HIGH QUALITY AIR POLLUTION DISPERSION MODELLING USING HIGH COMPUTATIONAL PERFORMANCE LAGRANGIAN PARTICLE MODEL

Boštjan Grašič[1], Primož Mlakar[1], Marija Zlata Božnar[1], Juš Kocijan[2] and Gianni Tinarelli[3]

[1] MEIS d.o.o., Mali Vrh pri Šmarju 78, SI-1293 Šmarje - Sap, Slovenia
[2] Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
[3] ARIANET Srl, Via Gilino n. 9, 20128 Milano, Italy

Abstract: High quality air pollution dispersion modelling over complex terrain using Lagrangian particle model consumes a lot of computational time. In this study has been determined that there are some air pollution situations where unnecessary computational time is spent like accumulation of air pollution in the domain during the calm or extra high emissions from sources due to temporary failure of exhaust filters. To reduce unnecessary spent computational time a solution based on the control of number of virtual particles has been found. Existing Lagrangian particle air pollution dispersion model has been upgraded into new high computational performance model using integration of three methods to control the number of particles in the Lagrangian simulations. First method based on artificial neural networks is used to directly control the number of emitted virtual particles by predicting meteorological situations when too much particles will start to accumulate in the domain. Second method represents the use of cell concentration estimation method based on kernel density to assure high quality of the results. And the third method is based on the clustering algorithm to decrease the number of virtual particles in special situations if the first method can not enough decrease the number of particles. This integration of three methods ensures approximately constant number of particles in the modelling domain during whole simulation and consequently also constant and high quality of modelling results. Enhanced air pollution dispersion model has also been validated on the Šaleška region data set which represents highly complex terrain located in Slovenia. Validation confirmed the good quality of modelling results where spent computational has been reduced up to 50%.

Key words: *Lagrangian particle air pollution dispersion model, computational performance and time, control virtual particles, artificial neural networks, cell concentration estimation method, kernel density, clustering*

INTRODUCTION

Air-pollution modelling based on the Lagrangian particle dispersion has evolved rapidly over the past ten years, and it has developed from being just a research tool to being used for operational regulatory purposes (Tinarelli, G. et al, 2000, Graff, A., 2002). In the past ten years significant progress was also achieved in the computer industry, thus making possible the use of Lagrangian particle-modelling techniques on personal computers. Simple problems that could be simulated only for a short period of time several years ago on dedicated workstations (or super computers), can these days be simulated in a much more complex form and for a longer period of time on personal computers. In recent years the general opinion is that the technique does not need any computational improvements because the time-consuming problems will be solved by the development of computers with greater processing power. But this is only a solution for today's requirements.

Unfortunately, however, future requirements are expanding rapidly: the areas of interest are becoming wider, the resolution demands are increasing, the accuracies must be better, a better sensitivity is demanded, longer periods of time must be reconstructed, and the complexity and the number of sources and species are increasing, etc. These increasing requirements bring us back to the initial problem of the Lagrangian particle-modelling technique, which is extremely expensive from the computational point of view.

In this research the air-pollution modelling methodology based on Lagrangian particle dispersion is evaluated. Its limit capacities, properties and performance are determined and evaluated for a complex terrain. Based on these results several methods are suggested to improve the computational performance. The new methods are developed in a manner such that the basic physical properties of the Lagrangian particle-dispersion model are not modified. All the parameters and methods of the model are preserved in their original form and there are no adjustments of the well tuned parameters. **Methods to determine and simplify computationally expensive situations presented in details and integration into the existing Lagrangian particle-dispersion model is proposed to improve the computational performance in order to optimally exploit the available computational capabilities.**

