PERGAMON

Atmospheric Environment 34 (2000) 1189–1196

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

# Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile

Patricio Pérez*, Alex Trier, Jorge Reyes

*Departamento de Física, Universidad de Santiago de Chile, Casilla 307, Correo 2, Santiago, Chile*

Received 26 October 1998; received in revised form 21 June 1999; accepted 25 June 1999

## Abstract

Hourly average concentrations of $PM_{2.5}$ have been measured at a fixed point in the downtown area of Santiago, Chile. We have focused our attention on data for the months that register higher values, from May to September, on years 1994 and 1995. We show that it is possible to predict concentrations at any hour of the day, by fitting a function of the 24 hourly average concentrations measured on the previous day. We have compared the predictions produced by three different methods: multilayer neural networks, linear regression and persistence. Overall, the neural network gives the best results. Prediction errors go from 30% for early hours to 60% for late hours. In order to improve predictions, the effect of noise reduction, rearrangement of the data and explicit consideration of meteorological variables are discussed. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Air pollution prediction; Neural networks; $PM_{2.5}$ pollution; Time series analysis

## 1. Introduction

Harmful health effects of particle matter suspended in the atmosphere are well established. Mass concentration of particles with an aerodynamic diameter less than 10 μm ($PM_{10}$), is usually used as a standard measure of air pollution. However, since recent studies have shown that finer particles have the greatest impact on health (Spurny, 1998), and that most of the secondary particles (produced by chemical reactions in the atmosphere) have an aerodynamic diameter less than 2.5 μm ($PM_{2.5}$) (Harrison et al., 1997), control of pollution due to this type of particle has become urgent. Increase in the level of $PM_{2.5}$ has been associated with increases in mortality (Borja-Aburto et al., 1998) and cardiorespiratory hospitalizations (Burnett et al., 1999). Vehicle exhaust emissions and resuspended surface dusts are among the main sources of $PM_{2.5}$ in urban areas. Carbon, ammonium sulfate and nitrate are found to be its major components.

In places where geographical and meteorological conditions allow for poor air circulation, and where there is a large population living in a not always well designed city, episodes of critically high atmospheric pollution enforce extreme actions as closing of schools and industries and restriction of motor vehicles circulation. If it were possible to predict these episodes one or two days in advance, more efficient actions could be taken in order to protect the citizens.

Our study is based on data obtained at an official monitor station in the city of Santiago, Chile. The city is located in a confined basin surrounded by hills and mountains. To the east we find the Andes and to the west a parallel coastal range. To the north and to the south we find transversal mountain chains. The slopes of the Andean and Coastal ranges are deforested due to scarce annual rainfall (of the order of 300 mm). During winter, temperatures range between 0 and 18°C, and winds have a daily average velocity of 1.2 m $s^{-1}$. In summer temperatures go from 10 and 30°C, and average wind velocity is of the order of 2.5 m $s^{-1}$. Population in the metropolitan area is about six million, with more than half of the national industries and near one million vehicles. Permanent anticyclonic thermal inversion layers and high

* Corresponding author.

1352-2310/00/$ - see front matter 
PII: S1352-2310(99)00316-7

amounts of solar radiation are observed (Rutllant and Garreaud, 1995). Geographical, meteorological and urban development conditions make Santiago one of the most heavily polluted cities in the world (Jorquera et al., 1998).

Classical statistical methods and neural network methods have been used by several authors for short term prediction of gas and particulate matter pollution. Prediction models have been proposed where average pollutant concentrations are given between one half and 24 h in advance. These models are phenomenological, and use the pollution data from the past plus some meteorological information. In one case predictions are one half hour in advance (Boznar et al., 1993), in other, up to 4 (Ruiz-Suarez et al., 1995) and in a third case, daily averages of particle concentrations are predicted one day in advance with an accuracy of the order of 30% (Hernandez et al., 1992). Recently, Gardner and Dorling (1998) have written a review on applications of multilayer neural networks in the atmospheric sciences where they show that it is a useful tool for prediction, function approximation and classification.

