18th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 October 2017, Bologna, Italy

SOURCE TERM ESTIMATION USING AN ADJOINT MODEL: A COMPARISON OF TWO DIFFERENT ALGORITHMS

Gianni Tinarelli[1], *Giuseppe Carlino*[2] *and Francesco Uboldi*[1]

[1]ARIANET s.r.l., Milan, Italy
[2]Simularia s.r.l., Turin, Italy

Abstract: The location of an unknown source and its pollutant emissions are estimated from concentration observations by means of two approaches, both making use of the adjoint of a Lagrangian particle dispersion model. In the first approach, plausible source locations are estimated by identifying areas with maximum spatial and temporal consistency among backward trajectories from each sensor. In the second approach, a variational method is used to minimize the objective function and to estimate emissions at each grid-box: the resulting map provides information on the source location and on its uncertainty. In both approaches, zero measurements are also taken into account to define upstream exclusion zones where emission sources cannot be located. The two methods are compared using a hierarchy of test cases, starting from a controlled field experiment up to real world operational cases.

Key words: *STE algorithm, inverse dispersion modeling, adjoint model, Lagrangian particle model.*

INTRODUCTION

Source Term Estimation (STE) are algorithms designed and used to predict release location and characteristics, such as the time of release and amount of emitted material in all situations where the information about pollutant sources is substantially unknown. These algorithms are often based on the use of local concentration measurements given as input to dispersion models used in their 'adjoint' configuration. That means models in which time is reversed, building backward trajectories starting from monitoring measurements. In this respect, RetroSPRAY (Armand *et al.*, 2013) is the adjoint of the Lagrangian particle dispersion model SPRAY. The adjoint of a Lagrangian particle dispersion model enables computing the Source-Receptor Matrix (SRM: Seibert and Frank, 2004) by actually performing the same calculations of the forward model with reverse sign of advection (Flesch *et al.*, 1995) so that the integration is actually performed backward in time. Considering that any Lagrangian particle dispersion model, including SPRAY, is intrinsically stochastic, RetroSPRAY can only be an approximate adjoint of SPRAY.

Backward particle trajectories starting at locations and times of observed pollutant concentration peaks compose "retro-concentration" fields that define areas where the emitting source could be located, provided that a good estimate of meteorological fields, particularly the mean wind, is available. Similarly, retro-concentrations starting from locations and times of observations with zero values (or values compatible with a possible environmental background) can define exclusion areas and time intervals, where the pollutant source cannot be located. These concepts have been applied to develop two approaches to the problem of locating an unknown pollutant source and estimating its emission. In the first, simple to implement approach, the retro-concentration fields are combined to provide a map of possible source positions based on the concept of maximum overlap of the simulated retro-plumes, and emission are estimated by appropriately combining the obtained retro-emitted values in the identified source area. The second approach is a more complex variational method, using the simulated retro-puffs minimizing a mean-square observational residual (objective function) to estimate emitted quantities and emission times in a grid of possible source locations. The resulting map of minimized objective function values is then used to estimate the position of the source. A comparison of the two methods is made in different conditions, using the data of a field experiment specifically designed to test STE algorithms, specific synthetic test cases in a real framework in complex terrain and a real case in the same framework, showing advantages and limitations of each of the two approaches.

METHODS

In the first, simpler method, releases form each measurement station are independently considered in the adjoint Lagrangian dispersion model, computing retro-plumes and related retro-concentration fields during an identified period. Each computed field thus provides an independent estimate of the potential mass release at all upwind locations and at all the timeframes considered and depicts a potential zone of emission. In addition, adjoint fields computed from "zero measurements" or measurements below a threshold representing a background concentration value, are also taken into account since they define upstream exclusion zones. The choice of the threshold for zero–release emission terms is appropriately tuned in order to get optimal results. Given all the adjoint fields, a consistency map is built on the whole domain by taking into account the maximum overlap and the exclusion areas for all the time frames, allowing for the identification of the most plausible areas where the source terms are located. Emission rates estimates are performed by statistical analysis of the adjoint concentrations in the maximum consistency areas.

