

# Pollutant's Dispersion in the Atmosphere. Parallel Models and Applications

Marcin Majer

Faculty of Electrical Engineering, Automatic Control  
and Computer Science  
Opole University of Technology  
Opole, Poland  
m.majer@po.opole.pl

Michał Podpora

Faculty of Electrical Engineering, Automatic Control  
and Computer Science  
Opole University of Technology  
Opole, Poland  
m.podpora@po.opole.pl

**Abstract**—The following article contains information about modeling of pollutants dispersion systems in the atmosphere. It also describes main types of pollutants, emission points and analysis of typical atmosphere pollutant dispersion models. It shows a way of using computer cluster systems in modeling process and proper programming libraries for parallel computing. Finally it presents a possibility of environment's protection method during the emission points' location planning with calculation of minimal destructive influence on a specified area (for example with definite restrictions of air pollution level).

**Keywords**—parallel computing; distributed computing; computer cluster; atmosphere; pollutant dispersion models; infrastructure expansion planning, MPI.

## I. INTRODUCTION

**T**O MODEL the process of pollution spreading in the natural environment, engineers and analysts extend typical atmosphere pollution monitoring systems to co-operate with computer clusters.

In a high technology age we live, single PC computers are often not enough for complicated computing problems contained in atmosphere pollutant dispersion models. It's also useful to use parallel programming methods in application creating process. There are many ready-to-use libraries for different programming languages but applying them is possible only if algorithm's model is designed for distributed architecture implementations.

The authors prepared an application to calculate hourly air pollutants concentrations in specified data points of an interest region. The application's algorithm involves solving emitter's location optimization problem in terms of fulfilling the levels of pollutants in a given area. Simulations and calculations were carried out on real input data describing specified area, including weather conditions. The following chapters describe the research and present the conclusions.

## II. POLLUTANTS DISPERSION IN THE ATMOSPHERE

### A. Air pollution

Air pollution becomes a huge problem. Generally, there are two main methods for pollutant's dispersion systems modeling: realized with physical models and mathematical models. Both of them have the same starting point: wide

knowledge about atmosphere's lower border layer characteristic. Physical models are based on executing atmospheric processes in laboratory conditions, whereas mathematical models contain mathematical description for physical and mathematical processes built on empiric measurement data. They are often called simulative deterministic models. In order to create a proper mathematical model it's important to classify substances in atmospheric air. It's not a challenge to find a number of substances in the atmosphere that are harmful for people, animals or plants. Emission of those pollutants leads to changing average established air accumulation.

TABLE I. AVERAGE ESTABLISHED AIR ACCUMULATION

| Gas type          |                  | Content [%]        |
|-------------------|------------------|--------------------|
| Nitrogen          | N <sub>2</sub>   | 78,09              |
| Oxygen            | O <sub>2</sub>   | 20,95              |
| Argon             | Ar               | 0,93               |
| Carbon dioxide    | CO <sub>2</sub>  | 0,03               |
| Neon              | Ne               | 2·10 <sup>-3</sup> |
| Helium            | He               | 5·10 <sup>-4</sup> |
| Methane           | CH <sub>4</sub>  | 2·10 <sup>-4</sup> |
| Krypton           | Kr               | 1·10 <sup>-4</sup> |
| Hydrogen          | H <sub>2</sub>   | 6·10 <sup>-5</sup> |
| Nitrogen monoxide | N <sub>2</sub> O | 3·10 <sup>-5</sup> |
| Xenon             | Xe               | 9·10 <sup>-6</sup> |

The main parameters for air pollution description are: air pollutant concentration (relation of the pollutant quantity and air capacity containing it) and pollutant's stream directed to the ground (pollutant quantity ground settled per unit of area in a unit of time).

### B. Air pollution spreading models

At the beginning of building atmosphere pollution monitoring system it is important to perform the parameters selection during the process of building the atmosphere's

pollution state describing equation. The parameters are strongly dependent on various factors (inter alia: territorial, meteorological, and factors describing the point of pollutant's emission). The following steps are to be performed:

- collecting the data with release of dirt from different sources (for example the chimney as a punctual source),
- checking the meteorological conditions (speed and direction of wind, coefficient of diffusion, temperature, humidity),
- gathering additional information (the form of terrain, stage of afforestation).

