PARALLEL IMPLEMENTATION OF A LARGE-SCALE 3-D AIR POLLUTION MODEL

Tzvetan Ostromsky *, **Zahari Zlatev** †

* Central Laboratory for Parallel Processing,
Bulgarian Academy of Sciences,
Acad. G. Bonchev str., bl. 25-A, 1113 Sofia, Bulgaria
E-mail: *ceco@copern.bas.bg*
Web-cite: http://copern.bas.bg/∼ceco

† National Environmental Research Institute,
Department of Atmospheric Environment,
Frederiksborgvej 399, DK–4000 Roskilde, Denmark
E-mail: *zz@dmu.dk*
Web-cite: http://www.dmu.dk/atmosphericenvironment/staff/zlatev.htm

Outline of the talk

- Mathematical background of the air pollution modeling
- The Danish Eulerian Model
- Splitting into submodels
- Space discretization
- Numerical methods and parallelization techniques
- Input data
- Numerical results on shared memory computers
- Numerical results on distributed memory computers
- Plans for future work

Mathematical background of the air pollution modeling

A system of PDE 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.

A huge computational task, because of the:

- size of the domain (should be large to reduce the boundary errors);
- dynamics of the processes – small time-step;
- complexity of the equations – decomposition.

The Danish Eulerian Model

$$\frac{\partial c_s}{\partial t} = -\frac{\partial (u c_s)}{\partial x} - \frac{\partial (v c_s)}{\partial y} - \frac{\partial (w c_s)}{\partial z}$$
$$+ \frac{\partial}{\partial x}\left(K_x \frac{\partial c_s}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial c_s}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial c_s}{\partial z}\right)$$
$$+ E_s + Q_s(c_1, c_2, \ldots c_q) - (k_{1s} + k_{2s}) c_s, \quad s = 1, 2, \ldots q \ .$$

- 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;
- $Q_s(c_1, c_2, \ldots c_q)$ – non-linear functions describing the chemical reactions between species under consideration (Gery et al. (1989)).

Splitting into submodels

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

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

$$\frac{\partial c_s^{(3)}}{\partial t} = E_s + Q_s(c_1^{(3)}, c_2^{(3)}, \ldots c_q^{(3)})$$ chemistry & emissions

$$\frac{\partial c_s^{(4)}}{\partial t} = -(k_{1s} + k_{2s}) c_s^{(4)}$$ deposition

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

Related work: Marchuk (1982), McRae, Goodin and Seinfeld (1984).

Space discretization and numerical treatment of the five submodels

— Five large ODE systems*:

$$\frac{dg^{(i)}}{dt} = f^{(i)}(t, g^{(i)}) , \quad i = 1, 2, 3, 4, 5$$

$$g^{(i)} \in \mathcal{R}^{N_x \times N_y \times N_z \times q}$$
$$f^{(i)} \in \mathcal{R}^{N_x \times N_y \times N_z \times q}$$

- N_x, N_y, N_z – the number of grid-points along the coordinate axes (different versions);
- $q = 35$ – the number of chemical species considered in the model;
- $g^{(i)}$ – functions, approximating the concentrations of the different species throughout the spatial domain.

* In general, these systems are simpler. For example, the chemical submodel splits into $N_x \times N_y \times N_z$ ODE systems of size q.

Size of the different versions of the model

Horizontal 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

* Coarse grid 3-D version with horizontal grid-step 50 km. :

- 5 ODE systems of order 3 225 600 ($N_x = N_y = 96$, $N_z = 10$);
- The ten layers in vertical direction are non-equidistant (the layers closer to the surface are thiner);
- With a primary time-step 15 min. more than 20 000 primary time-steps are to be carried out to cover a time period of 1 month;
- Smaller time-step (150 sec.) must be used in the chemical sub-model.

Numerical methods and parallelization techniques used in the main submodels

- Chemical submodel: An improved version of the QSSA (Quazi Steady-State Algorithm) (Hesstevedt et al. - 1978)

  Native parallel tasks: The calculations in a single grid-point. These are numerous, but rather small. Chunks are used to create somewhat coarser tasks in the parallel implementation of this part.

- Advection-diffusion part: Finite elements, followed by predictor-corrector schemes with several different correctors (Zlatev - 1984)

  Native parallel tasks: The calculations for a given pollutant in a given layer. The tasks are large enough and they are good ground for shared memory parallelism.

- Vertical transport: Finite elements, followed by θ-method

  Native parallel tasks: The calculations for a given pollutant in a given grid-point of the horizontal grid. Computing the vertical exchange along each vertical grid-line is a parallel task. The number of these tasks is $N_x \times N_y$, it is large when the grid is fine.

