AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD023949

Key Points:
- Development and testing for a system predicting severe $PM_{2.5}$ episodes in south-central Chile
- Forecasts are produced with skill up to 3 days in advance
- The system is being implemented in near real time for management of episodes

Correspondence to:
P. E. Saide,
saide@ucar.edu

Citation:
Saide, P. E., M. Mena-Carrasco, S. Tolvett, P. Hernandez, and G. R. Carmichael (2016), Air quality forecasting for winter-time $PM_{2.5}$ episodes occurring in multiple cities in central and southern Chile, *J. Geophys. Res. Atmos., 121*, 558–575, doi:10.1002/2015JD023949.

Received 16 JUL 2015
Accepted 10 DEC 2015
Accepted article online 15 DEC 2015
Published online 15 JAN 2016

Air quality forecasting for winter-time $PM_{2.5}$ episodes occurring in multiple cities in central and southern Chile

Pablo E. Saide[1,2,3], Marcelo Mena-Carrasco[1], Sebastian Tolvett[1,4], Pablo Hernandez[1], and Gregory R. Carmichael[5]

[1]Ministry of the Environment, Santiago, Chile, [2]Center for Climate and Resilience Research, University of Chile, Santiago, Chile, [3]Advanced Study Program and Atmospheric Chemistry Observations and Modeling lab, National Center for Atmospheric Research, Boulder, Colorado, USA, [4]Now at Department of Mechanical Engineering, Universidad Tecnológica Metropolitana, Santiago, Chile, [5]Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa, USA

Abstract Episodic air quality degradation due to particles occurs in multiple cities in central and southern Chile during the austral winter reaching levels up to 300–800 $\mu g/m^3$ hourly $PM_{2.5}$, which can be associated with severe effects on human health. An air quality prediction system is developed to predict such events in near real time up to 3 days in advance for nine cities with regular air quality monitoring: Santiago, Rancagua, Curicó, Talca, Chillan, Los Ángeles, Temuco, Valdivia, and Osorno. The system uses the Weather Research and Forecasting with Chemistry model configured with a nested 2 km grid-spacing domain to predict weather and inert tracers. The tracers are converted to hourly $PM_{2.5}$ concentrations using an observationally based calibration which is substantially less computationally intensive than a full chemistry model. The conversion takes into account processes occurring in these cities, including higher likelihood of episode occurrence during weekends and during colder days, the latter related to increased wood-burning-stove activity for heating. The system is calibrated and evaluated for April–August 2014 where it has an overall skill of 53–72% of episodes accurately forecasted (61–76% for the best initialization) which is better than persistence for most stations. Forecasts one, two, and three days in advance all have skill in forecasting events but often present large variability within them due to different meteorological initializations. The system is being implemented in Chile to assist authority decisions not only to warn the population but also to take contingency-based emission restrictions to try to avoid severe pollution events.

1. Introduction

Cities located in valleys across central (33°S to 37°S) and southern (37°S to 42°S) Chile are often affected by severe air pollution episodes which can be significantly associated with premature mortality and other negative health effects [e.g., *Díaz-Robles et al*., 2015; *Sanhueza et al*., 2009; *Valdés et al*., 2012]. These events are caused by a combination of complex topography, as most of these cities are located in between the Andes and a coastal cordillera, episodic meteorological conditions that produce poor ventilation in the valleys [e.g., *Garreaud et al*., 2002; *Rutllant and Garreaud*, 1995], and emissions due to anthropogenic activities. Historically, most of the attention has been centered on Santiago, the capital of Chile, where an air quality attainment plan has been in place since 1997 for PM_{10}. This plan and its updates include multiple long-term efforts to reduce emissions and the implementation of a requirement to forecast episodes both to warn the public and to try to reduce the episode's impact by invoking temporary measures (e.g., ban wood-burning stoves and restrict private transport and dirtier industrial operations) [*Troncoso et al*., 2012]. The current national $PM_{2.5}$ standard is 50 $\mu g/m^3$ for the 24 h average and episode categories correspond to Alert (80–110 $\mu g/m^3$), Preemergency (110–170 $\mu g/m^3$), and Emergency (>170 $\mu g/m^3$). Thus, forecasting tools are included in the decision-making process, and they are required to predict episodes skillfully for Santiago, which has motivated active research in the subject [e.g., *Cassmassi*, 1999; *Perez and Reyes*, 2002; *Perez and Salini*, 2008; *Saide et al*., 2011b].

Several midsized cities in Chile with similar or higher pollution levels than Santiago have not received as much attention. Therefore, the number of forecasting tools developed for these cities is limited [e.g., *Díaz-Robles et al*., 2008]. However, increased measurement capacities in these cities and the publication of a national $PM_{2.5}$ air quality standard in effect since 1 January 2012 indicate that these cities need

improved tools to support air quality management plans. Although the highest pollution loads in these cities also occur in winter, the sources of pollution contribute differently than in Santiago as in these cities household heating is primarily done by use of wood-burning stoves [*Sanhueza et al*., 2009], which needs to be considered when building prediction systems for these regions.

Since dispersion constraints affecting central and southern Chile (e.g., coastal lows and prefrontal conditions) typically have a scale that can affect multiple cities with severe pollution episodes simultaneously [*Rutllant and Garreaud*, 1995], it would be computationally cost efficient to have one combined episode prediction tool. However, due to the large heterogeneity of these cities (e.g., size, topography, and pollution sources), it is not evident that a single prediction system is feasible to implement. Forecasting pollution at a national scale is a common practice in other countries, where air quality predictions are generally based on chemical transport models (CTMs) or coupled chemistry-meteorology models (CCMMs) [*Baklanov et al*., 2014; *Zhang*, 2008]. Examples of these models include GEM-MACH [*Moran et al*., 2010] for Canada (https://weather.gc.ca/), WRF-CMAQ [*Foley et al*., 2010] for the United States (http://www.airnow.gov/), CCATT-BRAMS [*Longo et al*., 2013] for Brazil and South America (http://meioambiente.cptec.inpe.br/), and multiple ones for Europe (http://atmosphere.copernicus.eu/). CTMs and CCMMs have been used in the Chilean region for multiple applications, including studying transport of pollution [*Gallardo et al*., 2002; *Schmitz*, 2005], evaluating and constraining emission inventories at the city scale [*Jorquera and Castro*, 2010; *Saide et al*., 2011a], designing air quality networks [*Henríquez et al*., 2015], estimating health benefits of pollution control strategies [*Mena-Carrasco et al*., 2012], evaluating regional climate feedbacks [*Mena-Carrasco et al*., 2014], exploring aerosol sources and interactions with clouds in the Southeast Pacific [*Saide et al*., 2012b; *Spak et al*., 2010; *Twohy et al*., 2013; *Yang et al*., 2011], and developing novel data assimilation techniques [*Saide et al*., 2012a]. To our knowledge, there are no studies evaluating CTMs or CCMMs for pollution episodes in Chilean cities other than Santiago. Also, there are no studies assessing the day-to-day variability in wood-burning-stove emissions and their impacts on aerosol concentrations.

As episode categories are based on 24 h running averages and they are often associated with weather conditions that last longer than a day, episodes tend to occur in a consecutive manner, which is why persistence can often be an effective way to forecast pollution events. However, forecasts have to be issued by law no later than 8 P.M. local time; thus, a persistence forecast with a full record of observations is not feasible. As shown in *Saide et al*. [2011b], an advantage of having a forecasting tool able to predict episodes 2 or 3 days in advance (only possible with CTMs or CCMMs) is that it can allow authorities to effectively take contingency measures to try to avoid or ameliorate critical air pollution events. *Saide et al*. [2011b] show that episodes are the result of multiple days of accumulated emissions and that same day emissions can often contribute only a small fraction of the observed pollution during an episode. This allows, for example, decreeing wood-burning bans in advance of a forecasted pollution episode in order to prevent pollution accumulation instead of reacting to episodes that have already unfolded.

This study aims to answer the following scientific questions: Is it feasible to build a near-real-time forecasting system tailored at predicting severe episodes of pollution for multiple cities in central and southern Chile? What are the processes that need to be represented in the system to improve its predictive skill? Can the episodes be predicted more than 1 day in advance and is the skill of the system better than persistence? Can the system help guide future station deployments in cities with no monitors? The methods used in the development of the forecasting system are presented in section 2 and an evaluation of the model and analyses of case studies are discussed in section 3. Conclusions, future directions and needs for further developments of the system are presented in section 4.

2. Methods

In the following sections we describe the different components of the study, including the observations used, the modeling system and its configuration, the emissions driving the modeling system, and the strategy to forecast $PM_{2.5}$.

2.1. Observations

Air quality observations for April–August 2014 corresponding to nine cities (25 stations, Figure 1) in central and southern Chile (10 in Santiago, 2 in Rancagua, 1 in Curicó, 3 in Talca, 2 in Chillan, 2 in Los Ángeles, 3 in

