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

Locating a Source of Air Pollution Using Inverse Modelling and Pre-computed Scenarios

M. Vujadinovic[a], B. Rajkovic[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: An extension of the transition of a possible source of air pollution, as a combination of measurements and inverse modelling, based on Bayesian statistics, has been proposed. The method consists of two steps. The first one is creating a library of possible scenarios, where each scenario includes meteorological parameters, parameters related to the emission and calculated concentrations at measurement points. Once the library is formed, we calculate marginal probability function for the parameter that we would like to estimate. In the most cases parameters under consideration would be those related to the emission its position, intensity or duration. The parameter that we did estimation for was the position of the possible source, creating simulated measurements that had incorporated possible measuring errors. This, the second part of calculation, is extremely efficient and fast. The simplicity of the approach and its numerical efficiency qualifies it for the problem, especially in the operational mode. Members of the library and simulated measurements were generated using a puff model.

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

1. INTRODUCTION

At the first sight, we can infer parameters of the pollution field from measurements obtained from a network of measuring stations, capable to detect presence of pollutant. If we wanted to make quantitative assessment we would need measurements of the amount of pollutant i.e. its concentrations in several (the more the better) stations. Further examinations for the case of localized sources and pollutant released in the atmosphere is not really possible, except in some special cases. Two, the most important, reasons are turbulent nature of the motions in the Planetary Boundary Layer (PBL) and very large gradients that are characteristic for the concentration fields. The large gradients are the primary reason why it is difficult to synthesize (analyze in meteorology) concentration filed only from measurements and therefore it is impossible to localize source(s) of specific pollution field. For more detailed formulation, please see the accompanied paper by Rajkovic et al., in this conference.

Still inverse modeling, search for the source, position or strength of a passive substance, must have, as its starting point, at least detection of the presence of a pollutant. Measurements of the amount of a pollutant are even better. So, we may say that one aspect of inverse modeling could be design of an optimal network of measuring points, stations, relative to the envisioned sources of pollution.
Problem of the source detection can be approached using transport equation in its Eulerian or Lagrangian form. In both cases, we have 3D domain with lateral boundaries, bottom and top. Initialization and formation of boundary conditions may be quite complicated problem. On the other end of modeling spectrum are Gaussian plum models as the “bulkiest” ones.

Wind is treated as a known constant, while thermodynamics is heavily parameterized through stability classes [Pasquill and Smith 1983, Melli and Runca, 1979, Mocioaca and Stefan 2003]. The thermodynamics can be improved using Monin-Obukhov theory or even second-order closure theories [Sykes and Gabruk, 1997]. The most problematic part in Gaussian plum approach is constant wind assumption. The next generation of "bulk" models avoids this simplification introducing series of puffs, whose superposition gives the concentration field. But, even using a puff model, as relatively simple and efficient, to perform direct integrations, covering all possible variations of the atmosphere state and emission properties, is not a practical approach.

The most efficient and complete (accurate) approach is a combination of direct modeling and inverse techniques with the Bayesian probabilistic approach. The main point is to construct probability density function (PDF) of the problem and then, from its marginal distribution, calculate the value of considered parameter. The Bayesian method [West and Harrison 1997, Bernardo and Smith 1994, Gelman et al. 2004] combines observed data and possible prior knowledge, in order to calculate value of unknown model parameters. Such knowledge is inherently probabilistic and therefore more complete and closer to the real world, since, in principal, it can take into account all the uncertainties in the measurements and modeling. The main problem is an error distribution's shapes. Now, the most of them are assumed to be Gaussian.

Calculating PDF can be done in several ways, like Markov Chain Monte Carlo method, or better suited sequential Monte Carlo approach [Gliks et al. 1996, Liu 2001, Andrieuet et al. 2003, Doucet et al. 2001, Arulampalam et al. 2002]. We propose a much simpler version, where we create a library of possible scenarios. Each scenario includes meteorological parameters, parameters related to the emission and calculated concentrations at measuring points. Once the library has been created and stored, it is very easy to generate probability density function of the problem and estimate any desired parameter from the corresponding marginal distribution.

2. MODEL, RESULTS AND ITS ANALYSIS

2.1 The Probabilistic Approach

Conjunction of probability is the most complete description of a physical system. It has, as its building blocks, probability functions defined on three sets: the parameter space, simulated measurements space and actual measurements space. For more detailed formulation see accompanying paper Rajkovic et al.

2.2 The Model and Source Localization Procedure

The model that we used to construct the library was a puff model [Grsic and Milutinovic 2000, Rajkovic et al. 2008]. The grid had 301x301 points with spatial distance of 150 meters. The time interval between two consecutive puffs was one minute. For further details about the model, we refer to the accompanying paper Rajkovic et al.

