Science of the Total Environment 554–555 (2016) 155–166

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# A proposed methodology for the assessment of arsenic, nickel, cadmium and lead levels in ambient air

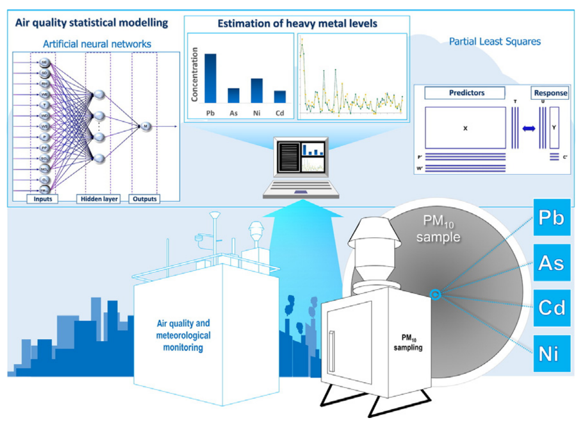
Germán Santos *, Ignacio Fernández-Olmo

*Dpto. de Ingenierías Química y Biomolecular, Universidad de Cantabria, Santander 39005, Spain*

## HIGHLIGHTS

- EU encourages modelling techniques over measurements for air quality assessment.
- A methodology for minor pollutants assessment by statistical modelling is presented.
- PLSR and ANNs represent valid approaches fulfilling EU uncertainty requirements.
- Inputs related to atmospheric stability can slightly improve models performance.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 16 December 2015
Received in revised form 24 February 2016
Accepted 25 February 2016
Available online xxxx

Editor: D. Barcelo

*Keywords:*
Air quality assessment
Atmospheric stability
Methodology
Metals
Statistical models

## ABSTRACT

Air quality assessment, required by the European Union (EU) Air Quality Directive, Directive 2008/50/EC, is part of the functions attributed to Environmental Management authorities. Based on the cost and time consumption associated with the experimental works required for the air quality assessment in relation to the EU-regulated metal and metalloids, other methods such as modelling or objective estimation arise as competitive alternatives when, in accordance with the Air Quality Directive, the levels of pollutants permit their use at a specific location. This work investigates the possibility of using statistical models based on Partial Least Squares Regression (PLSR) and Artificial Neural Networks (ANNs) to estimate the levels of arsenic (As), cadmium (Cd), nickel (Ni) and lead (Pb) in ambient air and their application for policy purposes. A methodology comprising the main steps that should be taken into consideration to prepare the input database, develop the model and evaluate their performance is proposed and applied to a case of study in Santander (Spain). It was observed that even though these approaches present some difficulties in estimating the individual sample concentrations, having an equivalent performance they can be considered valid for the estimation of the mean values – those to be compared with the limit/target values – fulfilling the uncertainty requirements in the context of the Air Quality Directive. Additionally, the influence of the consideration of input variables related to atmospheric stability on the performance of the studied statistical models has been determined. Although the consideration of these variables as additional inputs had no effect on As and Cd models, they did yield an improvement for Pb and Ni, especially with regard to ANN models.



* Corresponding author.
*E-mail address:* santosg@unican.es (G. Santos).

http://dx.doi.org/10.1016/j.scitotenv.2016.02.182
0048-9697/

## 1. Introduction

The member states of the European Union are obliged by the Air Quality Framework Directive, Directive 2008/50/EC (EC, 2008), to carry out the air quality assessment of certain specified pollutants according to a group of common methods and criteria. These methods vary depending on the population within each air quality zone or agglomeration and the relative levels of the mentioned pollutants compared with the respective assessment thresholds, which are expressed as a percentage of the corresponding limit/target value (EC, 2004; EC, 2008). Thus, fixed measurements are required when it has been confirmed that the levels of an atmospheric pollutant exceed the upper assessment threshold. A combination of fixed measurements and modelling techniques and/or indicative measurements is possible when the pollutant concentration is between both assessment thresholds. Moreover, air quality assessment based exclusively on modelling and/or objective estimation is solely reserved for zones and agglomerations with relatively good air quality and no large conurbations, in which at a preliminary assessment stage the levels of those pollutants whose air quality is to be assessed consistently fall sufficiently below their respective lower assessment thresholds (EC, 2008). In this sense, the minimum quality requirements for the objective estimation techniques to be acceptable in the context of Directive 2008/50/EC, which include uncertainty and minimal data capture and time coverage, are less strict than those for modelling techniques and experimental measurements, either fixed or indicative. Specifically, the relative uncertainty should be lower than 100% (EC, 2008). Although no indications on the nature of these methods are provided in the European Directives, the European Commission in its report for guidance on assessment under the European Union Air Quality Directives interprets objective estimation techniques as "mathematical methods to calculate concentrations from values measured at other locations and/or times, based on scientific knowledge of the concentration distribution". In this context, statistical models can be regarded as objective estimation techniques in the sense that they are based on statistical data analysis establishing empirical relationships between ambient concentrations and meteorological variables instead of simulating the relationship between emissions and immission concentrations through the description of the physical phenomena that rules the air transportation and dispersion of pollutants in the atmosphere, as with classical deterministic models (Daly and Zannetti, 2007). It consists of generating estimations of the European regulated pollutant concentrations directly from regional air quality information, which indirectly reflects atmospheric pollution in part due to local emission sources. Therefore, there is an interest in deepening the understanding of this sort of methods to be used as air quality assessment tools as they are cost-effective and simpler than other approaches, such as air dispersion modelling which can also be considered more affordable than experimental measurements, especially for certain pollutants, namely metals and polycyclic aromatic hydrocarbons, whose analytical determination, unlike other atmospheric pollutants, is expensive and time consuming.

