

Forecasting hourly values of PM2.5 concentrations

P. Perez

Departamento de Fisica, Universidad de Santiago de Chile, Chile

Abstract

We present the results of an hourly PM2.5 concentrations forecasting model in Santiago, Chile. The study concentrates on the comparison between model and observed values at the monitoring station with the highest concentrations for the time period between April and August, which is the season when high concentration episodes are frequent. The forecasting model is a feed forward neural network. The input variables are past values of hourly PM10 and PM2.5 concentrations measured at the city station with the highest values during episodes, concentrations from neighbouring stations and some observed and forecasted meteorological variables. Secondary particle formation is considered implicitly by including nitrogen oxides concentrations at previous hours as input. Training is performed with 2010 and 2011 data and the model is tested with 2012 values. Information is collected until 7 PM of the present day and percent error forecasting up to 15 hours in advance is of the order of 30%. Accuracy of forecasting is significantly better than different forms of persistence and is more accurate than results with a linear model. The model may be considered as a useful tool for anticipating episodes.

Keywords: particulate matter forecasting, neural networks, linear models.

1 Introduction

PM2.5 concentrations in Santiago, Chile are far above the limits recommended by international health organizations. An important reason for this is the unfavorable conditions for atmospheric pollutant dispersion that prevail specially during fall and winter. Coordinates of the city are 33.46°S and 70.65°W. It has 7 million habitants and is located in a valley with altitude varying between 900 m on the east side to 400 m on the west side. To the east we find the Andes



Mountains with altitudes up to 5000 m and to the west we observe a coastal range with altitudes up to 2000 m. Between April and August, dispersion of atmospheric pollutants in the valley is very poor, which is related to strong surface thermal inversions and weak winds, associated to anticyclone conditions at a regional scale. During this period of the year, when the 24 hour moving average exceeds $80 \mu\text{g}/\text{m}^3$, the Ministry of the Environment applies several restrictions to emission sources. During 2012, 25 days showed this exceedance at some time of the day.

The main sources of PM_{2.5} in Santiago are wood stoves, industries and public and private transportation [1]. Wood burning is a major source of PM_{2.5} in Santiago region [2]. Although, wood stoves in the urban area are not the main heating system, falling behind gas and kerosene, they contribute heavily to Santiago's pollution.

In recent years, several atmospheric particulate matter forecasting models have been developed. Models may be classified in two groups: statistical and deterministic models. In case of deterministic models, computational resources needed to run next-day forecast deterministic models high, and detailed information about emissions is not always available. On the other hand, statistical models are much easier, quicker and economical, computationally speaking [3].

In cities where episodes of atmospheric pollution are frequent, it seems very convenient for environmental authorities to have an operational air quality forecasting model available in order to anticipate situations that may put in danger the population. Most PM_{2.5} statistical models forecast daily averages or 24 hour moving averages rather than hourly averages [4–6]. Reported PM_{2.5} forecasting models that estimate hourly concentrations tend to lack precision, mainly when the goal is forecast values several hours into the future [7, 8].

In this paper we report the results of the implementation of a neural network model intended to forecast hourly values of PM_{2.5} concentrations from one to twenty hours in advance in the vicinity of the most polluted area of the city of Santiago, Chile. Given that several high concentrations episodes are usually observed in the city, emphasis is put in the accurate forecast of these cases.

2 Data

Particulate matter PM₁₀ and PM_{2.5} concentrations may be obtained from stations belonging to the official monitoring network, (Macam network). This network consists of eleven stations distributed through the city. They report one hour averages of several pollutants and meteorological data of one half hour (see figure 1). According to Chilean regulations, when the 24 h average exceeds $80 \mu\text{g}/\text{m}^3$, restrictions to emissions like limitations to car circulation and wood burning apply. This paper reports the results of a model aimed to forecast hourly concentrations of PM_{2.5} in Cerro Navia station, which shows the highest values during the high pollution season in Santiago. Figure 2 shows the typical onset of an episode. It corresponds to the evolution of hourly PM_{2.5} concentrations on June 6, 2012. We observe that after low concentrations during the early



