Environmental Modelling & Software 88 (2017) 183–187

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A user-friendly tool for comprehensive evaluation of the geographical origins of atmospheric pollution: Wind and trajectory analyses

J.-E. Petit [a, *], O. Favez [b], A. Albinet [b], F. Canonaco [c]

[a] Air Lorraine, 20 rue Pierre Simon de Laplace, 57070, Metz, France
[b] INERIS, Parc technologique ALATA BP2, 60750, Verneuil-en-Halatte, France
[c] Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, 5232, PSI Villigen, Switzerland

A R T I C L E I N F O

Article history:
Received 18 July 2016
Received in revised form
15 November 2016
Accepted 22 November 2016
Available online 29 November 2016

Keywords:
ZeFir
Atmospheric pollution
Wind
Air mass
Geographical origin

A B S T R A C T

Various receptor methodologies have been developed in the last decades to investigate the geographical origins of atmospheric pollution, based either on wind data or on backtrajectory analyses. To date, only few software packages exist to make use of one or the other approach. We present here ZeFir, an Igor-based package specifically designed to achieve a comprehensive geographical origin analysis using a single statistical tool. ZeFir puts the emphasis on a user-friendly experience in order to facilitate and speed up working time. Key parameters can be easily controlled, and unique innovative features bring geographical origins work to another level.

1. Introduction

The determination of the sources of atmospheric pollutants is a key parameter to improve air quality worldwide. Recent breakthroughs in statistical analysis now allow for more and more comprehensive source apportionment studies (e.g. Paatero, 1999; Paatero and Hopke, 2009). However, answering "What are the sources?" is not quite enough: defining their emission location represents an essential information.

While the backtrajectories of air masses are used to investigate potential advection of pollution over large geographical scales, they conceptually fail to provide a meaningful allocation of local sources. Coupling concentrations with measured wind data leads to refined information on local/regional sources, but the wind direction measured at a receptor site is not necessarily representative of the air mass origin. Both approaches are complementary, but are rarely performed together, mainly because they require specific scripts or programs, when they exist and/or are available. Indeed, the widespread use of a given methodology is essentially driven by its potential to solve a given problem, and also by the user-friendliness of the tool. For example, one of the reasons why Positive Matrix Factorization (PMF) analyses are nowadays regularly performed for source apportionment studies are partly linked to the development of user-friendly interfaces, like US EPA PMF software (Norris et al., 2008) or SourceFinder (SoFi, Canonaco et al., 2013).

In the case of geographical origin works, several receptor-based methodologies have been developed in the past decades such as Non-parametric Wind Regression (NWR, Henry et al., 2009), Conditional Bivariate Probability Function (CBPF, Uria-Tellaetxe and Carslaw, 2014) for wind analyses; Potential Source Contribution Function (PSCF, Polissar et al., 2001), or clusters for trajectory analyses. They all have their own pros and cons, and no single approach can be brought out as the one that should always be used. This emphasizes the need of tools providing the whole diversity of such methodologies; and because such tools serve science, it is critical to make them the most user friendly as possible. For instance, Openair (Carslaw and Ropkins, 2012) is a powerful R-package that offers, among others, the possibility to perform trajectory and wind analysis; but the lack of Graphical User Interface (GUI) requires the user to dig into the functions and get accustomed to R coding. Conversely, TrajStat (Wang et al., 2009) is a standalone program that is Geographical Information System (GIS)-based, but is able to perform trajectory analysis only. There is therefore a gap to fill with a GUI solution to perform both, wind and trajectory analyses.

---

* Corresponding author.
E-mail address: jeaneudes.petit@gmail.com (J.-E. Petit).

http://dx.doi.org/10.1016/j.envsoft.2016.11.022

Availability

Program name ZeFir
Developer Jean-Eudes Petit
Contact address zefir.contact@gmail.com
Year first available 2016
Software required Igor pro (wavemetrics) v6.3 or later installed.
Program language Igor
Package size 6.5 Mo.
Availability https://sites.google.com/site/zefirproject
Current version 3.10
Cost free of charge

This paper aims at describing a new Igor-based tool, named ZeFir, allowing to perform a comprehensive investigation of the geographical origins of air pollutants. The philosophy of ZeFir is to give users the possibility to interactively explore various methodologies for wind and trajectory analyses. NWR, Sustained Wind Incidence Method (SWIM), PSCF, Concentration-Weighted Trajectory (CWT) or Concentration Field (CF) can be easily performed from user-friendly panels, where some innovative features are implemented.

