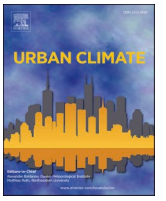
Urban Climate 38 (2021) 100906

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim



Forecasting $PM_{2.5}$ levels in Santiago de Chile using deep learning neural networks

Camilo Menares [a,b], Patricio Perez [c,*], Santiago Parraguez [a,d], Zoë L. Fleming [a,e]

[a] Center for Climate and Resilience Research (CR)^2, Santiago, Chile
[b] Departamento de Geofísica, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile
[c] Departamento de Física, Universidad de Santiago de Chile, Santiago, Chile
[d] Departamento de Ingeniería Mecánica, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile
[e] Envirohealth Dynamics lab, C+ Research Center in Technologies for Society, School of Engineering, Universidad Del Desarrollo, Santiago, Chile

ARTICLE INFO

Keywords:
Air quality forecasting
Meteorology forecast
Fine particulate matter
Deep neural networks
Machine learning
LSTM

ABSTRACT

Air pollution has been shown to have a direct effect on human health. In particular, $PM_{2.5}$ has been proven to be related to cardiovascular and respiratory problems. Therefore, it is important to have accurate models to predict high pollution events for this and other pollutants. We present different models that forecast $PM_{2.5}$ maximum concentrations using a Long Short-Term Memory (LSTM) based neural network and a Deep Feedforward Neural Network (DFFNN). Ten years of air pollution and meteorological measurements from the network of monitoring stations in the city of Santiago, Chile were used, focusing on the behaviour of three zones of the city. All missing values were rebuilt using a method based on discrete cosine transforms and photochemical predictors selected through unsupervised clustering. Deep learning techniques provide significant improvements compared to a traditional multi-layer neural networks, particularly the LSTM model configured with a 7-day memory window (synoptic scale of pollution patterns) can capture critical pollution events at sites with both primary and secondary air pollution problems. Furthermore, the LSTM model consistently outperform deterministic models currently used in Santiago, Chile.

1. Introduction

Santiago, the capital of Chile is known for its poor air quality, but over recent years a series of air quality plans have helped to reduce this pollution. Forecasting of pollution episodes is critical for informing the public of such episodes as well as activating pollution reduction efforts such as prohibiting the most polluting older vehicles and advising the public not to carry out outdoor exercise during these periods. Santiago de Chile (33° S, 70° W, 600 m a.s.l.) has an urban area of 640 km^2 and a population of 6 million and is situated in a 20 km wide valley between the Andes mountains and the coastal range (see Fig. 1). Due to an unfavourable topographic and meteorological conditions for pollutant dispersion and ventilation during the cold season (between April and August) (Toro et al., 2019), the Particulate Matter (PM) emissions from industry, transportation and heating do not disperse, increasing $PM_{2.5}$ concentrations.

An important fraction of the $PM_{2.5}$ (Particulate Matter smaller than 2.5 μm diameter) values recorded at the 10 monitoring stations

* Corresponding author.
E-mail address: patricio.perez@usach.cl (P. Perez).

https://doi.org/10.1016/j.uclim.2021.100906
Received 8 April 2021; Received in revised form 19 June 2021; Accepted 29 June 2021
Available online 5 July 2021

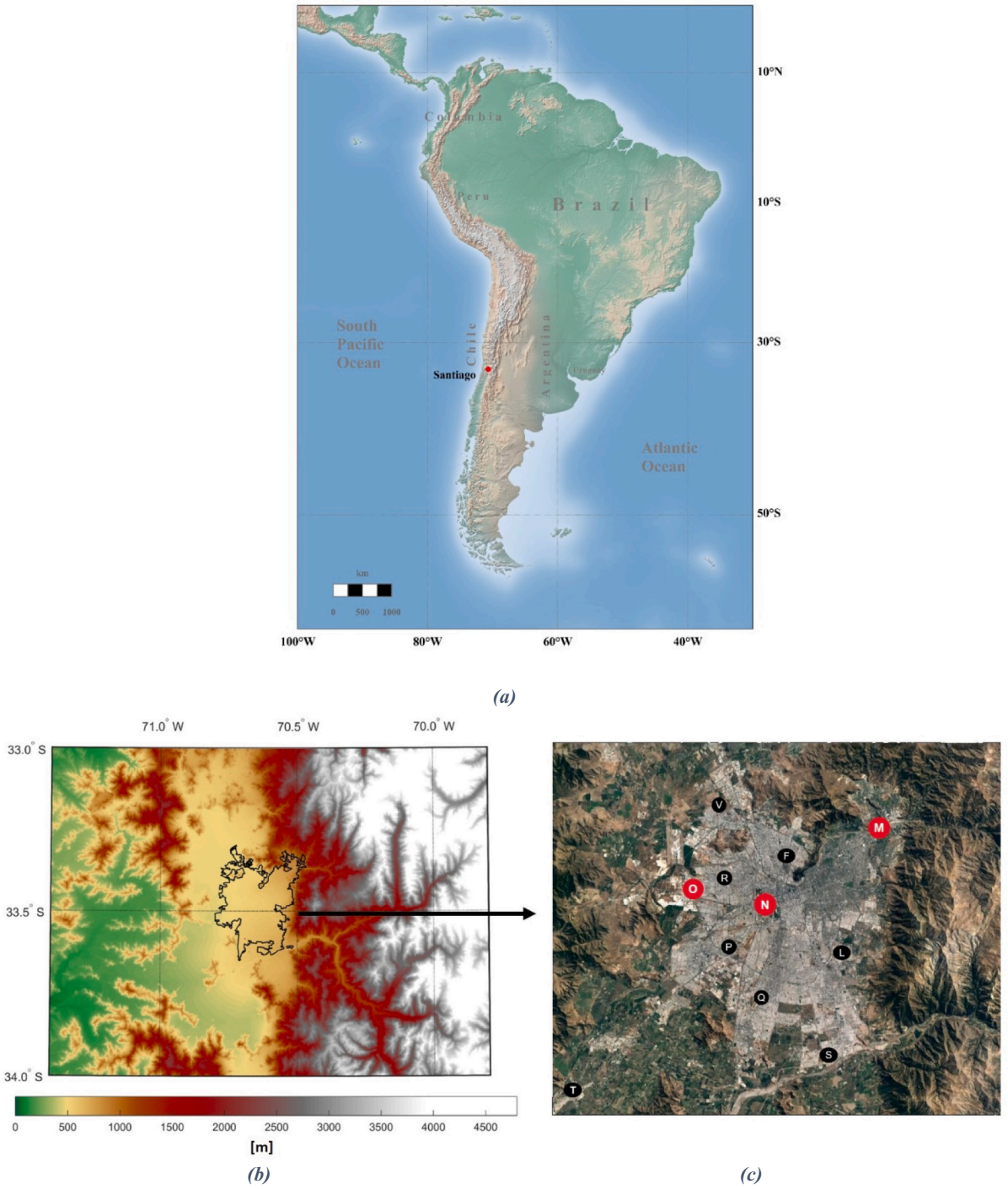


Fig. 1. a) Map of South Americas showing the location of Santiago, Chile. b) The city of Santiago with colors indicating the height above sea level. c) Circles indicate location of city's SINCA monitoring stations, circles in red indicate the stations with emphasis in this study. Stations P (many missing data) and T (far from urban area) are not considered in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is due to secondary particle formation from chemical reactions involving nitrogen and sulfur oxides. $PM_{2.5}$ has been identified as cause of cardiovascular and respiratory mortality throughout the world (Liu et al., 2019a; Hajizadeh et al., 2020). It has also been associated with an increase in the probability of developing Alzheimer's disease (Jung et al., 2015). Evidence exists of both short (day-to-day) and long term (years) exposure to gaseous and particulate air pollution (Kim et al., 2015).

