

Bayesian spatio-temporal modelling for estimating short-term exposure to air pollution in Santiago de Chile

Orietta Nicolis*, Mailiu Díaz, Sujit K Sahu, and Julio C. Marín

April 12, 2019

Abstract

Spatial prediction of exposure to air pollution in a large city such as Santiago de Chile is a challenging problem because of the lack of a dense air-quality monitoring network. Statistical spatio-temporal models exploit the space-time correlation in the pollution data and other relevant meteorological and land-use information to generate accurate predictions in both space and time. In this paper, we develop a Bayesian modelling method to accurately predict hourly $PM_{2.5}$ concentration in a one kilometer high resolution grid covering the city. The modelling method combines a spatio-temporal land-use regression model for $PM_{2.5}$ and a Bayesian calibration model for the input meteorological variables used in the land-use regression model. Using a 3-month winter-time pollution data set, the output of sample validation results obtained in this paper show a substantial increase in accuracy due to the incorporation of the linear calibration model. The proposed Bayesian modelling method is then used to provide short-term spatio-temporal predictions of $PM_{2.5}$ concentrations on a fine (one kilometer square) spatial grid covering the city. Along with the paper we publish the R code used and the output of sample predictions for future scientific use.

Keywords: spatio-temporal modelling; $PM_{2.5}$ pollution; WRF model; forecasting.

1 Introduction

Atmospheric air pollution from anthropogenic sources is of great concern due to its established harmful effects on human health [19, 8]. Such health concerns are exacerbated in all mega-cities around the world due to the large volume of human activities involved, such as transportation and energy usage.

Particulate matter with diameters less than $2.5 \mu m$ ($PM_{2.5}$), are tiny enough to penetrate the lungs and pulmonary alveoli, increasing the risk of premature mortality due to

*Faculty of Engineering, Andres Bello University, Calle Quillota, 980, Viña del Mar, 2520000, Chile, E-mail: orietta.nicolis@unab.cl

cardiopulmonary effects. Main sources of $\text{PM}_{2.5}$ pollution are cars, thermoelectric power plants, industrial and metallurgic processes, mining industry, and residential burning of coal, wood and kerosene.

Santiago de Chile (SCL) is affected by severe air pollution episodes due to its unique geographical location and climate. The city is located in a land depression called Santiago basin, which is surrounded by four main mountain ranges. Because of this, ventilation and air circulation are prevented efficiently, causing the pollutants to be contained inside the basin for long periods of time. In addition, air pollution is vertically spread over a very shallow layer of just several hundred meters because of large-scale subsidence from the south Pacific anticyclone located along the Chilean coast. These two factors work together in a nexus to produce very high levels of $\text{PM}_{2.5}$ in the city throughout the year, especially in winter months from June to August [23, 16, 45, 22], which are characterized by cold temperatures, high humidity, and low pressures. The main purpose of this paper is to estimate and predict hourly $\text{PM}_{2.5}$ concentration levels at any given location within the city.

Estimation of air pollution exposure in both space and time is a challenge as often there are only few active monitors which record air pollution continuously without fail for long periods of time. For example, in SCL there are only eleven active monitoring stations. Hence, sound statistical methods are needed to estimate and forecast pollution levels in areas away from the monitoring sites. There has been much recent activity in methodological developments in this area. Generic models for analyzing spatio-temporal data were developed by [6, 26, 24, 54, 50, 4, 7, 3, 30].

Space-time statistical modelling for short-term prediction of PM_{10} (particulate matter less than $10\mu\text{m}$ in diameter) and $\text{PM}_{2.5}$ has been proposed by [56, 46, 48, 40] and [38]. [10] and [41] proposed hierarchical space time models for the calibration of PM_{10} at heterogeneous monitoring networks. [42, 39], and [37] developed dynamical models for predicting ozone levels. Recently, [1] and [25] developed R packages for the Bayesian space time modelling of pollution data.

Several modelling efforts for estimating air pollution exposure in SCL are based on neural network models ([33]; [9], [32]). In particular, [33] developed an integrated artificial neural network model to forecast the maximum daily mean value of PM_{10} concentration at five monitoring stations in SCL, and [32] used a feed forward neural network to forecast critical episodes of $\text{PM}_{2.5}$ in the monitoring station with the highest concentration values (Cerro Navia) using historical hourly values of PM_{10} and $\text{PM}_{2.5}$ concentrations from nearby stations, weather variables, and a ventilation factor. [9] proposed a hybrid model combining autoregressive integrated moving average (ARIMA) and artificial neural network structures, for predicting the PM_{10} concentrations at Temuco city monitoring station, in Chile in 2006. [43] and [44] proposed a deterministic chemical based forecasting model for $\text{PM}_{2.5}$ using its high correlation with carbon monoxide as a tracer to predict critical night episodes.

A multivariate approach has been considered by [29] where they proposed a methodology based on a system of dynamic multiple linear equations that incorporates hourly, daily and seasonal characteristics to predict hourly $\text{PM}_{2.5}$ concentrations for eleven meteorological stations in SCL.

Air pollution exposure is highly affected by the prevailing weather and as a result, meteorological variables often highly correlate with air pollution levels, see e.g. [39, 55]. It is then natural to exploit these high correlations for modelling and predicting air pollution levels. Similar to air pollution concentrations, meteorological variables are only monitored at a handful of sites. However, regional weather forecasts can provide the meteorological information at locations where observations are not available. The Weather Research and Forecasting (WRF) model is a regional numerical weather prediction system that is used both for atmospheric operational forecasting and research. The model is freely available, flexible and computationally efficient, and offers advances in physics, numerics, and data assimilation contributed by the research community. It provides high-resolution forecasts using detailed databases for land use, topography and soil type. In this study we use version 3.6.1 of the WRF model [47]. WRF model forecasts at un-monitored sites are subject to uncertainty. Were we to use the high-resolution forecasts from the WRF model, we would be ignoring the uncertainty in them. The resulting air pollution estimates will also have unquantified uncertainty, which will reduce its usability.

This paper addresses the above issues in uncertainty quantification by adopting a Bayesian linear model linking the observed and forecasted meteorological variables by the WRF model. The Bayesian model enables us to assess the uncertainty of the meteorological variables forecasts at each of the corners of a 1-kilometer square grid. The calibrated forecasts are then used in a spatio-temporal regression model for air pollution exposure. Numerical cross-validation results obtained in this paper show better accuracy for the predicted air pollution exposure values using this proposed new method. The resulting model is able to forecast $\text{PM}_{2.5}$ concentrations at each corner of a 1-kilometer square grid.

The remainder of the article is organised as follows. Section 2 presents exploratory analysis of the study data set. Modelling details are provided in Section 3. Section 4 presents the main results. Conclusions and further developments are presented in Section 5.

2 Exploratory data analysis

2.1 $\text{PM}_{2.5}$ data description

We consider hourly mean $\text{PM}_{2.5}$ data obtained at $n = 11$ monitoring sites in SCL during three winter months, from June 1, 2011 to August 16, 2011. The choice of this period was motivated by the fact that $\text{PM}_{2.5}$ is higher in winter months and thereby ensuring a greater level of data availability. The monitoring sites belong to the national air quality infor-

mation system network (*Sistema de Información Nacional de Calidad del Aire (SINCA)*, <http://sinca.mma.gob.cl>) of Chile. Most of the monitoring sites are located in the center of the city as shown in Figure 1. The figure also shows the mountains surrounding the city, which impede air ventilation, especially during winter months, and the major roads carrying the bulk of the traffic in the city.

[Figure 1 about here.]

The boxplots in Figure 2 show the distribution of $PM_{2.5}$ concentrations by day of the week (a), by hour (b), and by monitoring station (c). Figure 2 (a) shows higher levels of $PM_{2.5}$ during the middle of the week when business activities are at their highest and the levels go down during the weekend. This indicates that there may be a weekday/weekend effect. A diurnal cycle is seen in Figure 2 (b). The boxplots together display a bi-modal pattern with significant peaks during the morning rush hour, 7-10AM and also during the night between 10PM and 2AM. This can be associated with the topography and the persistence of subsidence conditions in the winter months (June-August) that induce thermal inversions [15] and weak winds in the morning and evening, preventing the dispersion of pollutant concentrations and leading to the accumulation of gases (CO, NO_X and VOCs) and aerosols (PM_{10} and $PM_{2.5}$) in the air (see [35, 12, 53, 32, 55, 28] for more details). It is important to note that the high concentrations of $PM_{2.5}$ during the night and early morning hours may be also due to residential wood burning.

