Parallel Implementation of a Large-Scale 3-D Air Pollution Model

Tzvetan Ostromsky[1] and Zahari Zlatev[2]

[1] Central Laboratory for Parallel Processing, Bulgarian Academy of Sciences, Acad. G.Bonchev str., bl. 25-A, 1113 Sofia, Bulgaria; ceco@copern.bas.bg; http://copern.bas.bg/~ceco

[2] National Environmental Research Institute, Department of Atmospheric Environment, Frederiksborgvej 399, P. O. Box 358, DK-4000 Roskilde, Denmark, zz@dmu.dk; http://www.dmu.dk/AtmosphericEnvironment

Abstract. Air pollution models can efficiently be used in different environmental studies. The atmosphere is the most dynamic component of the environment, where the pollutants can be transported over very long distances. Therefore the models must be defined on a large space domain. Moreover, all relevant physical and chemical processes must be adequately described. This leads to huge computational tasks. That is why it is difficult to handle numerically such models even on the most powerful up-to-date supercomputers.

The particular model used in this study is the Danish Eulerian Model. The numerical methods used in the advection-diffusion part of this model consist of finite elements (for discretizing the spatial derivatives) followed by predictor-corrector schemes with several different correctors (in the numerical treatment of the resulting systems of ordinary differential equations). Implicit methods for the solution of stiff systems of ordinary differential equations are used in the chemistry part. This implies the use of Newton-like iterative methods. A special sparse matrix technique is applied in order to increase the efficiency. The model is constantly updated with new faster and more accurate numerical methods.

The three-dimensional version of the Danish Eulerian Model is presented in this work. The model is defined on a space domain of 4800 km $\times$ 4800 km that covers the whole of Europe together with parts of Asia, Africa and the Atlantic Ocean. A chemical scheme with 35 species is used in this version. Two parallel implementations are discussed; the first one for shared memory parallel computers, the second one – the newly developed version for distributed memory computers. Standard tools are used to achieve parallelism: OpenMP for shared memory computers and MPI for distributed memory computers. Results from many experiments, which were carried out on a SUN SMP cluster and on a CRAY T3E at the Edinburgh Parallel Computer Centre (EPCC), are presented and analyzed.

Keywords: air pollution model, system of PDE's, parallel algorithm, shared memory computer, distributed memory computer, OpenMP, MPI.

S. Margenov, J. Wasniewski, and P. Yalamov (Eds.): ICLSSC 2001, LNCS 2179, pp. 309–316, 2001.

1 Need of Parallel Computations in Large-Scale Air Pollution Modeling

High pollution levels can be harmful for plants, animals and human beings. This is why it is necessary to control the pollution and to take some preventive measures when necessary. Mathematical models can successfully be used in many different studies related to high pollution levels and the consequent damaging effects. The results from the models, however, must be reliable. This implies a demand for an adequate description of all relevant physical and chemical processes involved in the models. The consequence is that the models are normally very big and have to be treated on fast supercomputers.

The performance and scalability of the three-dimensional (3-D) versions of the Danish Eulerian Model (DEM) [16,19,20] on various parallel computers is discussed in this paper. The recently developed MPI version of the 3-D model is introduced. We concentrate our attention on the implementation of the parallel algorithms and the performance on high-speed computers. Different features of this model have been described in many publications. Comparisons of results obtained with DEM and measurements taken over land are discussed in [16,17,18]. Results obtained with this model have also been compared with measurements taken over sea, see [7]. Different environmental studies, in which DEM was successfully used, are described in [2,3,4,9,21,22,22]. Many other results can be found in the web-site of the model [14].

2 Mathematical Description of the Model

The 3-dimensional version of the Danish Eulerian Model (DEM) is presented in this work. It is represented mathematically by the following system of PDE's:

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

where c_s are the concentrations of the chemical species involved in the model, u, v and w are the wind components, K_x, K_y and K_z are diffusion coefficients, E_s are the emissions, κ_{1s} and κ_{2s} are the coefficients for dry and wet deposition, respectively, and $Q_s(c_1, c_2, \ldots, c_q)$ are expressions that describe the chemical reactions under consideration.