PROBLEM IDENTIFICATION

Atmospheric dispersion models are used to present how pollutants in the ambient atmosphere disperse and, in some cases, how they react in the atmosphere with other atmospheric compounds. They are important to governmental agencies whose task is protecting and managing the quality of the ambient air. Several different air-pollution modelling approaches and techniques can be used to reconstruct the state of the atmosphere around the air-pollution sources. The time evolution of the air-pollution dispersion of the emitted pollutant depends on

the wind speed and direction, the turbulence, temperature, humidity, air pressure, solar radiation and precipitations as well as on the terrain's complexity. The meteorological parameters (variables) are usually four-dimensional, which means that they depend on space and time. The terrain complexity and land use data should also be considered. Reconstruction of the air-pollution evolution is, in principal, understanding the meteorological conditions and the terrain's complexity. While some of these conditions can cause high concentrations of an air pollutant near the surface for an extended period of time, some other conditions can decrease the concentration levels much more rapidly.
Air-pollution models based on Lagrangian particle dispersion are generally accepted as the most powerful tools to model the dispersion of atmospheric pollutants in the boundary layer over the local scale domain in fine spatial and time resolution (Wilson, J. D. and B. L. Sawford, 1996, Schwere, S. et al., 2002). Their quality has continually improved over the past 10 years. Most of the improvements gradually increased the complexity of the models. This leads to their main disadvantage: air-pollution reconstruction based on the Lagrangian particle-dispersion model can become computationally very expensive. This disadvantage gets emphasized when complex terrain conditions are present. The time used for each air-pollution episode reconstruction is critical when used in on-line systems as well as for off-line systems. Both types of system, the on-line and the off-line, are of specific importance to our work.

The purpose of this research is to determine computationally expensive procedures in an air-pollution model based on Lagrangian particle dispersion and to implement new techniques to improve them. The need for improvements is important because the improvements of the computer performance cannot match the increasing requirements.

OTHER APPROACHES FOR OPTIMISING THE COMPUTATIONAL EFFICIENCY OF THE LAGRANGIAN PARTICLE-DISPERSION MODEL AND SIMULATION

The excessive computational demand of Lagrangian particle-dispersion models represents their major disadvantage. With the constant quality improvements of models, their complexity has gradually increased, which leads to enormous requirements for computational time. Not much work has been published so far on the aspect of saving computational time in comparison to the large amount of work that has been published on the quality of the model's results. In order to speed up simulations based on Lagrangian particle-dispersion models some work has already been done in the field of simplifying the underlying physics, for example, by Hurley, P. and W. Physick (1993), Ryall, D. B. and R. H. Maryon (1998), Ermak, D. L. and J. S. Nasstrom (2000). Another approach to simplifying the computational demands was proposed by Melheim, J. A. (2005). The clustering method for different purposes was presented in his paper to reduce the number of equations to be solved in collisional particle dynamics in the Lagrangian framework. To decrease spent computational time also parallelization has been implemented in the research by Oldrini et al. (2011). It has been successfully used to increase the computational efficiency on very large high resolution domains with large number of air pollution sources. In this study two approaches of parallelization have been implemented: firstly splitting geographically big domains through classical Eulerian tile splitting and secondly parallelization based on code specific properties to gain speed up. Similar approach is also presented at this conference in the study by Assennato et al. (2013). In this paper the focus is on analysing the efficiency of the numerical procedures involved in simulation procedures rather than analysing the underlying physics. Similar goals had been set in the contribution by Schwere, S. et al. (2002) where some specific tasks that waste computing time were outlined and three methods to reduce the computing time of these tasks were presented. In this paper additional supplementary methods are proposed.

LAGRANGIAN PARTICLE-DISPERSION SIMULATION CONTROL

When the number of particles in a single air-pollution reconstruction episode (representation of air pollution process for a limited time interval) exceeds the computational resources, the simulation process is interrupted and the reconstruction of the proceeding air-pollution episode begins without a consideration of results of the previous air-pollution situation. When such a problem occurs, the high air-pollution situation reconstruction can be lost. The number of new, active particles that are released during a single air-pollution reconstruction episode can be calculated from particle density and the emission during the episode. To reconstruct a single air-pollution episode correctly the emission cannot be changed. To avoid an excess of computational resources in a single air-pollution reconstruction episode the particle density should be controlled: when a high emission occurs, it should be decreased, and when a low emission occurs, it can be increased.
Proposed concept of **Lagrangian model simulation control is based on a model constructed of feed-forward neural network** and its main idea is to change particle density in the domain during the simulation to avoid the excess of computational resources.
Control of the particle density can depend on the number of active particles from the reconstruction of a previous air-pollution episode reconstruction and the emission in the current air-pollution episode reconstruction. To

improve the control an estimation of number of active particles that are lost during the proceeding air-pollution reconstruction episode can also be used:

- if many active particles are lost out of the domain during the simulation a larger particle density parameter should be used because more new active particles can be released to simulate the same amount of emission,
- if a very small number of active particles are lost out of the domain, a smaller particle density should be used because fewer new particles can be released to ensure that the computational resources will not be exceeded.