2. Background on Igor and ZeFir structure

Igor Pro (http://www.wavemetrics.com) is a data-analysis software, combining point-and-click interactions and programming. Because Igor is a licensed and proprietary software package, the choice of distributing ZeFir under this environment could, at first sight, constitute a major limitation. This choice has been constrained by two main reasons:

- First, the “hard science” behind ZeFir is easily programmable (i.e. not computationally intensive), and can therefore be implemented into an open-source software package, such as Openair. However, much more work would need to be done to build a GUI, which is essential to facilitate the use of such tool. Igor already benefits from a whole workspace to store, visualize, edit, and plot the data. Unlike some open-source software applications (e.g. R or Python), graphs are not “frozen” objects, which the user can easily interact with. Each point-and-click operation is linked to a displayed line code, which greatly simplifies script coding, especially for complex graphs and GUIs. A built-in debugger eventually makes the code more robust. The “Igor procedure” format makes ZeFir portable to any computer with an up-to-date version of Igor Pro 6 installed.
- Second, Igor is already well implemented within the atmospheric science community, notably through a growing interest in the aerosol mass spectrometer measurements (ACSM/AMS). Then, because i) the existing source apportionment user-friendly tools are unfortunately not open-source, but are compiled standalone programs - e.g., EPA PMF toolkit -, or Igor procedures, such as SoFi or PET (PMF Evaluation Tool - Ulbrich et al., 2009), and ii) the geographical origin work is complementary to source apportionment, the only way in order to have an integrated solution avoiding untimely copy-and-paste (or export-import) between a program to another was constrained to develop an Igor based procedure.

ZeFir structure is illustrated in Fig. 1. “Concentration” covers a large variety of data type: i) data already stored or produced within Igor experiments (e.g. aerosol mass spectrometer chemical composition, or outputs from SoFi or PET); or ii) external data which can be manually imported, such individual atmospheric chemical species, or outputs from other source apportionment software applications. These can be manually imported by copy-and-paste operations or dedicated built-in panels permitting the import of various types of files (e.g. Excel, text or NetCDF files).

3. Wind analysis

Wind analysis in ZeFir is performed by NWR. Originally developed by Henry et al. (2009), NWR couples ambient concentrations with co-located measurements of wind direction and speed. This approach can be simplified as a weighing average of the data at each predictive (θ, u) couple (respectively representing predictive wind direction and speed), where the weighing coefficients are determined through Gaussian-like functions; the global idea is to give weight to concentration values associated to wind direction and speed relatively close to (θ, u).

One methodological improvement of NWR, called SWIM (Sustained Wind Incidence Method) consists in taking wind direction and speed standard deviations into account instead of constant smoothing parameters (Vedantham et al., 2012). This allows to dynamically “downweight” data associated with high standard deviations. However, the main limitation of SWIM is that its scope of action is limited to high temporal resolution data (e.g. between 1 min and few hours), as lower time resolution, due to atmospheric variability, generally increases these standard deviations, leading to over-smoothed results. A variant of SWIM, called SWIM-2 in ZeFir, uses constant smoothing parameters and a weighing scalar, which depends on wind fluxes and wind direction standard deviation (Olson et al., 2012). Both SWIM and SWIM-2 methods are available in ZeFir.

As an example, Fig. 2 shows the results of different methods of wind analysis applied to 3-h PM_1 non-refractory (NR) chloride concentrations measured between June 2011 and May 2013 South-West of Paris, France.

NR-chloride in urban areas can mainly originate from anthropogenic activities, such as waste incineration. Fig. 3 illustrates the localization, around Paris, of such facilities, and the receptor site. The three wind analyses results are all in accordance with this map, as most of incinerators are located in the North to East sector. More specifically, SWIM-2 highlights hotspots in the N-NNE and NE-to-E sectors at wind speeds of about 20 and 18 km/h, respectively. This contrasts with NWR and CBPF, which give rather homogeneous concentrations between N and E over a large range of wind speeds. The benefit of SWIM-2 in this example is that data associated with unstable wind conditions were effectively downweighted. But it is noteworthy that SWIM approaches should not i) be always best suited, since each dataset is different, and ii) always meet the user's needs.

