Air pollution prediction models of particles, As, Cd, Ni and Pb in a highly industrialized area.

A. B. Vicente\(^{(1)}\), M. M. Jordán\(^{(2)}\), Teófilo Sanfeliu\(^{(1)}\) A. Sánchez\(^{(3)}\)& Mª D. Esteban\(^{(3)}\).

(1) Department of Agricultural and Environmental Sciences. Jaume I University. Campus Riu Sec s/n 12071 Castellón, Spain (e-mail: avicente@sg.uji.es).


ABSTRACT: The objective of this study was to elaborate a series of mathematical models with the aim of short-term prediction of TSP, PM10, As, Cd, Ni and Pb in air ambient. These pollutants depend on some known variables (meteorological variables). The goal is to provide a useful instrument to alert the population facing possible episodes of high concentrations of atmospheric pollutants. The study was carried out in a highly industrialized area in the ceramic cluster of Castellón during five years (2001-2005). The origin of the contamination in this area is both natural and anthropogenic. The natural origin is due to the resuspension of mineral materials from the surrounding mountains and from the long-range transport of materials from North Africa. The anthropogenic contamination sources that stand out include the non-metallic mineral material industries (ceramic production), chemical industries (color, frit and enamel manufacturing) as well as vehicular traffic. Once the particle samples were
collected in quartz fiber filters, the concentration levels of TSP and PM10 were determined gravimetrically. The chemical analysis of the filters was carried out by ICP-MS. Predictive models have been constructed by using Multiple Regression Analysis together with Time Series Models (ARIMA). The SPSS 14.0 statistical software has been employed to analyse the obtained experimental data.

**KEYWORDS:** air pollution, ambient air, prediction models, public health, TSP, PM10, heavy metals.

**INTRODUCTION**

The high development reached by our society, due to a constant increase in the demand of consumer goods and from the requirements of technological advances, brings as a consequence the use of raw materials and their subsequent industrial transformation. All this introduces large quantities of chemical substances into the atmosphere, whose behavior in the natural environment and the effects upon living organisms and material goods are unknown in some cases (Boix et al. 2001; Jordan et al. 2005). Furthermore, they influence increasing global problems affecting the climate (Kelessis 2001), such as global warming, environmental acidification, photochemical smog and ozone layer depletion (Sivakumar 2006, McMicahel et al. 2006, Kantarci et al. 2001).

The requirement for clean and pure air comes from treating it as a limited and common good, indispensable for life. The Earth’s atmosphere is finite and its self-cleansing capacity has limits. As such its utilization must be subjected to practices
that prevent the deterioration of its quality, from either use or abuse, in such a way that its purity is preserved for a guaranteed normal development of living beings (Vicente et al. 2007).

In the present study an assessment of atmospheric air quality was conducted, keeping in mind the atmospheric particulate, TSP (total suspended particles) and PM10 (particles less than 10 µm) in a Spanish coastal area (municipality of Vila-real, Fig. 1), for 5 years (2001-2005). The objective was to elaborate a series of mathematical models with the aim of short-term prediction of these pollutants that depend on some meteorological variables. The goal is to provide a useful instrument to alert the population facing possible episodes of high concentrations of atmospheric pollutants.

**DESCRIPTION OF THE STUDY AREA**

The study area is located in an important ceramic industrial zone in the province of Castellón, East Spain (Fig. 1). This area produces approximately 93% of the Spanish ceramic tiles and 95 % of its frits, enamels and colors. It is at the E-SE part of the province, and covers approximately 1000 Km². Given the density of industries in Vila-real, this municipality was chosen from the ceramic industrial cluster. This city, with approximately 49,000 inhabitants, is located at 39º 56’ North, 0º 56’ West, and is 46 meters above sea level. The population nucleus resides between two fluvial basins, the River Millars to the North and the Riu Sec to the South.
The origin of pollutants in this area is both natural and anthropogenic. The natural origin is due to the resuspension of mineral materials from the surrounding mountains (Gómez et al. 2004) and from the long-range transport of materials from North Africa (Rodríguez et al. 2004, Pérez et al. 2006). The anthropogenic pollution sources (Fig. 2) are the traffic (mobile sources), and also originate from the manufacturers of ceramic tiles and raw material as frits, enamel and ceramic colours (fixed sources) (Sanfeliu et al. 2002, Vicente et al. 2007).

