**Advances in
Science and
Research**

Air mass origin and its influence over the aerosol size distribution: a study in SE Spain

M. Cabello, J. A. G. Orza, and V. Galiano

SCOLAb, Física Aplicada, Universidad Miguel Hernández, Elche, Spain

Abstract. A k-means cluster analysis of 96 hour trajectories arriving in Southeast (SE) Spain at 3000, 1500 and 500 m for the 7-year period 2000–2006 has been performed to identify and describe the main flows arriving at the study area. The dependence of the aerosol size distribution on the air mass origin has been studied by using non-parametric statistics. There are statistically significant differences on aerosol size distribution and meteorological variables at surface level according to the identified clusters.

1 Introduction

Backward trajectory analysis is a commonly used method to identify synoptic-scale atmospheric transport patterns and/or determine the air pollutants origin (e.g., Dorling et al., 1992; Brankov et al., 1998; Cape et al., 2000; Stohl et al., 2002; Jorba et al., 2004; Salvador et al., 2004). The errors associated to trajectory calculation are on the order of 15–20% of the distance travelled but the accuracy of the trajectory analysis increases when a set of large number of trajectories with similar characteristics is considered (Stohl, 1998), so back-trajectory cluster analysis is a suitable technique to classify air masses arriving at a study site.

In this study, the number particle size distribution is related to different long range and regional/local contributions according to the identified air flows.

Cluster analysis is a multivariate statistical technique designed to classify a large data set into *non-predefined* dominant groups called clusters. However, clustering involves some subjective non-trivial decisions: the number of clusters to use, the selection of centroids in the initialization stage, etc. To determine the appropriate number of clusters and handle the sensitivity of the method to the initial centroids selection we have followed the procedures described by Dorling et al. (1992) and Mattis (2001) and considered some modifications to them in order to get smaller (better) values of the total Root Mean Square Deviation (RMSD), the clustering figure of merit.

2 Methodology

96-h backward air trajectories arriving at 12:00 UTC at the study site for the period 2000–2006 were computed using the HYbrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model v.4 with the FNL meteorological data, from the final run in the series of NCEP operational model runs, available at the NOAA ARL (Draxler and Rolph, 2003). Hourly latitude and longitude were used as input variables in the clustering procedures.

The trajectory classification has been performed by a k-means cluster analysis. We have followed the procedure described by Dorling et al. (1992) to reduce the subjectivity in the selection of the appropriate number of clusters: the algorithm was run for a range of cluster numbers k between 30 and 2, and the percentage change in the total RMSD (i.e. the sum of the RMSD of each cluster) when the number of clusters is reduced from k to $k-1$ was used to find out the proper number of clusters. Unlike Dorling et al. (1992), we define it as the smallest number of clusters for which the smallest total RMSD change is found.

Different approaches have been considered a) for the reduction from k to $k-1$ clusters, and b) to deal with the dependence of the final cluster solution on the initial centroids. The details of the clustering methodology we have followed and its comparison with the procedures of Dorling et al. (1992) and Mattis (2001) will be published elsewhere. Here we note that these authors reduce the number of clusters from k to $k-1$ by merging the two closest clusters, while we compute 100 000 k-means analysis for each k and then retain the solution with the smallest total RMSD. Calculations for the k clusters solution do not depend on the $k+1$ cluster solution as the initial centroids are taken from randomly chosen real

---

trajectories. This approach can provide smaller total RMSDs and hence better clustering solutions than the obtained using the Dorling and Mattis procedures.

Aerosol size distributions were measured on the roof of a building of the municipality of Agost, Spain (38.44° N, 0.64° W), a village of 4000 inhabitants located 18 km from the Mediterranean coast. The main activities in Agost are related to extractive operations, brick manufacturing and grape cultivation. Measurements were taken every ten minutes from January to May 2006 with a GRIMM 190 aerosol spectrometer. Measured size distributions range from 0.25 to greater-than-32 μm diameter in 31 size channels.