4. Trajectory analysis

PSCF, CWT and CF approaches investigate potential transport of pollution over large geographical scales (Polissar et al., 2001; Fleming et al., 2012). They couple atmospheric concentrations with backtrajectories and use residence time information (Fleming et al., 2012) to geographically identify air parcels that may be responsible of high concentrations observed at the receptor site.

Fig. 4 shows the results of a CWT calculation performed on hourly SO_2 concentrations measured in Dundee (Scotland, UK) during September 2014. Hourly 72-h backtrajectories arriving at 100 m above sea level were calculated from the PC-based version of HYSPLIT (Stein et al., 2015). Provided maps alternatively illustrate

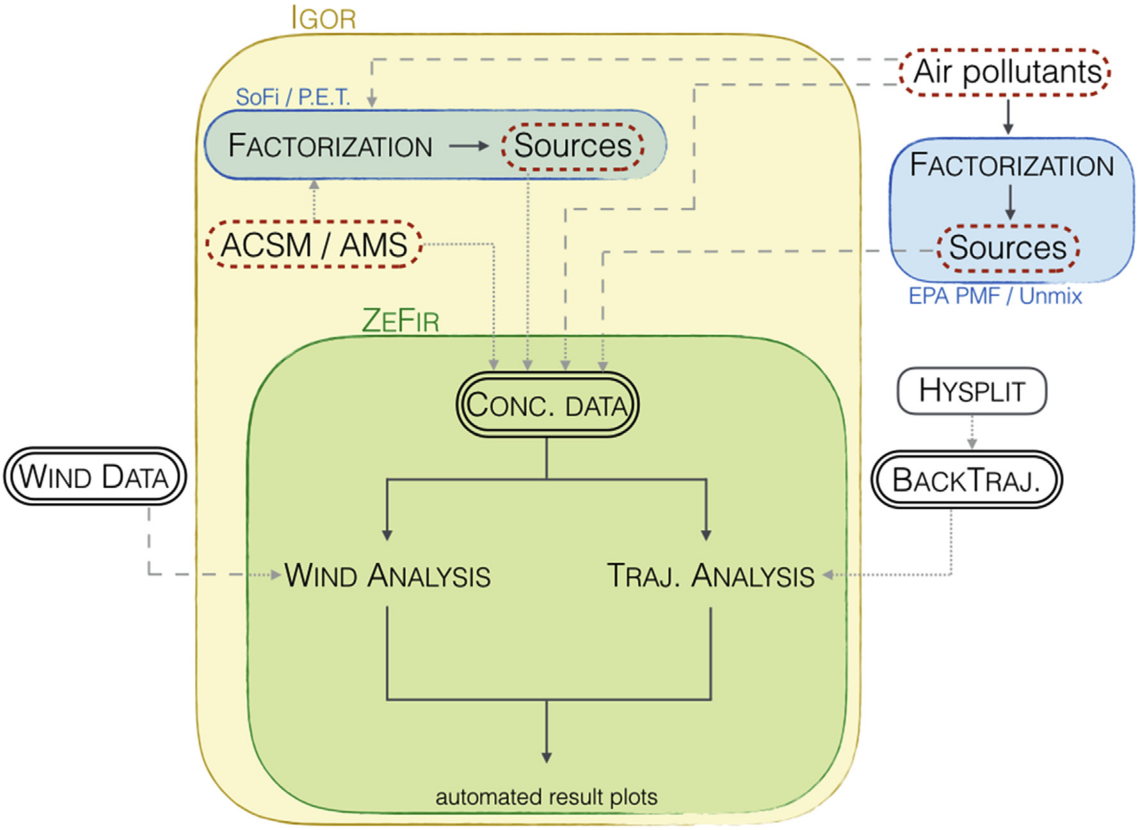


Fig. 1. Flowchart showing ZeFir structure. Green, yellow and blue rectangles respectively delimit the scope of action of ZeFir, Igor and source apportionment tools. Double-boxed rectangles highlight requested input data. Large and narrow dotted lines represent manual and automatic data import. Unboxed texts represent calculation steps.