The site-wise boxplots in Figure 2 (c) show that the most polluted locations in Santiago are the monitoring sites 5 (Pudahuel) and 8 (Cerro Navia). See also Table 1 where the site-wise summary statistics are presented along with the station names. Figure 1 shows that both stations 5 and 8 are located in the North-West area of Santiago, which is characterized by the presence of thermoelectric and copper smelter plants.

The high levels of the hourly $PM_{2.5}$ concentrations in the 11 monitoring stations lead to high levels of the daily averages which often exceed the threshold of $50 \mu g/m^3$ fixed by the Chilean Ministry of the Environment and $25 \mu g/m^3$ set by WHO (World Health Organization). Fig. 3 shows the daily averages for the 11 monitoring stations. Values for the three stations: 5-Pudahuel, 7-El Bosque, 9-Cerro Navia, plotted in the colors light blue, yellow, and grey respectively, show a high number of exceedances beyond the Chilean threshold of $50 \mu g/m^3$. This figure also shows that, except for the station 3- Las Condes (green color) the $PM_{2.5}$ levels in all the other stations exceed this threshold often during the peak pollution episodes seen as the peaks in the graph. Moreover, the WHO limit is seen to be exceeded at all the stations except for a few low-pollution days seen as the troughs in the graph.

[Figure 2 about here.]

[Figure 3 about here.]

[Table 1 about here.]

2.2 Meteorological covariates

Pollution levels are often associated with relevant meteorological variables. For example, low temperatures in the winter period increase the demand for residential wood-burning heating in Santiago city, leading to increased pollution levels; the dispersion of pollution particles depends on wind speeds, with higher values causing a more rapid dispersion [29]; and the relative humidity affects the particle-growth by hydration [5] and showed a large role explaining PM variations in Santiago [55]. In our study we have access to temperature, relative humidity (RH) and wind speed observations at the 11 monitoring sites. Table 2 shows the summary statistics for these variables.

The selection of these covariates have been motivated by the availability of these variables in each monitoring station, and according to other studies of air pollution in Santiago ([20], [15], [44]). [15, 55, 29] determined a negative correlation between temperature and wind speed with $PM_{2.5}$ and a positive correlation between relative humidity and $PM_{2.5}$ but a weak negative correlation was found for wind direction. Other authors have also considered weather variables to model the $PM_{2.5}$ in Santiago [35, 43, 44, 28].

[Table 2 about here.]

The use of meteorological variables, however, poses a problem using our hierarchical model in the prediction of pollution concentrations at locations where no meteorological observations are available. To overcome this problem the WRF model output can be used instead of meteorological observations, given that the WRF model can predict the meteorological variables on a regular grid at different spatial resolutions. We first describe below the WRF model configuration used in this study.

2.3 WRF model configuration

The WRF model was run using the fully compressible and non-hydrostatic options. The simulations were performed using 4 nested domains and 50 vertical levels. Results from the domain with higher horizontal resolution (d04 at 1 km), covering the city of Santiago, was used in the study. The Final (FNL) analysis run by NCEP (National Centers for Environmental Prediction) four times per day (0, 6, 12 and 18 UTC), at 1 degree x 1 degree horizontal resolution was used as initial and boundary conditions for the WRF simulations every 6 hours. The WRF simulation started on 1st June 2011 at 00 UTC and ended on 31st August 2011 at 23 UTC for the area included in the latitudes ($-32^{\circ}26'59''$; $-34^{\circ}8'6''$) and longitudes ($-69^{\circ}58'23''$; $-71^{\circ}53'35''$). WRF outputs were saved every 1 hour.

A nudging [49] analysis was applied over the period of the simulation to avoid the regional simulation to depart too much from the large-scale atmosphere provided by the

FNL. Nudging is a technique used in dynamical downscaling to improve the representation of atmospheric fields inside a regional atmospheric or climate simulation. It imposes a constraint in the interior grid-points of the regional simulation toward the large-scale atmospheric fields, by adding a term to the primitive equations based on the difference between the inside atmospheric regional fields and the large-scale fields.

The simulation was performed using the Rapid Radiative Transfer Model for a general circulation model (RRTM-G) scheme for long-wave and short-wave radiation [18], the Mellor-Yamada-NN 2.5 level turbulent kinetic energy (TKE) scheme to resolve the planetary boundary layer processes [31], the Noah land surface model [52] to solve the land-surface interaction in all domains, and the WSM 6-class graupel scheme [17] to solve microphysics in domains 3 and 4.

Figure 4 shows the 2 m temperature (a), relative humidity (b), and the 10 m wind speed from the WRF model in August 16th at 8:00 local time (LT) and 13:00 LT. Figure 5 shows strong positive correlations between some observed meteorological variables (such as temperature and relative humidity) and WRF predictions.

[Figure 4 about here.]

[Figure 5 about here.]

3 Modelling details

3.1 Hierarchical models

The main aim of the space-time regression approach is to model and analyze random observations of $\text{PM}_{2.5}$ at location \mathbf{s} and time t . The pollution observations are modeled on the square-root scale in order to encourage normality and stabilize the variance [39]. Hence, we use the notation $Y(\mathbf{s}, t)$ to denote the square-root of the observations at location \mathbf{s} and at time t which we model using a set of p covariates $\mathbf{x}(\mathbf{s}, t)$. Here we take the following covariates: the calibrated WRF predictions for the three meteorological variables (temperature, relative humidity, and wind speed), the mixing layer, the altitude, the distance from the nearest road, and the hour of the day. We will explain the details of the calibration procedure used for the meteorological variables in the next Section. The hierarchical models are specified by:

$$Y(\mathbf{s}, t) = U(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \tag{1}$$

where $U(\mathbf{s}, t)$ is a structured spatio-temporal process that we specify using a second level hierarchical specification and the non-spatial random error term $\varepsilon(\mathbf{s}, t)$ is assumed to have a zero mean and constant variance σ_ε^2 . We envision this error term to be the measurement error, although there are other possible interpretations, see, e.g., Chapter 6 in [2] for a discussion of the nugget term in spatio-temporal models.

The structured process $U(\mathbf{s}, t)$ is specified by three main components: (i) $\rho U(\mathbf{s}, t - 1)$ accounting for temporal correlation for an unknown value of the correlation parameter ρ , (ii) $\mathbf{x}(\mathbf{s}, t)' \boldsymbol{\beta}$ explaining large scale variation due to the p covariates at the same time point and the p - dimensional vector $\boldsymbol{\beta}$ denotes the regression coefficients; and (iii) $\eta(\mathbf{s}, t)$, a random space-time intercept that cannot be accounted for by the two previous terms. We do not consider a spatially varying temporal correlation coefficient, $\rho(s)$, as our data set from only $n = 11$ sites is not reach enough to estimate such parameters in the presence of the unobserved $U(\mathbf{s}, t)$.

The term $U(\mathbf{s}, t)$ also provides the spatial smoothing in the model and we assume an independent in time Gaussian process model for it. Thus, the second stage hierarchical specification is given by:

$$U(\mathbf{s}, t) = \rho U(\mathbf{s}, t - 1) + \mathbf{x}(\mathbf{s}, t)' \boldsymbol{\beta} + \eta(\mathbf{s}, t). \quad (2)$$