The methodology has two phases. The first phase is generating a set of many scenarios (realizations, in the probabilistic terminology) which we will call the library. These scenarios are runs of the puff model with variations of the problem parameters. There are two groups of parameters. The first one is consisting of parameters that define the source of the pollution, source's strength, its position and the duration of the release. In the second group are meteorological parameters: wind intensity, its direction and stability characterization of the PBL through Pasquill-Gifford categories. Using one combination of parameter's values, model produces concentrations in all measuring points. In our case, there were 13 measuring and 3 possible source points. The ranges of all parameters should be wide enough to cover all reasonable possibilities in the problem.

The computation of PDF, that characterizes the problem, is often done by some variation of the Monte Carlo method. But, if we really want to cover substantial part of possible ranges of all parameters using MC method, the computational effort will be extremely big. Instead, we have "digitized" values of various parameters into classes, appropriate for each parameter. For the wind intensity, we propose logarithmic scale, with winds of 0.1, 0.2, 0.5, 1, 2, 5, 10, 20 and 50 m/s, so that we certainly cover all possible situations. The stability classes are anyway quantized having values from 1 to 6. During testing of our method, we have found a very large sensitivity to the wind direction. Considering this, we propose wind direction interval of 1 degree, that is quite different from either 1/8 or even 1/16 of the full circle, as is standard way of characterizing wind direction. For the source locations we take positions of all possible sources in the particular configuration. For the source strength we take the exponential increment from 1 to 10^6 of the appropriate units. Finally, we have assumed that accident's duration is between 1 minute and 3 hours, which should be enough to cover realistic accidental release that can be treated using either Gaussian plum or puff models. Duration also depends on the size of the domain and relative positions of the sensors and possible pollution sources. At the end, we note that all the above values should be reexamined in each concrete case. The above choices are viewed as typical rather then all possible. This, first, phase can be quite computing expensive but it is done only once for a particular application. Please note that if we change the model or any of its components or parameterizations, the re-computing of the library is necessary.

The second phase can be conditionally called "operational". To carry it out, we must have meteorological observations of wind, its intensity and direction, decision which stability category is relevant for that moment (we do not want to discuss how this should be done, we only note that it can be done in many ways, in a simple manner or with some sub-model of small or quite high sophistication) and measured concentrations at measuring points. Starting from the implicit assumption, that deviation of the "true" values of all parameters are Gaussian, we form PDF as

$$\rho = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[\frac{(Qobs - Qlib)^2}{\sigma^2}\right], \tag{1}$$

for all members of the library whose meteorological parameters have been observed. The standard deviation (σ) measures the spread between observations (Q_{obs}) and values from the library (Q_{lib}). In general, it takes into account uncertainty in our a priori knowledge, uncertainties in measurements (σ_D) and modelling (σ_T). In this case, there is no a priori knowledge, so total σ is the sum of σ_D and σ_T. We have assumed that measurements errors were in the range of 5%, which together with values of simulated measurements gives an estimate for σ_D of 10^{-4}. On the other hand, we took total σ equal to $5 \cdot 10^{-2}$, which is two orders of magnitude larger than σ_D. This should take into account all modelling errors. Since wind intensity does not vary significantly, we could consider its average value during the accident. But, due to very high sensitivity of the model to wind direction, we should take the whole range of the observed directions. Once we have computed PDFs, its summation over all but desired parameter give us marginal probability for that parameter and its maximum gives the best estimate of it.

2.3 Results

Relative positions of the assumed source and measurement points are presented in the Figure 1.. Black dots are measurement points (13), the stars are sources (3) while the diamond is the source which we will look for. Instead of real measurements, we created simulated measurements using realistic (measured) winds, its intensity and direction, and temperatures at two meters height. These concentrations were then randomly perturbed by 5% mimicking the measurement errors. In most cases concentrations were "observed" at only two or three measuring points. The reason is that we have point source whose plume is quite narrow.

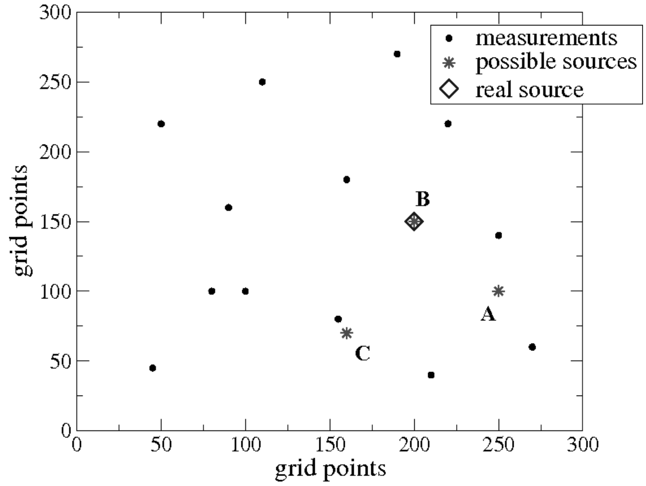

Figure 1. The position of measuring points (dots), possible sources A, B, C (stars) and real source (diamond).

