

Modelling System to Estimate Spatial and Temporal Variations in Urban Air Pollutants

Gregorio Andria, Giuseppe Cavone, Anna M. L. Lanzolla, Mario Savino

*Department of Electrics and Electronics (DEE) – Politecnico di Bari, Via Re David, 200, 70125 Bari, Italy
Phone/Fax: +39 080 5963436 – E-mail: [andria, cavone, lanzolla, savino]@misura.poliba.it*

Summary – This work proposes the development of a suitable air pollution model to estimate the temporal and spatial variations in concentration values of air pollutants measured in monitoring network located in urban area. The use of the proposed modelling technique in data environmental analysis allows characterising the pollutants behaviour, to validate measured data and to predict the values of contaminant substance emissions; so it results a very useful analysis tool especially when there are numerous missing or erroneous data to be validated suitably. The developed model has been tested against environmental data acquired during several years by verifying the agreement of the predicted and measured values. Finally, some information about reliability degree of estimate is provided.

Keywords: air pollution modelling, monitoring network, data validation, spatial correlation.

I. Introduction

In the last few years there is a growing interest in solutions designed to assess, prevent and reduce the impact on environment of human activities. In this context, a lot of public administrations have installed several monitoring systems able to carry out, in real time, qualitative and quantitative information about the characteristics of urban centres environment. In particular, a considerable importance is turned to the atmospheric pollution and the evaluation of air quality.

Often, the availability of measured environmental data for the determination of the air pollution level is not sufficient to describe the temporal and spatial behaviour of pollutant emissions. For this reason, it is very important to develop suitable mathematical model allowing estimating and reconstructing the concentration values even if the data are not available, due to instrumentation failure or absence of monitoring stations. Besides, given the complexity of the pollution process and the great number of factors involved, the use of suitable analysis technique produces considerable advantages in data management and interpretation [6].

Model development and testing tend to concentrate on how well models represent “reality” or reproduces measurements. However, there are many sources of uncertainty in modelling atmospheric pollution that can have significant consequences in environmental impact assessment, where the public authorities base their decisions on results of data analysis and calculation methods. So, it is necessary to characterize every used model in terms of uncertainty of estimate.

II. Proposed method

We propose the analysis of environmental data measured in monitoring network installed in Taranto area (south Italy). consisting of 7 automatic acquisition and recoding stations (six located in fixed sites and one mobile showed in Fig. 1) that have been able to measure continuously both chemical substances and meteorological quantity [3],[4]. Most of these stations have been located for the purpose of monitoring ‘hot spot’, either near busiest traffic roads or major local industrial plant. The data collected by the remote sensing must be carefully processed to identify possible anomalies or to indicate whether or not a prescribed limit or ceiling is exceeded [2]. In our study, the pollutants representing the main contributors to air pollution caused by vehicular traffic have been analysed since the road traffic has been identified as a major liable for the deterioration of air quality in urban areas.

In previous works [5] and [5], the Authors identified and validated a mathematical model, based on joined application of Kalman filter and Kriging technique, to describe the time-varying behaviour of analysed substances and to highlight the correlations between them.

Particularly, the authors proposed firstly the application of the Kalman filter to the analysis of environmental data in order to try overcoming the problem relevant to the eventual presence of not complete time series of measured values. As well known, this technique consists of a recursive procedure allowing the data filtering and providing the estimate of the analysed quantities [9], carrying out the best estimate of required parameters, even if the characteristics of the observed phenomena are unknown. Thanks to recursive approach of the filter, it was possible to analyse a considerable set of values and extrapolate the information contained in every single data, reducing the relevant noise content. Another important feature of this algorithm consists in its real-time forecasting capabilities of the system observable variables: the model parameters are