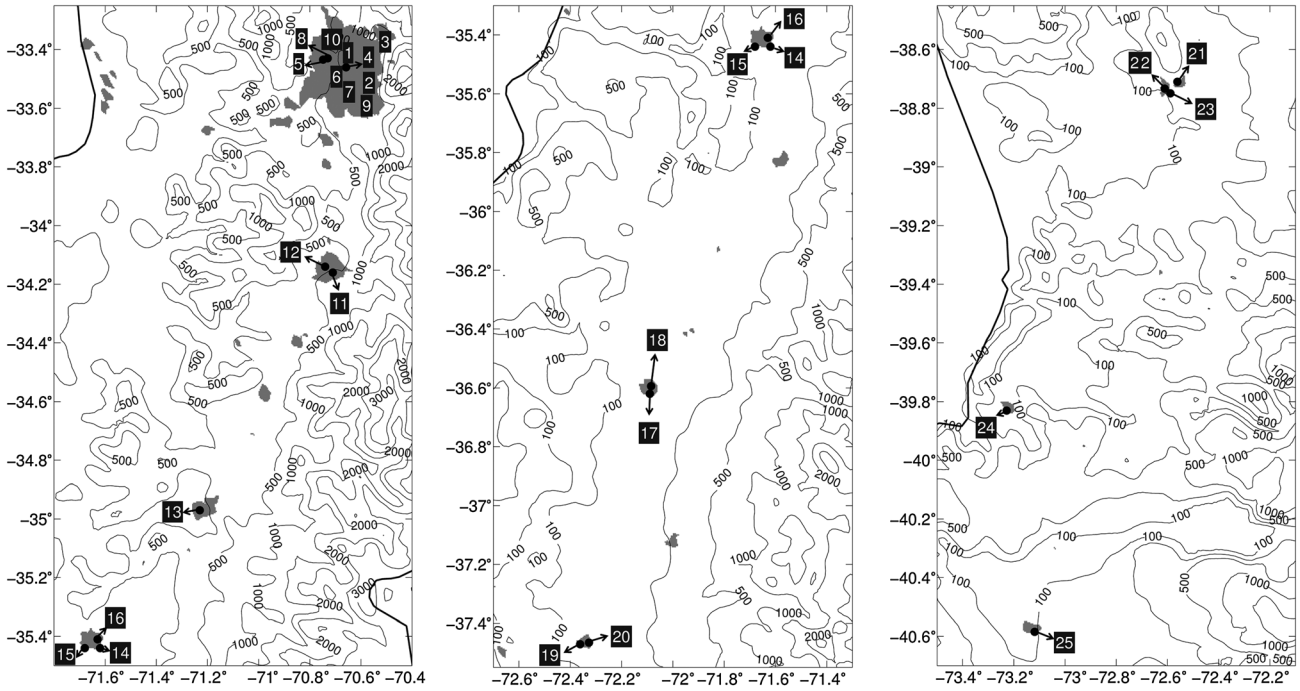

Figure 1. Maps of the Chilean cities and stations used in the study. Urban land use and terrain height (m) in the model are shown in grey and in contours, respectively. Western and eastern thick black solid lines represent the shore line and the border between Chile and Argentina, respectively. Station locations are indicated either with a numbered black square or with black dot, in which case the number is indicated with an arrow. Stations shown are Independencia (1), La Florida Santiago (2), Las Condes (3), Parque O'Higgins (4), Pudahuel (5), Cerrillos (6), El Bosque (7), Cerro Navia (8), Puente Alto (9), Quilicura (10), Rancagua I (11), Rancagua II (12), Curicó (13), U.C. Maule (14), La Florida Talca (15), Universidad de Talca (16), Purén (17), INIA Chillán (18), 21 de Mayo (19), Los Ángeles Oriente (20), Museo Ferroviario (21), Las Encinas Temuco (22), Padre de las Casas II (23), Valdivia (24), and Osorno (25). Cities (and their respective stations) analyzed correspond to Santiago (1–10), Rancagua (11–12), Curicó (13), Talca (14–16), Chillán (17–18), Los Ángeles (19–20), Temuco (21–23), Valdivia (24), and Osorno (25).

Temuco, 1 in Valdivia, and 1 in Osorno) are used in this study. The time the monitors have been in place depends on the site, with the older stations going back to 1997 in Santiago. All the observations are publicly available through the National System of Air Quality Information (http://sinca.mma.gob.cl/) where details on the full suite of instruments and time of deployment can be found. The stations were equipped with at least $PM_{2.5}$ (beta attenuation Met One BAM-1020), relative humidity, temperature, and wind speed (combination of LSI Lastem and Met One instruments) monitors during the study period. Data are available on an hourly basis. We also use observations of $PM_{2.5}$ and CO for 2009, which is the last year with a full record of $PM_{2.5}$ and CO observations for a city other than Santiago (Temuco). All of these stations are maintained by the Chilean Ministry of the Environment. The Superintendence of the Environment oversees quality control and whether these sites are representative of the air quality of the cities in which they are deployed. As beta attenuation monitors have been found to comply with U.S. Environmental Protection Agency $PM_{2.5}$ standards for hourly observations [*Gobeli et al*., 2008], we use in our analysis the thresholds of these standards as conservative lower and upper limits for the measurement uncertainties (i.e., ±10% or ±2 $\mu g/m^3$, whichever deviates more). Uncertainties of temperature and wind speed measurements used in this study are relatively much lower (e.g., LSI LASTEM DNA727 (wind speed) and DMA675 (temperature) used in multiple stations have uncertainties $< = 1.5\%$ and of 0.04°C, respectively); thus, they are not included in the analysis.

Additionally, daily precipitation observations maintained by the Chilean Ministry of Public Works (http://snia.dga.cl/BNAConsultas/reportes) and located in each city studied are used. These correspond to stations Oficinas Centrales DGA (Santiago), Cachapoal-DCP (Rancagua), Curicó, Universidad Católica de Talca, Chillán Viejo, Los Angeles, Temuco Centro, Universidad Austral (Valdivia), and Adolfo Matthei (Osorno).

2.2. Modeling System

The forecast system is designed based on the approach by *Saide et al*. [2011b] in which CO is used as a proxy of $PM_{2.5}$ in lieu of their high correlation during air quality episodes (see section 2.4). The online chemical

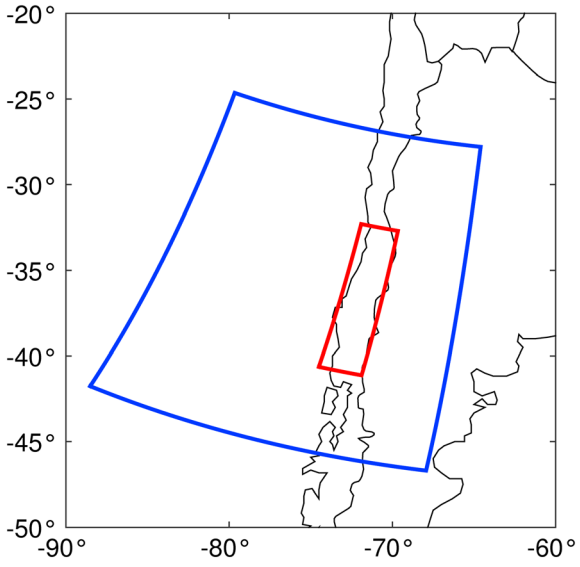

Figure 2. Modeling domains. The outer (dx = 10 km) and nested (dx = 2 km) domains are shown in blue and red solid lines, respectively. All maps in Figure 1 are contained in the nested domain.

and meteorology model Weather Research and Forecasting with Chemistry (WRF-Chem) [*Grell et al.*, 2005; *Skamarock et al.*, 2008] version 3.6.1 is used to advect inert tagged tracers. The domain configuration is shown in Figure 2, with a coarser and inner domain of 10 km and 2 km horizontal grid spacing, respectively, which interact through two-way nesting. The coarse domain coverage is chosen to resolve the regional meteorology and allow a sufficient area to spin-up clouds over the ocean, while the inner domain is focused on central and southern Chile. The standard longitude for the domain projections (equal to −57.0) is chosen to obtain a slightly tilted inner domain in order to have full coverage of central and southern Chile and to minimize the number of grid cells on the west–east axis (see Figure 2). The vertical resolution (39 levels up to 100 hPa) is chosen to have the first layer at 10 m and six levels below 100 m. *Saide et al.* [2011b] provide an evaluation of the parameterizations that can be used for the region; thus, we use a similar configuration: no parameterized horizontal diffusion in the inner domain, Mellor-Yamada Nakanishi Niino (MYNN) boundary layer [*Nakanishi and Niino*, 2004] with no lower limits on the vertical diffusion coefficients, Rapid Radiative Transfer Model (RRTM) longwave radiation [*Mlawer et al.*, 1997], Dudhia shortwave radiation [*Dudhia*, 1989], Noah Land Surface Model [*Chen and Dudhia*, 2001], and WRF Single-Moment 3 class scheme cloud microphysics [*Hong et al.*, 2004]. The Grell-Freitas cumulus scheme [*Grell and Freitas*, 2013] is used for the coarser domain as it is recommended at this resolution for 1–3 day forecasts (http://www2.mmm.ucar.edu/wrf/users/). U.S. Geological Survey (USGS) land use categories were used but the urban land cover was replaced with the Moderate Resolution Imaging Spectroradiometer one [*Friedl et al.*, 2002] as it is outdated in the USGS database [*Saide et al.*, 2011b; *Yu et al.*, 2012]. A similar model configuration was also used in previous studies [*Mena-Carrasco et al.*, 2012, 2014].

Once per day, multiday simulations are performed using an operational basis setup. Initial and boundary meteorological conditions are obtained from National Centers for Environmental Prediction (NCEP) Final Analysis (FNL, http://rda.ucar.edu/datasets/ds083.2/, available every 6 h) and from NCEP Global Forecasting System (GFS, http://rda.ucar.edu/datasets/ds084.6/, available every 3 h). The model is initialized from the 00 UTC FNL analysis and runs for 1 day with FNL boundary conditions. The simulation is then reinitialized from the WRF restart file and the 00 UTC GFS cycle is used as boundary conditions for a 4 day long simulation. Tracer concentrations, soil temperature profiles, and snow accumulations are used as initial conditions for the forecast the next day. On the operational setting, the first 2 days of the simulation account for spin-up and the last 3 days are considered as a forecast. These 3 days are labeled as day 1, 2, and 3 of the forecast. Thus, there are three forecasts for each specific day, which are used for calibration and performance statistics. The extended spin-up setup is based on previous experience of the Chilean Meteorological Service when performing near-real-time air quality forecasts for Santiago [*Delgado et al.*, 2014] using the system described in *Saide et al.* [2011b], as it was found it provided better performance for the first day forecast than initializing from the most current GFS without spin-up. This is consistent with our results for central Chile (see section 3).

2.3. Emissions Inventory

Two tagged tracers are implemented representing sources with strong day-time ("traffic" tracer) and night-time ("wood-burning" tracer) emissions. The traffic tracer emissions of CO are spatially distributed using population maps [*Saide et al.*, 2009; *Tuia et al.*, 2007] and temporally distributed using a diurnal traffic pattern

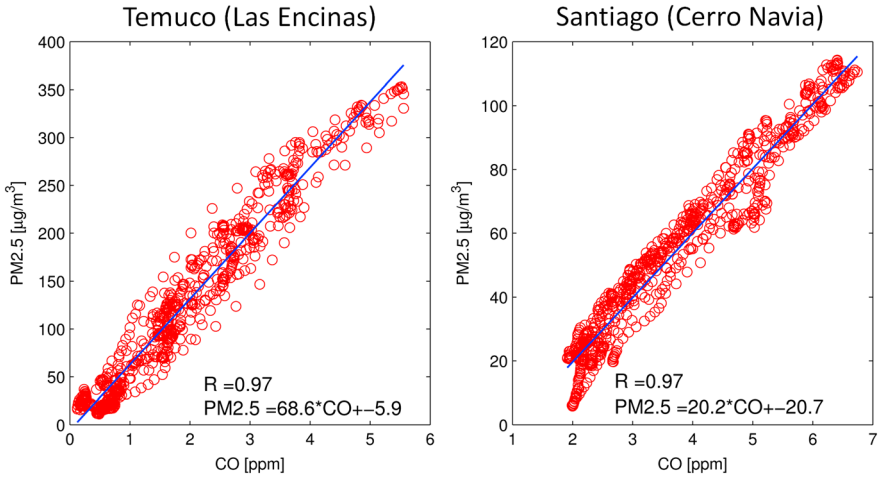

Figure 3. Scatterplots of observed 24 h average CO versus $PM_{2.5}$ for July 2009 in (left) Temuco and (right) Santiago. Regression line (blue solid line), correlation coefficient (*R*), and regression equation are also included in each panel.

provided by traffic authorities, which has been found to provide good agreement to air quality observations [*Saide et al*., 2011b]. Wood-burning-stove emissions are spatially distributed using population for all Chilean administrative divisions (Regions), except for the Metropolitan Region (which includes Santiago), where the distribution is based on the number of houses with wood-burning stoves per municipality obtained through surveys [*CDT*, 2012]. Two types of diurnal cycles are used for the temporal distribution of wood-burning-stove emissions. For cities in central Chile (Santiago, Rancagua, Talca, and Curicó) a diurnal cycle with only nighttime emissions is applied (peak use from 6 P.M. to 2 A.M. local time (LT), no emissions from 4 A.M. to 4 P.M. LT). For cities in southern Chile (including Chillán, Los Angeles, Temuco, Osorno, and Valdivia) wood-burning stoves are often active throughout the night and some are kept on during the whole day; thus, the cycle is modified by increasing the length of the peak hour window (6 P.M. to 4 A.M. LT) and by setting 20% of the peak activity from 6 A.M. to 4 P.M. LT. Using these spatial and temporal profiles, the same total amount of CO emissions by Region is distributed for each tagged tracer. The contribution of the tracers to modeled $PM_{2.5}$ in each city is set later in the calibration stage. CO emissions for each Region were obtained from estimates made for supporting Chilean air quality standards [*DICTUC*, 2011]. These correspond to 37 kton/yr for Region VI (Rancagua), 40 kton/yr for Region VII (Curico, Talca), 27 kton/yr for Region VIII (Chillan, Los Angeles), 26 kton/yr for Region IX (Temuco), 20 kton/yr for Regions X and XIV (Osorno and Valdivia), and 106 kton/yr for the Metropolitan Region.

