XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

MATHEMATIC MODEL FOR MEASUREMENT AND CHARACTERIZATION OF AIR POLLUTION IN AREAS WITH HIGH ROAD-TRAFFIC LEVELS

Gregorio Andria, Giuseppe Cavone, Vincenzo Di Lecce, Anna M. L. Lanzolla

Dept. of Electrics and Electronics, 1st & 2nd Engineering Faculty of Bari Polytechnic, Italy

Abstract – The paper deals with measurement and behaviour characterization of the main substances influencing air pollution in an urban area with high road traffic level. To this aim, the correlation existing between some pollutants is identified and a mathematic model able to forecast and describe the time varying behaviour of contaminating substances is implemented. Finally, some information about reliability degree of estimate is provided.

Keywords: measurement of environmental quantities, mathematic models, sensors of environmental pollutants.

1. INTRODUCTION

In the last years, in social and scientific area, we are assisting to a renewed and increasing interest turned to the environmental monitoring field, particularly to one relevant to air quality analysis and evaluation.

Air pollution is the result of a complex phenomenology involving several aspects, like emission and chemical-transformation processes, transport and dilution phenomena connected to the dynamic of atmospheric layers.

Therefore, it is difficult to measure the air quality and to define univocally a quality standard to guarantee a suitable human and environmental healthiness. In this context, a quite complex legal issue has been developed, differentiated for several nations, according to various social and geographic contexts. The national specific regulations identify and define the pollutants and provide, for each of them, the admitted maximum concentration level.

The usually analysed pollutants are classified as:

– *Primary* (like CO, NO, benzene), directly given out from source of emission (road traffic, industry, domestic heating). These substances are usually used as reliable indicators of the presence and type of emission sources
– *Secondary* (like O_3, NO_2), deriving from chemical transformation and used as indicators of photochemical smog

To evaluate methodically polling level and to forecast emergency situations, the national specific regulations require the implementation of suitable air pollution testing stations to measure pollutant levels and meteorological parameters of interest.

In this context, it is important to develop environmental data processing and interpretation methods with the aim to describe the peculiarly characteristics of polluting substances, to forecast the possible alarm situations or risk events and to plan effective technical operations.

2. ANALYSIS OF A REAL CASE

In this work, the environmental data, recoded in several monitoring stations installed in Taranto, Italy, are analysed. This considered country has been declared "*territory with high environmental risk*" owing to lots of industrial areas (cement factories, steelworks, oil refinery and so on) placed near the urban centre. In this context, considerable sensibilities and interests about environmental thematic have been developed. This leaded the municipal administrators to exercise a large control of quality air by means a continuous and monitoring of pollution level.

The examined network testing consists of 7 automatic acquisition and recoding stations that have been able to measure both chemical substances and meteorological quantities. These stations are placed on various typologies of urban areas (parks, residential areas, sub-urban areas), according to law national prescriptions.

To measure a large quantity and variety of substances, it is necessary to implement a complex measurement system consisting of several chemical substances analyzer (based on photo-ionization, gas chromatography, IR and UR absorption) and lots of sensors to acquire meteorological quantities too.

The analysed stations carry out about 50 measures per minute. The results are collected by means a sophisticated acquisition and measurement system that, every day, executes automatically the calibration of all devices.

The recorded data result from the acquisition process measuring and sampling the substances under test by using different frequency rates according to measurement technique. Then from a data post-elaboration process it is possible to carry out the hourly average values for each quantities.

A correct and suitable data collection is a very important and crucial phase for each scientific research, especially if it concerns environmental data, as the object of present study highly complex requires.

The data set are usually characterized of a lot of interpretation problems (that can generate some errors during elaboration and prediction phases), as:

– missing data, caused by periodic setting or stop function of instrumentations,

- incorrect data, caused by malfunction or non setting of equipments,
- incorrect recoding data (not interpretable symbols) caused by errors in transmission and recording data phases.

Here some interpolation and validation techniques are used to approach and try to solve these explained problems. The procedure develops in several steps, using data that, even if partially pre-treated and filtered by technicians employed in control of measure stations, contain equally irregular values.

The incorrect data, when they are present in a large number, can invalidate the carried out analysis, introducing errors in measurements of examined parameters. For this reason, some filters have been developed, to point out incorrect and discordant data. They are based on the elimination of both evident discontinuity in the collected data and values outside the admissibly ranges previously defined for each examined quantity.