We have collected data from $n = 11$ sites denoted by $\mathbf{s}_i, i = 1, \dots, n$ and T time points denoted by $t = 1, \dots, T$. Hence we assume the model equations (1) and (2) for each of the nT data points $\mathbf{s}_i, i = 1, \dots, n$ and $t = 1, \dots, T$. We define the notations: $\mathbf{Y}_t = (Y(\mathbf{s}_1, t), \dots, Y(\mathbf{s}_n, t))$ and $\mathbf{U}_t = (U(\mathbf{s}_1, t), \dots, U(\mathbf{s}_n, t))$, $\boldsymbol{\varepsilon}_t = (\varepsilon(\mathbf{s}_1, t), \dots, \varepsilon(\mathbf{s}_n, t))$, $\boldsymbol{\eta}_t = (\eta(\mathbf{s}_1, t), \dots, \eta(\mathbf{s}_n, t))$ and $\mathbf{X}_t = (\mathbf{x}'(\mathbf{s}_1, t), \dots, \mathbf{x}'(\mathbf{s}_n, t))'$. With the above notations, we re-write the model equations (1) and (2)

$$\mathbf{Y}_t = \mathbf{U}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_n), \quad (3)$$

$$\mathbf{U}_t = \rho \mathbf{U}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\eta}}) \quad (4)$$

where \mathbf{I}_n is the identity matrix of order n and the elements of the matrix $\Sigma_{\boldsymbol{\eta}}$ are specified using a Gaussian process as follows. We assume that $\Sigma_{\boldsymbol{\eta}} = \sigma_{\boldsymbol{\eta}}^2 S_{\boldsymbol{\eta}}$ where $\sigma_{\boldsymbol{\eta}}^2$ is the site-invariant common spatial variance and the i th row and j th column element of the correlation matrix $S_{\boldsymbol{\eta}}$ is obtained using the Matérn correlation function defined as:

$$S_{\boldsymbol{\eta}}(\mathbf{s}_i, \mathbf{s}_j; \phi, \nu) = \frac{1}{2^{\nu-1} \Gamma(\nu)} (2\sqrt{\nu} \|\mathbf{s}_i - \mathbf{s}_j\| \phi)^{\nu} K_{\nu}(2\sqrt{\nu} \|\mathbf{s}_i - \mathbf{s}_j\| \phi), \quad \phi > 0, \nu > 0, \quad (5)$$

where $K_{\nu}(\cdot)$ is the Bessel function of the second kind and ϕ and ν are parameters governing the strength and smoothness of the spatial correlation. In our implementation we take $\nu = 0.5$ which corresponds to the popular exponential correlation function. This model has been implemented in the R language in the spTimer package [1]. Our implementation will be based on this package too.

Bayesian model checking and verification of the model assumptions proceed by using the posterior predictive distributions, see e.g. [14]. Here one simply simulates replicated data from the fitted model and then compares those with the observed data. These authors also define Bayesian p-values based on the posterior predictive distributions. In this article, we do not adopt those formal Bayesian approaches as here the main objective is prediction

rather than model choice. We use leave one out cross validation approaches to validate and compare different modelling approaches in Section 4. However, before discussing those results we first provide the prediction details using our models.

3.2 Prediction details

Prediction at a new location \mathbf{s}_0 at any time point t must be performed using the model equations (1) and (2), i.e.,

$$\begin{aligned} Y(\mathbf{s}_0, t) &= U(\mathbf{s}_0, t) + \varepsilon(\mathbf{s}_0, t) \\ U(\mathbf{s}_0, t) &= \rho U(\mathbf{s}_0, t-1) + \mathbf{x}(\mathbf{s}_0, t)' \boldsymbol{\beta} + \eta(\mathbf{s}_0, t). \end{aligned}$$

Here we obtain the posterior predictive distribution of $Y(\mathbf{s}_0, t)$ given all the observed data and covariate values from the n data sites and also the prediction location \mathbf{s}_0 . However, as mentioned above, the values $\mathbf{x}(\mathbf{s}_0, t)$ are not available as we only have observed meteorological variables at the 11 data sites only. To overcome this problem, we assume a linear calibration model as follows. For each of the three meteorological variables, $\mathbf{x}_j(\mathbf{s}, t)$, $j = 1, 2, 3$ we assume that

$$x_j(\mathbf{s}_i, t) = \gamma_{0j} + \gamma_{1j} x_j^{(M)}(\mathbf{s}_i, t) + \omega(\mathbf{s}_i, t) \quad (6)$$

where γ_{0j} and γ_{1j} are intercept and slope for the j th meteorological variable, $j = 1, 2, 3$ and $\omega(\mathbf{s}_i, t)$ is a space-time error term and $x_j^{(M)}(\mathbf{s}_i, t)$ is the known WRF output for the j th meteorological variable at location \mathbf{s}_i and at time t . In our implementation, we have adopted an independent error distribution in space and time for $\omega(\mathbf{s}_i, t)$ since we do not want further smoothing of the WRF model output. That is, for the error term $\omega(\mathbf{s}, t)$, we do not assume any spatio-temporal dependence structure as we have assumed for $\eta(\mathbf{s}, t)$ in (2).

Consequently, we assume that $\omega(\mathbf{s}_i, t) \sim N(0, \sigma_\omega^2)$ independently for each $i = 1, \dots, n$ and $t = 1, \dots, T$. We assume independent vague prior distribution $N(0, 10^4)$ for γ_{0j} and γ_{1j} and assume a proper inverse gamma distribution with parameters 2 and 1 for σ_ω^2 . We then work out the posterior predictive distribution of $x_j(\mathbf{s}_0, t)$ given all the observed meteorological data $x_j(\mathbf{s}_i, t)$ for $i = 1, \dots, n$ and $t = 1, \dots, T$. From the posterior predictive distribution we obtain the mean and the 95% credible intervals, which we send to the pollution model (2) for assessing uncertainty due to the uncertainties in the model (6).

Our proposed prediction method runs algorithmically in two-steps as follows. In the first step, we fit the Bayesian linear model (6) and obtain samples $x_j(\mathbf{s}_0, t)^{(\ell)}$, for each \mathbf{s}_0 and t , and for each $\ell = 1, \dots, L$ where L is a large number, 5000 in our implementation. We then find summary statistics, i.e. the mean and the 95% credible limits from these samples and use those in the pollution model (2). Again we note that we perform this independently for each of the three meteorological variables separately.

In the second step, we use the `spTimer` package to fit the full spatio-temporal model. Once that model fitting has been performed, we go on to do the $\text{PM}_{2.5}$ predictions at each of the corners of the 1-kilometer square grid.

4 Results

4.1 Bayesian calibration of temperature, relative humidity and wind speed

We separately fitted the Bayesian linear model (6) (using the R package `spBayes`; see [11]), for each of the three meteorological variables, temperature, relative humidity and wind speed using the actual observed data from the monitoring stations and the outputs from the WRF model. The parameters of the model are estimated using the data for the period between June 1st, 2011 and August 15th, 2011 and the resulting fitted model is used to forecast the values for the next 24 hours (August 16, 2011).

Table 3 shows the Root Mean Square Error (RMSE), Bias, correlation (r) and Coverage percentage (Cov) for each monitoring station and each meteorological variable (wind speed, temperature, and relative humidity) for the 24h forecast in August 16th. The best correlation between observations and fitted values is obtained for temperature in all stations, and the best fitting (in terms of RMSE and BIAS) is shown in stations 4 (Parque O'Higgins) and 8 (Cerro Navia). This can be explained by the strong linear correlation between observations and WRF outputs as shown in Fig. 4a). A similar behavior (although with lower correlation coefficients) has been obtained for the relative humidity, except for stations 3 (Las Condes) and 9 (Puente Alto). This can be related to the fact that both stations are characterized by the highest values of altitude (768 and 670 meters respectively).

Our results are comparable to those obtained by [44] who found a higher correlation for temperature than for wind speed at Cerro Navia station 3. Other studies have also found a large bias in wind speed forecasts from WRF in complex terrain in other parts of Chile and the world [36, 21, 27, 34].

[Table 3 about here.]

In order to illustrate the calibration results for an individual station, we chose the Pudahuel station, since this station showed the highest mean $\text{PM}_{2.5}$ concentration (see Table 1). Figure 6 shows the calibration plots for temperature, RH and wind speed from August 1st to 15th. In general, the model calibrated values (red lines) are much closer to observations (black lines) than the uncalibrated (raw) WRF model output (blue lines).

Henceforth, the model fitted values for the three meteorological variables, wind speed, temperature, and relative humidity will be used as the calibrated values and the raw WRF model output will be referred to as the uncalibrated values. In order to propagate the uncertainty in the calibration values, we work not only with the mean calibrated values but

also with the 2.5% lower credible limit (LCL) and the 97.5% upper credible limit (UCL) of calibrated meteorological variables. The hierarchical models given by (1) and (2) with the un-calibrated and the calibrated (mean, LCL and UCL) WRF weather variables will be henceforth denoted by Un-CAL, Mn-CAL, Lo-CAL and Up-CAL, respectively.

[Figure 6 about here.]

4.2 Model estimation

In this section we show the parameter estimates of the Bayesian hierarchical model for predicting the $PM_{2.5}$ concentration in SCL using the R package `spTimer` from June 1st to August 15th, 2011. We consider spatial and/or temporal exogenous variables (the hour of the day as a factor, the altitude, the temperature, the distance to the principal road, the atmospheric boundary layer height, wind speed and relative humidity) as covariates in the matrix $X(\mathbf{s}, t)$. The four Bayesian models, Un-CAL, Mn-CAL, Lo-CAL and Up-CAL, have been fitted using MCMC and Table 4 provides the Bayesian predictive model choice criteria (PMCC) developed by [13]. The PMCC criteria clearly select the Mn-CAL model which uses the mean calibrated WRF weather variables.

