A Dynamic Multiple Equation Approach for Forecasting $PM_{2.5}$ Pollution in Santiago, Chile

[1], Stella Moisan[1], Rodrigo Herrera[1] and Adam Clements[2]

[1]Facultad de Economía y Negocios, Universidad de Talca, Chile.

[2]School of Economics and Finance, Queensland University of Technology, Australia.

Abstract

A methodology based on a system of dynamic multiple linear equations is proposed that incorporates hourly, daily and annual seasonal characteristics to predict hourly $PM_{2.5}$ pollution concentrations for 11 meteorological stations in Santiago, Chile. It is demonstrated that the proposed model has the potential to match or even surpass the accuracy of competing nonlinear forecasting models in terms of fit and predictive ability. In addition, the model is successful in predicting various categories of high concentration events, between 53% to 76% of mid-range, and around 90% of extreme-range events on average across all stations. This forecasting model is considered a useful tool for government authorities to anticipate critical episodes of air quality so as to avoid the detrimental economic and health impacts of extreme pollution levels.

Key Words: Air quality, Particulate matter, Dynamic multiple equations.

1 Introduction

In 2014, The Organisation for Economic Co-operation and Development (OECD) ranked Chile as the country with the highest air pollution among its 36 members. Furthermore, the capital Santiago, where 41% of the country's total population resides, is ranked fourth in terms of cities with the worst air quality on the continent (WHO, 2011). The components of pollution of most concern is particulate matter with a diameter less than either 10 or in particular 2.5 µm, PM_{10} and $\text{PM}_{2.5}$ respectively. In Chile at least 60% of the inhabitants are exposed to $\text{PM}_{2.5}$ concentrations over the annual US norm of 15 µg/m^3(Cifuentes, 2010), with the World Health Organization (WHO) suggesting an annual limit of 10 µg/m^3, since its effects on health are more severe than those of PM_{10} (Kelly & Fussell, 2012). In fact approximately 4,000 premature deaths due to chronic exposure to this component of pollution have been recorded according to the Chilean Ministry for the Environment, (Ministerio del Medio Ambiente, MMA), MMA (2011). There are also significant broader economic consequences of air pollution. In 2013, the World Bank estimated that lost work-related income due to air pollution was USD 225 billion. In Chile, the net economic benefits of effectively regulating $\text{PM}_{2.5}$ is estimated to be USD 7.1 billion according to their National System of Environmental Information (SINIA), SINIA (2010). Studies such as Böhringer & Jochem (2007) show that incorporating environmental quality into any analysis of economic and social issues, is key to the sustainable development of nations.

Given the negative impacts of $\text{PM}_{2.5}$, the central aim of this work is to propose a multiple linear equation model with dynamic coefficients, which can be easily interpreted and capable of capturing the stylized features of $\text{PM}_{2.5}$. The predictive ability of this approach is then compared to a number of more complex competing approaches. Predictive models are important so that government authorities can take efficient action to minimise the economic consequences of heightened pollution levels. It is found that the proposed dynamic multiple equation model provides more accurate forecasts then its more complex nonlinear competitor. A result in itself that is very encouraging. From a more practical point of view, the proposed model is only based on linear regression meaning that estimated coefficients are easily interpretable (an important issue if government agencies are trying to develop a deeper understanding of the factors driving pollution levels around the city), the model will reduce the risk of in-sample overfitting, it can be re-estimated in real-time as computational cost is extremely low (its main nonlinear competitor often takes over four times longer for coefficient estimation) and avoids the use of potentially complex or expensive software to implement.

In Chile, the regulated target level for $PM_{2.5}$ is $50\,\mu g/m^3$ over a 24-hour average. Above this threshold, three categories of critical episodes are defined: Alert ($80 – 109\,\mu g/m^3$), Pre-emergency ($110 – 169\,\mu g/m^3$) and Emergency ($>170\,\mu g/m^3$). Current forecasting methodology for particulate matter in Chile is based on a multiple linear regression model proposed by Cassmassi (1999). However, its prediction focuses on PM_{10} and its accuracy in forecasting extreme pollution levels in Santiago has been questioned (Delgado *et al.*, 2006), recording for example, a 44% rate of accuracy for alerts[1].

Different methodologies have been proposed to predict $PM_{2.5}$ concentrations in the short term, with the recent literature employing both linear and nonlinear econometric models. Linear specifications include Kalman filtering (Sahu & Mardia, 2005; Djalalova *et al.*, 2015), multiple linear regression models (Chaloulakou *et al.*, 2003; Genc *et al.*, 2010; Vlachogianni *et al.*, 2011) and autoregressive integrated moving average (ARIMA) models (Jian *et al.*, 2012). The last two are widely used in forecasting due to their accuracy and ease of interpretation of their coefficients (Zhou *et al.*, 2014). Nonlinear models include a support vector machine (Lu & Wang, 2005; Osowski & Garanty, 2007; Weizhen *et al.*, 2014), a hidden Markov model (Sun *et al.*, 2013) and artificial neural network (ANN) models (McKendry, 2002; Kukkonen *et al.*, 2003; Ordieres *et al.*, 2005; Prakash *et al.*, 2011). ANN have been found to successfully model time series with complex characteristics in different fields (Hill *et al.*, 1996; Hamzaçebi *et al.*, 2009; Zhang *et al.*, 2012) generating accurate forecasts over the long term given their capacity to make forecasts with multiple advanced notice periods (Tang & Fishwick, 1993).

In the case of Chile, studies focusing on forecasting $PM_{2.5}$ mainly employ nonlinear models (Perez & Reyes, 2006; Díaz-Robles *et al.*, 2008; Perez & Gramsch, 2016). For example, Perez & Gramsch (2016) use an ANN to forecast critical episodes of $PM_{2.5}$ during winter night periods using hourly historical PM_{10} and $PM_{2.5}$ data, concentrations from nearby stations and weather variables. They show that the model correctly predicts up to 70% of critical episodes of $PM_{2.5}$, which is attributed to the inclusion of a factor of ventilation

[1] Geophysics Department of the University of Chile (http://www.geofisica.cl/English/pics3/FUM6.htm).

as a covariate. Díaz-Robles *et al.* (2008) consider a hybrid model combining ARIMA and ANN structures, predicting PM_{10} for 2006 at the 'Las Encinas' station in Temuco, Chile. Their results show that the hybrid model captures 80% of the pre-emergency episodes at this location. Finally, Saide *et al.* (2011) propose a deterministic chemical based forecasting model for $\mathrm{PM}_{2.5}$ using carbon monoxide CO as a tracer due to its high correlation to predict critical night episodes. They conclude that the greatest benefit of the model is its ability to forecast up to 48 hours ahead.

The paper proceeds as follows. In Section 2, the stylized facts of the $\mathrm{PM}_{2.5}$ time series that motivate the proposed methodology are presented. Section 3 introduces the methodology including three specifications of the proposed model along with two competing approaches: a seasonal ARIMA model with exogenous variables (SARIMAX) and a nonlinear Artificial Neural Network (ANN) model. Section 4 discusses the estimation and prediction results. Finally, Section 5 provides concluding comments.

2 Description of the Data

The data used in this study are hourly historical observations of weather and environmental concentrations for 11 monitoring stations located in Santiago, Chile. The data was collected from the National Air Quality Information System (SINCA) for the period January 1, 2011 to August 31, 2015.

Figure 1 shows the geographic distribution of the monitoring stations where it is clear that the stations are not uniformly spaced across the Santiago city. There are stations separated by large distances, such as Talagante and Las Condes, where there is likely to be little relationship between their concentration levels. In contrast, there is likely to be interactions between the stations in Pudahuel and Cerro Navia, given their proximity. The values in parentheses report the annual average of $\mathrm{PM}_{2.5}$ concentrations for the 2011 – 2015 period at each station. An interesting pattern is that stations with high average concentrations are located in communes with a higher population density and industrialization. A clear example is the Cerro Navia station, which is surrounded by the greatest population density of the stations analyzed with a value of 13,361 inhab/km^2, where the annual average of $\mathrm{PM}_{2.5}$ in the period is 29.45 µg/m^3, the highest of all the stations studied.

To gain a deeper understanding of the stylized facts of the $\text{PM}_{2.5}$ concentrations, Figure 2 shows three violin plots characterizing the distribution of $\text{PM}_{2.5}$ at the Pudahuel monitoring station. According to the National Institute of Statistics (Instituto Nacional de Estadisticas, INE), this station is located in the commune with the largest geographical area, with a land surface of $197.4\,\text{km}^2$ and 195,653 inhabitants, INE (2007). In addition, it is one of the country's most polluted communes in terms of annual average $\text{PM}_{2.5}$ concentration, reaching $34\,\mu\text{g/m}^3$ in 2015.

The upper panel shows the hourly pattern of the time series: the highest hourly average $\text{PM}_{2.5}$ concentration is between 6:00 A.M. and 10:00 A.M, due to heavy traffic as the population begins the work day. Later, the greatest dispersion of hourly average $\text{PM}_{2.5}$ concentration corresponds to the time between 6:00 P.M. and 3:00 A.M. as the work day ends and people return to their homes between 6:00 P.M. and 9:00 P.M., and temperatures in winter decrease every day between 10:00 P.M. and 4:00 A.M.

The second panel characterizes the daily average $\text{PM}_{2.5}$ concentration according to the day of the week. While it is difficult to discern with the naked eye, weekend mean concentration levels are slightly lower, with greater dispersion in concentrations observed across Friday, Saturday and Sunday. This is likely due to the traffic following a stable pattern during weekdays, while on the weekends, according to the National Commission for the Environment (Comisión Nacional del Medio Ambiente, CONAMA), much of the transient population of 1,800,000 people made up of workers and students return to their homes from Santiago, CONAMA (2005).

Finally, the third panel highlights the annual seasonality in the form of monthly averages $\text{PM}_{2.5}$ concentrations. Note that this annual pattern is due to there being on average more pollution in the autumn-winter months, i.e., April to August, than at other times of the year. Different factors help explain this behavior. For example, the low temperatures registered in this period mean the demand for heating homes in Santiago increases, normally from the burning of fossil fuel material and firewood leading to increased pollution levels. Therefore, the primary source of $\text{PM}_{2.5}$ in Santiago, in terms of annual average, is firewood at 45%, followed by transport 33%, industry 16%, agriculture 4% and non-firewood heating 2% (MMA, 2012).