It is difficult to treat the PDE system (1) directly. Therefore, different kinds of splitting is used. A splitting procedure, based on ideas proposed in [7,8], leads to five sub-models, representing the main physical and chemical processes ($s = 1, 2, \ldots, q$): the horizontal advection (2), the horizontal diffusion (3), the

chemistry and the emission (4), the deposition (5) and the vertical exchange (6):

$$\frac{\partial c_s^{(1)}}{\partial t} = -\frac{\partial (u c_s^{(1)})}{\partial x} - \frac{\partial (v c_s^{(1)})}{\partial y} \tag{2}$$

$$\frac{\partial c_s^{(2)}}{\partial t} = \frac{\partial}{\partial x}\left(K_x \frac{\partial c_s^{(2)}}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial c_s^{(2)}}{\partial y}\right) \tag{3}$$

$$\frac{dc_s^{(3)}}{dt} = E_s + Q_s(c_1^{(3)}, c_2^{(3)}, \ldots, c_q^{(3)}) \tag{4}$$

$$\frac{dc_s^{(4)}}{dt} = -(\kappa_{1s} + \kappa_{2s}) c_s^{(4)} \tag{5}$$

$$\frac{\partial c_s^{(5)}}{\partial t} = -\frac{\partial (w c_s^{(5)})}{\partial z} + \frac{\partial}{\partial z}\left(K_z \frac{\partial c_s^{(5)}}{\partial z}\right) \tag{6}$$

If the model (1) is split into the five sub-models (2) - (6), then the discretization of the spatial derivatives in the right-hand-sides of the sub-models leads to the solution (successively at each time-step) of five systems ($i = 1, 2, 3, 4, 5$) of ordinary differential equations (ODE's):

$$\frac{dg^{(i)}}{dt} = f^{(i)}(t, g^{(i)}), \quad g^{(i)} \in R^{N_x \times N_y \times N_z \times N_s}, \quad f^{(i)} \in R^{N_x \times N_y \times N_z \times N_s}, \tag{7}$$

where N_x, N_y and N_z are the numbers of grid-points along the coordinate axes and $N_s = q$ is the number of chemical species. The functions $f^{(i)}$, $i = 1, 2, 3, 4, 5$, depend on the particular discretization methods used in the numerical treatment of the different sub-models, while the functions $g^{(i)}$, $i = 1, 2, 3, 4, 5$, contain approximations of the concentrations at the grid-points of the space domain (more details about the splitting procedure and about the numerical treatment of the obtained in this way sub-models can be found in [1,6,8,9,12,15,16]).

The space domain of the model is part of the hemisphere that covers Europe together with neighbouring parts of Asia, Africa and the Atlantic Ocean. It has been discretized by using a (96×96) grid when the basic version is used. This means that horizontaly the domain is divided into 9216 grid-squares of size approximately $(50\ km \times 50\ km)$. There exist versions of the model, which are discretized (in the horizontal plane) on a coarser or finer grids; see Table 1. In the vertical direction the grid is non-equidistant (the increments are smaller close to the surface, while large increments are used when the distance to the surface is increased). Ten layers are used at present. If the resolution in the horizontal direction is fine, then the model is currently used only as a two-dimensional model (see Table 1). If a long sequence of scenarios has to be run, then again the two-dimensional versions are used. This fact illustrates the need for further improvements (faster numerical algorithms, better exploitation of the potential power of the modern supercomputers, faster and bigger supercomputers, etc.).

Table 1. Information about the different versions of the Danish Eulerian Model (the size of the grid given in the third column is for the corresponding 2-D version of the model; this number should be multiplied by the number of layers to get the size of the domain in the 3-D version).

Grid	Grid-Squares	Size of the grid	3-D version
(32×32)	$(150\ km \times 150\ km)$	1024	Yes
(96×96)	$(50\ km \times 50\ km)$	9216	Yes
(288×288)	$(16.7\ km \times 16.7\ km)$	82944	No
(480×480)	$(10\ km \times 10\ km)$	230400	No

The numerical methods that are currently used in DEM are (i) finite elements in the treatment of the horizontal diffusion and advection followed by predictor-corrector schemes with several different correctors, (ii) an improved version of the Quasi-Steady-State-Approximation in the chemical part, and (iii) finite elements followed by θ-method in the vertical exchange part. More details about the numerical methods can be found in [1,6,8,9,12,15,16]).