In special situations when the smallest possible particle density is used, and it is still expected that the computational resources will be exceeded, the clustering as presented in following section must be activated to reduce the number of particles from previous air pollution. This is very important for situations when extreme air pollution is expected. Such situations occur very often in calm meteorological conditions when the air pollution starts to accumulate in the domain for a longer time interval. Control method consists of two main subsequent methods. In the first step the percentage of lost particles is predicted based on the meteorology, the emission and the situation of the air pollution at the end of the previous episode reconstruction. In the second step the clustering parameters are determined by a decision-making method.

FEEDFORWARD MULTILAYER PERCEPTRON NEURAL NETWORKS AS PROPOSED TOOL TO IMPLEMENT THE CONTROL

Artificial neural networks are selected for the proposed improvements because it was proven by Hornik, K. et al. (1989) that multilayer feed-forward networks are universal approximators. They have become a useful and efficient tool, in the past ten years, for establishing forecasting models in the field of air pollution. Artificial neural networks are very often used for system identification and control because of their ability to learn nonlinear relationships (Kocijan, J., 2007).

The feed-forward multilayer perceptron neural network was the first developed artificial neural network that was proved to be universal approximator by Hornik, K. et al. (1989): the universal approximation theorem for neural networks states that every non-linear continuous function that transforms intervals of real numbers to some output intervals of real numbers can be approximated arbitrarily closely by a multi-layer perceptron with just one hidden layer. The theorem holds only for restricted classes of non-linear activation functions (that neurons are constructed of) like, for example, the sigmoid functions**. In practice the feedforward neural networks represent a non-linear statistical data modelling tool and can be used to model complex relationships between inputs and outputs** (Engelbrecht, A. P., 2002)**.**
Artificial neural networks have become a useful and efficient tool in the past ten years for establishing forecasting models in the field of air pollution. The successful forecasting of air pollution using artificial neural networks in recent years was reported by many authors (Božnar, M. Z. et al., 1993, Mlakar, P. and M. Z. Božnar, 1996, Gardner, M. W. and S. R. Dorling, 1996, Grašič, B. et al., 2006).
From several above enumerated literature it is recommended that this type of neural network is one of the best suited solutions for the above explained problem that we had to solve.

METHOD FOR ESTIMATION OF A CELL CONCENTRATION BASED ON KERNEL DENSITY

In air-pollution modelling the methodology based on the Lagrangian particle dispersion concentrations are estimated by counting the particles in a cell that has a rectangular shape. The term “box counting” is used in the literature to denote this concentration estimation method. A study was performed by De Haan, P. (1999) where the effects of the different size and position of the boxes on the estimated point concentration was investigated. The outcomes of the study showed that if the boxes sizes are small, the concentration distribution becomes very noisy, having a large variance, and that if the sizes of the boxes are too large, the concentration becomes over smoothed, having a large bias. To minimise the sum of the variance and the bias, the kernel estimation method was proposed as an alternative, which also allows the number of particles to be reduced by an order of magnitude as compared to predictions made by the box counting method.
The presented study by De Haan, P. (1999) was focused on point-concentration estimations. For regulatory purposes the ground-level concentrations for a certain area of interest must be reconstructed. The focus must be extended from point concentration estimations to cell concentration estimations. Usually, the area of interest is split into a grid of rectangular boxes or cells, and for each cell at ground level the concentration is estimated by counting the particles caught in a certain cell. When not enough particles are used in the Lagrangian simulation, the concentrations between the neighbouring cells can become very different and the concentration distribution becomes very noisy over the area. From past work we have the experiences that the results become bubbled and less smooth when a smaller number of particles are used for the simulation. Because less heavier particles are released during the simulation, the concentrations in certain cells become very high, while the concentrations in neighbouring cells are zero. From the point of view of an inexperienced observer some erroneous conclusions

could be made, like, for example, that only half of the village or only few streets has been exposed to air pollution. A similar effect is also faced when clustering is introduced to the model to reduce the computational expense as presented in following section.