A varimax rotated principal components analysis (PCA) has been applied to the normalized aerosol size channels to reduce the 31 intercorrelated variables to a smaller set of independent factors that accounted for most of the observed variance. Each factor is a linear combination of the original variables and the coefficients of the linear combinations (factor loadings) represent the degree of correlation between the variables and the factor. This method allows detection of the modes of the size distribution; moreover it decomposes the distribution into size intervals that behave in a similar way (show significant correlations) upon changes in aerosol sources and meteorological conditions (Chan and Mozurkewich, 2007; Pugatshova et al., 2007).

To detect statistically significant differences in aerosol size distribution and meteorological variables at surface level according to the identified clusters (Brankov et al., 1998), the Kruskal-Wallis test and the pairwise Mann-Whitney test have been used. We conservatively adjusted p-values α_t=0.05 in the latter using the Dunn-Sidák correction for multiple comparisons to α=1−(1−α_t)$^{1/n}$, where n is the number of pairwise comparisons done between k categories.

A number of information sources were utilized to estimate when African dust outbreaks (ADOs) occur: the study of the backtrajectory pathways, the evaluation of daily maps of surface and column-integrated aerosol concentration from the DREAM, NAAPS and SKIRON aerosol dispersion models: DREAM (http://www.bsc.es/projects/earthscience/DREAM), NAAPS (http://www.nrlmry.navy.mil/aerosol/) and SKIRON (http://forecast.uoa.gr), as well as the inspection of the NASA SeaWiFS project satellite images (http://seawifs.gsfc.nasa.gov/SEAWIFS.html).

3   Results and discussion

Trajectories arriving at 3000 and 500 m are found to be clustered into 6 groups, while for 1500 m the number of clusters is 5 (Fig. 1). Most of the 3000 m trajectories correspond to westerly flows, identified as northwesterlies (NW) of different length, southwesterlies (SW) and zonal (W) flows. At 1500 and 500 m there is an elevated occurrence of slow flows. Although computed trajectories cannot resolve mesoscale phenomena, clusters composed of short trajectories are asso-

ciated to the absence of marked advective components due to low pressure gradient situations that last several days. Such weak synoptic forcing leads to situations where sea-breeze regime develops and it is intensified by the topography and the Iberian thermal low, thus inducing mesoscale recirculations as reported in (Millan et al., 1997). Stagnant situations are also associated to the slow flows. The corresponding flows include mainly regional Mediterranean recirculations (MedR), and slow westerlies (WR) and short trajectories coming from the North (N-Eu), respectively.

A short description of the identified air masses and flows is given in the following:

NWfast: Advection of polar maritime fast air masses starting in Canada and/or northern USA. This cluster is identified only for trajectories arriving at 3000 m. It is the smallest cluster in number of trajectories (7%). Almost 50% of the cases occur in wintertime.

NWmod: Advection of polar maritime air masses entering in the western Iberian Peninsula or coming from other countries in Western Europe (UK, France). It accounts for 13% of the trajectories arriving at 3000 m, 11.6% for 1500 and 9.5% at 500 m. More than 40% of them occur in wintertime.

NWslow: Slow NW advections or slow flows from Western Europe. 15% of the back-trajectories at 3000 and 1500 m, and 7% at 500 m fall in this cluster.

W: Advection of Atlantic tropical maritime air masses (around 17%). Cluster not identified at 500 m.

SW: Air masses composed by moderately short W-SW trajectories, many of them passing over the western coast of Morocco and the Straits of Gibraltar. 52% of its trajectories are associated with African dust intrusions. This cluster is identified only for trajectories arriving at 3000 m, being the major cluster in number of trajectories (30%), most of them occur in summertime.

N-EU: Composed by continental slow NE/N flows from Western Europe, and several Arctic and polar maritime air masses. This cluster is identified only for trajectories arriving at 500 m (19%).

MedR: Composed by trajectories recirculating over the Mediterranean Sea, and by slow advections from western Europe and northern Africa (3000 and 1500 m). These flows are associated with African dust intrusions (27% of its trajectories at 3000 m, 33% at 1500 m, 41% at 500 m). It is the major cluster at 500 m (29%).