In the rest of this paper we describe different ways for achieving high performance on two types of parallel computers: (i) computers with shared memory, represented by a SUN E-6500 system, and (ii) computers with distributed memory, represented by a CRAY T3E. It should be emphasized that no special properties of these particular computers are used in the codes. Thus, good results could be expected when the codes are run on other, either shared or distributed memory computers.

3 Runs on Shared Memory Computers

Only standard OpenMP [13] commands are used in the code for parallel computers with shared memory. Some results of experiments with this code are given in Table 2. It is important to identify the parallel tasks and to group them in an appropriate way when necessary. For the main parts of the code this is done as follows:

- **Horizontal advection and diffusion.** It can easily be seen that, after the splitting procedure, the performance of the horizontal advection can be carried out independently for every chemical compound (and for every layer in the 3-D version). This means that the number of parallel tasks is equal to the number of chemical compounds (multiplied by the layers when the 3-D version is used). The same is true for the horizontal diffusion. Moreover, the advection and the diffusion submodels can be treated together, as already mentioned. Thus, there are enough parallel tasks in this part of the code and the parallel tasks are very big.
- **Chemistry and deposition.** These two processes can be carried out in parallel for every grid-point. This means that there are many parallel tasks (the

Table 2. Time in seconds and speed-up (in brackets) of the main stages of the 3-D OpenMP version of DEM, (96×96×10) grid. The results are obtained on a SUN cluster in the EPCC and on an SGI ORIGIN 2000 at UNI•C, Denmark, by using chunks of size 48, which appears to be optimal for these machines.

3-D OMP version of DEM on SUN E6500 / 400MHz (NSIZE=48)				
Stage	**Time [sec.] (*Speed-up*)**			
	1 proc.	4 proc.	8 proc.	16 proc.
Wind+Sinks	78	80 (*1.0*)	73 (*1.1*)	106 (*0.7*)
Advection+Diffus.	8885	2393 (*3.8*)	1255 (*7.3*)	797 (*11.1*)
Chemistry+Depos.	25824	6490 (*4.0*)	3523 (*7.3*)	2069 (*12.5*)
Vertical transport	2459	616 (*4.0*)	310 (*7.9*)	172 (*14.3*)
Output operations	214	212 (*1.0*)	217 (*1.0*)	338 (*0.6*)
Total (SUN)	37890	9792 (*3.9*)	5379 (*7.0*)	3483 (*10.9*)
ORIGIN 2000	42406	11189 (*3.8*)	6257 (*6.8*)	3471 (*12.2*)

number of parallel tasks is equal to the number of grid-points), but each task is small. Therefore, the tasks should be grouped appropriately. This can be done by using chunks. Both the procedure of splitting the data into chunks and the effect of using chunks are discussed in detail in [5].

– **Vertical exchange.** The performance of the vertical exchange along each vertical grid-line is a parallel task. The number of these tasks is $N_x \times N_y$. If the grid is fine, then the number of these tasks is large. However, they are not very big and have to be grouped. This is done by trying to distribute equally the tasks among the processors.

4 Runs on Distributed Memory Computers

The MPI (Message Passing Interface, [6]) is used in the code for distributed memory computers. In the MPI implementation, the space domain is divided into several sub-domains (the number of these sub-domains being equal to the number of processors assigned to the job). Then each processor works on its own sub-domain. Similarly to the OpenMP version, chunks are used in the chemistry-deposition part in order to exploit the data locality in the big 3-D arrays. Some results of experiments with this version are given in Table 3 and in Fig. 1.

Two additional procedures, a *pre-processing* and a *post-processing*, are used for scattering the input data and gathering the results in the beginning and in the end of the run respectively.

– **Pre-processing.** In the beginning of the job the input data (the meteorological data and the emission data) are distributed (consistently with the sub-domains) to the processors. In this way, not only is each processor working on its own sub-domain, but it has also access to all meteorological and emission data for its sub-domain.