Emission fields and atmospheric chemistry in the city show spatial variations across different zones (Usach, 2014; Langner et al., 2020; Menares et al., 2020; Gallardo et al., 2018). The western zone is dominated by primary emissions generated by industries and heating, while the eastern part acts as an urban receptor (Tagle et al., 2018), due to the predominant westerly winds, where secondary particle formation, such as ozone becomes relevant.

Air quality monitoring has taken place in Santiago in a standardized manner since 1997 when the first attainment plan was implemented (Toro et al., 2019) and currently the monitoring is under the responsibility of the Ministry of the Environment. The latest attainment plan for Santiago can be found at (http://airesantiago.gob.cl/nuevas-medidas/, http://airechile.mma.gob.cl/comunas/santiago).

The World Health Organization (WHO, 2021) "Guideline level" is recommended at 25 $\mu g/m^3$ (24-h mean) as well as an annual Guideline level of 10 $\mu g/m^3$ for $PM_{2.5}$. Chilean National $PM_{2.5}$ standards are not so stringent as the international guidelines and are monitored as 24 h moving averages and defined in terms of 4 ranges of increasing concentrations:

- Level A (good to moderate): maximum of 24-h average of $PM_{2.5}$ is less than 80 $\mu g/m^3$.
- Level B (unhealthy): maximum of 24 h average is between 80 $\mu g/m^3$ and 110 $\mu g/m^3$.
- Level C (very unhealthy): maximum of 24 h average is between 110 $\mu g/m^3$ and 170 $\mu g/m^3$.
- Level D (hazardous): 24 h average is greater than 170 $\mu g/m^3$.

Levels B, C and D are considered as "episodes" and are reported on the news and fairly well understood by the population of the city. In order to anticipate the occurrence of episodes it has become important to generate reports with a forecasting model. These forecasting models predict the city's air quality with an anticipation of one or more days, thus alerting the population of forthcoming episodes and allowing restriction measures to be put into place.

The wide array of air quality forecasting models used elsewhere to predict potentially dangerous levels of pollution can be classified into two main families: Deterministic and statistical models. Deterministic models provide real time simulations of the behaviour of atmospheric components and include equations for particle interactions and chemical reactions. Details about local topography and the sources of pollutant emissions are important to feed these models (Rouïl et al., 2009; Mc Keen et al., 2007). In general, these models require that the meteorology is provided as an input (Honoré et al., 2008) or is embedded into the algorithm, as in the WRF-Chem model (Grell et al., 2005). Statistical models are based on the assumption that past values of relevant meteorological variables and pollutant concentrations determine the actual concentration of a given pollutant such as $PM_{2.5}$, such that similar combinations of the identified predictor variables associated with future values of the pollutant are inputted. Linear regressions, artificial neural networks (Thomas and Jacko, 2007; Oprea et al., 2016) and fuzzy systems (Ausati and Amamollahi, 2016) are some of the statistical models used for $PM_{2.5}$ forecasting. It has been reported that for short term particulate matter forecasting, artificial neural networks models have similar or even better accuracy than deterministic models (Kukkonen et al., 2003: Fernando et al., 2012, Samal et al., 2021, Chen and Li, 2021).

By the late 1990s, a multiple regression method was used to forecast the next day's probability of occurrence of heavy pollution episodes in Santiago. At present the official model used by the ministry of the environment uses a numerical model of CO, in combination with empirical relationships between CO and PM to forecast air quality for the next 72 h in Santiago as well as for many other cities in central and southern Chile (Saide et al., 2016).

In recent years, in addition to deterministic models (Saide et al., 2016), statistical models (Perez and Salini, 2008; Moisan et al., 2018) have been used for $PM_{2.5}$ forecasting in Santiago, oriented towards anticipating episodes. These studies show that WRF-Chem, linear models and feed forward artificial neural networks (FFNN) have the ability to forecast episodes with a comparable accuracy of the order of 60%. This accuracy is significant to provide a protection to the population when episodes of high pollution are foreseen (Borrego et al., 2008).

The computer resources necessary for the implementation of the deterministic models are in general substantially less than those needed for statistical models. Also, deterministic models require precise emission inventories, which are not always available (Zhou et al., 2017). Deep neural network models are convenient tools for air quality forecasting when several years of data, many input variables and several monitoring stations are available. A Deep Neural Network (DNN) is defined as a Feed Forward Multilayer Net (FFNN) that has from three to several tens of hidden layers. The simplest version of DNN is just an extension of the FFNN with a convenient adaptation of the learning rule and is known as DFFNN. Other approaches are based on the spatial relationship between data (Convolutional Neural Networks, CNN) or on the recurrence of each input (Recurrent Neural Networks, RNN). In particular, Long Short-Term Memory cells, a derivation of RNNs, have been shown to be more accurate than a DFFNN for forecasting $PM_{2.5}$ at Tehran, where 4 years of data from 9 stations were used in order to train and test the models (Kaimian et al., 2019). Mao and Lee (2019) used data from 77 stations in Taiwan to forecast Ozone, SO_2 and $PM_{2.5}$ concentrations with reasonable accuracy and extended these results with TS-LSTME models in Beijing, Tianjin, and Hebei (Mao et al., 2021). Li et al. (2020) have shown that a combination of deep algorithms is the best option for $PM_{2.5}$ forecasting in Taiyuan city, China.

In this paper we present two approaches to forecast $PM_{2.5}$ over Santiago. First a DFFNN model and second an LSTM network, thus enabling a comparison of the results of dense networks with a recurrent network, which has the capacity to retain long-term information in its memory. With both models we show that, based on an expanded database, we can significantly improve the accuracy of the episode forecast. In particular, it is shown that the LSTM model is more suitable in this case, thanks to its ability to retain information on past events.

Table 1
Species and measurement techniques used in the Santiago monitoring network. The monitoring stations where the species are measured are also indicated (for the location, see Fig. 1). This information can be found at http://sinca.mma.gob.cl/.

Species	Measurement technique	Measurement location
Ozone (O_3)	Ultraviolet photometry (THERMO 49i)	O, R, N, F, Q, M, L, S
Nitrogen monoxide (NO)	Chemiluminescence gaseous phase - THERMO 42i	O, R, N, F, Q, M, L, S
Nitrogen dioxide (NO_2)	Chemiluminescence gaseous phase - THERMO 42i	O, R, N, F, Q, M, L, S
Carbon monoxide (CO)	Photometry of gas correlation filter -THERMO 48i	O, R, N, F, Q, M, L, S
Sulfur dioxide (SO_2)	Pulsing fluorescence THERMO 43i	O, R, N, F, Q, M, L, S
Particulate Matter (<10 μm) PM_{10}	Tapered Element Oscillating Microbalance (TEOM)- THERMO 1400AB	V, O, R, N, F, Q, M, L, S
Particulate Matter (<2.5 μm) $PM_{2.5}$	Beta radiation attenuation - Met One Bam1020	V, O, R, N, F, Q, M, L, S
Meteorological variable	Measurement technique	Measurement location
Temperature	Sensor MET ONE HMP35A	O, R, N, F, Q, M, L, S
Relative humidity	Sensor MET ONE HMP35A	O, R, N, F, Q, M, L, S
Wind speed	Sensor MET ONE 020C	O, R, N, F, Q, M, L, S

