Statistical models for the assessment of ambient air quality and classification of areas under EU legislation

C. Trozzi & A. Bernetti
Techne Consulting srl, Via G. Ricci Curbastro, Italy

Abstract

In recent years new European Union (EU) new air quality directives have been introduced in the legislation. These directives define the legislative basis for assessment and management of air quality in Member States. The Framework Directive gives general requirements and the Daughter Directives specify the requirements for the various pollutants in more detail.

The Directive requires Member States to divide their territory into zones - whereby an agglomeration is a special type of zone - based on the results of the initial identification of the levels of pollution and to perform ongoing assessment requirements related to the levels of pollution within the zones.

In the paper statistical models used in a lot of Italy’s regions for classification of areas under the EU Framework Directive will be reported. Statistical models, usually used in the temporal domain to predict future concentrations based on historical data, are used in the spatial domain to predict concentration in areas not covered by monitoring using data from monitored areas.

1 European Union legislation

In recent years new European Union (EU) air quality directives have been introduced in the legislation; Directive 96/62/EC (the Framework Directive [1]) and a lot of Daughter Directives are adopted [2-4] or in way of adoption [5]. These directives define the legislative basis for assessment and management of air quality in Member States. The Framework Directive gives general requirements and the Daughter Directives specify the requirements for the various pollutants in more detail.
The Directive requires Member States to divide their territory into zones based on the results of the initial identification of the levels of pollution and to perform ongoing assessment requirements related to the levels of pollution within the zones.

The Framework Directive and the first Daughter Directive introduce, for the first time in European air quality directives, the use of modelling in assessment and management of air quality. The Framework Directive refers in its preamble to “the use of other techniques of estimation of ambient air quality besides direct measurement”, defines that assessment “shall mean any method used to measure, calculate, predict or estimate the level of a pollutant…” and then states specifically that modelling techniques may be used.

1.1 Limit values, margins of tolerance and assessment thresholds

The Daughter Directives introduce:
- **Limit value**, a level fixed on the basis of scientific knowledge, with the aim of avoiding, preventing or reducing harmful effects on human health and/or the environment as a whole, to be attained within a given period and not to be exceeded once attained;
- **Alert threshold**, beyond which there is a risk to human health from brief exposure and at which immediate steps shall be taken;
- **Margin of tolerance**, the percentage of the limit value by which this value may be exceeded subject to the conditions of Framework Directive;
- **Upper assessment threshold**, below which a combination of measurements and modeling techniques may be used to assess ambient-air quality;
- **Lower assessment threshold**, a level below which modelling or objective-estimation techniques alone may be used to assess ambient-air quality.

1.2 Assessment of ambient air quality under the Framework Directive

The Framework Directive declares that, once limit values and alert thresholds have been set:
1. Ambient air quality shall be assessed throughout the EU territory.
2. Measurement is mandatory in the following zones:
   - **Agglomerations** (shall mean a zone with a population concentration in excess of 250,000 inhabitants or with a population density per km² which for the Member States justifies the need for ambient air quality to be assessed and managed.),
   - **Zones** (shall mean part of their territory delimited by the Member States) in which levels are between the limit values and the upper assessment threshold,
   - Other zones where levels exceed the limit values.

The measures provided for may be supplemented by modelling techniques to provide an adequate level of information on ambient air quality.

3. A combination of measurements and modelling techniques may be used to assess ambient air quality where the levels over a representative period are
below the *upper assessment threshold*. 

4. Where the levels are below the *lower assessment threshold*, the sole use of modelling or objective estimation techniques for assessing levels shall be possible. This provision shall not apply to *agglomerations* in the case of pollutants for which alert thresholds have been fixed.

5. Where pollutants have to be measured, the measurements shall be taken at fixed sites either continuously or by random sampling.

1.3 Definition of zones and agglomerations under the Framework Directive

The *Framework Directive* prescribes that:

1. Member States shall draw up a list of *zones* and *agglomerations* in which the levels of one or more pollutants are higher than the limit value plus the margin of tolerance (or the limit value where no margin of tolerance has been fixed for a specific pollutant).

2. Member States shall draw up a list of *zones* and *agglomerations* in which the levels of one or more pollutants are between the limit value and the limit value plus the margin of tolerance.

3. Member States shall draw up a list of *zones* and *agglomerations* in which the levels of pollutants are below the limit values.

