Parallel Computations in Large Scale Air Pollution Models

Zahari Zlatev
Department of Environmental Science
Aarhus University
Frederiksborgvej 399, P. O. Box 358
DK-4000 Roskilde, Denmark,
E-mail address: zz@dmu.dk, Zlatevi@adslhome.dk
Web-site: http://www.dmu/atmosphericenvironment/staff/zlatev.htm
CONTENTS

- Are parallel computations still really needed?
- UNI-DEM – Mathematical description
- Achieving efficiency: Major requirements
- Numerical Treatment
- Parallel programming: Basic principles
- Parallel Computations in UNI-DEM
- Climatic Studies
- Some Conclusions
Are parallel computations needed?

Why is this question relevant?

- Many models, which 30 years ago could only be treated on parallel computers, can now be treated on PCs.
- Some problems cannot be run in parallel
- At least some parallel computers are expensive (both to buy them, but also to maintain them).

This means that the answer is: “at least not always”.

On the other side: if the problem is really big, then it can be handled only if suitable powerful parallel computers are available and efficiently used.
Need for fast computations versus available high-speed computers

“Although the fastest computers can execute millions of operations in one second, they are always too slow. This may seem a paradox, but the heart of the matter is: the bigger and better computers become, the larger are the problems scientists and engineers want to solve.”


The author of this article was right about 30 years ago, he is still right now and he will continue to be right in the future.
Different stages in the development of the Unified Danish Eulerian Model (UNI-DEM)

**First stage:** two chemical species (SO2 and SO4), linear chemistry, two-dimensional version, 35520 time-steps, 2048 equations, 150 km x 150 km grid-cells.

**Second stage:** 35 chemical species, non-linear chemistry, two-dimensional version, 35520 time-steps, 322560 equations, 50 km x 50 km grid-sells, three-dimensional version, 10 layers, 35520 time-steps, 3225600 equations, 50 km x 50 km horizontal grid-sells.

**Third stage:** 35, 56 and 168 chemical species, non-linear chemistry, two-dimensional version, 213120 time-steps, 8064000 equations, 10 km x 10 km grid-sells, three-dimensional version, 10 layers, 213120 time-steps, 80640000 equations, 10 km x 10 km horizontal grid-sells.

The number of equations is increased by a factor of 393750.
Nitrogen dioxide pollution in Europe

**Red colours:**
the most polluted areas in Europe

**Blue colours:**
low polluted regions

The year is **1997**, but the same trend was observed for any other year.
Nitrogen emissions in Denmark

Red colours: high emissions

The emissions are highest around the biggest cities of Denmark: Copenhagen, Århus and Odense

The emissions are also high around the Swedish city of Malmö
NO2 pollution in Denmark (coarse grid)

Red colours: indicate high pollution levels

Blue colours: low polluted regions

In this case, the coarse grid does not reproduce in a very realistic way the actual situation.
NO2 pollution in Denmark (fine grid)

Red colours: indicate high pollution levels

The fine grid produces qualitatively correct results

Why?
Because the high pollution levels are located where the emissions are also high.
Distribution of the mean values of the annual SO2 concentrations

Results for 1997 in ppb
Running DEM on a 96x96 grid
9216 cells in the European area
The resolution is 50 km by 50 km
Maximal value in Europe: 7
Minimal value in Europe: 0
Distribution of the mean values of the annual SO2 concentrations

Results for 1997 in ppb
Running DEM on a 480x480 grid
230400 cells in the European area
The resolution is 10 km x 10 km
Maximal value in Bulgaria: 14
Minimal value in Bulgaria: 1

Legend:
- Above: 4
- 3 - 4
- 2 - 3
- 1 - 2
- Below: 1
Six major tasks arising in the development of a large-scale air pollution model

- **Describe** in an adequate way all important physical and chemical processes
- **Ensure that the model runs efficiently on modern high-speed computers**
- **Use** high quality input data
- **Verify** the model results by comparing them with reliable measurements taken in different parts of the space domain
- Carry out some **sensitivity experiments** to check the response of the model to changes of different key parameters
- **Visualize and animate** the output results to make them easily understandable even for non-specialists
Need for large-scale mathematical models for studying air pollution levels in Europe

Two basic requirements in the modern society:

- Processes that can potentially be dangerous (for example for human health) should be carefully controlled.
- It is necessary to avoid situations in which some prescribed critical levels are exceeded.