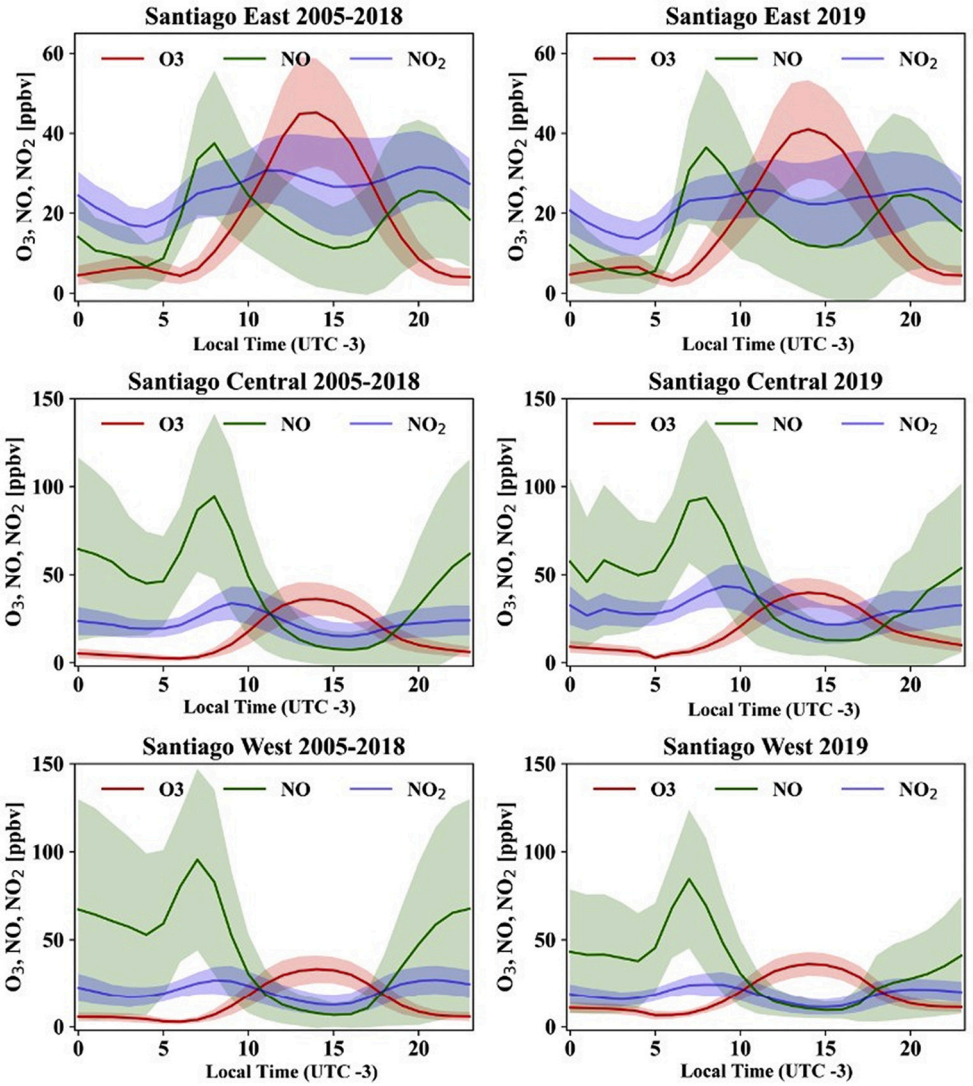


Fig. 2. Daily hourly average O_3, NO and NO_2 levels for stations M, N and O. On the left, years 2005–2018. On the right, year 2019. Shading is an indication of the standard deviation around mean values.

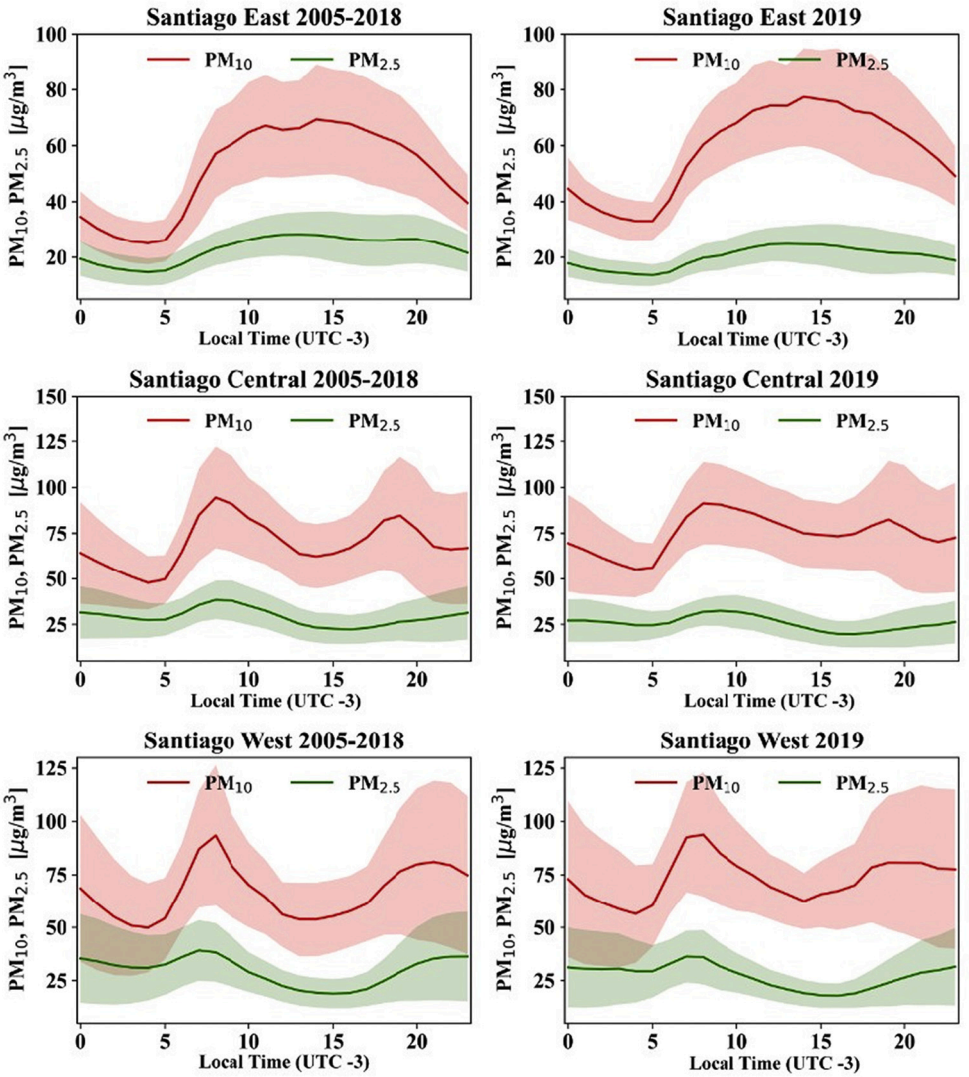


Fig. 3. Daily hourly average $PM_{2.5}$ and PM_{10} for stations M, N, O. On the left, years 2005–2018. On the right, year 2019. Shading is an indication of the standard deviation around mean values.

2. Data and methodology

2.1. Air quality and meteorology monitoring station data

Hourly averaged concentrations of air pollutants and meteorological information are collected at 9 monitoring stations distributed across the city of Santiago (Fig. 1). These stations are supervised by the Ministry of the Environment, which generates periodic quality controls on the data. Meteorological data includes Temperature, Humidity and Wind speed. Average temperature in Santiago are 20 °C in summer (DJF) and 9 °C in winter (JJA) and average annual precipitation is 300 mm, concentrated mostly in the cold season. Table 1 shows the pollutants and meteorological variables of interest measured in each station, including references to the instrumental method used.

This study used data between 2005 and 2019, with 2005–2018 used for training (adjusting the parameters of our models), and 2019 for the test (to check the performance of our models, see Section 2.3). For this reason, Fig. 2 displays the daily cycle of 3 pollutants for three representative stations from Eastern (station M), Central (station N) and Western (station O) Santiago, with specific and recurrent pollutant patterns. We also observe in Fig. 2 that Station East displays urban receptor conditions, where the production of ozone is enhanced, and secondary particle creation is relevant.

Average diurnal cycles for particulate matter $PM_{2.5}$ and PM_{10} are shown in Fig. 3. Central and West stations show similar behaviour, except that Santiago West has more frequent high concentration episodes. According to the behaviour of aerosols and the oxidative pollutants O_3 and NO_2 in Santiago Central it could be described as an urban transition zone.