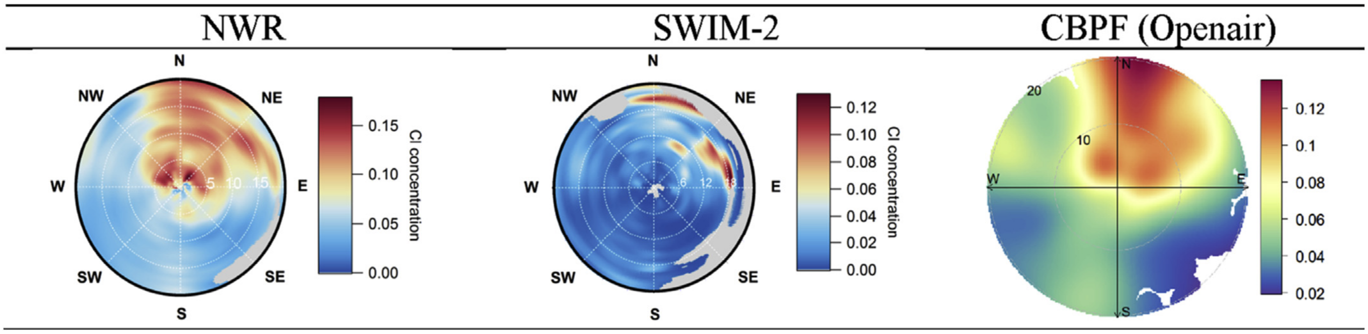


Fig. 2. Wind analysis results using NWR (left), SWIM-2 (middle) and CBPF (right) on 3-h Cl concentrations measured in Paris, France between June 2011 and May 2013.

the Bárdarbunga eruption (Iceland) that occurred from August to November 2014, which caused significant amount of volcanic-related material released within the atmosphere.

So far, ZeFir uses backtrajectory files from HYSPLIT only, since this model is extensively used in trajectory analyses (e.g. see Table 1 in Fleming et al., 2012). Trajectory calculation is critical regarding the uncertainty of final trajectory analyses results (Polissar et al., 1998). The use of different wind field files surely helps to down-weight these uncertainties, but it is, in the case of ZeFir, the user's responsibility to calculate his own backtrajectories, as long as HYSPLIT is used. The retrieval of backtrajectories from different dispersion models may also be a good alternative, and future versions of ZeFir will allow the import of backtrajectory files from other models than HYSPLIT.

Another indirect source of uncertainties is the temporal resolution of the input dataset. In the case of low time-resolution measurements (e.g. daily filter sampling), only one back-trajectory is associated to the concentration point, leading to poor temporal representativeness. A built-in interactive solution is available in ZeFir, and consists in enlarging the size of the input data in order to consider more trajectories, leading to more representative results. Fig. 5 illustrates the utility of such an approach: particulate sodium ion concentrations have been determined from 24-h PM_{10} high-volume sampling carried out every third day from April 2015 to March 2016 in Metz, France. Using the regular dataset, results indicate that sodium may have a rather continental origin with emission zones in the in-lands. This is in contradiction with the expected pattern, since, given the location of the sampling site, it should mainly originate from sea salt. However, when enlarging the dataset to take back-trajectory at +3, +6, +9, +12, +15, +18 and + 21 h into account, the oceanic pattern is much more highlighted and thus in

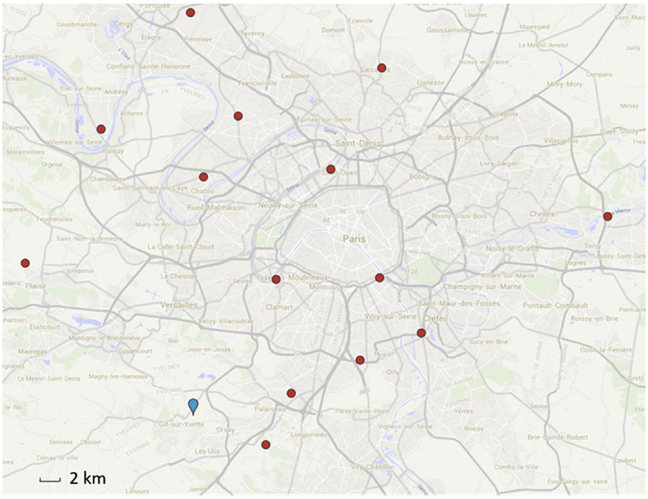


Fig. 3. Localization of incinerators around Paris (red dots), and receptor site (blue label pin).

accordance with what is expected. To date, no existing tool proposes this feature.