The use of modelling techniques is encouraged even when continuous monitoring is compulsory because of its capacity of providing supplementary information to air quality monitoring and future projections regarding public exposure under different emission scenarios. Furthermore, modelling is becoming a prominent air quality assessment tool as it has been reported that different modelling approaches have already been incorporated as part of the routine air quality network monitoring and assessment procedures in a number of countries, e.g., Finland, Norway, Sweden or United Kingdom. For those reasons, the use of models and other alternative air quality assessment tools for policy support is expected to continue increasing in the future.

Although the use of statistical modelling techniques to predict immission concentrations of major atmospheric pollutants has been extensively investigated in the literature, there are not many studies addressing the estimation of the levels of compounds bound to particulate matter; metals in particular. Artificial neural networks (ANNs), state-space modelling, time series autoregressive modelling and linear regression stand out among the approaches used to that end. With respect to ANNs, Chelani et al. (2002) estimated the ambient air levels of Cd, Cr, Fe, Ni, Pb, Zn as well as $PM_{10}$ in the city of Jaipur, India, by means of artificial neural network models obtaining low values of root mean square error (RMSE). In addition, Li et al. (2009) reconstructed occupational manganese exposure by means of back-propagation artificial neural networks and multiple linear regression. In a previous study, a state-space model coupled with Kalman filter and an autoregressive model with external input (ARX model) was used by Chelani et al. (2001) to predict the concentration of Pb, Fe and Zn and respirable suspended particulate matter in Delhi. Apart from ANNs, some research has been conducted to model metal concentrations in ambient air using other statistical approaches. Hernández et al. (1992) applied state-space modelling, Box-Jenkins modelling and time series autoregressive integrated moving average (ARIMA) models to estimate the daily concentrations of air-particulate Fe and Pb in Madrid (Spain). Predictions of daily Fe were better than those of Pb. No difference being found between State-space and Box-Jenkins models, their outcomes were better than those of ARIMA models in terms of root mean squared error (RMSE), correlation coefficient and efficiency. The state space model performed better than the ARX model. On the other hand, Vicente et al. (2012) developed predictive models based on multiple regression analysis together with time series (ARIMA) models to predict the concentration of total suspended particles (TSP), $PM_{10}$, As, Cd, Ni and Pb in the ambient air of Castellón (Spain). Arruti et al. (2011) developed statistical models based on multiple linear regression and principal component regression as objective estimation techniques to estimate the immission levels of As, Cd, Ni and Pb at four sampling sites in Cantabria (Northern Spain). Both techniques were found to be valid approaches; however, there was room for improvement with regard to their performance.

The implications of the use of objective estimation techniques by local air quality management authorities to assess air quality when it is possible are discussed in this work. For this purpose, we present a case of study focused on the estimation of ambient air levels of Pb, As, Ni and Cd in Santander (Spain), by means of statistical models based on partial least squares regression (PLSR) and artificial neural networks.

Additionally, given that normally deterministic models take into consideration directly or indirectly the atmospheric stability to simulate the dispersion of pollutants in the atmosphere, this work also aims at determining the influence of the use of input meteorological variables related to atmospheric stability on the performance of the studied statistical models.

## 2. Materials and methods

### *2.1. Case of study: European Union regulated metal(loids) in Santander, Spain*

Santander, with approximately 176,000 inhabitants in 2014 (INE, 2014), is the capital of the Cantabria Region in the North of Spain. The city is extended over a bay and in its suburbs, at 5–10 km, there is an industrial area mainly related to steel and ferroalloys manufacturing plants that makes Santander as the most complex urban area in the region. A sampling campaign was conducted by our research group to determine daily values of $PM_{10}$ (particulate matter with and aerodynamic diameter lower than 10 μm) in 2008, 2009 and 2011 in a coastal sampling site (ETSIIT) located in the rooftop of the building "Escuela Superior de Ingenieros Industriales y de Telecomunicación" (43° 28′ 24″ N, 3° 47′ 54″ W at 23 m.a.s.l) and approximately 1 km far from the sea. Glass microfiber filters (Whatman, 150 mm diameter) were used for $PM_{10}$ sampling during 2008 and 2009, whereas the 2011 campaign was carried out with quartz microfiber filters (Sartorius AG, 47 mm diameter) because of their lower detection limits, mainly for As (0.03 vs.