A TRAJECTORY STATISTICAL METHOD FOR THE IDENTIFICATION OF SOURCES ASSOCIATED WITH CONCENTRATION PEAK EVENTS

Rita Cesari[1], Paolo Paradisi[2], Paolo Allegrini[3,4]

[1]ISAC-CNR sede di Lecce, Str. Prov. Lecce-Monteroni km 1200 , 73100 Lecce (Italy);
[2] ISTI-CNR, via G. Moruzzi 1, 56124 Pisa (Italy);
[3] IFC-CNR, Via Moruzzi 1, 56124 Pisa (Italy);
[4] Centro EXTREME, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 7, 56127 Pisa (Italy).

Abstract: We briefly review the Trajectory Statistical Methods (TSMs) most used in literature for source identification, essentially based on the concept of Residence Time. Then, we introduce a statistical methodology that, starting from the Concentration Field method, takes into account only the peak values in the concentration time series measured at multiple receptor sites. We use virtual simulations to evaluate the performance of our approach. In order to derive concentration time series at multiple receptors, the Lagrangian Dispersion Model (LSM) FLEXPART is used, in the time forward mode, to simulate dispersion from a known emission source. Then, virtual concentration data are available in the receptor sites. As in many TSMs, Residence Times need to be computed and, to this goal, we use FLEXPART, but in the backward mode, that is, FLEXPART is applied to compute the backward trajectories from the receptor sites. Then, our proposed statistical method is applied to the computed Residence Times and to the concentration data to reconstruct the spatial distribution of emission sources. The numerical results show that our approach could overcome the problem of ghost sources. Further, the proposed method requires simulation times shorter that those required in other methods, since it makes use of a relatively small set of trajectories. This could be of some interest in the characterization of impact studies and local climatic scenarios.

Key words: *Trajectories Statistical Methods, Residence Time Analysis, Peak events*

1. INTRODUCTION

The analysis of source regions and their contributions to receptors is an important tool in the development of effective control strategies (Lin, C.H. and L.F.W. Chang, 2002). To this aim, receptor modelling techniques have been extensively developed. Currently, they can be classified into two types. The first class is that of receptor models (Henry, R.C., 1997), focused on finding the chemical composition of the sources, while the second class is based on Air Trajectories Analysis (Kabashnikov, V.P. et al., 2001), focused on evaluating the probability of source locations. These two approaches are often used together. However, here we are interested in the efficiency of the second class of models.

The trajectory-based approach directly indicates the linkages between sources and receptors. In a time backward approach, it is possible to follow the motion of a polluted air parcel from receptor to source sites. Since the air parcel arriving at a site could have followed different trajectories, the use of multiple trajectories could provide a measure of the uncertainty inherent to air mass pathway, arising for example from assumption regarding vertical transport, turbulence, interpolation of meteorological data. In the Air Trajectories Analysis, the trajectories can be computed with both deterministic and stochastic models. When considering turbulent transport and the boundary layer, it is natural to use an approach based on Lagrangian Stochastic Models (LSMs), which are the underlying models used for the development of Trajectory Statistical Methods (TSMs). The pathways of air parcels are then described in terms of probability distributions. Then, these models, taking into account turbulence, can be used to study the transport also in the planetary boundary layer. On the contrary, deterministic trajectories models are more suited to large scale circulation study.

Most TSMs reported in literature are based on the Residence Time (RT) analysis, initially presented by Ashbaugh (Ashbaugh L.L, 1983). This analysis is based on the estimation of the probability of finding a randomly selected air parcel in a given position during a given time period, along the pathway between source regions and receptor sites. The contribution of pollutant sources to measured concentration at receptor sites is then evaluated as a function of the time spent by air parcels in the polluted area. More quickly the air parcels pass through a pollutant source region, less time they have to accumulate pollutants. Starting from the RT analysis, other methods have been proposed: methods that determine the residence time for high or low concentrations measured at receptor site (e.g. Probability Source Concentration Field method) (Ashbaugh, L.L. et al., 1985, Zeng, Y. and P.K. Hopke, 1989) and methods that weight the residence time by measured concentration at the receptor point (e.g. Concentration Field method, Seibert, P. et al., 1994, and Redistributed Concentration Field methods, Stohl, A., 1996). The results of TSMs are spatial distributions of ratios as for Probability Source Concentration Field (PSCF) method, or concentrations, as for Concentration Field (CF) and Redistributed Concentration Field (RCF) methods, which can either be interpreted as spatial distributions of potential sources and sinks or as an indication of the spatial distributions of the trace substance itself.

