Multivariate statistical analysis for priority area localization in air quality management plans

C. Trozzi & R. Vaccaro
*Techne Srl., Via N. Zabaglia 3, 00153 Roma, Italy*

Abstract

In the paper the methodology established to locate priority areas for remediation actions in air quality management plans is reported. For classification of areas within the regions an innovative methodology based on multivariate statistical analysis techniques was developed. Particularly automatic classification criteria such as cluster analysis combined with simple statistical procedures were used.

In the methodology environmental sensibility and pollutants loads are evaluated and combined to obtain a classification of the areas based on remediation action priority within the regional territory. For priority areas localization two methodology are introduced: the first one takes into account overall sources, the second one keeps distinct area/line sources from point sources. Two different regional case studies are discussed and evaluated. Finally, balances and future studies direction are described.

Introduction

The following activities are necessary in the frame of the preparation of an Air Quality Management Plan:
- emissions inventory,
- use of models to estimate emissions in particular topics (road transport, airports, forests, ports and navigation lines, forest fires),
- air quality and meteorological data analysis,
- projection of emissions without emissions reduction measures,
- classification of territory and analysis of priority in remediation actions,
- planning of measures to reduce emissions,
- definition of scenario of reduction,
projection of emissions in the plan scenario,
use of air quality dispersion and photochemical models in the actual situation and in future projection with and without plan application.

These activities were recently reviewed [1], while specific aspects were discussed in the past also in Air Pollution conference series [2, 3, and 4]. In this paper we discuss one of most important tasks of air quality management: classification of territory and analysis of priority in remediation actions.

The problem

One of the most important problems in air quality management is the location of the areas where the remediation actions are priority. From monitoring data we know if air quality in selected locations meets the national and international air quality standards and from spatially resolved emissions inventories we can have an idea of the contribution of the different sources to air pollution. In optimal situation we have, with the use of air quality models, an approximate frame of air pollution distribution over the territory. The developed methodology was an attempt to define automatic criteria to obtain a classification of the areas based on priority in remediation action.

The methodology

The methodology was developed along the following lines:

- regional territory classification based on environmental sensibility of areas;
- regional territory classification based on pollutants loads;
- regional territory classification based on remediation action priorities.

For classification of areas within the regions an innovative methodology based on multivariate statistical analysis techniques was developed. Particularly automatic classification criteria such as cluster analysis combined with simple statistical procedures were used. For statistical analysis, and particularly for cluster analysis, the Statistica software from StatSoft was used [5].

Case studies

Italy Ministry of Environment in 1991 (Decree of May 20, 1991) introduced prescriptions and rules for the elaboration of regional air quality management plans. In December 1993 the Ministry of the Environment financed the realization of such plans. In 1999 Liguria, and in 2000 Umbria, regional administrations have completed the preparatory studies for the adoption of a regional air quality management plan according to the Decree of May 20, 1991.

In the framework of Liguria and Umbria Regional Air Quality Management Plans, experimentation of automatic classification of the areas within the territory, was carried out.

Liguria region is a thin plan strip between sea and mountains with very complex orography. The Region contains one of main Italy urban and industrial area, Genova, and includes the main national port and two other main ports,
three very large power plants, an integrated steel industry, chemical and petrochemicals industries, an airport and several highway sections.

Umbria region is an internal area at the very heart of the Italian peninsula, with 92 townships and no great urban agglomerations (Perugia, the capital, has only 151,000 inhabitants). The region has a good industrial development with iron and steel, metals and metallurgical, cement and chemicals industries. In the region there are also two large power plants.

Environmental sensibility

For the classification of territory based on environmental sensibility the regions were subdivided in 1 km x 1 km squared grid and for each squared area sensibility was evaluated using the following indicators:

- resident population,
- forests and other vegetation areas,
- beds available for tourists,
- biotopes.

Resident population and beds available for tourists were available on municipal level from ISTAT, the National Statistical Institute.

Forests coverage was available from the maps of European Union CORINE Land Cover programme, an inventory of land cover in 44 classes presented as a cartographic product, at a scale of 1:100,000. Italy CORINE land cover maps were available from the Centro Interregionale di coordinamento e documentazione per le informazioni territoriali.