We report here predictions of hourly average concentrations of $PM_{2.5}$ from one to 24 h into the future, based on data obtained at downtown Santiago, Chile. We show results obtained with three different methods, all of which use past values of $PM_{2.5}$ as input. The simplest method is persistence, which assigns hourly values on the next day equal to the values at the present day. Then we used a linear regression, which is a special case of a multilayer neural network, a linear perceptron, namely a two layer network with linear transfer function between input and output. The third method is a three layer feed forward neural network, where the transfer function between the input layer and the hidden layer and between the hidden layer and the output is a non linear, sigmoid type. This three layer network, once the right parameters are found, is a non linear algorithm. The hidden layer provides a way to improve our possibilities to find a correct mapping between input and output. We treated independently the 1994 and 1995 time series corresponding to hourly average concentrations of $PM_{2.5}$ for 5 months starting on May 1st. We have separated the data in a training set and a test set. Prediction errors are calculated from the difference between actual outputs and the outputs generated using the different methods on the test cases. Least errors are produced with the three layer neural network. We have chosen to work with the scalar time series on $PM_{2.5}$ concentrations keeping in mind the idea that if we use a large enough window of data as input, the effect of other pollutants or meteorological data should be implicit in its structure. For the linear perceptron and the three layer network we have chosen an input window of size 24. Three strategies in order to improve the predictions are analyzed: noise reduction previous to the implementation of the predictive methods, allowing superposition in the training sets, and explicit consideration of the values of wind velocity and relative humidity. Given the high correlation between $PM_{2.5}$ and other pollutants, in particular $PM_{10}$, the results of our study may be useful in order to take actions to control atmospheric pollution.

## 2. The data

Hourly averaged $PM_{2.5}$ mass concentrations for 1994 and 1995 were obtained continuously by means of a mechanically oscillated mass balance type instrument (TEOM 1400a supplied by Rupprecht and Pataschnick Co., Inc., Albany, NY, USA, having a 2.5 μm inlet cutoff). The temperature of the sample stream was maintained at 50°C. The $PM_{2.5}$ mass fraction of atmospheric particulate matter comprises particles having an aerodynamic diameter of 2.5 μm and less. The instrument was run at station A of the Santiago public air quality monitoring system, designated as MACAM; this station is located about 100 m NW of Government House (Palacio de La Moneda) in downtown Santiago, Chile.

We have chosen to work with data from 05/01 to 09/30 for both 1994 and 1995. It is during these months that, due to unfavorable conditions for the diffusion of particles, atmospheric pollution in Santiago reaches the highest levels. Then, our time series have 3672 points. For the 1994 data, the average concentration is 71 $\mu g\,m^{-3}$, the maximum is 331 $\mu g\,m^{-3}$ and the standard deviation is 45 $\mu g\,m^{-3}$. For the 1995 data, the average concentration is 61 $\mu g\,m^{-3}$, the maximum is 250 $\mu g\,m^{-3}$ and the standard deviation is 41 $\mu g\,m^{-3}$. Fig. 1 shows the hourly concentrations of $PM_{2.5}$ during June, 1994.

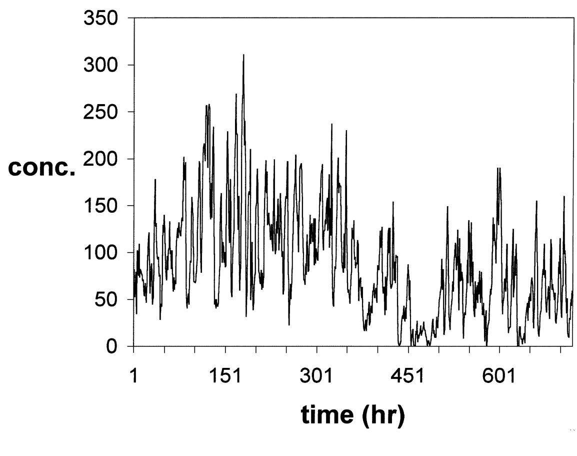


Fig. 1. Hourly concentrations of $PM_{2.5}$ data for June, 1994. Units are $\mu g\,m^{-3}$.

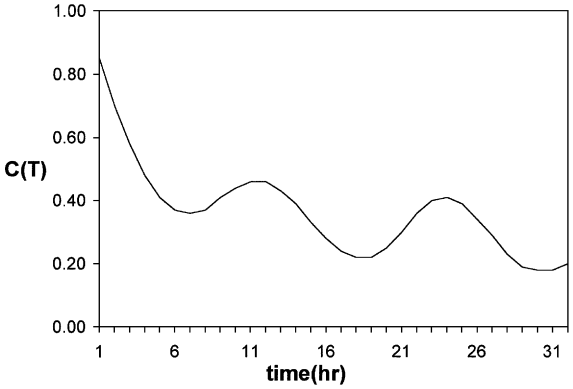


Fig. 2. Autocorrelation function for $PM_{2.5}$ data from 05/01 to 09/30, 1994.