A method for the concentration estimation is proposed to reduce the presented effect. The method is complementary to clustering as proposed in following section. It can be used independently to improve the results when not enough particles are used in the simulation. Another possibility is a combination with the clustering method when the number of particles after the clustering process is significantly reduced. To improve the results of the ground-level concentration estimations a method proposed by De Haan, P. (1999) is adopted and expanded from the point-concentration estimations to the cell-concentration estimations. In the study by De Haan, P. (1999) a quad-weight kernel was recommended as being optimal, but the results of the other kernels were not falling behind significantly. The cell-concentration estimation method that is implemented and tested in detail in this research is based on a Gaussian density kernel.

CLUSTERING METHOD

In practice the number of particles is constrained by the computer resources and the acceptable duration of the simulation. A larger number of particles usually results in more time being used for the simulation and to a greater computational cost. The clustering method is needed to decrease the computational cost while preserving the quality of results.

The number of particles in the area of interest during the simulation varies. It increases due to the emissions of new particles (species) from a different source in the area and decreases due to different meteorological conditions, especially the wind, which pushes the particles out of the domain. A smaller number of particles is also achieved due to their exposure to dry and wet deposition. In the current air pollution computer models the user can influence the number of particles only through the emission from sources where the same emission can be simulated with a smaller number of particles by increasing the weight of the emitted particles. This can be performed only to a certain level. When that level is reached the quality of the results begins to decline drastically.

An idea occurred to control the maximum number of particles in the domain by modifying the current air-pollution modelling methodology with the use of clustering methods generally. This means that certain particles are joined according to some rules into new, heavier particles. These new particles are introduced into the domain and the old, lighter particles are removed. The properties of the new particles are composed of the properties of the old, lighter particles. In this way the total number of particles in the domain (the area of interest) is significantly reduced and therefore the reconstructions of the proceeding air-pollution episodes (after the clustering) is significantly less computationally demanding. Within the conventional clustering method concept the number of clusters is equal to the maximum allowed number of particles. The particles are joined together only in clusters, which consist of more than one particle.

The clustering method for different purposes was already used, by Melheim, J. A. (2005) and presented in his paper, to reduce the number of equations to be solved in particle dynamics. The main idea is that only particles that interact or may interact during the next global time-step are integrated simultaneously, while the particles that are far from the other particles can be integrated alone. The clusters of particles were made of particles that interact or may interact during the next global time-step.

Clustering methods were proved to be a very effective tool also for environmental applications by Mlakar, P. and M. Z. Božnar (1996) in a cluster analysis of wind fields and SO_2 concentrations based on the Kohonen neural network

Two clustering methods are proposed as result of this research: the Kohonen neural network (SOM *clustering algorithm*) or the K-means clustering algorithm.

CONCLUSIONS

All the proposed methods in the previous sections have been validated on a complex terrain field data set air-pollution situation that is based on an experimental measuring campaign performed over the Šaleška region (Elisei, G. et al., 1992, Grašič, B. et al., 2011). The results of the comparisons show that the time used and the particle number in the original simulation are strongly depended on the current meteorological conditions and the emission rate, while in the the simulation using new high computational perfomance model are practically constant. A significant improvement of the correlation coefficients is achieved. Also, the fractional bias comparison showed that there is practically no underestimations in the results of the advanced simulation, which is crucial for any long-term evaluations of air pollution. The obtained results show that the use of the high computational perfomance computer model is recommended, not only in situations where the computational resources are constrained, but also in general to optimally take advantage of the available computational resources. Validation confirmed the good quality of modelling results where spent computational has been reduced up to 50%.

ACKNOWLEDGEMENT

The study was partially financed by the Slovenian Research Agency, Projects No. L1-4154.

REFERENCES

Assennato, G., M. Bevere, R. Giua, A. Morabito, T. Pastore, S. Spagnolo, A. Tanzarella, E. Valentini , G. Brusasca, G. Tinarelli, 2013: : Application of a Lagrangian particle model to source the apportionment for primary in Taranto area (south Italy). *HARMO15: proceedings of the 15th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 6-9 May 2013, Madrid, Spain*.

Božnar, M. Z., M. Lesjak and P. Mlakar, 1993: A neural network-based method for short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain. *Atmospheric Environment* Vol. **27 B**, 221-230.