Temperature should play an important role in the prediction of the $\text{PM}_{2.5}$ concentration,

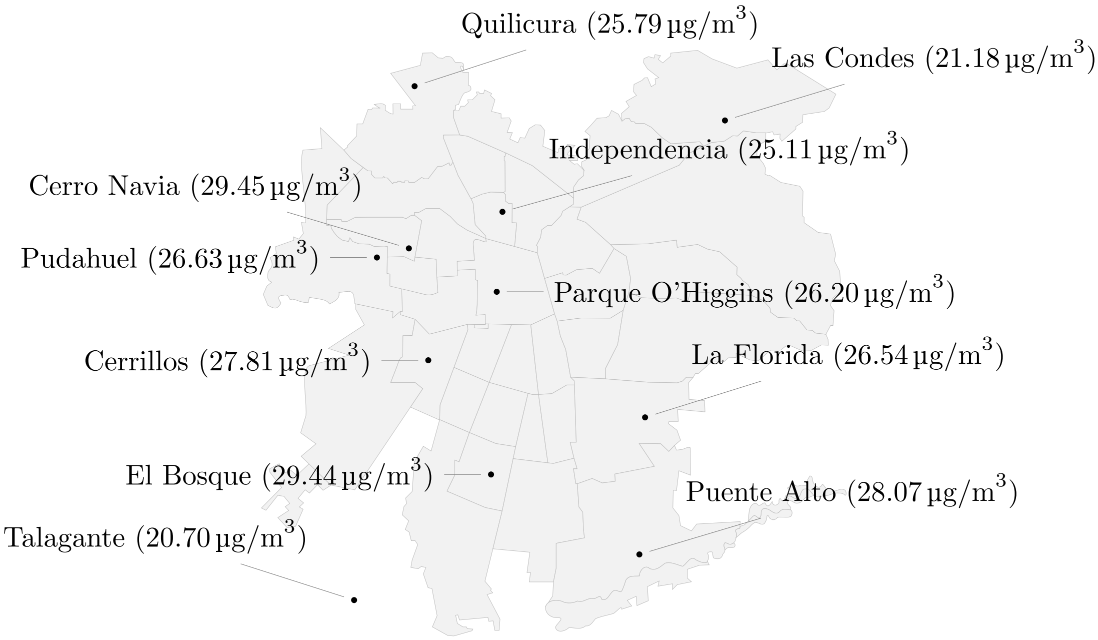

Figure 1: Map of 11 monitoring stations in Santiago, Chile. Values in parentheses indicate hourly average $\text{PM}_{2.5}$ in the study period (2011 – 2014) for each commune corresponding to the monitoring station.

given its impact on the atmospheric and ventilation conditions in the Santiago river basin and its impact on the demand for heating. Different studies also use temperature and relative humidity as explanatory variables in $\text{PM}_{2.5}$ prediction models (Kurt & Oktay, 2010; Zhou *et al.*, 2014; Feng *et al.*, 2015; Saide *et al.*, 2016).

In addition to weather variables, it is also possible to relate $\text{PM}_{2.5}$ to environmental concentrations. For example, studies such as Shah *et al.* (2004) and Wang *et al.* (2010) report that pollution from cars and burning firewood is associated with the current level of carbon monoxide (CO) in the atmosphere, which comprises up to 54% of the $\text{PM}_{2.5}$ concentration. In particular, Saide *et al.* (2011) reports a high correlation between the levels of CO concentration and $\text{PM}_{2.5}$ in Santiago, this even being over 0.95 during night periods in winter.

Figure 3 shows the dynamic behavior of $\text{PM}_{2.5}$ in relation to a set of weather and environmental covariates for whole sample period (2011 – 2015) at the Pudahuel monitoring station. Along with the variables themselves, a weekly moving average is shown as a thick red line. Indeed, a direct relation exists between CO and $\text{PM}_{2.5}$ with a positive correlation of 0.84, similar to what has been observed in the literature (Naeher *et al.*, 2001; Saide *et al.*,

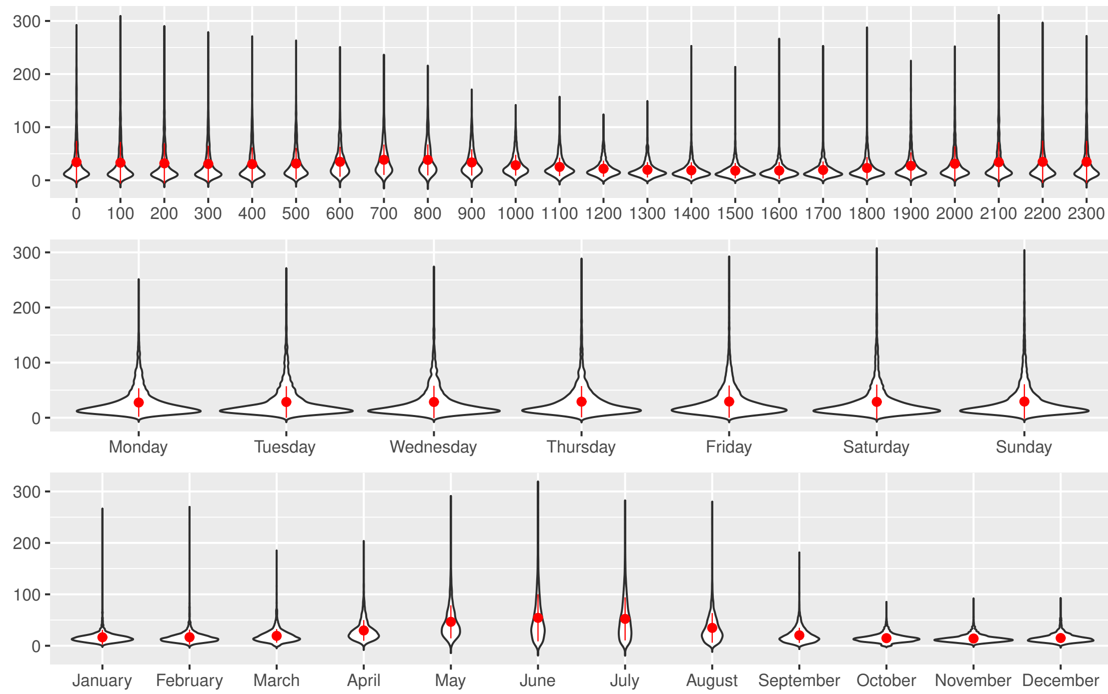

Figure 2: From upper to lower panel: violin plot for hourly average PM$_{2.5}$ concentration (upper), average for days of the week (middle) and monthly (lower) for the Pudahuel monitoring station from January 2011 to August 2015.

2011). Temperature ($Temp$) and wind speed (WS) are noteworthy among the weather covariates, having a strong relationship with PM$_{2.5}$ with a negative correlation of -0.41 and -0.38 respectively. On the other hand, the relative humidity (RH) exhibits a positive correlation with PM$_{2.5}$ of 0.23, whereas the wind direction (WD) shows a weak negative correlation of -0.09.

In particular, WS is an important weather variable as wind assists in dispersing pollution particles. Thus, low WS values favor the accumulation of contaminants; however, if WS is high, greater ventilation is experienced in the region (Saide *et al.*, 2016). This explains the negative correlation with the PM$_{2.5}$ concentration. When the WS values are high, the pollution particles dissipate faster, thereby reducing the PM$_{2.5}$ concentration; if this value is low, the ventilation of PM$_{2.5}$ decreases.

It should be pointed out that this effect across different stations will also depend on the WD at each particular station. Garreaud & Rutllant (2006) shows that southwesterly winds lead to the dispersion of pollution and the intake of clean air towards the Santiago river basin. Thus, an interaction term between wind speed and direction can capture the

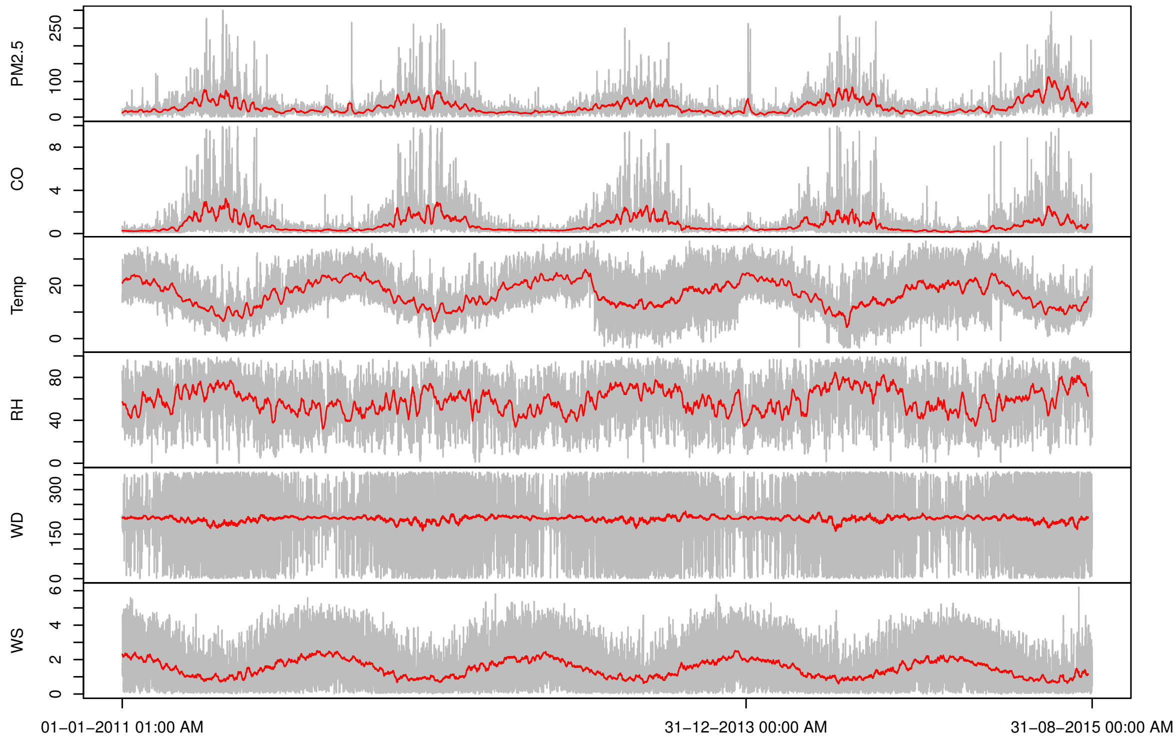

Figure 3: From upper to lower panel: hourly time series of $\text{PM}_{2.5}$ concentration, CO concentration [ppb], temperature [$^\circ$C], relative humidity [%], wind direction [$^\circ$] and wind speed [m/s] for Pudahuel monitoring station from January 2011 to August 2015.

natural ventilation conditions of surrounding a station (Horan & Finn, 2008).

The weekly moving averages highlight any annual cycles present and emphasizes the correlation between the variables. It is clear that CO, WS and $Temp$ share a very similar annual cycle with $\text{PM}_{2.5}$. While the annual cycle in RH is not as pronounced, RH is generally higher during the winter months when $\text{PM}_{2.5}$ is higher. In summary, the proposed model uses the following covariates: carbon monoxide (CO) measured in parts per billion (ppb), temperature ($Temp$) measured in degrees Celsius, the percentage of relative humidity (RH) in the atmosphere and wind speed (WS) in m/s interacting with wind direction (WD) defined in degrees.

3 Methodology

A dynamic multiple equation (DME) model is proposed for the purposes of forecasting PM2.5. The structure is designed to capture the salient features of PM2.5 and contains 24 equations, one for each hourly interval h within a day. Along with the regular patterns, a range of weather and environmental variables are included following a number earlier studies (Hien *et al.*, 2002; Saide *et al.*, 2011; Zhou *et al.*, 2014). A similar model was used by (Clements *et al.*, 2016) for forecasting electricity demand in the Australian National Electricity Market. Electricity demand exhibits broadly similar diurnal and seasonal patterns to PM2.5 levels. The importance of the proposed model lies in its ease of interpretation as it is linear in the parameters.

