Iterative Inverse Modelling Method for Locating a Source of Air Pollution and Its Robustness

B. Rajkovic

M. Vujadinovic

Z. Grsic

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Iterative Inverse Modelling Method for Locating a Source of Air Pollution and Its Robustness

B. Rajkovic[a], M. Vujadinovic[a] and Z. Grsic[b]

[a] *Institute of Meteorology, Faculty of Physics, Belgrade University, Dobracina 16, 11000 Belgrade, Serbia (bora@ff.bg.ac.yu)*
[b] *Institute of Nuclear Sciences Vinca, P.O. Box 522, 11001 Belgrade, Serbia (grsa@vin.bg.ac.yu)*

Abstract: Detection of a possible source of air pollution as a combination of measurements and inverse modelling, based on Bayesian statistics, has been proposed. The simplicity of the approach and its numerical efficiency qualifies this approach for the problem, especially when it is used in the operational mode. The method has been examined in its simplest form, with a single source and the implicit assumption that we know the moment of the release. The position of the possible source has been found as the maximum of the probability density function from an ensemble of possible sources that cover substantial part of the model domain. Members of the ensemble have been generated using a puff model. Search for the position consists of series of iterations converging toward the position of the source.

Keywords: point source; puff model; Bayesian statistics; probability density function.

1. INTRODUCTION

Detection of air pollution is a multi aspect problem. The first step is to design an optimal network of measuring points (stations), relative to the envisaged or possible sources of pollution. Then, based on measurements, quick and, as accurate as possible, detection of an accidental release should follow. It is important to emphasize that only measurements are not sufficient to create (analyze) concentration field, even if we have quite large number of sampling points. The main reason is that, generally, concentration fields have large gradients. We show this in Appendix, through series of idealized experiments. The only way to overcome this situation is by introducing Bayesian statistical method, as a component of the solution [Fuentes and Raftery 2001, Wikle et al. 2001, among others]. Its additional advantage is taking explicitly into account the measurement errors which are unavoidable and depended on the methodology, as well as on the quality of the instruments that are involved in the process.

Beside these, relatively straightforward ways of estimating quality of information from small number of monitoring points, there are several other approaches, although of higher sophistication. One is to use the spatial covariance structure of the quality patterns from the so called "sphere of influence" around each site as in Langstaff et al. [1987], Buell [1975] and Van Egdmond and Onderdielinden [1981], among others. Another approach is to generate data in a relatively large set of points (dense grid) and then analyze average concentration, its clustering, frequency and clustering of average concentration, as in Tseng and Chang [2001]. In this paper, we will concentrate only on the detection of possible source position, as an aspect of the inverse problem, assuming that there exist measurements of the concentrations of a passive substance.

2. MODEL

The central point is to calculate probability density function (PDF) as a most complete description of the system. It is defined as a conjunction of three spaces, space of possible values of a parameter (m), space of simulated measurements (d_m) and space of actual measurements (d). Using Bayesian statistics we combine then into a single PDF

$$p(m,d,d_m) = const \cdot \frac{\rho(m,d) \cdot f(m,d_m)}{\nu(m,d)} \quad , \qquad (1)$$

where ρ and f are PDFs defined on Cartesian product of parameters and real measurements space and parameters and simulated measurements space, respectively. *Const* is the normalization constant

$$const = \int_V \frac{\rho(m,d) \cdot f(m,d_m)}{\nu(m,d)} dV \quad , \qquad (2)$$

where $\boldsymbol{dV} = \boldsymbol{dm} \cdot \boldsymbol{dd} \cdot \boldsymbol{dd_m}$. Function $\nu(m,d)$ represents the homogeneous state of information. Its particular form doesn't play a role, except in some highly degenerated problems that are not considered here [Tarantola, 2005]. After taking into account that set of actual and simulated measurements are independent, we get the final form for PDF as

$$p(m,d,d_m) = const \cdot \frac{\rho_M(m) \cdot \rho_D(d) \cdot f(d_m / m)}{\nu_D(d)} \quad . \qquad (3)$$

