

Boundaries of air mass trajectory clustering: key points and applications

I. A. Pérez¹ · M. L. Sánchez¹ · M. A. García¹ · N. Pardo¹

Received: 3 May 2016/Revised: 27 July 2016/Accepted: 28 September 2016/Published online: 8 October 2016
© Islamic Azad University (IAU) 2016

Abstract Calculating air mass trajectories is common in atmospheric analyses. However, if explainable results are to be achieved, several procedures are needed to process the vast amount of information handled. Clustering methods are statistical tools usually considered for such a purpose. Although they are based on rigorous algorithms, certain questions still remain when these methods are applied. The current review is organised in sections according to the sequence followed by such procedures. First, the types of clustering methods are described, with their core being the distance used. One key point is the stopping rule, which determines the final number of clusters. A simple classification based on this number is then suggested. Finally, the graphical presentation of the results is examined and the main drawbacks are commented on. A range of applications and results are considered to illustrate each section, and certain caveats and recommendations are also presented.

Keywords Airflow · Atmospheric science statistics · Cluster analysis · Trajectory classification · Transport patterns

Introduction

Since the atmosphere is not at rest, the effects of air masses on receptors such as materials or living beings are determined by air mass history. When air masses travel over the sea, they are loaded with moisture, whereas those moving over the continent are drier (Sun et al. 2007).

Moreover, transport corridors of dirty and clean air may be identified, since pollutants may be injected when air masses sweep over sources such as cities or industrial facilities, and pollutants are recorded at distant regions from sources. Hence, the history of air masses needs to be investigated, for which purpose calculating their trajectory proves a suitable tool.

Two types of sites have been the object of such analyses. The first comprises remote areas where measurements are scarce or are assumed to be clean (Xu et al. 2011), and the second comprises areas that are especially polluted (Lv et al. 2015).

Although calculating trajectories complicates air mass analysis, this additional effort is necessary in certain situations, such as the precise determination of sources linked to singular events like volcanic eruptions (Karasiński et al. 2014) or pollen peaks (Lu et al. 2010). In fact, air mass trajectory calculation has been used in several applications and is a useful complement in atmospheric analyses (Pérez et al. 2015a).

Different models have been employed to determine air pollutant sources and atmospheric patterns such as the FLEXible TRAjectory model, FLEXTRA (Stohl 1999), recently used by Saini et al. (2014) to investigate air quality in Agra, India, or the METeorological data Explorer, METEX model (Zeng et al. 2010), applied by Kuramoto et al. (2008) to study snowfall in the Japanese Alps, and by Reddy et al. (2008) to analyse the characteristics of air

Editorial responsibility: M. Abbaspour.

✉ I. A. Pérez
iaperez@fa1.uva.es

¹ Department of Applied Physics, Faculty of Sciences, University of Valladolid, Paseo de Belén, 7, 47011 Valladolid, Spain

mass trajectories over the Bay of Bengal. However, the Hybrid Single Particle Lagrangian Integrated Trajectory Model, HYSPLIT (Draxler et al. 2014), is the most widely used model.

Isolated trajectories are frequently considered when investigating the source of certain atmospheric compounds such as aerosols (Freitas et al. 2009) or spores (Hermansen and Torp 1981). In similar cases, the origin of air masses may easily be determined (Egebäck et al. 2004). One advantage of this kind of analysis is the detailed tracing of the air mass trajectory evolution. One example is presented by Zhao et al. (2011) who, with a 10-day analysis, concluded that the arid regions of Central Asia were the main pollution source during this period at the remote site of Mount Bogda, China. Another important application is provided by Niemi et al. (2006), who considered six sampling periods to investigate background composition at a rural site in southern Finland, with the corresponding trajectories being grouped into two opposite directional patterns.

However, air mass trajectories form the main part of certain analyses. In such situations, different procedures should be considered to handle the significant amount of data provided by the models. Three methods are currently used: trajectory sector analysis, the potential source contribution function and trajectory cluster analysis (Li et al. 2012a).

This paper focuses on cluster analysis, which comprises a range of statistical techniques aimed at grouping similar objects. Its main advantage is that it replaces large numbers of observations with few groups. However, several difficulties are inherent in the process and the procedure is by no means easy since some elaboration is needed. Moreover, it is not a particularly fast method, especially when a lot of information is involved. Finally, results need to agree with observations. Continued use of this procedure in the future is ensured thanks to the varied applications of the HYSPLIT model, which also allows clustering of trajectories.

Bearing in mind these difficulties, the current study analyses different applications following the clustering procedure employed to form groups, the distance used to merge the air mass trajectories, the stopping rule that determines the end of the process and the final number of clusters considered. The main drawbacks involved are also discussed. Although reviews about air mass trajectories, such as Kulshrestha and Kumar (2014), are available, the current analysis emphasises the clustering procedure and shows its strengths and weaknesses. Since air mass trajectories may be applied in a range of fields, the present paper is not focused on specific processes or chemicals.

Clustering procedures

Since air trajectories are a kind of time series, certain clustering methods of time series may be used for trajectories (Warren Liao 2005). One early application of such a procedure was presented by Moody and Galloway (1988) to investigate the relationship between atmospheric flow patterns and precipitation composition in Bermuda. Clustering procedures may be divided into two kinds of groups, hierarchical and non-hierarchical methods.

Hierarchical methods

Donnelly et al. (2015) considered such a method to investigate the effects of long range transport on PM₁₀ and NO₂ concentrations in Ireland. They assumed an initial number of clusters equal to the total number of trajectories. They then merged two trajectories in one cluster following a minimum distance calculation and followed the same procedure until the last two clusters were merged. The process must be stopped at a certain stage if useful information is to be obtained.

