Parallel Computations with Large Atmospheric Models

Z. Zlatev¹, I. Dimov², K. Georgiev³, J. Wasniewski⁴

¹ National Environmental Research Institute, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark. email: luzz@sun2.dmu.dk.
² Institute of Informatics and Computer Technology, Bulgarian Academy of Sciences, Akad. G. Bonchev str., Bl. 25-A, 1113 Sofia, Bulgaria. email: dimov@iscbg.acad.bg
³ Institute of Informatics and Computer Technology, Bulgarian Academy of Sciences, Akad. G. Bonchev str., Bl. 25-A, 1113 Sofia, Bulgaria. email: georgiev@iscbg.acad.bg.
⁴ Danish Computing Centre for Research and Education (UNiC), DTU, Bldg. 304, DK-2800 Lyngby, Denmark. email: jerzy.wasniewski@unihp1.uni-c.dk

Abstract. Large atmospheric models appear often in important applications. Long-range transport of air pollutants is one of the phenomena that can be studied by using such models. These models lead, after some kind of discretization, to very huge systems of ordinary differential equations (up to order of $10^{**6}$). The use of modern parallel and/or vector machines is a necessary condition in the efforts to handle successfully big air pollution models. However, this is often far from sufficient. One should also optimize the code in order to be able to exploit fully the great potential power of the modern high-speed computers. The air pollution code developed at the National Environmental Research Institute has been optimized for several different types of high-speed computers (vector machines, parallel computers with shared memory and parallel computers with distributed memory; including in the last group the massively parallel computers).

Results obtained on two computers will be discussed in this paper. The first computer is a vector processor, a CRAY Y-MP C90A computer. The second one is a parallel computer with shared memory, a POWER CHALENGE product of Silicon Graphics.

Key words. Air pollution models, partial differential equations, ordinary differential equations, numerical algorithms, parallel and vector computers, speed-up, efficiency.

1 Need for large air pollution models

Air pollutants emitted by different sources can be transported, by the wind, on long distances. Several physical processes (diffusion, deposition and chemical transformations) take place during the transport. Regions that are very far from the large emission sources may also be polluted. It is well-known that the atmosphere must be kept clean (or, at least, should not be polluted too much). It is also well-known that if the concentrations of some species exceed certain acceptable (or critical) levels, then they may become dangerous for plants, animals and humans.
Mathematical models are needed in the efforts to predict the optimal way of keeping the air pollution under acceptable levels. It should be emphasized here that the mathematical models are the only tool by the use of which one can predict the results of many different actions and, moreover, one can attempt to choose the best solution (or, at least, a solution which is close to the best one) in the efforts to reduce the air pollution in an optimal way. The mathematical models are necessarily large (the transport of air pollutants is carried out over long distances and, thus, the space domains are very large).

2 The Danish Eulerian Model

The work on the development of the Danish Eulerian Model has been initiated in 1980. First a simple transport scheme was developed (one pollutant only and without chemical reactions). The next step was the development of a simple model with two pollutants (and linearized chemical reactions). An experimental model containing ten pollutants and non-linear chemical reactions (including here photochemical reactions) was the third step. The first operational version of the Danish Eulerian Model was based on a chemical scheme with 35 pollutants. Some experiments with chemical schemes containing 56 and 168 pollutants are carried out at present. Different versions of the Danish Eulerian Model are discussed in [3] and [6].

The reliability of the operational model with 35 pollutants has been tested by comparing model results both with measurements taken over land ([4], [5]) and with measurements taken over sea ([2]). Test-problems, where the analytical solution is known, have been used to check the accuracy of the numerical algorithms ([3]).

3 Mathematical description of the model

The Danish Eulerian model is described by a system of PDE’s (the number of equations in this system being equal to the number of pollutants involved in the model):

$$\frac{\partial c_s}{\partial t} = -\frac{\partial (uc_s)}{\partial x} - \frac{\partial (vc_s)}{\partial y} - \frac{\partial (wc_s)}{\partial z}$$

$$+ \frac{\partial}{\partial x} \left( K_x \frac{\partial c_s}{\partial x} \right) + \frac{\partial}{\partial y} \left( K_y \frac{\partial c_s}{\partial y} \right) + \frac{\partial}{\partial z} \left( K_z \frac{\partial c_s}{\partial z} \right)$$

$$-(\kappa_{1s} + \kappa_{2s})c_s$$

$$+ E_s + Q_s(c_1, c_2, \ldots, c_q)$$

$$s = 1, 2, \ldots, q. \quad (1)$$
The different quantities that are involved in the mathematical model have the following meaning:

- the concentrations are denoted by $c_s$;
- $u, v$ and $w$ are wind velocities;
- $K_x, K_y$ and $K_z$ are diffusion coefficients;
- the emission sources in the space domain are described by the functions $E_i$;
- $\kappa_{1s}$ and $\kappa_{2s}$ are deposition coefficients;
- the chemical reactions involved in the model are described by the non-linear functions $Q_s(c_1, c_2, \ldots, c_q)$.