Table 3. Time in seconds and relative weight of the main stages of the 3-D MPI version of DEM, $(96 \times 96 \times 10)$ grid. The results are obtained on a CRAY T3E computer at the EPCC by using chunks of size 24 in the chemistry part. The ratio between the times on 8 and 32 processors is given in the last column.

3-D MPI version of DEM on T3E computer (NSIZE=24)			
Stage	**Time [sec.]**	**(% of Total)**	**Scal. factor**
	8 processors	32 processors	T(8)/T(32)
Preprocess	44 (0.5 %)	39 (1.5 %)	1.1
Wind+Sinks	29 (0.3 %)	8.3 (0.3 %)	3.5
Advection+Diffusion	2060 (22.6 %)	647 (25.0 %)	3.2
Chemistry+Depos.	5945 (65.2 %)	1548 (59.9 %)	3.8
Vertical transport	502 (5.7 %)	126 (4.9 %)	4.0
Output operations	21 (0.2 %)	5.4 (0.2 %)	3.9
Communications	480 (5.3 %)	181 (7.0 %)	2.7
Postprocess	18 (0.2 %)	21 (0.8 %)	0.9
Total	9119 (100 %)	2585 (100 %)	3.5

– **Post-processing.** During the run each processor prepares its own output data. At the end of the run all the data are collected and prepared for future use by one of the processors during the post-processing procedure.

The use of the pre-processing and post-processing procedures is done in order to reduce as much as possible the communications during the actual computations. However, some communications are to be carried out during the computations. The time needed for these communications is very small (normally, several percent). This time includes some idle time as the tasks executed on different processors are not perfectly balanced.

More details about the runs of some versions of the Danish Eulerian Model on parallel computers with distributed memory can be found in [5].

5 Plans for Future Work

In the present MPI version the domain decomposition is performed in one direction only (the space domain is divided by planes orthogonal to the Oy axis). This limits the number of processors which can be used. If N_y is the number of grid-points along the Oy axis and p is the number of processors to be used, then N_y/p should not be too small. In addition, N_y/p should be an integer in order to achieve good load-balance. An implementation of an improved domain decomposition, based on splitting of the spatial domain in both horizontal directions, is under development. This version will be able to use more processors for the same grid size and, thus, to handle efficiently bigger problems.

Another important task is development of a refined grid 3-D version of the model, in which the spatial domain is discretized on a $(480 \times 480 \times 10)$ grid.

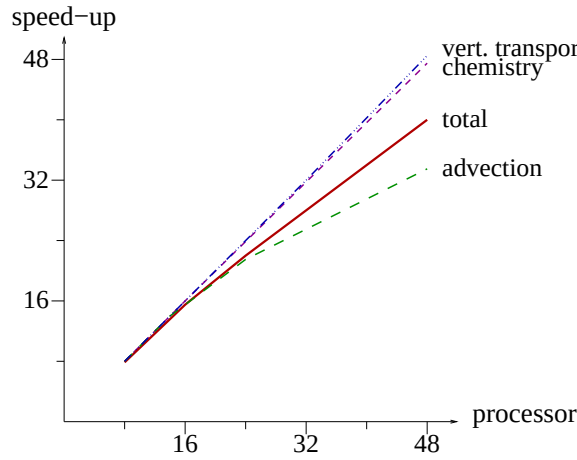

Fig. 1. Scalability of the main computational stages of the 3-D MPI version on the T3E. As there are no experiments on less than 8 processors (due to insufficient memory), the speed-ups are calculated under the assumption that it is 8 on 8 processors.

This leads to huge computational tasks, the treatment of which will be a a big challenge to the power of the existing now supercomputers. The total memory requirements of the refined grid version will be about 25 times bigger than these of the current version, so the number of processors by which this problem can be solved must necessarily be large. Preliminary calculations indicate that about 200 processors of the size of the T3E processors currently in use will do the job. Therefore, the improved domain decomposition version, mentioned in the previous paragraph, will be very useful in the development of a refined grid 3-D version for distributed memory computers.

