Position and Communication Papers of the 16th Conference on Computer Science and Intelligence Systems pp. 85–88
DOI: 10.15439/2021F51
ISSN 2300-5963 ACSIS, Vol. 26

Optimized stochastic methods for sensitivity analysis for large-scale air pollution model

Venelin Todorov
Bulgarian Academy of Sciensec
Institute of Mathematics and Informatics
ul. G. Bonchev 8, 1113 Sofia, Bulgaria
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria
Email: vtodorov@math.bas.bg,venelin@parallel.bas.bg

Tzvetan Ostromsky
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgria
Email: ceco@parallel.bas.bg

Ivan Dimov
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgria
Email: ivdimov@bas.bg

Rayna Georgieva
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
ul. G. Bonchev 25A, 1113 Sofia, Bulgria
Email: rayna@parallel.bas.bg

***Abstract*—Environmental security is rapidly becoming a significant topic of present interest all over the world, and environmental modelling has a very high priority in various scientific fields, respectively. Different optimizations of the Latin Hypercube Sampling algorithm have been used in our sensitivity studies of the model output results for some air pollutants with respect to the emission levels and some chemical reactions rates.**

I. Introduction

HIGH levels of pollution can disrupt ecosystems and cause harm to plants, animals and humans. Therefore, it is extremely important to investigate accurately the levels of contamination [9], [11]. It is necessary to know whether the pollution levels are below some critical values and if so - to develop a reliable control system to keep them within these limits. Mathematical models are used to study and predict the behavior of a variety of complex systems - engineering, physical, economic, social, environmental. They determine the most important quantities that control the state and behavior of a system, as well as the quantitative regularities, that is, the mathematical laws that underlie the change of these quantities. On the other hand, it is of significant importance to develop techniques to determine the reliability, robustness and efficiency of a mathematical model [10]. One such approach is sensitivity analysis (SA). Sensitivity studies are nowadays applied to some of the most complicated mathematical models from various intensively developing areas of application [1], [2], [3], [4]. In the present study we use UNI-DEM [12], [13], [14] which is a powerful large-scale air pollution model for calculation of the concentrations of a large number of pollutants and other chemical species in the air along a certain time period. Its results can be used in various application areas (environmental protection, agriculture, health care, etc.). The large computational domain covers completely the European region and the Mediterranean.

UNI-DEM simulates the long-range transport of the air pollutants, their transition in time as a result of the chemical and photochemical reactions between them and the interaction with the other components of the environment. It takes into account the basic physical processes - advection, diffusion, deposition, harmful emissions, as well as the chemical reactions. It provides an opportunity to study over time the concentrations of the main types of pollutants (sulfur, nitrogen, ammonia, ammonium ions, nitrogen, free radicals, hydrocarbons), which is important for environmental safety, agriculture, healthcare. The model domain includes the whole Europe and the Mediterranean plus parts of Asia and Africa. The approximate area covered is $4800 \times 4800\,km$.

II. Latin Hypercube Sampling

Latin Hypercube Sampling (LHS) is a type of stratified sampling and in the case of integral approximation we must simply divide the domain $[0,1]^d$ into m^d disjoint subdomains, each of volume $\frac{1}{m^d}$ and to sample one point from each of

Venelin Todorov is supported by the Bulgarian National Science Fund under Project KP-06-M32/2 - 17.12.2019 "Advanced Stochastic and Deterministic Approaches for Large-Scale Problems of Computational Mathematics" and by the National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICT in SES)", contract No DO1-205/23.11.2018, financed by the Ministry of Education and Science in Bulgaria. The work is also supported by Bulgarian National Science Fund under Project DN 12/5-2017 "Efficient Stochastic Methods and Algorithms for Large-Scale Problems", Project DN 12/4-2017 "Advanced Analytical and Numerical Methods for Nonlinear Differential Equations with Applications in Finance and Environmental Pollution", and by the Project KP-06-Russia/17 "New Highly Efficient Stochastic Simulation Methods and Applications" funded by the National Science Fund - Bulgaria.

them. Let this sample be $\mathbf{x}_{k,j}$, for dimensions $k = 1, \dots, m^d$, $j = 1, \dots, d$. LHS does not require more samples for more dimensions (variables) - it is one of the main advantages of this scheme. Examples of random, stratified and Latin Hypercube Samplings with 16 points are presented on Figure 1 [5].