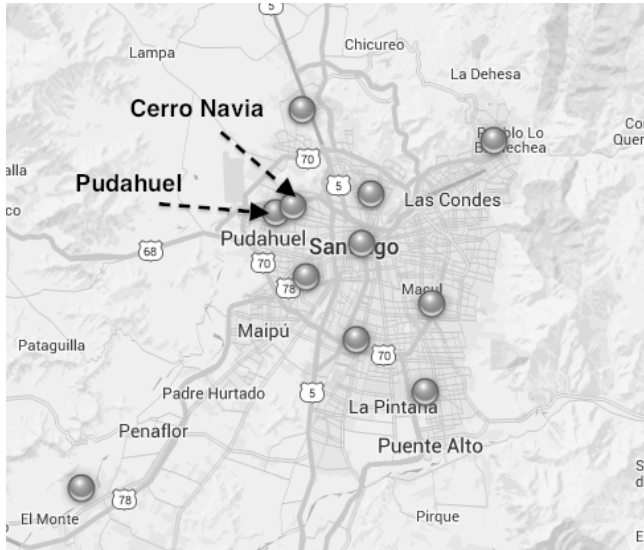


Figure 1: Location of air pollution monitoring stations in the city of Santiago, Chile. Hourly concentrations in Cerro Navia station are forecasted.

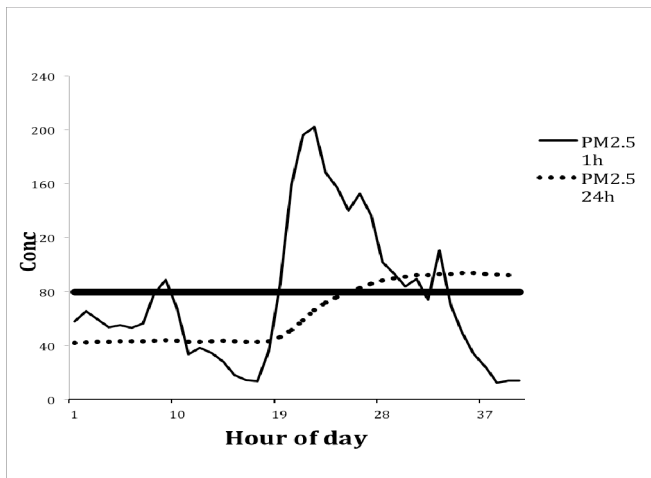


Figure 2: Evolution of PM_{2.5} concentrations in Cerro Navia station during an episode day in June, 2012.

afternoon, a sudden increase around 6 PM implies that the 24 h average exceeds $80 \mu\text{g}/\text{m}^3$ near midnight. If we are able to correctly forecast the evolution of hourly values, we can generate a report about expected 24 h average values for the following day. This report may provide an important information for

environmental authorities in order to decide to apply restrictions on the next day. An important meteorological variable that we have determined to be relevant for air quality forecasting in Santiago is the index called “meteorological potential of atmospheric pollution” [9]. This is a discrete variable with values between 1 and 5 and it is a measure of the ventilation conditions in the region. It is forecasted daily by local meteorologists as a result of a synoptic analysis. A value 1 corresponds to good conditions for pollutant dispersion (strong winds, intense rain). A maximum value 5 corresponds to poor ventilation associated with thermal inversions at low altitude and weak surface wind. The atmospheric conditions and PM2.5 episodes observed in Santiago are present in other regions of the world as for example in San Joaquin Valley, California, which has similarities with the central valley of Chile where Santiago is located [10].