We have calculated the autocorrelation function of the time series. It is obtained from:

$$C(T) = \sum_{n=0}^{N} s(n)s(n+T), \tag{1}$$

where $s(n)$ is the $PM_{2.5}$ concentration at time $n$, $N = 3672$ and $T = 0, 1, 2, \ldots$.

In Fig. 2 we can see this function evaluated for the five months considered in 1994, normalized to $C(0)$. There are peaks at 12, 24, 36, … h, which indicates that a given level of pollution has a tendency to be repeated every 12 h, at least qualitatively. We found a similar situation for the 1995 data. The long time coherence of the autocorrelation function, as observed here, is an indication of deterministic behavior (Kulkarni et al., 1997).

We have found that it is possible to relate the concentration of $PM_{2.5}$ at a given time of the day to the sequence of 24 points corresponding to the average hourly concentrations on the previous day. This was verified after performing several tests on the data using the methods described in the next section. When we used an input window of size less than 24, predictions were not good. For a size greater than 24, we were left with too many free parameters in order to adjust a model. A more systematical method to find the optimal size of the input window was not attempted, because all known methods are statistical and require huge amounts of data (Abarbanel et al., 1993). Since vehicle exhaust emissions contribute significantly to $PM_{2.5}$ concentrations, measured levels are expected to be lower on weekends and holidays as compared to normal working days. Fig. 3 shows average daily variations per day of the week for the analyzed period on 1994. Given that the difference between Sunday and Monday is not much greater than the difference between Monday and Tuesday or between Tuesday and Saturday and considering that there is no qualitative difference between them, we decided to treat all days in the same foot. Otherwise we would not have enough training cases in order to adjust the parameters of some of the predictive models studied. The two maxima of $PM_{2.5}$ concentrations at 10:00 A.M. and between 10:00 and 11:00 P.M. may be related to vehicle traffic peaks, because congestion occurs at rush hours (8:30 A.M. and 8:00 P.M.) in the area where the monitoring station is located. The delay of $PM_{2.5}$ maxima as compared to vehicle traffic peaks, is due in part to the fact that the monitoring station is located at a distance of 150 m from a stop light where the highest congestion is observed. The observed delay is within what is expected for turbulent transport of pollutants in air (Boeker and Grondelle, 1995). The additional delay observed at evenings may be due to cooling of the surface after sunset (around 6:00 P.M.) which decreases atmospheric turbulence, and makes transport slower. This evening peak is present in all stations with particle concentration measurements throughout metropolitan Santiago (5 stations in 1994), but occurs earlier when the station is located closer to the area of traffic congestion. However, it is observed only during the fall–winter season. The fact that in spring–summer time only the morning peak observed is probably due to more favorable conditions for particle dispersion at evenings because of increasing solar radiation (Prendez et al., 1995) and the stronger winds present in this season.

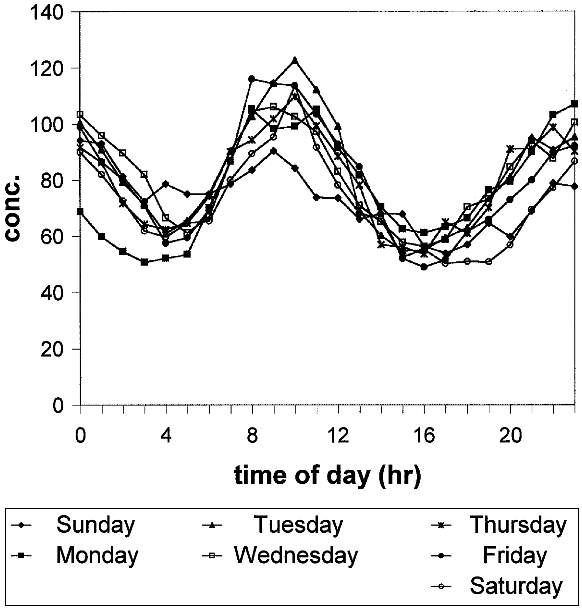


Fig. 3. Average daily variation of $PM_{2.5}$ concentrations ($\mu g\ m^{-3}$) from 05/01 to 09/30, 1994.