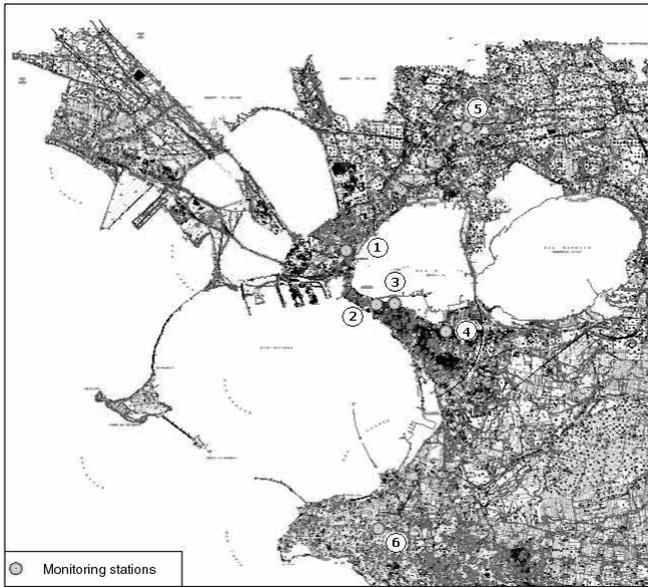


Figure 1. Plant of Taranto area with the network of monitoring stations

corrective coefficient that takes into account of wind conditions defined in a previous article [3], and $\gamma_3, \gamma_2, \gamma_1$ are the coefficients of the linear regression obtained by applying the Kalman filter to the measured data. The coefficients have been calculated in iterative way by means of the following recursive expression:

$$[\hat{\gamma}_i] = [\hat{\gamma}_{i-1}] + K_i \cdot \left\{ \hat{k}_{BE,i} - [k_{CO,i} \quad k_{TO,i} \quad 1] \cdot [\hat{\gamma}_{i-1}] \right\} \quad (2)$$

where $[\hat{\gamma}_i] = [\hat{\gamma}_{3i}, \hat{\gamma}_{2i}, \hat{\gamma}_{1i}]^T$ and $[\hat{\gamma}_{i-1}] = [\hat{\gamma}_{3,i-1}, \hat{\gamma}_{2,i-1}, \hat{\gamma}_{1,i-1}]^T$ are the estimates of coefficients at i -th and $(i-1)$ -th step respectively and K_i is the *Kalman gain* at i -th step. Then, the application of Kalman filter offers the typical advantages of recursive methods to estimate the state of a system from measurements that contain missing or erroneous data. Figure 2 shows the behaviour of daily averaged values of benzene and its estimate obtained by applying Kalman filter.

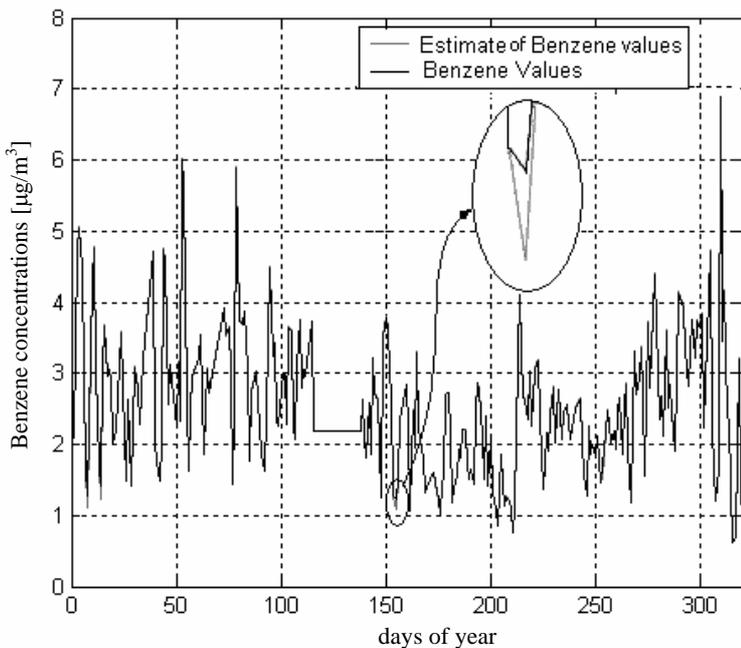


Figure 2. Daily average Benzene concentrations relevant year 2001 (dark line) and estimate values obtained by applying Kalman filter (grey line).

estimated from one initial set of measurements; subsequently, the filter model is applied, with the estimated parameters, to another set of data in order to calculate the forecasting performance of the same system variables.