2.4. $PM_{2.5}$ Forecasting

Figure 3 shows strong correlations ($R = 0.97$) between 24 h averages of observed $PM_{2.5}$ and CO for Temuco and Santiago. Although there is evidence of a substantial contribution of secondary aerosol to the particulate matter budget [*Jorquera and Barraza*, 2012; *Villalobos et al*., 2015], the strong correlation between CO and $PM_{2.5}$ for concentrations during episodes indicates that a proper episode forecast can be achieved by modeling an unreactive and primary-emitted pollutant (CO at the city scale) and using an empirically based conversion to obtain $PM_{2.5}$ concentrations [*Saide et al*., 2011b]. Therefore, computing power is dedicated to resolving the transport of passive tracers at high resolution (2 km) for a large area as opposed to full chemistry and aerosol simulations which are more computationally expensive.

The tagged tracers' concentrations representing traffic (Tr_T) and wood burning (Tr_WB) are converted to hourly $PM_{2.5}$ concentrations using the following equation:

$$\mathrm{PM2.5}_{t,s} = \mathrm{Tr\_to\_PM}_s * \mathrm{Weekend\_eff}_{t,s} * \left[F\_WB_s * Teff\_WB\left(T_{\mathrm{mean}_{t,s}}\right) * \left(\max_{i \in s \cup N_s} Tr\_WB_{t,i} \right) + (1 - F\_WB_s) * \left(\max_{i \in s \cup N_s} Tr\_T_{t,i} \right) \right], \quad (1)$$

where the indices *t* and *s* represent time and station, respectively, Tr_to_PM is the factor that converts tracer to $PM_{2.5}$ concentrations for each station, Weekend_eff is an inflation factor to account for differences within

weekday and weekend activity, *F*_WB denotes the fraction contributed by wood-burning stove sources to total concentrations, Teff_WB represents the fraction of wood-burning stoves started when the temperature drops to *T*_mean (running mean centered in the time *t*), and N_s is the set of eight spatial neighbors of each station *s* in the model grid. Each of these components is explained in detail in section 3.2. In equation (1), the maximum concentration within the model grid cell where each station is located and the neighboring grid cells ($s \cup N_s$) is used to take into account temporal (plume might be shifted in time) and spatial (plume might miss a station) modeling errors and has been found to improve model performance [*Kang et al*., 2007]. On the other hand, this choice might generate false alarms if a plume missed a monitor, in which case predicting an episode would still be desired to alert population or implement contingency measures.

We developed a calibration strategy oriented toward fitting episode statistics rather than hourly PM concentrations [*Saide et al*., 2011b], as predicting episodes is the main objective of this system. If Tr_to_PM factors are too small then they would produce a poor fitting to the observed episodes due to the lack of episodes forecasted by the model, while factors that are too large would better predict episodes but with too many false alarms. To find a balance, we chose the Tr_to_PM factor for each station that generates the same number of episodes (i.e., the total number of alerts, preemergencies and emergencies, all summed together) in the model as in the observations for the period studied (April–August 2014). However, as the forecast system produces three forecasts for each day, the criteria for calibration is to find the factors that match the number of observed episodes with the average number of episodes obtained by the three forecasts. The factors are adjusted for each station as there could be emission gradients and subgrid-scale processes not resolved by the 2 km resolution of the model which would not be considered if the calibration was made for each city.

A simple and efficient algorithm was developed to find the tracer to PM factors (Tr_to_PM) for given values of the other factors (Weekend_eff, *F*_WB, and Teff_WB). First, the number of modeled episodes (mean across the three forecasts) generated by three selected Tr_to_PM factors is computed. Two of these factors are selected as very low (10 $\mu g/m^3$/ppm) and very high (2000 $\mu g/m^3$/ppm) values so the optimal factor would be contained in between and the third factor is selected as the middle point between these two. Then, if the number of observed episodes is below (above) the number of modeled episodes obtained by the middle point, the lowest (highest) and middle point factors are selected as lower and upper thresholds for the next iteration and the process just described is repeated. At any iteration, if the number of observed episodes is equal to the modeled one for the middle point factor, then the algorithm has reached convergence. As the number of modeled episodes monotonically increases as Tr_to_PM increases, the solution is going to be given by a single range of factors, from which we chose the first value found by the iterative process. The Weekend_eff, *F*_WB, and Teff_WB factors are chosen by changing one at a time and computing a new set of Tr_to_PM factors (with the algorithm just described) until the number of episodes accurately predicted by the model (i.e., number of "hits") is maximized.

3. Results and Discussion

Simulations were performed using the configuration described in section 2. In the next sections these simulations were used along with observations to evaluate the meteorological performance of the forecasts, calibrate the $PM_{2.5}$ model, assess the episode predictive skill, and study the evolution of severe pollution events.

3.1. Evaluation of Meteorological Forecasts

It is important to evaluate the accuracy of the temperature forecasts over the cities studied as they are used to postprocess the wood-burning-stove tracer (see section 3.2). Figures 4a and 4b show the time series of 24 h running mean temperature for a station in Temuco as a representative example for cities in southern Chile. The three forecasts follow the observations generally with good skill (mean bias and error for all forecasts are ~−0.6 C and ~1.0 C, respectively). There are periods of 2 to 7 days where performance seems to degrade persistently for all forecasts, probably due to biases in the global model forcing the simulations. These multiday persistent biases need to be taken into consideration by the forecaster when producing a forecast in near real time as they could impact episode forecasts for places with large wood-burning-stove emission contributions (see next section).

A good indicator of how well the forecasts follow the temporal trends in the observations is the correlation coefficient (*R*), shown in Table 1 along with mean bias for meteorological variables measured at the most

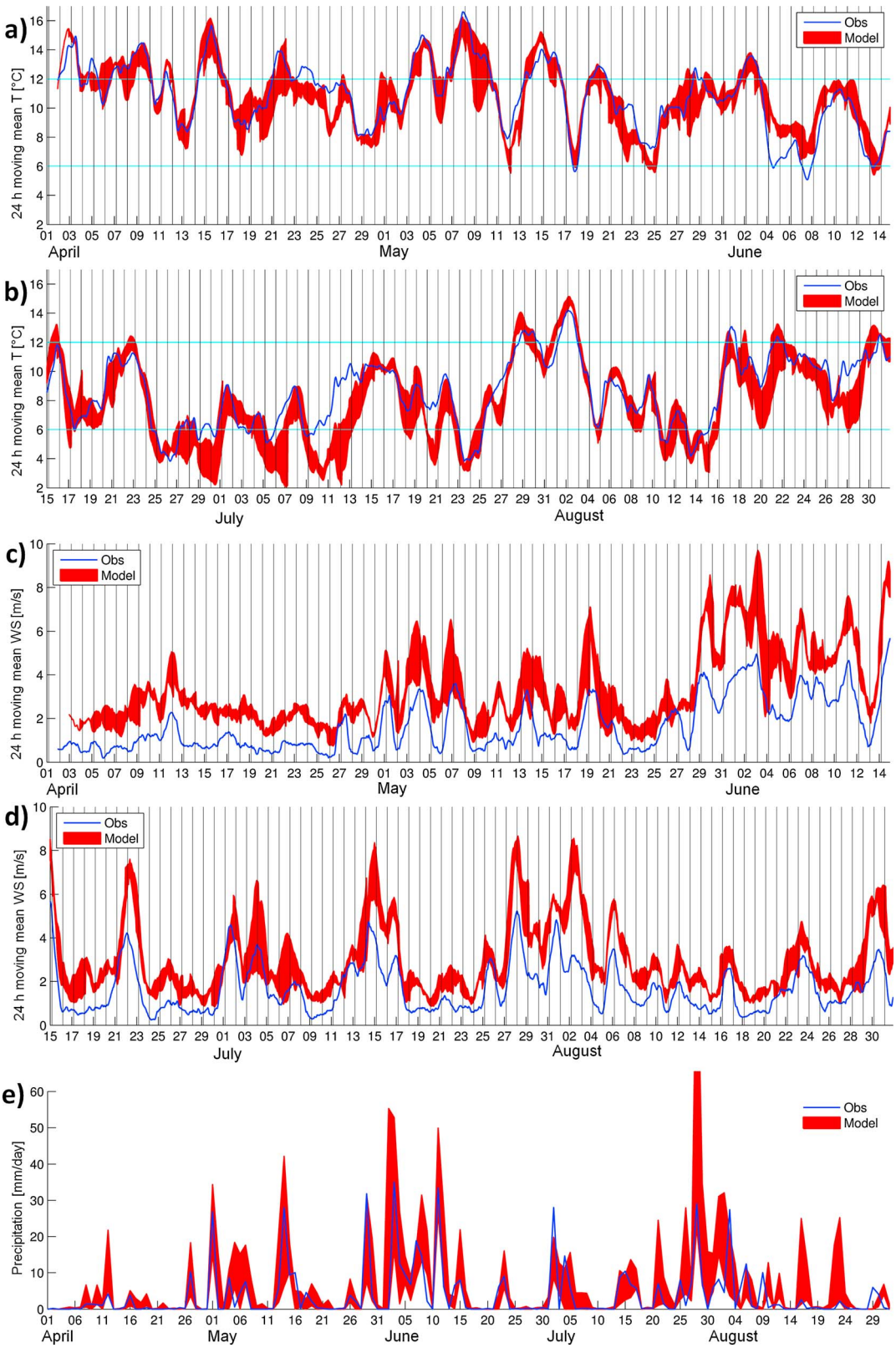

Figure 4. Time series of (a, b) 24 h running mean temperature and (c, d) wind speed for station Padre de las Casas II (Temuco) and (e) daily precipitation in Temuco Centro. Figures 4a and 4c show the time series for the first half of the study period, while Figures 4b and 4d for the second half. The solid blue line represents the observations and the red region represents the variability in the three forecasts available for each day. Cyan horizontal lines (a, b) represent the thresholds used for turning all wood-burning stoves on (6°C) and off (12°C) in the model (see section 3.2.3). Vertical lines represent the start of each day in local time.