De Haan, P., 1999: On the use of density kernels for concentration estimations within particle and puff dispersion models. *Atmospheric Environment* Vol. **33**, 2007-2022

Elisei, G., S. Bistacchi, G. Bocchiola, G. Brusasca, P. Marcacci, A. Marzorati, M. G. Morselli, G. Tinarelli, G. Catenacci, V. Corio, G. Daino, A. Era, S. Finardi, G. Foggi, A. Negri, G. Piazza, R. Villa, M. Lesjak, M. Božnar, P. Mlakar and F. Slavic, 1992: Experimental Campaign for the Environmental Impact Evaluation of Šoštanj Thermal Power Plant, Progress report. CISE Technologie innovative, Milano, Italia, ENEL S.p.A, CRAM-Servizio Ambiente, Milano, Italy, C.I.S.E. Tecnologie Innovative S.p.A, Milano, Italy, Institute Jozef Stefan, Ljubljana

Engelbrecht, A. P., 2002: Computational Intelligence: An Introduction. Chichester, England, John Wiley & Sons, Ltd

Ermak, D. L. and J. S. Nasstrom, 2000: A Lagrangian stochastic diffusion method inhomogeneous turbulence. *Atmospheric Environment* Vol. **34**, 1059-1068.

Gardner, M.W. and S. R. Dorling, 1996: Neural network modelling of the influence of local meteorology on surface ozone concentrations. *Proceedings 1st International Conference on GeoComputation, University of Leeds*, 359-370.

Graff, A., 2002: The new German regulatory model – a Lagrangian particle dispersion model. *8th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, October 14-17, Sofia, Bulgaria, 153-158.

Grašič, B., P. Makar and M. Z. Božnar, 2011: Method for validation of Lagrangian particle air pollution dispersion model based on experimental field data set from complex terrain. V: NEJADKOORKI, Farhad (ur.). *Advanced air pollution*. Rijeka: InTech, cop. 2011, 535-556.

Grašič, B., P. Mlakar and M. Z. Božnar, 2006: Ozone prediction based on neural networks and Gaussian processes. *Il Nuovo Cimento C*, Vol. **29**, Issue 6, 651-661.

Hornik, K., M. Stinchcombe and H. White, 1989: Multilayer feedforward networks are universal approximators. *Neural Networks* Vol. **2**, Issue 5: 359-366.

Hurley, P. and W. Physick, 1993: A skewed homogeneous Lagrangian particle model for convective conditions. *Atmospheric Environment* Vol. **27 A**, 619-624.

Kocijan, J., 2007: Modeliranje dinamičnih sistemov z umetnimi nevronskimi mrežami in sorodnimi metodami (in Slovene). Nova Gorica, Slovenia, Univerza v Novi Gorici

Melheim, J. A., 2005: Cluster integration method in Lagrangian particle dynamics. *Computer Physics Communications* Vol. **171**, 155-161.

Mlakar, P. and M. Z. Božnar, 1996: Analysis of winds and SO2 concentrations in complex terrain. Air pollution IV : monitoring, simulation and control. Southampton; Boston: Computational Mechanics Publications, 455-464.

Oldrini, O., C. Olry, J. Moussafir, P. Armand and C. Duchenne, 2011: Development of PMSS, the parallel version of MICRO-SWIFT-SPRAY. *HARMO14: proceedings of the 14th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 2-6 October 2011, Kos Island, Greece*., 443-447

Ryall, D. B. And R. H. Maryon, 1998: Validation of the UK Met. Office's NAME model against ETEX dataset. *Atmospheric Environment* Vol. **24**, 4265-4276.

Schwere, S., A. Stohl and M. W. Rotach, 2002: Practical considerations to speed up Lagrangian stochastic particle models. *Computers & Geosciences* Vol. **28**, 143-154.

Tinarelli, G., D. Anfossi, M. Bider, E. Ferrero and S. T. Castelli, 2000: A New High Performance Version of the Lagrangian Particle Dispersion Model SPRAY, Some Case Studies. *NATO challenges of modern society* Vol. **23A**, 499-508.

Wilson, J. D. and B. L. Sawford, 1996: Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere. *Boundary-Layer Meteorology* Vol. **78**, 191-210.