2.2. Data interpolation to deal with missing data

It is normal for meteorological and pollution data measurements to have missing values so for this reason, it is relevant to have a method that deals with these gaps in the dataset. For cases where the amount of missing data does not allow a reliable daily reconstruction of the information, we use an efficient data filling method based on discrete cosine transforms (Garcia, 2010). This method is a penalized least square regression based on the discrete cosine transform which describes the data in terms of the sum of cosine functions oscillating at different frequencies. This method has been used successfully for geophysical records, as shown by. This application is replicated by us to fill the gaps in the collected data.

The cosine transform (DCT) and its inverse (IDCT) are then used to iteratively find the $\widehat{X}$ of the form

$$\widehat{X} = IDCT\left(\Gamma \circ DCT\left(W \circ \left(X - \widehat{X}\right) + \widehat{X}\right)\right) \tag{1}$$

whit W the weights assigned to each value. Finally, Γ is a filtering tensor defined by

$$\Gamma_i = \left(1 + s\left(2 - cos\frac{(i-1)\,\pi}{n}\right)2\right)^{-1} \tag{2}$$

where i represents the ith element, s is the smoothing factor, which must be a positive scalar and n the amount of data used to predict.

2.3. Deep learning models

Given the positive results of particulate matter forecasting models using traditional multilayer neural networks developed for Santiago (Perez and Reyes, 2006; Perez and Salini, 2008, Perez and Gramsch, 2016, Perez and Menares, 2018), we are looking for an improvement in accuracy by introducing deep learning algorithms.

The variable we are aiming to forecast is the maximum of the 24 h moving average of $PM_{2.5}$ for the following day. We focus on two deep learning models, a Long- Short Term Memory model (LSTM) and a deep feed forward neural network (DFFNN).

2.3.1. Model architecture

Our algorithms LSTM and DFFNN are based on the backpropagation algorithm. This is implemented by means of the initiation of a forward propagation algorithm that computes the loss function L(θ) (where θ contains the weights ω and bias b of the Goodfellow model (Goodfellow et al., 2016)). The gradients g_t are obtained by using the Zinkevich version of backpropagation (Zinkevich, 2003) according to which:

$$g_t = \nabla_w L(Da_i, Dp_a) \tag{2}$$

where Da_i is the actual values of $PM_{2.5}$ and Dp_i is the output of the model.

Both networks are optimized by using the Adam algorithm (Kingma and Ba, 2015) where the weights are obtained by calculating:

$$\omega_t = \omega_{t-1} - \alpha\frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \in} \tag{3}$$

The first and second moments, $\widehat{m_t}$ and $\widehat{v_t}$ are estimated through:

$$\widehat{m_t} = \frac{\beta_1 \bullet m_{t-1} + (1-\beta_1)g_t}{1-\beta_1^t} \tag{4}$$

$$\widehat{v_t} = \frac{\beta_2 \bullet m_{t-1} + (1-\beta_2)g_t^2}{1-\beta_2^t} \tag{5}$$

Values used for β_1 and β_2 are 0.9 and 0.999 respectively, $\in$ is a stabilization parameter used as 1e-6 following (Kingma and Ba, 2015).

LSTM and DFFNN algorithms are adjusted by minimizing the mean square absolute percent error as a loss function:

$$Loss = \frac{1}{n}\sum_{i=1}^{n}\left(Y_{fi} - Y_{oi}\right)2 \tag{6}$$

Where Y_f is the forecasted value and Y_o is the observed value. This algorithm is based on a stochastic gradient descent which considers the adaptive estimation of first and second order moments and is more economic in computational time when compared with other training methods. In order to compare the results of the deep learning methods with traditional methods we implement a feed forward multilayer network (FFNN). The output layer in all models has one neuron and the activation function in this layer is linear. For other layers, depending on the model, the activation functions used are Sigmoid, ReLU (Eckle and Schmidt-Hieber, 2019) and Softplus (Dugas et al., 2000):

$$\text{Sigmoid}\ (x) = 1/(1+e^{-x}) \tag{7}$$

Table 2
Configuration of the neuronal models used. In the second column the number of hidden Layers are specified, as well as the activation function for each layer and the number of Neurons used in each of these.

Neural Network	Layers	Activation	Neurons	Batch
DFFNN	4	Relu-Softplus-Linear	45–25-01	20
LSTM	3	Relu-Relu-Sigmoid	64–64-01	34
MLP	3	Sigmoid-Linear	22–6-1	6

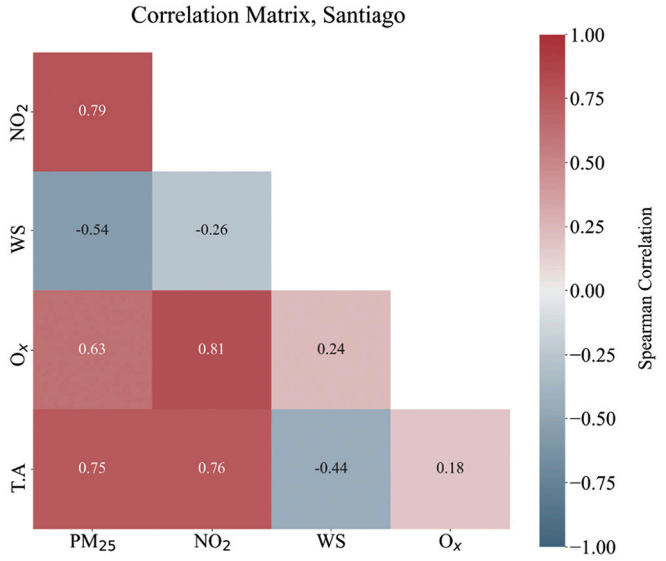


Fig. 4. Correlation matrix for $PM_{2.5}$, NO_2, O_x, ($O_3 + NO_2$), T.A (Thermal Amplitude) and WS (Wind speed) in day whit $PM_{2.5}$ above 80 μg/m^3. The spearman correlation is represented as a colour scale, with red indicating $R = 1$ and blue indicating R = -1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\text{ReLU}\ (x) = max\ (0, x) \tag{8}$$

$$\text{Softplus}\ (x) = ln\ (1 + e^x) \tag{9}$$

In Table 2 the architecture, activation functions and training details of the nets used in this study are shown. Additionally, our LSTM algorithm is configured with a 7-day memory window. This value of 7-days is selected because it is within the time of the synoptic factors of contamination (Liu et al., 2019b; Rutllant and Garreaud, 1995).

2.3.2. *Choice of $PM_{2.5}$ predictor*

After exploration (Fig. 4 and previous models) we found that the most convenient input variables for a given station are:

1) 24 h average of $PM_{2.5}$ concentrations
2) 24 h average of PM_{10} concentrations
3) 24 h average of $PM_{2.5}$ concentration at closest station
4) 24 h average of NO_2 concentrations
5) 24 h average of wind speed
6) Thermal amplitude in last 24 h
7) 24 h average of O_x concentration
8) maximum of O_3 hourly concentration in last 24 h
9) 24 h average of NO/NO_2 ratio
10) 24 h average of CO concentrations

All 24 h averages are calculated between 20:00 of the previous day and 19:00 of the present day. These times were chosen because when a $PM_{2.5}$ forecasting model is in operation, a report to the population is emitted at 20:30 of the present day in order to inform them about expected conditions for the following day. These input variables are the best predictor variables selected from a larger pool used in particulate matter concentration forecasting models in Chilean cities in the past (Perez, 2012; Perez and Menares, 2018; Perez et al., 2020). These variables have been selected with the following criteria:

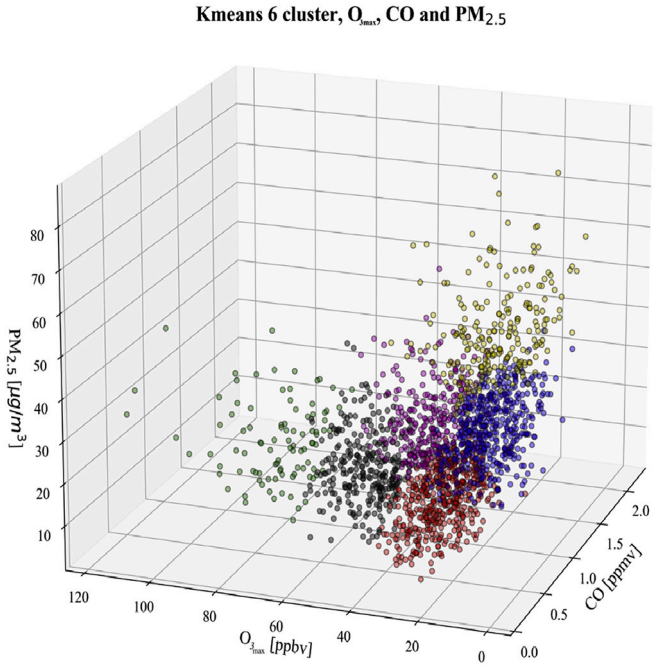


Fig. 5. K-means with K = 6 based on hourly values of $PM_{2.5}$, O_3 and CO in station M (East side).