3 Forecasting model

In order to forecast hourly concentrations of PM2.5 in Cerro Navia station several hours in advance we have used a feed forward three-layer neural network. Cerro Navia station (Figure 1) is the station that during the fall- winter period shows the highest concentrations among the network of Santiago’s air quality monitoring stations. After exploring correlations between several predictors and future PM2.5 values, we ended with a model with the following input variables:

- Hourly PM2.5 concentration at 6 PM in Cerro Navia station;
- Hourly PM2.5 concentration at 7 PM in Cerro Navia station;
- Hourly PM10 concentration at 6 PM in Cerro Navia station;
- Hourly PM10 concentration at 7 PM in Cerro Navia station;
- Hourly PM2.5 concentration at 6 PM in Pudahuel station;
- Hourly PM2.5 concentration at 7 PM in Pudahuel station;
- Hourly PM10 concentration at 6 PM in Pudahuel station;
- Hourly PM10 concentration at 7 PM in Pudahuel station;
- Wind speed at 7 PM in Cerro Navia station;
- Relative humidity at 7 PM in Cerro Navia station;
- Thermal amplitude of present day in Cerro Navia station;
- Forecasted thermal amplitude for the following day in Cerro Navia station;
- Forecasted ventilation factor for the following day.

We incorporate particulate matter information until 7 PM for operational reasons. Environmental authorities need to generate a public forecasting report for the following day between 8 PM and 9 PM of the present day.

PM10 and PM2.5 concentrations at 6 PM and 7 PM provide important information about incoming nightly concentrations (see fig. 2).

Training of the neural network model was performed with data from the period between April and August for the years 2010 and 2011. For every time delay, a different network was built with different weight connections. The



model was tested with new data from the same period on 2012. To evaluate the performance of the model, Pearson correlation and normalized percent error (NPE) were calculated, where:

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

where triangular bracket means average over the sample, y_{ta} is actual value and y_{tp} is the forecasted value. Normalized percent error is used in order to avoid the large errors in the case of low concentrations.

4 Results

In Figure 3 we display the Pearson correlation between modeled and observed data. The results of the neural network are compared with persistence (concentrations at any future time are the same than those at 7 PM of present day) and a linear regression with the same predictors. The figure shows the results from one to 21 hours in advance for all days in the period mentioned. Persistence in this case means that the delayed value is the same than the 7 PM value (this means that concentrations for hours later than 7 PM are all the same). We observe that correlations for the neural network model are reasonably good up to 15 hours in advance and slightly more accurate than the linear model and significantly better than persistence.

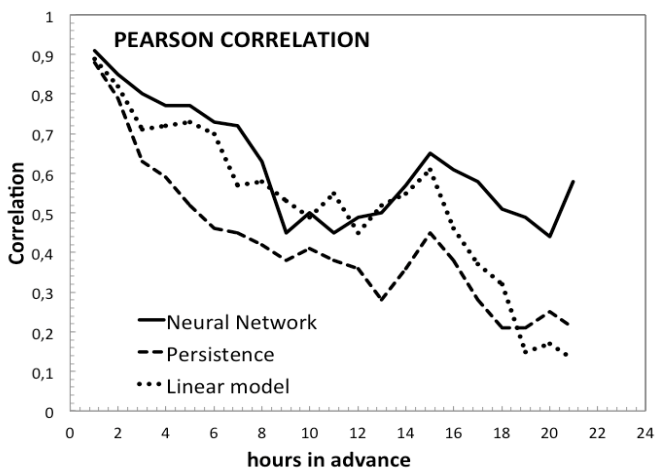


Figure 3: Pearson correlation of 2012 data. Comparison between neural network model, persistence and a linear model.

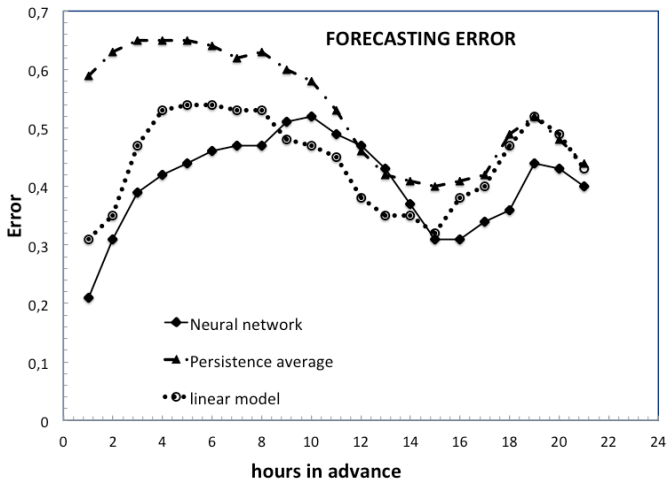


Figure 4: Normalized percent errors of three models for 2012 data in Cerro Navia station.