Once the Kalman filtering had been applied to environmental analysis on a noisy full set of data, we characterized the time-behaviour of specific analysed substances by means of suitable mathematical models. Then, we identified some temporal relationships between different pollutants, like the daily averaged concentration \hat{k}_{BE} of the benzene that has been estimated by using the averaged concentration values of other considered pollutants as carbon monoxide (CO) and toluene by the following simple expression [3]:

$$\hat{k}_{BE,i} = (\gamma_3 \cdot k_{CO,i} + \gamma_2 \cdot k_{TO,i} + \gamma_1) \cdot \theta(v_i) \quad (1)$$

where $k_{CO,i}$ e $k_{TO,i}$ represent the i -th daily averaged values of CO and toluene respectively, $\theta(v_i)$ is the

corrective coefficient that takes into account of wind conditions defined in a previous article [3], and $\gamma_3, \gamma_2, \gamma_1$ are the coefficients of the linear regression obtained by applying the Kalman filter to the measured data. The coefficients have been calculated in iterative way by means of the following recursive expression:

where $k_{CO,i}$ e $k_{TO,i}$ represent the i -th daily averaged values of CO and toluene respectively, $\theta(v_i)$ is the

corrective coefficient that takes into account of wind conditions defined in a previous article [3], and $\gamma_3, \gamma_2, \gamma_1$ are the coefficients of the linear regression obtained by applying the Kalman filter to the measured data. The coefficients have been calculated in iterative way by means of the following recursive expression:

corrective coefficient that takes into account of wind conditions defined in a previous article [3], and $\gamma_3, \gamma_2, \gamma_1$ are the coefficients of the linear regression obtained by applying the Kalman filter to the measured data. The coefficients have been calculated in iterative way by means of the following recursive expression:

month long) has been carried out, in which the data can be assumed quite stationary, so it is possible to calculate the experimental variogram as:

$$\gamma_k^*(h) = \frac{1}{2N_k} \cdot \sum_{j=1}^{N_k} (Z_{kj}(0) - Z_{kj}(h))^2 \quad (3)$$

where N_k is the number of measurements contained in class k ($N_k \cong 30$), $Z_{kj}(0)$ e $Z_{kj}(h)$ represent the j -th values of daily averaged of benzene relevant to class k , acquired in the station pairs with reciprocal distance h .

By analysing the different experimental variogram defined for each class, we have tried to identify an analytical function describing in an effective way the relationship between the spatial points. The function that best fit the calculated experimental variogram is the exponential one, defined as follows:

$$\gamma_k(h) = c_k \left[1 - \exp\left(-\frac{3h}{a_k}\right) \right] \quad (4)$$

Then, for each class, a weighed linear combination of the known sample values has been calculated, in order to estimate concentration data of substances acquired in a particular station under analysis. The model weights take into account spatial distances between the analysed stations.

The daily averaged values of benzene concentrations acquired in the station 2 have been estimated as function of the measurements relevant to all the stations different from 2, by means of the following relationship:

$$\widehat{k}_{BE,2i,k} = \sum_{j \neq 2} \lambda_{jk} k_{BE,ij,k} \quad (5)$$

where $k_{BE,ij,k}$ is the i -th daily averaged value of benzene (contained in class k) relevant to the i -th station, and λ_{jk} represent the weights of the linear combination for benzene values included in class k . These weights have been calculated by solving the following system:

$$\begin{bmatrix} \gamma_k(h_{11}) & \dots & \gamma_k(h_{1,n-1}) & 1 \\ \dots & \dots & \dots & \dots \\ \gamma_k(h_{n-1,1}) & \dots & \gamma_k(h_{n-1,n-1}) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_{1k} \\ \dots \\ \lambda_{n-1,k} \\ \mu_k \end{bmatrix} = \begin{bmatrix} \gamma_k(h_{12}) \\ \dots \\ \gamma_k(h_{n-1,2}) \\ 1 \end{bmatrix}; \quad \sum_{i=1}^n \lambda_{ik} = 1 \quad (6)$$

where the generic term $\gamma_k(h_{ij})$ represents the variogram relevant to the class k defined in eq. (4), calculated for the distance h_{ij} between the stations i and j . Besides, eq. (6) allows calculating μ_k , that represents the Lagrange parameter used to calculate the minimized standard deviation $\sigma_{BE_{2k}}$ of Benzene estimate in the class k by means the following expression:

$$\sigma_{BE_{2k}} = \sqrt{\mu_k + \sum_{j \neq 2} \lambda_{jk} \gamma_k(h_{j2})} \quad (7)$$