Several similar procedures have been developed and used. Kalkstein et al. (1987) evaluated three clustering methods (Ward's, average linkage and centroid) for classifying meteorological data and reported noticeable contrasts among the final groupings. Ward's method (Ward 1963) produced groups with a similar size, its main disadvantage being that unequal observations may be found in the same cluster. The centroid method produced contrasting clusters, with one group being very large and many groups containing only single observations. This study concluded that the most realistic groupings and most appropriate combinations of data were obtained by the average linkage procedure.

Despite these drawbacks, Ward's method was used by Eneroth et al. (2003) to reveal the source regions and transport pathways affecting Svalbard. They observed high CO₂ concentrations attributed to transport over Europe and Siberia during winter, whereas low CO₂ concentrations were linked with trajectories from the Atlantic. Additionally, Liu et al. (2013) applied the same technique to identify the transport pathways responsible for elevated pollutant concentrations in urban Lanzhou, China. Hernández-Ceballos et al. (2015) selected this procedure to investigate the impact of weather conditions on airborne pollen levels in south-western Spain due to the lack of previous knowledge available on the number of clusters to describe pollen variability.

Cape et al. (2000) considered the average linkage method to classify trajectories at Mace Head, Ireland, following the prevailing weather conditions, although they



needed to exclude observations affected by local perturbations. Pérez et al. (2012) used the same method to analyse CO₂ transport during the night at a rural site in northern Spain. Seven clusters were selected whose ranges covered different directions and distances. Another application is presented by Han et al. (2015), who considered six groups of air masses and their ozone concentrations at an observatory in the East China Sea.

Hierarchical methods may be suitable with a low number of observations. However, they evidence one noticeable inconvenience, since calculation time may increase substantially when handling a high number of trajectories. Consequently, they must be replaced by alternative algorithms when large databases are used. Another disadvantage of these procedures is that reassigning each element to another cluster is not possible (Seibert et al. 2007).

Non-hierarchical methods

The k-means procedure is an example of this kind of clustering method (MacQueen 1967). It was considered by Stock et al. (2014) to investigate the sources of aerosols in the European Arctic. In this case, air masses from Europe are not important. However, transport from the eastern Arctic is linked to Arctic haze, and air masses from Siberia and the central Arctic were responsible for increased aerosol optical depth.

Dorling et al. (1992) detailed this procedure to cluster 3-day air mass trajectories reaching Eskdalemuir, south Scotland, with the aim of investigating the relationship between pollutant concentrations and synoptic meteorology. They used a large number of seed trajectories to begin the procedure. Each real trajectory was assigned to its closest seed trajectory. Average trajectories were calculated in each group, and groups were reconsidered until all the trajectories were correctly distributed. Two clusters with the closest average trajectories were then merged, and the procedure continued until the conclusion, which was determined by a stopping rule. This paper is the basis of numerous studies, although a brief comment is provided for only a few. Cristofanelli et al. (2013) identified five circulation patterns affecting Mt. Portella, Italy. They concluded that transport from the Mediterranean Sea resulted in decreasing O₃ concentrations. Shi et al. (2014) formed six groups of air masses affecting precipitation at the Huangshan mountain range, China and found that trajectories from Mongolia were linked to highly polluted rainwater. Dimitriou et al. (2016) considered seven categories of atmospheric trajectory patterns and linked them to mortality levels in England.

One method modifying this was suggested by Mattis (2001), who replaced the real seed trajectories of Dorling's procedure with synthetic seed trajectories, which lead to

fast convergence since they uniformly cover the whole spread of the back trajectories. This procedure was employed by Jorba et al. (2004) to investigate air masses reaching the Barcelona area, Spain, and their seasonal evolution.

Borge et al. (2007) considered a non-hierarchical algorithm and proposed a two-stage clustering approach to overcome the strong influence of trajectory length in the usual one-stage cluster analysis, where short trajectories are grouped, even if they come from a large variety of regions, whereas long trajectories are disaggregated, even if they come from the same region. In this procedure, short trajectory clusters were first identified, then grouped and finally re-analysed.

The main inconvenience of non-hierarchical procedures is the influence of seed trajectories, which are fixed during the process.

Finally, a combination of hierarchical and non-hierarchical approaches has been proposed in certain studies. Zhu et al. (2011) considered hierarchical clustering as the first stage when defining the number of clusters and their mean trajectories, which acted as seeds for the second stage, the k-means clustering stage.

The distance used

Distance measures the similarity or dissimilarity between trajectories. The Euclidean distance between each air mass trajectory and the cluster mean has often been used to group trajectories, such as in Rozwadowska et al. (2010), who investigated the influence of air mass trajectories on aerosol optical properties at Svalbard. Moreover, Dimitriou and Kassomenos considered this distance in several recent studies focusing on particulate matter. The first investigates its sources in two German cities (Dimitriou and Kassomenos 2014a), whereas a further two analyses identify particulate matter sources in five major cities in northern Europe: London, Paris, Hamburg, Copenhagen and Stockholm (Dimitriou and Kassomenos 2014b), and four large cities in southern Europe: Lisbon, Madrid, Marseille and Rome (Dimitriou and Kassomenos 2013).

One noticeable drawback of using only geographical coordinates in distance calculation is that similar shaped trajectories are merged, whereas some important variables are ignored. Mace et al. (2011) added the chemistry from carbon monoxide data to propose an algorithm aimed at being more informative and involving a low level of subjectivity.

Engler et al. (2012) used a general expression of the Euclidean distance including the following four variables: geographical latitude and longitude, height above ground and pseudo potential temperature, which were multiplied

by certain weights to make them equitable magnitudes. This procedure was used to study PM_{10} concentration in Leipzig, Germany, the conclusion being the convenience of local, national and international reduction measures to avoid exceedances of the daily limit value.