The non-linear functions $Q_s$ are of the form:

$$Q_s(c_1, c_2, \ldots, c_q) = - \sum_{i=1}^{q} \alpha_{si} c_i + \sum_{i=1}^{q} \sum_{j=1}^{q} \beta_{sij} c_i c_j, \quad s = 1, 2, \ldots, q. \quad (2)$$

This is a special kind of non-linearity, but it is not clear how to exploit this fact during the numerical treatment of the model.

It is clear from the above description of the quantities involved in the mathematical model that all five physical processes (advection, diffusion, emission, deposition and chemical reactions) can be studied by using the above system of PDE’s.

The mathematical model described by (1) must be considered together with appropriate initial and boundary conditions ([3]).

The Danish Eulerian Model has mainly been run as a two-dimensional model until now, which can formally be obtained from (1) by removing the derivatives with regard to $z$. The computational difficulties connected with the development of a three-dimensional version will be discussed in the following sections. Some numerical results will be presented.

4 Numerical algorithms used in the 3-D version of the Danish Eulerian Model

The mathematical model described by the system of PDE’s (1) is split to five sub-models according to the physical processes involved in the model. The sub-models are: horizontal advection sub-model, horizontal diffusion sub-model, deposition sub-model, chemistry sub-model (the emission being included here) and vertical exchange sub-model. The splitting procedure leads to a series of parallel tasks. In the first four sub-models, each horizontal plane can be treated as a parallel task. In the fifth sub-model, each vertical line can be treated as a
parallel task. This shows that the splitting procedure leads to a lot of natural parallelism. More details about the splitting procedure and the possibilities to run the model in parallel are given in [3] and [6].

Various numerical algorithms have been tried in the treatment of the different sub-models. The numerical algorithms currently used are listed below; the algorithms used in the numerical treatment of the model are discussed in [3].

- A pseudospectral discretization algorithm is used in the horizontal advection sub-model. Experiments with other methods (finite elements, corrected flux transport algorithms and semi-Lagrangian methods) are presently carried out.

- A semi-analytical approach based on expansions of the unknown functions in Fourier series is used in the horizontal diffusion sub-model. Some experiments with finite elements are also carried out.

- The splitting procedure leads to a deposition sub-model consisting of independent linear ODE's; these are solved exactly.

- The QSSA (the quasi-steady-state-approximation) is used when the chemical sub-model is handled numerically. Experiments with several classical time-integration algorithms are also carried out.

- Linear finite elements are used in the treatment of the vertical exchange sub-model.

5 Computational difficulties connected with 3-D version of the Danish Eulerian Model

The development of three-dimensional air pollution models leads to huge computational tasks. The computational tasks are very large even if the model is considered as a two-dimensional model. Assume that some splitting procedure has been applied and that the spatial derivatives are discretized by using an appropriate numerical algorithm (see the previous section or [3], [6]). Then the system of PDE's (1) is transformed into five systems of ODE's corresponding to the five sub-models discussed in the previous section. The five ODE systems have to be treated successively at every time-step. The number of equations in each of these ODE systems is equal to the product of the number of grid-points and the number of pollutants. In the case where the model is considered as two-dimensional, the numbers of equations in the ODE systems for different space discretizations and for different chemical schemes are given in Table 1. The sizes of the grid-squares for the three grids used in Table 1 are given in Table 2. The total size of the square space domain used at present in the Danish Eulerian Model is (4800 km x 4800 km); the space domain contains the whole of Europe together with parts of Asia, Africa and the Atlantic Ocean.
<table>
<thead>
<tr>
<th>Number of pollutants</th>
<th>(32 x 32)</th>
<th>(96 x 96)</th>
<th>(192 x 192)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1024</td>
<td>9216</td>
<td>36864</td>
</tr>
<tr>
<td>2</td>
<td>2048</td>
<td>18432</td>
<td>73728</td>
</tr>
<tr>
<td>12</td>
<td>12288</td>
<td>110592</td>
<td>442358</td>
</tr>
<tr>
<td>35</td>
<td>35840</td>
<td>322560</td>
<td>1290240</td>
</tr>
<tr>
<td>56</td>
<td>57344</td>
<td>516096</td>
<td>2064384</td>
</tr>
<tr>
<td>168</td>
<td>172032</td>
<td>1548288</td>
<td>6193152</td>
</tr>
</tbody>
</table>

Table 1

Numbers of equations per ODE system in the two-dimensional versions of the Danish Eulerian Model.

\[
\begin{align*}
1 & (32 \times 32) (150 \text{ km} \times 150 \text{ km}) \\
2 & (96 \times 96) (50 \text{ km} \times 50 \text{ km}) \\
3 & (192 \times 192) (25 \text{ km} \times 25 \text{ km}) \\
\end{align*}
\]

Table 2

The sizes of the grid-squares for the three grids used in Table 1.