WR: Western recirculations composed by aged air masses coming from the western Iberian Peninsula or from north of Morocco. This cluster is not identified for trajectories arriving at 3000 m. Most of these trajectories are associated to days with African dust outbreaks (51% at 500 m and 39% at 1500 m).

Med: Mixed Mediterranean and continental European flows. This cluster is identified only at 500 m (11%).

The study of the influence of the air mass on the aerosol size distribution was simplified by a PCA that reduced the 31 size channels to four factors, accounting for 93.8% of the
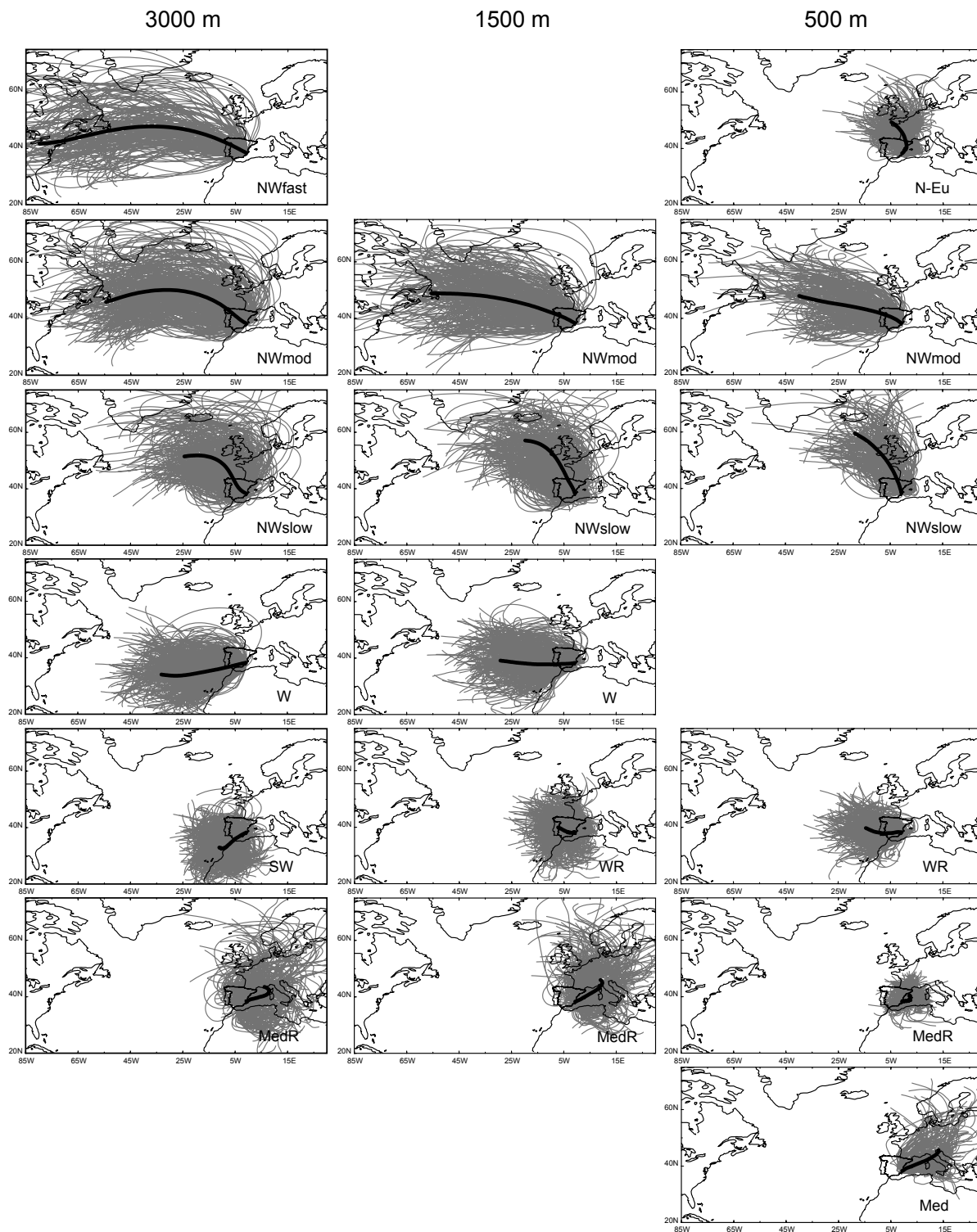
Figure 1. Clusters identified for 3000 m (left), 1500 m (center) and 500 m (right) for the 7-year study period. The representative centroid of each cluster is drawn as bold line while the single backtrajectories appear in grey.
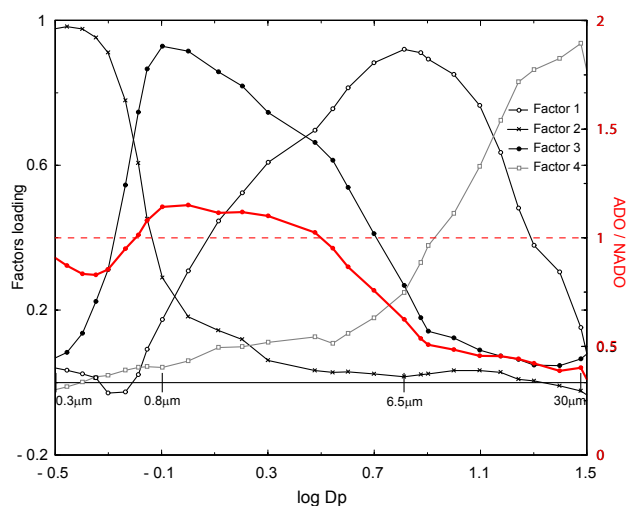
Figure 2. Factor loadings from the varimax rotated PCA analysis. Four representative size channels (0.3–0.35 μm, 0.8–1 μm, 6.5–7.5 μm and >30 μm) are selected. The ratio between the mean particle concentration on ADOs and NADOs for each size channel is shown in red; a ratio equal to one is drawn as a dashed line.