The missing data, when they are present in large number, can lead to perform the elaboration phase using reduced data sets. This restriction can wake difficult individuating peculiar dynamic of pollution phenomena (like daily and seasonal recurrences), distorting the obtained results. Therefore, reconstruction techniques, based on both continuity hypothesis of pollution values and interpolation methods, have been developed. By using these techniques, it is possible to work with a continuous and valid set of data, with the further aim of reducing the measurement uncertainty [1-3].

3. OBTAINED RESULTS

The results of environmental data analysis, recorded in Taranto, during year 2001, are here presented. The greatest attention is devoted to the substances playing the leading role in the road traffic pollution.

As well known, the problem of air pollution wakes particular worries in urban areas in which the polling emissions result elevated and people is more exposed to risk of human health injury.

3.1. Pollutants behaviour characterization

In urban areas, the traffic road is the main cause of a carbon monoxide (CO) emissions. So we started to analyse the data about the CO concentration measurements.

The availability of a great number of data permits us to individuate the trend on long time about concentrations of the examined substances. As example, analysing the annual trend of hourly concentrations of CO, allows us information about the presence of particular recurrences in specific hours of the day and period of year, as shown in Fig. 1. More exactly, we can notice the presence of:

- two distinguished daily peaks of carbon monoxide concentrations, corresponding to rush hours for road traffic, that result 9,00 a.m. and 8,00 p.m.
- a minimum of carbon monoxide concentrations, corresponding to summer days, when it is normally possible to observe a reduction in road traffic.

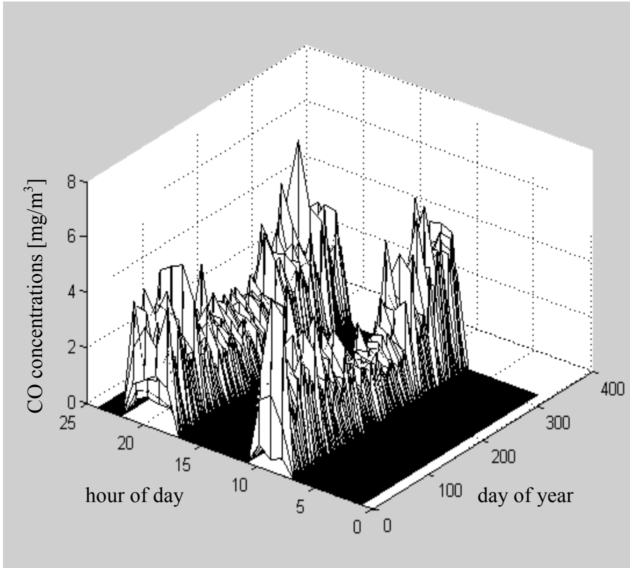

Fig. 1. Behaviour of CO concentrations relevant year 2001, filtered around rush hours (9,00 a.m. and 8,00 p.m.)

To characterize the CO behaviour relevant to a whole year, the trend of daily average and maximums values have been analysed.

By following executed observations and on the basis of numerous measurement elaborations performed on data sets, we have individuated a mathematic model, based on a suitable polynomial interpolation, able to describe the behaviour of daily average concentrations of pollutants examined during the days of year.

To evaluate the reliability of developed model, several tests have been carried out, by eliminating a subset of data in a random way, from the complete time series. By recalculating the interpolation polynomial obtained using the so reduced data set, and by comparing the results carried out using the original data set, it has been obtained a measurement error lower then 4%. As example, Fig.2 shows

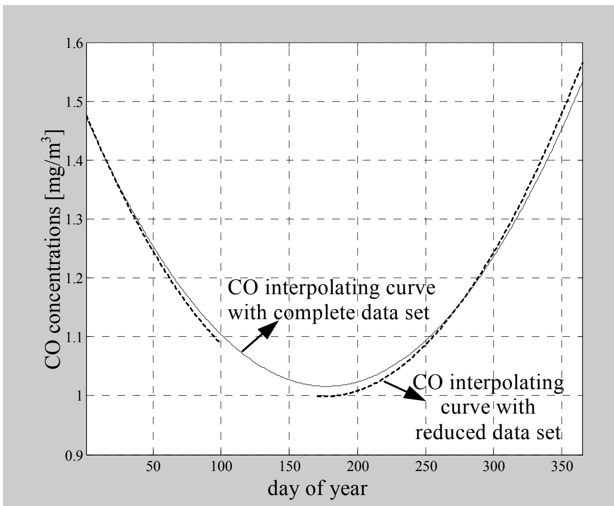

Fig. 2. Behaviour of fitting polynomial curves representing the daily average CO concentrations relevant year 2001, calculated by considering the complete set of data (continuous line) and the reduced one (dashed line).

the comparison between the interpolating curve of the daily average concentrations of CO relevant to year 2001 (obtained using the complete data set), and the curve carried out by eliminating the days of year included in the interval from 100 to 170. In this case, the calculated average error is resulted less than 2%.

