Available online at www.sciencedirect.com
ScienceDirect
Chemosphere 70 (2008) 1190–1195
CHEMOSPHERE
www.elsevier.com/locate/chemosphere

Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks

M.L. Martín [a,*], I.J. Turias [a], F.J. González [a], P.L. Galindo [a], F.J. Trujillo [a], C.G. Puntonet [b], J.M. Gorriz [b]

[a] *EPS de Algeciras, Universidad de Cádiz, Avda Ramón Puyol s/n, 11202 Algeciras (Cádiz), Spain*
[b] *E.T.S.I.I.T., Universidad de Granada, Spain*

Received 10 May 2007; received in revised form 14 August 2007; accepted 16 August 2007
Available online 24 October 2007

Abstract

The region of the Bay of Algeciras is a very industrialized area where very few air pollution studies have been carried out. The main objective of this work has been the use of artificial neural networks (ANNs) as a predictive tool of high levels of ambient carbon monoxide (CO). Two approaches have been used: multilayer perceptron models (MLPs) with backpropagation learning rule and k-Nearest Neighbours (k-nn) classifiers, in order to predict future peaks of carbon monoxide. A resampling strategy with twofold cross-validation allowed the statistical comparison of the different topologies and models considered in the study. The procedure of random resampling permits an adequate and robust multiple comparisons of the tested models and allow us to select a group of best models.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Air pollution forecasting; Multiple comparison; k-nn; Multilayer perceptron

1. Introduction

Bay of Algeciras region has one of the main industrial estates of Andalusia and Spain, with urban areas that reach more than 250 000 inhabitants ('agglomeration' in the sense of 96/62 EU Directive). This important industrial activity requires the necessary control of its environmental impact. Thus, 96/62 EU Directive states that Member States shall take the necessary measures to ensure compliance with the limit values and shall draw up action plans indicating the measures to be taken in the short term where there is a risk of the limit values and/or alert thresholds being exceeded, in order to reduce that risk and to limit the duration of such an occurrence. Therefore, the prediction of air pollutants has become an important task in the recent years, because the estimation of ground level concentrations gives valuable information for air pollution reduction policies. The dispersion mechanism of atmospheric pollutants in urban areas is quite complex and it depends on different factors: meteorological conditions, orography, 'heat island' effect, etc. All these factors, some of which are out of human control, are decisive in air quality conditions. It is well-known that air pollution episodes can cause important toxic effects especially on high-risk population. Then, the ability to predict the occurrence of peak concentrations will contribute to reduce these effects, either reducing emissions or warning population.

The main purpose of the present work is the application of classification techniques, such as artificial neural networks (ANNs) and k-nn method to identify peaks of carbon monoxide (CO) concentrations in the Bay of Algeciras region, for the period 1999–2001. The importance of input weights has been determined in order to study the relevance of input variables to neural models (Garson, 1991).

A procedure of resampling simulation was designed to avoid variation coming from different sources, thus

* Corresponding author. Tel.: +34 956 028015; fax: +34 956 028014. *E-mail address:* luz.martin@uca.es (M.L. Martín).

0045-6535/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2007.08.039

independence and randomness was guaranteed. In order to estimate the error produced by different models, we adopt the procedure of twofold cross-validation.

A detailed overview of the application of neural techniques on atmospheric sciences can be found in the work of Gardner and Dorling (1998). These authors conclude that models based on ANNs give better results than linear methods, especially when the problem under study is nonlinear. Different methods have been used in the prediction of atmospheric pollution: persistence models (Perez et al., 2000), regression models (Comrie and Diem, 1999), ANNs (Jorquera et al., 1998; Elkamel et al., 2001; Kolehmainen et al., 2001; Kukkonen et al., 2003; Brunelli et al., 2007; Dutot et al., 2007). Comrie and Diem (1999) studied the factors that influence CO concentrations in the area under study (meteorological variables, previous concentrations, traffic density) in order to develop a multiple regression model. Pelliccioni and Poli (2000) used multilayer perceptron models (MLP) for the prediction of CO and nitrogen dioxide (NO_2). Viotti et al. (2002) used an ANN model with a hidden layer to predict short-term and medium-term air pollutant concentrations (CO, ozone and benzene) in an urban area. Kolehmainen et al. (2001) compared different prediction models from meterological and NO_2 concentration data measured in the city of Stocolm for the period 1994–1998. These authors conclude that MLP models predict with a better accuracy than linear and Self Organising Map (SOM) models. Perez (2001) introduced a study to predict hourly mean sulphur dioxide (SO_2) concentrations in the city of Santiago de Chile. They compared the predictions made by persistence models, linear regression models and ANNs. The input variables were SO_2 concentrations, temperature, relative humidity and wind speed. Prediction errors were between 30% and 60%, for short-term and long-term predictions, respectively.