How to satisfy these requirements?

- Robust and reliable mathematical models have to be developed and used in different simulations in order to study under which conditions the appearance of very high pollution levels can be avoided.
Generic Formulation of an Air Pollution Model

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

\[s = 1, 2, \ldots, q\]

hor. transport
hor. diffusion
deposition
chem. + emis.
vert. transport
How to achieve computational efficiency

- Use some **splitting** procedure
- Apply **optimal numerical methods** in the treatment of each of the obtained sub-models
- **Parallelize** the computations on the available computer
- Utilize the **cache** memories

Some very big studies related to the impact of the climate changes on the high pollution levels in Europe can be handled only if the above four tasks are successfully solved.
Applying splitting techniques

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

hor. transport

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

deposition

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

vert. exchange

Questions: (a) How to couple the sub-models?
(b) What is the accuracy?
(c) How to deal with the boundary conditions?
Numerical treatment of the horizontal transport

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

hor. transport

\[
P \frac{dg^{(1)}}{dt} = Hg^{(1)}, \quad g^{(1)} \in \mathcal{R}^N
\]

Questions:
1. How to obtain the system of ODEs?
2. How to solve the system of ODEs?

Explicit methods with a stability control

New alternative: ERK + RE
How to obtain the ODE system

- Finite differences
- Finite elements
- Semi-Lagrangian method
- Pseudo-spectral discretization
- Wavelets

Zlatev 1995 (Kluwer / Springer)
How to solve the ODE system

- **Basic principles:** (a) keep the stepsize constant and (b) ensure stability control

- **Basic tools:** predictor-corrector schemes with several different correctors (Zlatev, 1984, BIT)

- **Stability check:**
  \[ \Delta t \leq \frac{h_{\text{imag}}}{U + V} (p \Delta x) \]
Numerical treatment of the chemical reactions

\[ \frac{\partial c_s^{(2)}}{\partial t} = -(k_{1s} + k_{2s})c_s^{(2)} \]

\[ + E_s + Q_s(c_1^{(2)}, c_2^{(2)}, \ldots, c_q^{(2)}) \]

Deposition

Chem. + emis.

\[ \frac{dg^{(2)}}{dt} = f(g^{(2)}, t), \quad g^{(2)} \in \mathbb{R}^N \]

Properties:
1. There are no spatial derivatives in this sub-model
2. The involved matrices are very badly scaled
3. Non-linear and stiff system of ODEs
4. Implicit numerical methods
5. It is possible to exploit the sparsity of the Jacobean matrix of the right-hand function \( f \).
**Numerical methods for the chemical sub-model**

<table>
<thead>
<tr>
<th>Method</th>
<th>Formula</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. θ-method (θ-M)</td>
<td></td>
<td>Hundsdorfer and Verwer, 1991</td>
</tr>
<tr>
<td>7. Second-order Rosenbrock method (ROS)</td>
<td></td>
<td>Verwer et al., 1999</td>
</tr>
</tbody>
</table>

All seven methods were extensively tested in:

*Zlatev and Dimov, 2006*
Stability results

<table>
<thead>
<tr>
<th>Numerical Method</th>
<th>Direct Implementation</th>
<th>Richardson Extrapolation</th>
<th>Sequential + Richardson</th>
<th>M-S + Richardson</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDF</td>
<td>LS</td>
<td>LS</td>
<td>LS</td>
<td>?</td>
</tr>
<tr>
<td>TR</td>
<td>AS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Θ-M</td>
<td>SAS, 0.5 &lt; θ &lt; 1</td>
<td>SAS, 2/3 &lt; θ &lt; 1</td>
<td>SAS, 0.64 &lt; θ &lt; 1</td>
<td>?</td>
</tr>
<tr>
<td>IMR</td>
<td>AS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>RK2</td>
<td>LS</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>RK5</td>
<td>LS</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>ROS</td>
<td>LS</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

