

A model for PM_{2.5} forecasting in Santiago, Chile

P. Perez & P. Rodriguez

Departamento de Fisica, Universidad de Santiago de Chile, Chile

Abstract

Given the evidence that concentrations of the fine fraction of particulate matter in the atmosphere are more strongly associated with health problems than concentrations of the coarse fraction, it seems very likely that in the near future air quality in large cities will be defined more on the basis of PM_{2.5} concentrations rather than PM₁₀. We have developed a forecasting model for the maximum of the 24 hour moving average of PM_{2.5}. The model is based on the neural network technique. Training is performed with historical data from four monitoring stations located in Santiago Chile and with relevant meteorological information in the city. Cross correlations between several candidate variables for input allowed the selection of a reduced number of them which are used in the final model. Results show that the developed model can be used as an operational tool for air quality management.

Keywords: air pollution, forecasting, PM_{2.5}, neural networks.

1 Introduction

At present, air quality in Chile is defined in terms of the 24 hour moving average (24MA) of concentrations of particulate matter with diameter smaller than 10 microns (PM₁₀). Although these particles are small enough as to penetrate the respiratory tract of humans, associations of long-term ambient PM₁₀ concentrations with mortality have indicated that the most significant harm can be attributed to the effect of the fine fraction of PM₁₀ (PM_{2.5}, particles with diameter less than 2.5 microns) [1]. Environmental policies in the recent years in Santiago, Chile, as for example more strict control of emissions in industry, and replacement of old buses by new cleaner vehicles in public transportation have produced significant improvements in air quality in the city. However, this city



can still be considered among the most polluted urban areas in the world, especially during fall and winter, when meteorological conditions are not favorable for pollutant dispersion.

Very few PM_{2.5} forecasting models have been reported in the literature. In the year 2000, a simple model for one hour average concentrations forecasting of PM_{2.5} in Santiago, Chile, based on neural network techniques was proposed [2]. More recently, a comparison of neural models, a linear model and a persistence model has been performed in order to implement a reliable tool to forecast daily averages of PM_{2.5} in El Paso (USA) and Ciudad Juarez (Mexico) [3]. In this paper, we report the results obtained with a neural network model aimed to forecast the maximum of the 24MA of PM_{2.5} in which correlations between particle matter concentrations and meteorological variables guide the selection of input variables.

2 The data

One hour average PM_{2.5} concentrations were obtained from the reports of four stations of the official Santiago monitoring network. We will refer to them as stations L, M, N, O. We have concentrated on data between April and August (which is the period with relatively high concentrations during the year) for years between 2001 and 2004. Meteorological information associated with air quality was obtained from the National Center of the Environment (CENMA). Our goal is the forecasting of the maximum of the daily 24 hour moving average of PM_{2.5} by using pollutant concentrations until 19:00 hr of a given day in order to forecast the value on the following day. Motivated by results on a PM₁₀ forecasting model previously developed [4], we have looked for a suitable neural network model for PM_{2.5}. An important step towards this objective is to determine the most appropriate input variables. We have calculated correlations between the quantity to forecast and the following variables: one hour average concentrations at different times on every station up to 19:00, present day and forecasted values of daily thermal amplitudes (difference between maximum and minimum temperature) and a “meteorological potential of atmospheric pollution” (PMCA in Spanish). This is a discrete variable which is a measure of ventilation in the Santiago area. It is calculated and forecasted daily by meteorologists of CENMA.

Guided by the information contained in these correlation calculations as can be seen in the following sample of figures and tables, we have implemented a forecasting neural network scheme.

Forecasting concentrations on station N is a priority because it is the closest to downtown, where a large population will be exposed during the day. From figure 1 we observe that the highest correlation up to 19:00 hr is precisely the concentration at this time. But correlation with station M which is several kilometres to the east, correlation with concentration at 18:00 is the highest, which should be taken in account when selecting the input variables. From table 1 we observe that correlations with the analyzed meteorological variables are all significant in the case of station N.



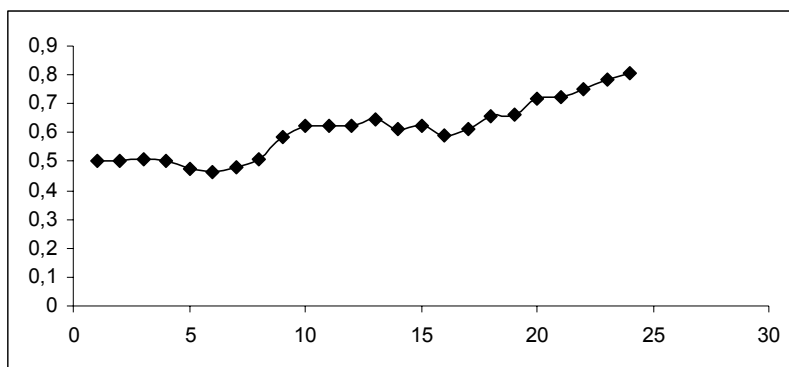


Figure 1: Correlations between one hour averages of PM_{2.5} with maximum of 24MA of PM_{2.5} on the next day for station N.

