$PM_{2.5}$ FORECASTING IN THE MOST POLLUTED CITY IN SOUTH AMERICA

PATRICIO PEREZ[1], CAMILO MENARES[2] & CAMILO RAMIREZ[1]
[1]Departamento de Fisica, Universidad de Santiago de Chile, Chile
[2]Departamento de Geofisica, Facultad de Ciencias Fisicas y Matematicas, Universidad de Chile, Chile

ABSTRACT
According to a recent study of WHO, Coyhaique, a small city in the south of Chile is the most polluted city in South America. With 70,000 habitants, the reasons for the high $PM_{2.5}$ concentrations in the city area during fall and winter are: topographic situation, stable atmospheric conditions and intense use of wood stoves for heating. During 2016, the 24h moving average exceeded 170 micrograms per cubic meter for 63 days. A neural network model that uses previous values of $PM_{2.5}$, meteorological information and previous concentrations of NO_2 and CO as input, which is trained with 2014 and 2015 data, is able to forecast 91% of these exceedances. This forecasting is very useful in order to alert the population and to motivate the authorities to take actions to control the emissions.
Keywords: air quality forecasting, wood stoves, neural networks, episodes.

1 INTRODUCTION

Coyhaique is a small city located in the south of Chile at 45°34’S and 72°04’W. Average altitude is 280 m over sea level and lies in a valley between two rivers. Nearby mountains have altitudes of the order of 1500 m. which contribute to poor dispersion of pollutants in the urban area especially in fall and winter when radiative cooling is more intense. At present, the city has a population approaching 70,000 habitants. Fig. 1 shows a satellite picture of the city in which we can observe surrounding elevations. Defining the cold season as the time between April and September, we observe that in the cold season, average minimum and maximum temperatures are 0°C and 8°C respectively. Average annual precipitation is 1000 mm, with 65% of it concentrated in the cold season. Average wind speed in the cold season is 2.5 m/s. Generalized heating fuel used in the city is wood. Given that most of the time, the wood used is not properly dried and that many stoves do not have an advanced technology to control emissions of particulate matter, during the cold season concentrations of $PM_{2.5}$ reach extremely high values. During 2016, average $PM_{2.5}$ in Coyhaique was 66 $\mu g/m^3$. According to a recent ranking of pollution in cities around the world, although Coyhaique is classified in position 143, behind many cities in Asia and Africa, it becomes the most polluted city in South America [1]. During the cold season, the 24 h average of $PM_{2.5}$ is in many occasions greater than 170 $\mu g/m^3$, which defines an “emergency” condition of air quality. In order to anticipate these situations, it appears convenient to have an operational forecasting model which can be used by authorities and people in general. A simple way to build a model to estimate local conditions without requiring significant resources is to use a statistical model. Among statistical models for particulate matter forecasting are linear regressions and neural network models [2]. Most of the studies have focused on the decrease of average forecasting errors but not on the correct estimation of episodes of high concentrations.

In the present study we describe the properties of two $PM_{2.5}$ forecasting models, a multilayer network and a linear model, adapted to Coyhaique conditions. An important goal for us is the detection of high concentrations episodes.

WIT Transactions on Ecology and the Environment, Vol 230,
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/AIR180181

Figure 1: The city of Coyhaique. Satellite picture.

2 DATA

We analyze data from years 2014, 2015 and 2016, with emphasis on the cold period(April-September). Pollution and meteorological information is obtained from a monitoring station located in a representative place of the city. We must mention that according to Chilean environmental regulations, four relevant ranges of 24 h average concentration are defined. Class A: when concentrations are les than 80 μg/m^3, class B: when concentrations are between 80 μg/m^3 and 109 μg/m^3, class C: concentrations between 110 and 169 μg/m^3 and class D for concentrations greater than or equal to 170 μg/m^3. For days in classes B, C, D, the Ministry of the Environment imposes restriction to emissions, which involves mostly wood use regulations, because wood combustion in stoves and kitchens is the main source of pollution. Table 1 shows the average concentrations for the cold period and the days when 24 h average is exceeding the exceeding the 80 μg/m^3 limit.

Table 1: $PM_{2.5}$ Statistics for cold period in Coyhaique. $PM_{2.5}$ in μg/m^3. B+C+D are the number of days when restrictions to emissions are imposed.