**LS**: L-stable  
**SAS**: Strongly A-stable  
**AS**: A-stable  
**NS**: not A-stable

?: Not investigated yet

Each cell in the table (excluding the cells with question marks) gives the assertion of a theorem.
Numerical treatment of the vertical exchange

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

\[
P \frac{dg^{(3)}}{dt} = Hg^{(3)}, \quad g^{(3)} \in \mathcal{R}^N
\]

Properties:
1. The matrices \( P \) and \( H \) are banded and depend on the spatial discretization
2. The system of ODEs is linear and stiff
3. Implicit numerical methods are to be used
4. This sub-model is cheaper than the other two
5. The uncertainties with the data for \( w \) and the modelling of the vertical diffusion coefficients \( K \) are enormous
Optimizing the code for runs on one processor

Why is optimizing needed?
- Computation time versus load-store time
- Cache memories: difficulties
- Main principle: organize the computations so that a small amount of data is used as long as possible in the computations

Owczarz and Zlatev (2002), Parallel Computing
Alexandrov et al. (2004), Mathematics and Computers in Simulation
Zlatev and Dimov (2006), Elsevier
Achieving parallelism

Need for standard tools

- OpenMP
- MPI

Parallel tasks when OpenMP is used
Parallel tasks when MPI is used

Owczarz and Zlatev (2002), Parallel Computing
Alexandrov et al. (2004), Mathematics and Computers in Simulation
Zlatev and Dimov (2006), Elsevier, Amsterdam
Naturally arising parallel tasks

<table>
<thead>
<tr>
<th>Sub-model</th>
<th>(480x480x10) grid</th>
<th>Size of the tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical trasport</td>
<td>8 064 000</td>
<td>very small</td>
</tr>
<tr>
<td>Horizontal transport</td>
<td>350</td>
<td>very large</td>
</tr>
<tr>
<td>Chemistry</td>
<td>2 304 000</td>
<td>very small</td>
</tr>
</tbody>
</table>

1. The exploitation of the natural (embarrassing) parallelism is **not** enough
2. Grouping together several of the small tasks in a bigger task (**N_CHUNKS**)
3. Parallelism within any of the very large tasks has somehow to be achieved (this is a much more difficult problem)

**General conclusion:** splitting is very important but cannot solve all the problems; it is necessary to do something more.
One-year runs

<table>
<thead>
<tr>
<th>Spatial resolution (horizontal)</th>
<th>2-D option (1-layer)</th>
<th>3-D option (10-layers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>96x96 grid</td>
<td>1.41</td>
<td>12.39</td>
</tr>
<tr>
<td>288x288 grid</td>
<td>19.56</td>
<td>152.72</td>
</tr>
<tr>
<td>480x480 grid</td>
<td>98.72</td>
<td>856.75</td>
</tr>
</tbody>
</table>

MPI option on 8 processors
Meteorological data for one year (1997)
Computing times in CPU-hours

More processors and more powerful processors are needed for the fine resolution 3-D options (about 36 days)
## OpenMP versus MPI

<table>
<thead>
<tr>
<th>Physical processes</th>
<th>OpenMP option</th>
<th>MPI option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advection-diffusion</td>
<td>2618</td>
<td>457</td>
</tr>
<tr>
<td>Chemistry-emission-deposition</td>
<td>1339</td>
<td>688</td>
</tr>
<tr>
<td>Total computing time</td>
<td>4011</td>
<td>1281</td>
</tr>
</tbody>
</table>

(480x480x1) grid on 8 processors

1000 time-steps

Computing times in **seconds**

**MPI performs better than OpenMP** in the solution of this problem

**Possible explanation: better utilization of caches**
Scalability

<table>
<thead>
<tr>
<th>Physical processes (and total time)</th>
<th>OpenMP option</th>
<th>MPI option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advection-diffusion</td>
<td>1001 (2.61)</td>
<td>120 (3.81)</td>
</tr>
<tr>
<td>Chemistry-emission-deposition</td>
<td>607 (2.21)</td>
<td>171 (4.02)</td>
</tr>
<tr>
<td>Total computing time</td>
<td>1771 (2.26)</td>
<td>348 (3.68)</td>
</tr>
</tbody>
</table>

Transition from 8 to 32 processors
480x480x1 grid, 1000 time-steps
Computing times in **seconds**, speed ups in brackets

Not only is the MPI option giving better times, but it also scales better than the OpenMP option.
Numerical Results: Hungary

Numbers of days in which the 8-hour averages of the ozone concentrations are greater than 60 ppb.