Table 5 shows the estimated parameters for the four models. The variations in the parameter estimates for the three models based on the calibrated weather variables show the effect of uncertainty in using the calibration method. For all the parameters we observe that the estimates do not vary a great deal and they all point to similar relationships between the $PM_{2.5}$ and the other covariates. As expected, the $PM_{2.5}$ concentration is inversely correlated with temperature, wind speed, boundary layer height and altitude. Similar results have been obtained for temperature and wind speed by [29], while the negative correlation between pollution and the mixing layer is explained by [51] from the meteorological point of view. By comparing the un-calibrated (Un-CAL) with the mean-calibrated model (Mn-CAL) we find a large difference in the estimated coefficient for wind speed. The estimates imply that the correlation with wind speed is stronger in the calibrated model. The spatial correlation is seen to be very similar in the two models.

[Table 4 about here.]

The spatio-temporal prediction performance of the models at each monitoring station has been assessed by implementing a leave-one-out cross-validation method for the period from June 1st to August 15th, 2011. Table 6 shows the RMSE, BIAS, and the percentage of coverage for the models using the un-calibrated and calibrated (mean) variables. As expected, the values for these indexes are generally better for the model using mean calibrated variables for the meteorological variables. In particular, the RMSE is only lower for UN-CAL for station 7 but for this station the BIAS is lower for Mn-CAL.

[Table 5 about here.]

4.3 Temporal prediction

In order to show the forecast ability of the model, we made a 24h prediction at each monitoring station on day August 16th by using the $PM_{2.5}$ data from June 1st to August 15th. In Figure 7 (a) we can observe the correlation between observations and predictions using calibrated and un-calibrated variables, whereas Figure 7 (b) shows the skill of the temporal prediction for station 5 (Pudahuel), which was the most polluted location according to the available observations. The figures show that the predictions are better for the model with the calibrated variables, especially for values of $PM_{2.5}$ concentrations between 15 and 25 $\mu g/m^3$ (Figure 7a), and, with reference to station 5 (Pudahuel), the predictions obtained with calibrated variables are much closer to the observations (Figure 7 b). These results are confirmed by the residual analysis of Table 7, which shows the lowest values of RMSE, MAE, MAPE, BIAS, rBIAS, and rMSEP for the model using the calibrated mean variables. In particular, we note that the bias of the model is considerably reduced in this case.

[Figure 7 about here.]

[Table 6 about here.]

4.4 Spatio-Temporal Prediction

Figures 8 and 9 show the spatio-temporal $PM_{2.5}$ prediction for August 16th at 8:00 and 13:00 LT, respectively, using a spatial resolution of $1km \times 1km$. We note that $PM_{2.5}$ concentration is higher in the morning which is normally characterized by higher traffic and lower temperature. In particular, Figure 8 (a) shows that the highest $PM_{2.5}$ concentrations are located in the North-West part of Santiago city where a large industrial sector is established. As expected, standard deviations are lower in the spatial points close to the monitoring stations (Figures 8b and 9b).

[Figure 8 about here.]

[Figure 9 about here.]

5 Conclusions

In this paper we have developed a hierarchical Bayesian land-use regression model for predicting the hourly $PM_{2.5}$ in conjunction with a calibration model for the covariates, given by the meteorological WRF variables (wind speed, temperature, and relative humidity). The results showed that the $PM_{2.5}$ forecasts using calibrated meteorological outputs are much more accurate than the predictions using un-calibrated ones. The proposed model is able to make good predictions for the next 24 hours on a regular grid of $1km \times 1km$. We

think these forecasts could be helpful to the local authorities to take decisions when high levels of pollution are predicted in the day and in case the daily average forecast level exceed the thresholds set by the national government and WHO for protecting the human health. The ability of the model to produce good forecasts on fine spatial and temporal resolutions make a significant contribution to the existing literature where most of the work focuses on forecasting daily average levels in a temporal domain and/or on a coarse spatial grid. Also, the proposed models resulting from the combination of the spatio-temporal hierarchical Bayesian model with the calibration one, contributes methodologically to a new area of research consisting on calibrating the covariates in a spatio-temporal approach.

6 Acknowledgements

Mailiu Diáz is grateful for the support of the National Commission for Scientific and Technological Research (CONICYT) of Chile under Grant No. 21150227. Orietta Nicolis and Julio C. Marn are partially supported by the Interdisciplinary Center of Atmospheric and Astro-Statistical Studies. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the National Laboratory for High Performing Computer (NLHPC) (ECM-02). National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2000: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. [Available online at <https://doi.org/10.5065/D6M043C6>.] Accessed 15 Aug 2016.

References

- [1] K. S. Bakar and S. K. Sahu. `spTimer`: Spatio-temporal bayesian modelling using `r`. *Journal of Statistical Software*, 63(15):1–32, 2015.
- [2] S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 2014.
- [3] G. Christakos. *Modern Spatiotemporal Geostatistics*. Dover Books on Chemistry and Earth Science. Dover Books on Chemistry and Earth Science, 2012.
- [4] J. S. Clark and A. E. Gelfand. *Hierarchical modelling for the environmental sciences: Statistical Methods and Applications*. Oxford University Press, Oxford, New York, 2006.
- [5] N. Clements, M. P. Hannigan, S. L. Miller, J. L. Peel, and J. B. Milford. Comparisons of urban and rural $\text{pm}_{10-2.5}$ and $\text{pm}_{2.5}$ mass concentrations and semi-volatile fractions in northeastern colorado. *Atmospheric Chemistry and Physics*, 16(11):7469–7484, 2016.
- [6] N. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, 1993.
- [7] N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 2011.
- [8] L. Díaz-Robles, S. Cortés, A. Vergara-Fernández, and J. C. Ortega. Short term health effects of particulate matter: A comparison between wood smoke and multi-source polluted urban areas in Chile. *Aerosol and Air Quality Research*, 15:306–318, 2015.
- [9] L.A. Díaz-Robles, J.C. Ortega, J.S. Fu, G.D. Reed, J.C. Chow, J.G. Watson, and J.A. Moncada-Herrera. 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, 2008.
- [10] A. Fassò, M. Cameletti, and O. Nicolis. Air quality monitoring using heterogeneous networks. *Environmetrics*, 18(3):245–264, 2007.
- [11] A. O. Finley, S. Banerjee, and A. E. Gelfand. `spBayes` for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software*, 63(13):1–28, 2015.
- [12] P. A. Garcia-Chevesich, S. Alvarado, D. G. Neary, R. Valdes, J. Valdes, J. J. Aguirre, M. Mena, R. Pizarro, P. Jofré, M. Vera, and C. Olivares. Respiratory disease and particulate air pollution in Santiago Chile: Contribution of erosion particles from fine sediments. *Environmental Pollution*, 187:202–205, 2014.
- [13] A. E. Gelfand and S. K. Ghosh. Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85:1–11, 1998.
- [14] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin.

- Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science Taylor & Francis, 2004.
- [15] E. Gramsch, D. Cáceres, P. Oyola, F. Reyes, Y. Vásquez, M. A. Rubio, and G. Sánchez. Influence of surface and subsidence thermal inversion on PM_{2.5} and black carbon concentration. *Atmospheric Environment*, 98:290–298, 2014.
- [16] E. Gramsch, F. Cereceda-Balic, P. Oyola, and D. Baer. Examination of pollution trends in Santiago de Chile with cluster analysis of PM₁₀ and ozone data. *Atmospheric Environment*, 40:5464–5475, 2006.
- [17] S.-Y. Hong and J.-O. J. Lim. The WRF single-moment 6-class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, 42(2):129–151, 2006.
- [18] M. J. Iacono, J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, 113(D13), 2008.
- [19] N.A. Janssen, M.E. Gerlofs-Nijland, T. Lanki, R.O. Salonen, F. Casee, G. Hoek, P. Fischer, B. Brunekreef, and M. Krzyzanowski. Health effects of black carbon. *Technical report*, 2012.
- [20] I. Jhun, P. Oyola, F. Moreno, M. A. Castillo, and P. Koutrakis. PM_{2.5} mass and species trends in Santiago, Chile, 1998 to 2010: The impact of fuel-related interventions and fuel sales. *Journal of the Air & Waste Management Association*, 63(2):161–169, 2013.
- [21] P. A. Jiménez and J. Dudhia. Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. *Journal of Applied Meteorology and Climatology*, 51(2):300–316, 2012.
- [22] H. Jorquera and F. Barraza. Source apportionment of ambient PM_{2.5} in Santiago, Chile: 1999 and 2004 results. *The Science of the total environment*, 435-436:418–429, 2012.
- [23] P. Koutrakis, S.N. Sax, J.A. Sarnat, B. Coull, P. Demokritou, P. Oyola, J. Garcia, and E. Gramsch. Analysis of PM₁₀, PM_{2.5}, and PM_{2.5-10} concentrations in Santiago, Chile, from 1989 to 2001. *Journal of the Air & Waste Management Association*, 55:342–351, 2005.
- [24] P. C. Kyriakidis and A. G. Journel. Geostatistical space-time models: A review. *Mathematical geology*, 31(6):651–684, 1999.
- [25] J. Lindström, A.A. Szpiro, P.D. Sampson, S. Bergen, and Oron A. **SpatioTemporal**: An r package for spatio-temporal modelling of air-pollution. *Journal of Statistical Software (in press)*, 2013.
- [26] K. V. Mardia and C. R. Goodall. Spatial-temporal analysis of multivariate environmental monitoring data. *Multivariate environmental statistics*, 6(76):347–385, 1993.
- [27] J. C. Marín, D. Pozo, E. Mlawer, D. D. Turner, and M. Curé. Dynamics of local

- circulations in mountainous terrain during the RHUBC-II project. *Monthly Weather Review*, 141(10):3641–3656, 2013.
- [28] A. Mazzeo, N. Huneus, C. Ordoñez, A. Orfanoz-Cheuquelaf, L. Menut, S. Mailler, M. Valari, H. Denier van der Gon, L. Gallardo, R. Muñoz, R. Donoso, M. Galleguillos, M. Osses, and S. Tolvett. Impact of residential combustion and transport emissions on air pollution in Santiago during winter. *Atmospheric Environment*, 190:195–208, 2018.
- [29] S. Moisan, R. Herrera, and A. Clements. A dynamic multiple equation approach for forecasting PM2.5 pollution in Santiago, Chile. *International Journal of Forecasting*, 34:566–581, 2018.
- [30] S. Mukhopadhyay and S. K. Sahu. A bayesian spatiotemporal model to estimate long-term exposure to outdoor air pollution at coarser administrative geographies in England and Wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(2):465–486, 2017.
- [31] M. Nakanishi and H. Niino. An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary Layer Meteorology*, 119(2):397–407, 2006.
- [32] P. Perez and E. Gramsch. Forecasting hourly PM2.5 in santiago de chile with emphasis on night episodes. *Atmospheric Environment*, 124:22–27, 1 2016.
- [33] P. Perez and J. Reyes. An integrated neural network model for PM10 forecasting. *Atmospheric Environment*, 40(16), 2006.
- [34] D. Pozo, J. Marín, G. Raga, J. Arevalo, D. Baumgardner, A. Cordova, and J. Mora. Synoptic and local circulations associated with events of high particulate pollution in Valparaiso, Chile. *Atmospheric Environment*, 196, 10 2018.
- [35] K. M. Ragsdale, B. S. Barrett, and A. P. Testino. Variability of particulate matter (PM10) in Santiago, Chile by phase of the Madden-Julian Oscillation (MJO). *Atmospheric Environment*, 81:304–310, 2013.
- [36] J. J. Ruiz, C. Saulo, and J. Nogués-Paegle. WRF model sensitivity to choice of parameterization over south america: Validation against surface variables. *Monthly Weather Review*, 138(8):3342–3355, 2010.
- [37] S. K. Sahu and K. S Bakar. Hierarchical bayesian auto-regressive models for large space time data with applications to ozone concentration modelling. *Applied Stochastic Models in Business and Industry*, 28:395–415, 2012.
- [38] S. K. Sahu, A. E. Gelfand, and D. M. Holland. Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:61–86, 2006.
- [39] S. K. Sahu, A. E. Gelfand, and D. M. Holland. High-resolution space-time ozone modeling for assessing trends. *Journal of the American Statistical Association*, 102:1221–

1234, 2007.

- [40] S. K. Sahu and K. V. Mardia. Recent trends in modeling spatio-temporal data. In *Proceedings of the Special meeting on Statistics and Environment*, pages 69–83. Università Di Messina, September 2005.
- [41] S. K. Sahu and O. Nicolis. An evaluation of european air pollution regulations for particulate matter monitored from a heterogeneous network. *Environmetrics*, 20(8):943–961, 2009.
- [42] S. K. Sahu, S. Yip, and D. M. Holland. Improved space-time forecasting of next day ozone concentrations in the eastern u.s. *Atmospheric Environment*, 43:494–501, 2009.
- [43] P. E. Saide, G. R. Carmichael, S. N. Spak, L. Gallardo, A. E. Osses, M. A. Mena-Carrasco, and M. Pagowski. 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, 2011.
- [44] P. E. Saide, M. Mena-Carrasco, S. Tolvett, P. Hernandez, and G. R. Carmichael. 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, 2016.
- [45] S. N. Sax, P. Koutrakis, P. A. Rudolph, F. Cereceda-Balic, E. Gramsch, and P. Oyola. Trends in the elemental composition of fine particulate matter in Santiago, Chile, from 1998 to 2003. *J Air Waste Manag Assoc.*, 57:845–855, 2007.
- [46] G. Shaddick and J. Wakefield. Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society Series C*, 51(3):351–372, 2002.
- [47] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers. A description of the advanced research WRF version 3. *NCAR Technical note*, page 125, 2008.
- [48] R. L. Smith, S. Kolenikov, and L. H. Cox. Spatiotemporal modeling of PM2.5 data with missing values. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.
- [49] D. R. Stauffer and N. L. Seaman. Multiscale four-dimensional data assimilation. *Journal of Applied Meteorology*, 33(3):416–434, 1994.
- [50] J. R. Stroud, P. Müller, and B. Sansó. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):673–689, 2001.
- [51] Jinqiang; Zhu Xiaowan; Song Tao; Mnel Christoph; Hu Bo; Schfer Klaus; Liu Zirui; Zhang Junke; Wang Lili; Xin Jinyuan; Suppan Peter; Wang Yuesi Tang, Guiqian; Zhang. Mixing layer height and its implications for air pollution over beijing, china. *Atmospheric Chemistry and Physics*, 16, 03 2016.
- [52] M. Tewari, F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca. Implementation and verification of the unified

noah land surface model in the WRF model. In *Proceedings of the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*, page 14.2A, 2004.

- [53] A. M. Villalobos, F. Barraza, H. Jorquera, and J. J. Schauer. Chemical speciation and source apportionment of fine particulate matter in Santiago, Chile, 2013. *Science of the Total Environment*, 512-513:133–142, 2015.
- [54] C. K. Wikle and N. Cressie. A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829, 1999.
- [55] M.A. Yañez, R. Baettig, J. Cornejo, F. Zamudio, J. Guajardo, and R. Fica. Urban airborne matter in central and southern Chile: Effects of meteorological conditions on fine and coarse particulate matter. *Atmospheric Environment*, 161:221–234, 2017.
- [56] J. Zidek, L. Sun, N. Le, and H. Ozkaynak. Contending with space-time interaction in the spatial prediction of pollution: Vancouver’s hourly ambient pm10 field. *Environmetrics*, 13:595–613, 2002.

Table 1: Summary statistics of the PM_{2.5} data.

No.	Station	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD	NA
1	La Paz	1.00	16.00	28.00	30.04	42.00	124.00	18.05	18
2	La Florida	1.00	14.00	28.00	34.44	48.00	152.00	25.23	8
3	Las Condes	1.00	10.00	18.00	20.67	29.00	78.00	14.38	10
4	Parque O'Higgins	1.00	16.00	31.00	36.37	52.00	154.00	25.90	5
5	Pudahuel	1.00	13.00	33.00	45.95	62.00	364.00	48.60	16
6	Cerrillos	1.00	17.00	36.00	43.98	64.00	187.00	33.96	5
7	El Bosque	1.00	15.00	33.00	41.92	58.00	241.00	36.55	11
8	Cerro Navia	1.00	14.00	36.00	51.47	73.00	346.00	50.94	9
9	Puente Alto	1.00	12.00	24.00	27.30	38.00	119.00	19.75	4
10	Talagante	1.00	11.00	27.00	33.66	49.00	202.00	28.54	4
11	Quilicura	1.00	13.00	29.00	33.99	47.00	210.00	27.59	11

Table 2: Summary statistics for the meteorological variables.