In the variational method, the forward (SPRAY) linear relation between emitted masses, **q**, and estimated concentrations (at station locations and observation times) $\boldsymbol{\chi}$ can be formally expressed as the application of a matrix **L** to the vector containing emitted masses as components: $\boldsymbol{\chi} = \mathbf{L}\,\mathbf{q}$. Apart from the intrinsic approximation due to the statistical nature of particle models, RetroSPRAY performs an application of the transpose matrix, $\mathbf{L}^{\mathrm{T}}$ through its integration backward in time from each observation locations and time, producing "retro-puffs". In this way, the **L** matrix components are directly and explicitly computed at each grid-box and at each possible emission time. Grid-boxes that are reached by retro-trajectories from significantly large observations can be considered as possible source locations. There, the emitted quantities **q** can be estimated by minimizing the square residual between estimated concentrations and observations. Variables **q** and $\boldsymbol{\chi}$ are converted to corresponding adimensional particle numbers, respectively **x** (emissions) and **y** (concentration estimates at observation times and locations). A threshold value x_{MIN} separates "zero" observations (and a possible environmental background) from large positive observations, actually used at estimation stage. To avoid estimating negative emissions, a composite (but differentiable) transformation Φ is used, logarithmic below x_{MIN} and linear above. The transformation is applied to observations: $\eta^{o}_{k,j} \equiv \Phi\left(y^{o}_{k,j}\right)$, estimated concentrations: $\eta_{k,j} \equiv \Phi\left(y_{k,j}\right)$ and emissions: $\xi_n \equiv \Phi\left(x_n\right)$. The minimized quantity is the objective function:

$$J(\boldsymbol{\xi}) = \frac{1}{2}\left[\boldsymbol{\eta}(\boldsymbol{\xi}) - \boldsymbol{\eta}^{o}\right]^{\mathrm{T}}\left[\boldsymbol{\eta}(\boldsymbol{\xi}) - \boldsymbol{\eta}^{o}\right] = MIN \qquad (1)$$

where the vector $\boldsymbol{\eta}$ depends on the vector $\boldsymbol{\xi}$ through the composition of three functions: 1) anti-transformation $x_n = \Phi^{-1}\left(\xi_n\right)$; 2) linear application $\mathbf{y} = \mathbf{A}\mathbf{x}$; 3) transformation $\eta_{k,j} = \Phi\left(y_{k,j}\right)$. Derivatives are analytically computed accordingly. Because of the logarithmic part of the transformation, the estimation is non-linear, then a non-linear minimization procedure (conjugate gradient) is used. Remark that, because of the linear part of the transformation, the value of *J* is effectively proportional to the mean square residual on "non-zero" (i. e. above x_{MIN}) observations.

"Zero" observations (below x_{MIN}) are used afterwards: the root-mean-square (RMS) residual on "zero" observations is computed and compared with a set of threshold values (application-dependent). This is done for all gridboxes where it is possible a minimization for *J*. Gridboxes that excessively over-estimate zero-observations, showing large (threshold exceeding) values of these RMS residuals, are discarded as possible source locations.

TEST CASES AND RESULTS

FFT07 experiment - Trial 54

The U.S. Department of Defense designed a specialized experiment, the Fusion Field Trial 2007 (FFT07) (Platt and Deriggi, 2012) to specifically provide a data base for use in developing and evaluating STE modeling systems. A series of experiment was set in Utah, U.S.A. in 2009 where a tracer (propylene) was released in prescribed quantities from a source located upstream of a network of (nominally) 100

concentration detectors set in a regular array, with flat orography. Among all the experiments available, we chose Trial 54: during this case the wind was quite intense from South-East, and the resulting plumes appear stretched along streamlines, with little lateral dispersion. The meteorological fields were reconstructed from the available observations by means of the diagnostic model SWIFT on a multilevel domain defined by a horizontal grid of 141 x 141 points with 10 m × 10 m spatial resolution and 1 min time resolution. The whole experiment lasts 12 minutes, from 14:15 to 14:27 of 2009/09/22, retro-concentrations are computed by RetroSpray at the time resolution of 1 min.