As we said, the first step is library creation. From the meteorological data we knew that the mean wind intensity was about 6 m/s and stability had the value of 5 for the Pasquill-Gifford stability class, while the wind direction was in the range from 90 to 120 degrees. Based on these values, we were able to select the appropriate members of the library. We have also varied the duration of the accident in seven time intervals from 1 minute to 181 minutes, every 30 minutes. The source parameters were as explained in the previous section. In all three cases of possible sources we have found the correct one.

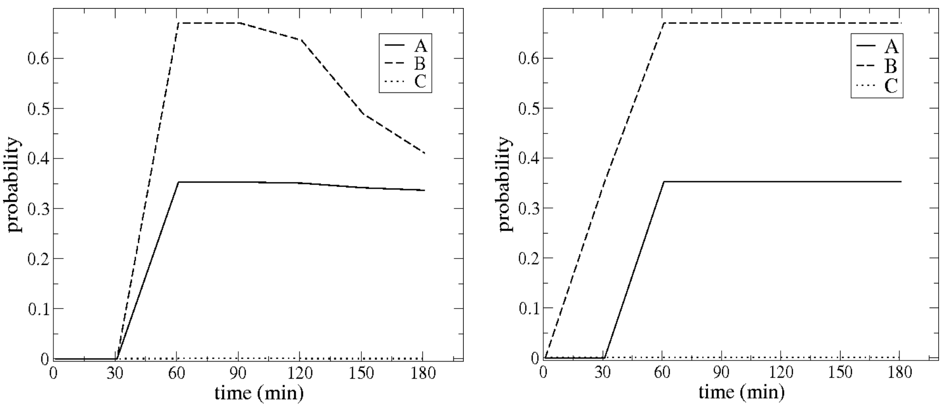

Figure 2. Marginal probability function for each of three sources (A, B, C) in 30 minutes intervals. On the left panel is its change in the case of measurements simulated using observed, while the right panel is for the case of measurements simulated using constant wind direction.

It is interesting to see how the probability changes with time for each of the source (Figure 2., left panel). Each source has different line style and the correct source (in this case it is source B) has the largest value, during entire integration. At the beginning, probability was very small, then reaches its maximum, thus indicating the duration time as well as the position, and then decreasing. The reason for this is that wind in the simulated measurements was realistic, changing its direction during the simulation, while the library records have constant wind direction. To check that, we redid our determination, but this time we kept the wind direction constant. Indeed, as shown in Figure 2. on the right panel, once the probability reaches its maximum, it stays constant.

3. CONCLUSIONS

Combination of pre-computed scenarios and simulated measurements, based on Bayesian statistical approach, results in an accurate and yet efficient method of detecting point like source of pollution. Time evolution of the problem indicates to the duration of the accident or how long ago, relative to the measurements, the accident occurred.

This methodology is independent of the choice of the air dispersion model, its complexity, parameterizations, etc.

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

Andrieue, C., da Freitas, N., Doucet, A. and M.I. Jordan, Introduction to MCMC for machine learning, *Machine Learning*, 50, 5-43, 2003.

Arulampalam, M.S., Maskell, S., Gordon, N. and T. Clapp, A tutorial on practice filters for on-line non-linear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing*, 50(2), 174-188, 2002.

Bernardo, J.M. and A.F.M. Smith, *Bayesian theory*, Wiley, Chichester, 1994.

Doucet, A., de Freitas, N. and N.J. Gordon, *Sequential Monte Carlo methods in practice*, Springer-Verlag, Berlin, 2001.

Gelman, A., Carlin, J.B., Stern, H.S. and D.B. Rubin, *Bayesian data analysis*, Chapman and Hall, 668 pp.,London, 2004.

Gliks, W.R., Richards, S. and D.J.E. Spieglhalter, *Markov Chain Monte Carlo in practice*, Chapman and Hall, London, 1996.

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.

Liu, J.S., *Monte Carlo strategies in scientific computing*, Springer-Verlag, 343 pp., Berlin, 2001.

Melli, P. and E. Runca, Gaussian plum model parameters for ground-level and elevated sources derived from the atmospheric diffusion equation in a neutral case, *Journal of Applied Meteorology*, 18(9), 1216-1221,1979.

Mocioaca G. and S. Stefan, Different parameterization of a Gaussian scheme: intercomparison study, *International Journal of Environment and Pollution*, 19(1), 32-45, 2003.

Pasquill, F. and F.B. Smith, *Atmospheric diffusion*, John Wiley, 437 pp., New York, 1983.

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

Sykes, R.I. and R.S. Gabruk, A second-order closure model for the effect of averaging time on turbulent plum dispersion, *Journal of Applied Meteorology*, 36(8), 1038-1045, 1997.

West, M. and J. Harrison, *Bayesian forecasting and dynamic models*, Springer, 680 pp., New York, 1997.