Covariates	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD	NA
Relative Humidity	0.00	58.08	74.08	69.96	84.71	100.00	18.48	573
Temperature	-4.27	4.96	7.66	8.26	11.17	26.83	4.96	73
Wind Speed	0.00	0.53	0.91	1.03	1.35	6.73	0.69	38

Table 3: Calibration results for the meteorological variables: wind speed, temperature, and relative humidity.

St.	Wind Speed				Temperature				Relative Humidity			
	RMSE	Bias	r	Cov	RMSE	Bias	r	Cov	RMSE	Bias	r	Cov
1	1.16	-0.26	0.44	100.00	3.86	-3.78	0.96	100.00	12.50	9.03	0.87	100.00
2	1.46	-0.46	0.11	100.00	2.36	-2.15	0.95	100.00	11.21	7.37	0.90	100.00
3	0.55	0.31	-0.04	100.00	5.95	-5.61	0.84	50.00	20.28	14.69	-0.01	79.16
4	1.41	-0.36	0.15	100.00	1.66	-1.32	0.94	100.00	10.29	5.27	0.85	100.00
5	1.14	-0.13	0.11	100.00	3.49	-3.35	0.95	100.00	17.46	14.81	0.83	87.50
6	1.62	-0.92	0.31	100.00	3.54	-3.42	0.95	100.00	9.78	2.85	0.84	100.00
7	1.37	-0.69	0.17	100.00	2.08	-1.77	0.95	100.00	10.21	5.71	0.87	100.00
8	1.63	-0.34	0.09	100.00	1.73	-1.36	0.96	100.00	10.60	5.41	0.88	100.00
9	1.73	-1.05	0.19	100.00	4.08	-3.89	0.94	95.83	17.81	13.63	0.04	87.50
10	0.66	0.50	0.32	100.00	2.67	1.87	0.69	95.83	7.65	2.07	0.56	100.00
11	0.87	0.06	0.35	100.00	3.66	-3.46	0.91	100.00	15.02	11.21	0.92	100.00

Table 4: Model Choice Criteria for the Un-CAL, Mn-CAL, Lo-CAL and Up-CAL models.

	Un-CAL	Lo-CAL	Mn-CAL	Up-CAL
Goodness of fit	13581	13569	13551	13713
Penalty	10140	10149	10136	10132
PMCC	23721	23718	23687	23845

Table 5: Parameter estimates for the four different models.

Parameters	Un-CAL		Mn-CAL		Lo-CAL		Up-CAL	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intercept	2.1919	0.1081	3.5259	0.2200	2.3540	0.1128	4.3642	0.3161
Hour (6:00-12:00)	0.1669	0.0305	0.1685	0.0323	0.1704	0.0299	0.1680	0.0328
Hour (13:00-16:00)	-0.0113	0.0416	-0.0109	0.0436	-0.0066	0.0419	-0.0106	0.0437
Hour (17:00-23:00)	0.6754	0.0309	0.6790	0.0330	0.6762	0.0303	0.6734	0.0335
Mixing Layer	-0.4365	0.0412	-0.4358	0.0426	-0.4451	0.0414	-0.4451	0.0427
Wind Speed	-0.1095	0.0195	-1.1659	0.1745	-1.0224	0.1633	-0.9867	0.1658
Altitude	-0.0011	0.0001	-0.0011	0.0001	-0.0010	0.0001	-0.0010	0.0001
Temperature	-0.0116	0.0037	-0.0122	0.0041	-0.0117	0.0039	-0.0123	0.0040
Relative Humidity	-0.0050	0.0007	-0.0095	0.0012	-0.0092	0.0012	-0.0096	0.0012
Distance to road	-0.0007	0.0013	-0.0005	0.0013	-0.0006	0.0014	-0.0006	0.0013
ρ	0.7964	0.0044	0.7960	0.0042	0.79620	0.0044	0.7961	0.0045
σ^2_ε	0.0096	0.0007	0.0098	0.0007	0.0098	0.0008	0.0097	0.0008
σ^2_η	1.1105	0.0139	1.1068	0.0141	1.1114	0.0156	1.1124	0.0143
ϕ	0.1368	0.0028	0.1375	0.0032	0.1370	0.0029	0.1361	0.0031

Table 6: Station-wise cross-validation performance of the Un-CAL and Mn-CAL models.

Station	Un-CAL			Mn-CAL		
	RMSE	BIAS	Coverage (%)	RMSE	BIAS	Coverage (%)
1	9.66	8.40	95.83	6.10	3.71	100.00
2	7.68	6.05	100.00	6.24	2.35	100.00
3	3.82	0.72	100.00	3.64	-1.49	100.00
4	10.20	9.05	100.00	7.03	4.38	100.00
5	16.75	15.81	45.83	10.81	8.41	66.67
6	9.28	7.56	100.00	6.85	1.88	100.00
7	7.90	5.59	100.00	8.14	1.23	100.00
8	11.49	9.80	58.33	8.67	3.05	75.00
9	6.13	4.97	100.00	5.46	1.69	100.00
10	27.26	25.77	62.50	20.05	19.23	79.17
11	12.18	11.21	95.83	7.35	4.62	100.00

Table 7: Residual analysis for the temporal prediction.

Model	RMSE	MAE	MAPE	BIAS	rBIAS	rMSEP
Un-CAL	9.57	7.94	278.97	7.29	0.81	0.82
Mn-CAL	7.09	5.60	202.36	2.40	0.27	0.78

List of Figures

1	Google map representation of Santiago city and the location of the 11 SINCA air-quality monitoring sites.	25
2	Boxplot of PM _{2.5} by day of the week (from Monday=1 to Sunday =7) (a), by hour (with midnight=0) (b), and by monitoring station (c).	26
3	Time series of the daily average PM _{2.5} concentrations for the 11 stations. The dashed horizontal line shows the the upper limit at 50 stipulated by the Chilean Government and the solid horizontal line shows the limit suggested by the WHO.	27
4	WRF model output for August 16th at 8:00 LT and 13:00 LT (bottom) for: temperature (a), relative humidity (b), and wind speed (c).	28
5	Scatter plots of 2 m temperature (a), 2 m relative humidity (b), and 10 m wind speed (c) from observations and WRF forecasts.	29
6	Calibration plots for the period August 1st to 15th at Pudahuel monitoring station: Panel (a) temperature, panel (b) relative humidity panel (c) wind speed. Observations are represented by black lines, the blue lines represent the WRF simulations, the red lines represent the mean fitted curve, and green dashed lines represent the 2.5% and 97.5% quantiles.	30
7	(a) Correlation plot between observations and predictions: calibrated predictions (red points) and un-calibrated predictions (black points). (b) Temporal prediction for Pudahuel. Black line shows the observations; blue line shows the predictions using model Un-CAL; red line shows the predictions using model Mn-CAL. The green dashed lines show the predictions from the Lo-CAL and Up-CAL models, respectively.	31
8	Spatio temporal prediction of PM _{2.5} concentration on August 16th at 8:00 (a), and its standard deviation (b).	32
9	Spatio temporal prediction of PM _{2.5} concentration on August 16th at 13:00 CLST (a), and its standard deviation (b).	33

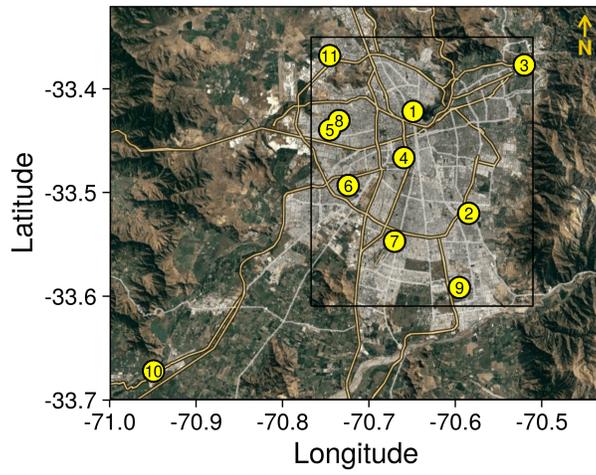


Figure 1: Google map representation of Santiago city and the location of the 11 SINCA air-quality monitoring sites.