**SAMPLING CONDITIONS**

The sampling station was set up, conforming to the guideline of the implantation of European Council Directive 1999/30/EC, in the south-western part of the city (UTM: X 746.543 Y 4.424.906). In order to avoid measuring microclimates, it was located in an open area covering at least 500 m². There were no restrictions to the airflow around the sampling entrance point, established approximately 3 meters above ground level on a special metallic platform. There were no local emission sources nearby, thus avoiding a distortion of the sample due to the influence of smoke plums from specific pollutants.

The technology of the equipment used consists in blowing air through an inlet with a vacuum pump. According to the characteristics of the inlet, different types of atmospheric particles were collected, TSP or PM10 (Vicente et al. 2007). Particles were trapped on a permeable support, being a 47mm diameter filter. Quartz fiber filters were used in the sampling. They are made of a pure SiO₂ base,
with a totally free of additives. Polystyrene cassettes were used to protect the filters during transport and storage.

A PM10 medium volume sampler was used, model IND-LVS3, manufactured by Kleinfiltergerät. This device is considered as a reference according to European regulations (European Council Directive 1999/30/EC; UNE-EN 12341:1999) for the sampling of PM10 particles. The particulate matter was blown in through the opening circumference between the frame and round cover mounted on top. Within the sampler inlet the airflow was accelerated by eight impactor nozzles and then directed toward the impacting surface. The device contains a temperature sensor, with a radiation protector that eliminates deviations in the reading due to solar radiation. It also has a pressure sensor. The sampling volume of flow was 2.3m$^3$/h during 24-h periods. A total of 887 PM10 samples were collected.

The captor for TSP sampling used was a high-volume MCV-8D sampler, to fulfill European Council Directive 80/779/EEC. The flow volume was 1.5m$^3$/h during 24-h periods. A total of 1003 TSP samples were collected.

Meteorological dates were also obtained from an automatic meteorological station (Weather Monitoring II-DAVIS). Daily dates were used to build a matrix for each sampling period with the following parameters: daily mean temperature, daily mean atmospheric pressure, daily mean humidity, wind direction, wind speed and daily total precipitation.
METHODOLOGY

Gravimetrical analysis

Particle concentration levels were determined gravimetrically. This method consists in weighing the empty filters and with sample. The filters must be kept for at least 48-h in a special chamber (Fig. 3 and 4). The conditions inside the chamber are 50% relative humidity and 20ºC temperature, according to normative UNE–EN 12341:1999.

Filters were weighed on an analytical scale with a precision of 0.1 mg (Fig. 5). The PM concentration levels are determined from the sample quantities obtained and the volume of air pumped using the expression:

\[ C_{PM} = \frac{(P_m - P_v) \times 10^6}{V_{air}} \]

Being:
- \( C_{PM} \): particle concentration in µg/m³
- \( P_m \): Weight of the sampled filter in gr
- \( P_v \): Empty weight of the filter in gr
- \( V_{air} \): Volume of air pumped in m³

Analytical analysis

The As, Cd, Ni and Pb levels in the PM10 samples were determined by ICP-MS. The equipment utilized is a PerkinElmer model Elan-6000, with a 3600 lines/mm holographic network, 1 meter focal distance, 0.26nm/mm linear dispersion, 27.12 MHz frequency and 1.60 kw maximum power. The instrumental technique allows a rapid way to determinate As, Cd, Ni and Pb concentration after dissolution of the sample. The dissolution is obtained by acid digestion in hermetic Teflon
recipients. This methodology has been used by many authors (Kubilay et al. 1995 and Querol et al. 2000).