1.4 Plan and programs under the *Framework Directive*

The *Framework Directive* prescribes that:

1. In the *zones* and *agglomerations* in which the levels of one or more pollutants are higher than the limit value plus the margin of tolerance, Member States shall take measures to ensure that a plan or programme is prepared or implemented for attaining the limit value within the specific time limit.

2. In the *zones* and *agglomerations* in which the levels of pollutants are below the limit values Member States shall maintain the levels of pollutants in these *zones* and *agglomerations* below the limit values and shall endeavour to preserve the best ambient air quality, compatible with sustainable development.

2 Air quality management in Italy

Before the adoption of the new European Union legislation the Italian 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 1997 the first regional plan was realized and in 1998–2000 several regional administrations have completed the preparatory studies. According to the law, the air quality management plan is the instrument of programming, coordination and control of human activities with atmospheric emissions, having as the primary aim the protection of human health and of the environment. A review of these experiences was recently published [6].
The following activities was performed in the frame of the plan preparation:

- Emissions inventory
- Air quality and meteorological data analysis;
- Classification of territory and analysis of priority in intervention;
- Projection of emissions without emissions reduction measures;
- Planning of measures to reduce emissions, definition of scenario of reduction and 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 the plan application.

3 Emissions inventory and projection

The nomenclature used in the emissions inventories, at the local level in Italy, follows the European Commission CORINAIR working group guidelines. CORINAIR nomenclature includes about 200 activities, grouped in 11 groups, and was adapted for local inventories.

The pollutants included in the inventories are:

- Main air pollutants, namely nitrogen oxides \( \text{NO}, \text{NO}_2, \text{N}_2\text{O} \), sulphur oxides \( \text{SO}_2, \text{SO}_3 \), non-methane organic compounds (VOC), carbon monoxide (CO) and suspended particles with diameter less than 10 µ (PM\(_{10}\));
- Heavy metals (As, Cd, Cu, Cr, Hg, Pb, Zn);
- Greenhouses gases: carbon dioxide (CO\(_2\)), methane (CH\(_4\)) and nitrogen protoxide (N\(_2\)O);
- Ammonia (NH\(_3\));
- Benzene (C\(_6\)H\(_6\));
- Polycyclic aromatic hydrocarbons (PAHs).

The sources are generally split in three categories: point sources, area sources and linear sources. Are considered point sources, the fixed sources for which the total annual emissions of one pollutant are larger than a fixed threshold value. Linear sources correspond to the main communication ways (road, railway, and waterway) and generally all the highways and all the main extra-urban roads are included. All the other sources are defined as area sources.

Area and line emissions at a future year of a specific pollutant produced by a given activity are evaluated through emission at base year, appropriate projection parameters for activity indicator and emission factor and specific extra projection parameters for specific zones, linear sources or units of point sources.

4 Spatial and temporal disaggregation of emissions inventory

The emissions are at first evaluated at the municipal level. Inside the municipality a square grid mesh is built up to represent pollutants emissions to allow the application of diffusion and statistical models. To disaggregate pollutant emissions from the municipal level to the mesh level, the methodology of proxy variables is used. The emissions on mesh \( k \) are obtained as:
\[ E_{ik} = \Sigma_j \left( \frac{E_{ij}P_{lkj}}{\Sigma_k P_{lkj}} \right) \]

where: i, activity; j, municipality; k, mesh; l, proxy variable appropriate to activity i; \( E_{ij} \), total emission; \( P_{lkj} \), value of the proxy variable. Proxy comes from land use maps.

Annual emissions are at first evaluated. Once annual emissions have been evaluated, it is also important to obtain an estimate of their hourly, monthly and daily distribution especially to allow the application of statistical and diffusion models. For main point sources temporal disaggregation may be evaluated directly at the plant, through the questionnaires. The disaggregation of other sources is estimated through the use of corrective factors that have a similar rule of the proxy variable in the case of spatial distribution (for example: typical working hours, wintertime, temperature, monthly selling of fuels, etc.).

5 Assessment of ambient air quality

5.1 Methodology

An innovative methodology for the assessment of ambient air quality of the whole regional territory, with reference to the protection of the health, is developed (Figure 1).

Figure 1: Methodology for assessment of ambient air quality under EU legislation.