Initial and boundary conditions: The first two variables take into account the trend in the air pollutant concentrations. The third variable takes the nearest neighboring station with $PM_{2.5}$ values and is used to identify the possible transport of aerosols.

High correlation: Fig. 4 shows the Spearman correlation map (chosen because it better represents non-normal variables and outlier data (de Winter et al., 2016)) for station O west of Santiago in days whit concentration of $PM_{2.5} > 80$ μg/m^3. Wind speed is an indicator of ventilation and shows a significant anticorrelation with $PM_{2.5}$ concentrations. Large thermal amplitudes in winter are an indicator of irradiative inversions which trap pollutants near the surface (positive correlation in Fig. 5). We also observe strongly positive correlation between $PM_{2.5}$ and indicators of secondary particle formation such as O_x and NO_2 (Menares et al., 2020). For this reason, the variables of wind speed, thermal amplitude, O_x and NO_2 are selected.

Pollution patterns: Fig. 5 shows a 3-D K-means (K = 6) clustering (Nainggolan et al., 2019) for hourly values of CO, O_{3max} and $PM_{2.5}$ for station O in the west of the town. The idea of this clustering is to identify atmospheric configurations relevant for pollution conditions. Affinity is obtained by calculating Euclidean distance with normalized variables. The six clusters have been chosen using the Nainggolan et al., 2019 approach, where the Sum Squared Error (SSE) of the values generated by distance between the data and the cluster center is calculated as indicator of optimization of clusters. We find that the SSE value is minimized with six clustering. Six clusters are generated which can be distinguished by colors. The green cluster is of particular interest because it is formed with high values of $PM_{2.5}$ and O_{3max} and low values of CO, which may represent the condition of high photochemical activity in the city. The yellow cluster by contrast is obtained with low values of O_{3max} and high values of $PM_{2.5}$ and CO, which identify situations of primary emissions from fossil fuels. The red and blue clusters may be identified with situations of low pollution. For this reason, the variables O_{3max} and CO are chosen. On the other hand, the NO/NO_2 ratio was included as an input variable so as to quantify primary and secondary emissions from transportation (Mavroidis and Chaloulakou, 2011).

2.4. Statistical evaluation of models

In order to evaluate the quality of the results of the different forecasting models we calculate the following statistical parameters: Correlation (R), Normalized Percent Error (NPE), Fractional Bias (FB) and Normalized standard deviation (NSD), which are defined by the expressions:

$$R = \frac{\langle (y_f - \langle y_f \rangle)(y_a - \langle y_a \rangle) \rangle}{\sqrt{\langle y_f - \langle y_f \rangle \rangle \langle y_a - \langle y_a \rangle \rangle}} \tag{9}$$

$$NPE = \frac{\langle |y_f - y_a| \rangle}{\langle y_a \rangle} \tag{10}$$

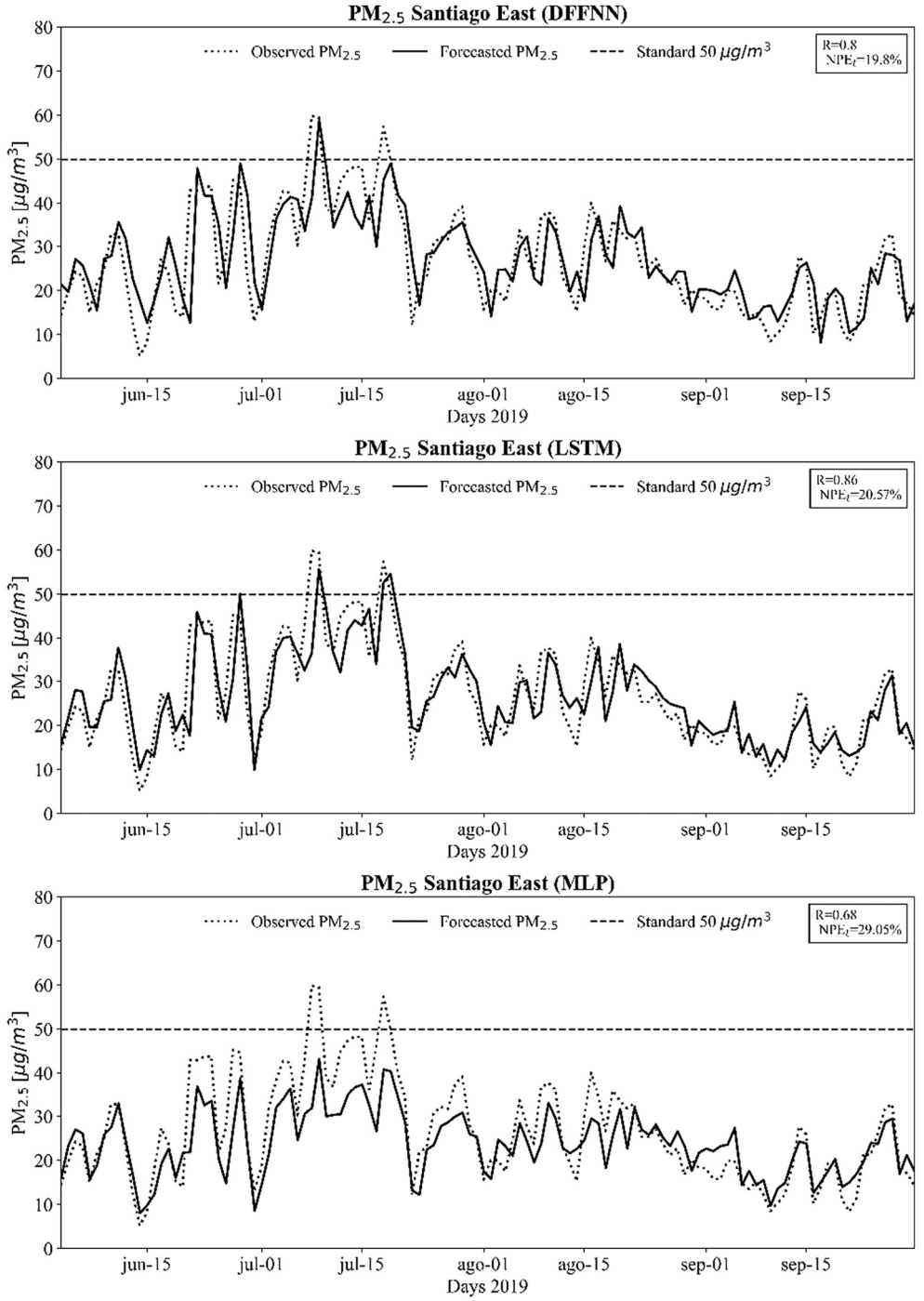


Fig. 6. Comparison between forecasted and observed values in station M (East) for three models: DFFNN, LSTM and MLP. Horizontal dotted line is the standard for 24 h average $PM_{2.5}$.