Figure 4 shows the NPEs obtained with the neural network mode the linear model and persistence average (the actual value is the average over the period for that hour). Again, the neural network shows the best performance. In Figure 5 we observe the comparison between observed and forecasted values using the neural network model at 11 PM (4 hours in advance). Results correspond to June 2012. During this month, episodes are more frequent and 11 PM is around the

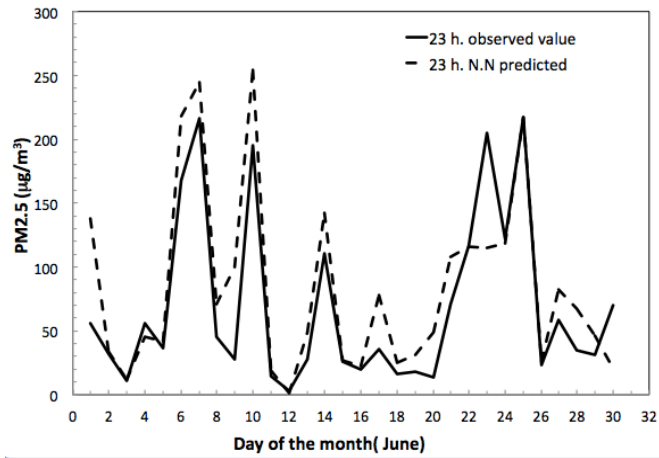


Figure 5: Forecasted v/s observed PM2.5 concentrations at 11 PM during June, 2012.



time when the maximum hourly concentration occurs on episode days (which normally affects the 24 h average during the following day). We can verify that most episodes are captured. During 2012, 17 days exceeding the $80 \mu\text{g}/\text{m}^3$ 24 h average occurred. From these, 70% were correctly forecasted on the basis of the hourly forecasted values.

5 NO₂ concentrations

It is well known that PM_{2.5} concentrations have an important contribution from secondary particle formation. Given that chemical analysis of particles in Cerro Navia station indicates a significant presence of nitrates, we have added a new input variable to the neural network model. This is the average NO₂ concentration between 10 AM and 5 PM for the present day, in order to forecast hourly PM_{2.5} concentrations starting at 8 PM of the present day.

Figure 6 displays a comparison between correlations using the neural network previously reported and the new neural network including NO₂ concentrations. We observe an improvement of performance from 9 to 20 hours in advance forecasting. The effect is also evident when comparing NPE, which is shown in Figure 7.

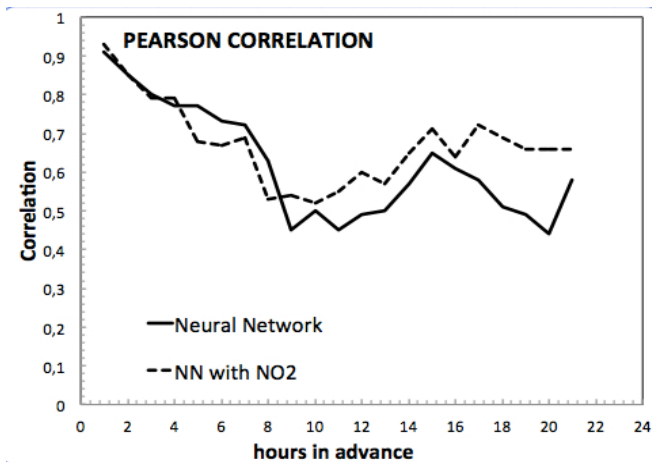


Figure 6: Comparison between Pearson correlations obtained using neural network model with and without NO₂ information as input.