## 3. Prediction

For each year we have built 24 matrices from the data set. All of them have 25 columns, the $i$th row of the $k$th matrix containing the average hourly concentrations of $PM_{2.5}$ for the $i$th day, the 25th element being the datum for the $k$th hour on day $i+1$. From each of these matrices we have extracted every 4th row, creating in this manner one matrix with 114 rows and other with 38 rows. The rows of $114 \times 25$ matrices will provide the training set and the rows of the $38 \times 25$ matrices will give rise to the test set which are needed in order to implement the methods described below.

The three methods we have implemented in order to predict $PM_{2.5}$ concentration at time $t$ on a given day may be described using the following scheme:

$$y_t = f_t(x_1, x_2, ..., x_{24}), \tag{2}$$

where $x_1, \ldots, x_{24}$ represent the pollution data on a given day and $y_t$ is the predicted value at time $t$ on the following day. For every $t$ ($t = 1, \ldots, 24$) there will be a different function $f_t$. The form of this function is obtained after adjusting a set of parameters defined within the respective method using the training set of data. Within the training matrix, every row is a sample case. The quality of the prediction is obtained from the performance of the test set of data. Percent errors on prediction (PE) are calculated according to

$$\mathrm{PE} = \frac{\langle |y_{\mathrm{tp}} - y_{\mathrm{ta}}| \rangle}{\langle y_{\mathrm{ta}} \rangle}, \tag{3}$$

where $y_{\mathrm{tp}}$ is the predicted value, $y_{\mathrm{ta}}$ is the actual value, and $\langle\ \rangle$ means average over the 38 test cases.

### 3.1. Persistence

This consists of a very simple prediction: tomorrow at time $t$ $PM_{2.5}$ mass concentration will be the same as today at time $t$. In this case Eq. (1) takes the form:

$$y_t = x_t, \tag{4}$$

which we would not expect to be very accurate. Here there are no parameters to adjust, and prediction errors are calculated from the performance of the test set (see Fig. 4 for 1994 results and Fig. 5 for 1995).

### 3.2. Linear perceptron

This method assumes that prediction of $PM_{2.5}$ at a given time on the next day is obtained from a linear combination of the values on the present day:

$$y_t = w_{t1}x_1 + w_{t2}x_2 + \ldots + w_{t24}x_{24} + w_{t0}. \tag{5}$$

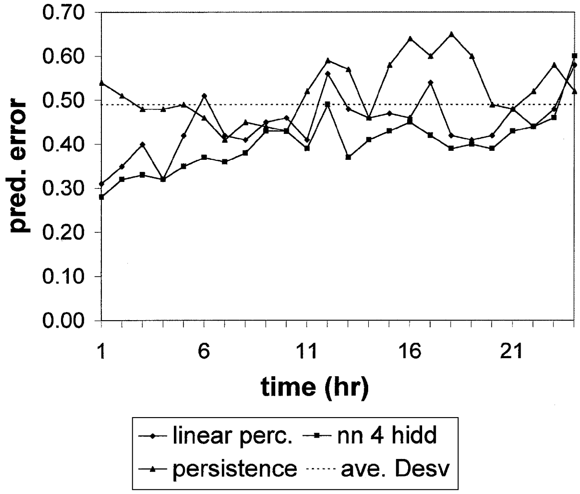


Fig. 4. Prediction errors for the 1994 $PM_{2.5}$ series. Predictions are from 1 to 24 h in advance using a three layer neural network (nn 4 hidd), a linear perceptron (linear perc.), and persistence. Points are averages over 38 cases. Horizontal line is the average deviation from the mean for the whole series (ave. Desv).

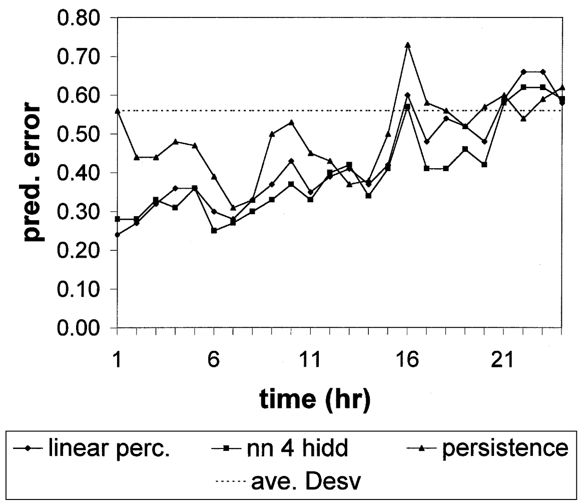


Fig. 5. Prediction errors for the 1995 $PM_{2.5}$ series. Predictions are from 1 to 24 h in advance using a three layer neural network (nn 4 hidd), a linear perceptron (linear perc.), and persistence. Points are averages over 38 cases. Horizontal line is the average deviation from the mean for the whole series (ave. Desv).