Finally, a hybrid model based on the joint application of both Kalman filter and Kriging techniques has been applied. Firstly, the behaviour of daily averaged values of CO concentrations relevant to the monitoring station 2 by using the Kriging model applied to the data of the remaining stations has been reconstructed. After, the daily averaged values of benzene concentrations in the station 2 by means of the eq. (1) have been evaluated, with an estimate error percentage equal about 3%. Figure 3 shows the results of benzene estimate obtained by applying the so-described hybrid model.

The previously-developed method has been improved tacking into account the relationship among the concentrations of different type of pollutants at a fixed station (indicated as f) and the concentrations of a same substance measured in all stations of monitoring network (indicated as g). In this way, it is possible to express the concentration values of analyzed pollutants as function of spatial and time coordinates, as follows:

$$\begin{aligned} f(k_{x_1}(t, h_j), k_{x_2}(t, h_j), \dots, k_{x_n}(t, h_j)) &= 0; & j = 1 \dots 7 \\ g(k_{x_1}(t, h_1), k_{x_2}(t, h_2), \dots, k_{x_n}(t, h_7)) &= 0; & i = 1 \dots n \end{aligned} \quad (8)$$

where $k_{x_i}(t, h_j)$ represent the concentration value of i -th analysed pollutant measured at time t and in the monitoring station located to distance h_j with respect to the spatial reference point; the subscripts i and j indicate the typology of pollutant and the number of monitoring station respectively; n is the total number of the analysed pollutants; f and g are the mathematical functions above mentioned. The temporal behaviour of acquired data is extremely variable, particularly for roadside sites, but there is evidence of a universal pattern for certain exposure periods [1], showing distinctive peaks or troughs across some or all monitoring sites. In particular, the data present a marked seasonal variation reaching a minimum in summer and a maximum in winter. In order to quantify the relationships between contributor factors affecting the spatial variations in concentrations of analysed pollutants, the values measured in all monitoring stations have been elaborated.

(1)

In the previous work [5], the Authors have been examined the environmental data relevant to only three near stations having same characteristics and local conditions very similar. The study has been carried out by using the already mentioned Kriging technique.

In this work, to improve the performances of developed model, the data analysis has been extended to all stations of monitoring network, taking into account the different typologies of areas where the acquired systems have been placed and their spatial distances.

In order to characterize properly the analysed territory a classification, in terms of pollutant spreading characteristics, has been carried out. In particular, the monitoring stations have been separated in two classes: *isotropy* (the pollutants disperse in all directions) and *anisotropy* (the pollutants disperse in some preferential directions only). Besides, it is important to differentiate the stations on the basis of significant features influencing the pollution spreading as different typology of area, street geometry (as canyon, open and background), driving condition expressed in terms of traffic speed and flow, topographic characteristics and presence of emissive sources. The effects of street geometry on atmospheric dispersion have been demonstrated in several studies [7] where it is highlighted that the narrow road widths and enclosing architecture of canyon streets inhibit of traffic emissions while the radial routes have a flatter more open topography producing good ventilation. Traffic in the centre is often congested and slow moving while free-flow conditional are more usual in others routes.

So, a crucial phase of the proposed study is the identification of the mayor categories of the analysed sites according to their function within the urban structure, ranging from busy roadside to rural background.

After the site analysis and classification, a lot of simulations have been carried out to quantify the influence of area typology and distance on air pollution level. The results of these simulations led to calculate the relationships describing the spatial variation of pollutant concentrations by means of a weighed linear combination of pollutant concentrations. The model weights take into account the different typologies of areas where the acquired systems have been placed and their spatial locations. Figure 3 illustrates an evaluation of daily-averaged benzene-concentrations against the time (in class $k=7$), carried out by using the proposed measurement and data-processing methodology, on the basis of all the available stations. The mean range of the relevant measurement standard-uncertainty has resulted within 2%.