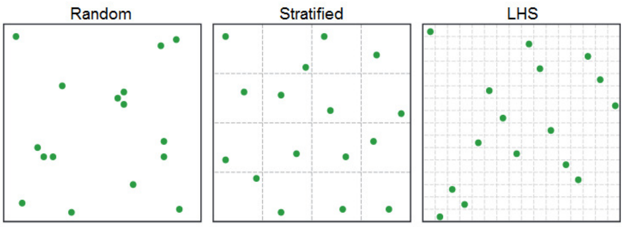

Fig. 1. Comparison of random, stratified and Latin Hypercube Samplings with 16 points (d = 2, m = 4).

Two different optimization versions of Latin Hypercube Sampling are compared with different seeds (seed is an integer used to initialize the corresponding pseudorandom number generator). The algorithms Latin Hypercube Sampling (random) [7], [8] LHSR1 (seed=1) and LHSR-1 (seed=-1) and Latin Hypercube Sampling (edge) [5], [6] LHSE1 (seed=1) and LHSE-1 (seed=-1) have been used in our comprehensive experimental study for the first time for the particular model. The difference is that the simple Latin edge algorithm returns edge points in a Latin square, while Latin random algorithm returns random points in a Latin square.

III. Sensitivity Studies with Respect to Emission Levels

In this section the results for the sensitivity of UNI-DEM output (in particular, the ammonia mean monthly concentrations) with respect to the anthropogenic emissions input data variation are shown and discussed. The anthropogenic emissions input consists of 4 different components

$\mathbf{E^A}$ – ammonia (NH_3);
$\mathbf{E^N}$ – nitrogen oxides $(NO + NO_2)$;

$\mathbf{E^S}$ – sulphur dioxide (SO_2);
$\mathbf{E^C}$ – anthropogenic hydrocarbons.

Results of the relative error estimation for the quantities f_0, the total variance $\mathbf{D}$, first-order (S_i) and total (S_i^{tot}) sensitivity indices are given in Tables I, II, III, IV, V, respectively. The quantity f_0 is presented by a 4-dimensional integral, while the rest of the above quantities are presented by 8-dimensional integrals, following the ideas of *correlated sampling* technique to compute sensitivity measures in a reliable way. The results show that the computational efficiency of the algorithms depends on integrand dimension and magnitude of estimated quantity. The order of relative error is different for different quantities of interest (see column *Reference value*) for the same sample size.

TABLE I
Relative error for the evaluation of $f_0 \approx 0.048$.

# of samples n	**LHSE1** Relative error	**LHSE-1** Relative error	**LHSR1** Relative error	**LHSR-1** Relative error
2^{10}	6.37e-05	2.45e-04	3.22e-04	2.43e-04
2^{12}	6.18e-05	1.55e-04	2.67e-04	7.11e-05
2^{14}	5.21e-06	8.44e-06	3.62e-05	8.30e-05
2^{16}	1.34e-05	1.58e-05	1.62e-05	1.01e-05
2^{18}	1.29e-05	5.31e-07	5.24e-05	1.52e-05
2^{20}	1.53e-05	8.48e-07	8.78e-06	2.24e-05

Table V is similar to Table IV and Table III, with the only difference – the increased number of samples $n = 2^{20}$ (instead of $n = 2^{16}$ in Table IV and $n = 2^{10}$ in Table III). In general, this increases the accuracy of the estimated quantities. The exceptions are S_4, S_4^{tot} and S_{15}, which have extremely small reference values. All LHS methods produce similar results, but for in small in value sensitivity indices LHSE-1 has the edge - see for example the value of S_2^{tot} in Table V.

TABLE II
Relative error for the evaluation of the total variance $\mathbf{D} \approx 0.0002$.