A limitation of PSCF-based approaches is the need to settle a weighing function in order to downweight cells associated with low residence time, usually observable as “trailing effects”. In ZeFir, a relative scale of the trajectory density is used (Waked et al., 2014), and allows a less empirical determination of this weighing function. Then, determining the most adapted coefficients is usually made by running as many runs as necessary. In ZeFir, one single run is needed, and a dedicated panel allows the user to interactively appreciate the changes in the weighing function, where the graphical visualization of the weighing function helps to save time in this process (Fig. 6).

Another innovative feature of ZeFir is to use additional parameters along each trajectory in order to remove meaningless endpoints that should not be associated with potential emission zones. There are several ways to perform this, and actually depend on the information provided by the user. In particular, precipitation values at each trajectory endpoint can be used to “cut” the trajectory where rain occurred, thus avoiding to take the rest of the back-trajectory into account, because the associated air parcel has been washed out. Similarly, an altitude threshold can be used, assuming that above a certain value, the emissions in the air parcel could not reach this trajectory. To our knowledge, no other tools propose this methodological improvement.

Merging different trajectory analyses is also available in ZeFir. This can be especially useful to combine results from different sampling sites, leading to higher trajectory density values (Biegalski and Hopke, 2004). This helps to get the bigger picture of a particular study, as it could identify emission hotspots that have an impact on larger geographical scales. Han et al. (2007) have also shown that multi-site (MS) PSCF or CWT could lead to the identification of different emission zones of gaseous mercury. Merging the results from the different approaches was then performed to identify major sources over the United States, and to refine the national emission inventory. Because each run is physically saved, combining results is made very easy through a built-in feature. Again, no other tools than ZeFir allow user-friendly multisite merging.

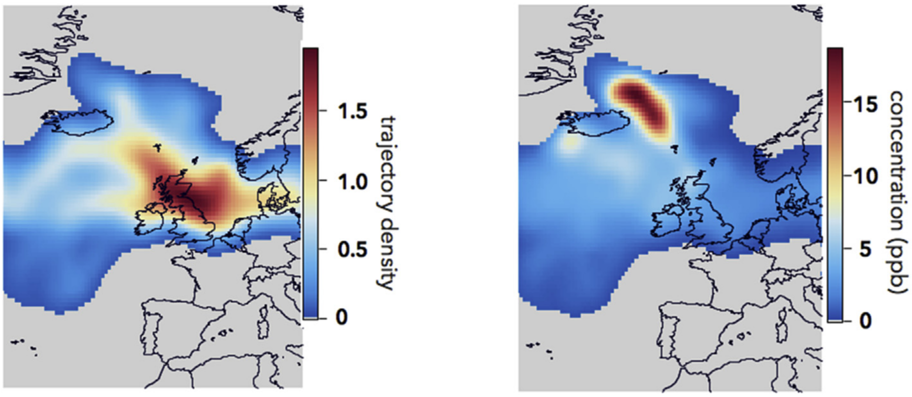


Fig. 4. CWT results for 1-h SO_2 concentrations measured at Dundee during September 2014 (left) trajectory density (log of residence time, no unit); (right) estimated concentration (in ppb).