Year	$PM_{2.5}$ ave	B+C+D
2014	112	111
2015	74	91
2016	115	118

Fig. 2 shows the hourly $PM_{2.5}$ concentrations for different hours of the day, averaged during the cold period. Peak concentrations around 9 PM may be explained by the presence of night thermal inversion and the most intense use of wood stoves. Morning peak signals stove use and traffic. Minimum around 3 PM has to do with more favourable conditions for pollutant dispersion under the presence of stronger winds.

We also pay attention to NO_2 and CO concentrations. The first may have an important effect on secondary particle formation [3]. The second is found to have a high correlation with particulate matter concentrations and may become a good predictor observing that the increase of evening hourly values of CO starts earlier than $PM_{2.5}$ (Fig. 3).

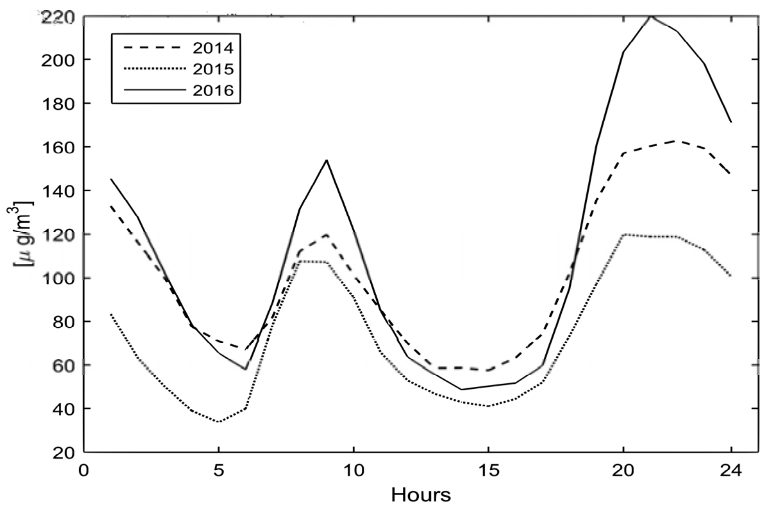

Figure 2: Hourly $PM_{2.5}$ concentrations as a function of hour of the day.

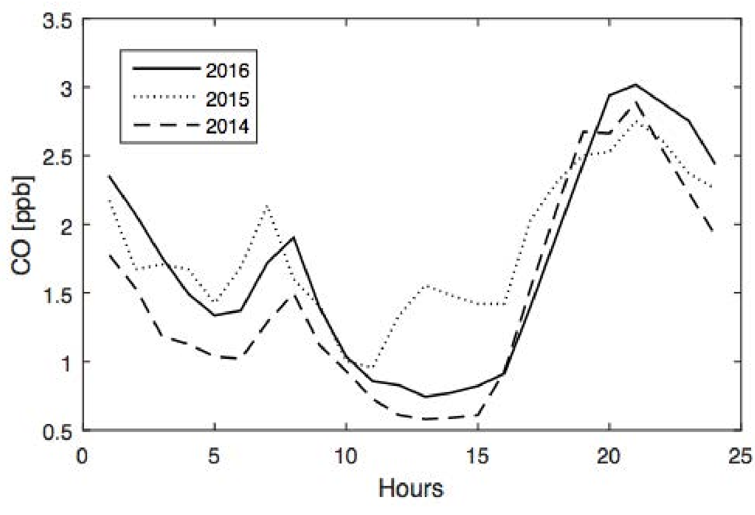

Figure 3: Hourly CO concentrations during cold season in Coyhaique.

3 MODELING

Based on our experience with particulate matter concentration forecasting for the city of Santiago de Chile [4,5] and from the positive results obtained elsewhere by many authors, we implemented a feedforward neural network and a linear model in order to forecast the maximum of the 24 h moving average of $PM_{2.5}$ one day in advance. Exploration of potential input variables was performed based on the case of Santiago, adding some variables that may be more specific to Coyhaique situation. The best results are obtained with the following 11 input variables:

- Hourly $PM_{2.5}$ measured at 6 PM of present day.
- Hourly $PM_{2.5}$ measured at 7 PM of the present day.
- Forecasted average temperature between 7 PM of present day and 7 AM of next day.
- Forecasted average wind speed between 7 PM of present day and 7 AM of next day.
- Thermal amplitude for present day.
- Accumulated precipitation for present day (including forecast between 7 PM and midnight of present day).
- Average NO_2 concentration between 6 AM and 7 PM of present day.
- Average CO concentration between 6 AM and 7 PM of present day.
- Today`s maximum temperature.
- Today`s minimum temperature.