For every $t$, the coefficients $w_{t_i}$, which allow global adjustment of the training cases, may be calculated using any linear regression technique. Since Eq. (5) represents a special case of a two layer neural network, we have applied an iterative algorithm (the Delta rule) widely used in this area (Rumelhart et al., 1986). Results represent a considerable improvement with respect to persistence (see Figs. 4 and 5). In average, the fractional prediction error using a linear perceptron goes down by 0.08 in 1994 and 0.07 in 1995 as compared with persistence.

### *3.3. Three layer neural network*

Recent studies have shown the advantage of neural network methods over traditional statistical methods for time series prediction (Weigend and Gershenfeld, 1994).

If in Eq. (2) the function $f_t$ were a non linear function of a linear combination of the inputs, we would have a non linear perceptron. Additional room for a good fitting of the data may be achieved by introducing a set of hidden nodes $z_{tk}$, $(k = 1, \ldots, n)$, in such a way that:

$$z_{tk} = f(w_{tk1}x_1 + \ldots + w_{tk24}x_{24} + w_{tk0}) \tag{6}$$

and

$$y_t = f(v_{t1}z_{t1} + \ldots + v_{tn}z_{tn} + v_{t0}). \tag{7}$$

The function $f$ we have used is

$$f(X) = \frac{1}{1 + e^{-X}}. \tag{8}$$

In order to find the $w$'s and $v$'s we have used a generalized Delta rule (Rumelhart et al., 1986), which is a back propagation of errors, starting from the difference between calculated and actual outputs. The number of hidden units $n_h$ is determined by trial and error. Since the number of adjustable parameters in a three layer feed forward neural network with $n_i$ input units $n_o$ output units and $n_h$ hidden units is $n_o + n_h(n_i + n_o + 1)$, for $n_i = 24$, $n_o = 1$ and with 114 training cases, we cannot use an $n_h$ greater than 4. The best results were obtained for $n_h = 4$.

In Figs. 4 and 5 we compare the predictions produced by the three methods described above, for years 1994 and 1995, respectively. In both cases, least average prediction errors are obtained with the three layer network. Standard deviation at each point for all methods is rather large, reaching in some cases to 100% of the present value. However, the systematical relative shift of the curves gives us confidence that from better to worse the methods are: three layer network, linear perceptron and persistence. Using persistence as a reference, we can say that significant prediction is possible only until approximately 20 h into the future. Considering that the three layer network represents a non linear algorithm, we can say that non linear effects are more pronounced for the 1994 series, which may be due to a higher complexity for the data on this year. In average, the fractional prediction error with the three layer network decreases by 0.04 in 1994 and 0.03 in 1995 with respect to the linear perceptron. We have also included in these plots an horizontal line that represents the average deviation from the mean of the respective complete series of data, which is calculated using Eq. (3) for the 3672 data points.

## 4. Discussion and possible improvements

The best accuracy of the predictions produced with the three layer network seems still poor in order for the results to have practical application in environmental pollution policies. However, we have performed additional studies aimed to improve these predictions.

### *4.1. Noise reduction*

We have assumed that the data on $PM_{2.5}$ are contaminated with white noise. There exist several methods to clean a signal from this noise, and for this purpose we have chosen to use a neural network technique that has been used for image compression (Cotrell et al., 1989). The method consists of the following: the series is divided into intervals containing $n_I$ points each. Training sets are constructed which have $n_I$ inputs, $n_I$ outputs and $n_H$ hidden units. Output values are the same as inputs and $n_H < n_I$. During training the network learns to reproduce at the output the input values after extracting the relevant features which are encoded in the hidden layer. The difference between the output produced and actual output is attributed to noise. Then the output produced by the network is used to create a new series which is assumed to be cleaner than the original. $n_H$ should be the smallest value that does not destroy the relevant information in the series. This may be monitored checking the form of the autocorrelation function. We have used $n_H = 6$ after observing that in this case there is only a constant vertical shift of this curve, preserving otherwise its qualitative form. As we can see from Fig. 5, for the new series, for 12 h into the future, for 1994 data, and using again a three layer network for prediction, errors go down significantly. In this case errors are calculated with respect to the noise reduced series, and it will remain open the question if we are dealing with really representative data.