Distributed memory parallelizm

The distributed memory parallelization strategy is based on domain decomposition of the horizontal grid. This strategy induces some boundary dependencies of the advection-diffusion subproblems, that require frequent communications. As a result, the complexity of the advection-diffusion part in the distributed memory version of the model is a little bit more complex, compared to the shared memory version.

Input data sets

Emission data: More or less constant over a long time period. Loaded at the beginning, updated with internal constants if necessary.

← Antropogenic emissions:

SO_2, NO_x, NH_3,

Antr. hydro-carbons

← Natural hydro-carbons

Geographical data: Constant. Loaded at the beginning.

← Latutudes and longitudes

← Land and sea areas

Meteorological data: Quickly changing, updated every 6 hours. New data must be read regularly after certain number of time steps.

⇐ Horizontal wind

⇐ Precipitations

⇐ Mixing hights

⇐ Temperatures

⇐ Humidities

⇐ Cloud cover

⇐ Surface temperatures

⇐ Vertical wind

⇐ Heat flux

⇐ Surface stress

Basic conditions of the numerical experiments

- For portability reasons, only standard OpenMP directives and standard MPI routines are used in the parallel codes for shared and for distributed memory machines respectively.
- The $(96 \times 96 \times 10)$ version of the grid is used in all experiments on shared and distributed memory machines.
- All experiments are for a time period of 1 month.
- Characteristics of the machines, used in the experiments:
  - SUN E-6500 UltraSPARC, 400 MHz (shared memory)
  - CRAY T3E-900 (distributed memory)
- Compiler options:
  - On the SUN : *"-fast -O4 -xarch=v8plusa -xparallel"*.
  - On the T3E: *"-O 3 "*.
- The results are obtained in multiuser mode on normal priority batch queues.

Parallel runs on shared memory computers

3-D OpenMP version of DEM on SUN E6500/400MHz				
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*

Time (in seconds) and speed-up of the main stages of the 3-D OpenMP version of DEM, $(96 \times 96 \times 10)$ grid. The results are obtained on a SUN cluster in the EPCC and on an SGI ORIGIN 2000 at UNI•C, Denmark. Chunks of size 48 are used in the chemistry part, which appears to be optimal for these machines.

Parallel runs on distributed memory computers

3-D MPI version of DEM on a CRAY T3E computer			
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+Diffus.	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

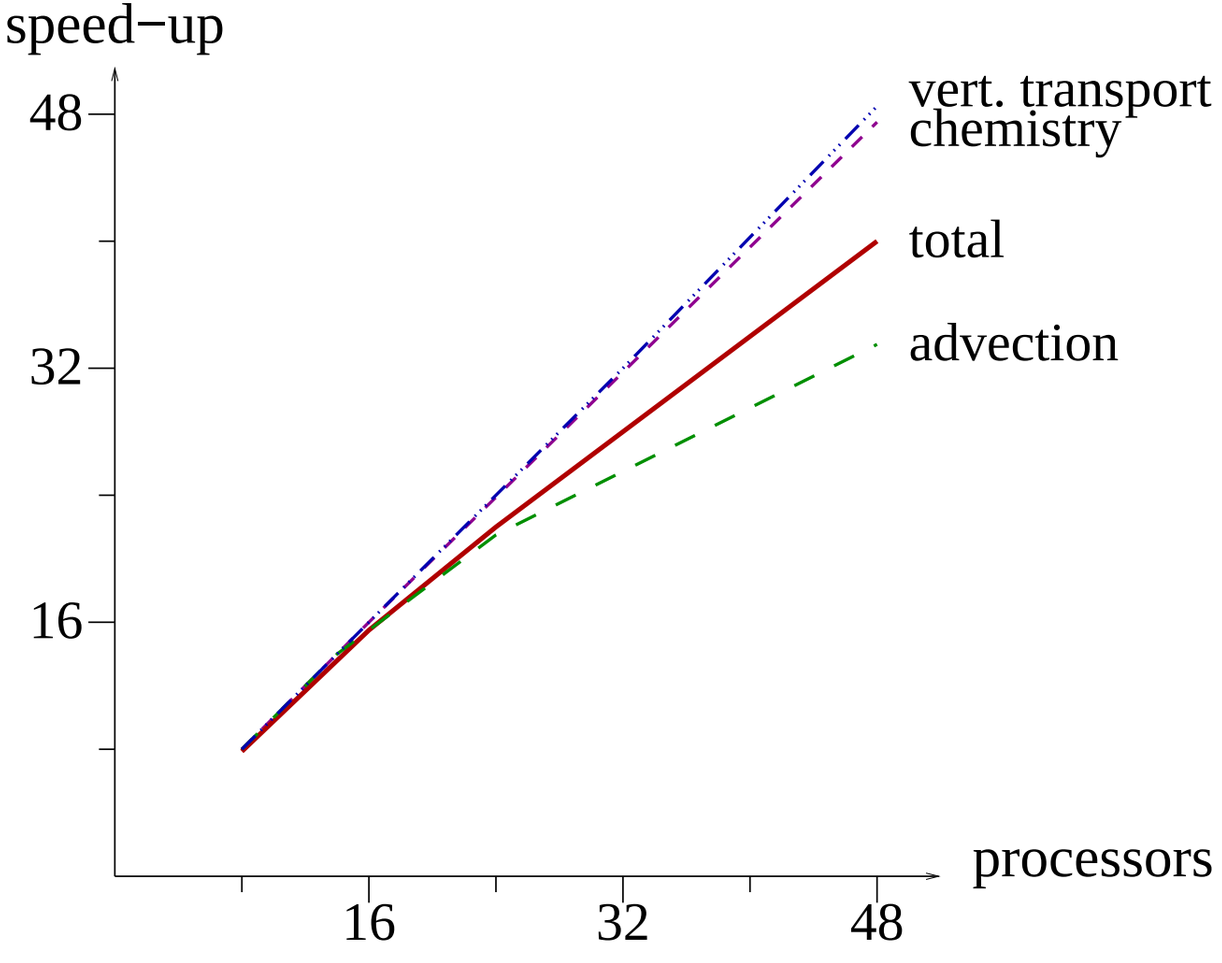
In the previous table: Time in seconds and relative weight of the main stages of the 3-D MPI version of DEM, $(96 \times 96 \times 10)$ grid. The results are obtained on a CRAY T3E computer at the EPCC by using chunks of size 24 in the chemistry part.

