PM_{2.5} forecasting in a large city: Comparison of three methods

Patricio Perez^{a,*}, Giovanni Salini^{a,b}

^aDepartamento de Física, Universidad de Santiago de Chile, Av. Ecuador 3493, Santiago, Chile

^bFacultad de Ingeniería, Universidad Católica de la Santísima Concepción, Caupolicán 491, Concepción, Chile

ARTICLE INFO

Article history:

Received 6 March 2008

Received in revised form 21 July 2008

Accepted 24 July 2008

Keywords:

Air quality forecasting

Particulate matter

PM_{2.5}

Santiago, Chile

Neural networks

Hybrid clustering algorithm

ABSTRACT

There is an increasing awareness for the toxic effects produced by the inhalation of fine particles present in the air. It is important then to provide precise information to the population about the concentrations of this pollutant expected for the incoming hours. We present here a study about the capability of three types of methods for PM_{2.5} forecasting one day in advance: a multilayer neural network, a linear algorithm and a clustering algorithm. Input variables are past concentrations measured in four monitoring stations and actual and forecasted meteorological information. Outputs are the maxima of the 24 h moving average of PM_{2.5} concentrations for the following day at the site of the monitoring stations. By training with data from the three previous years, we are able to generate results for the fall–winter period for each year from 2004 to 2007. Although the three methods may be used as operational tools, the clustering algorithm seems more accurate in detecting high concentration situations.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

There is evidence that fine particulate matter (particles with aerodynamic diameter less than 2.5 μm , PM_{2.5}) is more harmful than coarse particulate matter (particles with aerodynamic diameter between 2.5 μm and 10 μm , PM_{10–2.5}) for humans (Cifuentes et al., 2000; Schwartz et al., 2000; Klemm et al., 2000). PM_{2.5} particles can penetrate the respiratory tract and have toxic effects in heart (Brook et al., 2002) and lungs (Pope et al., 2002). However, very few countries have a PM_{2.5} standard. In the United States, a 24-h PM_{2.5} standard of 65 $\mu\text{g m}^{-3}$ was established in 1997, a value that was recently lowered to 35 $\mu\text{g m}^{-3}$. The standard for the one year average was maintained at 15 $\mu\text{g m}^{-3}$. The Canadian Council of Ministers of the Environment (CCME) enforced in 2000 a 30 $\mu\text{g m}^{-3}$ PM_{2.5} (24 h) standard. In Mexico, the current 24-h PM_{2.5} standard is 65 $\mu\text{g m}^{-3}$, and the one year value is 15 $\mu\text{g m}^{-3}$. In Australia the 24-h value is 25 $\mu\text{g m}^{-3}$ and the one year average is

8 $\mu\text{g m}^{-3}$. The World Health Organization has recently recommended that in order to avoid significant harmful effects in the population, the PM_{2.5} 24-h average should not exceed 25 $\mu\text{g m}^{-3}$ and that the annual average must not exceed 10 $\mu\text{g m}^{-3}$ (WHO, 2006).

At present, Chile does not have a PM_{2.5} standard and concentrations in the capital city of Santiago often exceed the recommended maximum levels. As an example, during 2007, the maximum of the 24-h moving average exceeded the 65 $\mu\text{g m}^{-3}$ value on 79 days, and on 22 days it was greater than 100 $\mu\text{g m}^{-3}$. These exceedances were concentrated during the period between April and August indicating a clear seasonal fall–winter effect. In this period strong thermal inversions are observed (which do not occur in spring and summer). On the basis of usual PM_{2.5} emissions in a large city (intense traffic, heating and a number of industries in the suburbs) and a geographical area confined by mountains, it is the meteorology that dominates the occurrence of events of significantly high concentrations of fine particles in the atmosphere (for more details on Santiago geography and climate, see Perez et al., 2000). The relevance of meteorology and geography over emissions is illustrated by the situation of Logan, Utah,

* Corresponding author. Tel.: +56 2 7181257.

E-mail address: pperez@usach.cl (P. Perez).

USA, a city with only 50,000 habitants, where winter thermal inversions during 2004 are strongly correlated with the observation that the 24-h $\text{PM}_{2.5}$ concentration was greater than $100 \mu\text{g m}^{-3}$ on nine occasions (Silva et al., 2007).