Table 1. Correlation Coefficient (*R*) and Mean Bias (MB) Between Observed and Modeled Variables (Temperature, Wind Speed, and Precipitation) Split by Different Model Initialization Times (Day 1, Day 2, and Day 3)[a]

		Temperature 24 h Mean						Wind Speed 24 h Mean						Daily Precipitation					
		R			MB (°C)			*R*			MB (m/s)			*R*			MB (mm/day)		
City	Station	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
Santiago	Cerro Navia	0.87	0.84	0.82	2.37	2.37	2.50	0.55	0.59	0.51	0.57	0.53	0.48	0.75	0.82	0.38	0.29	−0.08	−0.47
Rancagua	Rancagua II	0.82	0.75	0.71	1.65	1.67	1.67	0.63	0.64	0.62	−0.07	−0.13	−0.11	0.50	0.44	0.27	0.51	0.25	0.46
Curico	Curico	0.85	0.82	0.79	−0.16	−0.31	−0.53	0.79	0.79	0.75	0.52	0.50	0.47	0.59	0.69	0.68	0.15	−0.07	−0.52
Talca	La Florida	0.88	0.84	0.78	−1.42	−1.73	−2.12	0.81	0.74	0.72	0.83	0.78	0.75	0.69	0.56	0.79	0.35	0.08	−0.35
Chillan	Puren	0.88	0.87	0.85	−0.66	−0.79	−1.00	0.84	0.81	0.73	0.70	0.71	0.65	0.59	0.63	0.51	0.50	0.35	−0.32
Los Angeles	21 de Mayo	0.86	0.84	0.83	−0.86	−1.01	−1.14	0.88	0.85	0.76	0.96	0.94	0.87	0.70	0.64	0.58	−0.47	−0.46	−0.90
Temuco	Padre Las Casas II	0.89	0.88	0.88	−0.32	−0.41	−0.50	0.90	0.85	0.81	1.45	1.38	1.27	0.59	0.62	0.56	1.57	1.25	0.65
Valdivia	Valdivia	0.91	0.88	0.86	0.75	0.71	0.62	0.84	0.82	0.78	1.23	1.16	1.13	0.31	0.27	0.21	3.41	2.68	3.01
Osorno	Osorno	0.88	0.85	0.84	0.50	0.42	0.42	0.86	0.82	0.77	1.83	1.74	1.74	0.65	0.62	0.57	0.01	0.12	−0.08

[a]Temperature and wind speed statistics are shown for the air quality stations indicated in the second column, which corresponds to the most polluted station in each city studied, while precipitation statistics are shown for the stations available in each city (see section 2.1).

polluted station for each city studied. Large correlations between model and observations are found for temperature for all cities and initializations (0.71 to 0.91) showing that temperature forecasts can be used with confidence when modeling the wood-burning-stove tracer. Also, the correlation and mean bias for temperature is similar for all three forecast days at each station and generally slowly degrades for earlier initialization times. This is encouraging as a future implementation of the forecasting system could use the previous day forecast to compute temperature dependent wood-burning-stove emissions.

Another important meteorological variable to assess for air pollutant transport is the wind speed, as large values will result in the ventilation of the region, while low values will favor the accumulation of pollutants. This is depicted by the correlations found between observed $PM_{2.5}$ and wind speed (*R* ranges from 0.49 to 0.62 for the 24 h mean). Figures 4c and 4d show 24 h running average wind speed. Modeled wind speed is biased high for most sites (bias ranges from 0.2 to 1.8 m/s for stations with a full record) which is typically found for WRF simulations in complex terrain if the effects that the unresolved topographic features exert over the momentum flux are not parameterized [*Jiménez and Dudhia*, 2011]. A parameterization for topographic effects is not included in the configuration as it is currently not implemented for the boundary layer scheme used. Also, the model is not configured to include an urban canopy parameterization (detailed urban land cover was not available) which could contribute to the bias in wind speed as the air quality stations are located within city limits and observe local winds. Parameterizations that improve the representation of urban canopy and topographic effects should be explored in the future to assess if it helps in reducing the wind speed bias. Regardless of these biases, the model shows good representation of strong and calm wind events; it is able to represent the observed correlations between $PM_{2.5}$ and wind speed (modeled *R* range is 0.48–0.62 for day 1 forecasts), and, as will be shown in section 3.3, it presents good skill for predicting pollution episodes.

Wind speed performance shows distinct patterns that differ from those for temperature (Table 1). First, the range of correlations is wider (0.51 to 0.9) with a tendency toward lower correlations for cities in central Chile and increasing toward the south. Part of this difference is due to larger observed wind speeds in the south, which the model is able to resolve better compared to low wind speeds (e.g., *R* drops from 0.9 to 0.77 for the first day forecast in Padre de las Casas II

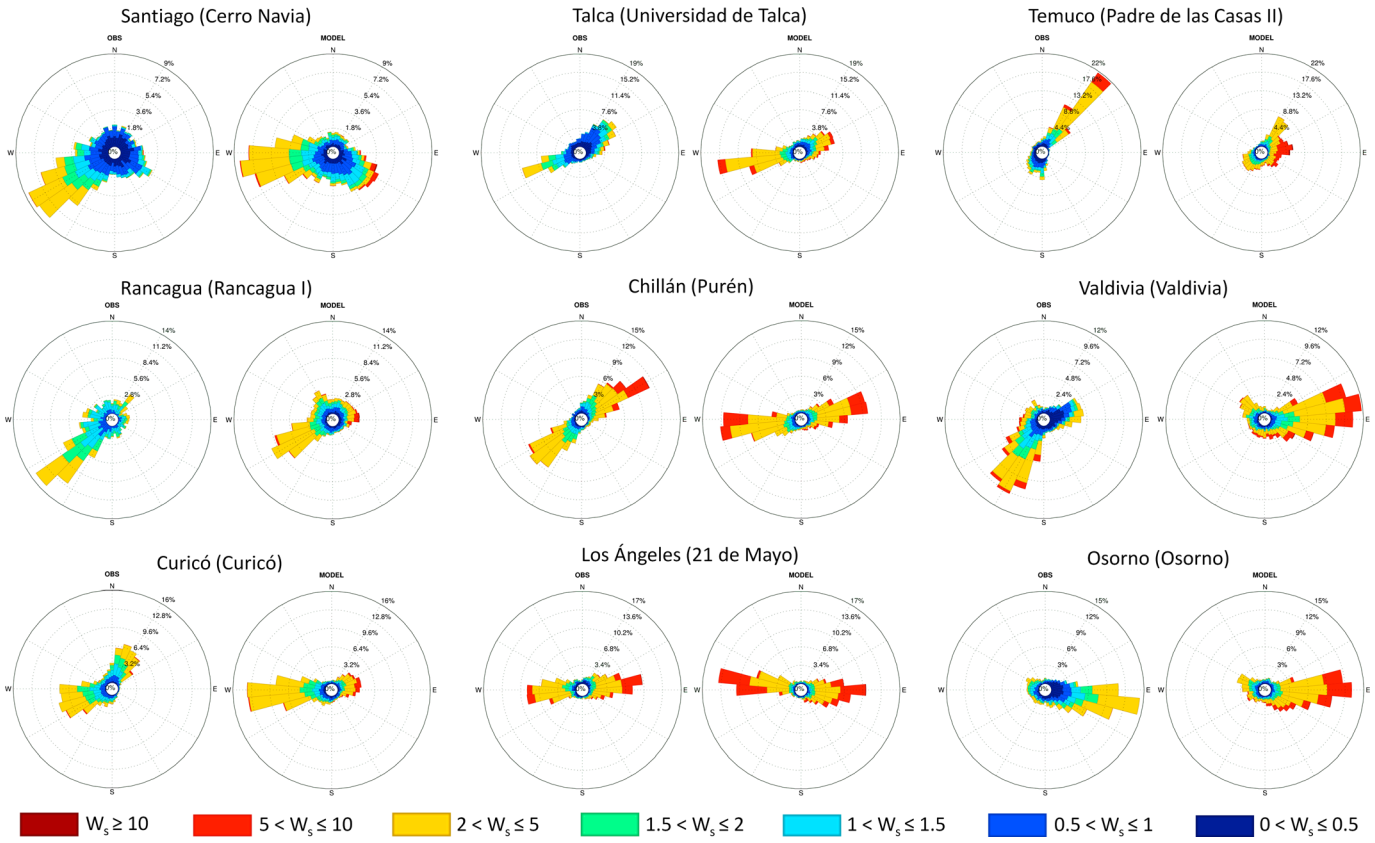

Figure 5. Observed and modeled (day 1 initialization) wind roses for selected stations at each city studied. Wind speeds (W_s) in m/s. Stations were selected as the ones with the larger amount of episodes in each city when its wind data was available for the whole period. When record was incomplete, another station was selected. Stations in Rancagua show wind records only for July–August.

when considering only wind speeds lower than 2.5 m/s). For instance, wintertime 24 h mean wind speed in Santiago (Cerro Navia) is always lower than 2.5 m/s for the period analyzed, while this value is commonly exceeded in the cities to the south. However, even after computing correlations for observations below 2.5 m/s, stations in the south still show better correlations than the ones in central Chile (R for all forecasts in Santiago and Rancagua is lower than the R of 0.77 cited earlier for Temuco). The relatively less complex topography toward the south (Figure 1) which the model is able to better resolve might explain the better performance in the south. Second, there can be substantial differences in wind speed performance for different initializations in each station. From Talca to the south there is a clear decreasing trend in performance for forecasts initialized earlier, while for Santiago and Rancagua the forecast for day 2 shows better performance than for days 1 and 3. This behavior is also found for most stations in Santiago (only Cerro Navia shown in Table 1, not shown for the rest) and will have implications in the episode forecasting skill (section 3.3). As stated in the methods section, the system performs 2 days of spin-up. Although this is beneficial for predictions in Santiago and Rancagua (likely due to the need for spin-up to let the high resolution model better resolve the more complex topography and land-sea transitions), future implementations should assess if simulations with less spin-up (i.e., initialized directly from GFS) produce improved forecasts for the cities in the south due to the better performance shown by the most current forecasts in these cities.

Wind roses (Figure 5) can be used to assess the performance of modeled wind speed and direction. Modeled wind roses do not vary significantly with initialization (not shown), so only day 1 forecasts are shown in Figure 5. In general, the area covered by the $<$1 m/s wind speeds in the observations is larger than in the model, which is indicative of the high bias in wind speeds mentioned earlier. The model captures the predominant wind directions in each station within 30° with two exceptions: Curicó and Valdivia. For Curicó there is a large bias in the representation of the northerly winds, which the model shows as easterly. This is likely because Curicó station is located ~300 m to the northwest of an isolated hill found within the city boundaries

(Condell Hill, ~80 m high) which the model is not able to resolve. The model shows poor skill in representing the wind directions for Valdivia, likely due to the inability to reproduce the topography at 2 km resolution as Valdivia is surrounded by steep hills (Figure 1). Also, Valdivia is the only city studied that is next to the ocean; thus, problems resolving the sea-land breeze through the river connecting the ocean and the city might also contribute. Biases in wind direction in other stations are also likely related to the representation of topography; thus, future configurations should explore if increasing model resolution resolves these biases and improves air quality forecasts.