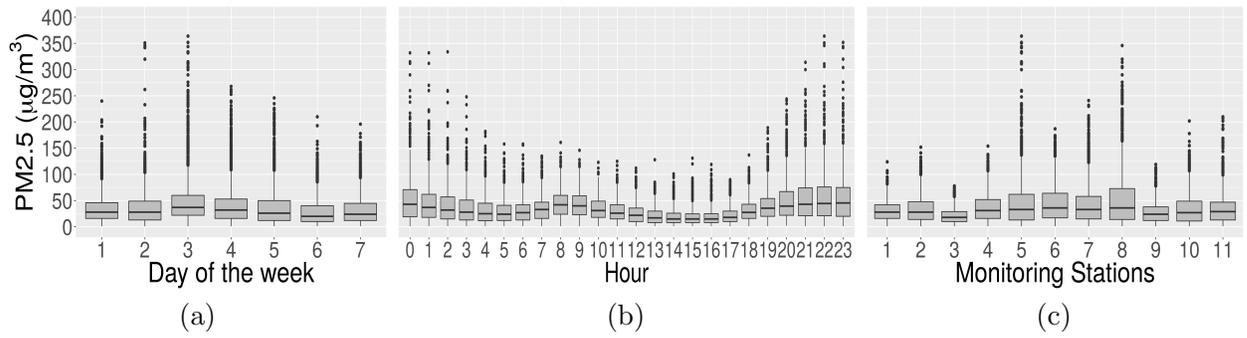


Figure 2: Boxplot of PM_{2.5} by day of the week (from Monday=1 to Sunday =7) (a), by hour (with midnight=0) (b), and by monitoring station (c).

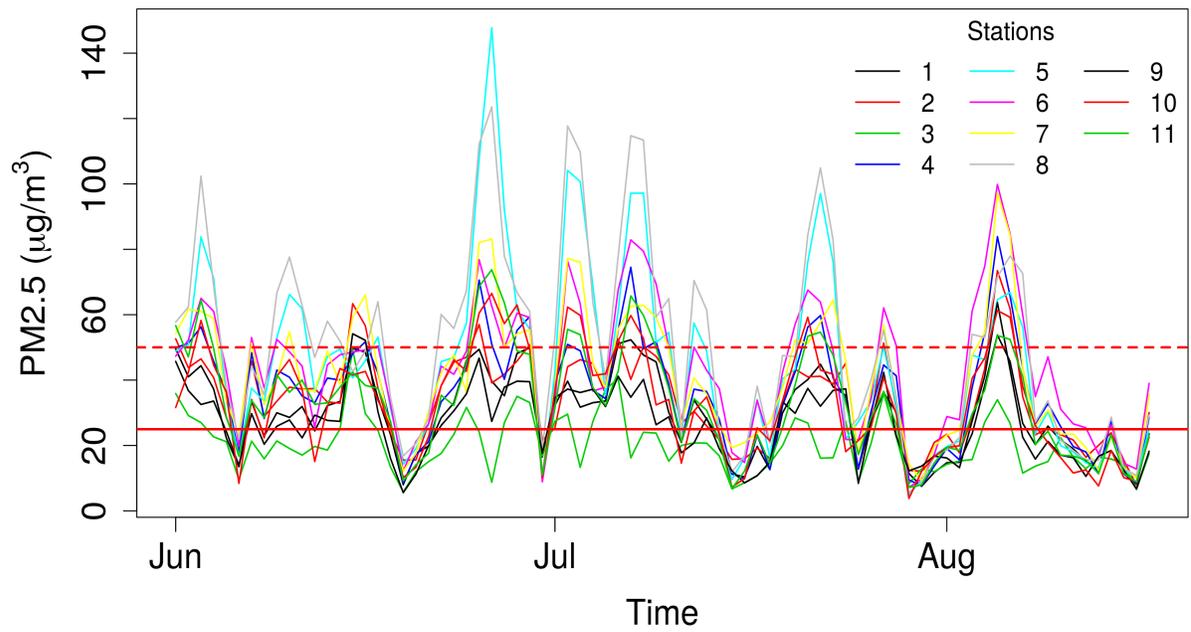


Figure 3: Time series of the daily average PM_{2.5} concentrations for the 11 stations. The dashed horizontal line shows the the upper limit at 50 stipulated by the Chilean Government and the solid horizontal line shows the limit suggested by the WHO.

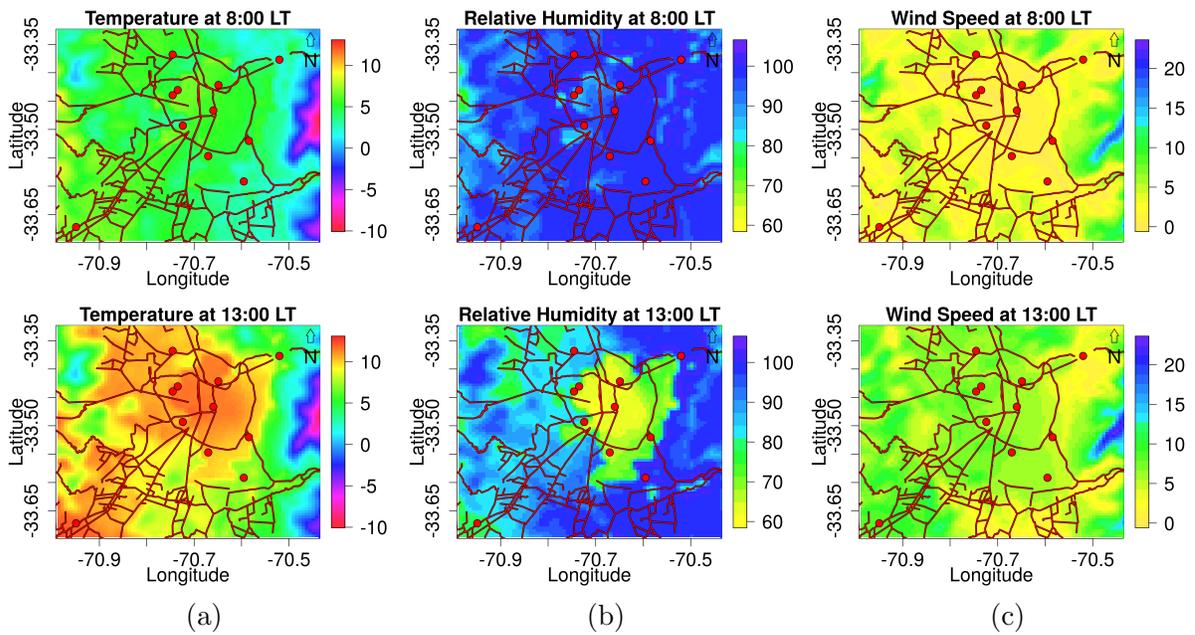


Figure 4: WRF model output for August 16th at 8:00 LT and 13:00 LT (bottom) for: temperature (a), relative humidity (b), and wind speed (c).