# of samples n	**LHSE1** Relative error	**LHSE-1** Relative error	**LHSR1** Relative error	**LHSR-1** Relative error
2^{10}	4.91e-02	2.78e-03	4.08e-02	4.68e-02
2^{12}	5.10e-03	9.60e-03	1.68e-02	1.40e-02
2^{14}	8.82e-03	7.55e-03	1.26e-02	1.12e-02
2^{16}	8.36e-03	9.72e-03	4.43e-03	4.27e-03
2^{18}	1.32e-04	7.90e-04	7.55e-04	1.14e-03
2^{20}	1.03e-03	8.64e-04	3.43e-05	7.19e-04

IV. Sensitivity Studies with Respect to Chemical Reactions Rates

In this section we will study the sensitivity of the ozone concentration values in the air over Genova with respect to the rate variation of some chemical reactions of the condensed CBM-IV scheme [12], namely: # 1, 3, 7, 22 (time-dependent)

TABLE III
Relative error for estimation of sensitivity indices of input parameters ($n \approx 2^{10}$).

Est. qnt.	Ref. val.	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
S_1	9e-01	1.13e-02	5.33e-03	4.39e-02	1.85e-01
S_2	2e-04	8.53e+00	7.23e+00	1.22e+00	5.36e+00
S_3	1e-01	1.48e-01	4.78e-02	4.53e-01	4.62e-02
S_4	4e-05	1.85e+01	1.22e+01	1.55e+01	1.06e+01
S_1^{tot}	9e-01	1.54e-02	4.63e-03	5.32e-02	6.91e-03
S_2^{tot}	2e-04	1.21e+01	4.87e+00	1.15e+01	6.36e+00
S_3^{tot}	1e-01	1.27e-01	4.53e-02	3.74e-01	3.26e-02
S_4^{tot}	5e-05	2.25e+01	1.71e+01	1.35e+01	5.66e+00

TABLE IV
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS ($n \approx 2^{16}$).

Est. qnt.	Ref. val.	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
S_1	9e-01	1.19e-03	2.72e-03	2.75e-03	5.98e-03
S_2	2e-04	6.16e-01	4.53e-01	1.99e-01	5.40e-02
S_3	1e-01	7.19e-03	1.49e-02	1.38e-02	3.66e-03
S_4	4e-05	9.65e-01	2.22e+00	4.11e+00	5.30e-01
S_1^{tot}	9e-01	7.04e-04	1.79e-03	1.81e-03	5.15e-04
S_2^{tot}	2e-04	9.78e-01	3.45e-01	2.36e-01	1.06e+00
S_3^{tot}	1e-01	7.67e-03	1.76e-03	2.02e-02	6.24e-04
S_4^{tot}	5e-05	8.66e-01	2.72e+00	1.08e+00	7.54e-01

TABLE V
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS ($n \approx 2^{20}$).

Est. qnt.	Ref. val.	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
S_1	9e-01	4.66e-04	1.46e-04	1.40e-04	2.99e-04
S_2	2e-04	6.38e-03	5.63e-02	1.98e-01	2.55e-01
S_3	1e-01	4.01e-03	1.27e-03	4.06e-04	4.05e-04
S_4	4e-05	7.09e-01	4.81e-01	5.24e-01	1.77e-01
S_1^{tot}	9e-01	4.97e-04	1.34e-04	2.26e-05	1.31e-04
S_2^{tot}	2e-04	5.40e-02	1.59e-03	3.09e-01	3.49e-01
S_3^{tot}	1e-01	4.42e-03	3.09e-02	6.25e-04	1.71e-03
S_4^{tot}	5e-05	5.85e-01	1.08e+00	3.11e-01	1.02e-01

TABLE VI
RELATIVE ERROR FOR THE EVALUATION OF $f_0 \approx 0.27$.