In Chile, rain is usually associated with frontal passages which tend to bring marine air into the basins resulting in a reduction of pollution, which is why it is also important to evaluate the model skill to capture precipitation. Figure 4e shows that for Temuco the model captures most rain events with precipitation amounts usually within the spread of the initializations, which is the case for all cities. In terms of the correlations (Table 1) they tend to show large spread (0.21–0.82). Rain occurs less often in central Chile compared to the south. For instance, daily totals of over 10 mm were recorded only on 6 days in Santiago and Rancagua for the whole period, while to the south it happened on 15–28 days. This is why there is large variability in the correlation coefficients from different initializations in Santiago and Rancagua, as R is going to be driven by the skill in predicting a few large values. To the south (Curico-Osorno) the spread in correlation values among different initializations is less and there is no clear trend on which initialization shows the best skill. Similarly to wind direction, Valdivia shows the worst skill likely for the same reasons (very complex topography and proximity to the ocean).

3.2. Processes Affecting the $PM_{2.5}$ Forecast Skill

3.2.1. Traffic and Wood-Burning Tracers Contribution

According to previous studies, the contribution of wood-burning-stove emissions to $PM_{2.5}$ in winter is estimated to be in the range of 10–40% for Santiago [*Jorquera and Barraza*, 2012; *USACH*, 2014; *Villalobos et al*., 2015] and up to 90% for cities in southern Chile [e.g., *Sanhueza et al*., 2006]. The difference in contribution of sources is also evident when comparing the slopes of the CO to $PM_{2.5}$ relationship (Figure 3), where the slope for Temuco is more than three times larger than for Santiago, as the $PM_{2.5}$ to CO emission ratio is much larger for wood-burning stoves than for other sources [*CONAMA*, 2007]. Thus, the contribution of wood-burning to total concentrations (F_WB) by city needs to be considered in the $PM_{2.5}$ computation. By applying the algorithm described in section 2.4, we found a progressive increase in F_WB from north to south, with values of 0.3 for Santiago and Rancagua, 0.8 for Curicó, Talca, Chillán, and Los Ángeles, and 0.9 for Temuco, Valdivia, and Osorno. Compared to *Saide et al*. [2011b], the wood-burning smoke contribution remains the same for Santiago.

3.2.2. Weekend Effect

There is higher likelihood of episode occurrence during weekends compared to weekdays for multiple cities in central Chile for the period studied. For instance, considering the number of episodes for all stations in a city, there were 137%, 34%, and 38% more episodes on average for weekend days than for weekdays in Santiago, Rancagua, and Talca, respectively. The cities in the south do not show this behavior (e.g., for Temuco there is only a 2% difference) probably because the main source of pollution comes from wood-burning stoves. The weekend inflation factors are obtained by applying the calibration approach with values of 1.4 for Santiago, 1.2 for Rancagua, 1.3 for Talca and Chillán, and 1 for the rest of the cities. Weekend inflation factors are incorporated into the $PM_{2.5}$ model (equation (1)) and applied from 6 P.M. LT on Friday to 6 P.M. LT on Sunday to scale activity, which is hypothesized to increase during Friday and Saturday nights and to play a major role in episode occurrence [*Saide et al*., 2011b]. The skill in predicting episodes improves when including a weekend inflation factor, as the overall (i.e., all forecasts combined) percentage of correctly forecasted episodes increases from 67% to 72% in Santiago, 64% to 69% in Rancagua, 57% to 63% in Talca, and 63% to 67% in Chillan for the most polluted stations.

3.2.3. Temperature Effect

The use of wood-burning stoves in central and southern Chile is mainly related to heating [*Gómez-Lobo et al*., 2006]. Thus, it is expected that pollution episodes are associated with colder days due to larger wood-burning emissions. This is shown in Figure 6a for Temuco (Padre de las Casas II), where significant differences (statistical testing was performed using a two-sample t test to assess statistical differences between temperature distributions at the $p = 0.05$ level) in temperature distributions are found for days with episodes compared to all days ($p = 0.0001$), with temperatures being shifted toward lower values for episode days. On the other

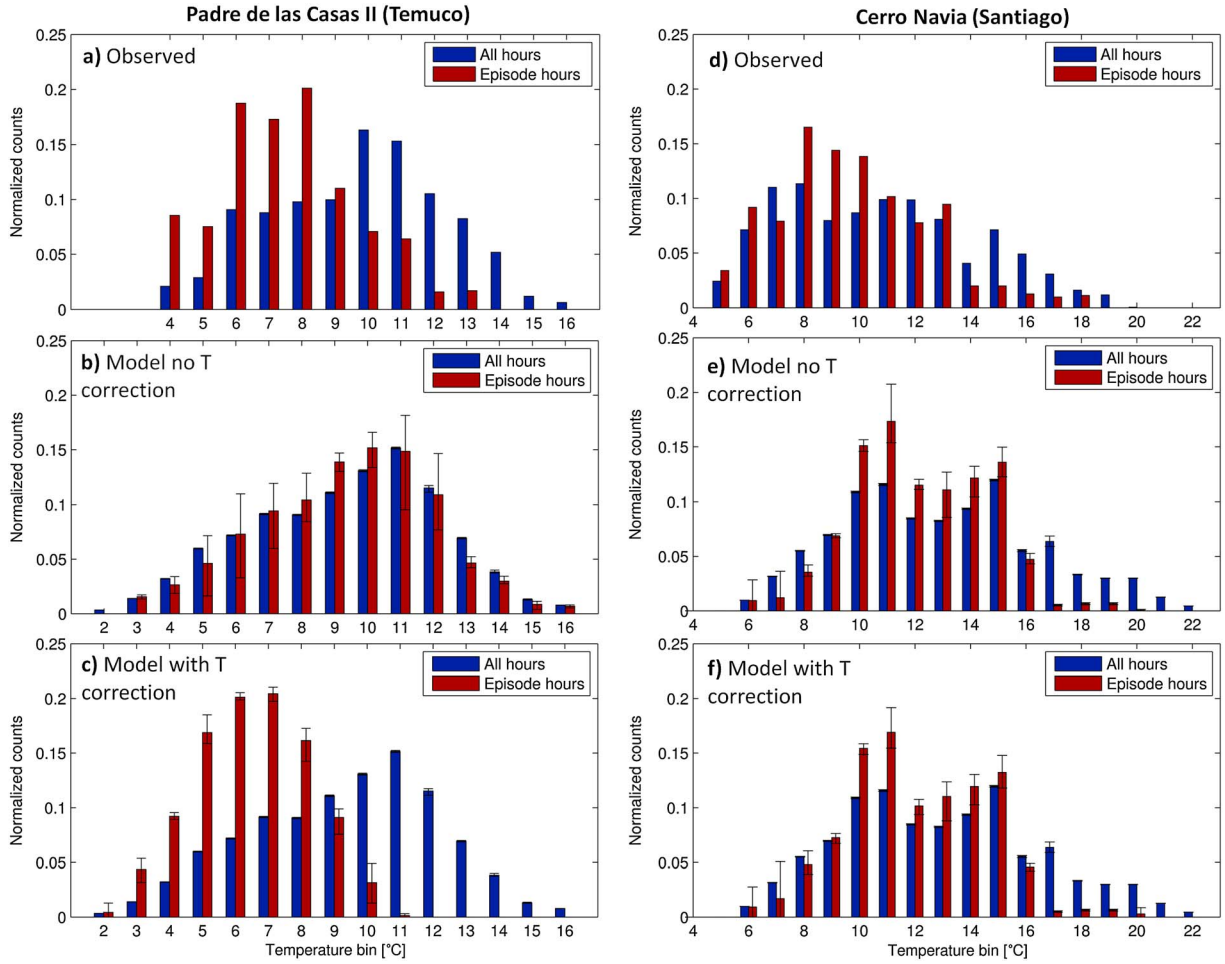

Figure 6. Normalized histograms of temperature for all hours (blue) and for hours with $PM_{2.5}$ episodes (concentrations over 80 $\mu g/m^3$, dark red) at (a–c) Padre de las Casas II station in Temuco and (d–f) Cerro Navia station in Santiago. The upper, middle, and bottom rows show the histograms for the observation, model without temperature correction and model after the temperature correction. All three forecast days are included in the modeled histograms, with the bars and error bars representing the mean and the spread, respectively.

hand, the modeled distribution of temperature for all times and during episodes is not significantly different ($p = 0.32$) when a temperature correction is not included (Figure 6b), which suggests that episodes at low temperature are associated not only with a large-scale stability condition but also to increased emissions during these days.

The temperature dependence is included by applying a factor (Teff_WB in equation (1)) to the wood-burning tracer concentrations to model the fraction of stoves that are started as a function of ambient temperature. This factor is not applied directly to emissions as a temperature forecast would be needed to estimate emissions before starting the simulation. Also, applying the factor to concentrations instead of emissions makes the calibration process faster as it is done in a postprocessing step. Otherwise, if the factor were to be applied to emissions, simulations for the 5 months of study would have to be performed for each Teff_WB tested in the calibration stage, which would be restrictive in terms of the computational resources. The temperature running mean (*T*_mean) is used to smooth the temperature diurnal cycle and to buffer the response of changes in wood-burning stove use to sharp temperature changes, as there is a delay between the change in ambient temperature and the action to turn on or off wood-burning stoves.

Teff_WB is equal to 1 (all stoves turned on) for low temperature values up to a minimum temperature threshold, then it decreases down to a maximum temperature threshold, where Teff_WB = 0 (all stoves are off) for all temperatures above that threshold. The minimum and maximum temperature thresholds and the shape of Teff_WB for connecting them are found by using the strategy explained in section 2.4. The possible values

Table 2. Summary of Tracer to PM Factors (Tr_to_PM, in μg/m^3/ppm), Episode Prediction Skill (Percentage of Episodes Accurately Predicted) and False Alarm Rate for Stations With More Than 20% of the Days With an Episode[a]

			Episode	% Correct Episode					% False Alarms				
City	Station	Tr_to_PM	Days (%)	Day 1	Day 2	Day 3	All	Pers	Day 1	Day 2	Day 3	All	Pers
Santiago	Pudahuel	57	21%	63%	72%	69%	68%	56%	23%	30%	41%	32%	44%
	Cerro Navia	60	30%	71%	69%	76%	72%	67%	20%	30%	33%	28%	33%
Rancagua	Rancagua II	64	35%	68%	72%	68%	69%	64%	23%	33%	35%	31%	36%
Curico	Curico	236	21%	61%	48%	48%	53%	42%	42%	50%	50%	47%	58%
Talca	La Florida	188	38%	66%	64%	59%	63%	57%	39%	35%	37%	37%	42%
Chillan	Puren	938	40%	65%	67%	62%	64%	65%	33%	37%	37%	36%	35%
Los Angeles	21 de Mayo	1273	33%	67%	67%	65%	67%	63%	31%	33%	36%	33%	37%
Temuco	Museo Ferroviario	563	33%	73%	69%	71%	71%	63%	23%	32%	30%	29%	37%
	Las Encinas	358	33%	68%	62%	68%	66%	64%	33%	37%	32%	34%	36%
	Padre Las Casas II	343	38%	73%	71%	68%	71%	73%	27%	29%	32%	29%	27%
Valdivia	Valdivia	404	31%	59%	61%	61%	60%	52%	45%	36%	38%	40%	47%
Osorno	Osorno	613	38%	71%	64%	63%	66%	70%	30%	33%	39%	34%	30%

[a]Days 1, 2, and 3 represent the different initialization times, while "All" shows the overall skill. The persistence forecast skill is also shown (Pers).

of the temperature thresholds are selected by comparing the histograms for all and episode hours (Figure 6a), with upper (10–14°C) and lower limits (4–8°C) when a small and a large fraction of episodes were observed, respectively. The temperature thresholds that provide the best model skill are 6°C and 12°C, while the shape selected was parabolic with its symmetry axes at 12°C (linear and inverse exponential shapes were also tested). Thus, the temperature correction factor is defined as

$$\text{Teff\_WB} = \begin{cases} 0.0278 \times T\_\text{mean}^2 - 0.6667 \times T\_\text{mean} + 4.0 & 6 < T\_\text{mean} < 12 \\ 1 & 6 \geq T\_\text{mean} \\ 0 & T\_\text{mean} \geq 12 \end{cases}.$$

As seen in Figure 6c, the modeled temperature histograms for all hours and episode hours are significantly different after applying Teff_WB ($p = 0.0002$) and show the same trend of shifting episodes toward lower temperatures as in the observed histograms (Figure 6a). Also, observed and modeled temperature histograms for episode days are very similar to each other after the temperature correction (i.e., distributions go from significantly different ($p = 0.0003$) to no significant differences ($p = 0.6$)). Significant improvements are found in forecasting skill when including the temperature correction for wood burning. For instance, the percentage of episodes accurately predicted in Temuco (Padre de las Casas II) increases from 45% to 71% considering all forecast days, showing the importance and power of including the temperature dependence of wood burning in regions which rely heavily on it as fuel for heating.