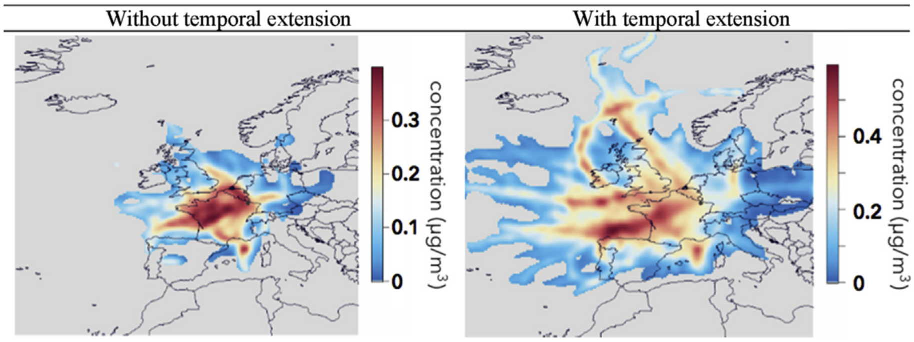
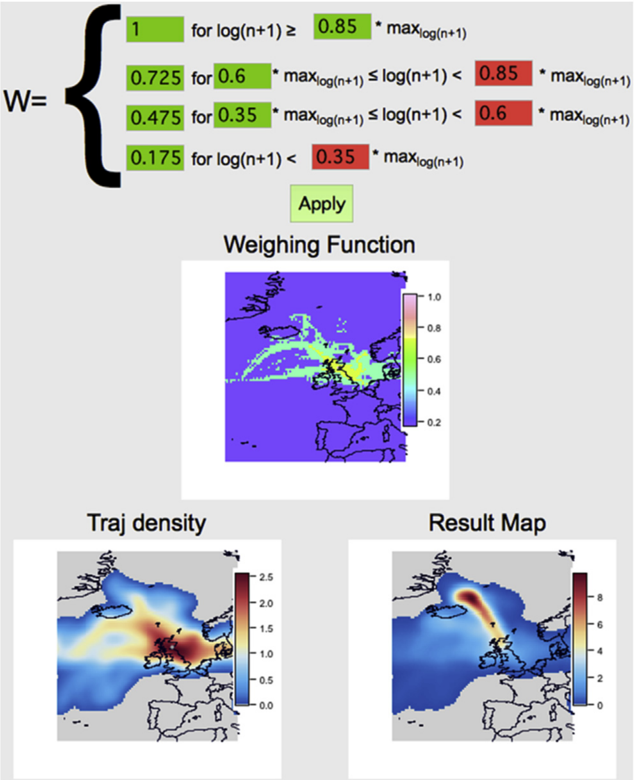


Fig. 5. CWT approach on Sodium ion measured in Metz, France with (right) and without (left) temporal extension of the dataset.



$$W=\begin{cases} 1 & \text{for } \log(n+1) \geq 0.85 * \max_{\log(n+1)} \\ 0.725 & \text{for } 0.6 * \max_{\log(n+1)} \leq \log(n+1) < 0.85 * \max_{\log(n+1)} \\ 0.475 & \text{for } 0.35 * \max_{\log(n+1)} \leq \log(n+1) < 0.6 * \max_{\log(n+1)} \\ 0.175 & \text{for } \log(n+1) < 0.35 * \max_{\log(n+1)} \end{cases}$$

Fig. 6. Weighing function panel in ZeFir. Log(n+1) represents the trajectory density.

4.1. Future developments

ZeFir falls in with existing tools whose philosophy is to make various methodologies accessible to the scientific community. For wind analyses, further developments may be focused on implementing other methodologies. In particular, one disadvantage of wind analyses is that wind speed cannot be directly related to a distance, leading to a risk of misinterpretation when plotting results on top of a map. Henry et al. (2011) proposed to calculate short backtrajectories (few hour durations) from wind data to convert the results into a distance, and thus to better allocate local sources. Also, there would be a great interest in coding CBPF in Igor.

For trajectory analyses, cluster calculation may be implemented, as well as the import from other dispersion models than HYSPLIT. Then, further methodological improvements may be proposed (3-D backtrajectories, or the consideration of e.g. the mixing layer height along trajectory).

Finally, in an open-source effort, collaborative work could for instance include some of the methodological improvements that ZeFir proposes into Openair as functions.

Acknowledgements

The authors would like to thank Air Lorraine for financial support, the Igor community for their precious help, and the Scottish air quality network for SO_2 data.

References

Biegalski, S.R., Hopke, P.K., 2004. Total potential source contribution function analysis of trace elements determined in aerosol samples collected near lake huron. Environ. Sci. Technol. 38 (16), 4276–4284. http://dx.doi.org/10.1021/es035196s.

Canonaco, F., Crippa, M., Slowik, J.G., Baltensperger, U., Prévôt, A.S.H., 2013. SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data. Atmos. Meas. Tech. 6 (12), 3649–3661. http://dx.doi.org/10.5194/amt-6-3649-2013.

Carslaw, D.C., Ropkins, K., 2012. Openair — an R package for air quality data analysis. Environ. Model. Softw. 27–28, 52–61. http://dx.doi.org/10.1016/j.envsoft.2011.09.008.

Fleming, Z.L., Monks, P.S., Manning, A.J., 2012. Review: untangling the influence of air-mass history in interpreting observed atmospheric composition. Atmos. Res. 104–105, 1–39. http://dx.doi.org/10.1016/j.atmosres.2011.09.009.