After completing these steps it's necessary to choose one of possible air pollution spreading models.

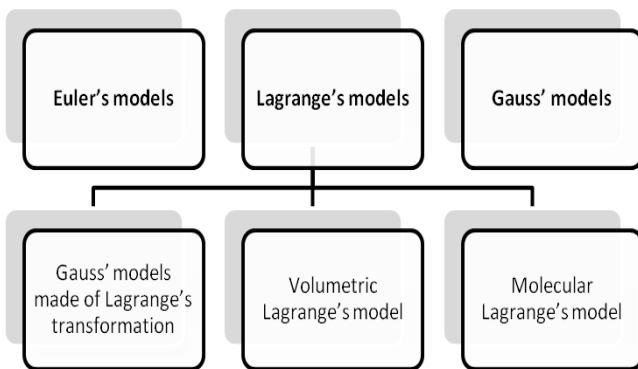


Fig. 1. Typical air pollution spreading models

There are three main types of pollution spreading models (fig. 1):

- Euler's models – the most important issue in these models is the pollution transport equation (1)[1]:

$$\frac{\partial \tilde{C}}{\partial t} + \frac{\partial(\tilde{u}\tilde{C})}{\partial x} + \frac{\partial(\tilde{v}\tilde{C})}{\partial y} + \frac{\partial(\tilde{w}\tilde{C})}{\partial z} = D_C \left( \frac{\partial^2 \tilde{C}}{\partial x^2} + \frac{\partial^2 \tilde{C}}{\partial y^2} + \frac{\partial^2 \tilde{C}}{\partial z^2} \right) + \tilde{S}_C \quad (1)$$

where:  $\tilde{C}$  - air pollutant concentration (relation of pollutant quantity and air capacity containing it),  $\tilde{u}, \tilde{v}, \tilde{w}$  – direction of wind vector's ( $\tilde{U}$ ) ingredients in square system of coordinates related with Earth along axes:  $OX, OY, OZ$ ;  $D_C$  – pollution diffusion coefficient,  $\tilde{S}_C$  – losses and pollutants source describing module.

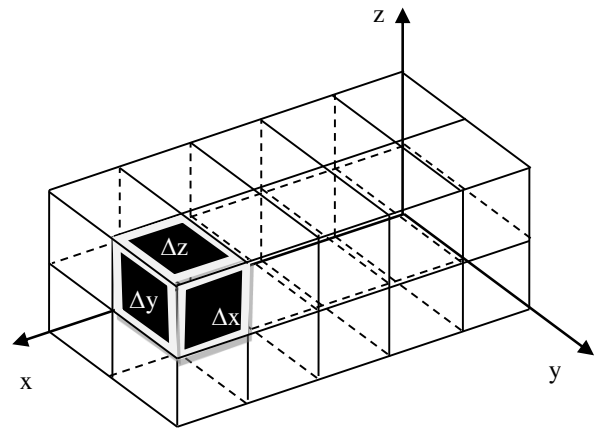


Fig. 2. Computation's scheme grid in Euler's air pollution spreading models (moving "box" of pollutants)

The most known Euler's formula implementations are: MISCAM (Germany 1996), MOD4 (Poland 1989), STEM-II (USA 1991).

- Lagrange's models – based on Lagrange's pollution dispersion equation (2):

$$\overline{C(r, t)} = C(r, t) = \int_{-\infty}^t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(r, t | r', t') S_C(r', t') dr', dt' \quad (2)$$

where:  $C(r, t)$  - air pollutant concentration averaged in realization class in  $r$  point for time  $t$ ;  $S_C(r', t')$  – losses and pollutants source describing module;  $p(r, t | r', t')$  - probability's density function that hypothetical air capacity that is placed in point  $r'$  at the time  $t'$  will be placed in point  $r$  at the time  $t$ .

The most known Lagrange's model implementations are: MATHEW/ADPIC (USA 1988), ARCO (Italy 1992), MDMS (USA 1983).

