Accounting for complex environmental exposure situations: a classification approach

Benoît Lalloué[1,2,3,4], Jean-Marie Monnez[3,4], Cindy Padilla[1,2], Wahida Kihal[1], Denis Zmirou-Navier[1,2,5], Séverine Deguen[1,2]

[1] EHESP Rennes, Sorbonne Paris Cité, France
[2] Inserm, UMR 1085-IRSET (Institut de recherche sur la santé l'environnement et le travail), France
[3] Lorraine University, CNRS UMR 7502, Institut Elie Cartan, France
[4] Lorraine University, INRIA, CNRS UMR7502, BIGS (INRIA Nancy - Grand Est / IECN), France
[5] Lorraine University Medical School, France

Background: Everyone is subject to environmental exposures from various sources, with negative health impacts (air, water and soil contamination, noise…) or with positive effects (e.g. green space). Studies considering such complex environmental settings in a global manner are rare.

Aims: We propose to use data mining techniques to create a composite exposure index with a data-driven approach, in view to assess the environmental burden experienced by populations. We illustrate this approach on a large French metropolitan area.

Methods: The study was carried out in the Great Lyon area (France, 1,260,348 inhabitants in 2007, 527 km²) at the census block group (BG) scale. Indicators on NO_2 annual concentrations, noise levels, proximity to green spaces and to industrial plants, and road traffic were synthetized using Multiple Factor Analysis (MFA), which allows to explore the relations between exposures and BGs without a priori knowledge, and to synthetize indicators of different types. Hierarchical clustering was then used to create BG classes.

Results: The four first components of the MFA explained respectively 23, 14, 13 and 12% of the total variance. Clustering in 3 classes group: 1) BGs far from industries, greener and with less noise and air pollution than the average; 2) BGs close to industries with less green spaces and air pollution than the average; and 3)BGs far from industries but with higher levels of noise, air pollution and traffic. Greater numbers of classes were tested in order to assess a variety of clustering.

Conclusions: We present a data driven approach using data mining techniques, which seem overlooked for cumulative exposure assessment in complex environmental settings. Although it cannot be applied directly for risk or health effect assessment, the resulting index can help to identify hot spots of cumulative exposure, to prioritize urban policies or to compare the environmental burden across study areas in an epidemiological framework.