In order to determine the possible traces of As, Cd, Ni and Pb that the reagents and quartz filter fibers might contain, which give rise to sample contamination, digestions with only reagents (blank reagents) and filters without a sample (blank filters) are performed. The SRM 1648 “urban particulate matter” pattern was used to validate the results, whose composition is particulate matter of anthropogenic origin which was collected in an industrialized urban atmosphere and was adequate for use as a reference standard.

With the values obtained in ppb from the aforementioned technique described, the As, Cd, Ni and Pb concentration levels were determined, keeping in mind the possible traces that the reagents and quartz fiber filters utilized in the collection of samples may contain. The values are expressed in micrograms of As, Cd, Ni and Pb in PM10 per cubic meter of air.

Statistical analysis
As available data is time dependent, predictive models have been constructed by using Multiple Regression Analysis (Rencher 2000) together with Time Series Models (Wei 1990). We recall that one of the standard hypotheses in Multiple Regression Analysis is the independence of the residual errors. However if variables are periodically observed over time, a lack of independence occurs and
therefore straightforward linear regression predictions are inefficient. To avoid this difficulty, Time Series analysis, and more concretely ARIMA models, were used for modeling the behavior of residual errors. The SPSS 14.0 statistical software has been employed to analyze the experimental data obtained.

ATMOSPHERIC POLLUTING AGENTS PREDICTION MODELS

The objective of this study is to construct a model to obtain predictions of a dependent variable based on several independent variables with information over time. The dependent variables are: TSP and PM10 in µg/m³, As, Cd, Ni, and Pb in ng/m³. Independent variables are the following meteorological variables: average wind speed (Vav, m/s), maximum wind speed (Vmax, m/s), predominant wind direction (DirDomi, degrees), average temperature (Tav, degrees Celsius), relative humidity (RH in %) and pressure (P, milibares). All the independent variables were initially considered in the model, but only those that were significant ones are finally used to obtain predictions. The values of the different variables used were daily averages.

The models obtained are the following, where ε(t) is the prediction error, and is estimated as the difference between the observed value and the predicted value in the model. a(t) is white noise with standard normal distribution.

**TSP model**

\[
TSP(t) = -1256.970 + 2.545 \cdot Tav(t) - 0.055 \cdot DirDom(t) + 1.282 \cdot P(t) + \varepsilon(t)
\]

\[
\varepsilon(t) - \varepsilon(t-1) = -0.549 \cdot \varepsilon(t-1) - 0.316 \cdot \varepsilon(t-2) + a(t)
\]

\[
\varepsilon(t) = 0.451 \cdot \varepsilon(t-1) - 0.316 \cdot \varepsilon(t-2) + a(t)
\]

\[
TSP(t) = -1256.970 + 2.545 \cdot Tav(t) - 0.055 \cdot DirDom(t) + 1.282 \cdot P(t) + 0.451 \cdot \varepsilon(t-1) - 0.316 \cdot \varepsilon(t-2) + a(t)
\]
The correlation coefficient obtained as a result of the primary multiple regression fitting is $R^2=0.55$. The residuals obtained do not follow a normal distribution after being tested with the corresponding goodness-of-fit test in this case. With regard to fitting residuals with the ARIMA model, the autocorrelation and partial autocorrelation functions of the residuals are presented in the figure 6, after analysing several ARIMA models, modelling with an AR(2) model was decided, which is also a totally valid model for making predictions. In this case, we note how the stationary conditions are fulfilled as far as the estimations of the calculated model parameters are concerned. In table 1, we may also observe how the coefficients significantly differ from zero.

TSP are solid or liquid particles with high sedimentation velocities and relatively short residence periods in the atmosphere. Their size range is between 0.1 and 32μm.