- Gauss models – based on either Lagrange's or Euler's pollution dispersion method with additional assumption equations for essential simplification. The main assumptions are:

- concentration's pole is stabile in time:  $\frac{\partial C}{\partial t} = 0$

-air moves are mostly horizontal, in  $OX$  axis direction with speed:  $u \geq 0, v = w = 0$

- there are no volumetric pollution sources:  $S_C = 0$

-pollutant advection's module is much stronger than turbulent module:

$$u \frac{\partial C}{\partial x} \gg \frac{\partial}{\partial x} \left( K_x \frac{\partial C}{\partial x} \right) \quad (3)$$

After all transformations the Gauss formula takes the form as follows (4):

$$(x, y, z) = \frac{E}{2\pi\sigma_y\sigma_z\bar{u}} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left[ \exp\left(-\frac{(z-H_e)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H_e)^2}{2\sigma_z^2}\right) \right] \quad (4)$$

where:  $C(x, y, z)$  - pollutant in air concentration at  $(x, y, z)$  point;  $E$  - pollutant emission's intensity;  $H_e$  - effective emission's height.

Gauss formula, defined in bibliography as Pasquill's formula (4) is strongly recommended to be applied in cases where average wind speed  $\bar{u} > 1 \frac{m}{s}$ . Furthermore, it is one of the most frequently realized and implemented models. The most known Gauss formula's implementations are: RAM (USA 1989), KOMIN (Poland 1998), AERMOD (USA 1995), CALPUFF (USA 1977, later modified).

### III. PARALLEL IMPLEMENTATION

There are plenty implementations of described models but most of them are sequential. Nowadays, extending these models to the capability of parallel/distributed implementation is highly recommended because of extremely long time of computer calculation. The best hardware solution is to use computer clusters if possible.

To implement a model in a distributed environment it is important to choose a suitable parallel programming library, and a supported programming language. Authors have chosen to adopt a sequential model by redesigning and implementing it using the LAM-MPI parallel library. The Message Passing Interface daemon executes the same code on all machines (nodes) in the same time; the nodes can compute individual parts of data and to exchange information by passing messages. The MPI specification-based libraries (e.g. LAM-MPI, MPICH2, etc.) are supposed to fully support heterogeneous computer clusters (i.e. of various operating systems and/or hardware architectures), the algorithm should be designed to support a variety of platforms (i.e. hardware configurations of a specific node, varying computational resources among the computer cluster's nodes). According to Flynn's classification [2] MPI implements MIMD model (Multiple Instruction stream - Multiple Data stream). Furthermore, MPI allows to use all the benefits of computer clusters:

- beneficial rate of price to performance,
- fault tolerance,

- high accessibility,
- hardware scalability,
- high performance.

Another way for minimizing computation time on computer clusters is by using OpenMP for additional parallelization within a specific node. OpenMP is an API that supports multi-platform shared memory multiprocessing programming in C, C++, and Fortran. An application implemented to support the hybrid model of parallel programming may be executed on a computer cluster using both OpenMP and Message Passing Interface (MPI), or more transparently through the use of OpenMP extensions for non-shared memory systems.

Supercomputers can bring different way profits. It's because of their hardware architecture. 2-level structure (figure 3) of some clusters gives multiply possibilities for parallel programming using MPI. It is a great step towards minimization of communication time. It is also possible to get additional acceleration of parallel algorithms by using simple methods like CPU gathering or shared memory utilization [3],[5].