Output is the forecasted value of maximum of the 24 h moving average for the next day. The neural network has 7 units in a unique hidden layer, and the linear model has the same eleven inputs. The reason to include measured data only until 7 PM, is because the goal is to generate a forecasting report at 8 PM of the present day, such that with this information, the authority can transmit it to the population. We have trained the models, determining connection weights with 2014 and 2015 data. The test was made with independent data corresponding to year 2016. In order to analyse the performance of the models we have calculated the normalized percent error, defined as:

$$NPE = \frac{\langle |y_o - y_F| \rangle}{\langle y_o \rangle} \cdot 100 , \qquad (1)$$

where y_O is the observed value and y_F is the forecasted value. The triangular bracket is the average over all the test sample, and the denominator has the shown value in order to eliminate the distortion that may appear in the case of low concentrations. For the feedforward neural network, error obtained was 18% and for the linear model it was 22%. More important than obtaining a small average error is to be able to forecast the high concentration values contained in classes B, C and D. To test this, we display contingency tables of results for both models.

We can verify that for the forecasting of episodes of classes B, C and D the neural network model is slightly more accurate. Overall class correct forecasting is 75% for the neural model and 72% for the linear model. One more class D case is detected by the neural network. Both models have a reasonable good performance for class forecasting and may be used as an operational model for the city. The neural model forecasted 58 class D days,

Table 2: Contingency table for the 2016 performance of neural network forecasting model.

			Forecast Neural model				
		A	B	C	D	TOT	%O
O	A	**51**	9	2	0	62	82
B	B	7	**15**	3	0	25	60
S	C	2	7	**19**	5	33	50
	D	1	2	**7**	**53**	63	84
	TOT	61	33	31	58	**183**	*75*
	%F	84	45	61	91		

Table 3: Contingency table for the 2016 performance of linear forecasting model.

			Forecast Linear model				
		A	B	C	D	TOT	%O
O	A	**51**	9	2	0	62	82
B	B	8	**12**	5	0	25	48
S	C	1	10	**17**	5	33	52
	D	0	3	**8**	**52**	63	83
	TOT	61	34	32	57	**183**	*72*
	%F	85	35	53	91		

from which 53 occurred and 5 of them turned out to be class C. For class C, we observe 5 over forecasting, two verified as class A and 3 in class B.

4 CONCLUSIONS

We have shown that with an appropriate choice of input variables it is possible to implement an operational $PM_{2.5}$ statistical forecasting model which can be used by authorities and citizens in general. We believe that restriction to emissions applied in the city of Coyhaique during episodes do not have a significant impact in air quality conditions, because otherwise we would expect a much higher percentage of over forecasting, which indeed is not evident from the calculated contingency tables. It is likely that high concentrations of particulate matter will prevail in the cold season unless the habitants change their heating fuel from wood to something cleaner as natural gas or electricity.

www.witpress.com, ISSN 1743-3541 (on-line)

ACKNOWLEDGEMENT

We would like to thank the support of the Research Department of Universidad de Santiago de Chile (DICYT) through project 091731PJ.

REFERENCES

[1] WHO Global Urban Ambient Air Pollution Database (update 2016). http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/.

[2] Shahraiyni, H.T. & Sodoudi, S., Statistical modelling approaches for PM_{10} prediction in urban areas: A review of 21st studies. *Atmosphere*, **7**(2), p. 15, 2016.

[3] Perez, P., Forecasting hourly values of $PM_{2.5}$. *Air Pollution*, 2016.

[4] Perez, P. & Salini, G., $PM_{2.5}$ forecasting in a large city: Comparison of three methods. *Atmos. Environ*. **42**, pp. 8219–8224, 2008.

[5] Perez, P. & Gramsch, E., Forecasting hourly $PM_{2.5}$ in Santiago de Chile with emphasis on night episodes. *Atmos. Environ*., **124**, pp. 22–27, 2016.