### *4.2. Using a larger data set*

A limitation in order to generate a good prediction with low dispersion using the methods described above is the small size of the data set. Training and test sets are built using non overlapping groups of data. We have investigated also the effect of using overlapping groups of data. We have again used an input window of size 24, but in this case we displaced it over the series, where on every displacement the earliest value is left out and the next value is appended. In this way, consecutive cases differ slightly, and we are able to generate 3171 training cases

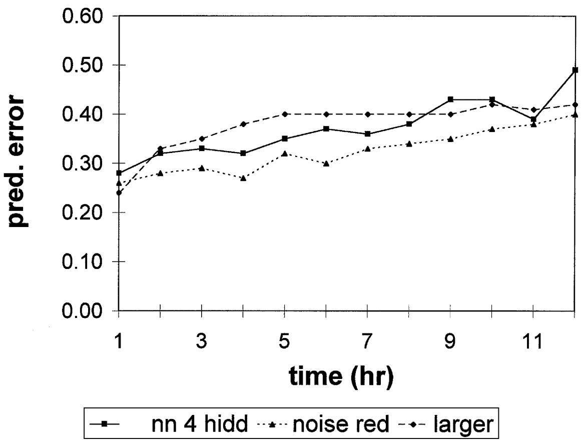


Fig. 6. Comparison of prediction errors for 1994 $PM_{2.5}$ using a three layer neural network on the original data (nn 4 hidd), a three layer network on noise reduced data (noise-red.), and a non linear perceptron on enlarged training and test sets (larger).

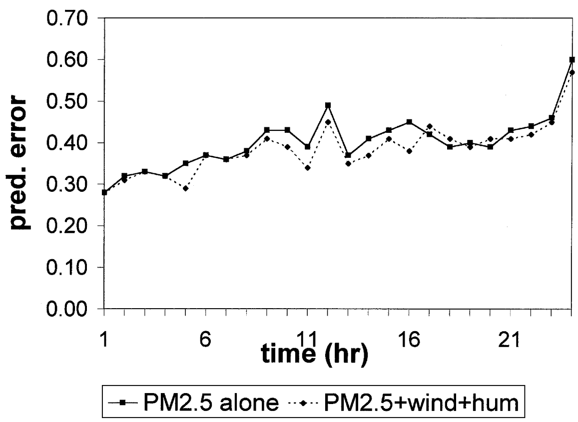


Fig. 7. Comparison of prediction errors for 1994 $PM_{2.5}$ using a three layer neural network with 24 inputs ($PM_{2.5}$ alone) and a three layer network with 26 inputs ($PM_{2.5}$ + wind + hum). Wind velocity and relative humidity used as input are the values at the time of the intended prediction.

and 453 test cases. The best fitting is obtained with a non linear perceptron, i.e. a two layer network with a sigmoid (Eq. (7)) as the transfer function. Prediction errors shown in Fig. 6, are not smaller than those with the smaller data set using a three layer network, except for the first hour. In the new situation there is a great redundancy of information, but what is positive is that now we obtain a smoother curve which is a consequence of lower dispersion around each data point, which in some way validates our previous results.

### 4.3. *Effect of considering meteorological variables*

It is a well established fact that atmospheric pollution depends strongly on weather conditions. It may be that the effect of different meteorological variables is implicit in the structure of the time series of a given pollutant, like $PM_{2.5}$. However, due to the complexity of the correlation, and also because of the presence of noise, an explicit consideration of the effect of variables like temperature, wind velocity and relative humidity may allow a better prediction of particle concentrations. In order to estimate the effect of these meteorological variables, we have calculated the cross correlation coefficient between the $PM_{2.5}$ series and the series of the three meteorological variables mentioned above. The cross correlation coefficient $\rho_{ab}$ between series $s_a$ and $s_b$ is defined as

$$\rho_{ab} = \frac{\langle s_a s_b \rangle - \langle s_a \rangle \langle s_b \rangle}{\sqrt{(\langle s_a^2 \rangle - \langle s_a \rangle^2)(\langle s_b^2 \rangle - \langle s_b \rangle^2)}}, \tag{9}$$