The ratio between the times on 8 and 32 processors is given in the last column.

Due to insufficient memory, there are no experiments on less than 8 processors.

In the next figure: Scalability of the main computational stages and the whole MPI implementation of the 3-D DEM on the T3E. The speed-ups are calculated under the assumption that it is 8 on 8 processors.

Scalability of the 3-D MPI code on the T3E



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 MPI code started while the first author was on a TRACS research visit in the Edinburgh Parallel Computer Centre, the University of Edinburgh, 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.

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. 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 (2000), 297–311.

3. A. Bastrup-Birk, J. Brandt and Z. Zlatev, Using partitioned ODE solvers in large air pollution models, Systems Analysis Modeling Simulation 32 (1998), 3–17.

4. K. Georgiev and Z. Zlatev, Parallel sparse matrix algorithms for air pollution models, Parallel and Distributed Computing Practices, to appear.

5. M. W. Gery, G. Z. Whitten, J. P. Killus and M. C. Dodge, A photochemical kinetics mechanism for urban and regional modeling, J. Geophys. Res. 94 (1989), 12925-12956.

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 (1994) 497–516.

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 (1978), 971–994.

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 (1988), 967–983.

10. G. I. Marchuk, Mathematical modeling for the problem of the environment, Studies in Mathematics and Applications, No. 16, North-Holland, Amsterdam, 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. Phys., **45,** 1984, 1-42.

12. Tz. Ostromsky, Z. Zlatev, Application of sparse matrix techniques in the chemical part of a large air pollution model, NNFM 62 (1998), 189–197.

13. Z. Zlatev, Application of predictor-corrector schemes with several correctors in solving air pollution problems, BIT 24 (1984), 700–715.

14. Z. Zlatev, Computer treatment of large air pollution models, Kluwer Academic Publishers, Dordrecht-Boston-London, 1995.

15. Z. Zlatev, Partitioning ODE systems with an application to air pollution models, Internal report. National Environmental Research Institute, Department of Atmospheric Environment, Frederiksborgvej 399, P. O. Box 399, DK-4000 Roskilde, Denmark, 1997.

16. Z. Zlatev, I. Dimov, K. Georgiev, Studying long-range transport of air pollutants, Computational Science and Engineering, 1 (1994) 45–52.

17. Z. Zlatev, I. Dimov, K. Georgiev, Three-dimensional version of the Danish Eulerian Model, Zeitschrift für Angewandte Mathematik und Mechanik, 76 (1996) 473–476.

18. 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.