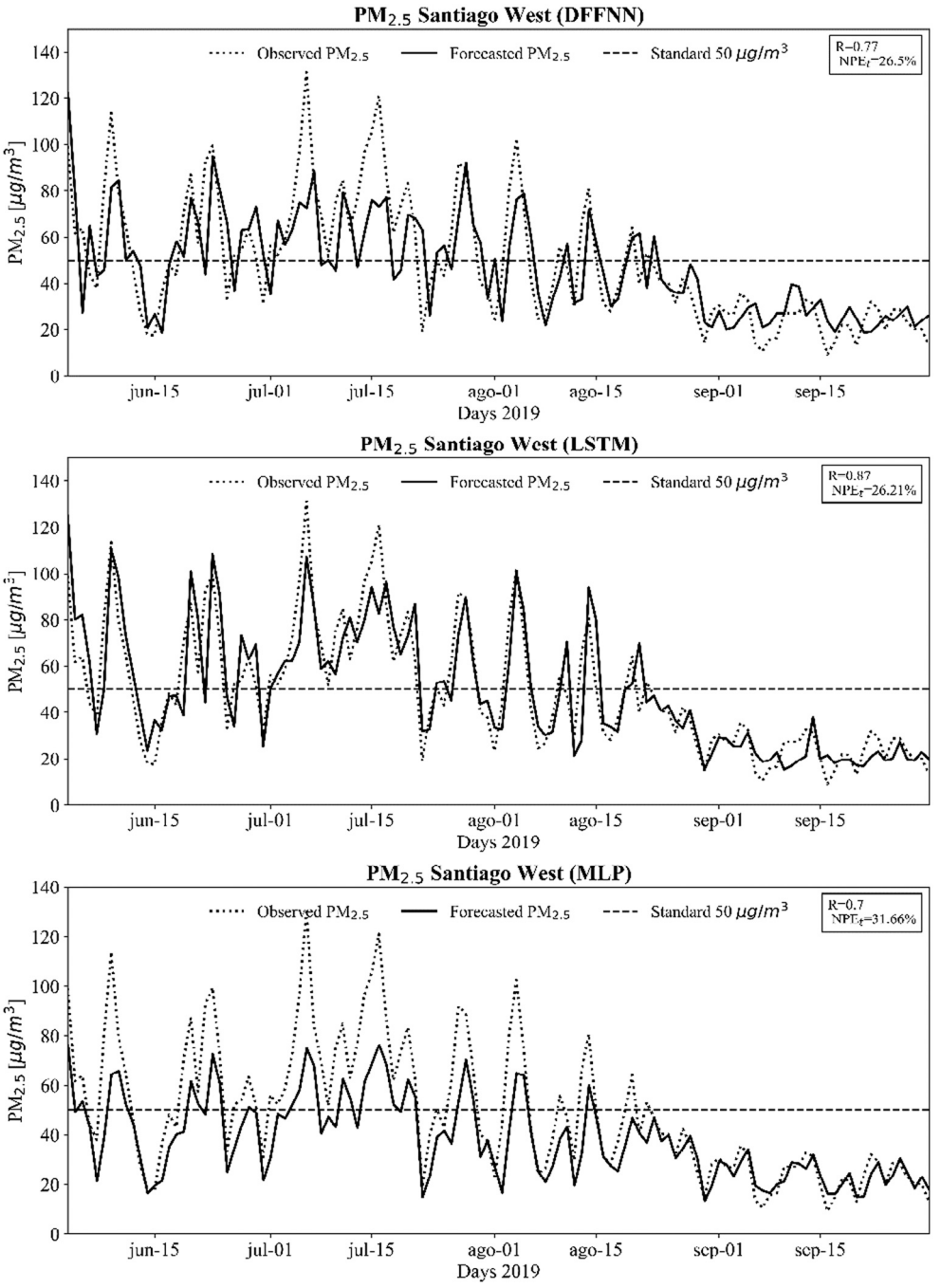


Fig. 7. Comparison between forecasted and observed values in station O (west side) for three models: DFFNN, LSTM and MLP. Horizontal dotted line is the standard for 24 h average $PM_{2.5}$.

$$FB = 2 \bullet \frac{\langle y_a \rangle - \langle y_f \rangle}{\langle y_f \rangle + \langle y_a \rangle} \tag{11}$$

$$NSD = \frac{\sigma_f}{\sigma_a} \tag{12}$$

Where triangular brackets indicate average over test set, yf are the forecasted 24 h average values and ya are the actual values. The reason to consider NPE is because this quantity avoids amplification of errors in the case of low concentrations. Bias measures systematic errors, where a positive value indicates a tendency of the model to under-forecast and a negative value a tendency to over-

Table 3
Presents the relevant statistical indices measuring performance of the three models for the case of station M (east side), station N (center) and station O (west side).

Station M				
MODEL	NPE	R	FB	NSD
DFFNN	20%	0.80	0.031	0.82
LSTM	21%	0.86	0.048	0.78
MLP	29%	0.68	0.060	0.68
Station N				
MODEL	NPE	R	FB	NSD
DFFNN	26%	0.78	0.19	0.82
LSTM	24%	0.87	0.13	0.84
MLP	38%	0.68	0.42	0.69
Station O				
MODEL	NPE	R	FB	NSD
DFFNN	27%	0.77	0.70	0.80
LSTM	26%	0.88	0.45	0.89
MLP	32%	0.70	0.80	0.66

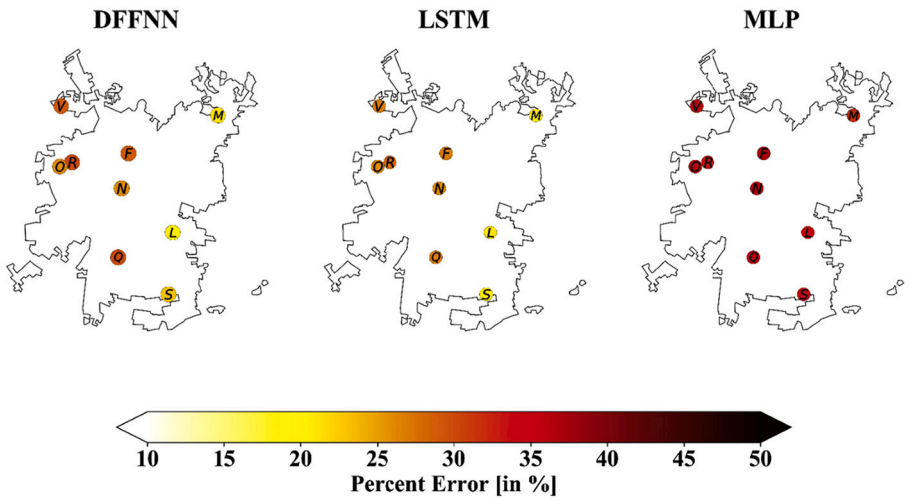


Fig. 8. Normalized percent errors for all stations in Santiago using the three models.

Table 4
DFFNN confusion matrix for station O in 2019 where A, B, C and D represent range of concentration of Level A (good to moderate), Level B (unhealthy), Level C (very unhealthy) and Level D (hazardous) respectively.

	A	B	C	D	%
A	55	10	2	0	82%
B	13	21	6	1	51%
C	0	1	8	1	80%
D	0	0	1	1	50%
%	81%	66%	47%	33%	71%

forecast. NSD gives a measure of the dispersion of the model results as compared with the observed values.

3. Results and discussion

After training the different models using 2005–2018 data, we performed a test with 2019 winter data (from June 1st to September 30) where the target value is the maximum of the 24 h average of $PM_{2.5}$ for the following day. In Fig. 6 we display the comparison between forecasted and observed values for a station located in Eastern Santiago (M station).