EU-critical value: 25 days

More than 25 days might cause damages on people with asthmatic difficulties.
Numerical Results: Budapest

RESULTS OBTAINED AT BUDAPEST
BASIC SCENARIO VERSUS CLIMATIC SCENARIO 3
NUMBERS OF BAD DAYS ARE COMPARED
THE BLUE LINE SHOWS THE EU LIMIT

CLIMATIC SCENARIO 3:  
BASIC SCENARIO:
Numerical Results: Balkan Peninsula

Increase in percent of the numbers of days in which the 8-hour averages of the ozone concentrations are greater than 60 ppb when climatic changes are taken into account.

EU-critical value: 25 days

More than 25 days might cause damages on people with asthmatic difficulties

In many parts of the region the increase is greater than 15%.
Numerical Results: Big cities

INCREASES OF BAD DAYS DUE TO CLIMATIC CHANGES
PERCENTAGES: 100.0*(CLIMATIC 3 - BASIC))/BASIC)
UNITS: PERCENTAGES
BLUE LINES: 30 PERCENT AND 0 PERCENT LEVELS

Increased numbers of bad days

YEARS

Paris: MOSCOW
Madrid: MILAN
Berlin: ROME
Vienna: LONDON
MAJOR CONCLUSIONS

- **Faster and sufficiently accurate** numerical methods are needed
- **Faster and “bigger”** computers will help us to solve some problems which cannot be treated on the available at present computers
- There are urgent requirements for **new modules** (as, for example, a module for treatment of aerosols) which will lead to even more time-consuming and storage consuming computational tasks
- **Data assimilation** would be useful in some cases (under the assumption that sufficiently many, high-quality measurements are available)
Geographical locations of the cities and the corresponding stations

Figure 2: The selected cities (left-hand-side plot) and measurement stations (right-hand-side plot).
### Table: Difference Between Climates

<table>
<thead>
<tr>
<th>Difference</th>
<th>AT02</th>
<th>DE02</th>
<th>ES01</th>
<th>FR06</th>
<th>GB36</th>
<th>IT01</th>
<th>IT04</th>
<th>RU18</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0</td>
<td>4</td>
<td>21</td>
<td>23</td>
<td>26</td>
<td>73</td>
<td>6</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>0</td>
<td>23</td>
<td>38</td>
<td>116</td>
<td>79</td>
<td>52</td>
<td>56</td>
<td>58</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
<td>39</td>
<td>34</td>
<td>33</td>
<td>26</td>
<td>52</td>
<td>46</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>24</td>
<td>5</td>
<td>20</td>
<td>11</td>
<td>31</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>21</td>
<td>3</td>
<td>11</td>
<td>10</td>
<td>15</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>15</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>&gt; 9</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>23</td>
</tr>
</tbody>
</table>
Introducing grey zones

**Introduction of a Grey Zone**

Instead of the sharp limit of 60 PPB for the 8-hours averaged ozone concentrations

Station: DE02, Units: PPB

Maximal 8-hour averaged

<table>
<thead>
<tr>
<th>Days</th>
<th>(August 2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Bad Days:</td>
<td>4</td>
</tr>
<tr>
<td>Number of Grey Days:</td>
<td>22</td>
</tr>
<tr>
<td>Number of Good Days:</td>
<td>5</td>
</tr>
</tbody>
</table>
Conclusions related to the computations

- Need for better numerical methods
- Need for faster and “bigger” computers
- The use of data assimilation leads to an increase of both the computing time and the storage needed
- Necessity to run sets of different models using huge sets of different data: the use of computer grids might be appropriate for such a combination of large tasks.