CLUSTER ANALYSIS OF BACKTRAJECTORIES ARRIVING AT THE BARCELONA AIR BASIN

Jorba[1], O., Pérez [1], C., Baldasano [1], J.M., Rocadenbosch [2], F., Lopez [2], M. A.

Laboratorio de Modelización Ambiental

(1) Universitat Politècnica de Catalunya (UPC). Laboratory of Environmental Modelling.

(2) Universitat Politècnica de Catalunya (UPC). Dep. of Signal Theory and Communications, Lidar Group.

Introduction

Trajectory and cluster analysis have been increasingly employed to study the movement of air parcels carrying pollutants from sources situated at long distances. Cluster analysis is a statistical tool which can be used to find out relationships between the large-scale weather regime patterns and the pollution climatology of a site. This method classifies a large set of trajectories into dominant groups called clusters.

In the frame of the European Aerosol Research Lidar Network (EARLINET), lidar vertical profiles are derived from routine measurements in order to characterize the horizontal and vertical distribution of aerosols over Europe. The cluster classification of backtrajectories will serve to group lidar measurements of a site, in order to investigate the dependence of the optical properties of aerosols on the origin and the pathways of the air masses.

The atmospheric trajectories, which are used for the EARLINET project, are calculated by the German Weather Service for all EARLINET lidar sites for two arrival times each day, which correspond approximately to the times of the routine lidar observations at noon and at sunset. The analytical as well as the prognostic trajectories are 4-day backward trajectories and are calculated from the wind fields of the global numerical weather prediction model of the German Weather Service (Kottmeier and Fay, 1998). They are available since May 2000 for all EARLINET participants.

A cluster algorithm has been implemented to analyze the backtrajectories arriving at the Barcelona air basin. Results for the 500, 700, 850 and 975 hPa EARLINET backtrajectories are presented.

Trajectory data

Trajectory model

Data Base: European numerical weather prediction model of the German Weather Service

Spatial resolution: 0.5º

Time resolution: 6 h

Trajectory data set

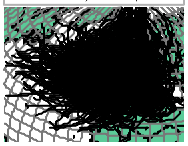
Trajectory length: 4 days

Arrival times: 13 UTC, 19 UTC each day

Arrival pressure levels: 975, 850, 700 and 500 hPa

Period: May 2000 – April 2002

4 days backtrajectories arriving at Barcelona at 850 hPa from may 2000 to april 2002



Cluster algorithm

The algorithm adopted to perform the cluster analysis is based on Dorling et al. (1992). It's a non-hierarchical clustering algorithm designed for large databases due to its requirement for relatively small computational storage.

Step 1- It generates a large number of "seed" trajectories (n) covering the spread of the real trajectories. The algorithm allows the use of real seed trajectories, as Dorling et al. (1992), or synthetic seed trajectories, as Bösenberg et al. (2001).

Step 2- It assigns each trajectory to that seed which is closest in distance. The distance calculation is based on the Haversine formula (Sinnott, 1984) concerning only latitude and longitude. It recalculates the new seeds (centroids) by averaging the trajectories of each cluster.

Step 3- It checks that each trajectory is in the right cluster, and reassigns the trajectories if necessary. Then, recalculates the centroids after completing the check. Several passes may be needed till all trajectories are correctly assigned.

Step 4- It calculates the root mean square deviation (RMSD) of each real trajectory from its cluster mean and sum these to give total RMSD.

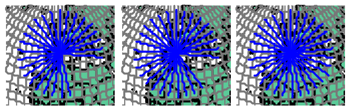
Step 5- It Merges the two closest clusters and calculates the new centroids. It checks as in step 3 that all trajectories are correctly assigned.

Step 6- Repeat step 4 to find the RMSD for **n**-1 clusters.

Step 7- Repeat step 5 consecutively to find the RMSD for 1 – **n** clusters.

Step 8- Repeat several times all the process (step 1 to 7) with different initial conditions (e.g., **n**-1, **n**-2, **n**-3, … , seed trajectories) to check the convergence in the solution.

Step 9- Finally, plot the percentage change in RMSD with cluster number.