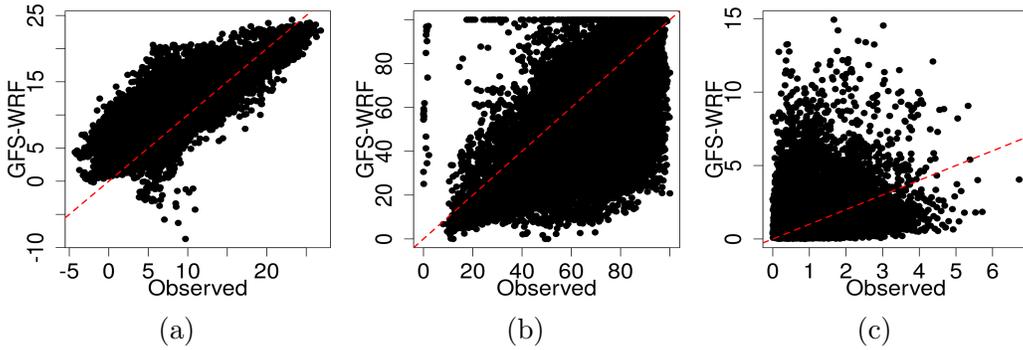
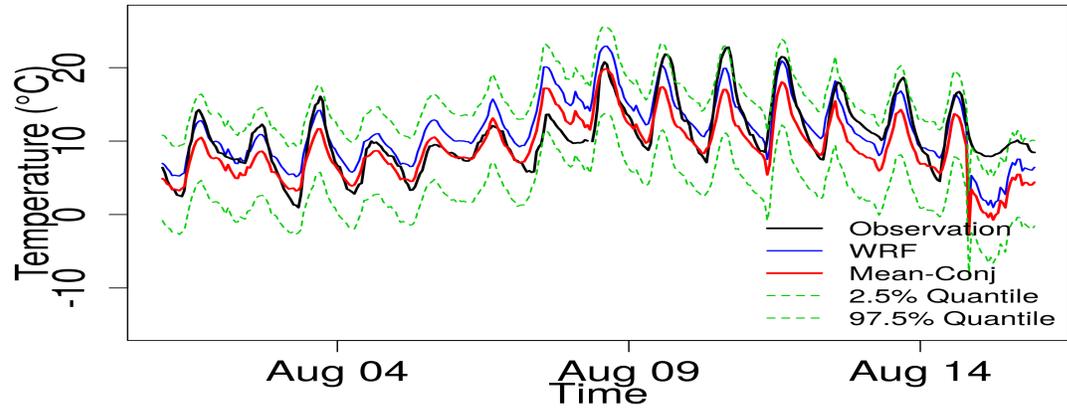
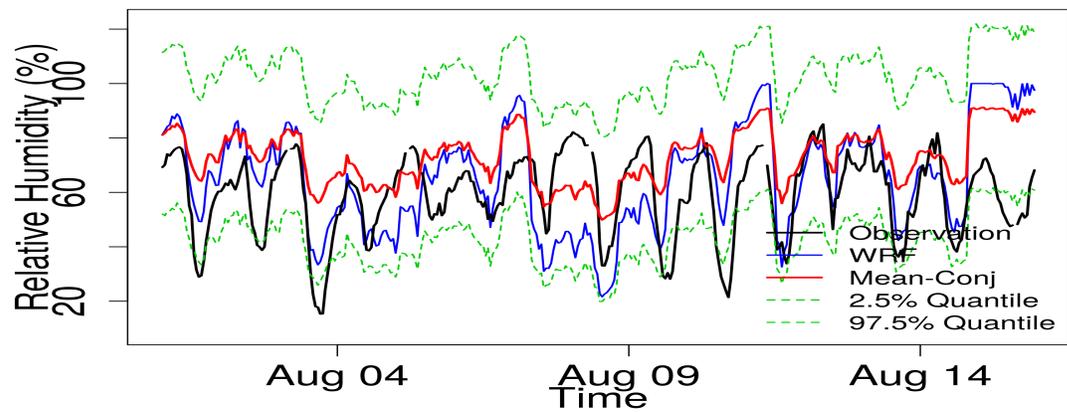


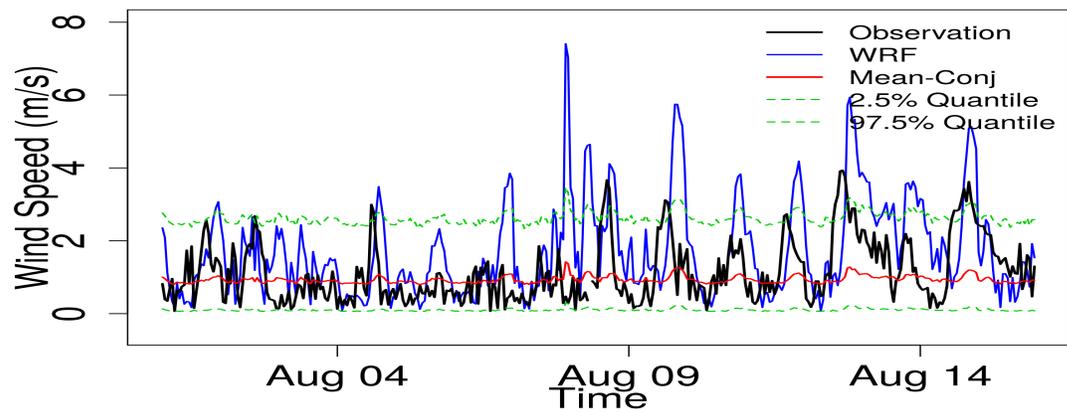
Figure 5: Scatter plots of 2 m temperature (a), 2 m relative humidity (b), and 10 m wind speed (c) from observations and WRF forecasts.



(a)



(b)



(c)

Figure 6: Calibration plots for the period August 1st to 15th at Pudahuel monitoring station: Panel (a) temperature, panel (b) relative humidity panel (c) wind speed. Observations are represented by black lines, the blue lines represent the WRF simulations, the red lines represent the mean fitted curve, and green dashed lines represent the 2.5% and 97.5% quantiles.

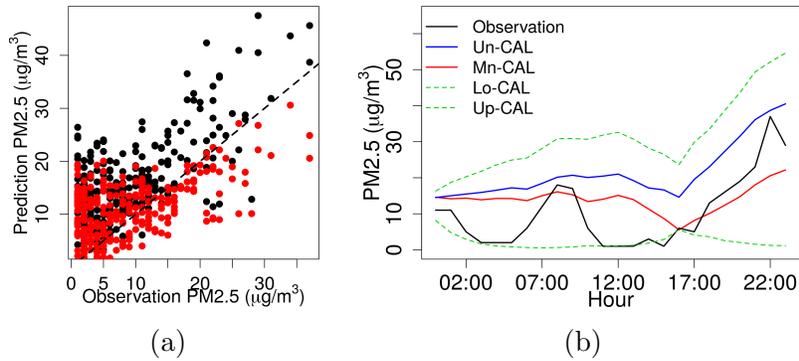


Figure 7: (a) Correlation plot between observations and predictions: calibrated predictions (red points) and un-calibrated predictions (black points). (b) Temporal prediction for Pudahuel. Black line shows the observations; blue line shows the predictions using model Un-CAL; red line shows the predictions using model Mn-CAL. The green dashed lines show the predictions from the Lo-CAL and Up-CAL models, respectively.

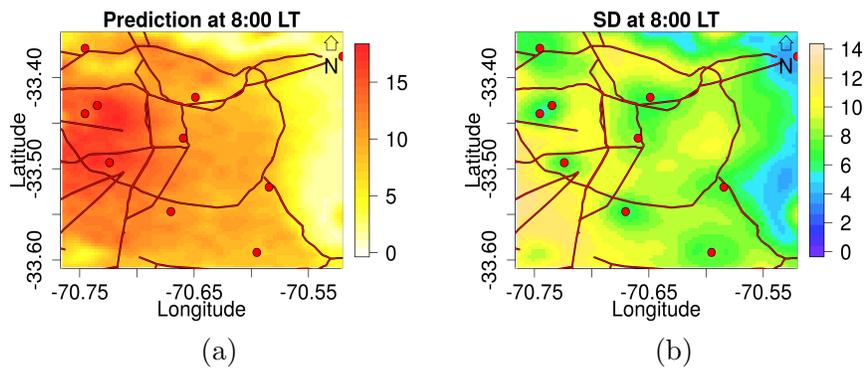


Figure 8: Spatio temporal prediction of PM_{2.5} concentration on August 16th at 8:00 (a), and its standard deviation (b).

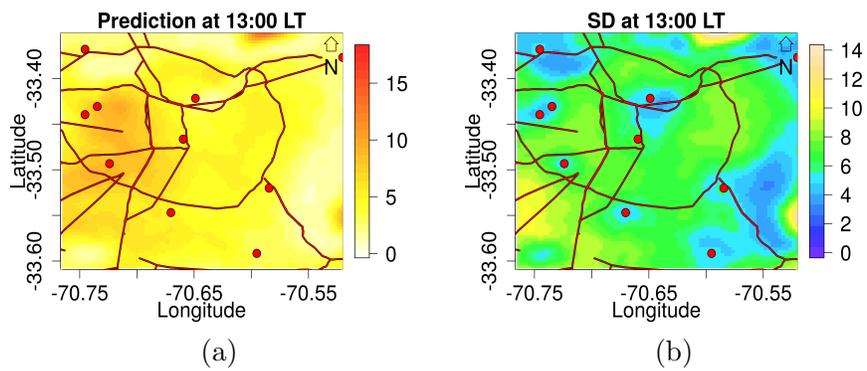


Figure 9: Spatio temporal prediction of PM_{2.5} concentration on August 16th at 13:00 CLST (a), and its standard deviation (b).