

Functional approach to land classification according to air quality legislation ¹

S. Bande^{2,3}, S. Ghigo² and R. Ignaccolo²

²Dip. di Statistica e Matematica Applicata “Diego de Castro”, Università di Torino

³Area Previsione e Monitoraggio Ambientale, ARPA Piemonte

email:{bande, ghigo, ignaccolo}@econ.unito.it

Riassunto: Il Decreto Legislativo n. 351 del 4 agosto 1999, recepimento italiano della Direttiva 96/62/CE, richiede la suddivisione del territorio in differenti zone sulla base della valutazione della qualità dell’aria ai fini di predisporre adeguate politiche di azione, risanamento o mantenimento. Per tale classificazione si propone l’approccio funzionale applicato ai dati relativi agli inquinanti atmosferici a maggiore criticità, ottenuti dal sistema modellistico implementato dall’ARPA Piemonte.

Keywords: land classification, functional cluster analysis, air pollution

1. Introduction and data

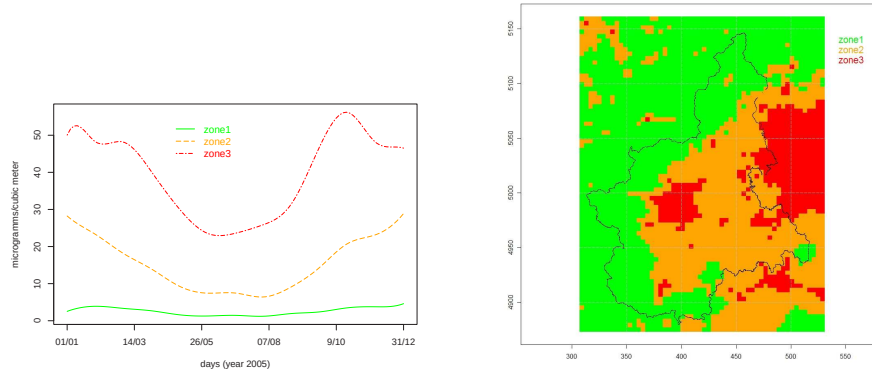
European and Italian legislation (Dir. 96/62/CE and D. Lgs. 351/99 art. 6) imposes land classification in relationship with air quality features, to distinguish zones that need actions or only hold their conditions. Thus, a “zoning” of Piemonte region through a functional approach is proposed, using information about hourly concentration levels for the biggest critical atmospheric pollutants (CO , SO_2 , PM_{10} , NO_2/NO_x e C_6H_6) in 2005. Data are produced using a three-dimensional modelling system (Bande *et al.* (2007)) which simulates air pollutant emission, transport, diffusion and transformation. It provides concentration fields for the main atmospheric pollutants on a hourly basis, over a regular grid that has an horizontal resolution of 4 km and covers Piemonte, neighbouring Italian Regions and foreign countries. Hence available data are not observed but artificial, however the european law allows their use in the air quality assessment. These pollutant concentrations are converted to functional data for every grid point and so these points are classified by functional cluster analysis (*FCA*).

2. Method and results

Pollutant time series can be considered as realisations of continuous processes that are recorded in discrete time. The conversion from discrete to functional data allows to reduce even thousands of observations to a few coefficients preserving their temporal patterns. It involves smoothing, using linear combinations of B-spline functions (Ramsay and Silverman (2005)). Then the set of curves is summarized by a set of coefficient vectors of $\mathbb{R}^{d+K+1}$, where K is the number of knots and d the degree of B-splines. Degree is fixed to 3 because pollutant time series are well fitted by a cubic spline; knots are located at

¹Work partially supported by PRIN project n.2006131039 and Regione Piemonte project CIPE 2004

Figure 1: NO_2 in 2005: Medoids of clusters (left), clustering results (right)



the end of every month since the data are equally spaced and according to pollutant time profiles. Thus we have functional data that can be classified by clustering procedures. Differently by the classical *FCA* (Abraham *et al.* (2003)), the non hierarchical algorithm PAM (Partitioning Around Medoid) is used to cluster objects (Kaufman and Rousseeuw (1990)). A typical feature of PAM is that it provides a graphical representation of the obtained clusters by so-called *silhouettes* showing which objects lie well within the group and which ones merely hold an intermediate position. Moreover PAM suggests the optimal number of clusters, but here we follow the regional law imposing a classification in three clusters. For NO_2 pollutant, Figure 1 shows cluster medoids (left) and the three resulting zones in Piemonte. Medoids' curves are representative of different concentration levels and time profiles of NO_2 in the three zones, that are different also for land use and morphology.

Concluding, the parameterization of functions using B-splines can replace the classical analysis by synthetic indicators (as average levels, standard deviations, quantiles) to respect the functional form of pollutant concentrations through time. Synthesizing features contained within time series in a little number of coefficients, we preserve in the “zoning” most information about pollutant concentration levels and time profiles. This approach is applied to functional data of the biggest critical atmospheric pollutants in Piemonte. Thus, it is possible to compare the resulting “zonings” for different pollutants and to obtain a joint classification of municipalities, as law requires.

References

- Abraham C., Cornillon P.A., Matzner-Løber E. and Molinari N. (2003) Unsupervised curve clustering using B-splines, *Scand. J. Statist.*, 30, 581–595.
- Bande S., Clemente M., De Maria R., Muraro M., Picollo M., Arduino G., Calori G., Finardi S., Radice P., Silibello C. and Brusasca G. (2007) The modelling system supporting Piemonte region yearly air quality assessment, *UAQ 2007, Cipro, March 2007*.
- Kaufman L. and Rousseeuw P.J. (1990) *Finding Groups in Data. An introduction to cluster analysis*, Wiley.
- Ramsay J.O. and Silverman B.W. (2005) *Functional data analysis*, Springer.