	0.72	0.72	0.72	29	30	30	0.72	0.72	0.70
July	0.49	0.63	0.64	0.62	22	23	20	0.82	0.83	0.83
August	0.56	0.73	0.74	0.75	23	24	25	0.63	0.63	0.71
September	0.60	0.85	0.88	0.86	29	29	30	0.82	0.82	0.81
October	0.58	0.68	0.69	0.67	16	17	15	0.75	0.75	0.81
November	1.32	1.59	1.60	1.58	17	18	17	0.29	0.28	0.45
December	1.41	2.89	2.91	2.64	51	52	47	0.39	0.39	0.52
mean	0.91	1.38	1.39	1.32	34	34	31	0.71	0.71	0.77

Table 2. SO_2 results (concentrations, discrepancy and correlation factor) for 1998.

UNI-DEM results for $NH_3 + NH_4$ on $(96 \times 96 \times 1)$ grid, in comparison with the measurements of 24 stations in Europe										
Month	Concentration [mol/l]				‖Discrepancy %‖			Correlation		
	Obs.	**Seq.**	**Weig.**	**M.-S.**	**Seq.**	**Weig.**	**M.-S.**	**Seq.**	**Weig.**	**M.-S.**
January	1.02	1.27	1.37	1.04	20	26	2	0.78	0.78	0.80
February	1.51	1.43	1.51	1.18	5	0	28	0.79	0.79	0.79
March	1.19	1.42	1.52	1.16	16	22	2	0.67	0.64	0.67
April	1.36	1.29	1.35	1.12	6	0	21	0.65	0.64	0.63
May	1.46	1.67	1.75	1.31	12	17	11	0.76	0.75	0.69
June	1.20	0.99	1.02	0.93	21	18	29	0.74	0.74	0.74
July	1.09	1.00	1.03	0.97	9	6	12	0.72	0.72	0.68
August	1.01	1.17	1.18	1.06	13	15	4	0.83	0.83	0.82
September	1.39	1.86	1.96	1.53	25	29	9	0.78	0.77	0.80
October	0.73	0.89	0.93	0.76	18	21	4	0.65	0.64	0.67
November	1.19	1.82	1.95	1.59	34	39	25	0.65	0.64	0.66
December	1.13	1.70	1.83	1.46	33	38	22	0.76	0.77	0.77
mean	1.19	1.38	1.45	1.18	13	17	1	0.78	0.77	0.77

Table 3. Amonia results (concentrations, discrepancy and correlation factor) for 1998.

4.2 Performance results

Some time and speed-up results, showing the performance of UNI-DEM on Sunfire 6800 computing system, are presented in Table 4.

Time / **(speed-up)** results of UNI-DEM on a SunFire 6800 at DTU					
Grid size / stage	**1 proc.**	**2 proc.**	**4 proc.**	**8 proc.**	**16 proc.**
$[96 \times 96 \times 1]$ – total	36270	18590	10296	4788	2492
		(1.95)	**(3.52)**	**(7.58)**	**(14.55)**
– **advection**	9383	4882	2512	1540	792
		(1.92)	**(3.74)**	**(6.09)**	**(11.85)**
– **chemistry**	24428	13120	7122	3088	1544
		(1.86)	**(3.43)**	**(7.91)**	**(15.82)**
$[96 \times 96 \times 10]$ – total	415442	211370	108620	52920	30821
		(1.97)	**(3.82)**	**(7.85)**	**(13.48)**
– **advection**	132660	64759	31367	15422	8612
		(2.05)	**(4.23)**	**(8.60)**	**(15.40)**
– **chemistry**	263584	133496	68138	34255	17621
		(1.97)	**(3.87)**	**(7.69)**	**(14.96)**
$[288 \times 288 \times 1]$ – total	699622	352470	197811	98033	44465
		(1.98)	**(3.54)**	**(7.14)**	**(15.95)**
– **advection**	348024	169773	85812	42558	20782
		(2.05)	**(4.05)**	**(8.18)**	**(16.75)**
– **chemistry**	294155	150088	84504	37923	18498
		(1.96)	**(3.48)**	**(7.76)**	**(15.90)**

Table 4. Time in seconds and the **(speed-up)** (given in brackets below) of UNI-DEM for running an 1-year experiment on different size grids (2-D and 3-D versions) on a Sunfire 6800 machine at DTU.