TSP presents a direct relationship with the temperature and atmospheric pressure. In the case of the temperature, this is due to increase that cause the dryness of the substrate in the area and which favors resuspension of the particles from this substrate (Colombo et al. 1999). This fact causes an increase of TSP concentration levels in the atmosphere. On the other hand, when the pressure increases there is less expansion in the lowest layers of the atmosphere, which carry less dispersion of the pollutants with it, and their concentration increase (Wark et al. 2000).
The relationship between TSP and predominant direction is inverse. The concentration of TSP increases under low degree directions, for example NE (45°) winds, and decreases if the predominant wind is W (270°) or NW (315°). This behavior is logical if we bear in mind that the greatest sources of TSP are the ceramic sector businesses located between the SW (225°) and NW (315°) in relation to the sampling point (Fig.2).

**PM10 model**

\[ PM10(t) = -1110.954 + 1.168 \ Tav(t) - 0.024 \ DirDom(t) + 0.257 \ RH(t) + 1.112 \ P(t) + \epsilon(t) \]

\[ \epsilon(t) - \epsilon(t-1) = -0.552 \ \epsilon(t-1) - 0.342 \ \epsilon(t-2) + \ a(t) \]

\[ \epsilon(t) = 0.448 \ \epsilon(t-1) - 0.342 \ \epsilon(t-2) + \ a(t) \]

The correlation coefficient obtained as a result of the primary multiple regression fitting is \( R^2 = 0.45 \). The residuals obtained do not follow a normal distribution after being tested with the corresponding goodness-of-fit test in this case too. With regard to fitting residuals with the ARIMA model, the autocorrelation and partial autocorrelation functions of the residuals are presented in the figure 7, after analysing several ARIMA models, modelling with an AR(2) model was decided, which is also a totally valid model for making predictions. In this case, we note how the stationary conditions are fulfilled as far as the estimations of the calculated model parameters are concerned. In table 2, we may also observe how the coefficients significantly differ from zero.
Like in the TSP model, PM10 (particles between 0.1-10µm) depend in the same way upon the meteorological variables of average temperature, predominant wind direction and atmospheric pressure. The explanation is the same for TSP.

Unlike the TSP model however, the PM10 model reveals a direct relationship between this pollutant and the relative humidity (RH). This new variable introduces the influence of traffic on the PM10. The particles from traffic that contribute to PM10 present a small particle size, that is, those within the PM10 range are in the smallest region, of around 1-2µm. These particles are influenced by condensation processes which, at the same time, are governed by environmental humidity. More humidity leads to greater condensation (McGregor 1999), and to increase concentrations of these particles.

**Arsenic model**

\[
As(t) = 13.104 + 0.064 \, PM10(t) - 0.333 \, Tav(t) + \varepsilon(t)
\]

\[
\varepsilon(t) = \varepsilon(t-1) + 0.576 \, \varepsilon(t-1) - 0.239 \, \varepsilon(t-2) + a(t)
\]

\[
As(t) = 13.104 + 0.064 \, PM10(t) - 0.333 \, Tav(t) + 0.448 \, \varepsilon(t-1) - 0.342 \, \varepsilon(t-2) + a(t)
\]

For arsenic, the correlation coefficient obtained as a result of the primary multiple regression fitting is $R^2=0.52$, and the residuals obtained do not follow a normal distribution after being tested with the corresponding goodness-of-fit test. The autocorrelation and partial autocorrelation functions of the residuals (Fig. 8) show that the ARIMA AR(2) model is the best option. We note how the stationary conditions are fulfilled as far as the estimations of the calculated model
parameters are concerned. In the table 3, we may also observe how the coefficients significantly differ from zero.