On the other hand, Figures 6d–6f show the histograms for Cerro Navia station in Santiago. The model histograms are shifted toward lower temperatures compared to the observed ones (which generate significant differences between model and observations, $p < 0.0001$) because model temperatures are biased high (see also Table 2). This bias is a general trend for all stations in Santiago (1–3°C bias) and occurs likely due to an overestimation of the sensible heat flux over large urban areas in the NOAH land surface model [*Lee et al*., 2011]. The two observed histograms for all times and episode times (Figure 6d) are much more similar to each other than the ones in Temuco (Figure 6a). However, there are still significant differences within the observed histograms for Santiago (Figure 6d, $p = 0.0044$), as episodes are more frequent for middle temperatures (8–10°C), while they are much less likely to happen for the higher temperatures (14–19 C). Despite there being a high bias, the same trend is shown by the two model histograms of temperature that do not include a temperature correction for wood burning (Figure 6e, $p < 0.0001$), i.e., episodes are more and less common at middle (8–14°C) and higher (17–22°C) temperatures, respectively. Thus, in Santiago the differences between observed temperature histograms for all times and episode times are likely not related to greater wood-burning emissions occurring on colder days as in the south but to specific synoptic patterns controlling episode conditions [e.g., *Garreaud and Rutllant*, 2003; *Garreaud et al*., 2002]. Also, there is little change in the histograms when applying the temperature correction (Figure 6e versus Figure 6f) as wood-burning emissions in Santiago are considered to contribute 30% to $PM_{2.5}$ forecasts (see section 3.2.1).

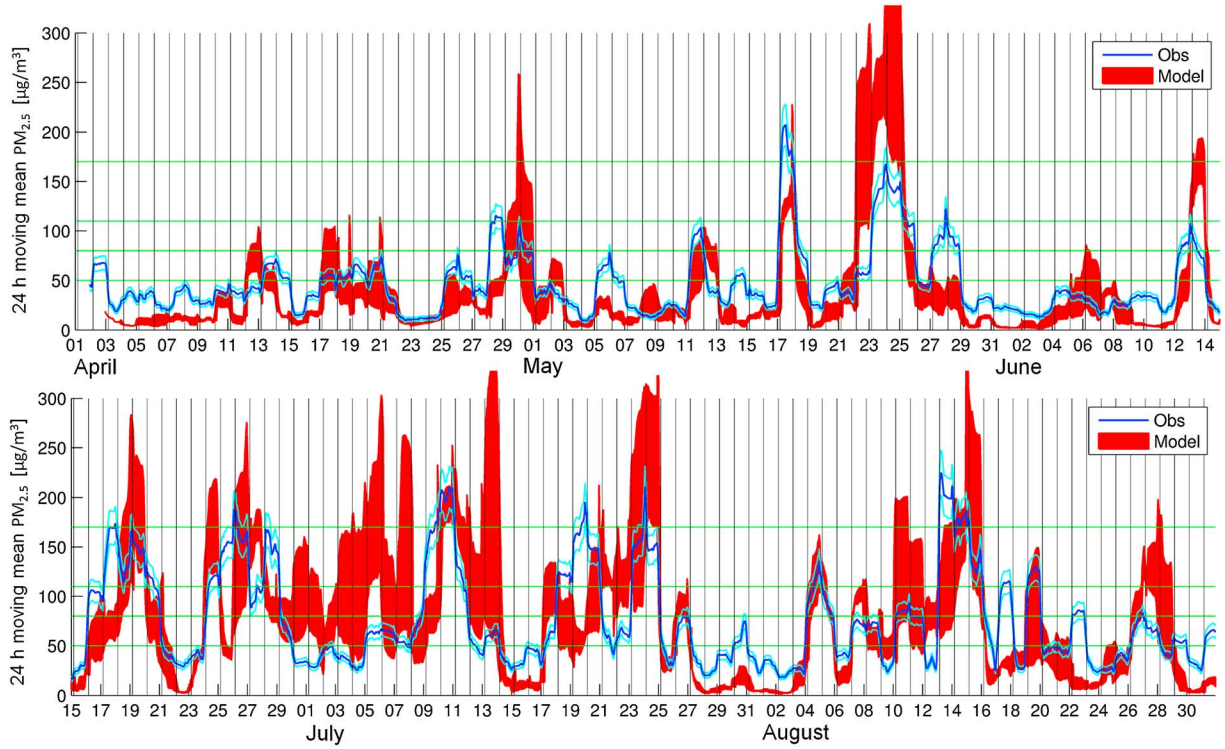

Figure 7. As Figure 4 but for 24 h running mean $PM_{2.5}$ in Padre de las Casas II station (Temuco). An estimate of the measurement uncertainty is shown in cyan solid lines. Green horizontal lines divide different pollution categories (Good: $<50\ \mu g/m^3$, Fair: 50–80 $\mu g/m^3$, Alert: 80–110 $\mu g/m^3$, Preemergency: 110–170 $\mu g/m^3$, and Emergency: $>170\ \mu g/m^3$).

3.2.4. Tracer to PM Factors

Using the calibration process described in section 2.4 and applying equation (1) with conversion factors outlined in sections 3.2.1, the final Tr_to_PM factors are obtained (Table 2). The range of Tr_to_PM values is wide (57–1273 $\mu g/m^3$/ppm) which can be due to multiple reasons. For instance, there are differences in the CO emission inventories assigned to each city (see section 2.3) and there could be biases in them (e.g., observed slopes in Figure 2 do not match the corresponding Tr_to_PM values). Other potential reasons for the wide range of Tr_to_PM values include differences in emissions sources between cities (e.g., cities dominated by wood-burning-stove emissions have larger PM to CO ratios, see section 2.4) and subgrid processes not resolved by the 2 km resolution (e.g., large spread of factors for stations in Temuco, probably due to the close proximity of Museo Ferroviario to the northern hills).

3.3. Predictability of Pollution Episodes

Figure 7 shows an example of 24 h running average $PM_{2.5}$ concentrations for the most polluted station in Temuco. The ensemble of forecasts generally follows the observed trend but with substantial variability among initialization times. The ensemble variability is generally larger than the measurement uncertainty and often results in different categories of pollution predicted for the same day. Figure 7 also shows that concentrations at the episode level ($>80\ \mu g/m^3$) can last from a few hours to multiple days. The model is able to represent the observed distribution of episode lengths (not shown). For instance, for Padre de las Casas II station there were 13 (13–17 in the forecast) episodes less than 24 h long, 5 (6–10 in the forecast) episodes within 1–2 days long, and the rest (6 observed, 6–8 modeled) 3–6 days long. The model tendency of slightly overestimating the amount of shorter episodes could be related to the overestimation of wind speeds, as pollution could get advected out of the basins faster in the model.

The skill of the system in predicting episodes and the associated false alarm rate is shown in Table 2. The model's ability to predict episodes depends on multiple factors. First, more accurate weather forecasts (including wind speeds and temperatures) should provide better representation of pollutant transport and

wood-burning emissions corrections and lead to better $PM_{2.5}$ predictions. As mentioned in section 3.1, wind speeds from the most current forecasts for stations from Curicó to the south tend to have better skill, which is also shown for many of these stations for episode prediction. For Santiago and Rancagua the trend shifts and better skill are shown by the day 2 wind speed forecast, which is also replicated for stations Pudahuel and Rancagua II for episode prediction skill. Although modeled wind speeds in Cerro Navia (Santiago) show the lowest correlation for day 3 forecasts, they also show the lowest bias (Table 1) as the forecasts initialized earlier tend to show lower wind speeds and thus reduce the high bias with respect to observations. Lower wind speeds produce higher chances of producing episodes (day 1 and day 3 initializations forecast episodes for Cerro Navia 26% and 35% of the time, respectively), and thus, higher likelihood of accurately predicting episodes as more episode days are forecasted, which also generates higher chances of false alarms. This explains why the day 3 initialization shows the best performance but the higher false alarm rate for Cerro Navia (Table 2). Skill in representing wind direction is also an important factor when reproducing episodes at Valdivia and Curicó, which are the stations with the lowest overall episode predictive skill (Table 2), and are also the ones with problems resolving wind direction (see section 3.1). Finally, the more common episodes are for a given station (i.e., larger number of days with observed episodes, Table 2) the higher the chances for being correct when predicting them as the chances of guessing correctly (i.e., episode prediction skill) increase from a probabilistic point of view. For instance, stations Pudahuel and Cerro Navia in Santiago and Las Encinas and Padre de las Casas II in Temuco are less than 2 km away from each other (Figure 1) which results in them being in neighboring grid cells, so model predictions should be similar. However, episodes are observed more often at Cerro Navia and Padre de las Casas II stations; thus, the model shows better prediction skill in these sites (Table 2).

As episodes can often last longer than 24 h, persistence (i.e., forecast the pollution category that was observed on the previous day to occur the next day) might be a good forecasting strategy; thus, we compare it to the forecasting system skill (Table 2). The overall prediction skill and false alarm rate from all three initializations generally do better than persistence (in 9 out of 12 stations shown in Table 2), and when the overall scores do worse, the first day forecast does similar or better than persistence.

3.4. Episode Dynamics and Evolution

An advantage of using a model that predicts three-dimensional concentrations is that it can be used to help understand how episodes develop. We selected episodes occurring in Temuco and Osorno as two cities representing locations of moderate to complex topography. One episode for each city was chosen that represents typical patterns of stagnation before sundown and that was well captured by the forecasts.