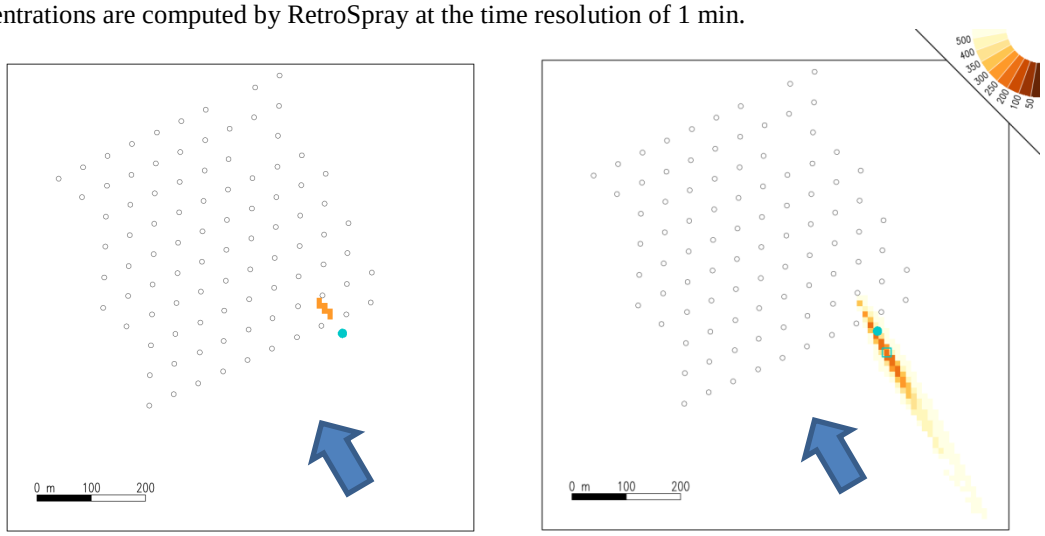

Figure 1. Left panel: possible source locations as estimated by the overlap method, map of maximum overlap region averaged during the entire period at the 95% level. Right panel: variational method, map of minimized objective function values after zero obs. exclusion. Black empty circles: observing stations, filled circle: true source location, empty square in right panel: best estimate of source location. Arrows represent the average incoming wind direction.

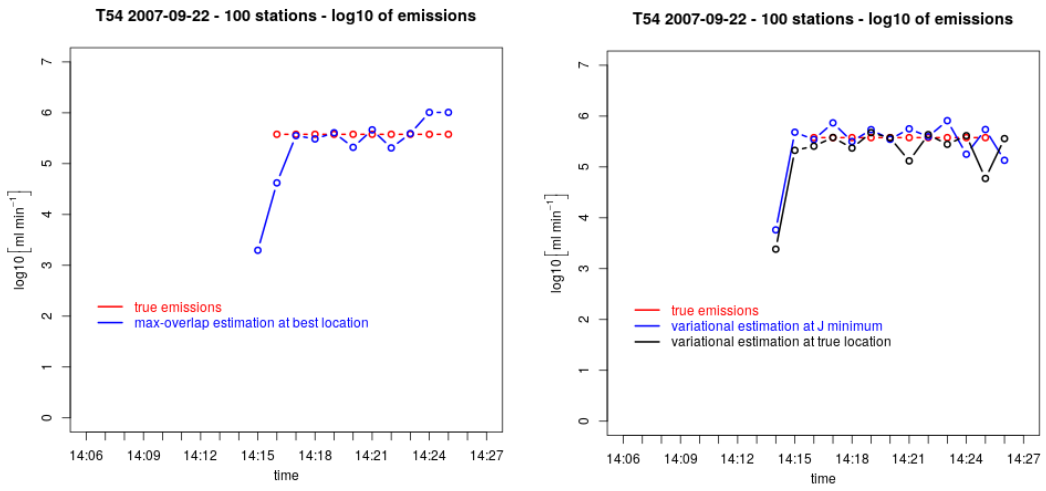

Figure 2. Base-10 logarithm of estimated emissions. Left panel: overlap method. Right panel: variational method. True emissions: red (both panels), total = 3.775×10^6 ml. Maximum overlap method: blue, left panel, total = 4.410×10^6 ml. Variational method: right panel; estimation at true source location: black, total = 3.606×10^6 ml; estimation at best location 5.405×10^6 ml.

Fig. 1 shows the estimated area of the source using the maximum overlap method (left) and the variational method (right). The max-overlap method estimates the source position few gridboxes downwind with respect to the real one. The variational methods includes the real source among the possible positions and locates its best estimate a couple of gridboxes upwind. Even in this abundantly observed situation, zero observations play an important role, defining useful exclusion areas and reducing

the uncertainty on the source location. Fig. 2 shows the time dependent estimation of emission rates: both methods show a good comparison with real data, the variational method a better homogeneity during the entire considered period.