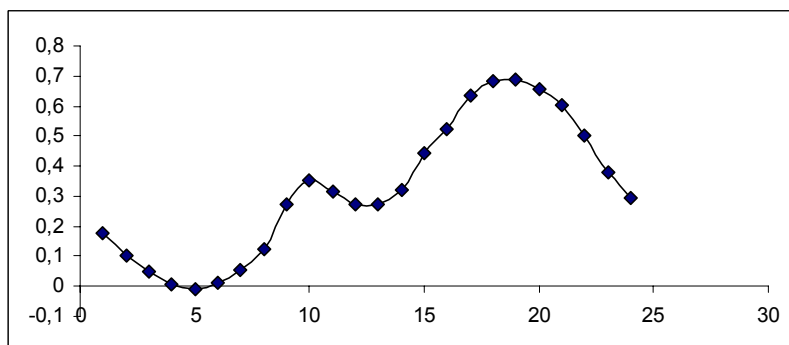


Figure 2: Correlations between one hour averages of PM_{2.5} in station M with maximum of 24MA of PM_{2.5} on the next day in station N.

Table 1: Correlations between maximum of tomorrow's 24MA of PM_{2.5} and tomorrow's PMCA, and today's and tomorrow's thermal amplitude for the four monitoring stations.

	PMCA	ΔT today	ΔT tomorrow
STATION L	0.38	0.26	0.14
STATION M	0.27	0.15	0.03
STATION N	0.52	0.43	0.25
STATION O	0.49	0.48	0.26

3 Results with the neural model

The neural network algorithm consists of a set of nodes or neurons distributed in layers. Information flows sequentially from an input layer to an output layer, passing eventually through one or more hidden layers. Connections weights between nodes on a given layer and nodes on the subsequent layer are fixed after a training stage based on a sample of the data of interest by means of an optimization algorithm called backpropagation [5]. The output of a given node is obtained upon the application of a transfer function to a linear combination of the outputs of the previous layer, the coefficients being the connection weights. After training, the net is prepared to generate the output when input data of new cases is presented. The performance of the model may be measured by calculating a percent error function defined as:

$$PE = \frac{\langle |y_{tp} - y_{ta}| \rangle}{\langle y_{ta} \rangle} \times 100 \quad (1)$$

where y_{tp} is the forecasted value, y_{ta} is the actual value, and $\langle \rangle$ means average over the sample cases.

The best results for station N are obtained with a net that has 7 input variables, 4 hidden nodes and one output node. The seven input variables are:

- one hour average of PM2.5 in station M at 18:00
- one hour average of PM2.5 in station L at 19:00
- one hour average of PM2.5 in station O at 19:00
- one hour average of PM2.5 in station N at 19:00
- today's thermal amplitude
- tomorrow's thermal amplitude (forecasted)
- tomorrow's PMCA (forecasted)

After training with 2001, 2002 and 2003 data, we have performed a test with 2004 data, obtaining a percent error of 17% for the forecasting of tomorrow's maximum of 24 h average of PM2.5. This is significantly better than the 22% error obtained with a persistence model.

We must mention that we have implemented similar forecasting models for the other three stations, and preliminary results show that forecasting neural models give results of an accuracy of the same order of magnitude.

Acknowledgements

We would like to thank the support of the Fondo Nacional de Ciencia y Tecnología (FONDECYT) through project 1050015, to the Comision Nacional del Medio Ambiente (CONAMA), Chile, for providing the data pollutant used in this work, and to CENMA for providing the relevant meteorological information.



References

- [1] McDonnell, W. F., Nishino-Ishikawa, N., Petersen, F. F., Chen, L. H., Abbey, D. E., Relationships of mortality with the fine and coarse fractions of long term ambient PM₁₀ concentrations in nonsmokers. *J. Expo Anal Environ Epidemiol* 10, 427- 436 (2000).
- [2] Perez, P., Trier, A., Reyes, J., Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile. *Atmospheric Environment* 34, 1189- 1196 (2000).
- [3] Ordieres, J. B., Vergara, E. P., Capuz, R. S., Salazar, R. E. Neural network prediction model for fine particulate matter (PM_{2.5}) on the US-Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua). *Environmental Modelling & Software* 20, 547- 559. (2005).
- [4] Perez, P., Reyes, J., PM₁₀ forecasting in Santiago, Chile, *Air Pollution XIII*, Córdoba, Spain, WIT Press 2005, 33- 37.
- [5] Rumelhart, D. E., Hinton, G. E., Williams, R. J., Learning Internal Representations by Error Propagation. *Parallel Distributed Processing*, The MIT Press, Cambridge, London, pp 318-364 (1986).