PM_t is used to denote the PM2.5 concentration observed at any each station at hour $h = 1, \ldots, 24$, where the index t is used to indicate the pooled time of the series, with one calender year containing 8,760 hourly observations. Based on this time index t, it is easy to determine the current day, week or month if necessary, and hence the simple notation is used. The base specification proposed for the DME is given by the following hourly equation h:

$$PM_t = \theta_h^0 + \theta_h(t)PM_{t-24} + \gamma_h(t)PM_{t-168} + \phi_{1h}\varepsilon_{t-24} + \phi_{1h}\varepsilon_{t-168} + \varepsilon_t + \delta^\top \mathbf{Z_{t-j}} \tag{1}$$

As each hour of the day is governed by a separate equation, the intercepts, θ_h^0, $h = 1, \ldots, 24$ control the diurnal pattern within a calendar day. $\varepsilon_t \sim N(0, \sigma_h^2)$ is the residual term, and moving averages at one-day ε_{t-24} and seven-days ε_{t-168} are also included to complete the vector ARMA structure. The weekly cycle is incorporated by allowing the autoregressive coefficient on one-day lagged PM2.5, PM_{t-24} to be a function of the day of the week in the following way:

$$\theta_h(t) = \sum_{d=1}^{7} \eta_{hd} W_d(t) \tag{2}$$

where W_p is a dummy variable, taking a value of 1 if the pooled time t corresponds to the day of the week d and zero in the other case; η_{hd} corresponds to the coefficients to be estimated. This structure allows for differences for example, between using Sunday to forecast Monday and using Tuesday to forecast Wednesday.

Similar to the inclusion of the weekly pattern through the $\theta_h(t)$ coefficients, the annual cycle is incorporated through the interactions of the $\gamma_h(t)$ coefficients on the one-week lag PM_{t-168}. This annual cycle is captured through a Fourier polynomial with annual cycles:[2]

$$\gamma_h(t) = a_{h0} + \sum_{j=1}^{4} \left[a_{hj} \sin\left(2j\pi\left(\frac{t}{8760}\right)\right) + b_{hj}\cos\left(2j\pi\left(\frac{t}{8760}\right)\right)\right] \tag{3}$$

where a_{h0}, a_{hj} and b_{hj} are the coefficients of the polynomial for $h = 1, \ldots, 24$ and $j = 1, \ldots, 4$.

The proposed model also includes a set of environmental and weather covariates related to the $\text{PM}_{2.5}$ concentration level with j hours of delay, with three specifications proposed. The first of these is given by:

$$\begin{aligned} \delta^{\top}\mathbf{Z_{t-j}} = \delta_{1h}PM_{t-j} + \delta_{2h}PMmax_{t-j} + \delta_{3h}CO_{t-j} + \delta_{4h}Temp_{t-j} \\ +\delta_{5h}RH_{t-j} + \delta_{6h}(WD_{t-j})WS_{t-j} \end{aligned} \tag{4}$$

The first two covariates correspond to the $\text{PM}_{2.5}$ concentration with a one-hour delay and the maximum $\text{PM}_{2.5}$ concentration in last the 24 hours ($PMmax$). The latter, given that if it reflects an above-standard value, it is likely to be repeated or it will be difficult to reduce in the following 24 hours, therefore acting as a good predictor of the following day's concentration. The third covariate is the hourly environmental concentration of CO, which is directly related to the $\text{PM}_{2.5}$ concentration, as discussed in the previous section[3]. The final covariates are $Temp$, RH and WS as an interaction term with WD, which is defined by using a combination of dichotomous variables given by:

$$\delta_{6h}\left(WD_t\right) = \pi_{1h}N_t + \pi_{2h}S_t + \pi_{3h}E_t + \pi_{4h}W_t \tag{5}$$

where North (N) is between 45 $^\circ$ and 135 $^\circ$, South (S) between 225 $^\circ$ and 315 $^\circ$, East (E) between 315 $^\circ$ and 45 $^\circ$ and West (W) between 135 $^\circ$ and 225 $^\circ$. These each take the value of 1 if the wind is blowing from that specific direction. As discussed in Section 2, WD is designed to reflect the atmosphere's ventilation conditions around the meteorological

[2]A fourth-degree Fourier polynomial is found to be a good compromise between goodness of fit and simplicity of the model.

[3]Note that in the main specification in Eq.(1), we have not included the hourly delay, since the autoregressive model is defined in daily terms.

station.

The second specification is designed to capture spatial effects and includes a covariate PMc, which represents the average PM$_{2.5}$ concentrations at the neighboring stations:

$$\begin{aligned}\delta^\top \mathbf{Z_{t-j}} &= \delta_{1h}PM_{t-j} + \delta_{2h}PMmax_{t-j} + \delta_{3h}CO_{t-j} + \delta_{4h}Temp_{t-j} \\ &+\delta_{5h}RH_{t-j} + \delta_{6h}(WD_{t-j})WS_{t-j} + \delta_{7h}PMc_{t-j}\end{aligned} \tag{6}$$

where,

$$PMc_t = \sum_{m=1}^{10} w_m PM_{mt}. \tag{7}$$

The other stations are denoted by $m = 1, \ldots, 10$. w_m is a specific weight corresponding to the Euclidean distance between the station under study and the other stations, standardized such that their total is one, and PM_{mt} is the level of PM$_{2.5}$ concentration at the neighboring stations. Thus, the closer another station is, the greater impact its concentration will likely have on PM$_{2.5}$ at the station of interest. The main idea of this specification is to control for possible spatial correlations among the concentrations at the different monitoring stations.

The third specification also captures the impact of the PM$_{2.5}$ concentrations at stations close to the station under study, but dynamically using the wind direction at those stations. The idea is to determine whether the wind at the stations nearby is moving in the direction of the station under study, and if so, this station would more likely be an influence, although this influence would be inversely proportional to the distance between these stations. This third model is specified as follows:

$$\begin{aligned}\delta^\top \mathbf{Z_{t-j}} &= \delta_{1h}PM_{t-j} + \delta_{2h}PMmax_{t-j} + \delta_{3h}CO_{t-j} + \delta_{4h}Temp_{t-j} \\ &+\delta_{5h}RH_{t-j} + \delta_{6h}(WD_{t-j})WS_{t-j} + \delta_{7h}(WDc_{t-j})PMc_{t-j}\end{aligned} \tag{8}$$

In this case, the spatially weighted concentrations PMc_t interact with the average direction of the wind at each of the nearby stations, WDc_t. To achieve this, the wind direction WD_t from each station m is decomposed into vectors x and y, giving greater weight to those closest to the station under study. These are determined by:

$$\partial x_{mt} = w_m \cos\left(\pi \frac{WD_{mt}}{180}\right), \partial y_{mt} = w_m \sin\left(\pi \frac{WD_{mt}}{180}\right) \tag{9}$$

Thus, the direction for this control station is obtained:

$$WDc_t = \cos^{-1}\left(\frac{\sum_{m=1}^{10} \partial x_{mt}}{\sqrt{\left(\sum_{m=1}^{10} \partial x_{mt}\right)^2 + \left(\sum_{m=1}^{10} \partial y_{mt}\right)^2}}\right) \tag{10}$$

Similar to Eq. (5), dummy variables are used to determine whether the wind direction at the nearby stations plays an important role in predicting the $PM_{2.5}$ concentration.

The multiple equation model in all its variants can be estimated equation-by-equation using the iterative ordinary least squares method proposed by (Spliid, 1983). Each equation is initially estimated ignoring the moving-average error terms and the regression residuals stored. The equations are then re-estimated using the regression residuals from the previous step as observed moving average error terms. This process is then iterated until convergence which is defined as the difference in parameter values in successive iterations being less than a user supplied tolerance, in this case the square root of machine precision for floating-point arithmetic.

3.1 Competing Models

Here, two competitors to the DME are presented, a SARIMAX model and an ANN model.

3.1.1 SARIMAX Model

A multiplicative double seasonal ARIMA model with exogenous variables (SARIMAX) is proposed (Box *et al.*, 2015) as $PM_{2.5}$ series exhibits, in addition to hourly patterns, daily and weekly seasonality. The general structure for the model is as follows:

$$\begin{aligned} \phi^p(L)\phi^{P1}(L_{S1})\phi^{P2}(L_{S2})(1-L)^d(1-L_{S1})^{D1}(1-L_{S2})^{D2}PM_t \\ = \delta_h + \Theta^q(L)\Theta^{Q1}(L_{S1})\Theta^{Q2}(L_{S2})\varepsilon_t \end{aligned} \tag{11}$$

Where PM_t is the $PM_{2.5}$ concentration in the period t, L is the delay operator, ϕ^p and Θ^q are standard autoregressive polynomials and moving averages of orders p and q respectively. Likewise, $\phi^{P1}(L_{S1})$ and $\phi^{P2}(L_{S2})$ determine the autoregressive polynomials of the orders P_1 and P_2, while $\Theta^{Q1}(L_{S1})$ and $\Theta^{Q2}(L_{S2})$ are the moving average polynomials

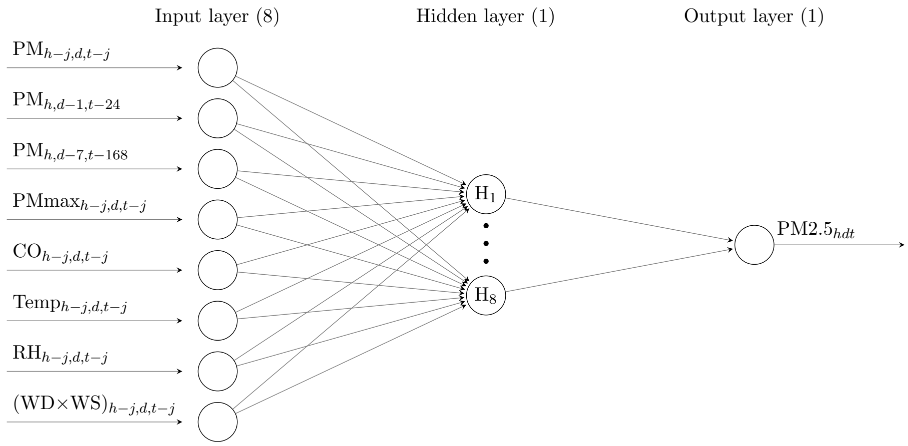

Figure 4: Artificial neural network (ANN) with feedforward structure and backpropagation learning algorithm used in this investigation.

of the orders Q_1 and Q_2. The order of integration for each component is defined as d, D_1 and D_2. Note that δ_h in this model represents the group of exogenous covariates according to three specifications used in the DME model defined in Eq. (4)–(8). Thus, the previous model is built for each station in such a way that it is effectively comparable to the proposed DME model, being expressed as SARIMAX $(p,d,q) \times (P_1,D_1,Q_1)_{S1} \times (P_2,D_2,Q_2)_{S2}$. In this case, the seasonal cycles S_1 and S_2 capture the daily and weekly patterns with $S_1 = 24$ and $S_2 = 168$, respectively with $D_1 = D_2 = 0$. Overall, the structure similar to the DME is SARIMAX $(1,0,1) \times (1,0,1)_{24} \times (1,0,0)_{168}$.

