EMS7/ECAM8 Abstracts,
Vol. 4, EMS2007-A-00545, 2007
7th EMS Annual Meeting / 8th ECAM

On the methodology of backtrajectory cluster analysis. Origin of air masses and their contribution to PM levels

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

SCOLAb, Física Aplicada, Universidad Miguel Hernández, E-03202, Elche, Spain

A k-means cluster analysis of 96 hour trajectories arriving in Elche at 3000, 1500 and 500m for the 7-year period $2000 - 2006$ has been performed to describe the main flows arriving in SE Spain and to relate them to suspended particulate matter (PM10 and PM2.5) values and meteorological variables in this area.
Backtrajectories arriving at 12 UTC were computed using the HYSPLIT model v.4 with the FNL meteorological data. We have utilized two frequently used clustering procedures, that differ on the type of initial centroids selected to begin the clustering process: (synthetic) radial trajectories (Mattis, 2001) starting in Elche and randomly chosen real trajectories (Dorling et al., 1992).
The subjectivity in the selection of the appropriate number of clusters is reduced using the percentage change in total RMSD (i.e. the sum of the Root Mean Square Deviation of every cluster) when the number of clusters is reduced from k to $k - 1$ by merging the two closest clusters. Unlike these methods, we define this number as the smallest quantity that corresponds to the smallest total RMSD change.
These approaches gave identical results for 3000 and 1500m, and quite similar solutions for 500m. Although this would show the robustness of the procedures, they do not necessarily lead to the smallest total RMSD, due to a lack of sampling of the trajectories' ensemble (especially for small number of clusters) when selecting the initial seeds. The calculation of $100,000$ clustering analyses for each k, taking the initial seeds from randomly chosen real trajectories, provides smaller total RMSDs and hence better solutions. Moreover, this last method obtains a different number of clusters for 3000 and 1500m than the other referred procedures when the percentage change in total RMSD to extract the appropriate number of air mass types is applied.
We have considered as best solution the one with the smallest RMSD; the 96h back-

trajectories arriving at 3000, 1500 and 500m are found to be clustered into 6, 5, and 6 groups, respectively. The Dorling and Mattis' classifications coincide with this method in the 75, 83 and 65% of the trajectories for 3000, 1500 and 500m, respectively.
For the trajectories arriving at 3000m the southwesterly flows cluster is the major one. For 1500 and 500m there is an elevated occurrence of slow flows and recirculations due to situations with a low baric gradient that last several days.
There are statistically significant differences on PM and meteorological variables according to the identified clusters. Air masses influence atmospheric parameters like precipitation, mixing height or surface wind speed, these influencing PM levels as well.

Work supported by MEC under the CGL2004-04419/CLI (RESUSPENSE) project.

Dorling, S. R., Davies, T. D. Pierce, C. E. (1992). Atmos. Environ., 26a, 2575-2581.
Mattis, I. (2001). http://lidarb.dkrz.de/earlinet/scirep1.pdf