The possibility that on some days the population of the city will be exposed to unhealthy concentrations of $\text{PM}_{2.5}$, implies the convenience of a forecasting system which can be used as a prevention tool. Atmospheric particulate matter concentration forecasting models presented in the scientific literature are mostly statistical models. Perez et al. (2000) developed a multilayer neural network (MLP) that predicts 1-h average concentrations of $\text{PM}_{2.5}$ from 1 to 24 h in advance, where errors go from 30% for early hours to 60% for late hours. Initially, only lagged values of the pollutant were considered as input, and it was shown that some improvement was possible by including the values of meteorological variables. Forecasted values of $\text{PM}_{2.5}$ using the MLP were shown to be more precise than those generated with a linear model and persistence. Ordieres et al. (2005) have shown that a neural network approach outperforms persistence and linear regression models when predicting $\text{PM}_{2.5}$ concentrations up to 16 h in advance for an urban area on the US–Mexico border. Recently, a model for forecasting expressway fine particulate matter was reported (Thomas et al., 2007), where 1-h average concentrations 1-h in advance were obtained with a linear model and an MLP neural network. Their conclusion is that with a proper choice of input variables, there is not a noticeable difference between the two approaches. A hybrid clustering algorithm (HCA) has also been proposed for forecasting tasks, and it is claimed that it can outperform neural network models (Sfetsos and Siriopoulos, 2004). This approach was used for PM_{10} forecasting, and the results showed a 10% improvement over neural network models (Vlachogiannis and Sfetsos, 2006). In this work we analyze the $\text{PM}_{2.5}$ data measured in Santiago, Chile between 2001 and 2007 and implement three types of forecasting models: linear, neural network and a clustering algorithm. The outputs are the maxima of the 24-h moving average of $\text{PM}_{2.5}$ concentrations for the following day as measured in four monitoring stations, and the inputs are lagged values of $\text{PM}_{2.5}$ concentrations, and measured and forecasted values of some meteorological variables.

2. The data

Between 2001 and 2007, the official network of air pollution monitoring stations in the city of Santiago, Chile, consisted of seven stations which measured particulate matter and gas pollutants. The location of these stations is shown in Fig. 1. Station B stopped operation on 2003. From these, only four stations measure $\text{PM}_{2.5}$: L, M, N, O. The instrument used to measure fine particle matter concentrations is a mechanical oscillated mass balance (TEOM). Altitude over sea level is increasing from west to east. In average, highest concentrations are observed in station N and lowest concentrations occur in station M. Table 1 shows the one year and fall–winter averages for the four stations during the period under study. This puts in evidence the seasonal effect that is likely to occur in

Fig. 1. Location of atmospheric pollution monitoring stations in the city of Santiago, Chile. Black area represents the urban region. To the right we observe the Andes mountains which spread in the north–south direction to the east of the city. $\text{PM}_{2.5}$ is measured in stations O, N, M, L.

a southern hemisphere location like Santiago. Fig. 2 shows Santiago in the context of the southern part of South America. A typical map of sea level pressures (in units in excess of 1000 hPa) for the region is displayed. A is for high pressures and B for low. The low near Santiago has its minimum by the coast, at approximately 200 km from the capital of Chile. This situation implies air descent over the Santiago region which compresses the mixing layer, with the result that air pollutants emitted in the city become trapped. This type of configuration is observed more frequently and with more intensity during fall and winter.

Table 1

$\text{PM}_{2.5}$ averages for years between 2001 and 2007 in $\mu\text{g m}^{-3}$

	Station L	Station M	Station N	Station O
2001 (12)	35	26	34	33
2001 (5)	44	31	48	48
2002 (12)	32	25	35	34
2002 (5)	43	30	50	51
2003 (12)	34	28	40	36
2003 (5)	48	34	60	55
2004 (12)	36	26	35	34
2004 (5)	47	30	48	48
2005 (12)	34	27	33	32
2005 (5)	41	29	42	41
2006 (12)	35	25	34	32
2006 (5)	46	31	48	47
2007 (12)	34	26	34	33
2007 (5)	46	33	50	50

Number (5) means that the average is taken between April and August and (12) means that the average is taken for the whole year, between January and December.