3.1.2 Artificial Neural Network Model

Among nonlinear prediction models, artificial neural networks (ANN) are a popular choice given their flexibility when dealing with seasonal patterns (Franses & Draisma, 1997). The most frequently used ANN is the feedforward type and the backpropagation learning algorithm, following the works by Feng *et al.* (2015) and Perez & Gramsch (2016). Figure 4 presents the structure of the ANN used. It consists of 8 unit input layers, using the same inputs as the exogenous variables used in the proposed DME model. These inputs feed into a hidden layer of 8 neurons, which are transformed to one output, corresponding to the $PM_{2.5}$ concentration over a specified forecast horizon.

The feedforward neural network refers to information only moving forward through the network in one direction, as represented by the arrows in Figure 4. Parameter estimation is based on the backpropagation learning algorithm of (Rumelhart *et al.*, 1986). Backpropagation minimizes error between the predicted and target values by propagating the errors back through the network to the hidden neurons where the weights are adjusted according to their previous contributions to the output. Observations over a two year and three month period prior to forecasting are used as a training set where this algorithm is used to minimize prediction error in this period before the subsequent forecasting exercise.

3.2 Measures of Fit and Comparing Predictive Ability

The mean absolute error and the root mean square error are used as measures to evaluate the fit of the forecast for the $PM_{2.5}$ time series. The simple prediction error is interpreted as $\varepsilon_t = Y_t - F_t$, where Y_t represents the observed values and F_t the predicted values.
Eq. (12) and (13) show the standard mean absolute error (MAE) and root mean square error (RMSE) measures respectively:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t| \tag{12}$$

$$RMSE = \left[\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t)^2\right]^{\frac{1}{2}} \tag{13}$$

In addition to these simple loss measures, two tests of predictive ability are used to statistically distinguish between the forecast accuracy of the competing models. The first is the test proposed by Diebold & Mariano (1995), a traditional test of unconditional predictive ability (DM test) to reveal whether there is a statistically significant difference between the forecast accuracy of two models and is based on the null hypothesis of no difference in the squared errors of the alternative models, $H_0 : E[(\varepsilon_t^1)^2 - (\varepsilon_t^2)^2]$. The test compares the performance of the proposed DME to either the ANN or SARIMAX, with a rejection of the null hypothesis indicating that the DME provides significantly more accurate predictions. Similarly, the test of conditional predictive ability proposed by Giacomini & White (2006), denoted as the GW test, is performed. This test is based on the same null hypothesis as the DM test, but its evaluation includes the backtesting period (2 years and 3 months) and all the prediction periods which vary according to the re-estimation periods (1, 3 and 24 hours). The GW test is more powerful than the DM test which only considers differences in average forecast performance.

These tests are based on point forecasts, however provide little information about the precision of the prediction when only a subset of observations are the focus. In contrast, probabilistic forecasts generated from predictive distributions provide a complete description of the uncertainty surrounding a point prediction. To overcome the shortcomings of relying on point forecasts, proper scoring rules can be used to compare different specifications by means of probabilistic forecasts of critical episodes.

The first scoring rule used is the continuous ranked probability score (CRPS) which is defined as

$$CRPS\left(F, y\right) = \int_{-\infty}^{\infty} \left(F\left(z\right) - \mathbb{1}\left\{y \leq z\right\}\right)^2 \mathrm{d}z,$$

where $F\left(z\right)$ is the predictive CDF, (Winkler *et al.*, 1996). The second scoring rule is the threshold-weighted continuous ranked probability score (twCRPS)

$$twCRPS\left(F, y\right) = \int_{-\infty}^{\infty} w(z) \left(F\left(z\right) - \mathbb{1}\left\{y \leq z\right\}\right)^2 \mathrm{d}z.$$

Here $w(z)$ is a nonnegative weight function which emphasizes a specific region of interest, critical episodes here. Following Amisano & Giacomini (2007) and Gneiting & Ranjan (2011), a weight function based on the normal distribution $\Phi\left(z \mid r, \sigma^2\right)$ with variance $\sigma^2 > 0$ is used. Here r denotes a high threshold as the focus here lies in the right tail of the distribution. In the empirical section three high thresholds defined in terms of quantiles (0.95, 0.99, 0.999) are used to assess the accuracy of the competing probabilistic forecasts. Note, that these scoring rules imply more accurate predictions result in lower scores. In order to implement the scoring rules, the numerical method proposed by Gneiting & Ranjan (2011) is used.

Under the DME model, the conditional distribution of PM_t is given by

$$\begin{aligned} &PM_t \mid PM_{t-24}, PM_{t-168}, \mathbf{Z_{t-j}} \\ &\qquad \sim \mathcal{N}\left(\omega_h + \theta_h(t) PM_{t-24} + \gamma_h(t) PM_{t-168} + \delta^\top \mathbf{Z_{t-j}}, \sigma_h^2\right) \end{aligned}$$

while for the SARIMA model the conditional distribution of PM_t is given by

$$
\begin{aligned}
PM_t \mid PM_{t-1}, PM_{t-24}, PM_{t-168}, \mathbf{Z_{t-j}} \\
\sim \mathcal{N}\left(\omega + \alpha PM_{t-1} + \theta(t) PM_{t-24} + \gamma(t) PM_{t-168} + \delta^\top \mathbf{Z_{t-j}}, \sigma^2\right)
\end{aligned}
$$

Given that no assumption is made with respect to the distribution of errors under the ANN approach, for the purposes of computing the scoring rules, it is assumed that the forecast errors are normally distributed with a constant variance.

4 Empirical Results

This section presents the empirical analysis in terms of in-sample fit and predictive power, in context of both the level of $\text{PM}_{2.5}$ and the occurrence of periods of extreme levels. Here, the performance of the DME model will be compared to the SARIMAX and ANN approaches.

4.1 Specification of the models

Three different periods are used for estimation and prediction, 2011–2013, 2012–2014 and 2013–2015. Within each, a period of 2 years and 3 months is used for model estimation, beginning January 1 at 1:00 A.M. and ending March 31 at 6:00 P.M. of the subsequent year. Then, the quality of the prediction is evaluated for 2013, 2014 and 2015, from March 31 at 7:00 P.M. to August 31 at 6:00 P.M. of each year. This stage is called critical episode management (GEC, in Spanish) because it is the period where the highest $\text{PM}_{2.5}$ concentration levels are recorded, (Perez & Gramsch, 2016), and where the government authorities take mitigation measures through environmental alert, pre-emergency or emergency, according to the levels defined in Section 1. The data used to forecast the following 24 hours ends at 6:00 P.M. every day. This is because the primary quality regulation of $\text{PM}_{2.5}$ demands that a critical episode of air pollution be reported between 8:00 P.M. and 9:00 P.M. on the day prior to its occurrence (MMA, 2011).

The prediction is made by the hour, re-estimating the model every 1 hour, 3 hours and 24 hours. For the 1 hour ahead forecast, the covariates are included a one-hour delay $j = 1$. For 3 and 24 hour forecast horizons, the covariates are included with $j = 24$ hours of delay to be consistent with the previously explained forecasting structure. This is because using

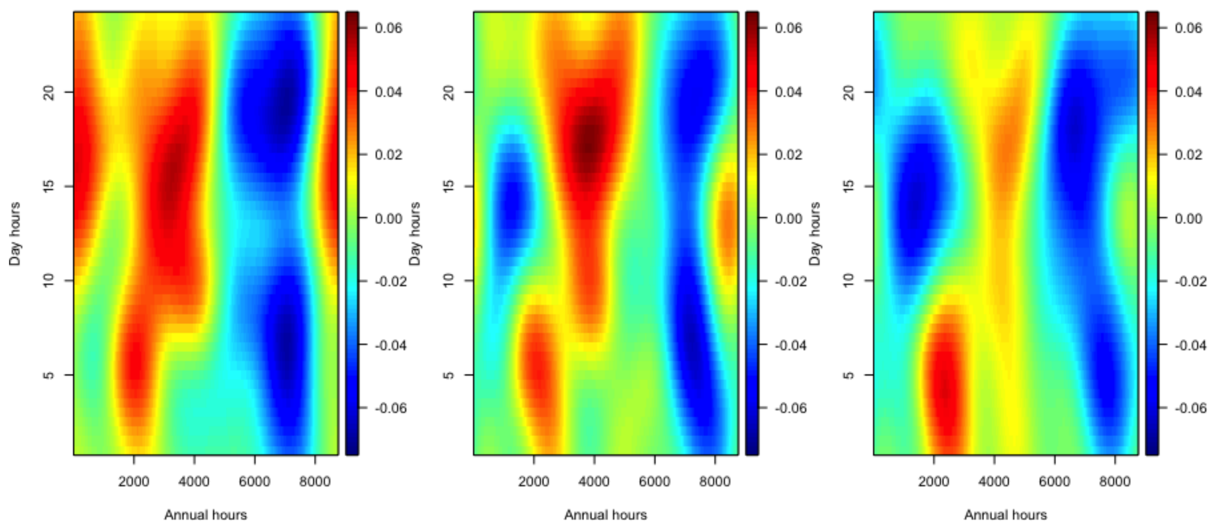

Figure 5: From the left panel, heat map plots showing for annual seasonality of $\mathrm{PM}_{2.5}$ series with 4 Fourier coefficients for the Pudahuel monitoring station in 2013, 2014 and 2015, respectively. The horizontal axis represents the total hours in a year, and the vertical axis represents the 24 hours in a day.

the same variable with a one-day delay provides greater explanatory power than including it at a 3-hour delay given the diurnal pattern discussed in Section 2. When the longer 3- and 24-hour forecasts are generated, 1-hour ahead predictions of $\mathrm{PM}_{2.5}$ are recursively constructed and used as lagged information in the longer forecasts.

4.2 In-sample fit of the models

The in-sample RMSE and MAE measures of fit were evaluated for each of the three specifications of the three models, across all 11 stations and time periods. A full set of these results are available in the online appendix. Overall, at a 1-hour horizon, Specification 2 produces the best in-sample fit in all three periods, on average across the 11 stations, with the differences falling moving out to the longer forecast horizons. Table 1 presents a summary of these results based on the best specification for each of the models. The DME offers superior in-sample fit relative to the SARIMAX and ANN approaches across all periods, loss functions and forecast horizons. Only in the case of 24-hour forecasts in 2015, do the DME and ANN models exhibit equal RMSE. The superiority of Specification 2 implies that the interaction between WS and WD acts effectively as a ventilation factor for each station, with their influence being important only if accompanied by the $\mathrm{PM}_{2.5}$ concentration of the

Model	2013				2014				2015			
	1 hour		24 hours		1 hour		24 hours		1 hour		24 hours	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
DME	7.88	4.97	8.38	5.28	7.86	5.02	8.37	5.31	8.16	5.20	8.63	5.47
SARIMAX	8.66	5.37	9.07	5.62	8.58	5.36	8.58	5.36	8.79	5.47	9.21	5.71
ANN	7.98	5.04	8.46	5.32	7.96	5.09	8.38	5.34	8.12	5.17	8.57	5.46

Table 1: Average in-sample fit for the 11 stations for the three proposed models. Columns exhibit the estimates for 1 and 24 hours in the three estimation years.

stations near the one being studied. This result is consistent with earlier studies such as Perez & Salini (2008), Jollois *et al.* (2014) and Perez & Gramsch (2016) where geographical proximity is found to help explain the relationship between concentrations at nearby stations.

