Cluster Analysis Applied to Optical Particles Counter Data Collected in Apulia Region

L. Trizio[1], M. Amodio[1], E. Andriani[1], G. de Gennaro[1], P.R. Dambruoso[1], A. Demarinis Loiotile[1], A. Di Gilio[1], C.M. Placentino[1], M. Tutino[1]
[1]Chemistry Department, Bari University, via Orabona, 4, 70126 Bari
livia.trizio@uniba.it

Abstract

Air quality data were collected in Bari, South of Italy, during 2008 by means of a prototype consisted of different instruments, such as a Swam dual-channel sampler, an Optical Particle Counter Monitor (20 channels), a sonic anemometer and a Planetary Boundary Layer Mixing monitor.

Attention was focused on OPC hourly data; cluster analysis was applied in order to reduce the number of available channels and to improve Particulate Matter data interpretation to better identify and distinguish long range transport, regional and other local contributions on PM.

.

Introduction

Several epidemiological studies have shown the negative effects of air pollution on human health, which range from respiratory and cardiovascular disease to neurotoxic effects, and cancer. Most recent investigations have been focused on health toxicological features of Particulate Matter (PM) and its interactions with other pollutants: it was found that fine particles (PM2.5) could be an effective media to transport these pollutants deeply into the lung and to cause many kind of reactions which include oxidative stress, local pulmonary and systemic inflammatory responses [1].

In this work, the results obtained by applying an integrated approach in PM monitoring will be shown. A prototype, located close to Chemistry Department of University of Bari (Bari, Apulia Region, South of Italy) was assembled: it consists of several kinds of instruments, such as a Swam dual-channel sampler, an OPC Monitor (20 channels), a sonic anemometer and a PBL Mixing monitor. Attention was focused on OPC hourly data collected from January 2008 to December 2008. Cluster analysis was applied in order to reduce the number of available channels and to improve the data interpretation. In fact, this information is useful to evaluate which sources contribute to PM levels. Moreover, result interpretation was confirmed by data obtained from satellites (Modis) and remote sensing (Aeronet); in this way, it will be possible to identify and distinguish long range transport, regional and other local contributions on Particulate Matter.

Materials & Methods

OPC monitor (FAI Instruments) is a real optic multichannel particles counter working in time able to characterize the granulometric fractions beyond 0.3 μm with spectrum given up to 20 granulometric channels. It allowed the characterization of the temporal trends of the particulate matter granulometric distribution as a support and integration to information about air quality.

Cluster analysis groups data objects based only on information found in the data that describes the objects and their relationships. The goal is that the objects within a group be similar (or related) to one other and different from (or unrelated to) the objects in other groups. The greater the similarity within a group and the greater the difference between groups, the better or more distinct the clustering. [2, 3]. Cluster toolbox implemented in Matlab (version 7.0.4) was used to perform the cluster analysis. The method used to generate the hierarchical cluster tree was the average linkage, that uses the average distance between all pairs of objects in clusters.

Results

OPC hourly data collected from January 2008 to December 2008 were used to perform cluster analysis in order to reduce the available channels for the interpretation of PM sources and concentrations. Granulometric range were selected as variables for clustering (1: 0.30-0.32 μm; 2: 0.320.35μm; 3: 0.350.38 μm; 4: 0.38-0.41 μm; 5: 0.41-0.45 μm; 6: 0.45-0.49 μm; 7: 0.49-0.53 μm; 8: 0.53-0.68 μm; 9: 0.68-0.74 μm; 10: 0.74-0.94 μm; 11: 0.94-1.21 μm; 12: 1.21-1.54 μm; 13: 1.54-1.98 μm; 14: 1.98-2.53 μm; 15: 2.53-3.23 μm; 16: 3.23-4.13 μm; 17: 4.13-5.28 μm; 18: 5.28-6.75 μm).

Before processing, a column normalization was performed on data.

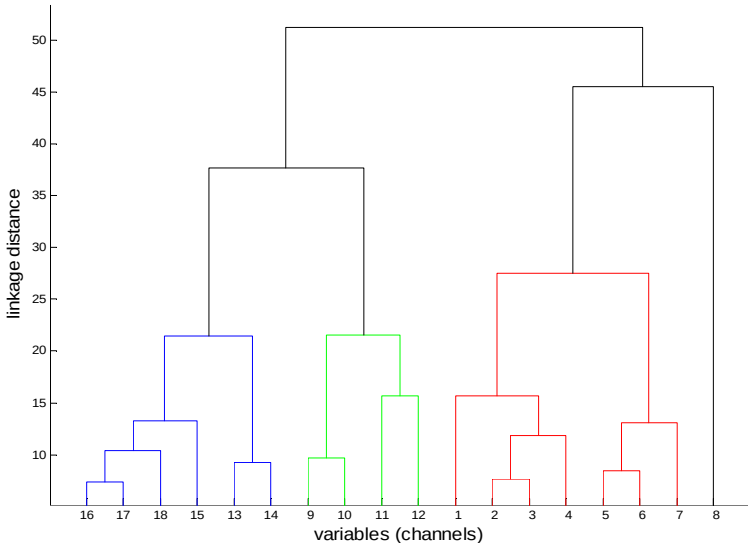

Fig. 1; Cluster analysis to OPC data collected in Apulia Region (South of Italy) from January 2008 to December 2008.

The results shown in Fig. 1 allow to identify five clusters at the average distance linkage equal to 23: 1: 0.30-0.41 μm; 2: 0.41-0.53 μm; 3: 0.53-0.68 μm; 4: 0.68-1.54 μm; 5. > 1.54 μm. The clusters will be confirmed by taking into account other information about PM data (PM concentrations, natural radioactivity, meteorological conditions, satellite and remote sensing data); in particular, the variable (channel) 8 will be considered because it does not group with the closest variables.

Conclusions

OPC data collected in Bari during 2008 were analyzed in order to obtain information useful to PM sources and concentration.

These data were stored in 18 channels; hourly mean data were used to perform cluster analysis. Preliminary results allowed to identify 5 different clusters that will be confirmed by means of other statistical methods. Moreover, additional information about PM data and metereological conditions, could be useful to improve the interpretation of statistical results.

This work was supported by the Strategic Project PS_122 founded by Apulia Region.

References

1) N. Künzli, L. Perez, Evidence based public health - the example of air pollution, *Swiss Med. Wkly.*, 139(17-18), (2009) 242-250

2) P. Arabie, L. J. Hubert, G. D. Soete, An overview of combinatorial data analysis. In P. Arabie, L. J. Hubert, and G. D. Soete, *Clustering and Classification*, 188-217, World Scientific Publishing Company, (1996) ISBN 9810212879

3) B. S. Everitt, S. Landau, M. Leese, *Cluster Analysis*, 4th edition, Arnold Publishers, London, (2001), ISBN 0-340-761199