The starting point of the methodology is represented by the presence on the territory of an air quality-monitoring network that satisfies to criterions of completeness and reliability and to the realization of a detailed air pollutants emissions inventory on zone basis with specification of the sources type (diffuse,
linear and punctual). The experimental approach consists, following the indication of the Framework Directive, in the integration of:

- Continuous monitoring data from air quality monitoring network;
- Mobile monitoring station data;
- Results of emissions inventory and diffusion and statistical models to integrate the results of monitoring.

The evaluation with use of models consists in the application of three complementary approaches:

- Evaluation of the concentrations on the 1km x 1km network generated from the linear and diffused emissions through a statistic model;
- Integration of the evaluations of which to the preceding point with specific diffusion models for point sources;
- Integration, if applicable, of the evaluations of which to the preceding points with specific diffusion models for linear sources.

5.2 Statistical model

Statistical models are usually used in temporal domain to predict future concentrations based on historical data (see for example [7-9]). In the following a statistical model is used in spatial domain, to predict concentration in areas not covered by monitoring using data from monitored areas.

The model is based on multiple linear regression methods through interpolation in the space of the monitored concentrations. The function obtained by interpolation is used to predict the concentrations in the areas where monitoring is not active.

The model is the following:

\[ C(x,y,t) = F(f_i(x,y,t)) \]

where \( x, y \) are the spatial variables, \( t \) is the time, \( C \) the concentration and \( f_i \) a lot of explaining variables.

In the following only emissions for area and linear sources are used as explaining variables. The analysis is performed on emissions on a network of 1 km x 1km areas and concentration inside the same area. The concentration of an area is associated with the concentration of one or more monitoring stations inside the area. Analysis is applied therefore to a data set in which the dependent variable is the concentration of pollutant and the independent variables are the emissions of the same pollutant for each of 11 CORINAIR nomenclature group.

For each pollutant, the initial hypothesis is adopted that the concentration depends from the emissions of all the macro-sectors in the inventory. An analysis of correlation is performed with the purpose to find the relevant macro-sectors in the determination of the concentrations. The results are that the only traffic and non-industrial combustion macro-sectors are relevant.

Emissions are evaluated on the specific 1 km x 1 km area where concentrations are measured and in the surrounding territory. In such way there is also an indirect account of some meteo-climatic factors responsible of the effect of edge.

The assumption that the concentrations depend in non-linear way on the emissions is finally taken. In consequence, is taken the hypothesis of a
polynomial dependence of the concentrations from the emissions from the traffic ($E_{\text{traffic}}$), from the non industrial combustion ($E_{\text{dom}}$), and from the total emissions of the neighboring cells ($E_{\text{neig}}$). The presence of not-linear mixed terms is excluded.

The final model used is the following:

$$C = (d_1E_{\text{dom}} + d_2E_{\text{dom}}^2 + d_3E_{\text{dom}}^3 + d_4E_{\text{dom}}^4) + (t_1E_{\text{traffic}} + t_2E_{\text{traffic}}^2 + t_3E_{\text{traffic}}^3 + t_4E_{\text{traffic}}^4) + (n_1E_{\text{neig}} + n_2E_{\text{neig}}^2 + n_3E_{\text{neig}}^3 + n_4E_{\text{neig}}^4)$$

A stepwise procedure is used in the application of regression model. The stepwise procedure introduces variables at steps and use analysis of variance to decide if introduces a new variable and which variable is to be introduced.

Figure 2 give a resume of the model definition.

\[ C = (E_{\text{dom}} + E_{\text{dom}}^2 + E_{\text{dom}}^3 + E_{\text{dom}}^4) + (E_{\text{traffic}} + E_{\text{traffic}}^2 + E_{\text{traffic}}^3 + E_{\text{traffic}}^4) + (E_{\text{neig}} + E_{\text{neig}}^2 + E_{\text{neig}}^3 + E_{\text{neig}}^4) \]

Figure 2: Statistical model.

The justifications of the use of the only emissions as independent variable and not to consider other variables, for the meteorology of the site or the transport to long distance are the followings:

- The model is used only to evaluate the concentrations in the urban areas;
- In urban area local emissions sources (traffic and domestic heating) are the most important;
- Possible point sources are separately evaluated.
- The emissions of surrounding areas are taken into account in the model.

From an operative point of view the following procedure is used.