Biotopes map was available from European Union funded project Bioitaly: Biotopes Inventory of Italy. A biotope is defined following the Council of the European Communities Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora as a site which, in the biogeographical region or regions to which it belongs, contributes significantly to the maintenance or restoration at a favorable conservation status of:

- a natural habitat type of Community interest (means those which are: in danger of disappearance in their natural range, or have a small natural range following their regression or by reason of their intrinsically restricted area, or present outstanding examples of typical characteristics of one or more of the five following biogeographical regions: Alpine, Atlantic, Continental, Macaronesian and Mediterranean)
- a specie of Community interest (means species which are: endangered or vulnerable or rare or endemic and requiring particular attention by reason of the specific nature of their habitat and/or the potential impact of their exploitation on their habitat and/or the potential impact of their exploitation on their conservation status).

The following procedure was used for data evaluation on 1 km x 1 km grid:

- municipal data on resident population and on tourists beds available were assigned to 1 km x 1 km squared grid using CORINE land cover class "Urban fabric" through topologic queries;
biotopes were evaluated on 1 km x 1 km squared grid through topologic query on BioItaly sites of regional, national or community interests.

- forests and other vegetation were evaluated on 1 km x 1 km squared grid through topologic query on CORINE land cover map "Forests";

In Figure 1 Liguria and in Figure 2 Umbria case studies are reported.

Figure 1 - Liguria territory classification based on Environmental Sensibility

![Liguria Region Environmental sensibility](image)

- high urbanization, high tourism, few forests
- high tourism and/or average urbanization, few forests
- forests with low urbanization and/or low tourism
- forests - low density (between 20% and 40%)
- forests - average density (between 40% and 75%)
- forests - high density (greater than 75%)
- others

Figure 2 - Umbria territory classification based on Environmental Sensibility

![Umbria Region Environmental sensibility](image)

- average
- average-high
- high - vegetation no biotopes
- high - biotopes no vegetation
- very high - vegetation and biotopes
- very high - urban areas

In Liguria case study the areas with no population and tourists beds available are classified in empirically way. Particularly, using vegetation
coverage of the grid units, three classes are defined: forest - low density
(coverage of the single units between 20% and 40%), forest - medium density
(coverage between 40% and 75%) and forest - high density (coverage greater
than 75%). The other grid units were classified using cluster analysis on
population, forests and tourism (no data on Biotopes was available at the time of
elaboration).

In Umbria case study automatic territory classification (cluster analysis)
based on all four previous introduced variable, was carried out.

Pollutants loads

For the classification of territory based on pollutants loads area, line and point
sources from regional emissions inventories were taken into account. Emissions
of the following main five pollutants were used: carbon monoxide (CO), volatile
organic compounds (VOC), nitrogen oxides (NOx), fine particulate or suspended
particles with diameter less than 10 micron (PM10), sulphur oxides (SOX). The
nomenclature used for the inventories follows the guidelines coming from the
European Commission CORINAIR working group. CORINAIR nomenclature
includes about 200 activities grouped in 11 groups [6]. This nomenclature was
enlarged to take into account locally relevant activities. The sources are split in
point sources, area sources and linear sources. The fixed sources, for which the
total annual emissions of one pollutant are larger than a fixed threshold value,
are considered point sources. Linear sources correspond to the main
communication ways (road, river, railway, and seaway) and generally all the
highways and all the main extra-urban roads are included. All the other sources
are defined as area sources.

The pollutant loads methodology uses multivariate statistical analysis
for classification of area/line sources and simple statistical methods for
autonomous classification of point sources. The area and line emissions were
evaluated on 1 km x 1 km squared grid using EC CORINE Land Cover maps
and/or others available land use maps, while point sources are individually
treated.

For area and line emissions classification was used the same
methodology as in environmental sensibility case; the methodology is based on
multivariate statistical analysis techniques, and particularly on automatic
classification criteria such as cluster analysis. The cluster analysis was applied
using as variables the values of 5 (pollutants) x 11 (groups of activities)
emissions and as cases the grid values.

Point sources are individually treated using simple statistical methods
for classification. Particularly an indicator was build for each point source k,
using the emissions $E$ of the pollutants $j$, as:

$$P_k = \sum_j (E_{jk}/(1 + 4*\delta_{j,CO})) \text{ where } \delta_{j,CO} = 1 \text{ if } j=CO, \text{ 0 otherwise}$$

MapInfo "Natural Break" algorithm was used for classification [7]; the algorithm
distributes the values so that the average of each class is as close as possible to
each of the class values in that class.

In Figure 3 Liguria and in Figure 4 Umbria case studies are reported.
Pollutants loads
area and line sources

very high
high
average - high
average
average - low
low

Figure 3 - Liguria pollutants loads territory classification based on area and line sources

In Table 1 average area and line sources emissions in different load classes are reported.