3.4.1. Temuco

Observed 24 h average $PM_{2.5}$ concentrations are accurately predicted by the model for the episode on 3–5 August that affected Temuco, as for these days the spread of the forecasts is narrow and the observations are contained within the spread (Figure 7). This pollution episode is not one of the most extreme ones in the study period but it is still severe as it reached preemergency levels at Padre de las Casas II station and Alert levels at the other two stations in Temuco for two consecutive days. Also, this episode is a good example of the abrupt transition from fairly clean to severely polluted conditions. Hourly concentrations generally peak at nighttime, and Figures 8a–8c show the evolution of the first night of the episode. Temuco is located in an area of complex topography in the Cautín river basin in between two hills, the Ñielol to the north, and the Conunhueno to the south. As shown in Figure 8a, low concentrations are found in the afternoon due to deeper boundary layer mixing and persistent westerly flow across the domain. When the boundary layer collapsed on the evening of 3 August the winds suddenly became stagnant in the valleys which, together with the evening emissions associated with wood-burning stove use and traffic, increased evening $PM_{2.5}$ from ~5 $\mu g/m^3$ to over 500 $\mu g/m^3$ in a period of 5 h, which are accurately predicted by the model (Figure 8b). After reaching peak concentrations, winds from the northeast slowly flushed pollutants out of the basin throughout the rest of the night (Figure 8c). Other episodes in the period of study reached similar peak concentrations but they remained high for longer periods due to a slower flushing of the basin, thus increasing the 24 h average and reaching Emergency pollution levels (mean 24 h $PM_{2.5}$ above 170 $\mu g/m^3$).

As shown in Figure 7, the model simulation of the 4–5 August episodes shows some variability with respect to initialization (~50 $\mu g/m^3$ spread). Similar analysis as those shown in Figures 8a–8c was performed for the 1 August initialization (not shown) presenting similar patterns of pollution accumulation but slightly stronger

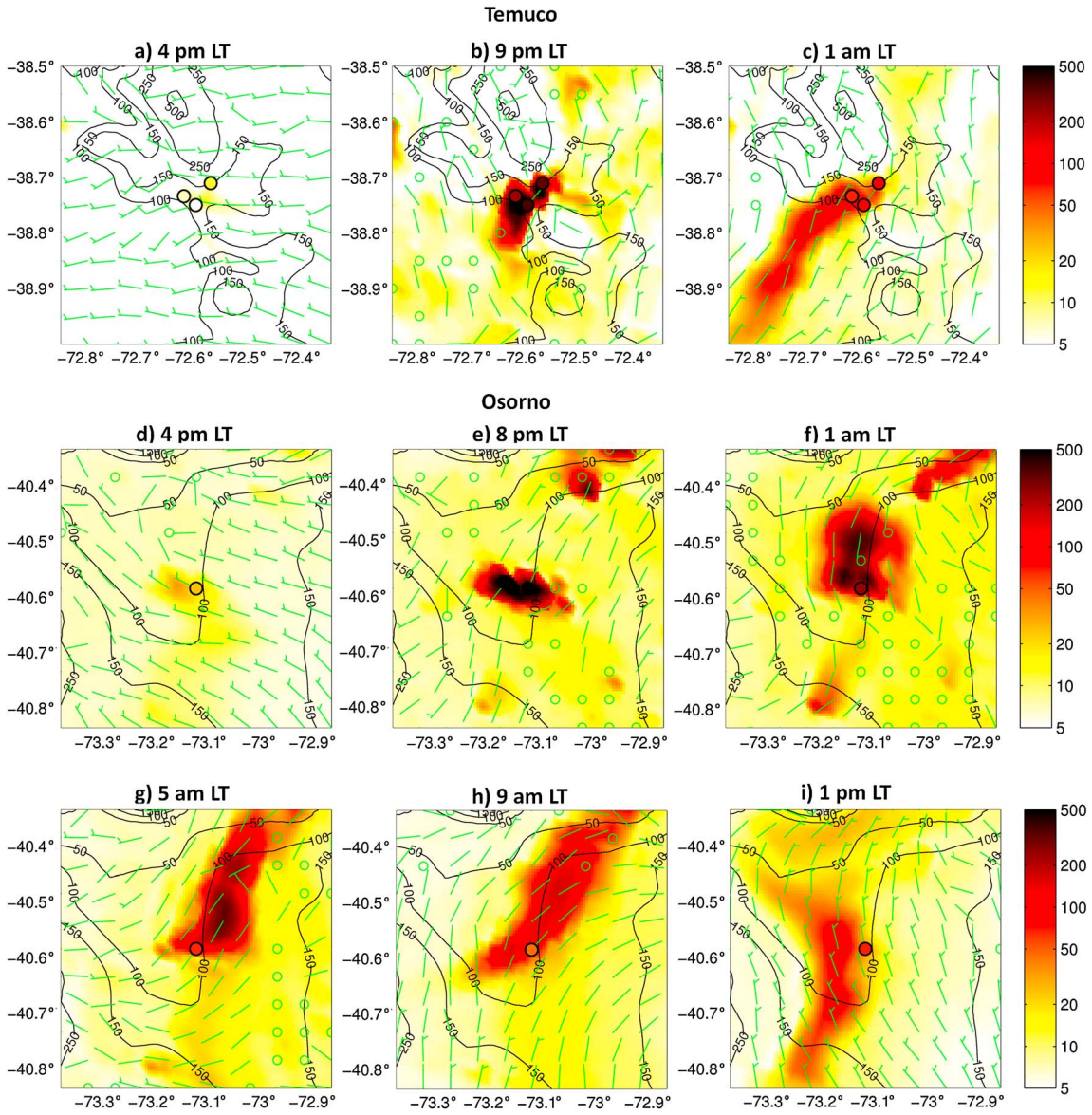

Figure 8. Hourly $PM_{2.5}$ ($\mu g/m^3$) forecasted maps showing the evolution of two pollution episodes. (a–c) An episode on 3–4 August 2014 in Temuco as forecasted by the 2 August initialization. (d–i) An episode on 7–8 August 2014 in Osorno as forecasted by the 5 August initialization. Observed values are shown in black circles and color coded in the same scale of the maps. Topography is shown as black contours (m) and wind barbs are shown in green. $PM_{2.5}$ was obtained using equation (1) with Tr_to_PM values from Padre de las Casas II and Osorno stations (Table 2).

regional winds, which resulted in model concentrations reaching Alert pollution levels instead of preemergency levels as shown by the 2 August initialization. The spread between initializations can be much larger for other episodes (Figure 7), sometimes resulting in one initialization predicting no episode, while another one Emergency for the same day (e.g., 7 and 13 July).

3.4.2. Osorno

On 8 August, the Osorno station reached 24 h average values of ~200 $\mu g/m^3$ surpassing the Emergency level threshold. The model was able to capture the evolution of this pollution episode very accurately (Figures 8d–8i). Although Osorno is not located in a region of such complex terrain as Temuco, it lays in a basin which, given conditions of poor vertical mixing, can trap pollution and produce episodes. Some accumulation of pollutants was found on the night of 6 August; thus, at 4 P.M. on 7 August background pollution levels were higher than usual (Figure 8d versus Figure 8a). Similarly as in Temuco, severe stagnation occurred producing a rapid increase in pollution levels reaching hourly concentrations of 800 $\mu g/m^3$ by 10 P.M. (Figure 8e).

Stagnation continued throughout the night and the plume meandered in the basin from Osorno toward the northeast without being able to exit the basin (Figures 8f and 8g). This plume together with fresh morning emissions also produced large pollution levels during the day (Figures 8h and 8i). As shown in Figures 8d–8i, smaller neighboring towns without air quality monitors also were predicted to have large levels of pollution and be affected by the Osorno plume because they are located in the same basin. Thus, implementing monitors in towns within the same basin of highly polluted cities should be considered in the future.

4. Conclusions

Severe air pollution episodes regularly affect several cities in central and southern Chile during the winter season. However, the skill of models in simulating meteorological and air quality measurements has not been evaluated for most of these cities and forecasting tools are limited. Here we develop and evaluate a near real time forecasting system able to predict the episodes 1 to 3 days in advance for nine cities.

The system is based on the WRF-Chem model configured at fine horizontal resolution (2 km) to resolve the complex topography, stability and stagnation conditions, and anthropogenic sources. Due to the high correlation between particles and CO, two inert tracers (traffic and wood burning) are used to represent primary pollutants from anthropogenic sources. $PM_{2.5}$ concentrations are then estimated by applying factors to these tracers which are calibrated with observations. The factors are based on physical phenomena including the relative contribution of wood-burning stove and traffic tracers to episodes by city, the increase in emission activity on weekend nights, and the temperature dependence of wood-burning-stove emissions. The calibration is performed by matching observed and modeled statistics on the total number of episodes predicted by station and by selecting the calibration with the best skill in predicting pollution events.

The calibration and testing are performed for the latest air pollution period available, April–August 2014. Temperature prediction skill tends to be similar from one initialization to another; thus, future implementations could assess modeling emissions using temperature predictions from previous days instead of applying correction factors. The skill of wind speed predictions can change depending on the time of initialization, with the recent ones having better skill from Curicó to the south and an opposite trend in Santiago and Rancagua, likely due to spin-up required to resolve the meteorology for the more complex topography in Central Chile. As $PM_{2.5}$ concentrations tend to be correlated to wind speeds, similar trends are found for episode predictions. Skill in predicting wind direction and precipitation also affects the ability to forecast episodes, which might be improved by testing finer grid cell resolution in future implementations. Episode forecasting skill is also modified by other factors such as how often episodes are observed and the number of episodes predicted by each initialization. While overall episode prediction skill varies from 53% to 72%, which is generally better than or similar to persistence, with the best initializations per station presenting skills of 61% to 76%.

An analysis of the $PM_{2.5}$ concentrations and wind maps at two cities and for two example events shows that the model can represent episode evolution. For these two events, the largest concentrations occur at nighttime after wind stagnation and rapid accumulation of freshly emitted pollution, which is generally how episodes unfold. This accumulation can often be significant over large regions, including towns with no monitors; thus, the model can be used to suggest locations for future stations. Plumes from the main cities lingering through the night can influence neighboring towns located in the same basin and can add to fresh emissions the next day, increasing pollution levels.

$PM_{2.5}$ 24 h running means can vary drastically between initialization start times as a result of differences in wind fields. Thus, an ensemble approach with multiple initializations per day (e.g., other than 00 UTC, using different global models, using ensemble members from a single global model) could be implemented in the future to assess uncertainties due to initialization and to provide a probabilistic forecast. For the current system, as all initializations show forecasting skill, they should all be used when deciding which episode category to forecast.

The system could be implemented in other parts of the world with complex topography, where strong correlations between particulate matter and anthropogenic tracers exist for severe pollution episodes [e.g., *Pataki et al*., 2005] and where observations and computational resources may be limited. Positive model performance for 2014 has led the Chilean Ministry of the Environment to implement the system for near real time use during 2015, making it the first deterministic and also the largest national air quality forecasting system in South America used for management of pollution episodes.

Acknowledgments
This research was supported in part through computational resources provided by the University of Iowa, Iowa City, Iowa, and with the aid of CONICYT/Fondap grants 15110017 and 15110009 (Center for Climate and Resilience Research), and NASA Grant #NNX11AI52G. The National Center for Atmospheric Research is supported by the National Science Foundation. Thanks to Marcelo Corral (Division of Air Quality and Climate Change of the Chilean Ministry of the Environment) and Louisa Emmons (NCAR) for their comments that helped improve the study. Contact P.E. Saide (saide@ucar.edu) for data and code requests.