Acknowledgments

This research was supported by EC via Access to Research Infrastructure action of the Improving Human Potential Programme (contract No. HPRI-1999-CT-00026). The development of the code started while the first author was on a TRACS research visit in EPCC, and was completed during the period of extended access to the EPCC supercomputing facilities. It was supported also by grant I-901/99 from the Bulgarian NSF and by Center of Excellence BIS-21 grant ICA1-2000-70016.

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, 365–377, 1997.

2. C. Ambelas Skjøth, A. Bastrup-Birk, J. Brandt, and Z. Zlatev. Studying variations of pollution levels in a given region of Europe during a long time-period, *Systems Analysis Modeling Simulation*, 37, 297–311, 2000.
3. A. Bastrup-Birk, J. Brandt, I. Uria, and Z. Zlatev. Studying cumulative ozone exposures in Europe during a 7-year period, *J. Geophysical Research*, 102, 23917–23935, 1997.
4. A. Bastrup-Birk, J. Brandt and Z. Zlatev. Using partitioned ODE solvers in large air pollution models, *Systems Analysis Modeling Simulation*, 32, 3–17, 1998.
5. K. Georgiev and Z. Zlatev. Parallel sparse matrix algorithms for air pollution models, *Parallel and Distributed Computing Practices*, (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. R. M. Harrison, Z. Zlatev, and C. J. Ottley. A comparison of the predictions of an Eulerian atmospheric transport chemistry model with experimental measurements over the North Sea, *Atmospheric Environment*, 28, 497–516, 1994.
8. 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, 971–994, 1978.
9. Ø. Hov, Z. Zlatev, R. Berkowicz, A. Eliassen and L. P. Prahm. Comparison of numerical techniques for use in air pollution models with non-linear chemical reactions, *Atmospheric Environment*, 23, 967–983, 1988.
10. G. I. Marchuk. *Mathematical modeling for the problem of the environment*, *Studies in Mathematics and Applications*, North-Holland, Amsterdam, 16, 1985.
11. 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, 1–42, 1984.
12. Tz. Ostromsky and Z. Zlatev. Application of sparse matrix techniques in the chemical part of a large air pollution model, *NNFM*, 62, 189–197, 1998.
13. WEB-site for OPEN MP tools, http://www.openmp.org
14. WEB-site of the Danish Eulerian Model, http://www.dmu.dk/AtmosphericEnvironment/DEM
15. Z. Zlatev. Application of predictor-corrector schemes with several correctors in solving air pollution problems, *BIT*, 24, 700–715, 1984.
16. Z. Zlatev. *Computer Treatment of Large Air Pollution Models*, Kluwer, 1995.
17. Z. Zlatev, J. Christensen, and A. Eliassen. Studying high ozone concentrations by using the Danish Eulerian model, *Atmos. Environment*, 27A, 845–865, 1993.
18. Z. Zlatev, J. Christensen, and Ø. Hov. An Eulerian air pollution model for Europe with non-linear chemistry. *Journal of Atmospheric Chemistry*, 15, 1–37, 1992.
19. Z. Zlatev, I. Dimov, and K. Georgiev. Studying long-range transport of air pollutants, *Computational Science and Engineering*, 1, 45–52, 1994.
20. Z. Zlatev, I. Dimov, and K. Georgiev. Three-dimensional version of the Danish Eulerian Model, *Zeitschrift für Angewandte Mathematik und Mechanik*, 76, 473–476, 1996.
21. Z. Zlatev, I. Dimov, Tz. Ostromsky, G. Geernaert, I. Tzvetanov, and A. Bastrup-Birk. Calculating losses of crops in Denmark caused by high ozone levels, *Env. Modeling and Assessment*, (to appear).
22. Z. Zlatev, J. Fenger, and L. Mortensen. Relationships between emission sources and excess ozone concentrations, *Computers and Math. with Appl.*, 32, 101–123, 1996.
23. Z. Zlatev, G. Geernaert, and H. Skov. A Study of ozone critical levels in Denmark, *EUROSAP Newsletter*, 36, 1–9, 1999.