In this work, we briefly review some TSMs reported in literature. Then, starting from the CF analysis, we introduce a statistical methodology, that takes into account the concentrations measured at multiple receptors only during peak events. As a validation test, we apply the described methodology to artificial receptor data that are derived from numerical simulations of the Lagrangian Dispersion model FLEXPART (Stohl, A. et al., 2005).

2. METHODOLOGY

Ashbaugh (Ashbaugh, L.L., 1983, Ashbaugh et al., 1985) was one of the first to use back trajectories analysis to identify source regions, introducing the Residence Time (RT) Analysis. This analysis is based on the estimation of the probability P[A], i.e. the probability of finding the particle in the position (X,Y), where X and Y are the air parcel's longitude and latitude, respectively, at some time $t' < t$. Given the position of an air parcel at time t, the position of the air parcel at time t', i.e., a time point on the trajectory, defines a set of points in the space of all possible coordinates on the Earth's surface. The estimation of this probability from actual trajectory is referred to as a Residence Time Analysis (Ashbaugh, L.L., 1983) and can be calculated as:

$$P\left[A_{ij}\right] \approx \frac{n_{ij}}{N} \tag{1}$$

where n_{ij} is the number of positions of the air parcels that, at each time increments (referred to as trajectory segment endpoints), fell in the ij^{th} cell (i.e. the grid cell of latidude-longitude coordinates *(i,j))*, during a time interval T, and N is the total number of endpoints computed for the time interval. The probability P $[A_{ij}]$ represents the residence time of a randomly selected air parcel in the ij^{th} cell relative to the total time interval T.

It is also interesting to determine the residence time for high concentrations, that is the probability of the event that an air parcel with a high pollutants concentration is in the ij^{th} cell. We can then calculate the conditional probability of the event B_{ij} (trajectories that arrived at the receptor site with pollutant concentration higher than a prescribed value), given that the event A_{ij} (that is the residence time of a randomly selected air parcel on the ij^{th} cell relative to time period T) occurs. If n_{ij} endpoints are into the ij^{th} cell during the period T, the probability of this event A_{ij} is given by $P[A_{ij}]=n_{ij}/N$. If in the same ij^{th} cell there are m_{ij} endpoints that correspond to the trajectories that arrive at the receptor site with pollutant concentration higher than a prescribed value, the probability of this event, B_{ij} is $P[Bij]=m_{ij}/N$. The conditional probability is:

$$P\left[B_{ij} \mid A_{ij}\right] = \frac{P[B_{ij}]}{P[A_{ij}]} = \frac{m_{ij}}{n_{ij}} = \frac{\sum_{l=1}^{M_t} \tau_{ijl}}{\sum_{l=1}^{M} \tau_{ijl}} \tag{2}$$

where M is the total number of trajectories, M_t is the total number of trajectories arriving to the receptor point when concentration is above the threshold value, and τ_{ijl} the time spent in grid cell *(i,j)* by trajectory l . Zeng, Y. and P.K. Hopke (1989) called this method Potential Source Contribution Function (PSCF), since a high value of the conditional probability in a cell indicates that this region has a high potential to influence the air pollution concentration observed at receptor site.