Arsenic is a very volatile element that is presented in the lowest grain size fractions of PM10, 74.8% of arsenic is found in the fraction <0.6µm (Fernández et al. 2001). It inversely depends on temperature. Arsenic oxides, from different local sources, condense upon contact with the atmosphere; the aerosol-vapor equilibrium depends on the temperature. At low temperatures the aerosol phase is favored and at high temperatures the vapor phase (Pallarés et al. 2007). Because of this, the arsenic concentration increases at low temperatures just as the model demonstrates. Arsenic is shed by several of aforementioned sources, it must be noted that it stems from the combustion of combustible fossils, which are used in many fields. From here on it does not depend upon meteorological factors other than wind direction.

**Cadmium model**

\[
Cd(t) = 1.541 + 0.034 \, Vmax(t) + \varepsilon(t)
\]

The correlation coefficient obtained as a result of the primary multiple regression fitting is \( R^2 = 0.6 \), the residuals obtained follow a normal distribution after being tested with the corresponding goodness-of-fit test. In this case, and given the autocorrelation functions of the residuals (Fig. 9), the use of the former regression model to make predictions was decided rather than any type of ARIMA model.
Cadmium, according to the proposed model, is related with the maximum wind velocity. This element is the chemical element studied that appears in lesser concentrations in PM10. Cadmium is found in 61.3% of that associated with the fraction <0.6µm (Fernández et al. 2001), from which it has long residence times in the atmosphere (Röösli et al. 2001). By being the element with the lowest concentration, the variations it can undergo are influenced by external contributions when suitable atmospheric conditions take place with the arrival of pollutants from locations at considerable distance from the sampling point. Thus, there the only significant variable that appears in the proposed model is maximum wind velocity. Therefore, cadmium concentration levels increase with the velocity, disperse a large amount of particles and arrive at the sampling point from external sources in large numbers (Fig. 2).

Nickel model

\[
Ni(t) = -2.682 + 0.045 \, PM10(t) + 0.083 \, Tav(t) + 0.009 \, DirDom(t) + 0.044 \, RH(t) + \varepsilon(t)
\]

\[
\varepsilon(t) - \varepsilon(t-1) = -0.540 \, \varepsilon(t-1) - 0.256 \, \varepsilon(t-2) + a(t)
\]

\[
\varepsilon(t) = 0.460 \, \varepsilon(t-1) - 0.256 \, \varepsilon(t-2) + a(t)
\]

The correlation coefficient obtained as a result of the primary multiple regression fitting is \(R^2=0.40\), and the residuals obtained do not follow a normal distribution after being tested with the corresponding goodness-of-fit test. Once again, modelling with an AR(2) model was done as it was considered to be the best option for making predictions (Fig. 10). In this case, we note how the stationary
conditions are fulfilled as far as the estimations of the calculated model parameters are concerned. In the table 4, we may also observe how the coefficients significantly differ from zero.

Nickel, like the previous chemical elements, is found in a high percentage, 57.5% in the fraction <0.6µm (Fernández et al. 2001). The main origin of nickel is the combustion of combustible fossils, as it is found in petroleum in traces. The main sources of this pollutant are the power station and the petrochemical plant in the “Grao de Castellón”, located to the SE of the sampling point (Fig. 2). For this reason the model presents a direct relationship with the dominant wind direction. At wind directions between SW (225º) and NW (315º), the arrival of pollutants from the “Grao de Castellón” is favored and the nickel concentration levels increase at the sampling point. Nickel also presents a direct relationship with the temperature. The main emission sources of nickel, the power station and petrochemical plant, emit this pollutant in the form of airborne ashes and not in a gaseous state (Boix et al. 2001). These airborne ashes have a very fine grain size and the temperature influences their dispersion. A greater atmospheric turbulence is noted at high temperatures (Wark et al. 2000), and this is due to a greater dispersion, which make that these ashes arrive from greater distance, and increase the nickel concentration levels increase at the sampling point, just as the model presents. Relative humidity is another significant variable in this model; it presents a direct relationship with the nickel. A higher relative humidity means greater condensation (McGregor 1999), and increases concentration levels of this pollutant.
**Lead model**