The remaining discussion of in-sample fit is based on Specification 2 of the DME model. One of the main stylized facts described in Section 2 is the strong seasonal component present in different forms of persistence (daily, weekly and annually). First, using heat graphs, Figure 5 presents the results of the estimation of the Fourier series proposed in Eq. (3) to capture the annual cycle, based on the hourly $\text{PM}_{2.5}$ data at the Pudahuel station for 2013, 2014 and 2015, respectively. The horizontal axis corresponds to the 8,760 hourly observations in a year, and the vertical axis represents the 24 equations of the model, one for each hour of the day. Colors close to the blue end of the spectrum indicate lower estimates of the $\gamma_h(t)$ coefficient implying lower persistence in $\text{PM}_{2.5}$ concentration, whereas colors closer to red reflect an increase in persistence. Thus, the annual seasonal component is characterized for each hour of the day. Additionally, in the period between 2000 and 5000 hours, corresponding to the colder months of April to August when GEC is needed, increases in $\gamma_h(t)$ are observed, indicating higher persistence in the $\text{PM}_{2.5}$ concentration. Another important pattern is the time of day when the greatest persistence of $\gamma_h(t)$ is observed. In this case, the highest values for $\gamma_h(t)$ are seen around 9:00 A.M. and 5:00 P.M. approximately, the peak periods of vehicular traffic in the city.

In relation to the weekly component, Figure 6 shows the coefficients on the dummy variables according to the day of the week and the multiplicative variable according to the daily delay $PM_{h,d-1,t}$. These stay positive for most of the hourly periods, except between 1:00 A.M. and 4:00 A.M., where the influence of the daily delay is lower and even negative in some cases, a night-time period in which there is no major traffic or movement in the

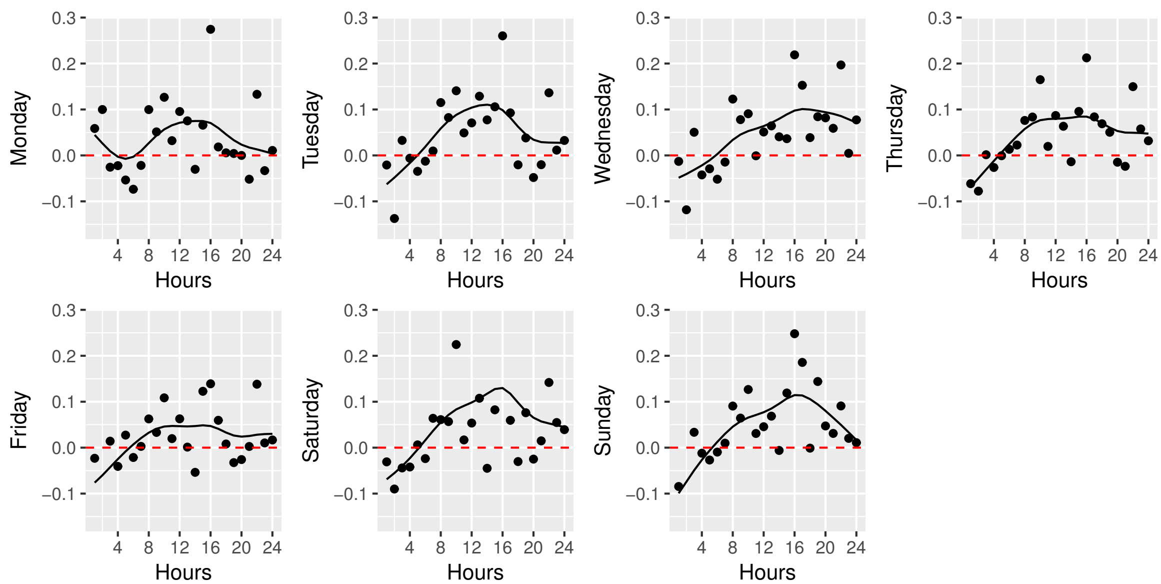

Figure 6: Dummy variables coefficients for each day and hour in the DME for the Pudahuel station in 2013. The horizontal axis represents the evaluated equation (24 equations, one per hour of the day) and the vertical axis the value of the coefficient. Graph smoothed on the basis of the average of observed values.

city. A smoothed fit to the coefficients is also shown to highlight the average value of these coefficients, revealing that the persistence is also a function of the time of the day.

To highlight the role played by the proposed covariates, Figure 7 shows the smoothed coefficients in graphical form, again using the Pudahuel monitoring station and the year 2013 for illustrative purposes. They show that on average, the coefficients on both the daily lag of $\text{PM}_{2.5}$ and concentrations at nearby stations, PM_c are positive. In addition, CO positively influences $\text{PM}_{2.5}$ on average, while the coefficients on $Temp$ and RH take negative values, consistent with the logic discussed in Section 2, cold days imply a greater use of wood for heating, while a lower percentage of RH means less dispersion of particulate matter. Figure 8 presents the coefficients of the interaction term of $WS \times WD$, according to north, south, east and west. The results reflect the average negative coefficients, which may be associated with the component of wind speed which shown in Section 2 reflects ventilation conditions. This effect is all the more important if it is moving in a southwesterly direction.

In relation to the overall fit obtained by the DME model, Figure 9 shows the average R^2 statistic for the 11 stations. The three colors represent 2013, 2014 and 2015, respectively. The upper limit of each band represents the 1-hour estimation and the lower limit the 24-

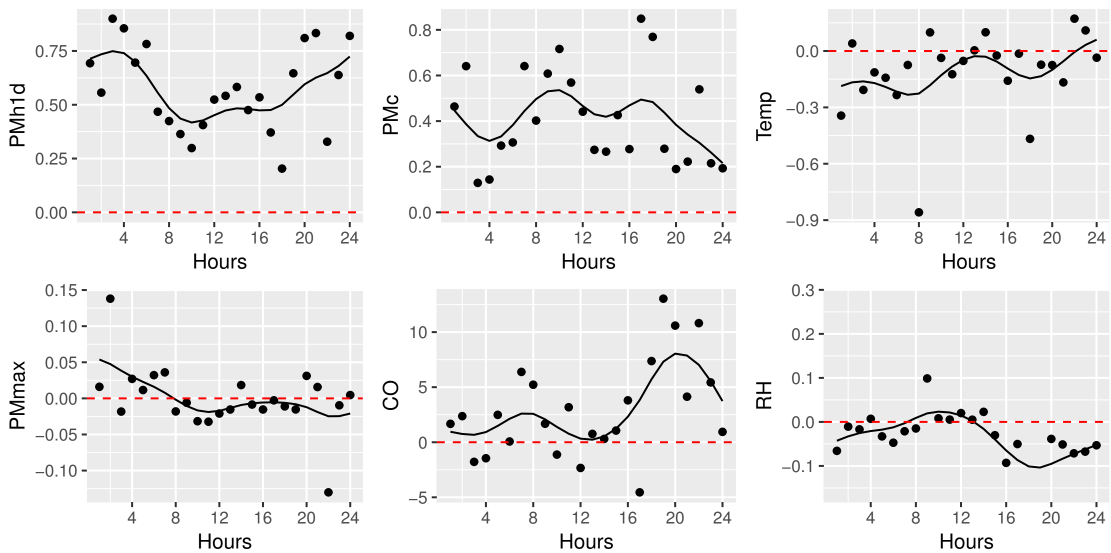

Figure 7: Coefficients corresponding to the covariates included for each hour in the DME for the Pudahuel station in 2013. The horizontal axis represents the evaluated equation (24 equations, one per hour of the day) and the vertical axis the value of the coefficient for each covariate. Graph smoothed on the basis of the average of observed values.

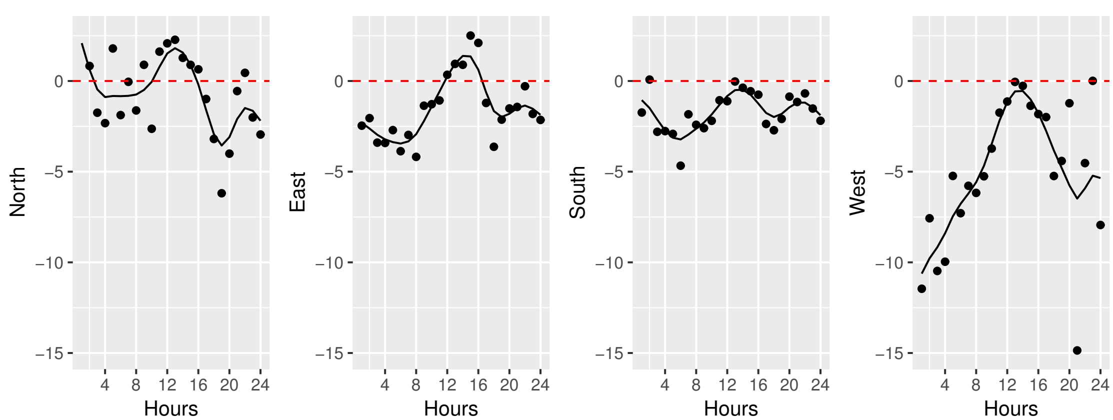

Figure 8: Coefficients on the interactive variable $WS \times WD$ according to north (N), south (S), east (E) and west (W) for each hour in the DME for the Pudahuel station in 2013. The horizontal axis represents the evaluated equation (24 equations, one per hour of the day) and the vertical axis the value of the coefficient. Graph smoothed on the basis of the average of observed values.