To characterize the behaviour of daily maximums of CO concentrations, a similar analysis has been carried out. We have obtained similar results with respect to daily average values (the new interpolation polynomial has the same degree of the previous one and the coefficients of two curves are tied of 2 multiplying factor).

3.2. Analysis of correlations

By analysing the benzene concentrations, we have obtained very similar results (same daily and seasonal trends) with respect to CO data. After a suitable fitting of the available measured data sets, it has been possible to identify a strong correlation between the two substances (the correlation coefficient between the two interpolation curves is resulted equal to 0,97 [3]) and to permit us to individuate a good linear relationship between the average concentrations of examined pollutants.

To supply a fist validation of obtained results, we have carried out an annual reconstruction of daily average concentrations of benzene, starting from CO values and coefficients of the achieved linear regression straight line.

To evaluate the reliability of the so identified correlation, two indexes relevant to the mean square deviations have been defined as follows:

$$ind_1 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - x_{i_1}}{x_i}\right)^2}\ ;$$
$$ind_2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - x_{i_2}}{x_i}\right)^2} \qquad (1)$$

where N is the samples number of data set, x_i is the i-th sample of the collected data set relevant to the concentration values of benzene, x_{i_1} is the i-th sample, relevant to its interpolating curve and x_{i_2} is the i-th sample relevant to the curve obtained using the interpolating curve of CO and applying the identified linear correlation.

The percentage difference between the cited two indexes is resulted equal to 4,5% and this can represent a first validation of the so-identified correlation.

Then, the accuracy of benzene behaviour estimate, obtained by using of linear relationship with CO, has been calculated. The following expression indicates the carried out relation between CO and benzene.

$$Be(t) = m \cdot CO(t) + q \qquad (2)$$

where *Be(t)* and *CO(t)* represent the values of interpolation curves relevant to average of benzene and CO concentrations, respectively, while m and q indicate the linear regression coefficients.

By supposing to consider the concentration measurements of the two analysed pollutants as random variables with constant variance, the coefficients m and q can be considered random variables with normal distribution too [4]. In this case, the standard deviations of m and q are expressed by means the following relationships:

$$\sigma_m = \sqrt{\frac{1}{N-2}\cdot\frac{\sum_{i=1}^{N}\left(y_i-\bar{y}\right)^2}{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}}$$
$$\sigma_q = \sqrt{\frac{1}{N\cdot(N-2)}\cdot\frac{\sum_{1=1}^{N}\left(y_i-\bar{y}\right)^2\sum_{i=1}^{N}\left(x_i\right)^2}{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}} \qquad (3)$$

where x_i and y_i represent the i-th value of CO and benzene interpolation curves, respectively, $\bar{x}$ and $\bar{y}$ are the average values of CO and benzene interpolation curves, respectively, and N is the number the collected values of the two considered pollutants concentrations.

By means of eq. (3), the standard deviation of benzene interpolation polynomial, based on the use of linear regression, can be carried out and expressed by following relationship:

$$\sigma_{Be} = \sigma_m \cdot CO + \sigma_q \qquad (4)$$

where all the quantities are relevant to the same time instant.

So, it is possible to estimate the benzene behaviour by using equation (2) and to obtain a curve that is always inside

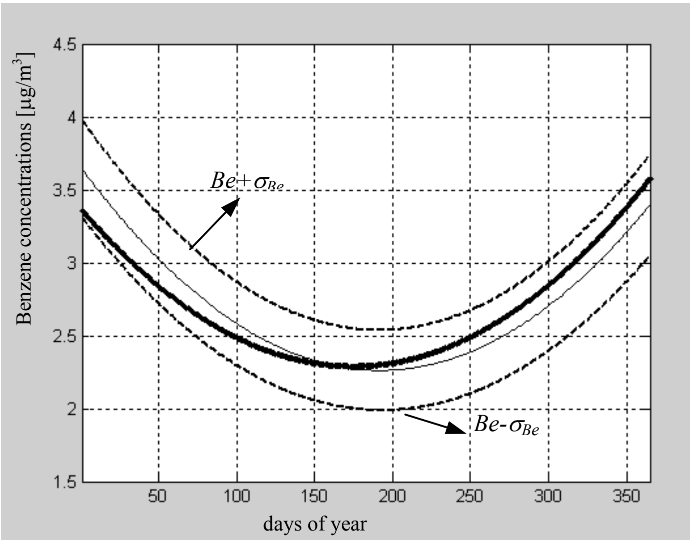

Fig. 3 Behaviours of benzene interpolation curve (continuous line), of estimate obtained by using linear regression (bold line), of benzene curve ± σ_{Be} (dashed lines)

the uncertainty band defined as $\left[Be-\sigma_{Be}\ ,\ Be+\sigma_{Be}\right]$ (see Fig. 3).