A similar equation was used by Pérez et al. (2015b) to investigate the seasonal evolution of the airflow pattern in the northern plateau of the Iberian Peninsula, although distances between corresponding points of the trajectories were calculated using spherical trigonometry. Five clusters were considered, three of which were associated with long range transport.

Makra et al. (2013) used the Mahalanobis metric to investigate transport patterns which impact on PM_{10} levels in two cities of Eastern Europe, Szeged, Hungary, and Bucharest, Romania. They found that long range transport played a key role in the PM_{10} concentration measured.

Wang et al. (2006) preferred the angle distance, since their analysis focused on the direction from which the air masses reached XiAn, China. This distance has been applied in several recent studies. Byčenkienė et al. (2014) considered the long range transport of aerosols in the south-eastern Baltic region and reported a significant contribution of southern Europe in winter and wildfires in spring. Yan et al. (2015) investigated the pollution recorded in Beijing, China, and concluded that the surrounding provinces to the south and south-west were the major sources.

The stopping rule

Once the clustering process has commenced, the main question is to decide when it should be stopped. Wilks (2011) stressed that selecting the best number of clusters is not obvious, since some subjectivity is required depending on the goals pursued.

In practice, several procedures have been proposed to select this optimum number of clusters. One possibility is based on the increase in the root-mean-square distance between clusters. Another option is the sharp decrease in the coefficient of determination, and the third is the number of clusters containing more than 3 % of total trajectories. Morgan et al. (2009) used these methods in their study into the chemical composition of sub-micron aerosols from north-western Europe and the north-east Atlantic.

Kumar et al. (2011) considered a sharp increase in the total spatial variance as a stopping rule. This quantity is the sum of the cluster spatial variances, which were calculated as the sums of squared distances between the endpoints of the cluster trajectories and the mean trajectory in each cluster. Piñero-García et al. (2015) used this procedure when analysing the influence of air masses reaching the south-east of Spain on aerosol radioactivity.

Another approach is the L-method suggested by Salvador and Chan (2005). In this procedure, the distance or an evaluation quantity of the clusters is represented versus the number of clusters. Three regions may be considered in the resulting curve: a region with a sharp slope to the left, a curved transition region or the “knee” of the graph, in the middle, and a nearly flat region to the right. The aim is to determine the number of clusters associated with the exact location of the knee, since it represents a balance between adjacent regions. The L-method fits these adjacent regions to straight lines, and its intersection determines the number of clusters. Kassomenos et al. (2010) used this procedure to identify air masses affecting Athens, Greece, and to compare three clustering methods.

Another procedure, albeit one which is scarcely used, is silhouette analysis (Rousseeuw 1987), whose peaks reveal optimum cluster size. This method was used by Notaro et al. (2013) to investigate dust sources reaching 13 stations in Saudi Arabia.

Final number of groups of trajectories

Although some subjectivity may be removed in the aggregation process, cluster analysis is not an objective method for classification. Sources of subjectivity are the algorithm selection, the choice of distance and the number of clusters.

One example which evidences that selecting the cluster number is not easy is given by Cheng et al. (2013), who investigated the impact of trans-boundary air pollutant transport on air quality at Guangzhou in southern China and tested numbers of clusters ranging from three to seven in order to choose the number that best represented transport pathways at the site.

Several classifications have been suggested to simplify the information provided by trajectories. The simplest do not involve algorithms. Classifications made with geographical or time criteria are widely used due to the lack of specific requirements. Additionally, plots are used to clearly communicate large amounts of information. One desirable property is that isolated trajectories should not constitute clusters. Moreover, mean trajectories of clusters should appear separately so as to clearly reveal the path followed by the air masses.

Geographical classifications

The simplest classification of air mass trajectories comprises only two groups, one example being the classification presented by Li et al. (2011a), which focused on the range and height of the trajectories reaching the measuring site at a mountain summit in east China.

Another classification which depends on geographical criteria is employed by Zhang et al. (2010), who investigated the impact of anthropogenic air pollutants over Xiaoyangshan Island, East China Sea. Air trajectories revealed that the most polluted day corresponded to flow from the continent.

Three groups of air mass trajectories were identified by Costa et al. (2014) in their analysis of Ethiopian precipitation sources. They found that the main source was the Indian Ocean, 80 %, followed by the Congo Basin, 13 % and the Northern Seas, 6 %, over a 12-year period, 2000–2011.

Ma et al. (2013) presented four groups of trajectories to investigate the transport of polycyclic aromatic hydrocarbons (PAH) in Lhasa, China, the first originating from Southeast Asia, the second from India, the third from Northwest China and the fourth from West Asia.

Temporal classifications

Limiting the time extension to certain intervals may sometimes simplify the interpretation of the trajectory plot. Pavuruli et al. (2010) considered just a few days in early winter, late winter and summer, to obtain three transport pathways reaching Chennai, on the southeast coast of India. Similarly, Prijith et al. (2012) used two periods of air trajectories, the first and second halves of the campaign, in their study of the role played by wind parameters in the long range transport of aerosols over the Bay of Bengal in winter.

One development from the preceding classification is formed by trajectories grouped by seasons following the climate at the site. Co et al. (2014) considered dry, wet and transitional seasons in northern Vietnam. Similarly, Yongjie et al. (2009) presented the distribution of trajectories reaching Mount Gongga, China, and concluded that northeast India is the source of fourteen elements in the particulate matter recorded in spring and winter.

Clustering procedures as an exploratory analysis tool

When clustering algorithms are used to explore air mass trajectories, practical reasons determine the final number of groups, since only the number of clusters that reveal a meaningful structure of air mass trajectories should be accepted.