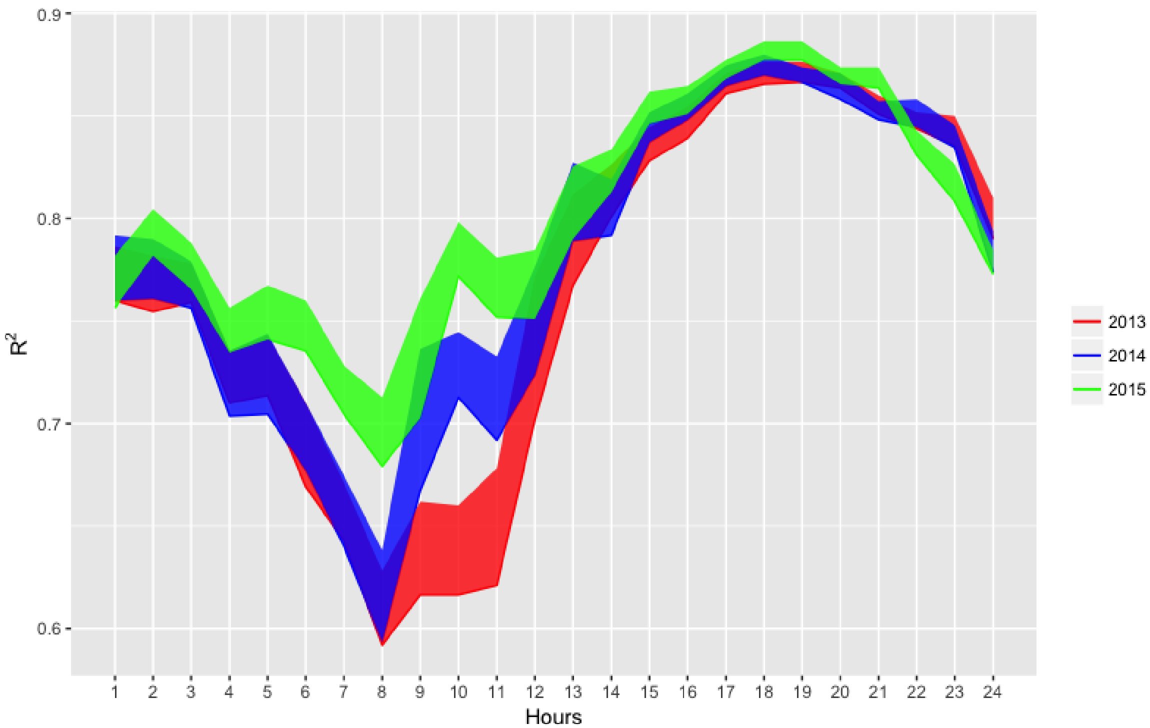

Figure 9: Average R^2 statistic for the 11 stations in 2013, 2014 and 2015 for each equation (one per hour). The upper limit of each band represents the 1-hour estimation and the lower limit the 24-hour estimation.

hour estimation. Note that between 5:00 and 10:00 A.M. the explanatory capacity of the model decreases, which is consistent with the greater dispersion of the series in that period (see Section 2). The peak in average R^2 for the 11 stations is at 5:00 P.M., with the model explaining up to 90% of the variance of the $\text{PM}_{2.5}$ in individual cases.

4.3 Prediction

In this section, the predictive accuracy of the proposed DME model is compared with its competitors. In the context of the DME and SARIMAX models, Specification 2 produces the most accurate forecasts, consistent with the in-sample results reported earlier. However under the ANN, Specification 1 described in Eq. (4) is preferred, a result at odds with the in-sample results. This demonstrates a possible problem of overfitting when using the ANN, an issue widely cited in the literature (Tetko *et al.*, 1995). This occurs when the supervised training algorithm, backpropagation here, memorizes the training set in such a way that when there are new observations and hence patterns in the data to which it

Model	1 hour		3 hours		24 hours	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
			2013			
DME	10.48	7.30	14.36	10.06	19.29	13.89
SARIMAX	10.87	7.58	15.18	10.56	23.33	16.49
ANN	11.64	7.36	16.24	10.08	20.57	13.51
			2014			
DME	11.04	7.45	15.14	10.45	21.75	15.80
SARIMAX	11.14	7.47	15.56	10.45	24.53	17.24
ANN	11.43	7.23	18.69	10.45	21.30	15.31
			2015			
DME	11.19	7.63	15.45	10.64	21.95	15.98
SARIMAX	11.28	7.61	15.95	10.77	25.23	18.00
ANN	14.50	7.74	21.17	10.77	23.04	16.27

Table 2: Out-of-sample fit for the three models evaluated (DME, SARIMAX and ANN) for 1, 3 and 24 hours in three prediction years. Values in bold indicate average for the 11 stations according to the category.

cannot recognize.

Table 2 reports a summary of out-of-sample forecast accuracy in terms of MAE and RMSE based on averages across the 11 stations. Again, a full set of results based on each station is available in the online appendix. Overall, irrespective of the model it is clearly possible to generate more accurate forecasts at the shorter 1-hour horizon, with relatively little differences between the precision of the forecasts over the 3- and 24-hour horizons. Even though the models are re-estimated every 3-hours, the covariates are included at a lag of 24-hours which results in the accuracy of both longer horizon forecasts being similar due to the diurnal patterns in the data, which is less of an issue for 1-hour forecasts that use 1-hour lags. On average across the 11 stations, at a 1-hour horizon the DME produces the most accurate forecasts in nearly all combinations of loss function and evaluation periods. At a 3-hour horizon, the DME models continues to dominate the others in most cases, somewhat more frequently under the RMSE loss function. Even at the 24-hour horizon, the DME provides more accurate forecasts in two of the three periods, 2013 and 2015, with the ANN and DME exhibiting similar performance in 2014.

Although the previous results provide a preliminary view of the relative forecast accuracy, they do not reveal whether the performance of the models are significantly different. Therefore, tests of unconditional and conditional predictability, the DM and GW tests are performed, respectively (Diebold & Mariano 1995; Giacomini & White 2006). Table 3 shows the p-values of the GW test statistic in every period, emphasizing in bold the values that conclude that the DME approach produces significantly more accurate forecasts than the alternative model, SARIMAX or ANN. As the GW test offers more power, results based of the DM test (they reveal the same patterns as the GW test) are reported in the online appendix to economise on space.

It is clear in Table 3 that the DME produces significantly more accurate forecasts than the SARIMAX model for most stations and periods, with the superior performance of the DME more pronounced at the longer 24-hour horizon. At the shorter horizons of 1- and 3-hours, while there are a number of cases where the DME is superior to ANN, in many cases there are no significant differences in forecast accuracy. Unreported results (when the direction of the DM and GW tests are reversed to identify if the ANN is significantly more accurate than the DME) show that there are very few instances where the ANN models produces significantly more accurate forecasts. Overall, these results indicate that the DME model produces forecasts that are at least as accurate as the more complex ANN model, and in many instance are significantly more accurate even if they are produced from a set of simpler linear regressions.

4.4 Analysis of critical episodes

Given the potential impact of periods of extreme pollution, this section considers how well the DME predict episodes of alert, pre-emergency and emergency, based again on Specification 2. Even though extreme levels are observed only on a small number of days, the practice of analyzing critical episodes while discarding periods of low air pollution levels appears to be a natural approach. Table 4 shows the success rate, in the form of percentage of episodes correctly predicted during the GEC periods, along with number of episodes recorded. The average for the 11 stations (in the final column headed by $\bar{X}\,EST$) indicates that the proposed model achieves a 76% success rate in alert episodes and 100% in pre-emergency and emergency episodes for 2013. The lowest average success rate for the three critical episodes is observed in 2015, 53% for alert and 91% for emergency episodes, a year that exhibited the highest average $PM_{2.5}$ at 46.19, and greatest standard deviation of 29.82.

In terms of analysis by individual stations, Las Condes station has the lowest average

Models	Station	1 hour			3 hours			24 hours		
		2013	2014	2015	2013	2014	2015	2013	2014	2015
SARIMAX vs. DME	Cerrillos	0.36	1.00	0.93	**0.00**	**0.05**	0.46	**0.00**	0.07	0.06
	Independencia	**0.00**	0.07	**0.00**	**0.03**	1.00	0.99	**0.00**	0.27	**0.01**
	Las Condes	**0.00**	**0.02**	**0.04**	**0.00**	0.98	**0.00**	**0.00**	**0.00**	**0.00**
	El Bosque	**0.00**	0.57	0.67	0.00	0.18	**0.04**	**0.00**	**0.00**	**0.00**
	Parque O'Higgins	**0.00**	0.93	1.00	**0.00**	**0.00**	0.23	**0.00**	0.11	0.06
	Talagante	0.32	0.68	1.00	**0.00**	0.39	0.70	**0.00**	**0.01**	**0.00**
	Quilicura	0.06	0.24	0.88	**0.05**	0.93	0.44	**0.00**	**0.02**	**0.00**
	Pudahuel	0.23	0.74	**0.05**	0.42	0.70	**0.00**	**0.00**	0.12	**0.00**
	Cerro Navia	0.21	0.26	0.89	0.04	0.10	**0.05**	**0.00**	**0.02**	**0.00**
	La Florida	**0.00**	**0.00**	0.11	**0.00**	0.14	0.13	**0.00**	**0.00**	**0.00**
	Puente Alto	**0.00**	0.13	0.44	**0.00**	0.99	0.45	**0.00**	0.12	**0.00**
ANN vs. DME	Cerrillos	0.47	**0.00**	**0.00**	0.15	0.18	0.28	0.16	0.79	**0.00**
	Independencia	1.00	0.96	1.00	0.17	0.97	0.95	0.11	0.81	0.30
	Las Condes	0.21	0.11	0.08	0.28	0.46	0.11	0.11	0.09	**0.04**
	El Bosque	**0.01**	0.51	0.19	0.65	0.52	0.53	0.45	0.08	0.26
	Parque O'Higgins	**0.00**	1.00	0.11	0.61	1.00	0.28	**0.03**	0.86	0.83
	Talagante	0.88	1.00	0.96	0.24	0.87	1.00	**0.00**	**0.00**	0.99
	Quilicura	0.13	**0.00**	**0.01**	0.80	0.23	0.24	**0.03**	0.48	0.12
	Pudahuel	0.43	0.91	0.37	0.98	0.83	0.08	**0.05**	0.46	0.07
	Cerro Navia	0.94	0.66	**0.04**	0.90	0.35	0.28	0.13	0.49	0.06
	La Florida	0.08	0.99	0.17	0.76	0.26	0.65	0.59	**0.00**	0.92
	Puente Alto	0.29	0.74	0.97	0.81	0.99	0.87	0.57	0.07	0.19

Table 3: p-values for the GW Test at 11 monitoring stations for three forecasting years (2012, 2013 and 2014) and in three re-estimation periods: prediction every 1 hour, 3 hours and 24 hours. Values in bold indicate p-values lower than 0.05.