# of samples	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
n	Relative error	Relative error	Relative error	Relative error
2^{10}	4.90e-05	5.86e-04	9.90e-05	1.19e-03
2^{12}	7.47e-05	1.05e-04	7.21e-04	1.04e-04
2^{14}	2.43e-04	1.89e-04	3.38e-04	2.26e-04
2^{16}	2.99e-05	1.73e-05	2.92e-05	6.15e-05
2^{18}	1.29e-04	2.09e-05	3.04e-05	3.68e-05
2^{20}	2.07e-05	2.99e-06	1.31e-05	1.21e-05

TABLE VII
RELATIVE ERROR FOR THE EVALUATION OF THE TOTAL VARIANCE $\mathbf{D} \approx 0.0025$.

# of samples	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
n	Relative error	Relative error	Relative error	Relative error
2^{10}	5.30e-02	5.01e-02	3.86e-03	1.04e-01
2^{12}	1.70e-02	2.90e-03	6.63e-02	4.79e-02
2^{14}	1.47e-02	1.41e-02	2.48e-02	3.25e-02
2^{16}	5.81e-03	2.91e-03	7.13e-03	1.51e-02
2^{18}	6.91e-03	1.73e-03	7.22e-04	4.72e-03
2^{20}	2.30e-03	2.84e-04	9.33e-05	1.07e-03

and # $27, 28$ (time independent). The simplified chemical equations of those reactions are:

$$[\#\mathbf{1}] \quad NO_2 + h\nu \Longrightarrow NO + O;$$
$$[\#\mathbf{3}] \quad O_3 + NO \Longrightarrow NO_2;$$
$$[\#\mathbf{7}] \quad NO_2 + O_3 \Longrightarrow NO_3;$$

$$[\#\mathbf{22}] \quad HO_2 + NO \Longrightarrow OH + NO_2;$$
$$[\#\mathbf{27}] \quad HO_2 + HO_2 \Longrightarrow H_2O_2;$$
$$[\#\mathbf{28}] \quad OH + CO \Longrightarrow HO_2.$$

The relative error estimation for the quantities f_0, the total variance $\mathbf{D}$ and some sensitivity indices are given in Tables VI, VII, VIII, IX, X respectively. The quantity f_0 is presented by 6-dimensional integral, while the rest are presented by 12-dimensional integrals. Table X is similar to Table IX and Table VIII, with the only difference – the increased number of samples $n = 2^{20}$ (instead of $n = 2^{16}$ in Table IX and $n = 2^{10}$ in Table VIII). In general, this increases the accuracy of the estimated quantities. Exceptions are S_5, S_5^{tot} and S_{15}, which have extremely small reference values. All LHS methods produce similar results, but for in small in value sensitivity indices LHSE-1 sometimes has the edge - see for example the value of S_{45} in Table X and sometimes LHSE1 gives the best results - see for example the value of S_5 and S_4^{tot} in Table X.

V. CONCLUSION

The present study focuses on the so-called environmental safety. The computational efficiency (in terms of relative error and computational time) of the optimization versions of Latin Hypercube Sampling Random and Edge algorithms with different seeds for multidimensional numerical integration have been studied to analyze the sensitivity of UNI-DEM model output to variation of input emissions of the anthropogenic pollutants and of rates of several chemical reactions. The various optimization versions of Latin Hypercube Sampling techniques have been successfully applied to compute global Sobol sensitivity measures corresponding to the influence of several input parameters on the concentrations of important air pollutants. The novelty of the proposed approaches is that Latin Hypercube Sampling Edge algorithm with different seeds have been applied for the first time to sensitivity studies of the particular air pollution model. The numerical tests show that the presented stochastic approaches is efficient for the multidimensional integrals under consideration and especially for computing small by value sensitivity indices.

TABLE VIII
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS ($n \approx 2^{10}$).