Synthetic test cases

Both methods were also tested in the context of synthetic cases, built on a computational domain in complex orography corresponding to a coastal site in southern Italy where an industrial complex is present. Here, a situation closer to a real condition was taken into account for all the simulations, considering a 4-hour stationary emission event close to the ground, a time resolution for concentration data of 1h and an horizontal spatial resolution of 250 m. In these conditions, a set of cases considering different source positions, prescribed wind blowing from different incoming directions and some configurations of pseudo-station locations giving ground level concentration was built using the forward SPRAY model. Station locations were chosen to ensure the availability of both downstream and upstream (zero) observations with respect to the given source location. In all the considered cases, both the STE methods under test show to be robust, giving similar information, better for the variational method. These tests put also into evidence the importance of the spatial station distribution with respect to the wind direction and that of zero observations availability, especially in upstream areas, to exclude impossible source locations. Fig. 3 shows the source locations estimated by the two methods, maximum overlap on the left and variational method on the right in one of the considered cases. Both methods correctly identify the area containing the source position, the variational one with less spatial spread.

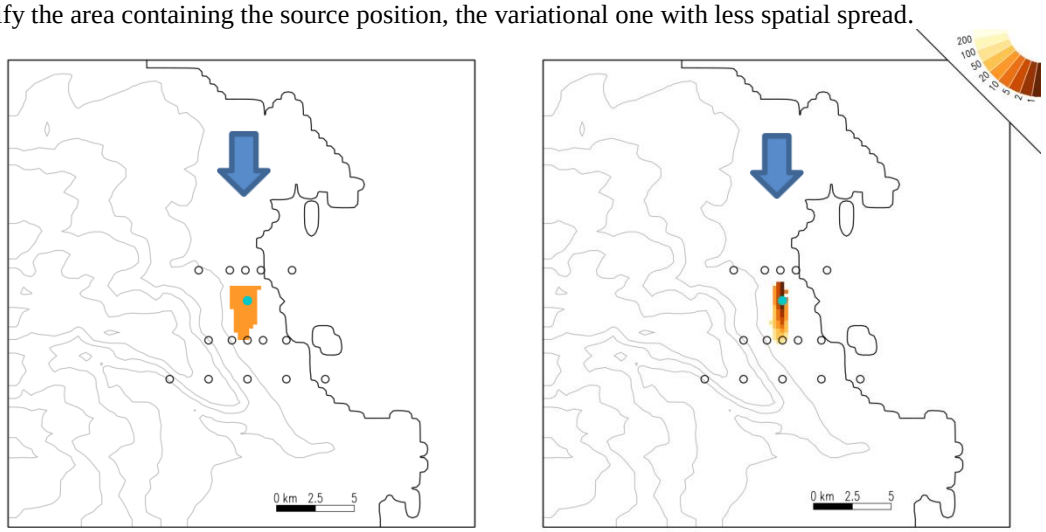

Figure 3. Left panel: possible source location area as estimated by the overlap method. Right panel: variational method, map of minimized objective function values after exclusions from zero obs. Filled circle: true source location, empty square: best estimate of source location. Arrows represent the incoming wind direction. Black empty circles mark observing stations.

Real world operational cases: industrial area

In the site considered for the synthetic test cases a large industrial complex is operational. Here, various industrial plants are located including an industrial harbor. A modeling system operationally runs: 1) reconstructing, on hourly basis, the meteorological conditions based on local measurements and 2) estimating the quality impact of the main sources present. This system also implements retroSPRAY with the maximum-overlap method to estimate a source reconstruction, to be activated by an operator in case of particular peak events involving an accidental or unknown release of substances having a potential odor or health impact (H2S, hydrocarbons). A local network of in-situ samplers measures hourly averaged concentrations of relevant chemical species. Station locations are shown in both panels of Fig. 4. The system is currently under test, to also verify the difference between the two STE methods and a possible operational implementation of the variational system.

The event illustrated showed a large peak of benzene on 2016/05/29. Observed concentrations ($\mu g\ m^{-3}$) are hourly averaged and emissions are estimated as $\mu g\ h^{-1}$. Wind intensity was rather weak and a wind

rotation due to a breeze cycle was present. Among the observing stations present in the area, 6 detected benzene. The most critical element influencing the estimation of source location and emissions are the definition of the wind field, generally more uncertain when the intensity is weak, and the stations spatial distribution with respect to the main wind direction. Fig. 4 shows the estimated position of the source by the two methods for the considered event. Again, the two methods broadly show to be consistent, the variational system have a larger significant area, that could correspond to the presence of ships located in front of the harbor.