Year	GEC (% Success)	CERRILLOS	INDEPENDENCIA	LAS CONDES	EL BOSQUE	PARQUE O'HIGGINS	TALAGANTE	QUILICURA	PUDAHUEL	CERRO NAVIA	LA FLORIDA	PUENTE ALTO	$\overline{X}$ EST
	Alert	50%	100%	100%	20%	100%	100%	100%	57%	60%	100%	50%	76%
		2	**0**	**0**	**5**	**0**	**0**	**0**	**7**	**15**	**0**	**2**	**2.82**
	Pre-Emergency	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
		0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	Emergency	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
		0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
2013	$\overline{X}$ $PM_{2.5}$	36.40	33.04	25.53	40.30	36.48	30.26	33.55	36.15	40.65	33.59	43.28	35.38
	Std. Dev. $PM_{2.5}$	23.59	15.77	15.66	27.71	20.99	22.06	20.78	28.08	32.04	20.37	26.46	23.05
	Alert	50%	67%	100%	30%	100%	100%	43%	67%	29%	13%	50%	59%
		14	**3**	**0**	**20**	**0**	**0**	**14**	**24**	**24**	**8**	**4**	**10.09**
	Pre-Emergency	0%	100%	100%	60%	100%	100%	0%	45%	58%	100%	0%	60%
		4	**0**	**0**	**5**	**0**	**0**	**1**	**11**	**19**	**1**	**1**	**3.82**
	Emergency	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	91%
		0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**0**	**0**	**0.18**
2014	$\overline{X}$ $PM_{2.5}$	41.13	39.79	30.04	48.61	32.14	26.74	43.21	46.71	53.84	43.55	36.39	40.19
	Std. Dev. $PM_{2.5}$	30.29	20.74	20.54	34.62	19.96	20.19	28.64	36.47	42.08	26.87	24.76	27.74
	Alert	61%	60%	100%	35%	65%	53%	37%	55%	44%	38%	40%	53%
		36	**15**	**0**	**34**	**23**	**17**	**27**	**38**	**34**	**16**	**5**	**22.27**
	Pre-Emergency	67%	100%	100%	100%	100%	100%	50%	75%	50%	100%	100%	86%
		3	**0**	**0**	**1**	**0**	**1**	**2**	**12**	**14**	**0**	**0**	**3**
	Emergency	100%	100%	100%	100%	100%	100%	100%	50%	50%	100%	100%	91%
		0	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**2**	**0**	**0**	**0.36**
2015	$\overline{X}$ $PM_{2.5}$	49.94	45.16	33.73	50.86	46.78	41.21	46.15	54.81	52.77	46.40	40.28	46.19
	Std. Dev. $PM_{2.5}$	32.58	23.78	20.49	33.24	26.68	29.56	30.23	39.97	41.79	26.49	23.23	29.82

Table 4: Critical episode management for the three prediction periods: March 31 to August 30, 2013, 2014 and 2015, respectively. The results found for each station analyzed in the study. The percentage in the rows corresponds to the success rate of the DME model in effectively forecasting the different critical episodes. The values in bold are the real number of critical episodes in their different categories. Std. Dev. is the standard deviation of $PM_{2.5}$. $\overline{X}$ EST corresponds to the average observed for all the stations in relation to each of the previously described statistics.

PM$_{2.5}$ and the least variability compared to the other stations and is the only station to have not experienced an alert (or above) episode. This is due to the fact that Las Condes station is located at a higher altitude than other communes thus having the benefit of better atmospheric ventilation. In this case, the DME has never predicted a critical episode of PM$_{2.5}$ concentration. In contrast, Cerro Navia shows the lowest success rate on average over the three years, with 44%, 69% and 50% of alerts, pre-emergencies, emergencies, correctly predicted respectively. This is because this station exhibits the worst contamination rates, mainly due to the high consumption of firewood in this commune and the high population density. Moreover, this station exhibits the highest dispersion in terms of the standard deviation of the PM$_{2.5}$ concentration (average 53.84 in 2014), which seems to reflect the large differences between levels of pollution between winter and other seasons of the year. In fact, in 2014 firewood accounted for 45% of all sources of pollutants, while in the winter period its contribution increased to 70%.

Table 5 shows the estimates of mean CRPS and twCRPS scoring rules for the three models during the forecasting period, from April to August, 2015. In other periods the results were very similar. As expected, all models produce less accurate forecasts according to both the unrestrictive (CRPS) and restrictive (twCRPS) scoring rules as the forecast horizon increases. The results in terms of mean CRPS (relating to the full distribution) across the horizons of 1-, 3- and 24-hours are now compared. Overall, the DME model exhibits the best predictive performance in the vast majority of cases, as indicated by the results i the bottom row of Table 5 tht report the number of stations at at which th DME is the most accurate prediction. In relation to the critical episodes, we restrict our attention to the twCRPS results. At the 1- and 3-hour horizons, the DME produces the most accurate forecasts at the three different extreme thresholds, followed by the ANN specification. While at a 24-hour horizon, the DME is also considered best, although the second best approach seems to be SARIMAX. Even when focusing on the more extreme critical episodes the DME remains the most accurate forecast in the majority of cases. Again the DM test is used to differentiate between the forecasts based on their performance under the probability scoring rules. A $*$ or $\dagger$ is used in Table 5 to indicate whether the DME is significantly more accurate than the SARIMAX or ANN models respectively. At the 1- and 3-hour horizons, the DME approach is a significantly more accurate forecast than the either the SARIMAX or ANN in many cases. At the 24-hour horizon, the DME is significantly more accurate than its competitors under CRPS in most cases. However, at the extreme quantiles the differences are less pronounced, the DME is statistically superior to the ANN in nearly half the cases (of equal predictive accuracy in the others) and only superior to the SARIMAX in a small number of instances.

Overall, the DME model, based on a simple system of linear regressions, is shown to produce forecasts that are as least as accurate, and in many cases more accurate than a number of common competitors. Beyond its relative forecast performance, the ease with which the coefficients can be interpreted is beneficial as the impact of a range of exogenous covariates can easily be examined.

Acknowledgements

Herrera acknowledges the Chilean CONICYT funding agency for financial support (FONDECYT 1150349) for this project.

5 Conclusions

Air pollution is a major environmental, health and economic issue in many large urban areas around the world. In Santiago, Chile, this issue is exacerbated by it unique geographical location in the Central Valley nestled between the Andes to the west and a smaller range to the east. Given the negative impacts of air pollution, and in particular $PM_{2.5}$, much research attention has been paid to developing predictive models.

This paper developed a multiple linear equation model (DME) with dynamic coefficients for the purposes of forecasting $PM_{2.5}$ in Santiago. The model is structured with an linear equation for each hour of the day, with dynamic coefficients using an annual Fourier component to capture the annual cycle, well as dummy variables to capture the day of the week effect. The advantages of this approach lie in the model being linear, meaning that it is less susceptible to overfitting issues associated with nonlinear models such as ANNs, and the coefficients can easily be interpreted. A forecasting exercise has demonstrated that the proposed multiple equation approach is a competitive forecasting alternative to the two alternative models, ANN and SARIMAX, which are often applied in the literature.

Station	Model	1 hour				3 hour				24 hour			
		CRPS	$twCRPS_{0.95}$	$twCRPS_{0.99}$	$twCRPS_{0.999}$	CRPS	$twCRPS_{0.95}$	$twCRPS_{0.99}$	$twCRPS_{0.999}$	CRPS	$twCRPS_{0.95}$	$twCRPS_{0.99}$	$twCRPS_{0.999}$
Cerrillos	DME	**6.323**†	0.875	0.414	0.169	**8.996***†	**1.286***	**0.610***	**0.247***†	**12.886***†	**1.765***	**0.791**†	**0.285**†
	SARIMA	6.343	0.925	0.461	0.211	9.315	1.380	0.688	0.309	13.560	1.817	0.815	0.302
	ANN	7.055	**0.868**	**0.389**	**0.153**	9.655	1.481	0.813	0.473	14.199	1.910	0.870	0.333
Independencia	DME	**4.587***†	**0.597***	0.271	0.103	6.223	**0.828**	**0.369**	**0.132**†	**9.031***†	**1.154**†	**0.500**†	**0.175**†
	SARIMA	4.848	0.698	0.338	0.144	**6.209**	0.831	0.386	0.150	10.503	1.209	0.507	0.165
	ANN	5.549	0.543	**0.209**	**0.071**	6.335	0.909	0.465	0.238	9.406	1.311	0.621	0.252
Las Condes	DME	**4.520***†	**0.560***	0.254	0.102	**5.733***†	**0.738***	0.333	0.131	**8.045***†	**0.965***†	**0.396**†	**0.124**†
	SARIMA	4.729	0.692	0.356	0.174	6.151	0.872	0.436	0.203	10.791	1.057	0.411	0.125
	ANN	5.957	0.558	**0.224**	**0.096**	6.720	0.654	**0.248**	**0.080**	8.545	1.036	0.437	0.145
El Bosque	DME	**6.908***†	**0.889***†	**0.426***†	**0.192***†	**9.588***†	**1.328***†	**0.633***†	**0.269***†	**13.126***†	**1.749***†	**0.787***†	**0.297***†
	SARIMA	7.079	0.989	0.505	0.248	10.484	1.593	0.850	0.432	15.285	1.953	0.906	0.376
	ANN	7.240	1.006	0.511	0.252	10.141	1.534	0.834	0.451	13.638	1.892	0.933	0.430
Parque O'Higgins	DME	5.058	**0.707***†	0.341	0.145	**7.492***†	**1.050***	**0.483***	**0.184***	**11.101***	1.557	0.719	0.277
	SARIMA	**5.012**	0.744	0.382	0.183	7.796	1.110	0.539	0.231	11.659	**1.519**	**0.676**	**0.251**
	ANN	12.229	0.839	**0.228**	**0.037**	8.019	1.065	0.505	0.225	11.219	1.542	0.708	0.278
Talagante	DME	**6.264***†	**0.938***†	**0.458***†	**0.194***†	**8.080***	**1.197***†	**0.570***†	**0.230***†	**11.382***	**1.560***	**0.690***	**0.243***
	SARIMA	6.379	1.083	0.592	0.302	8.511	1.449	0.787	0.396	12.527	1.708	0.780	0.299
	ANN	6.403	1.033	0.555	0.276	8.046	1.286	0.671	0.319	11.317	1.503	0.689	0.268
Quilicura	DME	**5.769**†	**0.819***†	**0.403***†	**0.182***†	**7.959***†	**1.156***†	**0.558***†	**0.234***†	**11.447***†	1.672	0.805	0.331
	SARIMA	5.832	0.889	0.472	0.244	8.284	1.280	0.706	0.390	12.737	**1.611**	**0.746**	**0.309**
	ANN	6.116	0.902	0.479	0.253	8.418	1.309	0.725	0.403	12.047	1.860	0.968	0.465
Pudahuel	DME	**6.188***†	**0.849***	**0.410***	**0.176***	8.716	1.220	0.581	0.237	**14.249***†	**2.197***	**1.089**	**0.464**†
	SARIMA	6.424	0.930	0.482	0.239	**7.796**	**1.110**	**0.539**	**0.231**	15.984	2.319	1.153	0.516
	ANN	6.471	0.863	0.437	0.220	18.912	2.007	1.062	0.750	15.208	2.242	1.154	0.549
Cerro Navia	DME	**5.910**†	**0.780**	**0.375***	**0.164***	**8.633***†	**1.212***	**0.582***	**0.247***†	**14.598***†	**2.204**†	**1.071**†	**0.440**†
	SARIMA	5.990	0.828	0.425	0.214	9.297	1.378	0.736	0.389	16.328	2.237	1.076	0.461
	ANN	9.572	1.183	0.689	0.510	9.190	1.256	0.641	0.324	16.961	2.732	1.607	1.005
La Florida	DME	**5.107***†	**0.674***†	**0.314***†	**0.127***†	**7.158***†	**0.979***†	**0.459***†	**0.187***†	10.702	1.441	0.651	0.242
	SARIMA	5.396	0.828	0.440	0.223	7.646	1.126	0.581	0.287	13.637	**1.369**	**0.540**	**0.171**
	ANN	5.373	0.734	0.363	0.168	7.460	1.073	0.551	0.271	**10.634**	1.473	0.695	0.288
Puente Alto	DME	**5.550***	**0.680***†	**0.303***†	**0.120***†	**7.262***†	**0.956***†	**0.430***†	**0.165***†	**9.956***†	1.324	0.581	0.205
	SARIMA	5.748	0.789	0.384	0.173	7.613	1.044	0.502	0.221	11.437	**1.302**	**0.545**	**0.185**
	ANN	5.629	0.759	0.366	0.164	7.432	1.028	0.500	0.223	10.116	1.332	0.602	0.234
	#Forecast	10	10	7	7	9	10	9	9	10	7	7	7

Table 5: Results for the probability scoring rules, *CRPS* and *twCRPS* (at 0.95, 0.99 and 0.99 quantiles) for each station for the critical episode management prediction periods: March 31 to August 30, 2015. For each station, the most accurate model is denoted in bold. The number of station where the DME is the most accurate forecast is shown the bottom row. The symbols $*$ and $\dagger$ indicate when the Diebold Mariano test is rejected at 5% meaning the DME is significantly more accurate than the SARIMAX or ANN models respectively.