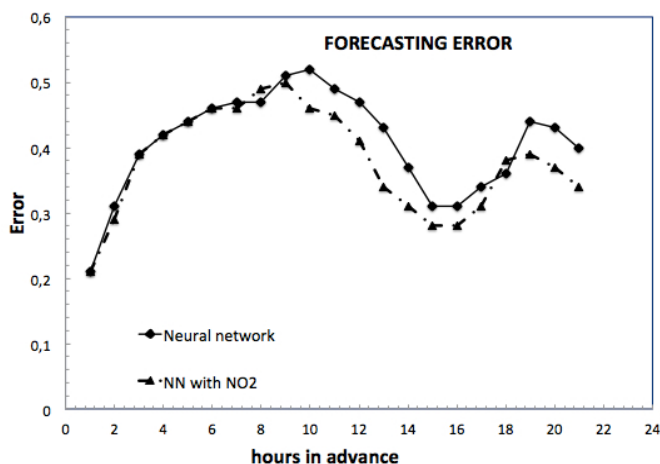


Figure 7: Normalized percent errors using neural network with and without NO₂ information as input.

6 Conclusions

Results reported here represent a refinement and improvement with respect to a recent publication [11]. The ability to forecast hourly PM_{2.5} concentrations in cities with unfavourable conditions for pollutant dispersion may be used to focus efforts in order to avoid in part harmful effects on population health. Including gas concentrations that are precursors of secondary particle formation may be important.

Acknowledgements

We would like to express our thanks for the support from the research office, Universidad de Santiago de Chile (DICYT) through project 091531PJ. We also appreciate the collaboration of C. Menares for his help in processing the data.

References

- [1] Jorquera, H., Barraza, F. Source apportionment of PM_{2.5} in Santiago, Chile: 1999 and 2004 results. *Sci Total Environ* 435–436, 418–429 (2012)
- [2] Gramsch, E., Reyes, F., Oyola, P., Rubio, M. A., López, G., Pérez, P. and Martínez, R. Particle size distribution and its relationship to black carbon in two urban and one rural site in Santiago de Chile, *Journal of the Air & Waste Management Association*, 64(7): 785–786 (2014)
- [3] Fernando, H. J. S., Mammarella, M. C., Grandoni, C., Fedele, P., Di Marco, R., Dimitrova, R., Hyde, P. Forecasting PM₁₀ in metropolitan areas: Efficacy of neural networks. *Environmental Pollution* 163, 62–67 (2012)



- [4] 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)
- [5] Perez, P., Salini, G., PM_{2.5} forecasting in a large city: Comparison of three methods. *Atmospheric Environment* 42, 8219–8224 (2008)
- [6] Cobourn, W. G., An enhanced PM_{2.5} air quality forecast model based on nonlinear regression and back-trajectory concentrations. *Atmospheric Environment* 44, 3015–3023 (2010)
- [7] 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)
- [8] Thomas, S., Jacko, R. B., Model for forecasting expressway fine particulate matter and carbon monoxide concentrations: application of regression and neural network models. *J. Air Waste Manag Assoc* 57, 480–488 (2007)
- [9] Ruttlant, J., Garreaud, R., Meteorological air pollution potential for Santiago, Chile: towards an objective episode forecasting. *Environmental Monitoring and Assessment* 34, 223–244 (1995)
- [10] Chen, J., Lu, J., Avise, J. C., DaMassa, J. A., Kleeman, M. J., Kadwela, A. P., Seasonal modeling of PM_{2.5} in California's San Joaquin Valley. *Atmospheric Environment* 92, 182–190 (2014)
- [11] Perez, P., Gramsch E., Forecasting hourly PM_{2.5} in Santiago de Chile with emphasis on night episodes. *Atmospheric Environment* 124, 22–27 (2016)