where $\langle s_a \rangle = 1/N \sum_{n=1}^{N} s_a(n)$, $\langle s_a^2 \rangle = 1/N \sum_{n=0}^{N} s_a^2(n)$, and $\langle s_a s_b \rangle = 1/N \sum_{n=0}^{N} s_a(n) s_b(n)$. In our case, $N = 3672$. We found that for 1994 series, the cross correlation coefficient between $PM_{2.5}$ and temperature is zero, between $PM_{2.5}$ and wind velocity there is an anti correlation of $-0.1$ and between $PM_{2.5}$ and relative humidity there is also an anti correlation of $-0.1$. The anti correlation between wind velocity and $PM_{2.5}$ may be explained by the relation between wind velocity and ventilation. Strong winds will imply unfavorable conditions for particle accumulation in a given region. The anti correlation between relative humidity and $PM_{2.5}$ may be associated to the occasional penetration of coastal air (the coast is at a distance of the order of 100 km towards the West), a relatively uncontaminated air, into the Santiago region. Correlation between $PM_{2.5}$ and wind direction was found to be negligible. Given these results, we attempted a neural network scheme to predict $PM_{2.5}$ concentrations from 1 to 24 h in advance in which to the 24 $PM_{2.5}$ inputs used earlier we added two new inputs, wind velocity and relative humidity at the time of the day we intend to predict $PM_{2.5}$. This approach assumes that we have an independent method to predict these meteorological variables. Due to the increase in the amount of weights to determine, in this situation we put three units (instead of four) in the hidden layer. The new predictions are slightly better than those based exclusively on $PM_{2.5}$ information (see Fig. 7), results that seem consistent with the degree of correlation between the variables considered. Independent predictions of wind velocity and relative humidity will be desirable, as those produced by meteorologists. However, from the structure of the time series of these variables, we can expect good predictions with simple methods, at least in the case of relative humidity. For humidity, the average value for 3672 points is 78.1, standard deviation is 15.2 and the fractional average

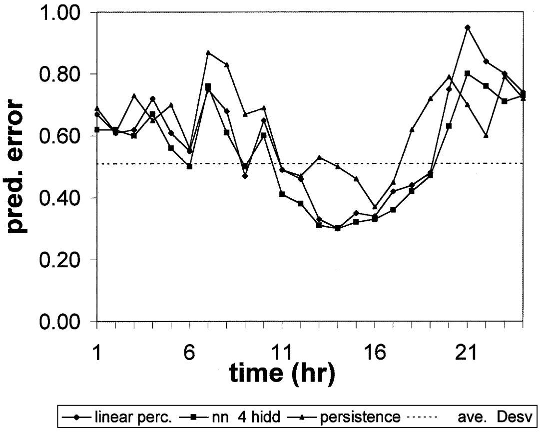


Fig. 8. Prediction errors for the 1994 wind velocity series. Predictions are from 1 to 24 h in advance using a three layer neural network (nn 4 hidd), a linear perceptron (linear perc.), and persistence. Points are averages over 38 cases. Horizontal line is the average deviation from the mean for the whole series (ave. Desv).

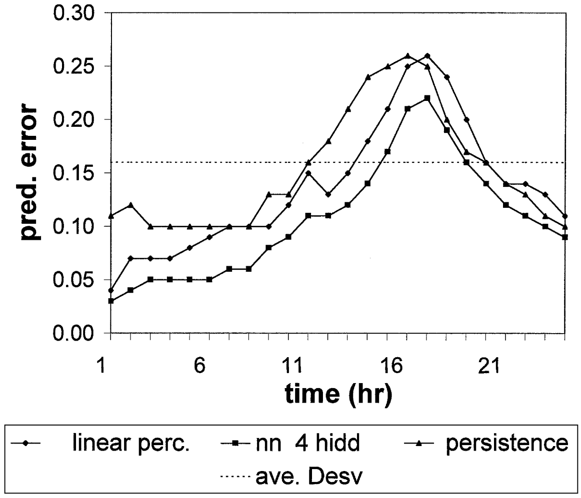


Fig. 9. Prediction errors for the 1994 relative humidity series. Predictions are from 1 to 24 h in advance using a three layer neural network (nn 4 hidd), a linear perceptron (linear perc.), and persistence. Points are averages over 38 cases. Horizontal line is the average deviation from the mean for the whole series (ave. Desv).

deviation from the mean is 0.16. For wind velocity, taking into account the same amount of points, the average value is $1.16\,m\,s^{-1}$, standard deviation is $0.75\,m\,s^{-1}$ and the fracional average deviation from the mean is 0.51.