Having determined the relationship between interpolation curves of two pollutants and having verified the reliability of obtained estimate, the reconstruction of benzene daily-average values has been carried out, applying eq. (2) to CO daily-average values. Once more, the reconstruction curve is inside in uncertainty band previous calculated as shown in Fig 4.

To increase the reliability of benzene behaviour reconstruction, possible correlations with other substances have been investigated. In this context a 20-order vector correlation coefficients, between all quantities analysed (20 in all) in measurement stations, has been calculated.

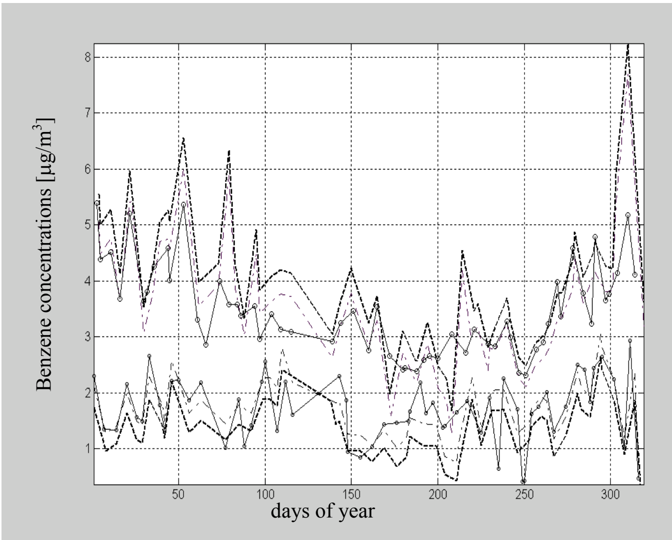

Fig.4 Envelops Behaviours of benzene concentrations (dash-dot lines) of estimate obtained using the linear regression (lines with circle markers), of uncertainty bands benzene $\pm\,\sigma_{Be}$ (bold dashed lines)

Particularly, a strong correlation between with toluene has been individuated. So, a linear relationships between the trend of these three pollutants has been implemented by means the following expression:

$$Be(t)=c_1\cdot CO(t)+c_2\cdot Tou(t)+c_3 \qquad (5)$$

where *Tou(t)* indicates the behaviour of interpolation polynomial relevant to of toluene daily average concentrations and c_1 ,c_2, c_3 are the liner interpolation coefficients.

Using eq. (5) permits us to halve the error of benzene concentrations estimate with respect the error relevant the use of a correlation with a single substance CO.

4. CONCLUSION

Environmental data relevant to the concentration measurements of air pollutants have been analysed and elaborated. Particular attention has been turned on behaviour of substances that more influence the pollution in very high road traffic. Some mathematic models has been developed to characterize the trend of quantities analysed and to estimate the reliability of implemented method.

Finally, using correlations techniques and calculating the obtained accuracy, we have proposed an estimate of concentrations of substance benzene, only on the basis of knowledge of the concentrations of substance CO, and/or other ones.

The so-obtained correlation-based estimate is very useful for operations of function-reconstruction, especially when the incorrect and/or missing data are available.

REFERENCES

[1] M. Chin., D.J. Jacob, J.W. Munger, D.D. Parrish, B.G. Doddrige, *Relationship of ozone and carbon monoxide over North America.* Journal of Geophysical Research Vol. 99, July, N. D7: 14565-14573, 1994.

[2] S. Bendat, A.G. Piersol, "Analysis and measurements procedures", J. Wiley & Sons, New York, 1971

[3] R. Paroli, L. Spezia, "*Analysis of the dynamics of an air pollution by Gaussian Hidden Markov Models* " Proceedings of the International Workshop of statistical Modelling, pp. 493-496, Bilbao, 2000.

[4] J.R. Taylor, "An introduction to error analysis", *University Science Books*, Sausalito, CA, 1997.

Authors: Gregorio Andria, Giuseppe Cavone, Vincenzo Di Lecce, Anna M. L. Lanzolla, Department of Electrics and Electronics, Bari Polytechnic, 4, Via Orabona, 70125, Bari, Italy phone +39080 5963436, fax +39080 5963410, E-mail labmis@poliba.it