According to the simplest classification procedures, a low number of clusters, such as three at most, are sometimes proposed since it proves easiest to explain. Li et al. (2011b) considered only three clusters whose mean trajectories presented contrasting ranges and directions to investigate the origin of aerosols in central Asia. Similarly,

Kang et al. (2013) investigated a long-lasting haze episode in Nanjing, China. In this study, they found three main air pathways, the first and least frequent was linked to long range transport processes. The second comprised the greatest frequency of trajectories. However, the third pathway, although with an intermediate number of trajectories, was the most noticeable, since many fire spots were linked with it, implying the partial influence of biomass burning in the haze pollution episode. Song et al. (2016) also used three clusters to investigate the long range transport of air pollutants from atmospheric aerosol analysis at Jeju Island, Korea.

A lower intermediate number of clusters, such as four or five, are chosen as a simple option that retains some degree of detail. Xin et al. (2016) presented a seasonal analysis where four clusters are considered to investigate transport pathways at Xining, China. Yang et al. (2014) investigated the differences between the particulate matter recorded at an urban and a rural site in Beijing, China. They considered five clusters and concluded that around 60 % of air masses reaching both sites originated from the south and south-east.

A higher intermediate number of clusters, from six to eight, provide a greater degree of detail, although complexity also increases.

One desirable result would be for the mean trajectories of the clusters to be clearly separated so as to yield information about different air mass origins. Several applications illustrate this result. Li et al. (2012b) presented six mean trajectories reaching Lushan Mountain, China, in a study of precipitation chemistry, five of them with varied directions. Hermanson et al. (2010) considered six clusters of air masses reaching Svalbard, Norway. Five of the mean trajectories were of a similar length, and their directions were similarly distributed. However, the mean trajectory of the sixth cluster corresponded to the shortest range and its direction did not differ from the rest. Park et al. (2007) used seven clusters in their study of the relationship between pollution sources and health outcomes in Boston, Massachusetts. Since the cluster number is not low, some overlapping is observed.

Another desirable result is that different ranges and directions should be covered by the mean trajectories of the clusters, such as in Pietruczuk and Jarosławski (2013), whose study into transport patterns in the Mazovia Region, Poland, considers five clusters. The shortest mean trajectory corresponds to local effects, whereas the longest comes from near Canada. Moreover, different ranges are obtained for trajectories in the continent and over the sea.

Masiol et al. (2012) used seven clusters linked to geographical regions to analyse the influence of air masses on particulate matter and PAH affecting Venice-Mestre, Italy. They reported the influence of the Po Valley on PM_{2.5}

concentrations, whereas the contribution to PAH was low for external sources. Three pairs of mean trajectories may be considered. Each pair is formed by trajectories with the same direction, but with a different range, the lowest frequencies corresponding to the longest ranges.

A high number of clusters provide a large degree of detail. However, interpreting the results may be more complex and frequency distribution may be formed by contrasted values. Im et al. (2013) used twelve clusters to investigate ozone levels in Istanbul, Turkey. They concluded that most transport was observed from the Eastern Europe and Mediterranean region, together with the recirculation over the city associated with high-pressure systems. Markle et al. (2012) proposed subjectively classifying a greater number than usual. They considered that around twenty clusters were desirable to prevent meteorologically distinct although spatially similar paths from being grouped.

Drawbacks

When a high number of clusters are used, one inconvenience is that some may overlap. Pongkiatkul and Kim Oanh (2007) used six clusters made up of 10-day back trajectories, which differed in length, shape, height and departure locations, to assess the possible contribution of long range transport from populated regions to particulate air pollution in Bangkok, Thailand. However, two pairs were formed by very close mean trajectories.

The same number was used by Xu et al. (2014) to demonstrate that the northern boundary of the Tibetan Plateau is affected by mineral dust from arid regions of northwest China, while urban areas to the east of the measurement site were the source of anthropogenic aerosols. In this case, two mean trajectories were similar in range and direction.

Katragkou et al. (2009) considered eight clusters to investigate the origin of PM₁₀ recorded in Thessaloniki, Greece. They found that high concentrations were recorded with north-eastern and southern flows superposed to biomass burning and Sahara dust events, respectively. Crossing some of the mean trajectories presented in this analysis evidences the noticeable mixing of air masses.

Clusters are sometimes presented separately, with each spaghetti plot showing the trajectories contained in each cluster, such as in Harpaz et al. (2014), who presented the relationship between the cool summer temperatures in the East Mediterranean and four circulation regimes.

However, including trajectory means is a very common approach. One example is the study by Petrou et al. (2015), who analysed the relationship between atmospheric patterns and heat-induced mortality in five regions in England. They found a link between human thermal stress and

atmospheric stagnation. Another example is the analysis by Dimitriou and Kassomenos (2015) to investigate air masses linked to episodes of tropospheric ozone in Ioannina and Athens, Greece. The information provided concerning the dispersion of trajectories and their geographical distribution are the main advantages of this kind of representation, with the impossibility of comparing mean trajectories being the major drawback. In order to overcome this inconvenience, in their analysis into transport of ⁷Be to Vladivostok, Russia, Neroda et al. (2016) presented spaghetti plots for the isolated clusters accompanied by spaghetti plots of all the air mass trajectories together with the cluster means.

One alternative to spaghetti plots is given by Kahl et al. (1997) in their analysis of the seasonal transport pattern of air masses reaching Summit, Greenland, between 1946 and 1989. They presented the mean trajectories surrounded by envelopes comprising the areas traversed by individual trajectories. Pérez et al. (2015b) represented the average trajectory bounded by its horizontal standard deviations to study air mass trajectories on the northern plateau of the Iberian Peninsula. A simpler choice was used by Dumka et al. (2013), drawing vertical bars over the mean line to represent the latitudinal spread of each cluster when analysing the evolution of air flow over Hyderabad, India and its relationship with observed aerosols. In contrast, Makra et al. (2010) considered a more complex representation with convex hulls that envelop the trajectories of each cluster. They used said analysis to investigate the atmospheric pathways influencing pollen levels in three European cities.