2. Meteorological and air pollution data

CO is a primary air pollutant produced by incomplete combustion processes. In urban areas, traffic is its main source, while its reaction with the OH radical is the main removal mechanism. CO concentrations are usually well correlated with traffic density. 2000/69 EU Directive sets the maximum 8-h CO concentration in 10000 $\mu g\,m^{-3}$. CO concentration in a given time t, CO(t), depends on three main factors: dispersion capacity of the atmosphere, emissions, and previous CO concentrations (accumulated CO) (Maffeis, 1999). CO emissions depend on ambient temperature and source features. This gaseous pollutant exhibits a clear cyclical behaviour, with a characteristic variation in the daytime and night-time and during the week (González Gallero, 2003). Two significant peaks can be observed during working days, which correspond with the hours of higher traffic intensity, the first between 8:00 and 10:00 am and the second one between 7:00 and 9:00 pm (Fig. 1a). In contrast, weekends are characterized by lower CO concentration levels during day-time and higher levels during night-time (Fig. 1b).

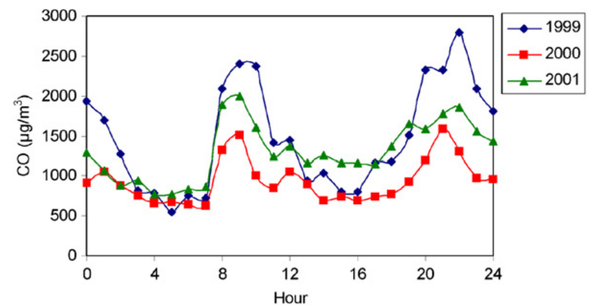

Fig. 1a. CO daily mean concentration (1999, 2000 and 2001).

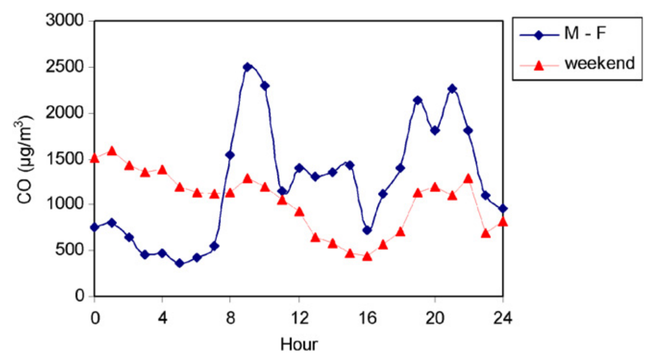

Fig. 1b. CO daily mean concentration for working days and weekends (1999).

Data used in this work have been provided by the 'Consejería de Medio Ambiente' from the Andalusian Regional Government for a period of three years (1999–2001). CO concentration data are from the monitoring station located at the town of Algeciras, which gives data every half an hour. Only the exogen variables measured at the same monitoring station have been used in the present study: wind speed (V) and wind direction (D), temperature (T) and NO concentration. In order to avoid discontinuities, wind direction was transformed using the expression $1 + \sin(D + \pi/4)$ (Chelani et al., 2002).

The maximum 8-h CO concentration of 10000 $\mu g\ m^{-3}$ established by the 2000/69 EU Directive has not been surpassed during the period of analysis. However, local authorities are very interested in assessing the impact of urban traffic on air quality in Algeciras.

In a first step, data have been validated removing data measured during periods of malfunction of the monitoring station. Linear interpolation was used to fill data series when gap was not too large.

3. Prediction models

The classification analysis allows the development of models that are able to predict if one object belongs to a specific class or category, considering different features of the object. The data matrix has at least one categorical

Download English Version:

https://daneshyari.com/en/article/4414148

Download Persian Version:

https://daneshyari.com/article/4414148

Daneshyari.com