total variance. The factor loadings (Fig. 2) indicate that these factors correspond to four particulate size intervals, so the size channel with the highest loading in each factor is regarded as its representative size. Then the 0.3–0.35 μm, 0.8–1 μm, 6.5–7.5 μm, and >30 μm channels have been related to the results of the clustering analysis.

The autocorrelation function of the selected representative size channels shows stronger one-week periodicity (anthropogenic) the smaller the particle size; no periodicity is found for the >30 μm size.

The variation of the particle concentration on these size channels according to the different clusters is significant as shown by the Kruskal-Wallis test. After detection of pairwise significant differences by the corrected Mann-Whitney test air flows were grouped and labeled with a superscript number by increasing particle concentration. In each row in Table 1 (i.e. for each representative particle size), mean values not sharing the same superscript denote significant differences between particle concentration by air flow.

NW fast (3000 m) and moderate (1500, 500 m) advections are associated to the renovation of the air masses at the study site due to the entrance of clean air flows, thus the fine fraction (0.3 and 0.8 μm factors) shows low concentration levels. These situations correspond to the highest surface winds that promote local suspension of soil dust with the highest levels in the coarse fraction (6.5 and >30 μm factors).

Slow flows and regional recirculations, however, are related to smaller concentrations of the coarse fraction, while they show the highest fine fraction levels. There is an elevated number of days with such slow flows (54% of the days

for trajectories arriving at 1500 m and 72% at 500 m), that result in the accumulation of aerosols either at a regional level (by mesoscale recirculations occurring mainly in the March–October term) or at a local level (in stagnant conditions that occur mainly in the November–February term) (Rodríguez et al., 2003). We note that in coastal SE Spain the mean number of sea-breeze days is found to be between 15 and 20 a month in the study period: January–April and some days in May, as shown in (Azorín-Molina and Martín–Vide, 2007).