The first step in the definition of the model is the selection of the monitoring stations, positioned in the urban areas and relevant from the point of view of the generalization of the results of the measures to the whole regional territory with reference to the correlation between the measured data and the emissions of the air pollutants; from the selection will be eliminated the stations positioned in green areas or finalized to the evaluation of the pollution in extraurban areas.
For these stations are extracted the available data for all the parameters relatively to a year of interest.

The second step is the definition of the geographical domain of application of the model through selection of the urban 1km x 1km areas using land cover information and the assignation of monitoring stations to these areas.

The third step is the extraction from the inventory of the emissions: linear and diffuse, on the 1km x 1km network, on hourly base, for macro-sectors traffic and non industrial combustion in the 1km x 1km area where one or more stations are present, and total for the surrounding area.

The fourth step is the application of the regression model, with stepwise technique, that connects, for every pollutant, the hourly emissions and the hourly concentrations.

The final step is the application of the model to all the "urban" areas and the definition of maps of the estimate pollutants concentration.

5.3 Final assessment

Once concluded the elaboration of the statistic model, the results are integrated with those coming from the monitoring and from the application of diffusion models. The final evaluation is made:

• Assigning monitoring data to the area where monitoring data are available;
• Integrate in the evaluation, in the area where no monitoring data are available, the results of the diffusion and statistical models.

6 Definition of zones and agglomerations for plan and programs

The procedure for definition of zones and agglomerations for plan and programs consists, for every polluting, in:

• To compare in each area the results from air quality assessment with limits;
• To define the administrative areas in which the limits are not respected;
• To unify the administrative areas in homogeneous areas.

7 Definition of zones for monitoring

The method has been used also for the definition of the areas for the planning of the monitoring of the quality of the air. Using Upper and lower assessment thresholds are defined the zones where measurements are mandatory, the zones where a combination of measurements and modeling techniques may be used and the zones where modelling or objective-estimation techniques alone may be used to assess ambient-air quality.

8 Use of the statistical model for air quality prediction

Finally is in progress the experimentation of the use of the methodology in forecast, using emissions in projection to evaluate the improvement of the quality of the air in plan and programs using projection of emissions.
9 Case studies

Methodology has been applied, in respective autonomy and with some methodological differences, in the Autonomous Province of Trento, Liguria, Umbria and Friuli Venezia Giulia regions. The pollutant taken into consideration is sulphur dioxide (SO$_2$), nitrogen dioxide (NO$_2$), suspended particles with diameter less than 10 μ (PM$_{10}$), carbon monoxide (CO) and Benzene (C$_6$H$_6$). In Table 1, Figure 3 and Figure 4 some examples of results are reported.

Table 1: $R^2$ coefficients for different case studies (CS1,…, CS4) and pollutants.

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>CS1</th>
<th>CS1 (°)</th>
<th>CS2</th>
<th>CS3</th>
<th>CS4 (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO$_2$</td>
<td>0.68</td>
<td>0.41</td>
<td>0.56</td>
<td>0.44</td>
<td>0.68</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>0.70</td>
<td>0.69</td>
<td>0.69</td>
<td>0.64</td>
<td>0.79</td>
</tr>
<tr>
<td>PM$_{10}$</td>
<td>0.92</td>
<td>0.88</td>
<td>0.80</td>
<td>0.73</td>
<td>0.58</td>
</tr>
<tr>
<td>CO</td>
<td>0.57</td>
<td>0.54</td>
<td>0.63</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>C$_6$H$_6$</td>
<td>0.62</td>
<td>0.60</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

(°) model without total emissions of the neighboring cells.

10 Conclusions

The paper presents a methodology for the assessment of ambient air quality and the classification of areas under air quality EU legislation. The methodology contains a statistical regression model of air pollutant concentrations versus air pollutant emissions. The model can be used for air quality assessment under EU air quality directive. Work are in progress to compare results from regression techniques with results from neural network application.
Acknowledgements

The work has been realized for in the framework of regional air quality management planning activities. We wish to thank the regional administration that contributed to the work. Particularly acknowledgement to L.Badalato (Liguria Region), C.Brescianini, U.Gasparrino, M.Beggiato (Liguria ARPA), G.Anderle, G.Tonidandel (Trento APPA), P.Gubertini (Friuli Venezia Giulia Region), M.Trinei (Umbria Region), G.Marchetti, S.Curcuruto (Umbria ARPA).

References