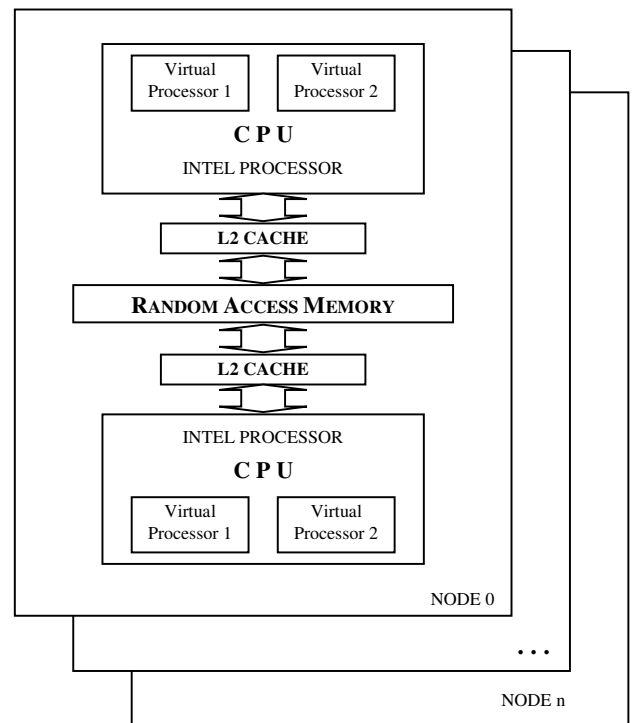


Fig. 3. Possible cluster's node architecture

The results of applying CPU gathering and shared memory utilization are shown in figure 4.

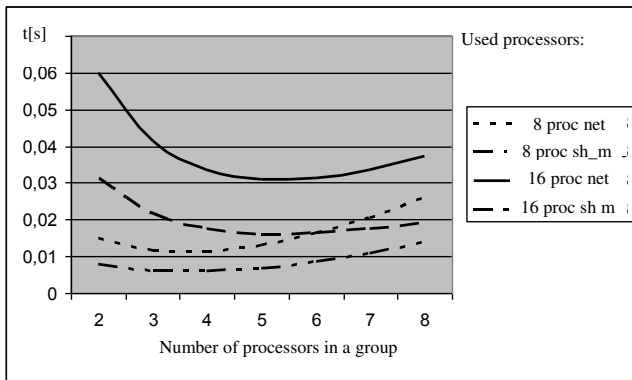


Fig. 4. Times of full data broadcast grouping processors (using shared memory and network) [own work]

#### IV. RESEARCH METHODOLOGY

During the application development, it was determined that the main task (further called “the solver”) was the calculation of hourly air pollutant concentrations in the specified data points of the interest region.

In relation to the assumed regional reach of the system controlling the proliferation of pollutants, a field grid was established for conducting measurements of weather factors. The span for it was 84.17 km per 83.56 km, however average distance of measuring knots was 27.67 km. Spreading test points of the grid was referred to GPS coordinates as it was shown in figure 5.

|                                    |                                    |                                    |                                    |
|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| [x=171; y=423]<br>16°26'E, 51°34'N | [x=178; y=423]<br>16°50'E, 51°34'N | [x=185; y=423]<br>17°14'E, 51°35'N | [x=192; y=423]<br>17°39'E, 51°35'N |
| [x=171; y=430]<br>16°27'E, 51°19'N | [x=178; y=430]<br>16°51'E, 51°19'N | [x=185; y=430]<br>17°15'E, 51°20'N | [x=192; y=430]<br>17°39'E, 51°20'N |
| [x=171; y=437]<br>16°28'E, 51°4'N  | [x=178; y=437]<br>16°52'E, 51°4'N  | [x=185; y=437]<br>17°16'E, 51°5'N  | [x=192; y=437]<br>17°40'E, 51°5'N  |
| [x=171; y=444]<br>16°29'E, 50°49'N | [x=178; y=444]<br>16°53'E, 51°49'N | [x=185; y=444]<br>17°16'E, 50°50'N | [x=192; y=444]<br>17°40'E, 50°50'N |

Fig. 5. GPS coordinates describing points of the weather factors grid [own work]

The factors describing weather conditions were gathered during 30 days period. After the analysis of the collected meteorological data, it was essential to average the data and to determine the dominating weather conditions in points of the measuring grid. After that process, the weather conditions data were ready to be applied to computation algorithms (fig. 6).