It should be mentioned here that (i) if a three-dimensional model with ten vertical layers is used, then all figures that are given in Table 1 must be multiplied by ten, (ii) 3456 time-steps have been used for a typical run (with meteorological data for one month + five days to start up the model; [3]) when the (96 x 96) grid is used and (iii) the chemical reactions lead to a very stiff (and also very badly scaled) ODE systems; this fact causes extra computational difficulties.

6 Need for high-speed computers

The size of the ODE systems shown in Table 1 (these systems have to be treated during several thousand time-steps) explains why the use of high-speed computers is absolutely necessary when large air pollution problems are to be handled. It is even more important to perform carefully the programming work in order to try to exploit better the great potential power of the modern high-speed computers. As a rule, this is not an easy task (especially on the newest parallel computers). Finally, it should be emphasized that even when the fastest computers are available and even when the programming work is very carefully done, it is still not possible, at present, to solve some of the biggest problems listed in Table 1 (especially in the case where the model is treated as a three-dimensional model). Therefore faster and bigger computers are needed in order to be able to treat successfully big air pollution models.

Different versions of the Danish Eulerian Model have been used in runs on several high-speed computers ([1], [3], [6]). Some results obtained by running the three-dimensional version on a CRAY Y-MP C90A will be presented in the next
section. Preliminary results obtained by running three important modules of the model (the advection sub-model, the chemistry sub-model and a combination of the advection and the chemistry sub-models) on a POWER CHALLENGE Silicon Graphics computer will be discussed in Section 8.

7 Runs on a CRAY Y-MP C90A computer

The three-dimensional version of the Danish Eulerian Model has until now been run only on a CRAY Y-MP C90A computer. The performance achieved in the different parts of the model (corresponding to the five physical processes) is shown in Table 3. The total computing time and the overall speed of computations, measured in MFLOPS (millions of floating point operations, additions and multiplications, per second), are given in Table 4.

<table>
<thead>
<tr>
<th>Physical process</th>
<th>Computing time</th>
<th>In percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advection</td>
<td>638</td>
<td>22.2</td>
</tr>
<tr>
<td>Diffusion</td>
<td>295</td>
<td>10.2</td>
</tr>
<tr>
<td>Deposition</td>
<td>138</td>
<td>4.8</td>
</tr>
<tr>
<td>Chemistry</td>
<td>1652</td>
<td>57.4</td>
</tr>
<tr>
<td>Vertical exchange</td>
<td>50</td>
<td>1.7</td>
</tr>
<tr>
<td>Overhead</td>
<td>107</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Table 3
Computing times (in seconds) for the different parts of the model obtained on a CRAY Y-MP C90A computer.

<table>
<thead>
<tr>
<th>Quantity measured</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing time</td>
<td>2880</td>
</tr>
<tr>
<td>Speed in MFLOPS</td>
<td>304</td>
</tr>
</tbody>
</table>

Table 4
The total computing time (in seconds) and the overall computational speed obtained in a run with one-month meteorological data (plus five days to start up the model) on a CRAY Y-MP C90A computer.

The performance of several subroutines on CRAY Y-MP C90A has recently been improved. The improvements made lead to an increase of the performance of the whole code by about 30%.
8 Runs on a POWER CHALLENGE product of Silicon Graphics

As mentioned above, three modules of the Danish Eulerian model:

1. the chemical sub-model,
2. the transport (advection) sub-model
3. the combination of the advection and the chemical sub-models,

have been run on a POWER CHALLENGE product of Silicon Graphics. It should be emphasized that the results produced by these modules are not directly comparable with the results presented in the previous section, where the whole three-dimensional model has been run. The modules are used to check the accuracy of the numerical methods used in the advection part and in the chemistry part as well as to check the reliability of the coupling procedure between the advection and the chemical parts of the Danish Eulerian Model (see more details in [1], [3] and [6]).