As an indicator of municipality pollutant load the number of grid areas in the different classes can be counted ad a rank of municipalities in function of the number of grid areas in very high, high, average-high and average areas classes can be performed.

Figure 4 - Liguria Pollutants loads territory classification based on point sources
Table 1 – Average area and line sources emissions in different load classes (Mg)

<table>
<thead>
<tr>
<th></th>
<th>Low-average</th>
<th>Average</th>
<th>Average-high</th>
<th>High</th>
<th>Very high</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CO</strong></td>
<td>37.9</td>
<td>289.5</td>
<td>893.0</td>
<td>3063.3</td>
<td>8298.0</td>
</tr>
<tr>
<td><strong>COV</strong></td>
<td>14.4</td>
<td>75.9</td>
<td>239.6</td>
<td>749.4</td>
<td>1887.7</td>
</tr>
<tr>
<td><strong>NO_x</strong></td>
<td>15.1</td>
<td>46.8</td>
<td>133.5</td>
<td>276.5</td>
<td>662.8</td>
</tr>
<tr>
<td><strong>PM_{10}</strong></td>
<td>0.7</td>
<td>5.6</td>
<td>11.0</td>
<td>21.6</td>
<td>43.2</td>
</tr>
<tr>
<td><strong>SO_x</strong></td>
<td>3.8</td>
<td>13.7</td>
<td>76.1</td>
<td>75.1</td>
<td>118.4</td>
</tr>
<tr>
<td><strong>Areas N.</strong></td>
<td>955</td>
<td>188</td>
<td>22</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

**Remediation actions priorities**

The combination of environmental sensibility and pollutant loads classification produce a regional classification based on priorities. This classification can take into account overall sources (integrated methodology) or area/line sources and separately point sources (area and line methodology).

In the area and line methodology the contribution of point sources was taken into account separately from area and line sources. An example of the application of this method is reported in Table 2 and Figure 5.

Figure 5 - Liguria (Genova area) remediation action priority classification
In this approach great attention was devoted at the problem of dispersion of pollutants emitted at high quote. Consequently classification of territory in terms of priority in remediation actions was performed only with area and line sources while priority in point sources was separately stated. In the integrated methodology the contribution of all the sources was taken into account at the same time. In this approach great attention was devoted at the local impact of sources and at the origins of pollution. Consequently classification of territory was performed in terms of where the emissions are generated. An example of the application of this method is reported in Table 3 and Figure 6.
<table>
<thead>
<tr>
<th>Priority</th>
<th>Sensibility</th>
<th>Area and line pollutant load</th>
<th>Point sources load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average-high</td>
<td>very high urban</td>
<td>highways</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>average</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>average low</td>
<td>average</td>
</tr>
<tr>
<td></td>
<td>high - vegetation no biotopes</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>average high</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>average low</td>
<td>low</td>
</tr>
<tr>
<td>Average</td>
<td>high - vegetation no biotopes</td>
<td>average low</td>
<td>average</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>highways</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>average</td>
<td>zero</td>
</tr>
<tr>
<td>Average-low</td>
<td>high - biotopes no vegetation</td>
<td>average low</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>very high - vegetation/biotopes</td>
<td>average low</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>very high urban</td>
<td>average low</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>high - vegetation no biotopes</td>
<td>highways</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>high - vegetation no biotopes</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>high - vegetation no biotopes</td>
<td>average</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>high - biotopes no vegetation</td>
<td>highways</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>low</td>
<td>very high</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>average low</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>very high - vegetation/biotopes</td>
<td>low</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>high - vegetation no biotopes</td>
<td>average low</td>
<td>Zero</td>
</tr>
<tr>
<td></td>
<td>high - biotopes no vegetation</td>
<td>average low</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>highways</td>
<td>average</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Very low</td>
<td>average</td>
<td>low</td>
<td>Zero</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6 - Umbria territory classification based on remediation action priority
Conclusion

The methodology presented was an useful support for priority area individuation in air quality management plans. The methodology can be further improved for better representation of point sources.

Future research will be directed toward:

- improve the classification of territory based on environmental sensibility through sub-municipal population data;
- use of local indicator of air pollutant dispersion characteristics;
- use of regional photochemical models results as an indicator of pollutant loads;
- use of multivariate correspondences analysis instead of empirical methods for remediation actions priority analysis.

Acknowledgement

The work has been realized for Liguria and Umbria regional administration in the framework of regional air quality management plan.

References