We can verify that LSTM and DFFNN have a similar performance in Normalized Percent Error (NPE $\approx$ 20%) but LSTM has the best performance in terms of Pearson Correlation ($R = 0.86$). It is evident from the graphs that LSTM is more accurate than the other models for the forecasting of the highest concentrations that occur during the first 20 days of the month of July. In Fig. 7 we show the results

Table 5
LSTM confusion matrix for station O in 2019 where A, B, C and D represent range of concentration of Level A (good to moderate), Level B (unhealthy), Level C (very unhealthy) and Level D (hazardous).

	A	B	C	D	%
A	62	3	1	0	94%
B	6	20	5	0	65%
C	0	8	10	1	53%
D	0	0	1	2	67%
%	91%	65%	59%	67%	79%

Table 6
MLP confusion matrix for station O in 2019 where A, B, C and D represent range of concentration of Level A (good to moderate), Level B (unhealthy), Level C (very unhealthy) and Level D (hazardous).

	A	B	C	D	%
A	68	19	3	0	76%
B	0	12	10	3	48%
C	0	1	4	0	80%
D	0	0	0	0	–
%	100%	38%	24%	0%	71%

Table 7
Ability of different models to achieve correct level forecasting for station O in Santiago.

Model	Reference	Correct level forecasting
WRF-CHEM	Saide et al., 2011	55%
WRF-CHEM	Saide et al., 2016	63%
MLP	Perez and Salini, 2008	68%
LSTM	This paper	79%

for the forecasting of the three models for station O (west of the city), where more high concentration episodes are observed as compared to in the East. Here the LSTM gives the best results with $R = 0.87$ and NPE = 26,21%. Similar results have been recently found based on LSTM models, combined with Support Vector Regression (Janarthanan et al., 2021) or modifying the connection layers (Lag-FLSTM, Ma et al., 2020). It can be seen that all three models are not very efficient at forecasting the high concentrations that occur during July. Although the traditional MLP does not have the best average performance, it seems to have a similar accuracy to DFFNN but is less efficient than LSTM for the detection of these episodes.

Moving from east to west, the accuracy of forecasting tends to decrease (Table 3). This may be due to the fact that the central and west zone of the city is dominated by primary emissions, and here some of the non-vehicular sources show variations and episodic peaks (Lapere et al., 2020) that may not be captured by the models. This behaviour can be illustrated by representing the values of normalized percent errors in different stations across the city with degrees of coloring (Fig. 8).

These confusion matrices (Table 4, Table 5, and Table 6) not only allow the visualization of whether the level forecasted is correct but also, to what levels the erroneously forecasted levels go. In the case of station O, where an important number of episodes are observed, LSTM has a 91% agreement for level A forecasting and 94% of the level A forecasted really occurred. Considering all the four levels, 79% of them are forecasted correctly (Table 5).

We observe that LSTM is the model that has the best performance for $PM_{2.5}$ forecasting in Santiago, both in the east of the city, where secondary particle formation is more relevant and, in the west, where primary emissions are predominant. Additionally we verified the non-stationarity of the observed series of $PM_{2.5}$ in 2019 with a Dickey-Fuller test (Cheung and Lai, 1995) in Central, West and East stations (p value <0.005). Similar results to ours show that the extreme events of non-stationary series are usually better represented by non-linear models with temporary dependence (Guoyan et al., 2021; Lu et al., 2021). This suggests that LSTM models with temporal units are a good approach for the forecasting of air pollutant non-stationarity time series with extreme events.

We can compare the results of our LSTM model with those of deterministic and MLP models previously reported for Santiago (Table 7). The current operating model correctly predicts 63% of $PM_{2.5}$ levels (Saide et al., 2016), LSTM model obtains a 79% success rate. Our results show that a deep learning model approach can improve air quality and be adaptive to changes and evolution of the city.

4. Conclusions

Operating a reliable air quality forecasting model in populated cities is one of the most important tools for health protection. We have shown that deep learning neural networks are among the recommended methodologies when creating public policies that seek to improve pollution conditions which will help to develop more sustainable cities.

Deep Learning offers the scope and possibility for optimization based on historical data. Therefore, it is very important to have a network of monitoring stations that generate continuous data with supervision guaranteeing only a small percentage of missing values. The feasibility of reconstructing these gaps in the data with the DCT-based method has been shown, allowing continuity of temporal data to carry out the model training.

We show that LSTM models can better respond to $PM_{2.5}$ event prediction by selecting the appropriate pollutant and meteorological variables. In particular, the LSTM models through their memory units can remember important synoptic patterns over time which are useful for $PM_{2.5}$ forecasting. The results obtained with this model configuration are comparable with that obtained in other modified LSTM models for air quality.

In stations where secondary pollutants such as O_3 are predominant (east of the city) our models obtain better results than in stations where there is a preponderance of primary sources (west of the city). Stations with primary sources of pollution have larger variations from the typical patterns, making it difficult to forecast aerosols.

Overall, for Santiago de Chile LSTM models were the best available tool for the development of statistical air pollution forecasting models and these are easily adaptable to different locations.

Declaration of Competing Interest

The authors of this manuscript declare no conflict of interest.

Acknowledgements

We would like to thank the support from the research office, Universidad de Santiago de Chile (DICYT) through project 091931PJ. CM and ZLF acknowledge Center for Climate and Resilient Research (FONDAP15110009) support.

The models were performed in infrastructure provided by Leftraru supercomputer from the National Laboratory of High-Performance Computing Chile (NLHPC).

References

Ausati, S., Amamollahi, J., 2016. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR and MLR models in predicting PM2.5. Atmos. Environ. 142, 465–474.

Borrego, C., Monteiro, A., Ferreira, J., Miranda, A., Costa, A., Carvalho, A., Lopes, M., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. Environ. Int. 34 (5), 613–620. https://doi.org/10.1016/J.ENVINT.2007.12.005.

Chen, Y.C., Li, D.C., 2021. Selection of key features for PM2.5 prediction using a wavelet model and RBF-LSTM. Appl. Intell. 51, 2534–2555. https://doi.org/10.1007/s10489-020-02031-5.

Cheung, Y.-W., Lai, K.S., 1995. Lag order and critical values of the augmented dickey–fuller test. J. Bus. Econ. Stat. 13 (3), 277–280. https://doi.org/10.1080/07350015.1995.10524601.

de Winter, J.C.F., Gosling, S.D., Potter, J., 2016. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. Psychol. Methods 21 (3), 273–290. https://doi.org/10.1037/met0000079.

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., Garcia, R., 2000. Incorporating second-order functional knowledge for better option pricing. Adv. Neural Inf. Proces. Syst. 13, 472–478.

Eckle, K., Schmidt-Hieber, J., 2019. A comparison of deep networks with ReLu activation function and linear spline-type methods. Neural Netw. 110, 232–242.

Fernando, H.J.S., Mammarella, M.C., Grandoni, C., Fedele, P., Di Marco, R., Dimitrova, R., Hyde, P., 2012. Forecasting PM10 in metropolitan areas: efficacy of neural networks. Environ. Pollut. 163, 62–67.

Gallardo, L., Barraza, F., Ceballos, A., Galleguillos, M., Huneeus, N., Lambert, F., Ibarra, C., Munizaga, M., O'Ryan, R., Osses, M., Tolvett, S., Urquiza, A., Veliz, K.D., 2018. Evolution of air quality in Santiago: the role of mobility and lessons from the science-policy interface. Elementa 6, 38. https://doi.org/10.1525/elementa.293.

Garcia, D., 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. Comput. Stat. Data Anal. 54 (4), 1167–1178. https://doi.org/10.1016/j.csda.2009.09.020.

Goodfellow, I., Bengio, Y., Courville, A., 2016. DeepLearning. MITPress. http://www.deeplearningbook.org.

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

Guoyan, H., Li, X., Zhang, B., Ren, J., 2021. PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Sci. Total Environ. 768, 144516. https://doi.org/10.1016/j.scitotenv.2020.

Hajizadeh, Y., Jafari, N., Mohammadi, A., Momtaz, S.M., Fanaei, F., 2020. Abdolahnejad, A. 2020. Concentrations and mortality due to short- and long-term exposure to PM2.5 in a megacity of Iran (2014–2019). Environ. Sci. Pollut. Res. 27, 38004–38014. https://doi.org/10.1007/s11356-020-09695-z.