While it is difficult to find a good numerical algorithm for the chemical sub-model, it is not very difficult to achieve parallel computations with any numerical algorithm. Indeed, the computations within the chemical sub-model can schematically be described (at every time-step and for any numerical algorithm applied in the chemical sub-model) by the following loop.

\[
\text{DO } i = 1, \text{NUMBER OF GRID-POINTS} \\
\quad \text{Perform all chemical reactions at the } i^{th} \text{ grid-point.} \\
\text{END DO}
\]

It is clear that the number of parallel tasks is equal to the number of grid-points and that the loading balance is perfect. Therefore one should expect to obtain good speed-ups. The results shown in Table 5 show that the efficiency with regard to the speed-ups are very close to the optimal values that could be achieved. Let

- \( N_p \) be the number of the processors used,
- \( t_1 \) be the computing time obtained on one processor only,
- \( t_p \) be the computing time obtained when \( p \) processors are used,

then the speed-up and the efficiency (in percent) are defined by the following two formulae

\[
\text{Speed-up} = \frac{t_1}{t_p},
\]

\[
\text{Efficiency} = \left( \frac{t_1}{t_p} \right) \times 100%.
\]

(3)
\[ Efficiency - 1 = \frac{100 t_1}{N_p t_p}. \] (4)

It is very important to achieved parallel computations, however this is very often not sufficient (in other words even when the computations are carried out in parallel, the performance may be rather low). Therefore another measure is needed in order to achieve a full evaluation of the efficiency of the code. This second measure is the efficiency with regard to the peak performance of the architecture on which the code is run. This quantity is defined by

\[ Efficiency - 2 = \frac{100 \ast MFLOPS}{260 \ast N_p}. \] (5)

where MFLOPS is the computational speed achieved by the code (millions of floating point operations per second), while 260 MFLOPS is the peak performance of one POWER CHALLENGE processor.

It is seen from Table 6 that the efficiency with regard to the peak performance is rather poor. The results obtained on one processor of CRAY Y-VP C90A are:

\[ t_1 = 284, \quad MFLOPS = 422.5, \quad Efficiency - 2 = 46.5\%. \] (6)

The POWER CHALLENGE results can perhaps be improved by better tuning of the code and by trying to exploit the cash more efficiently.

The transport (advection) part can also easily be run in parallel. The computations within the transport sub-model can schematically be described (again at every time-step and for any numerical algorithm applied in this module) by the following loop.

\begin{verbatim}
DO I=1,NUMBER OF COMPOUNDS
   Perform the transport for the I\textsuperscript{th} compound.
END DO
\end{verbatim}

The number of parallel tasks is now equal to the number of compounds (in the present case the number of compounds is 56). This may cause problems if the number of processors is large. In the runs presented in Table 7 and Table 8, where the number of processors is up to 8, there are no problems with the above loop. The tasks are not very well balanced for the algorithm used (because there is an iterative process, and for the different compounds the number of iterations may be different). Therefore one should expect lower speed-ups (comparing with the chemical module). This effect can be seen by comparing the results given in Table 7 with those in Table 5. However, however, the speed-ups given in Table 7 are nevertheless very high (the efficiency with regard to speed-ups is higher than 80%, which is an excellent result).
The efficiency obtained with regard to the peak performance of the configurations of the POWER CHALLENGE used in the runs is greater for the transport module (comparing with the chemical sub-model); compare the results given in Table 8 with those in Table 6. Nevertheless, the results show that also here improvements are highly desirable. This becomes very clear by comparing the results given in Table 7 - Table 8 with the results obtained for this module on CRAY Y-MP C90A:

\[ t_1 = 205, \quad MFLOPS = 440.0, \quad Efficiency - 2 = 48.8\%. \] (7)

The third module, which is very important in the checks of the accuracy of the coupling procedure (transport with chemistry), is carried out by performing, at every time-step, successively the two loops given above. Therefore one should expect the efficiency results for the third module to be between the efficiency results of the first two modules (the chemical module and the transport module). The results given in Table 9 - Table 10 confirm such a conclusion; compare these results with the corresponding results in Table 5 - Table 8.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Time</th>
<th>Speed-up</th>
<th>Efficiency-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4576.1</td>
<td>-</td>
<td>100.0%</td>
</tr>
<tr>
<td>2</td>
<td>2390.9</td>
<td>1.91</td>
<td>95.6%</td>
</tr>
<tr>
<td>4</td>
<td>1179.8</td>
<td>3.88</td>
<td>96.9%</td>
</tr>
<tr>
<td>8</td>
<td>602.7</td>
<td>7.59</td>
<td>94.9%</td>
</tr>
</tbody>
</table>