Other methods use the pollutant concentration measured at receptor point to weight the residence time. We recall here the Concentration Field method (CF) (Siebert, P. et al., 1994) based on the computation of the logarithmic mean concentration for each grid cell of the domain of trajectories simulation:

$$C_{ij} = \frac{1}{\sum_{l=1}^{M} \tau_{ijl}} \sum_{l=1}^{M} \log\left(c_l\right) \tau_{ijl} \tag{3}$$

where *(i,j)* are the indices of the horizontal grid, l the index of trajectory, M the total number of trajectories, c_l the concentration observed on arrival of trajectory l and τ_{ijl} the time spent in grid cell *(i,j)* by trajectory l. A high value of C_{ij} means that, on average, air parcels passing over cell *(i,j)* result in high concentrations at the receptor site. The CF method leads to underestimate the spatial gradients of the true source fields, since the concentration measured at the receptor sites is attributed with equal weight to all segments of the trajectory, whereas pollution sources are usually concentrated in ''hot spots'' (Stohl, A., 1996). Stohl, A. (1996) proposed the redistribution concentration field method (RCF), based on an iterative redistribution of the concentrations along the trajectories. A first guess concentration field is calculated by the CF method, then the concentrations values along each trajectory are re-weighted according to the ratio of the concentration of that grid cell to the mean

concentration of all grid cells along the path of that particular trajectory. After the redistribution is performed for all trajectories, a new concentration field can be computed with the equation:

$$C_{ij} = \frac{1}{\sum_{l=1}^{M}\sum_{i=1}^{N_l}\tau_{ijkl}}\sum_{l=1}^{M}\sum_{k=1}^{N_l}\log(C_{kl})\tau_{ijkl} \tag{4}$$

where C_{kl} is now the redistributed concentration (and not the measured concentration) and τ_{ijkl} is the residence time of segment k of trajectory l in grid cell (i,j). With this new concentration field, a second redistribution of the concentration along the trajectory is done. The complete procedure is iteratively repeated until the average difference between the concentration fields of two successive iteration is below 0.5%. The major assumption of the redistribution method is that the measured species is directly emitted or produced by linear chemistry.

We are now interested in evaluating the relationship between peak events in the pollutant concentration, measured at the receptor sites, and source emission intensity. To this goal, we propose the following modified CF method, herein called Peak Events Concentration Field (PECF):

$$C_{ij} = \frac{1}{\sum_{l=1}^{M_t}\tau_{ijl}}\sum_{l=1}^{M_t}\log(c_l)\tau_{ijl} \tag{5}$$

where (i,j) are the indices of the horizontal grid, l the index of trajectory, M_t the total number of trajectories that arrived at the receptor site when pollutant concentration is higher than a prescribed value, c_l the concentration observed on arrival of trajectory l and τ_{ijl} the time spent in grid cell (i,j) by trajectory l. Note that this modified CF method, at variance with Eq. 3, includes the residence times of trajectories only when the measured concentration field, at the receptor site, is above a given threshold value. The PECF method of Eq. (5) can be considered, in some sense, as a bridge between Eq. (2) (PSCF) and Eq. (3) (CF). Note also that, in Eq. (5), at variance with Eq. (2), theres' no conditional probability, and thus there's no need to compute all the trajectories for all time values of the concentration at the receptor sites. This allow to reduce the simulation time. Basically, in Eq. (5) the role of the statistical weights given by the denominator in Eq. (2) has been substituted by the logarithm of the peak values c_l, which is based on the hypothesis that the peak events constitute a complete set of random events or, in other words, on the hypothesis that under-threshold events determine a small contribution to the statistical evaluation of the source spatial distribution.

3. NUMERICAL RESULTS AND DISCUSSION

Before to apply a new method to a real world situation, its reliability should be checked under simplified and controlled conditions (Scheifinger, H. and A. Kaiser, 2007). To this aim, we have generated daily artificial inert tracer concentrations using the FLEXPART model (Stohl, A. et al., 2005), with wind field from ECMWF at horizontal resolution of 0.25° and temporal resolution of three hours. The simulation domain extends over Europe for the period march 2006. The simulations, in forward mode, have been performed with a continuous release of 10^6 inert particles between ground and 500m, from an aerial source. Figure 1 shows the portion of simulation domain around source position (square) and receptor points (crosses). The simulated time series concentrations at two receptor points and FLEXPART model, used in backward mode, have been used to implement the CF, PSCF and PECF methods.