We have analyzed the possibility to predict both wind velocity and relative humidity using a scheme similar to that used with $PM_{2.5}$ before introducing the effect of meteorological variables, this means using the hourly averages of a given day as an input to adjust a prediction of values at the next day, from one to 24 h in advance. In Fig. 8 we show the comparison between persistence, linear perceptron and a 24–4–1 neural network for wind velocity. We observe that a reasonable prediction is only possible between 11 and 19 h. The reason for this may be that it is during afternoon hours that the wind velocity is higher, so the relative errors after prediction are smaller. We can also verify that the three layer neural network is slightly better than the linear perceptron, which implies that non linear effects are not very important. Fig. 9 displays the results for relative humidity. Here we find that average prediction errors are rather small for the three methods, except between 13 and 19 h, which may be explained noticing that in most cases, during afternoon hours relative humidity is smaller and then relative errors are higher. For this variable, predictions using a three layer neural network are considerably better than those with a linear perceptron, signaling the importance of non linear effects.

As a summary, we can say that a non linear method seems convenient for $PM_{2.5}$ prediction. Noise reduction prior to modeling appears to be necessary. Relevant information is contained in the structure of the time series of the same variable, and some improvement on prediction is possible by taking into account related meteorological variables explicitly. Still, different schemes to treat the past values as input should be investigated.

## Acknowledgements

We would like to thank the support from the Research Department, Universidad de Santiago de Chile (DICYT), through project 049631PJ, and Fondo Nacional de Ciencia y Tecnologia (FONDECYT) through project 1970418.

## References

Abarbanel, H.D.I., Brown, R., Sidorowich, J.J., Tsimring, L.S., 1993. The analysis of observed chaotic data in physical systems. Rev. Mod. Phys. 65, 1331–1392.

Boeker, E., van Grondelle, R., 1995. Environmental Physics. Wiley, New York.

Borja-Aburto, V.H., Castillejos, M., Gold, D.R., Bierzwinski, S., Loomis, D., 1998. Mortality and ambient fine particles in southwest Mexico City, 1993–1995. Environ Health Perspect 106 (12), 849–855.

Boznar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for short term predictions of ambient $SO_2$ concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment 27B, 221–230.

Burnett, R.T., Smith-Doiron, M., Stieb, D., Cakmak, S., Brook, J.R., 1999. Effects of particulate and gaseous air pollution on cardiorespiratory hospitalizations. Archives of Environment Health 54 (2), 130–139.

Cotrell, G.W., Munro, P., Zipser, D., 1989. Image compression by back propagation: an example of extensional programming, models of cognition. In: Sharley, N.E. (Ed.), A Review of Cognitive Science. Ablex, Norwood, NJ, pp. 209–240.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. Atmospheric Environment 32, 2627–2636.

Harrison, R.M., Deacon, A.R., Jones, M.R., 1997. Sources and processes affecting concentrations of $PM_{10}$ and $PM_{2.5}$ particulate matter in Birmingham (UK). Atmospheric Environment 31, 4103–4117.

Hernandez, E., Martin, F., Valero, F., 1992. Statistical forecast models for daily air particulate iron and lead concentrations for Madrid, Spain. Atmospheric Environment 26B, 107–116.

Jorquera, H., Perez, R., Cipriano, A., Espejo, A., Letelier, M.V., Acuña, G., 1998. Forecasting ozone daily maximum levels at Santiago, Chile. Atmospheric Environment 32, 3415–3424.

Kulkarni, D.R., Parikh, J.C., Pratap, R., 1997. Simulation of characteristics and artificial neural network modeling of electroencephalograph time series. Physics Review E 55, 4508–4511.

Prendez, M.M., Egido, C., Tomas, C., Seco, J., Calvo, A., Romero, H., 1995. Correlation between solar radiation and total suspended particulate matter in Santiago, Chile – Preliminary results. Atmospheric Environment 29, 1543–1551.

Ruiz-Suarez, J.C., Mayora-Ibarra, O.A., Torres-Jimenez, J., Ruiz Suarez, L.G., 1995. Short-term ozone forecasting by artificial neural networks. Advances in Engineering Software 23, 143–149.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning Internal Representations by Error Propagation. Parallel Distributed Processing. The MIT Press, Cambridge, London, pp. 318–364.

Rutllant, J., Garreaud, R., 1995. Meteorological air pollution potential for Santiago, Chile: towards an objective episode forecasting. Environmental Monitoring and Assessment 34, 223–244.

Spurny, K.R., 1998. On the physics, chemistry and toxicology of ultrafine anthropogenic, atmospheric aerosols (UAAA): new advances. Toxicological Letters 96, 253–261.

Weigend, A., Gershenfeld, N., 1994. Time Series Prediction, Santa Fe Series. Addison Wesley, New York.