African dust is transported to the Western Mediterranean basin in layers at altitudes between 1500 and 5000 m (Rodríguez et al., 2001). SW flows arriving at 3000 m are closely related to ADOs. At lower altitude (500 m), a high percentage of the MedR flows corresponds also to days with African outbreaks, but in these cases trajectories do not pass over Africa. In this respect we note that on days with ADO there is also accumulation of aerosols due to the absence of advection at low altitude.

The contribution of African dust events is observed in the range 0.65–2 μm in diameter as shown in Fig. 2, where the ratio between the mean particle concentrations for ADOs and that for non African dust outbreaks (NADO) for each size channel is plot. From the daily contribution of each factor (not shown) it is obtained that Factor 3 (0.8 μm) is the only one that contributes positively to days with ADOs and negatively to NADOs. African dust intrusions are therefore related to the factor labeled as 0.8 μm. Lyamani et al. (2005) found a mode centered at a radius of 0.6 μm in the number size distribution under two ADOs in Granada, southern Spain, with a sun photometer. We note that most of the SW trajectories arriving in our study site at 3000 m have passed previously over Granada, which is 300 km away from the study area. Blanco et al. (2003), from size and shape measurements made by scanning electron microscopy, found a mode in 2 μm in diameter for two ADOs in Lecce, southern Italy. We believe that the contribution of ADOs to larger particle sizes in the study area might be hampered by the high crustal dust load due to the brick industry.

4 Conclusions

We have identified the main flows arriving at SE Spain at 3000, 1500 and 500 m by means of cluster analysis of back-trajectories for a 7-year period. In spite of the relative simplicity of the single-particle Lagrangian model utilized, that is based on the mean flow field and that considers neither deposition nor growth nor chemical processes on the aerosol particles, significant differences in particle size distribution have been found according to the main flows. Such differences indicate that flows of distinct origin, arriving at different heights at the study site, contribute differently to the aerosol load at surface level.

While the arrival of loaded air masses contribute to an increase of PM according to its typical particulate size, the

Table 1. Mean particle number concentrations (particles cm^{-3}) corresponding to the different air flows identified at each height. Mean values within a row with unlike superscript numbers are significantly different as stated from the corrected Mann-Whitney test.

	3000 m				
	NWfast	NWmod	NWslow	W	SW	MedR
0.3 μm	86243.85[1]	205914.96[4]	273800.90[6]	168876.07[2]	203013.88[3]	209414.94[5]
0.8 μm	2771.47[3,4]	2570.29[2]	2901.53[3]	3160.31[4]	2982.27[2,3]	2159.22[1]
6.5 μm	211.88[5]	121.59[4]	112.13[3]	90.90[3,4]	54.61[1]	80.22[2]
30 μm	0.17[3]	0.05[2]	0.16[2]	0.03[1]	0.04[1]	0.06[2]

	1500 m			
	NWmod	NWslow	W	WR	MedR
0.3 μm	135421.37[1]	233612.94[4]	141975.40[2]	257998.50[5]	224371.05[3]
0.8 μm	2964.42[3]	2835.41[2]	3027.13[3]	2797.01[2]	2569.04[1]
6.5 μm	223.85[4]	161.03[3]	96.87[2]	56.33[1]	77.40[2]
30 μm	0.13[3]	0.26[3]	0.04[1]	0.04[1]	0.05[2]

	500 m				
	N-EU	NWmod	NWslow	WR	MedR	Med
0.3 μm	283912.30[5]	113064.72[1]	111092.29[1]	146540.79[2]	253657.66[4]	203168.32[3]
0.8 μm	3007.33[3,4]	2810.18[2,3]	2191.58[1]	2896.07[4]	2705.33[2]	2774.19[3,4]
6.5 μm	97.10[2]	182.71[4]	167.46[4]	113.24[3]	67.95[1]	74.79[2]
30 μm	0.14[2]	0.13[3]	0.16[3]	0.04[1]	0.05[1]	0.03[1]

arrival of clean air masses (NWs) is related in the study area to moderately high surface wind speeds that increase the concentration of the coarse particles (those are entrained by the wind) and dilute the finest ones. Slow flows produced under weak synoptic forcing and stable conditions are associated to mesoscale recirculations and stagnant situations, respectively, that lead to an increase in the smallest particles while the concentration of the coarse ones decreases.