Conclusion

Although simple procedures based on geographical or temporal criteria are used to group air mass trajectories, cluster algorithms are commonly used. However, certain questions linked to the complexity and subjectivity of these methods may arise.

Due to the great number of variables involved when calculating air mass trajectories, hierarchical methods may be recommended when trajectories are short or their number is low. In contrast, non-hierarchical procedures are preferred for high numbers of trajectories, since their calculation time is shorter.

The Euclidean distance is a simple procedure to merge trajectories in clusters. Moreover, it is flexible enough to include additional variables other than geographical coordinates.

Although varied algorithms have been developed to determine the final number of clusters, the number chosen must agree with the interpretation of the airflow patterns. A



low number allows a simple explanation at the expense of the lack of detail. In contrast, large numbers may lead to clusters that are barely representative.

Although one desirable property is that directions and ranges of mean trajectories should be regularly distributed, in practice, this objective may be not so easy to achieve. Overlapping and crossing mean trajectories should be justified.

Scant information is provided by local trajectories, since they are normally associated with low wind speeds or stagnant conditions. In addition, their range is small and they are extremely curved and mixed. Mean trajectories of clusters with this kind of trajectories only reveal spatial extent, but are not illustrative of airflow direction.

Although spaghetti plots are widely used, these graphs should be complemented with a numerical statistics. Moreover, mean trajectories are usually considered as representative of clusters. However, the inclusion of spread, skewness and flatness of trajectory distribution in each cluster is barely used.

The application of clustering procedures should consider that eliminating certain drawbacks may prove difficult or even impossible, since assigning complex trajectories to a specific group may not be easy, particularly if the number of clusters is low, even though a numerical procedure is used. However, despite these constraints, these methods may produce results that are illustrative of airflow patterns. Several procedures should be assayed to select the method whose results may best be justified.

Finally, existing algorithms of clusters should be used when analysing air mass trajectories, taking into account the observations commented on in the current paper. However, efforts to develop new clustering procedures adapted to the specificities of air mass trajectories should be made so as to overcome the drawbacks of known methods.

Acknowledgments The financial support of the Ministry of Economy and Competitiveness, Spain and ERDF funds is gratefully acknowledged (grant numbers CGL2009-11979 and CGL2014-53948-P).

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- Borge R, Lumbreras J, Vardoulakis S, Kassomenos P, Rodríguez E (2007) Analysis of long-range transport influences on urban PM₁₀ using two-stage atmospheric trajectory clusters. *Atmos Environ* 41:4434–4450
- Byčenkienė S, Dudoitis V, Ulevicius V (2014) The use of trajectory cluster analysis to evaluate the long-range transport of black carbon aerosol in the south-eastern Baltic region. *Adv Meteorol* 2014:137694. doi:10.1155/2014/137694
- Cape JN, Methven J, Hudson LE (2000) The use of trajectory cluster analysis to interpret trace gas measurements at mace head, Ireland. *Atmos Environ* 34:3651–3663
- Cheng S, Wang F, Li J, Chen D, Li M, Zhou Y, Ren Z (2013) Application of trajectory clustering and source apportionment methods for investigating trans-boundary atmospheric PM₁₀ pollution. *Aerosol Air Qual Res* 13:333–342
- Co HX, Dung NT, Oanh NTK, Hang NT, Phuc NH, Le HA (2014) Levels and composition of ambient particulate matter at a mountainous rural site in northern Vietnam. *Aerosol Air Qual Res* 14:1917–1928
- Costa K, Russell J, Konecky B, Lamb H (2014) Isotopic reconstruction of the African humid period and Congo air boundary migration at lake Tana, Ethiopia. *Quat Sci Rev* 83:58–67
- Cristofanelli P, di Carlo P, Altorio AD, Dari Salisburgo C, Tuccella P, Biancofiore F, Stocchi P, Verza GP, Landi TC, Marinoni A, Calzolari F, Duchi R, Bonasoni P (2013) Analysis of summer ozone observations at a high mountain site in central Italy (Campo Imperatore, 2388 m a.s.l.). *Pure Appl Geophys* 170:1985–1999
- Dimitriou K, Kassomenos P (2013) The fine and coarse particulate matter at four major Mediterranean cities: local and regional sources. *Theor Appl Climatol* 114:375–391
- Dimitriou K, Kassomenos P (2014a) Decomposing the profile of PM in two low polluted German cities —mapping of air mass residence time, focusing on potential long range transport impacts. *Environ Pollut* 190:91–100
- Dimitriou K, Kassomenos P (2014b) Local and regional sources of fine and coarse particulate matter based on traffic and background monitoring. *Theor Appl Climatol* 116:413–433
- Dimitriou K, Kassomenos P (2015) Three year study of tropospheric ozone with back trajectories at a metropolitan and a medium scale urban area in Greece. *Sci Total Environ* 502:493–501
- Dimitriou K, McGregor GR, Kassomenos PA, Paschalidou AK (2016) Exploring winter mortality variability in five regions of England using back trajectory analysis. *Earth Interact* 20:27p. doi:10.1175/EI-D-15-0012.1
- Donnelly AA, Broderick BM, Misstear BD (2015) The effect of long-range air mass transport pathways on PM₁₀ and NO₂ concentrations at urban and rural background sites in Ireland: quantification using clustering techniques. *J Environ Sci Health Part A Toxic Hazard Subst Environ Eng* 50:647–658
- Dorling SR, Davies TD, Pierce CE (1992) Cluster analysis: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry-method and applications. *Atmos Environ Part A* 26:2575–2581
- Draxler R, Stunder B, Rolph G, Stein A, Taylor A (2014) Hysplit4 user Guide. http://www.arl.noaa.gov/documents/reports/hysplit_user_guide.pdf. Accessed 11 Apr 2016
- Dumka UC, Manchanda RK, Sinha PR, Sreenivasan S, Moorthy K, Suresh Babu S (2013) Temporal variability and radiative impact of black carbon aerosol over tropical urban station Hyderabad. *J Atmos Sol Terr Phys* 105–106:81–90
- Egeback AL, Wideqvist U, Järnberg U, Asplund L (2004) Polychlorinated naphthalenes in Swedish background air. *Environ Sci Technol* 38:4913–4920
- Eneroth K, Kjellström E, Holmén K (2003) A trajectory climatology for Svalbard; investigating how atmospheric flow patterns influence observed tracer concentrations. *Phys Chem Earth* 28:1191–1203
- Engler C, Birmili W, Spindler G, Wiedensohler A (2012) Analysis of exceedances in the daily PM₁₀ mass concentration (50 µg m⁻³) at a roadside station in Leipzig, Germany. *Atmos Chem Phys* 12:10107–10123