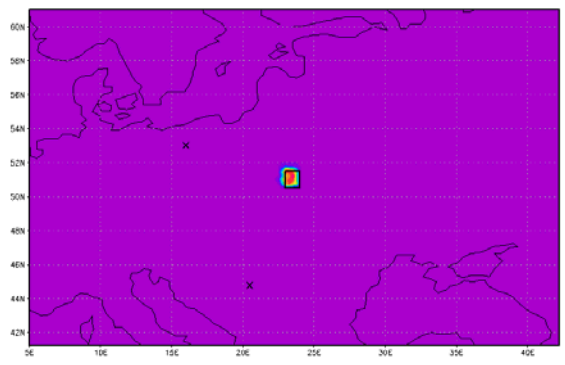

Figure 1. Portion of the simulation domain used for backward and forward simulations. The square indicates the source region used in forward simulation, crosses indicate the receptor points.

Figure 2 shows the source spatial distribution, as output of the CF method. Formula 3 has been implemented including all trajectories starting from receptor point for value of concentration different from zero.

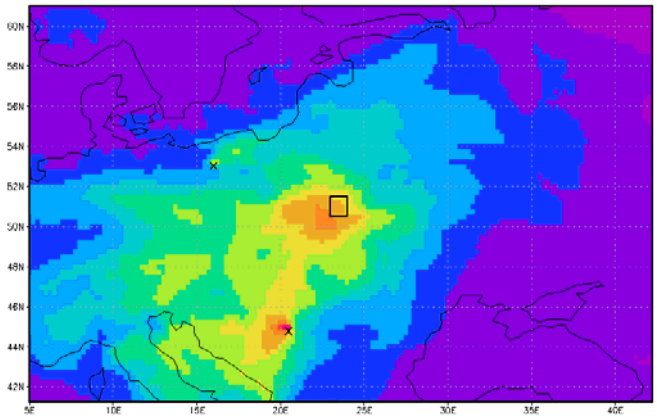

Figure 2. Source spatial distribution from the CF method.

Figure 3 and Figure 4 show the source spatial distribution reconstructed respectively with PSCF and PECF methods. The threshold on measured concentration at the receptor sites is such that only the concentration values that reside in the highest 6% of all non-zero measurements are included.

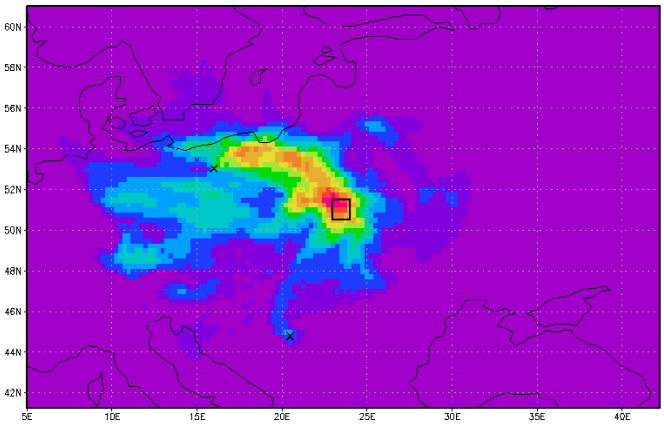

Figure 3. Source spatial distribution from the PSCF method.

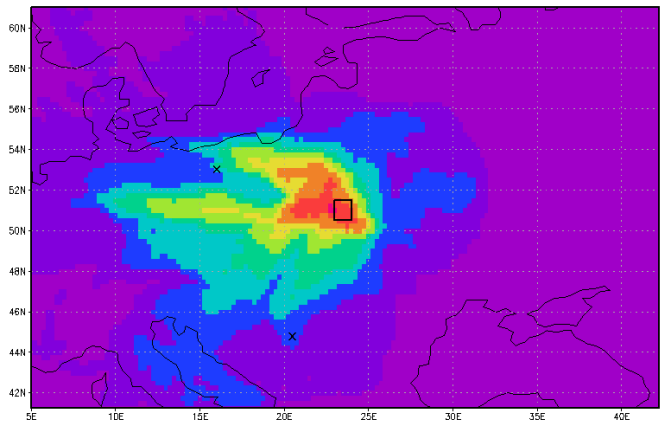

Figure 4. Source spatial distribution from the PECF method**.**

From Figures 2 and 3 we find the emergence of the well-known ghost potential sources in the wake of the real emission source in both CF and PSCF methods. These spurious sources are known to emerge in the Residence Time approaches. The PSCF method was derived from the CF method as a tentative to avoid spurious effects

associated with low concentration values and, actually, it seems to better estimate the position of the real emission source (see Figure 3). However, also the emergence of ghost sources in the wake of the real source seems to be more enhanced in the PSCF method with respect to the CF one. Interestingly, ghost sources in the CF method are more evident along the way between source and receptor at latitude/longitude=20.50°E/44.80°N, while the PSCF method shows a much more evident wake between source and receptor at latitude/longitude=16.00°E/53.00°N.