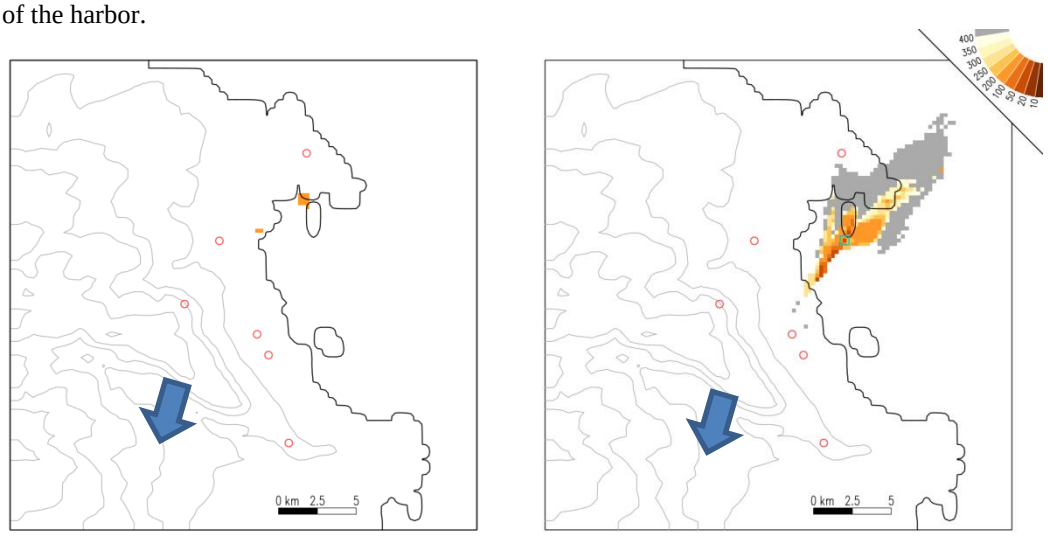

Figure 4. Real event of benzene peak on 2016/05/29. Left panel: possible source location as estimated by the overlap method. Right panel: variational method, map of minimized objective function values after exclusions from zero obs. Empty square: best estimate of source location. Red empty circles: locations of observing stations. Thin contour lines: orography; thick line: coastline. Arrows represent the recurring wind direction during the considered period.

CONCLUSIONS

Two STE (Source Term Estimation) algorithms, based on retroSPRAY, the adjoint of the Lagrangian particle dispersion model SPRAY, are tested on real and synthetic test cases and on real operational cases. The main critical information is the definition of the meteorological fields, in particular the wind field responsible for advection and transport. The spatial distribution of available observing stations can be critical depending on the main wind direction, and, as a consequence, on their ability to detect the main puffs emitted from the unknown source location. An important result of the study has been to highlight the role of "zero" observations: these mark positions and times that have not been reached from the pollutants, and resulted to be very important to exclude impossible, or very unlikely, source locations. When the observational information is sufficient and the wind uncertainty small, both method appear to be able to locate the source position with acceptable accuracy, even considering different spatial and temporal time scales. The variational method, in particular, showed to provide more accurate estimates of emitted quantities at the estimated source position (and at the true source position when this is known).

REFERENCES

Armand, P., C. Olry, A. Albergel, C. Duchenne and J. Moussafir, 2013, Development and application of Retro-Spray, a backward atmospheric transport and dispersion model at the regional and urban scale. Proceedings of the 15th Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes – 6-9 May 2013, Madrid, Spain, pp. 789-893.

Platt, N. and D. Deriggi, 2012: Comparative investigation of source term estimation algorithms using FUSION field trial 2007 (FFT-07) data: linear regression analysis. *Int. J. Environ. Pollut.*, **48**, 13-21.

Seibert, P. and A. Frank, 2004: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. *Atmos. Chem. Pyhs.*, **4**, 51-63.

Flesch, T.K., J.D. Wilson and E. Yee, 1995: Backward-time Lagrangian stochastic dispersion models and their application to estimate gaseous emissions. *J. Appl. Meteorol.*, **34**, 1320-1332.