Est. qnt.	Ref. val.	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
S_1	4e-01	8.83e-03	2.40e-03	1.35e-01	1.41e-01
S_2	3e-01	1.68e-01	3.04e-02	6.03e-02	1.17e-01
S_3	5e-02	1.30e-01	4.19e-01	2.52e-01	2.20e-02
S_4	3e-01	1.65e-01	1.76e-02	7.06e-02	8.60e-02
S_5	4e-07	3.81e+03	5.17e+03	3.10e+03	6.53e+03
S_6	2e-02	5.97e-01	1.13e+00	4.29e-01	1.82e-01
S_1^{tot}	4e-01	6.77e-02	5.56e-02	1.04e-02	1.82e-02
S_2^{tot}	3e-01	2.39e-01	2.50e+00	1.47e-01	1.80e-01
S_3^{tot}	5e-02	3.74e-02	1.36e-01	3.18e-01	4.92e-01
S_4^{tot}	3e-01	2.40e-01	1.17e-02	4.71e-02	1.10e-01
S_5^{tot}	2e-04	4.79e+00	6.65e+00	8.59e+00	2.65e+01
S_6^{tot}	2e-02	3.72e-01	5.90e-01	7.67e-01	3.64e-02
S_{12}	6e-03	2.97e+00	1.48e+00	9.64e+00	5.18e+00
S_{14}	5e-03	4.97e+00	6.74e-01	2.61e+00	5.77e-01
S_{15}	8e-06	8.60e+02	9.12e+02	1.00e+03	8.17e+02
S_{24}	3e-03	8.40e-02	3.58e+00	3.49e+00	2.72e+00
S_{45}	1e-05	1.33e+01	2.64e+01	1.12e+02	9.94e+01

TABLE IX
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS ($n \approx 2^{16}$).

Est. qnt.	Ref. val.	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
S_1	4e-01	3.19e-03	5.38e-03	9.94e-03	1.49e-02
S_2	3e-01	2.23e-02	2.13e-02	4.19e-03	2.34e-03
S_3	5e-02	8.71e-02	6.84e-02	1.24e-02	1.58e-02
S_4	3e-01	5.16e-03	5.60e-03	2.14e-02	1.62e-02
S_5	4e-07	1.02e+03	1.01e+03	2.37e+01	2.29e+02
S_6	2e-02	7.27e-02	4.22e-02	1.71e-02	1.03e-01
S_1^{tot}	4e-01	7.30e-03	3.85e-03	1.88e-02	9.89e-03
S_2^{tot}	3e-01	1.05e-02	1.73e-02	1.10e-02	4.60e-04
S_3^{tot}	5e-02	1.00e-01	1.07e-01	2.71e-03	6.85e-03
S_4^{tot}	3e-01	6.82e-03	1.57e-02	5.89e-03	8.41e-04
S_5^{tot}	2e-04	5.13e-01	1.21e+00	9.24e-01	1.61e+00
S_6^{tot}	2e-02	2.59e-02	1.01e-01	4.23e-02	1.82e-01
S_{12}	6e-03	2.75e-02	1.42e-01	5.70e-01	2.68e-01
S_{14}	5e-03	2.35e-02	1.10e-01	8.29e-01	1.29e+00
S_{15}	8e-06	9.25e+02	9.33e+02	9.06e+02	9.25e+02
S_{24}	3e-03	3.52e-01	6.26e-02	1.53e-01	5.18e-01
S_{45}	1e-05	2.55e+00	3.88e+00	2.29e+00	4.13e+00

TABLE X
RELATIVE ERROR FOR ESTIMATION OF SENSITIVITY INDICES OF INPUT PARAMETERS ($n \approx 2^{20}$).

Est. qnt.	Ref. val.	**LHSE1**	**LHSE-1**	**LHSR1**	**LHSR-1**
S_1	4e-01	7.83e-04	7.26e-05	6.12e-03	2.73e-03
S_2	3e-01	4.62e-03	3.65e-03	1.70e-03	2.91e-03
S_3	5e-02	6.86e-03	5.05e-03	7.73e-04	7.33e-03
S_4	3e-01	2.98e-04	2.65e-03	2.46e-03	2.21e-03
S_5	4e-07	9.28e+00	1.28e+02	1.53e+02	3.06e+02
S_6	2e-02	5.68e-03	6.60e-03	2.54e-02	1.09e-02
S_1^{tot}	4e-01	1.23e-03	1.23e-03	2.39e-03	2.37e-03
S_2^{tot}	3e-01	2.25e-03	3.68e-03	7.43e-03	4.91e-03
S_3^{tot}	5e-02	1.39e-02	8.33e-03	1.21e-02	1.16e-02
S_4^{tot}	3e-01	2.91e-04	3.47e-03	2.88e-03	3.23e-03
S_5^{tot}	2e-04	4.68e-01	1.12e+00	2.68e-01	3.52e-01
S_6^{tot}	2e-02	2.33e-02	6.06e-03	4.22e-02	9.54e-03
S_{12}	6e-03	3.06e-01	1.22e-01	3.05e-01	1.39e-03
S_{14}	5e-03	9.18e-02	5.75e-02	9.69e-02	6.95e-02
S_{15}	8e-06	9.31e+02	9.32e+02	9.29e+02	9.27e+02
S_{24}	3e-03	7.99e-02	3.96e-01	3.16e-02	2.87e-01
S_{45}	1e-05	2.01e+00	4.83e-01	1.96e+00	2.41e+00