All the meteorological data, as it was shown above, were taken from measurements and observations. That methodology enables full result's verification and enables to examine the solver's behavior on a real model data.

|         | col=171                  |                             |                                  |                            |
|---------|--------------------------|-----------------------------|----------------------------------|----------------------------|
|         | Average wind speed [m/s] | Dominating wind's direction | Average air temperature [degr C] | Average air pressure [hPa] |
| row=423 | 3,67                     | SE                          | 20,33                            | 1018,25                    |
| row=430 | 3,50                     | SE                          | 20,05                            | 1018,28                    |
| row=437 | 3,32                     | SE                          | 19,78                            | 1018,35                    |
| row=444 | 3,10                     | SE                          | 19,25                            | 1018,57                    |
|         | col=178                  |                             |                                  |                            |
| row=423 | 3,60                     | SE                          | 20,20                            | 1018,05                    |
| row=430 | 3,50                     | SE                          | 19,95                            | 1018,17                    |
| row=437 | 3,27                     | SE                          | 19,80                            | 1018,20                    |
| row=444 | 3,17                     | SE                          | 19,62                            | 1018,35                    |
|         | col=185                  |                             |                                  |                            |
| row=423 | 3,60                     | SE                          | 20,20                            | 1018,05                    |
| row=430 | 3,50                     | SE                          | 19,95                            | 1018,17                    |
| row=437 | 3,27                     | SE                          | 19,80                            | 1018,20                    |
| row=444 | 3,17                     | SE                          | 19,62                            | 1018,35                    |
|         | col=192                  |                             |                                  |                            |
| row=423 | 3,50                     | SE                          | 19,87                            | 1017,72                    |
| row=430 | 3,40                     | SE                          | 19,93                            | 1017,72                    |
| row=437 | 3,28                     | SE                          | 20,08                            | 1017,68                    |
| row=444 | 3,28                     | SE                          | 20,32                            | 1017,67                    |

Fig. 6. Averaged and dominating weather factors according to field grid - used in computations [own work]

Technical parameters of the potential emitter in the form of the chimney were determined in the following way:

- height [m]: 44;
- inside diameter of the emitter's wire outlet: 1,6 m;
- speed of gasses on the exit of the emitter: 14,4 m/s
- gas temperature on the exit: 493,8 K;
- maximum emission of a gaseous substance: 43,9 mg/s.

There are also many other parameters needed for the solver, one of them is the aerodynamic coarseness coefficient of the area. It is selected according to the scheme shown in table 2.

TABLE II. THE AERODYNAMIC COARSENESS COEFFICIENT OF THE AREA SELECTION PROCESS [6].

|    | Type of the area's covering    | Aerodynamic coarseness coefficient of the area |
|----|--------------------------------|--|
| 1. | Water                          | 0,00008  |
| 2. | Meadows, pastures              | 0,02   |
| 3. | Farmland                       | 0,035  |
| 4. | Orchards, thickets, groves     | 0,4  |
| 5. | Forests                        | 2,0  |
| 6. | Country high-housing           | 0,5  |
| 7. | City up to 10000 residents     | 1,25   |
| 8. | City 10000 to 100000 residents | 2,75   |
| 9. | City over 500000 residents     | 3,75   |

To complete all the necessary entry values for calculating model one more factor is needed - the atmosphere's balance factor (ATM). It is determined by the observations of atmosphere's stormy nature and additionally it's described by scopes of the permissible wind speeds for individual states. The principle of atmosphere's balance factor's selection is shown in the table 3.

TABLE III. THE ATMOSPHERE'S BALANCE FACTOR SELECTION PROCESS [6].

| Atmosphere's balance factor ATM | Atmosphere's balance state | Scope of the wind speed m/s |
|---------------------------------|----------------------------|-----------------------------|
| 1                               | highly unstable            | 1 – 3                       |
| 2                               | unstable                   | 1 – 5                       |
| 3                               | medium unstable            | 1 – 8                       |
| 4                               | indifferent                | 1 – 11                      |
| 5                               | medium stable              | 1 – 5                       |
| 6                               | stable                     | 1 – 4                       |

Preliminary computational tests of proper working and computer cluster's usage reasonability were investigated during Pasquill's formula model implementation in a Linux environment. The solver was implemented in C++ using an MPI library and OpenMP directives.