References

Baklanov, A., et al. (2014), Online coupled regional meteorology chemistry models in Europe: Current status and prospects, *Atmos. Chem. Phys.*, *14*, 317–398.

Cassmassi, J. (1999), Improvement of the forecast of air quality and of the knowledge of the local meteorological conditions in the Metropolitan region Informe final, CONAMA RM.

CDT (2012), Propuestas de medidas para el uso eficiente de la leña en la Región Metropolitana de Santiago.

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system, Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*, 569–585.

CONAMA (2007), Update of the Atmospheric Emission Inventories for Santiago 2005 Final Technical Report by DICTUC, Jan. [Available at http://www.sinia.cl/1292/articles-49590_inf_final_inventario.pdf.]

Delgado, R., P. Hernandez, and M. Mena-Carrasco (2014), Operational implementation and performance evaluation of a PM10 and PM2.5 model for Santiago de Chile The World Weather Open Science Conference, Montreal, Canada.

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

Díaz-Robles, L., S. Cortés, A. Vergara-Fernández, and J. C. Ortega (2015), Short term health effects of particulate matter: A comparison between wood smoke and multi-source polluted urban areas in Chile, *Aerosol Air Qual. Res.*, *15*, 306–318.

DICTUC (2011), Guia Metodologica para elaboración de un AGIES para instrumentos de gestión de calidad de aire Chilean Ministry of the Environment. [Available at http://www.sinia.cl/1292/articles-54428_guia_metodologica.pdf.]

Dudhia, J. (1989), Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, *46*, 3077–3107.

Foley, K. M., et al. (2010), Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7, *Geosci. Model Dev.*, *3*, 205–226.

Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, *83*, 287–302.

Gallardo, L., G. Olivares, J. Langner, and B. Aarhus (2002), Coastal lows and sulfur air pollution in Central Chile, *Atmos. Environ.*, *36*, 3829–3841.

Garreaud, R., and J. Rutllant (2003), Coastal lows along the subtropical west coast of South America: Numerical simulation of a typical case, *Mon. Weather Rev.*, *131*, 891–908.

Garreaud, R., J. Rutllant, and H. Fuenzalida (2002), Coastal lows along the subtropical west coast of South America: Mean structure and evolution, *Mon. Weather Rev.*, *130*, 75–88.

Gobeli, D., H. Schloesser, and T. Pottberg (2008), Met one instruments BAM-1020 beta attenuation mass monitor US-EPA PM2. 5 federal equivalent method field test results The Air & Waste Management Association (A&WMA) Conference, Kansas City, Mo.

Gómez-Lobo, A., J. L. Lima, C. Hill, and M. Meneses (2006), Diagnóstico del mercado de la leña en Chile, Centro de Microdatos, Departamento de Economía, Universidad de Chile.

Grell, G. A., and S. R. Freitas (2013), A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, *Atmos. Chem. Phys. Discuss.*, *13*, 23,845–23,893.

Grell, G., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder (2005), Fully coupled "online" chemistry within the WRF model, *Atmos. Environ.*, *39*, 6957–6975.

Henríquez, A., A. Osses, L. Gallardo, and M. Díaz Resquin (2015), Analysis and optimal design of air quality monitoring networks using a variational approach, *Tellus B*, *67*.

Hong, S.-Y., J. Dudhia, and S.-H. Chen (2004), A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, *Mon. Weather Rev.*, *132*, 103–120.

Jiménez, P. A., and J. Dudhia (2011), Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model, *J. Appl. Meteorol. Climatol.*, *51*, 300–316.

Jorquera, H., and F. Barraza (2012), Source apportionment of ambient PM2.5 in Santiago, Chile: 1999 and 2004 results, *Sci. Total Environ.*, *435–436*, 418–429.

Jorquera, H., and J. Castro (2010), Analysis of urban pollution episodes by inverse modeling, *Atmos. Environ.*, *44*, 42–54.

Kang, D., R. Mathur, K. Schere, S. Yu, and B. Eder (2007), New categorical metrics for air quality model evaluation, *J. Appl. Meteorol. Climatol.*, *46*, 549–555.

Lee, S. H., S. W. Kim, W. M. Angevine, L. Bianco, S. A. McKeen, C. J. Senff, M. Trainer, S. C. Tucker, and R. J. Zamora (2011), Evaluation of urban surface parameterizations in the WRF model using measurements during the Texas Air Quality Study 2006 field campaign, *Atmos. Chem. Phys.*, *11*, 2127–2143.

Longo, K. M., et al. (2013), The Chemistry CATT-BRAMS model (CCATT-BRAMS 4.5): A regional atmospheric model system for integrated air quality and weather forecasting and research, *Geosci. Model Dev.*, *6*, 1389–1405.

Mena-Carrasco, M., E. Oliva, P. Saide, S. N. Spak, C. de la Maza, M. Osses, S. Tolvett, J. E. Campbell, T. E. Tsao, and L. T. Molina (2012), Estimating the health benefits from natural gas use in transport and heating in Santiago, Chile, *Sci. Total Environ.*, *429*, 257–265.

Mena-Carrasco, M., P. Saide, R. Delgado, P. Hernandez, S. Spak, L. Molina, G. Carmichael, and X. Jiang (2014), Regional climate feedbacks in Central Chile and their effect on air quality episodes and meteorology, *Urban Clim.*, *10*(Part 5), 771–781.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res.*, *102*, 16,663–16,682.

Moran, M. D., S. Ménard, D. Talbot, P. Huang, P. A. Makar, W. Gong, H. Landry, S. Gravel, S. Gong, and L. P. Crevier (2010), *Particulate-Matter Forecasting with GEM-MACH15, a New Canadian Air-Quality Forecast Model, Air Pollution Modelling and its Application XX*, pp. 289–292, Springer, Dordrecht.

Nakanishi, M., and H. Niino (2004), An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification, *Bound.-Layer Meteorol.*, *112*, 1–31.

Pataki, D. E., B. J. Tyler, R. E. Peterson, A. P. Nair, W. J. Steenburgh, and E. R. Pardyjak (2005), Can carbon dioxide be used as a tracer of urban atmospheric transport?, *J. Geophys. Res.*, *110*, D15102, doi:10.1029/2004JD005723.

Perez, P., and G. Salini (2008), PM2.5 forecasting in a large city: Comparison of three methods, *Atmos. Environ.*, *42*, 8219–8224.

Perez, P., and J. Reyes (2002), Prediction of maximum of 24-h average of PM10 concentrations 30 h in advance in Santiago, Chile, *Atmos. Environ.*, *36*, 4555–4561.

Rutllant, J., and R. Garreaud (1995), Meteorological air pollution potential for Santiago, Chile: Towards an objective episode forecasting, *Environ. Monit. Assess.*, *34*, 223–244.

Saide, P. E., G. R. Carmichael, S. N. Spak, L. Gallardo, A. E. Osses, M. A. Mena-Carrasco, and M. Pagowski (2011b), Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model, *Atmos. Environ.*, *45*, 2769–2780.

Saide, P. E., G. R. Carmichael, S. N. Spak, P. Minnis, and J. K. Ayers (2012a), Improving aerosol distributions below clouds by assimilating satellite-retrieved cloud droplet number, *Proc. Natl. Acad. Sci. U.S.A.*, *109*, 11,939–11,943.

Saide, P. E., et al. (2012b), Evaluating WRF-Chem aerosol indirect effects in Southeast Pacific marine stratocumulus during VOCALS-REx, *Atmos. Chem. Phys.*, *12*, 3045–3064.

Saide, P., R. Zah, M. Osses, and M. Ossés de Eicker (2009), Spatial disaggregation of traffic emission inventories in large cities using simplified top–down methods, *Atmos. Environ.*, *43*, 4914–4923.

Saide, P., M. Bocquet, A. Osses, and L. Gallardo (2011a), Constraining surface emissions of air pollutants using inverse modelling: Method intercomparison and a new two-step two-scale regularization approach, *Tellus B*, *63*, 360–370.

Sanhueza, P. A., M. A. Torreblanca, L. A. Diaz-Robles, L. N. Schiappacasse, M. P. Silva, and T. D. Astete (2009), Particulate air pollution and health effects for cardiovascular and respiratory causes in Temuco, Chile: A wood-smoke-polluted urban area, *J. Air Waste Manage.*, *59*, 1481–1488.

Sanhueza, P., C. Vargas, and P. Mellado (2006), Impact of air pollution by fine particulate matter (PM10) on daily mortality in Temuco, Chile, *Rev. Med. Chile*, *134*, 754–761.

Schmitz, R. (2005), Modelling of air pollution dispersion in Santiago de Chile, *Atmos. Environ.*, *39*, 2035–2047.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers (2008), A description of the advanced research WRF version 3, *NCAR Tech. Note NCAR/TN-475+ STR*.

Spak, S., M. Mena, and G. Carmichael (2010), Atmospheric transport of anthropogenic oxidized sulfur over the Southeast Pacific during VOCALS REx, *CLIVAR Exch.*, *15*, 20–21.

Troncoso, R., L. de Grange, and L. A. Cifuentes (2012), Effects of environmental alerts and pre-emergencies on pollutant concentrations in Santiago, Chile, *Atmos. Environ.*, *61*, 550–557.

Tuia, D., M. Ossés de Eicker, R. Zah, M. Osses, E. Zarate, and A. Clappier (2007), Evaluation of a simplified top-down model for the spatial assessment of hot traffic emissions in mid-sized cities, *Atmos. Environ.*, *41*, 3658–3671.

Twohy, C. H., et al. (2013), Impacts of aerosol particles on the microphysical and radiative properties of stratocumulus clouds over the southeast Pacific Ocean, *Atmos. Chem. Phys.*, *13*, 2541–2562.

USACH (2014), Actualización y sistematización del inventario de emisiones de contaminantes atmosféricos en la Región Metropolitana. [Available at http://www.sinia.cl/1292/articles-56914_Inf_Inventarios_FINAL.pdf.]

Valdés, A., A. Zanobetti, J. Halonen, L. Cifuentes, D. Morata, and J. Schwartz (2012), Elemental concentrations of ambient particles and cause specific mortality in Santiago, Chile: A time series study, *Environ. Health*, *11*, 1–8.

Villalobos, A. M., F. Barraza, H. Jorquera, and J. J. Schauer (2015), Chemical speciation and source apportionment of fine particulate matter in Santiago, Chile, 2013, *Sci. Total Environ.*, *512–513*, 133–142.

Yang, Q., et al. (2011), Assessing regional scale predictions of aerosols, marine stratocumulus, and their interactions during VOCALS-REx using WRF-Chem, *Atmos. Chem. Phys.*, *11*, 11,951–11,975.

Yu, M., G. R. Carmichael, T. Zhu, and Y. Cheng (2012), Sensitivity of predicted pollutant levels to urbanization in China, *Atmos. Environ.*, *60*, 544–554.

Zhang, Y. (2008), Online-coupled meteorology and chemistry models: History, current status, and outlook, *Atmos. Chem. Phys.*, *8*, 2895–2932.