REFERENCES

[1] G. Dimitriu: Global Sensitivity Analysis for a Chronic Myelogenous Leukemia Model: Proc. 9th International Conference NMA'2018, Borovets, Bulgaria, August 20-24, 2018, LNCS 11189, Springer, Jan 2019. DOI: 10.1007/978-3-030-10692-8_42

[2] Gocheva-Ilieva, Snezhana G., Atanas V. Ivanov, and Ioannis E. Livieris. "High Performance Machine Learning Models of Large Scale Air Pollution Data in Urban Area." Cybernetics and Information Technologies 20.6 (2020): 49-60.

[3] Gocheva-Ilieva, S. G., Voynikova, D. S., Stoimenova, M. P., Ivanov, A. V., & Iliev, I. P. (2019). Regression trees modeling of time series for air pollution analysis and forecasting. Neural Computing and Applications, 31(12), 9023-9039.

[4] H. Hamdad, Ch. Pézerat, B. Gauvreau, Ch. Locqueteau, Y. Denoual, Sensitivity analysis and propagation of uncertainty for the simulation of vehicle pass-by noise, Applied Acoustics Vol. 149, Elsevier, pp. 85-98 (June 2019). DOI: 10.1016/j.apacoust.2019.01.026

[5] Kroese, D.P., Taimre, T., Botev, Z.: Handbook of Monte Carlo Methods, Wiley Series in Probability and Statistics, (2011)

[6] McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2), 239–45 (1979)

[7] Minasny B., McBratney B.: A conditioned Latin hypercube method for sampling in the presence of ancillary information Journal Computers and Geosciences archive, Volume 32 Issue 9, November, 2006, Pages 1378-1388.

[8] Minasny B., McBratney B.: Conditioned Latin Hypercube Sampling for Calibrating Soil Sensor Data to Soil Properties, Chapter: Proximal Soil Sensing, Progress in Soil Science, pp. 111-119, 2010.

[9] Pencheva, Velizara, Ivan Georgiev, and Asen Asenov. "Evaluation of passenger waiting time in public transport by using the Monte Carlo method." AIP Conference Proceedings. Vol. 2321. No. 1. AIP Publishing LLC, 2021.

[10] I. M. Sobol', Sensitivity estimates for nonlinear mathematical models, *Matem. Modelirovanie* **2** (1) (1990), 112–118.

[11] S. L. Zaharieva, I. Radoslavov Georgiev, A. N. Borodzhieva and V. Angelov Mutkov, "Classical Approach For Forecasting Temperature In Residential Premises Part 1," 2021 20th International Symposium Infoteh-Jahorina (infoteh), 2021, pp. 1-6.

[12] Z. Zlatev, Computer treatment of large air pollution models, KLUWER Academic Publishers, Dorsrecht-Boston-London, 1995.

[13] Z. Zlatev, I. T. Dimov, Computational and Numerical Challenges in Environmental Modelling, Elsevier, Amsterdam, 2006.

[14] Z. Zlatev, I. Dimov, K. Georgiev, Three-dimensional version of the Danish Eulerian Model, *Zeitschrift für Angewandte Mathematik und Mechanik*, **76** (1996) S4, 473-476.