Synthetic seed trajectories initial conditions. The figures show 30, 27 and 25 seed trajectories for the initialization of the process. The synthetic seed trajectories are removed from the southeast quadrant in a clockwise direction (Bösenberg et al., 2001)

Barcelona's location

Barcelona is located at the east coast of the Iberian Peninsula. The major orographic features that influence the flows of the Barcelona air basin are the Pyrenees and the Ebro valley. The Pyrenees range from 2000 m to 3000 m acting as a natural barrier to the flows producing important orographic forcings in the low troposphere. The Ebro valley has a length of 350 km, channeling the flows of the Cantabric sea to the Mediterranean or vice versa. Typical flows reinforced by these features are the Tramontana (N) and Mistral (NW) winds.

Results

CLUSTER 1

CLUSTER 2

CLUSTER 3

CLUSTER 4

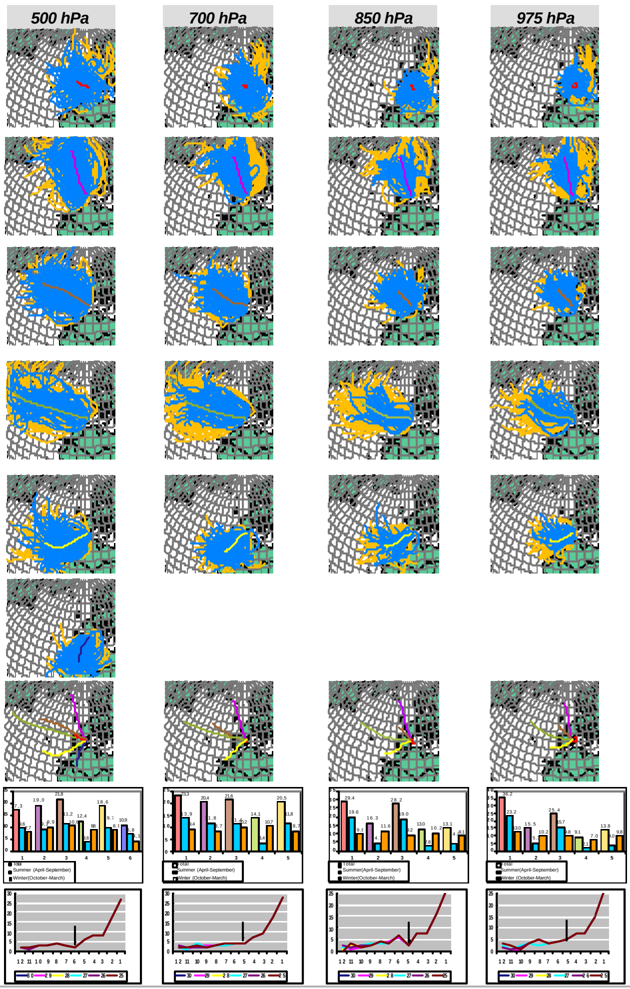
CLUSTER 5

CLUSTER 6

CENTROIDS

STATISTICS

RMSD



Cluster 1: Regional Western Mediterranean recirculations

- Major cluster at 700, 850 and 975 hPa (23.2 %, 29.4 % and 36.2%).
- Increasing percentage of trajectories with higher pressures.
- Higher percentage of summer trajectories.
- Includes the typical summer barometric swamp situations and anticyclonic situations with weak pressure gradient related to decrease in air quality.
- Includes some winter eastern-northeastern trajectories with continental origin.
- Advection of Mediterranean subtropical maritime air masses.

Cluster 2: Northerly flows

- Includes mainly fast trajectories from NW directions.
- Includes also winter NE trajectories.
- The percentage of summer trajectories, decreases at lower levels (influenced by orography).
- Advection of artic maritime (N), polar maritime (NW) and polar continental (NE) air masses.

Cluster 3: Western-northwestern slow advections

- Major cluster at 500 hPa, and the second at 700, 850 and 975 hPa.
- Similar percentage of summer and winter trajectories.
- Slow trajectories.
- Influence of orography at low levels. Canalization by the Ebro valley and the Pyrenees.
- Advection of polar maritime air masses.

Cluster 4: Western fast advections

- Mainly includes winter trajectories.
- Typical winter westerly advections.
- It is the minor cluster in number of trajectories.
- Advection of Atlantic subtropical maritime air masses.