References

Azorín-Molina, C. and Martín-Vide, J.: Methodological approach to the study of the daily persistence of the sea breeze in Alicante (Spain), Atmósfera, 20, 57–81, 2007.

Blanco, A., Dee Tomasi, F., Filippo, E., Manno, D., Perrone, M. R., Serra, A., Tafuro, A. M., and Tepore, A.: Characterization of African dust over southern Italy, Atmos. Chem. Phys., 3, 2147–2159, 2003,
http://www.atmos-chem-phys.net/3/2147/2003/.

Brankov, E., Rao S. T., and Porter, P. S.: A trajectory-clustering-correlation methodology for examining the long-range transport of air pollutants, Atmos. Environ., 32, 1525–1534, 1998.

Cape, J. N., Methven, J., and Hudson, L. E.: The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland, Atmos. Environ., 34, 3651–3663, 2000.

Chan, T. W. and Mozurkewich, M.: Simplified representation of atmospheric aerosol size distributions using absolute principal component analysis, Atmos. Chem. Phys., 7, 875–886, 2007,
http://www.atmos-chem-phys.net/7/875/2007/.

Dorling, S. R., Davies, T. D., and Pierce, C. E.: Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry-Method and applications, Atmos. Environ., 26, 2575–2581, 1992.

Draxler, R. R. and Rolph, G. D.: HYSPLIT Model access via NOAA ARL READY Website (http://www.arl.noaa.gov/ready/hysplit4.html), NOAA Air Resources Laboratory, 2003.

Jorba, O., Pérez, C., Rocadenbosch, F., and Baldasano, J. M.: Cluster analysis of 4-day back trajectories arriving in the Barcelona Area, Spain, from 1997 to 2002, J. Appl. Meteorol., 43, 887–901, 2004.

Lyamani, H., Olmo, F. J., and Alados-Arboledas, L.: Saharan dust outbreak over southeastern Spain as detected by sun photometer, Atmos. Environ., 39, 7276–7284, 2005.

Mattis, I.: Compilation of trajectory data. EARLINET: A European Aerosol Research Lidar Network to Establish an Aerosol Climatology, Scientific Report for the period Feb. 2000 to Jan. 2001, J. Bösenberg, Max Planck Inst. für Meteorologie, 26–29, 2001,

available at: http://lidarb.dkrz.de/earlinet/scirep1.pdf, 2005.

Millán, M. M., Salvador, R., Mantilla, E., and Kallos, G.: Photo-oxidant dynamics in the Mediterranean basin in summer: results from European research projects, J. Geophys. Res., 102, 8811–8823, 1997.

Pugatshova, A., Reinart, A., and Tamm, E.: Features of the multi-modal aerosol size distribution depending on the air mass origin in the Baltic region, Atmos. Environ., 41, 4408–4422, 2007.

Rodríguez, S., Querol, X., Alastuey, A., Kallos, G., and Kakak-liagou, O.: Saharan dust contributions to PM10 and TSP levels in Southern and Eastern Spain, Atmos. Environ., 35, 2433–2447, 2001.

Rodríguez, S., Querol, X., Alastuey, A., Viana, M. M., and Mantilla, E.: Events affecting levels and seasonal evolution of airborne particulate matter concentrations in the Western Mediterranean, Environ. Sci. Technol., 37, 216–222, 2003.

Salvador, P., Artíñano, B., Alonso, D. G., Querol, X., and Alastuey, A.: Identification and characterisation of sources of PM10 in Madrid (Spain) by statistical methods, Atmos. Environ., 38, 435–447, 2004.

Stohl, A.: Computation, accuracy and applications of trajectories – a review and bibliography, Atmos. Environ., 32, 947–966, 1998.

Stohl, A., Eckhardt, S., Forster, C., James, P., Spichtinger, N., and Seibert, P.: A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements, Atmos. Environ., 36, 4635–4648, 2002.