The PECF method can be considered a combination of the CF and PSCF methods, in the sense that it includes the possibility of using above-threshold concentration values, such as in the PSCF method, but it also includes the concentration values, measured at the receptor sites, as statistical weights, such as in the CF method.

At variance with both CF and PSCF methods, the PECF method does not use all the trajectories arriving at receptor sites, which in the PSCF method are included in the denominator while in the CF method are in both the numerator and the denominator. On the contrary, PECF method assumes that the set of above-threshold concentration values includes the essential information about the emission sources or, using a probability theory "jargon", it is a complete set of random events. Apart from the aspects related to the ghost sources, this also allow form much faster simulations with respect to the other approaches.

The numerical simulations seem to prove that this approach could overcome the problem of ghost sources in the wake of the real emission source, at least for the used configuration of receptor points. This is probably due to the fact that the PECF method, in addition to considering the above-threshold concentration values (as in the PSCF method), takes into account also the concentration value at receptor sites. In PSCF method, grid cells having same PSCF can result from trajectories starting from receptor points when measured concentration is slightly higher than threshold value or extremely high concentration (Hsu, Y.K. et al., 2003). These first positive results need further work that has to be done to test the PECF method with different spatial distributions of receptor points and emission sources, before it can be applied to real situations.

REFERENCES

Ashbaugh, L.L, 1983: A statistical trajectory technique for determining air pollution source regions. *J. Air Pollut. Control Assoc.*, **33**, 1096-1098

Ashbaugh, L.L., W.C. Malm and W.Z. Sadeh, 1985: A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmos. Environ.*, **19**, 1263–1270

Henry, R.C., 1997: History and fundamentals of multivariate air quality receptor models. Chemometr. Intell. Lab. (37), 37-42

Hsu Y.K., T.M. Holsen and P.K. Hopke, 2003: Comparison of hybrid receptor models to locate PCB sources in Chicago. *Atmos. Env.*, **37(4)** , 545-562

Kabashnikov,V.P., A.P. Chaikovsky, T.L. Kucsera and N.S. Metelskaya, 2011: Estimates accuracy of three common trajectory statistical methods. *Atmos.Env.*, **45(31)**, 5425-5430

Lin, C.H. and Chang, L.F.W., 2002: Relative source contribution analysis using an air trajectory statistical approach. *Journ. of Geophys. Res.-Atmos.*, **107(D21)**, DOI: 10.1029/2001JD001301

Scheifinger H. and A. Kaiser, 2007: Validation of trajectory statistical methods., *Atmos.Env.*, **41(39)**, 8846-8856

Seibert, P., H. Kromp-Kolb, U. Baltensperger, D.T. Jost, M. Schwikowski, A. Kasper, H. Puxbaum, 1994: Trajectory analysis of aerosol measurements at high alpine sites. In Borrell, P.M., Borrell, P., Cvitai, T., Seiler, W. (Eds.), Transport and Transformation of Pollutants in the Troposphere. Academic Publishing, Den Haag, 689–693

Stohl. A., 1996: Trajectory statistics – A new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulphate in Europe. *Atmos. Env.*, **30(4),** 579-587

Stohl, A., C. Forster, A. Frank, P. Seibert and G. Wotawa, 2005: Technical note: the Lagrangian particles dispersion model FLEXPART version 6.2. *Atmos. Chem. Phys.*, **5**, 2461-2474.

Zeng, Y. and P.K. Hopke, 1989: A Study of the sources of acid precipitation in Ontario, Canada. *Atmos. Env.*, **23(7)**, 1499-1509