- Freitas MC, Pacheco AMG, Dionísio I, Vieira BJ (2009) Aerosol concentrations and remote sources of airborne elements over Pico mountain, Azores, Portugal. In: Kim YJ, Platt U, Gu MB, Iwahashi H (eds) Atmospheric and biological environmental monitoring. Springer, New York, pp 137–157
- Han J, Shin B, Lee M, Hwang G, Kim J, Shim J, Lee G, Shim C (2015) Variations of surface ozone at leodo ocean research station in the east China sea and the influence of Asian outflows. Atmos Chem Phys 15:12611–12621
- Harpaz T, Ziv B, Saaroni H, Beja E (2014) Extreme summer temperatures in the East Mediterranean-dynamical analysis. Int J Climatol 34:849–862
- Hermansen JE, Torp U (1981) *Erysiphe graminis* f.sp. *hordei* conidia of remote origin in the spring in Denmark. Grana 20:209–213
- Hermanson MH, Isaksson E, Forsström S, Teixeira C, Muir DCG, Pohjola VA, Van De Wal RSV (2010) Deposition history of brominated flame retardant compounds in an ice core from Høltedahlfonna, Svalbard, Norway. Environ Sci Technol 44:7405–7410
- Hernández-Ceballos MA, García-Mozo H, Galán C (2015) Cluster analysis of intradiurnal holm oak pollen cycles at peri-urban and rural sampling sites in southwestern Spain. Int J Biometeorol 59:971–982
- Im U, Incecik S, Guler M, Tek A, Topcu S, Unal YS, Yenigun O, Kindap T, Odman MT, Tayanc M (2013) Analysis of surface ozone and nitrogen oxides at urban, semi-rural and rural sites in Istanbul, Turkey. Sci Total Environ 443:920–931
- Jorba O, Pérez C, Rocadenbosch F, Baldasano JM (2004) Cluster analysis of 4-day back trajectories arriving in the Barcelona area, Spain, from 1997 to 2002. J Appl Meteorol 43:887–901
- Kahl JDW, Martinez DA, Kuhns H, Davidson CI, Jaffrezo JL, Harris JM (1997) Air mass trajectories to Summit, Greenland: a 44-year climatology and some episodic events. J Geophys Res 102:26861–26875
- Kalkstein LS, Tan G, Skindlov JA (1987) An evaluation of three clustering procedures for use in synoptic climatological classification. J Clim Appl Meteorol 26:717–730
- Kang H, Zhu B, Su J, Wang H, Zhang Q, Wang F (2013) Analysis of a long-lasting haze episode in Nanjing, China. Atmos Res 120–121:78–87
- Karasiński G, Posyński M, Bloch M, Sobolewski P, Małarzewski L, Soroka J (2014) Lidar observations of volcanic dust over polish polar station at Hornsund after eruptions of Eyjafjallajökull and Grímsvötn. Acta Geophys 62:316–339
- Kassomenos P, Vardoulakis S, Borge R, Lumbreras J, Papaloukas C, Karakitsios S (2010) Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories. Theor Appl Climatol 102:1–12
- Katragkou E, Kazadzis S, Amiridis V, Papaioannou V, Karathanasis S, Melas D (2009) PM₁₀ regional transport pathways in Thessaloniki, Greece. Atmos Environ 43:1079–1085
- Kulshrestha U, Kumar B (2014) Air mass trajectories and long range transport of pollutants: Review of wet deposition scenario in South Asia. Adv Meteorol 2014:596041. doi:10.1155/2014/596041
- Kumar S, Devara PCS, Manoj MG, Safai PD (2011) Winter aerosol and trace gas characteristics over a high-altitude station in the Western Ghats, India. Atmosfera 24:311–328
- Kuramoto T, Shah SK, Tanaka M, Suzuki K (2008) Chemical characteristics of snowpack due to differences in snowfall type in Japan Alps. Bull Glaciol Res 26:15–21
- Li Y, Wang Y, Ding A, Liu X, Guo J, Li P, Sun M, Ge F, Wang W (2011a) Impact of long-range transport and under-cloud scavenging on precipitation chemistry in East China. Environ Sci Pollut Res 18:1544–1554
- Li Z, Zhao S, Edwards R, Wang W, Zhou P (2011b) Characteristics of individual aerosol particles over Ürümqi Glacier No. 1 in eastern Tianshan, central Asia. China. Atmos Res 99:57–66
- Li M, Huang X, Zhu L, Li J, Song Y, Cai X, Xie S (2012a) Analysis of the transport pathways and potential sources of PM₁₀ in Shanghai based on three methods. Sci Total Environ 414:525–534
- Li Y, Tang J, Yu X, Xu X, Cheng H, Wang S (2012b) Characteristics of precipitation chemistry at Lushan mountain, East China: 1992–2009. Environ Sci Pollut Res 19:2329–2343
- Liu N, Yu Y, He J, Zhao S (2013) Integrated modeling of urban-scale pollutant transport: application in a semi-arid urban valley, Northwestern China. Atmos Pollut Res 4:306–314
- Lu X, Herrmann M, Mosbrugger V, Yao T, Zhu L (2010) Airborne pollen in the Nam Co Basin and its implication for palaeoenvironmental reconstruction. Rev Palaeobot Palynol 163:104–112
- Lv B, Liu Y, Yu P, Zhang B, Bai Y (2015) Characterizations of PM_{2.5} pollution pathways and sources analysis in four large cities in China. Aerosol Air Qual Res 15:1836–1843
- Ma WL, Qi H, Baidron S, Liu LY, Yang M, Li YF (2013) Implications for long-range atmospheric transport of polycyclic aromatic hydrocarbons in Lhasa, China. Environ Sci Pollut Res 20:5525–5533
- Mace A, Sommariva R, Fleming Z, Wang W (2011) Adaptive K-means for clustering air mass trajectories. Lect Notes Comput Sci 6936:1–8
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Le Cam LM, Neyman J (eds) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley, pp 281–297
- Makra L, Sánta T, Matyasovszky I, Damialis A, Karatzas K, Bergmann KC, Vokou D (2010) Airborne pollen in three European cities: detection of atmospheric circulation pathways by applying three-dimensional clustering of backward trajectories. J Geophys Res 115:D24220
- Makra L, Ionel I, Csépe Z, Matyasovszky I, Lontis N, Popescu F, Sümegey Z (2013) The effect of different transport modes on urban PM₁₀ levels in two European cities. Sci Total Environ 458–460:36–46
- Markle BR, Bertler NAN, Sinclair KE, Sneed SB (2012) Synoptic variability in the Ross Sea region, Antarctica, as seen from back-trajectory modeling and ice core analysis. J Geophys Res 117:D02113
- Masiol M, Centanni E, Squizzato S, Hofer A, Pecorari E, Rampazzo G, Pavoni B (2012) GC-MS analyses and chemometric processing to discriminate the local and long-distance sources of PAHs associated to atmospheric PM_{2.5}. Environ Sci Pollut Res 19:3142–3151
- Mattis I (2001) WP4, Compilation of trajectory data. In: Bosenberg J (ed) EARLINET: A European aerosol research lidar network to establish an aerosol climatology, compiled by Jens Bösenberg. Max-Planck-Institut für Meteorologie, Hamburg, pp 26–29
- Moody JL, Galloway JN (1988) Quantifying the relationship between atmospheric transport and the chemical composition of precipitation on Bermuda. Tellus Ser B 40:463–479
- Morgan WT, Allan JD, Bower KN, Capes G, Crosier J, Williams PI, Coe H (2009) Vertical distribution of sub-micron aerosol