The model that we used to construct probability density function of the possible source is a puff model [Grsic and Milutinovic 2000, Rajkovic et al. 2008]. The model starts with measured (10m) winds and extrapolate it to the height of each puff's center. In several previous simulations, measured winds were weaker than 1m/s, but results were still quite reasonable. The model has capability to calculate the rise of a warm gas to its neutral buoyancy height, but in these runs, we assumed that the gas has the ambient temperature. Vertical mixing is calculated using the Pasquill-Gifford stability classes. The effects of the relief are included according to the stability class. In the case of A, B, C and D class, a puff ascends over an obstacle with the same height as in front of the obstacle. In the case of stable stratification (E, F), a puff moves up the obstacle at fixed height of 10m. At the moment, the model does not estimate the stochastic part of the motion of particles. The grid had 301x301 points with spatial distance of 60 meters thus spanning the area of about 40 square kilometers. The time interval between two consecutive puffs was one minute.

3. METHODOLOGY AND RESULTS

Relative position of the assumed source and measuring point are presented in the Figure 1, upper panel. The first step was to generate field of a passive substance by its release during 120 minutes. Values at the measurement points were then randomly perturbed by 5%, mimicking the measurement errors. The second step was to create the cluster of points, possible sources, with center positioned in the measuring point with the highest concentration observed. From the cluster points we calculated PDF of the source position, assuming its Gaussian shape [Tarantola, 2005]. In the second iteration, we translated the center of a cluster to the position of the first PDF's maximum. Again, we calculated PDF and its maximum and translated the center to the position of the new maximum. Since now the cluster member with the maximum PDF was at the inside area of the previous cluster, we decided also to halve the spread among the third cluster's members. In the next two iterations we again translated the cluster and halved the distance between its members.

The error in the calculated distance after these five iterations was about 120 meters (two grid points). The whole procedure takes 12 minutes on a comparatively new personal computer thus clearly fulfilling the efficiency requirement. The accuracy is also acceptable. After this, we increased margin of errors from 5% to 30% and increased variability of the Gaussian distribution. In both cases convergence was very close to the first case. In the case of the increase of margin of errors, found position was one grid point away from the original margin of errors. In the case of σ, the result was the same as in the standard case.

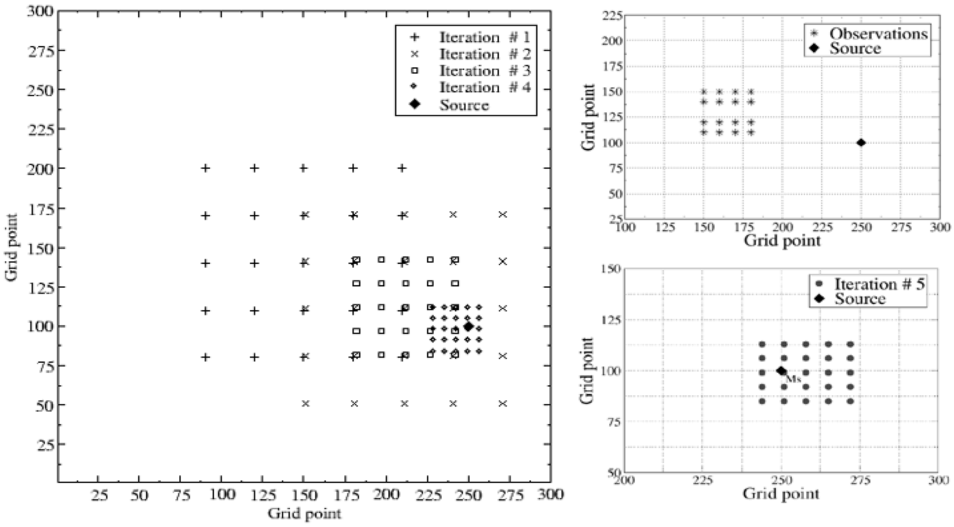

Figure 1. The right upper panel shows relative positions of the source and observation points. The left panel presents changes of the cluster position during four iterations, while the position of the final, fifth, cluster is shown on the right lower panel, where Ms denotes modelled position of the source

4. CONCLUSIONS

Combination of simulated measurements and Bayesian statistical approach through several iterations results in a very accurate yet efficient method of detecting a point like source of pollution. Sensitivity to the assumed margin of errors and σ, Gaussian distribution parameter, was quite small.