Cluster 5: Southwestern advections

- Higher number of winter trajectories at 975 and 850 hPa.
- Related to some Saharan dust intrusions.
- Advection of tropical maritime air masses.

Cluster 6: South advections

- Only 500 hPa cluster.
- Higher number of summer trajectories.
- Includes trajectories from the north of Africa. Related to Saharan dust intrusions at high levels.
- Advection of tropical continental air masses.

Arrival pressure	*500 hPa*			*700 hPa*			*850 hPa*			*975 hPa*		
Cluster	Total (%)	S (%)	W (%)	Total (%)	S (%)	W (%)	Total (%)	S (%)	W (%)	Total (%)	S (%)	W (%)
1 Reg.	17,3	9,6	7,7	23,3	13,9	9,4	29,4	19,6	9,8	36,2	23,2	13,0
2 N	19,0	9,1	9,9	20,4	11,8	8,7	16,3	4,7	11,6	15,5	5,3	10,2
3 W-NW	21,8	11,2	10,6	21,6	11,4	10,2	28,2	19,0	9,2	25,4	15,6	9,8
4 W	12,4	3,6	8,8	14,1	3,4	10,7	13,0	2,8	10,2	9,1	2,1	7,0
5 SW	18,6	9,7	8,9	20,5	11,8	8,7	13,1	4,0	9,1	13,8	4,0	9,8
6 S	10,9	6,8	4,1									

Once the evolution of the percentage change in RMSD is plotted against number of clusters, one can appreciate a sudden increase of the curve. This is interpreted as the merging of clusters of trajectories which are significantly different in terms of wind direction and speed.
For the 500 hPa backtrajectories the plot shows a sudden increase at 5 clusters, indicating an optimum number of 6 clusters with the methodology applied.
For the 700, 850 and 975 hPa the process drives to a 5 cluster solution.

Conclusions

A package to perform cluster analysis have been developed based on Dörling et al. (1992). The algorithm has been applied to EARLINET backtrajectories arriving at the Barcelona air Basin site. Data from May 1st of 2000 to April 30th 2002 have been used.

Results show five clusters at 975, 850, 700 hPa and six at 500hPa. The clusters are regional western Mediterranean recirculations, northerly flows, northwestern slow advections, western fast advections, southwesterly flows and for the 500 hPa data a group of southern flows. The influence of the complex orography of the region with the Pyrenees mountainous range and the Ebro valley as the more important orographical features is detected in the cluster solution of different levels.

Low levels present more influence of the canalizations by the Ebro valley and the Pyrenees mountainous range.

The most dominant situations at low levels are the recirculations over the western Mediterranean, specially in summer. The northwestern situations present an important frequency, being the second more usual group at 975 and 850 hPa. In height, 500 and 700 hPa, the trajectories are more egurally distributed within all groups.

For the Barcelona air basin, recirculations and northwestern flows are more usual than zonal flows. The southwestern and south advections are more important at 700 and 500 hPa.

Acknowledgements

The authors acknowledge the support of the EARLINET project by the European Commission under grant EVR1-CT1999 -40003 and the projects IMMPACTE and CICYT REN2000-1754-C02. The atmospheric trajectories used in this study were calculated by the German Weather Service. Thanks to Inna Mattis for her useful comments and help.

References

Bösenberg et al. (2001) The German Aerosol Lidar Network: Methodology, Data, Analysis, ISSN 0937-1060, 62-67.

Dorling, S.R., Cavies, T.D., Pierce, C.E. (1992) Cluster analysis: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry-Method and applications, Atm. Env., 26A, 2575-2581.

Kottmeier, C. and Fay, B. (1998) Trajectories in the antarctic lower troposphere JGR, 103(D9), 10947–10959.

Sinnott, R.W. (1984) Virtues of the Haversine, Sky and Telescope, v. 68, 2, 159.

Corresponding author: José Maria Baldasano. Laboratory of Environmental Modelling. Dpt Proyectos de Ingeniería, ETSEIB Universitat Politècnica de Catalunya (UPC) Avda. Diagonal 647, P 10, 08028 Barcelona, Spain. Phone +34 934011746. Fax +34 933340255. E-mail address: jose.baldasano@upc.es