Han, Y.-J., Holsen, T.M., Hopke, P.K., 2007. Estimation of source locations of total gaseous mercury measured in New York State using trajectory-based models. Atmos. Environ. 41 (28), 6033–6047. http://dx.doi.org/10.1016/j.atmosenv.2007.03.027.

Henry, R., Norris, G.A., Vedantham, R., Turner, J.R., 2009. Source region identification using Kernel smoothing. Environ. Sci. Technol. 43 (11), 4090–4097. http://dx.doi.org/10.1021/es8011723.

Henry, R.C., Vette, A., Norris, G., Ved, R., Kimbrough, S., Shores, R.C., 2011. Separating the air quality impact of a major highway and nearby sources by nonparametric trajectory analysis. Env. Sci. Technol. 45, 10471–10476. http://dx.doi.org/10.1021/es202070k.

Norris, G., Vedantham, R., Wade, K. S., Brown, S. G., Prouty, J. D., and Foley, C.: EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals and User Guide. Prepared for the US. Environmental Protection Agency, Washington, DC, by the National Exposure Research Laboratory, Research Triangle Park; Sonoma Technology, Inc., Petaluma, CA; and Lockheed Martin Systems Engineering Center, Arlington, VA, EP-D-05–004; STI-907045.053347-UG, October, 2008.

Olson, D.A., Vedantham, R., Norris, G.A., Brown, S.G., Roberts, P., 2012. Determining source impacts near roadways using wind regression and organic source markers. Atmos. Environ. 47, 261–268. http://dx.doi.org/10.1016/j.atmosenv.2011.11.003.

Paatero, P., 1999. The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. J. Comput. Graph. Stat. 8 (4), 854–888. http://dx.doi.org/10.1080/10618600.1999.10474853.

Paatero, P., Hopke, P.K., 2009. Rotational tools for factor analytic models. J. Chemom. 23 (2), 91–100. http://dx.doi.org/10.1002/cem.1197.

Polissar, A.V., Hopke, P.K., Paatero, P., Malm, W.C., Sisler, J.F., 1998. Atmospheric aerosol over Alaska: 2. Elemental composition and sources. J. Geophys. Res. Atmos. 103 (D15), 19045–19057. http://dx.doi.org/10.1029/98JD01212.

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001. Atmospheric aerosol over Vermont: chemical composition and sources. Environ. Sci. Technol. 35 (23), 4604–4621. http://dx.doi.org/10.1021/es0105865.

Stein, A.F., Draxler, R.R., Rolph, G.D., Stunder, B.J.B., Cohen, M.D., Ngan, F., 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling System. Bull. Am. Meteorol. Soc. 96 (12), 2059–2077. http://dx.doi.org/10.1175/BAMS-D-14-00110.1.

Ulbrich, I.M., Canagaratna, M.R., Zhang, Q., Worsnop, D.R., Jimenez, J.L., 2009. Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data. Atmos. Chem. Phys. 9, 2891–2918.

Uria-Tellaetxe, I., Carslaw, D.C., 2014. Conditional bivariate probability function for source identification. Environ. Model. Softw. 59, 1–9. http://dx.doi.org/10.1016/j.envsoft.2014.05.002.

Vedantham, R., Norris, G., Brown, S.G., Roberts, P., 2012. Combining continuous near–road monitoring and inverse modeling to isolate the effect of highway expansion on a school in Las Vegas. Atmos. Pollut. Res. 3 (1), 105–111. http://dx.doi.org/10.5094/APR.2012.010.

Waked, A., Favez, O., Alleman, L.Y., Piot, C., Petit, J.-E., Delaunay, T., Verlinden, E., Golly, B., Besombes, J.-L., Jaffrezo, J.-L., Leoz-Garziandia, E., 2014. Source apportionment of PM10 in a north-western Europe regional urban background site (Lens, France) using positive matrix factorization and including primary biogenic emissions. Atmos. Chem. Phys. 14 (7), 3325–3346. http://dx.doi.org/10.5194/acp-14-3325-2014.

Wang, Y.Q., Zhang, X.Y., Draxler, R.R., 2009. TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data. Environ. Model. Softw. 24 (8), 938–939. http://dx.doi.org/10.1016/j.envsoft.2009.01.004.