The computational resources for the research included the computer cluster available on the Faculty of Electrical Engineering, Automatic Control and Computer Science at Opole University of Technology. It is a processing unit with following technical parameters:

TABLE IV. CLUSTER'S TECHNICAL SPECIFICATION

|  |                               |
|--|-------------------------------|
| Processor architecture:                | x86 64 bit                    |
| Processor type and frequency:          | AMD Phenom II X6 1090T s. AM3 |
| Number of processors in a single node: | 1                             |
| Number of cores in a single processor: | 6                             |
| Number of nodes:                       | 16                            |
| Number of cores (summary):             | 48                            |
| RAM available in a single node:        | 8 GB                          |
| Hard drive capacity (summary)          | 5 TB                          |
| Network:                               | 2 x Gigabit Ethernet          |

The operating system installed in the nodes of the computer cluster is Linux. The main solver tests concerned legitimacies of the computer clusters' application in this kind of computation as a pollutant's dispersion in the atmosphere's parallel models and applications issues.

V. RESULTS

Calculating hourly concentrations of chosen pollutant is one of the examples for CPU consuming calculations

benchmarking the cluster computational efficiency through a potential modeled chimney, for the domains adjacent to the chimney towards the dominating wind on the given area. The values were calculated for the points distant from oneself at 1 m in the distance of 10000 m in directions parallel to the x axis and 5000 m perpendicularly to the x axis (symmetrically towards axis).

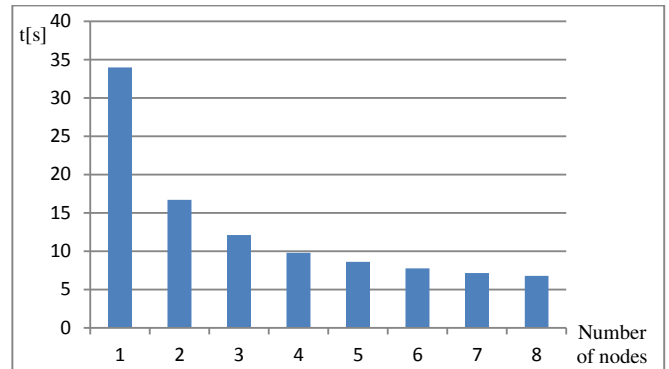


Fig. 7. Relation of the computation time of the hourly concentration of pollutants for grid of 10000 x 5000 points to the number of computational nodes [own work]

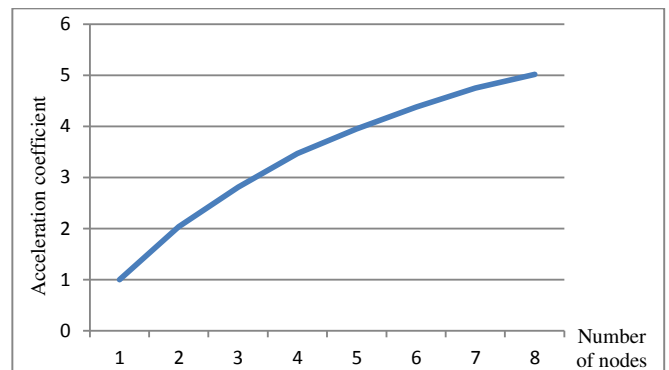


Fig. 8. Dependence of the coefficient of the acceleration on the number of computational nodes for calculation times introduced in figure 7.

In another test problem the MPI library was replaced with OpenMP directives. It required redesigning of the algorithm.

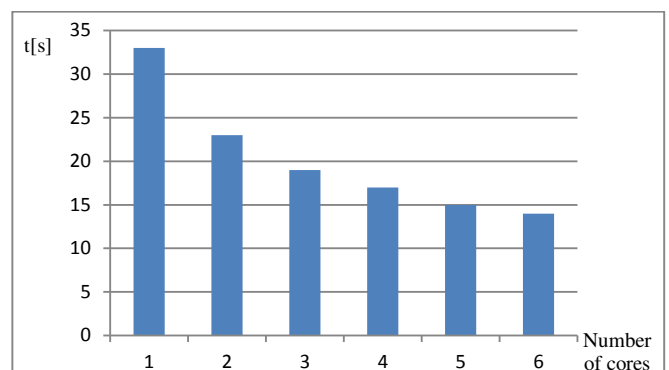


Fig. 9. Relation of the computation time of the hourly concentration of pollutants for the grid of 10000 x 5000 points to the number of processor's cores in a single node.

It allowed for checking the advantages of multicore processors programming on the level of single node. The test's results were shown at the following figures.