- chemical composition from North-Western Europe and the North-East Atlantic. *Atmos Chem Phys* 9:5389–5401
- Neroda AS, Goncharova AA, Goryachev VA, Mishukov VF, Shlyk NV (2016) Long-range atmospheric transport Beryllium-7 to region the Sea of Japan. *J Environ Radioact* 160:102–111
- Niemi JV, Saarikoski S, Tervahattu H, Mäkelä T, Hillamo R, Vehkamäki H, Sogacheva L, Kulmala M (2006) Changes in background aerosol composition in Finland during polluted and clean periods studied by TEM/EDX individual particle analysis. *Atmos Chem Phys* 6:5049–5066
- Notaro M, Alkolibi F, Fadda E, Bakhrjy F (2013) Trajectory analysis of Saudi Arabian dust storms. *J Geophys Res D: Atmos* 118:6028–6043
- Park SK, O'Neill MS, Stunder BJB, Vokonas PS, Sparrow D, Koutrakis P, Schwartz J (2007) Source location of air pollution and cardiac autonomic function: trajectory cluster analysis for exposure assessment. *J Expos Sci Environ Epidemiol* 17:488–497
- Pavuluri CM, Kawamura K, Swaminathan T (2010) Water-soluble organic carbon, dicarboxylic acids, ketoacids, and α -dicarbonyls in the tropical Indian aerosols. *J Geophys Res D: Atmos* 115:D11302
- Pérez IA, Sánchez ML, García MA, Pardo N (2012) Spatial analysis of CO₂ concentration in an unpolluted environment in northern Spain. *J Environ Manage* 113:417–425
- Pérez IA, Artuso F, Mahmud M, Kulshrestha U, Sánchez ML, García MA (2015a) Applications of air mass trajectories. *Adv Meteorol* 2015 Article ID 284213
- Pérez IA, Sánchez ML, García MA, Pardo N (2015b) Analysis of air mass trajectories in the northern plateau of the Iberian Peninsula. *J Atmos Sol -Terr Phys* 134:9–21
- Petrou I, Dimitriou K, Kassomenos P (2015) Distinct atmospheric patterns and associations with acute heat-induced mortality in five regions of England. *Int J Biometeorol* 59:1413–1424
- Pietruczuk A, Jarosławski J (2013) Analysis of particulate matter concentrations in Mazovia region, central Poland, based on 2007–2010 data. *Acta Geophys* 61:445–462
- Piñero-García F, Ferro-García MA, Chham E, Cobos-Díaz M, González-Rodelas P (2015) A cluster analysis of back trajectories to study the behaviour of radioactive aerosols in the south-east of Spain. *J Environ Radioact* 147:142–152
- Pongkiatkul P, Kim Oanh NT (2007) Assessment of potential long-range transport of particulate air pollution using trajectory modeling and monitoring data. *Atmos Res* 85:3–17
- Prijith SS, Aloysius M, Mohan M, Beegum N, Krishna Moorthy K (2012) Role of circulation parameters in long range aerosol transport: evidence from winter-ICARB. *J Atmos Sol Terr Phys* 77:144–151
- Reddy LAK, Kulshrestha UC, Satyanarayana J, Kulshrestha MJ, Moorthy KK (2008) Chemical characteristics of PM₁₀ aerosols and airmass trajectories over Bay of Bengal and Arabian Sea during ICARB. *J Earth Syst Sci* 117:345–352
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
- Rozwadowska A, Zielinski T, Petelski T, Sobolewski P (2010) Cluster analysis of the impact of air back-trajectories on aerosol optical properties at Hornsund, Spitsbergen. *Atmos Chem Phys* 10:877–893
- Saini R, Singh P, Awasthi BB, Kumar K, Taneja A (2014) Ozone distributions and urban air quality during summer in Agra—a world heritage site. *Atmos Pollut Res* 5:796–804
- Salvador S, Chan P (2005) Learning states and rules for detecting anomalies in time series. *Appl Intell* 23:241–255
- Seibert P, Frank A, Formayer H (2007) Synoptic and regional patterns of heavy precipitation in Austria. *Theor Appl Climatol* 87:139–153