\[ Pb(t) = -5072.687 + 1.474 \times V_{\text{max}}(t) - 0.175 \times \text{DirDom}(t) + 1.651 \times RH(t) + 5.105 \times P(t) + \varepsilon(t) \]

\[ \varepsilon(t) - \varepsilon(t-1) = -0.555 \times \varepsilon(t-1) - 0.297 \times \varepsilon(t-2) + a(t) \]

\[ \varepsilon(t) = 0.445 \times \varepsilon(t-1) - 0.297 \times \varepsilon(t-2) + a(t) \]

\[ Pb(t) = 5072.687 + 1.474 \times V_{\text{max}}(t) - 0.175 \times \text{DirDom}(t) + 1.651 \times RH(t) + 5.105 \times P(t) + 0.445 \times \varepsilon(t-1) - 0.297 \times \varepsilon(t-2) + a(t) \]

The correlation coefficient obtained as a result of the primary multiple regression fitting is \( R^2 = 0.58 \), and the residuals obtained do not follow a normal distribution after being tested with the corresponding goodness-of-fit test. Once again, modelling was done an AR(2) model as it was considered to be the best option for making predictions (Fig. 11). In this case, we note how the stationary conditions are fulfilled as far as the estimations of the calculated model parameters are concerned. In table 5, we may also observe how the coefficients significantly differ from zero.

Lead has the highest concentration levels in PM10 than the other studied elements. Notwithstanding, the variation of its concentration levels with the PM10 variation is not sufficiently significant for this variable to appear in the proposed model. Like the other chemical elements, it is associated with the finest PM10 fraction, 69.0% is found in the fraction < 0.6µm (Fernández et al. 2001). The main emission sources of lead are the ceramic industries located between the NW (315º) and SW (225º) from the sampling point (Fig.2). This pollutant presents an inverse relationship with the dominant wind direction and a direct relationship
with the maximum wind velocity. We must be in main that this element is heavier than those previous ones and presents a greater concentration. Although its emission source is closer than the previous elements, greater wind velocities are needed to displace it. This model also presents a direct relationship between the lead and the relative humidity and the atmospheric pressure. As explained in previous models, higher relative humidity favors condensation (McGregor 1999), and increases the concentration levels of this pollutant. Likewise an increase in pressure reduces the expansion of the lowest atmospheric layers, and pollutant dispersion decreases (Wark et al. 2000), leads to increase concentration increases of lead.

**CONCLUSIONS**

Some easy applied prediction models of the studied atmospheric pollutants have been developed and are presented as a useful to alert the population for possible episodes of high contamination.

The independent variables, considered to be significant in the final prediction models, mark the importance of the types of anthropogenic emissions that occur in the study area, and the influence they exercise upon the concentration levels of the different atmospheric pollutants. Therefore, the variables of predominant direction and maximum wind velocity indicate from which industrial park the contamination comes; the average temperature indicates the relationship with the types of industrial processes developed in the area, and this variable along with
the atmospheric pressure and relative humidity mark the convective dynamic of
the atmosphere that carries a greater or lesser dispersion of the pollutants.

REFERENCES

Boix A., Jordán M.M., Querol X., Sanfeliu T. 2001 Characterization of total
suspended particles around a power station in an urban coastal area in eastern

Colombo J.C., Landoni P., Bilos C. 1999. Sources, distribution and variability of
airborne particles and hydrocarbons in La Plata, Argentina. *Environmental

values and guide values for sulphur dioxide and suspended particulates. Official

for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter
and lead in ambient air. Official Journal of the European Union **L163**, 26th
September, 1999, 41-60.


UNE-EN 12341:1999 Air quality: Determination of the PM10 fraction of the suspended particulate matter. Reference method and field test procedure to demonstrate reference equivalence of measurement methods.