5. APPENDIX

Here we will show that even high number of measurements can not solve the problem of finding source(s) position. Let us generate concentration field with a release of seventeen sources. The geometry resembles the situation in the petrochemical zone of city Pancevo nearby Belgrade. In Figure A1, on the right panels, is a field of 24 hour-average concentration field obtained with a Gaussian plum diffusion model [Rajkovic at all, 2008] for the continuous source, using 10 minutes averages of the wind field and stability parameters with Pasquill-Gifford stability categories. The only deviation from the actual situation is the assumption that all sources are at the same height. The grid used for calculations was very dense one, with 90601 points (301x301). The idea is that such high density grid will reduce problems with the accuracy of the model and thus calculated field can be regarded as an “observed” one, i.e. we regard values in grid points as the measured ones. Formation of the isolines (graphical representation of the field) was done using the Kriging method. In this case of very dense grid we can clearly see positions of all sources as the positions of the local maximums.

If we now reduce the number of grid points, and therefore number of “sampling” points, by two orders of magnitude (31x31), we get situation depicted in the first row of Figure A1, left panel. We use again Kriging method to form the “continuous” field. Comparison with the starting situation, right panel of the same figure row, shows that we still have very high resemblance with the original field. This situation is still far from a realistic one regarding possible density of actual sampling points and tells us more about the possibility of reduction of computation effort.

Grid with 36 points (6x6) is closer to the regulatory recommendations and economical feasibility. Again, with the help of Kriging method, we arrive to the result in Figure A1, the second row. Same as in the first row, on the left we have the original concentration filed and on the right is the result from 36 points. Dots on the left panel are positions of the "sampling" locations. We now see that the distribution field is very different from the starting one, showing only the gross characteristics of the pollution concentration fields, but with complete loss of its local details. This still may be acceptable as a very rough assessment of the long term influences of pollution sources for that area, but totally insufficient in an accident situation.

Finally, we have created irregular set of "sampling" stations that qualitative represents the actual situation in that area. As it is often in the real life situations, several points are very close to each other, while others are relatively far away (Figure A1, the third row, left panel). Using the same analysis method, we get the distribution shown in Figure A1, the third row, right panel. The general structure is similar to the previous case with a few exceptions. This distribution is somewhat closer than the one with 36 points, due to the fact that several points "samplings" are very close to the position of the sources and thus accidentally giving better result. It is also possible that this particular wind direction has "helped" in this case.