The DME can adequately capture the seasonality in $PM_{2.5}$, and surpasses the ANN and SARIMAX competitors in most cases, in terms of both in-sample and out-of-sample performance. Performance relating to both point and distributional predictions were analyzed across a number of horizons, with day ahead forecasts, and in particular, extreme events of particular importance. Based on such forecasts, government authorities can take prompt strategic measures in response to the forecasts of critical $PM_{2.5}$ episodes to restrict emissions of this pollutant.

This research can be extended in a number of different directions including more complex multivariate models to take into account spatial interactions, forecasts of covariates, and other information such as traffic flows. This methodology could also be applied to air quality forecasting in other cities and countries where the time series exhibit broadly with similar characteristics.

References

Amisano, G. & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* **25**(2), 177–190.

Böhringer, C. & Jochem, P. (2007). Measuring the immeasurable: A survey of sustainability indices. *Ecological Economics* **63**(1), 1–8.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley and Sons.

Cassmassi, J. (1999). Improvement of the forecast of air quality and of the knowledge of the local meteorological conditions in the metropolitan region. Tech. rep.

Chaloulakou, A., Kassomenos, P., Spyrellis, N., Demokritou, P. & Koutrakis, P. (2003). Measurements of PM 10 and PM 2.5 particle concentrations in Athens, Greece. *Atmospheric Environment* **37**(5), 649–660.

Cifuentes, L. (2010). Relación de la Norma de Calidad Primaria PM2.5 con la Norma de Calidad Primaria de PM10. Tech. rep., Comision Nacional del Medio Ambiente, Region Metropolitana de Santiago.

Clements, A., Hurn, A. & Li, Z. (2016). Forecasting day-ahead electricity load using a multiple equation time series approach. *European Journal of Operational Research* **251**(2), 522–530.

CONAMA (2005). Sostenibilidad urbana en la comuna de santiago de chile. Tech. rep., Observatorio de Sostenibilidad, Comisión Nacional del Medio Ambiente.

Delgado, M., Sanchez, F., Zapata, C., Escobar, R. & Goday, M. (2006). Intelligent analysis to the Contamination in the city of Santiago from Chile. In: *Advances in Systems, Computing Sciences and Software Engineering*. Springer, pp. 345–352.

Díaz-Robles, L., Ortega, J., Fu, J., Reed, G., Chow, J., Watson, J. & Moncada, J. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment* **42**(35), 8331–8340.

Diebold, F. & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**, 253–263.

Djalalova, I., Delle Monache, L. & Wilczak, J. (2015). PM2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmospheric Environment* **119**, 431–442.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L. & Wang, J. (2015). Artificial neural networks forecasting of PM 2.5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment* **107**, 118–128.

Franses, P. H. & Draisma, G. (1997). Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics* **81**(1), 273–280.

Garreaud, R. & Rutllant, J. (2006). Factores meteorológicos de la contaminación atmosférica en Santiago. *Contaminación Atmosférica Urbana: Episodios Críticos de Contaminación Ambiental en la Ciudad de Santiago* , 36–53.

Genc, D., Yesilyurt, C. & Tuncel, G. (2010). Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. *Environmental Monitoring and Assessment* **166**(1-4), 11–27.

GIACOMINI, R. & WHITE, H. (2006). Tests of conditional predictive ability. *Econometrica* **74**(6), 1545–1578.

GNEITING, T. & RANJAN, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics* **29**(3), 411–422.

HAMZAÇEBI, C., AKAY, D. & KUTAY, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications* **36**(2), 3839–3844.

HIEN, P., BAC, V., THAM, H., NHAN, D. & VINH, L. (2002). Influence of meteorological conditions on PM 2.5 and PM 2.5- 10 concentrations during the monsoon season in Hanoi, Vietnam. *Atmospheric Environment* **36**(21), 3473–3484.

HILL, T., O'CONNOR, M. & REMUS, W. (1996). Neural network models for time series forecasts. *Management Science* **42**(7), 1082–1092.

HORAN, J. & FINN, D. (2008). Sensitivity of air change rates in a naturally ventilated atrium space subject to variations in external wind speed and direction. *Energy and Buildings* **40**(8), 1577–1585.

INE (2007). División Politico-Administrativa y Censal, 2007. Publicación Bi-anual. Tech. rep., Instituto Nacional de Estadisticas.

JIAN, L., ZHAO, Y., ZHU, Y., ZHANG, M. & BERTOLATTI, D. (2012). An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. *Science of the Total Environment* **426**, 336–345.

JOLLOIS, F. X., POGGI, J. M. & PORTIER, B. (2014). Three non-linear statistical methods for analyzing PM10 pollution in rouen area. *Case Studies In Business, Industry And Government Statistics* **3(1)**, 1–17.

KELLY, F. & FUSSELL, J. (2012). Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmospheric Environment* **60**, 504–526.

KUKKONEN, J., PARTANEN, L., KARPPINEN, A., RUUSKANEN, J., JUNNINEN, H. *et al.* (2003). Extensive evaluation of neural network models for the prediction of NO 2 and PM 10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment* **37**(32), 4539–4550.

KURT, A. & OKTAY, A. (2010). Forecasting air pollutant indicator levels with geographic models 3days in advance using neural networks. *Expert Systems with Applications* **37**(12), 7986–7992.

LU, W. & WANG, W. (2005). Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere* **59**(5), 693–701.

MCKENDRY, I. (2002). Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. *Journal of the Air and Waste Management Association* **52**(9), 1096–1101.

MMA (2011). Norma primaria de calidad ambiental para material particulado fino respirable MP2,5. Reporte Oficial. Tech. rep., Ministerio del Medio Ambiente.

MMA (2012). Contribución a emisiones primarias de MP2.5 para la Región Metropolitana. Tech. rep., Ministerio del Medio Ambiente. Centro Mario Molina.

NAEHER, L. P., SMITH, K. R., LEADERER, B. P., NEUFELD, L. & MAGE, D. T. (2001). Carbon monoxide as a tracer for assessing exposures to particulate matter in wood and gas cookstove households of highland Guatemala. *Environmental Science and Technology.* **35(3)**, 575–581.

ORDIERES, J., VERGARA, E., CAPUZ, R. & SALAZAR, R. (2005). Neural network prediction model for fine particulate matter on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Modelling and Software* **20**(5), 547–559.

OSOWSKI, S. & GARANTY, K. (2007). Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Applications of Artificial Intelligence* **20**(6), 745–755.

PEREZ, P. & GRAMSCH, E. (2016). Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. *Atmospheric Environment* **124**, 22–27.

Perez, P. & Reyes, J. (2006). An integrated neural network model for PM10 forecasting. *Atmospheric Environment* **40**(16), 2845–2851.

Perez, P. & Salini, G. (2008). PM 2.5 forecasting in a large city: Comparison of three methods. *Atmospheric Environment* **42(35)**, 8219–8224.

Prakash, A., Kumar, U., Kumar, K. & Jain, V. (2011). A wavelet-based neural network model to predict ambient air pollutants concentration. *Environmental Modeling and Assessment* **16**(5), 503–517.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* **323(6088)**, 533–536.

Sahu, S. K. & Mardia, K. V. (2005). A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(1), 223–244.

Saide, P., Carmichael, G., Spak, S., Gallardo, L., Osses, A., Mena, M. & Pagowski, M. (2011). Forecasting urban PM10 and PM2. 5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model. *Atmospheric Environment* **45**(16), 2769–2780.

Saide, P., Mena, M., Tolvett, S., Hernandez, P. & Carmichael, G. (2016). Air quality forecasting for winter-time PM2.5 episodes occurring in multiple cities in central and southern Chile. *Journal of Geophysical Research: Atmospheres* **121**(1), 558–575.

Shah, S., Cocker, D., Miller, J. & Norbeck, J. (2004). Emission rates of particulate matter and elemental and organic carbon from in-use diesel engines. *Environmental Science and Technology* **38**(9), 2544–2550.

SINIA (2010). Análisis del impacto económico y social de la norma de calidad primaria de material particulado 2.5. Tech. rep., Sistema Nacional de Información Ambiental.

Spliid, H. (1983). A fast estimation method for the vector autoregressive moving average model with exogenous variables. *Journal of the American Statistical Association* **78**, 843–849.

Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W. & Liu, S. (2013). Prediction of 24-hour-average PM 2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. *Science of the Total Environment* **443**, 93–103.

Tang, Z. & Fishwick, P. A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing* **5(4)**, 374–385.

Tetko, I., Livingstone, D. & Luik, A. (1995). Comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Sciences* **35(5)**, 826–833.

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S. & Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. *Science of the Total Environment* **409**(8), 1559–1571.

Wang, Y., Chung, A. & Paulson, S. (2010). The effect of metal salts on quantification of elemental and organic carbon in diesel exhaust particles using thermal-optical evolved gas analysis. *Atmospheric Chemistry and Physics* **10**(23), 11447–11457.

Weizhen, H., Zhengqiang, L., Yuhuan, Z., Hua, X., Ying, Z., Kaitao, L. & Yan, M. (2014). Using support vector regression to predict PM10 and PM2.5. In: *IOP Conference Series: Earth and Environmental Science*, vol. 17.

WHO (2011). Urban outdoor air pollution database. Tech. rep., Department of Public Health and Environment.

Winkler, R. L., Munoz, J., Cervera, J. L., Bernardo, J. M., Blattenberger, G., Kadane, J. B., Lindley, D. V., Murphy, A. H., Oliver, R. M. & Ríos-Insua, D. (1996). Scoring rules and the evaluation of probabilities. *Test* **5**(1), 1–60.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. & Baklanov, A. (2012). Real-time air quality forecasting, part i: History, techniques, and current status. *Atmospheric Environment* **60**, 632–655.

Zhou, Q., Jiang, H., Wang, J. & Zhou, J. (2014). A hybrid model for pm 2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Science of the Total Environment* **496**, 264–274.