Table 5

Computing times (in seconds), speed-ups and efficiency-1 results obtained during runs of the chemical module on the POWER CHALLENGE product of Silicon Graphics.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Time</th>
<th>MFLOPS</th>
<th>Efficiency-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4576.1</td>
<td>26.2</td>
<td>10.1%</td>
</tr>
<tr>
<td>2</td>
<td>2390.9</td>
<td>50.2</td>
<td>9.7%</td>
</tr>
<tr>
<td>4</td>
<td>1179.8</td>
<td>101.7</td>
<td>9.8%</td>
</tr>
<tr>
<td>8</td>
<td>602.7</td>
<td>199.1</td>
<td>9.6%</td>
</tr>
</tbody>
</table>

Table 6

Computing times (in seconds), MFLOPS and efficiency-2 results obtained during runs of the chemical module on the POWER CHALLENGE product of Silicon Graphics.
Table 7
Computing times (in seconds), speed-ups and efficiency-1 results obtained during runs of the transport module on the POWER CHALLENGE product of Silicon Graphics.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Time (s)</th>
<th>Speed-up</th>
<th>Efficiency-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1963.5</td>
<td>1</td>
<td>100.0%</td>
</tr>
<tr>
<td>2</td>
<td>1041.3</td>
<td>1.09</td>
<td>94.6%</td>
</tr>
<tr>
<td>4</td>
<td>594.6</td>
<td>3.00</td>
<td>82.5%</td>
</tr>
<tr>
<td>8</td>
<td>305.8</td>
<td>6.62</td>
<td>80.3%</td>
</tr>
</tbody>
</table>

Table 8
Computing times (in seconds), MFLOPS and efficiency-2 results obtained during runs of the transport module on the POWER CHALLENGE product of Silicon Graphics.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Time (s)</th>
<th>MFLOPS</th>
<th>Efficiency-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1963.5</td>
<td>46.0</td>
<td>17.7%</td>
</tr>
<tr>
<td>2</td>
<td>1041.3</td>
<td>86.6</td>
<td>16.7%</td>
</tr>
<tr>
<td>4</td>
<td>594.6</td>
<td>151.7</td>
<td>14.6%</td>
</tr>
<tr>
<td>8</td>
<td>305.8</td>
<td>295.2</td>
<td>14.2%</td>
</tr>
</tbody>
</table>

Table 9
Computing times (in seconds), speed-ups and efficiency-1 results obtained during runs of the third module (chemistry + transport) on the POWER CHALLENGE product of Silicon Graphics.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Time (s)</th>
<th>Speed-up</th>
<th>Efficiency-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5990.3</td>
<td>1</td>
<td>100.0%</td>
</tr>
<tr>
<td>2</td>
<td>3040.0</td>
<td>1.97</td>
<td>98.5%</td>
</tr>
<tr>
<td>4</td>
<td>1632.2</td>
<td>3.67</td>
<td>91.8%</td>
</tr>
<tr>
<td>8</td>
<td>897.4</td>
<td>6.68</td>
<td>83.5%</td>
</tr>
</tbody>
</table>

Table 10
Computing times (in seconds), MFLOPS and efficiency-2 results obtained during runs of the third module (chemistry + transport) on the POWER CHALLENGE product of Silicon Graphics.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Time (s)</th>
<th>MFLOPS</th>
<th>Efficiency-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5990.3</td>
<td>35.0</td>
<td>13.4%</td>
</tr>
<tr>
<td>2</td>
<td>3040.0</td>
<td>69.0</td>
<td>13.3%</td>
</tr>
<tr>
<td>4</td>
<td>1632.2</td>
<td>128.5</td>
<td>12.4%</td>
</tr>
<tr>
<td>8</td>
<td>897.4</td>
<td>233.7</td>
<td>11.2%</td>
</tr>
</tbody>
</table>
9 Concluding remarks

It is seen that the chemistry is the most time-consuming part of the model. Therefore improvements of the chemical subroutines are most desirable (however, also the performance of the mathematical modules describing the other physical processes are to be improved). It should be emphasised that the three-dimensional model can be treated numerically at present only on a relatively coarse spatial grid, the (32 x 32) grid, and only for the relatively simple chemical scheme with 35 pollutants. It is necessary to improve both the numerical algorithms and the computer programs for the different modules in order to be able to run also some of the more complicated cases (refined grids and chemical schemes with more pollutants). It is crucial to select (or develop) algorithms that perform well on the new high-speed computers. Some work in these directions is carried out at present.

Acknowledgements

This research was partially supported by NMR (Nordic Council of Ministers), EMEP (European Monitoring and Evaluating Programme), NATO (North Atlantic Treaty Organization), SMP (Danish Strategic Environmental Programme) and the Danish Natural Sciences Research Council.

References