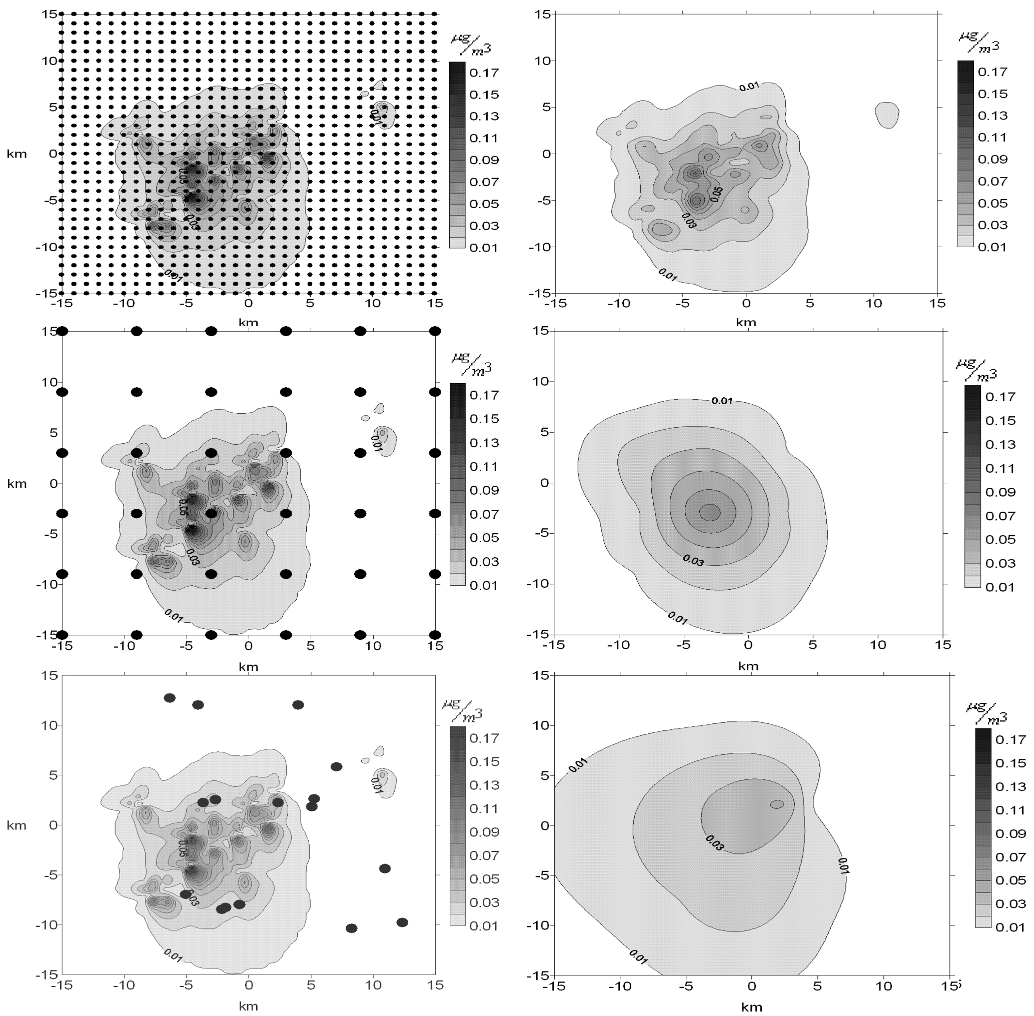

Figure A1. The right panel in every row shows the same concentration field, obtained from a release of 17 sources using a Gaussian plum model with very dense grid (301x301 grid point). Black dots on each right panel represent "sampling" points. In the first row there is 31x31 "sampling" point, in the second row 6x6, and in the third row, there is 16 irregularly spaced "sampling points". Left panels show concentration fields interpolated, using Kriging method, from "sampling" points on the right.

ACKNOWLEDGMENTS

This work has been partially funded by Serbian Ministry of Science, Technology and Environment, grant no. 1197 and Italian Ministry of Environmental and Territories through its two projects, SINTA and ADRICOSM-STAR. The third author is partially funded through the project of the decommission of the nuclear reactor in Vinca.

REFERENCES

Buell, C.E., The topography of the empirical orthogonal functions. *4th Conference on Probability and statistics,* American Meteorological Society, Boston, 188-193, 1975.

Grsic, Z., and P. Milutinovic, Automated meteorological station and the appropriate software for air pollution distribution assessment, *Air Pollution Modelling and Its Application XIII*, 2000.

Fuentes, M. and A. Raftery, Model evaluation and spatial interpolation by combining observations with outputs from numerical models via Bayesian Melding, *Technical report,* No 403, 2001.

Langstaff, J., Seigneur, C., Liu, M.K., Behar, J.V. and J.L. McElroy, Design of an optimum air monitoring network for exposure assessments, *Atmospheric Environment*, 21, 1393-1410, 1987.

Rajkovic, B., Arsenic, I. and Z. Grsic, *Fluid Mechanics of Environmental Interfaces*, Taylor and Francis, 348 pp., London, 2008.

Tarantola, A., *Inverse problem theory and methods for model parameter estimation*, SIAM, 342 pp., Philadelphia, 2005.

Tseng, C.C and N.B. Chang, Assessing relocation strategies of urban air quality monitoring stations by GA-based compromise programming, *Environment International*, 26 (7), 523-541, 2001.

Van Egmond, N.D. and D. Onderdelinden, Objective analysis of air pollution monitoring network data; spatial interpolation and network density, *Atmospheric Environment*, 15, 1035-1046, 1981.

Wikle, C.K., Milliff, R.F., Nychka, D. and L.M. Berliner, Spatiotemporal hierarchical Bayesian modelling: Tropical ocean surface winds, *Journal of the American Statistical Association*, 96(454), 382-397, 2001.