- Shi C, Deng X, Yang Y, Huang X, Wu B (2014) Precipitation chemistry and corresponding transport patterns of influencing air masses at Huangshan mountain in East China. *Adv Atmos Sci* 31:1157–1166
- Song JM, Bu JO, Kim WH, Kang CH (2016) Influences of long-range transported air pollutants on atmospheric TSP aerosol compositions at Jeju Island of Korea during 2011–2013. *Bull Korean Chem Soc* 37:626–631
- Stock M, Ritter C, Aaltonen V, Aas W, Handorff D, Herber A, Treffeisen R, Dethloff K (2014) Where does the optically detectable aerosol in the European arctic come from? *Tellus Ser B Chem Phys Meteorol* 66:21450
- Stohl A (1999) The FLEXTRA trajectory model version 3.0, user's guide. In: Lehrstuhl für Bioklimatologie und Immissionsforschung. University of Munich, Munich, Germany, pp 1–41
- Sun L, Shen B, Gao Z, Sui B, Bai L, Wang SH, An G, Li J (2007) The impacts of moisture transport of East Asian monsoon on summer precipitation in Northeast China. *Adv Atmos Sci* 24:606–618
- Wang YQ, Zhang XY, Arimoto R (2006) The contribution from distant dust sources to the atmospheric particulate matter loadings at XiAn, China during spring. *Sci Total Environ* 368:875–883
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
- Warren Liao T (2005) Clustering of time series data—a survey. *Pattern Recogn* 38:1857–1874
- Wilks DS (2011) Cluster analysis. *Statistical methods in the atmospheric sciences*, 3rd edn. Academic Press, Amsterdam/Boston, pp 603–616
- Xin Y, Wang G, Chen L (2016) Identification of long-range transport pathways and potential sources of PM₁₀ in Tibetan Plateau uplift area: case study of Xining, China in 2014. *Aerosol Air Qual Res* 16:1044–1054
- Xu Y, Zhang G, Li J, Chakraborty P, Li H, Liu X (2011) Long-range atmospheric transport of persistent organochlorinated compounds from south and mainland south-eastern Asia to a remote mountain site in south-western China. *J Environ Monit* 13:3119–3127
- Xu J, Wang Z, Yu G, Qin X, Ren J, Qin D (2014) Characteristics of water soluble ionic species in fine particles from a high altitude site on the northern boundary of Tibetan Plateau: mixture of mineral dust and anthropogenic aerosol. *Atmos Res* 143:43–56
- Yan R, Yu S, Zhang Q, Li P, Wang S, Chen B, Liu W (2015) A heavy haze episode in Beijing in february of 2014: characteristics, origins and implications. *Atmos Pollut Res* 6:867–876
- Yang Y, Yu Y, Zhou R, Ma Z, Ren L, Zhang L, Di Y (2014) Elemental compositions and size distributions of PM between urban and rural site in Beijing during the spring of 2012. *Asian J Chem* 26:3180–3186
- Yongjie Y, Yuesi W, Tianxue W, Wei L, Ya'nan Z, Liang L (2009) Elemental composition of PM_{2.5} and PM₁₀ at Mount Gongga in China during 2006. *Atmos Res* 93:801–810
- Zeng J, Matsunaga T, Mukai H (2010) METEX—a flexible tool for air trajectory calculation. *Environ Model Softw* 25:607–608
- Zhang Y, Yu Q, Ma WC, Chen LM (2010) Influence of continental outflows on marine aerosols over Xiaoyangshan Island in the



- East China Sea: an episode of heavy air pollution. *Terr Atmos Ocean Sci* 21:939–953
- Zhao S, Li Z, Zhou P (2011) Ion chemistry and individual particle analysis of atmospheric aerosols over Mt. Bogda of eastern Tianshan mountains. *Central Asia Environ Monit Assess* 180:409–426
- Zhu L, Huang X, Shi H, Cai X, Song Y (2011) Transport pathways and potential sources of PM_{10} in Beijing. *Atmos Environ* 45:594–604