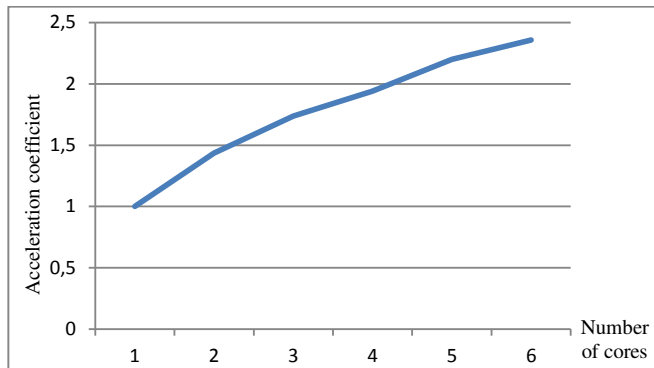


Fig. 10. Dependence of the coefficient of the acceleration on the number of computational processor's cores for computation times introduced in figure 9.

The following figure presents the solver's results of hourly pollutant's concentration calculation. That kind of data can be used in further computation of optimizing placing new emitter to fulfill minimal destructive influence assumptions on a specified area.

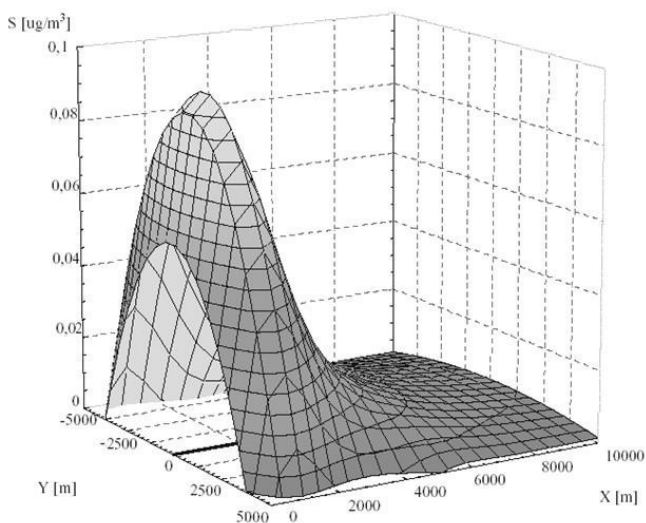


Fig. 11. Results of hourly pollutant's concentration calculation

## VI. SUMMARY

The specificity of computer clusters and software configuration creates a real basis for efficiency improvement of solving the computational pollution's spreading and accumulation process in the atmosphere issues. Parallel programming models can be used to effectuate a higher degree of parallelism via the simultaneous execution of separate parts of application's code on different processors, also grouping processors in calculations tasks. Parallelization of programs continues to remain a human challenge but it can bring high profits. Furthermore, this type of results of calculation can be used to emitter's location optimization problem in terms of fulfilling the levels of pollutants in a given area. It can be useful during infrastructure expansion planning process.

## REFERENCES

- [1] Findeisen W., Szymanowski J., Wierzbicki A.: Metody obliczeniowe optymalizacji. WPW, W-wa 1973
- [2] Flynn M.J.: Some Computer Organizations and Their Effectiveness, IEEE Trans. on Comp. vol C-21, No 9, 1972
- [3] Kaliczyńska M., Sadecki J.: Obliczenia równoległe – klastry obliczeniowe. Elektryka z. 57, Politechnika Opolska, Opole 2006
- [4] Karbowski A., Niewiadomska – Szykiewicz E.: Obliczenia równoległe i rozproszone. OWPW, W-wa 2001
- [5] Sadecki J.: Algorytmy równoległe optymalizacji i badania ich efektywności; systemy równoległe z rozproszoną pamięcią. Studia i monografie. Z.126, Politechnika Opolska, Opole 2001
- [6] Dz. U. 2010 nr 16. poz 87. 2010. Rozporządzenie Ministra Środowiska z dnia 26 stycznia 2010 roku w sprawie wartości odniesienia dla niektórych substancji w powietrzu. Warszawa: Dz. U. 2010, nr 16. poz 87., 2010