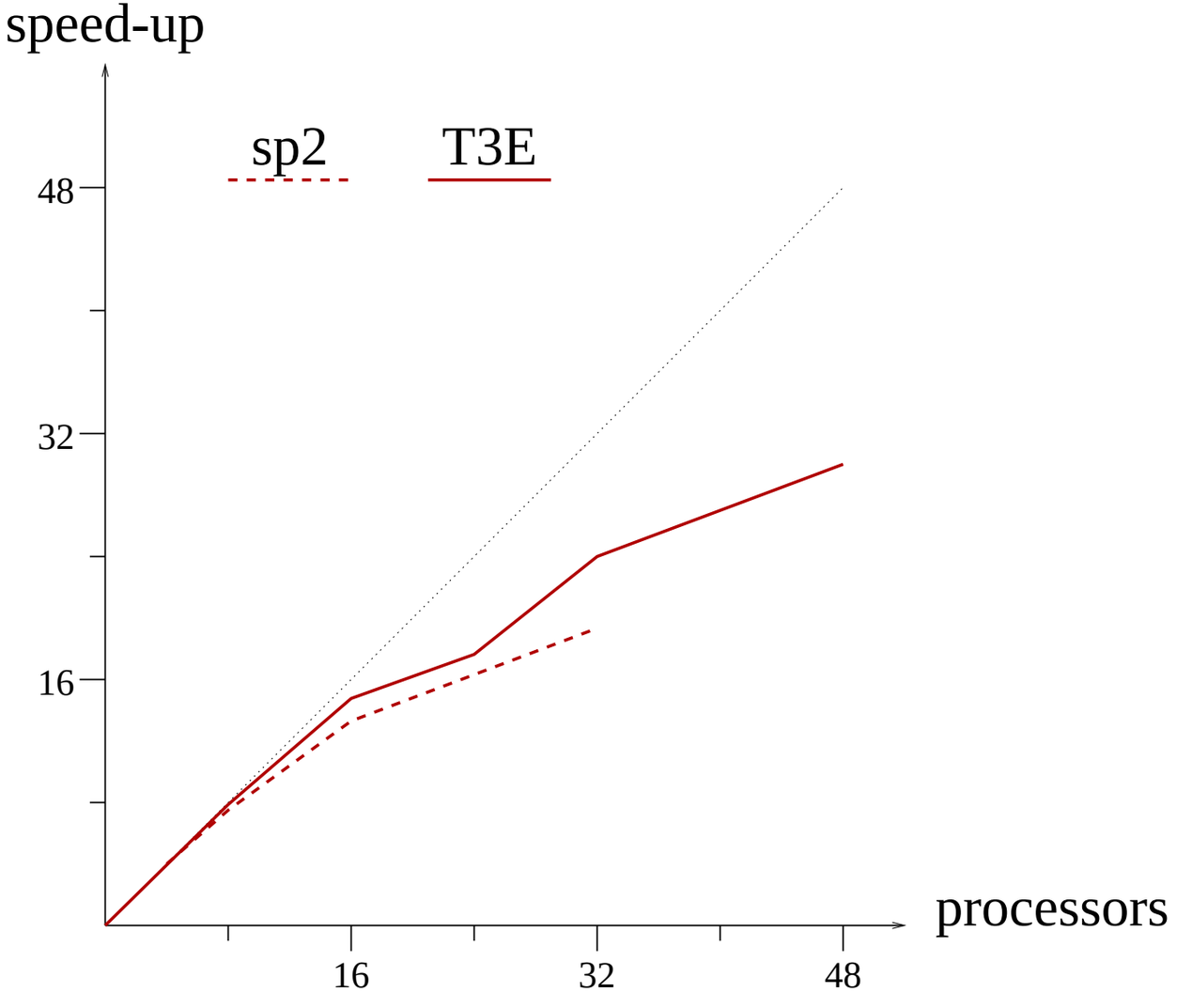
19. Z. Zlatev, G. Geernaert and H. Skov, A Study of ozone critical levels in Denmark, EUROSAP Newsletter 36 (1999), 1–9.

Times and performance of the 2-D coarse-grid codes on the SUN cluster

no. of proc.	OpenMP code		MPI code	
	Time [s]	Speed -up	Time [s]	Speed -up
1	4356		5067	
4	1199	3.6	1293	3.9
8	636	6.9	629	8.0
16	391	11.2	363	14.0

Time and speed-up of the OpenMP and the MPI versions of DEM for $(96 \times 96 \times 1)$ grid on the SUN HPC cluster at EPCC.

Scalability of the MPI code for the $(96 \times 96 \times 1)$ grid



Times and performance of the 3-D code

Stage	SUN E3500		SGI Origin-2000	
	Time [s]	Part [%] of Total	Time [s]	Part [%] of Total
W+S (Input)	105	0.3 %	59	0.1 %
Adv.+Diff.	8602	22.9 %	10458	21.1 %
Chem.+Dep.	25381	67.6 %	35232	71.2 %
Vertical tr.	3307	8.8 %	3465	7.0 %
Output	202	0.5 %	277	0.6 %
Total	37565	100 %	49483	100 %

Time profile for sequential runs of the 3D version of DEM on two shared-memory supercomputers