III. Conclusive considerations

The approach suggested in the paper allows to analyse and to characterize concentration measurements of air pollutants, relevant to very high road-traffic areas. Some mathematic models, based on analysis of temporal and spatial variations of pollutant concentrations, have been developed and the relevant uncertainty has been estimated. By using these techniques, it is possible to obtain a detailed map of pollutant behaviour on analysed territory and to estimate the pollution level even in areas where are not installed monitoring stations, with a suitable confidence level. The research is actually in progress, owing to a lot of both time- and spatial-varying environmental data we collected in the period 2001-2005. In a next work, a complete report, evidencing the overall performance of the proposed technique applied to actual data jointly with their suitable characterization, will be presented and emphasized.

References

- [1] P. Zanetti: *Air pollution modelling, theories, computational methods and available software*, Van Nostrand Reinhold, New York 1990.

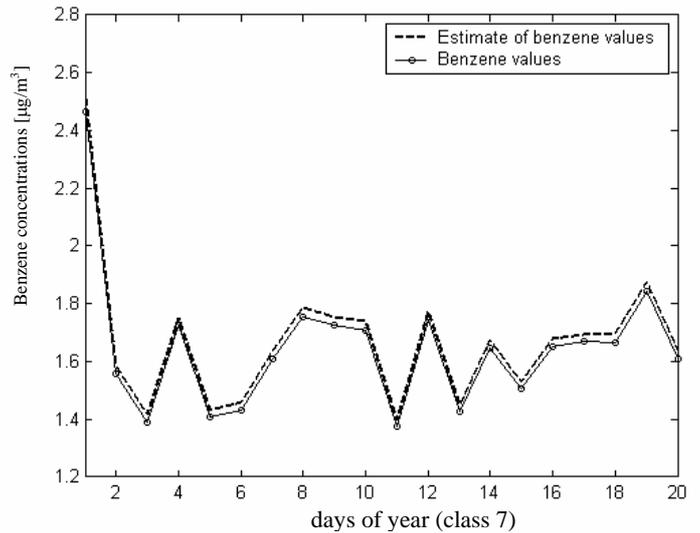


Figure 3: Daily averaged of benzene concentrations relevant year 2001 (line with circle marker) and their estimate values obtained by applying the combined algorithms (dashed line).

- [2] S. Brandini, D. Bogni, S. Manzoni: "Knowledge based environmental data validation", *IEMSS 2002 Integrate Assessment and Decision Support*, pp. 330-335, Lugano, June 2002.
- [3] G. Andria, G. Cavone, V. Di Lecce, A.M.L. Lanzolla: "Measurement and Characterization of Environmental Pollutants via data Correlation of Sensor outputs", *Proc. of the IMTC/03 Instr. & Meas. Techn. Conf.*, Vail-CO, USA, May 2003.
- [4] G. Andria, G. Cavone, V. Di Lecce, A.M.L. Lanzolla "Mathematic model for measurement and characterization of air pollution in areas with high road-traffic level", *Proc. of the XVII IMEKO World Congress*, Dubrovnik, June 2003.
- [5] G. Andria, G. Cavone, A.M.L. Lanzolla , A. Rubino "Modelling study for characterizing and predicting urban air pollution", *Proc. of IMEKO TC4 Symposium*, Athens, Greece, September 2004
- [6] A. Karppinen, J. Kukkonen, T.elolahde, T. Koskentalo: "A modelling system for predicting air pollution: comparison of model predictions with the data of an urban measurement network in Helsinki", *Atmospheric Environment*, vol. 34, pp. 3735-3743, 2000.
- [7] C. Kirby, A. Greig, T. Drye: "Temporal and spatial variations in nitrogen dioxide concentrations across an urban landscape: Cammbrige, U.K.", *Environmental Monitoring Assessment*, vol. 52, pp. 65-82, 1998
- [8] P. Brooker: "Kriging", *Engineering and Mining Journal*, vol. 180, No. 9, 1979
- [9] H. W. Soreson: *Kalman Filtering: Theory and Application*, IEEE Press, New York, 1985.