5 CONCLUSIONS AND PLANS FOR FUTURE WORK

- By using high performance parallel computers to run the variable grid-size code UNI-DEM, reliable results for a large region (whole Europe) and for a very long period (one or several years) can be obtained within a reasonable time.
- In most cases the results, obtained with the weighted sequential splitting, are quite similar to those of the sequential splitting. This could be an indication for a small commutator of the two operators.
- The splitting is not the only source of error in UNI-DEM. Nevertheless, its contribution in the total error seems to be significant. For some species (NO_2, SO_2, amonia) the Marchuck-Strang scheme gives results, closer to the measuremets than the other two splitting methods. More experiments are needed in order to investigate the consistency of such behaviour.
- The parallel code, created by using MPI standard library, appears to be highly portable and shows good efficiency and scalability. The limited size of the fast cache-memory causes superlinear speed-up in some cases.

ACKNOWLEDGMENTS

This research was supported in part by grant I-901/99 from the Bulgarian NSF, grant ICA1-CT-2000-70016 – Center of Excellence and by NATO grant NATO ARW "Impact of future climate changes on pollution levels in Europe". A grant from the Danish Natural Sciences Research Council gave us access to all Danish supercomputers.

References

1. V. Alexandrov, A. Sameh, Y. Siddique and Z. Zlatev, Numerical integration of chemical ODE problems arising in air pollution models, Env. Modeling and Assessment, 2 (1997) 365–377.
2. M. Botchev, I. Faragó, A. Havasi, Testing weighted splitting schemes on a one-column transport-chemistry model, Int. J. Env. Pol., Vol. 22, No. 1/2.
3. P. Csomós, I. Faragó, A. Havasi, Weighted sequential splittings and their analysis, Comput. Math. Appl., (to appear)
4. I. Dimov, I. Faragó, A. Havasi, Z. Zlatev, L-commuataivity of the operators in splitting methods for air pollution models, Annales Univ. Sci. Sec. Math., 44 (2001) 127-148.
5. I. Faragó, A. Havasi, On the convergence and local splitting error of different splitting schemes, Progress in Computational Fluid Dynamics, (to appear)
6. W. Gropp, E. Lusk and A. Skjellum, Using MPI: Portable programming with the message passing interface, MIT Press, Cambridge, Massachusetts, 1994.
7. E. Hesstvedt, Ø. Hov and I. A. Isaksen, Quasi-steady-state approximations in air pollution modeling: comparison of two numerical schemes for oxidant prediction, Int. Journal of Chemical Kinetics 10 (1978), pp. 971–994.
8. G. I. Marchuk, Mathematical modeling for the problem of the environment, Studies in Mathematics and Applications, No. 16, North-Holland, Amsterdam, 1985.
9. G. J. McRae, W. R. Goodin and J. H. Seinfeld, Numerical solution of the atmospheric diffusion equations for chemically reacting flows, J. Comp. Physics 45 (1984), pp. 1–42.
10. Tz. Ostromsky, W. Owczarz, Z. Zlatev, Computational Challenges in Large-scale Air Pollution Modelling, Proc. 2001 International Conference on Supercomputing in Sorrento, ACM Press (2001), pp. 407-418
11. Tz. Ostromsky, Z. Zlatev, Parallel Implementation of a Large-scale 3-D Air Pollution Model, Large Scale Scientific Computing (S. Margenov, J. Wasniewski, P. Yalamov, Eds.), LNCS-2179, Springer, 2001, pp. 309-316.
12. Tz. Ostromsky, Z. Zlatev, 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, Springer (2002), 545-554.
13. G. Strang, On the construction and comparison of difference schemes, SIAM J. Numer. Anal. **5**, No. 3 (1968), 506-517.
14. WEB-site of the Danish Eulerian Model, available at: *http://www.dmu.dk/AtmosphericEnvironment/DEM*
15. Z. Zlatev, Computer treatment of large air pollution models, Kluwer, 1995.
16. Z. Zlatev, I. Dimov, K. Georgiev, Three-dimensional version of the Danish Eulerian Model, Zeitschrift für Angewandte Mathematik und Mechanik, 76 (1996) pp. 473–476.