Honoré, C., Rouil, L., Vautard, R., Beekmann, M., Bessagnet, B., Dufour, A., Elichegaray, C., Flaud, J.M., Malherbe, L., Meleux, F., Menut, L., Martin, D., Peuch, A., Peuch, V.-H., Poisson, N., 2008. Predictability of European air quality: the assessment of three years of operational forecasts and analyses by the PREV'AIR system. J. Geophys. Res. 113, D04301 https://doi.org/10.1029/2007JD008761.

Janarthanan, R., Partheeban, P., Somasundaram, K., Navin Elamparithi, P., 2021. A deep learning approach for prediction of air quality index in a metropolitan city. Sustain. Cities Soc. 67, 102720. https://doi.org/10.1016/j.scs.2021.102720.

Jung, C.R., Lin, Y.T., Hwang, B.F., 2015. Ozone, particulate matter, and newly diagnosed Alzheimer's disease: a population-based cohort study in Taiwan. J. Alzheimers Dis. 44 (2), 573–584.

Kaimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X., Sachdeva, S., 2019. Evaluation of different machine learning approaches to forecasting PM2.5 mass concentrations. Aerosol Air Qual. Res. 19, 1400–1410.

Kim, K.-H., Kabir, E., Kabir, S., 2015. A review on the human health impact of airborne particulate matter. Environ. Int. 74, 136–143.

Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization. ICLR 2015. https://arxiv.org/abs/1412.6980](https://arxiv.org/abs/1412.6980).

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in Central Helsinki. Atmos. Environ. 37, 4539–4550.

Langner, J., Gidhagen, L., Bergstrom, R., Gramsch, E., Oyola, P., Reyes, F., Segersson, D., Aguilera, C., 2020. Model-simulated source contributions to PM2.5 in Santiago and the central region of Chile. Aerosol Air Qual. Res. 20, 1111–1126. https://doi.org/10.4209/aaqr.2019.08.0374.

Lapere, R., Menut, L., Mailler, S., Huneeus, N., 2020. Soccer games and record-breaking PM2.5 pollution events in Santiago, Chile. Atmos. Chem. Phys. 20, 4681–4694. https://doi.org/10.5194/acp-20-4681-2020.

Li, S., Xie, G., Ren, J., Guo, L., Yang, Y., Xu, X., 2020. Urban PM2.5 concentration prediction via attention-based CNN-LSTM. Appl. Sci. 10, 1953. https://doi.org/10.3990/app10061953.
Liu, C., Chen, R., Sera, F., et al., 2019a. Ambient particulate air pollution and daily mortality in 652 cities. N. Engl. J. Med. 381 (8), 705–715.
Liu, Y., Wang, B., Zhu, Q., et al., 2019b. Dominant synoptic patterns and their relationships with PM2.5 pollution in Winter over the Beijing-Tianjin-Hebei and Yangtze River Delta Regions in China. J. Meteorol. Res. 33, 765–776. https://doi.org/10.1007/s13351-019-9007-z.
Lu, G., Yu, E., Wang, Y., Li, H., Cheng, D., Huang, L., Liu, Z., Manomaiphiboon, K., Li, L., 2021. A novel hybrid machine learning method (OR-ELM-AR) used in forecast of PM2.5 concentrations and its forecast performance evaluation. Atmosphere 12, 78. https://doi.org/10.3390/atmos12010078.
Ma, J., Ding, Y., Cheng, J.C.P., Jiang, F., Gan, V.J.L., Xu, Z., 2020. A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant PM2.5 prediction. Sustain. Cities Soc. 60, 102237. https://doi.org/10.1016/j.scs.2020.102237.
Mao, Y., Lee, S., 2019. Deep Convolutional Neural Network for Air Quality Prediction. IOP Conf. Series: Journal of Physics: Conf. Series, 1302 032046. IOP Publishing. https://doi.org/10.1088/1742-6596/1302/3/032046.
Mao, W., Wang, W., Jiao, L., Zhao, S., Liu, A., 2021. Modeling air quality prediction using a deep learning approach: method optimization and evaluation. Sustain. Cities Soc. 65, 102567. https://doi.org/10.1016/j.scs.2020.102567.
Mavroidis, I., Chaloulakou, A., 2011. Long-term trends of primary and secondary NO2 production in the Athens area. Variation of the NO2/NOx ratio. Atmos. Environ. 45, 6872–6879.
Mc Keen, S., Chung, S.H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G.R., Mathur, R., Yu, S., 2007. Evaluation of several PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study. J. Geophys. Res. 112. D10S20.
Menares, C., Gallardo, L., Karakidu, M., Seguel, R., Huneeus, N., 2020. Increasing trends (2001–2018) in photochemical activity and secondary aerosols in Santiago, Chile. Tellus Ser. B Chem. Phys. Meteorol. 72 (1), 1–18.
Moisan, S., Herrera, R., Clements, A., 2018. A dynamic multiple equation approach for forecasting PM2.5 pollution in Santiago, Chile. Int. J. Forecast. 34, 566–581.
Nainggolan, R., Perangin-angin, R., Simarmata, E., Tarigan, F.A., 2019. Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the elbow method. J. Phys. Conf. Ser. 1361 https://doi.org/10.1088/1742-6596/1361/1/012015.
Oprea, M., Mihalache, S.F., Popescu, M., 2016. Applying artificial neural networks to short-term PM2.5 forecasting modelling. IFIP Adv. Inform. Commun. Technol. 475, 204–2011.
Perez, P., 2012. Combined model for PM10 forecasting in a large city. Atmos. Environ. 60, 271–276.
Perez, P., Gramsch, E., 2016. Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. Atmos. Environ. 124, 22–27.
Perez, P., Menares, C., 2018. Forecasting of hourly PM2.5 in south-west zone in Santiago de Chile. Aerosol Air Qual. Res. 18, 2666–2679.
Perez, P., Reyes, J., 2006. An integrated neural network model for PM10 forecasting. Atmos. Environ. 40 (16), 2845–2851.
Perez, P., Salini, G., 2008. PM2.5 forecasting in a large city: comparison of three methods. Atmos. Environ. 42 (35), 8219–8224. https://doi.org/10.1016/j.atmosenv.2008.07.035.
Perez, P., Menares, C., Ramirez, C., 2020. PM2.5 forecasting in Coyhaique, the most polluted city in the Americas. Urban Clim. 32 https://doi.org/10.1016/j.uclim.2020.100608.
Rouïl, L., Honoré, C., Vautard, R., Beekmann, M., Bessagnet, B., Malherbe, L., Meleux, F., Dufour, A., Elichegaray, C., Flaud, J.-M., Menut, L., Martin, D., Peuch, A., Peuch, V.-H., Poisson, N., 2009. Prev'air: an operational forecasting and mapping system for air quality in Europe. Bull. Amer. Meteor. Soc. 90, 73–83.
Rutllant, J., Garreaud, R., 1995. Meteorological air pollution potential for Santiago, Chile: towards an objective episode forecasting. Environ. Monit. Assess. 34, 223–244. https://doi.org/10.1007/BF00554796.
Saide, P.E., Carmichael, G.R., Spak, S.N., Gallardo, L., Osses, A.E., Mena-Carrasco, M., Pagoswski, M., 2011. Forecasting PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model. Atmos. Environ. 45, 2760–2780.
Saide, P.E., Mena-Carrasco, M., Tolvett, S., Hernandez, P., Carmichael, G.R., 2016. Air quality forecasting for winter- time PM2.5 episodes occurring in multiple cities in central and southern Chile. J. Geophys. Res.-Atmos. 121 (1), 558–575.
Samal, K.K.R., Babu, K.S., Das, S.K., 2021. Multi-directional temporal convolutional artificial neural network for PM2.5 forecasting with missing values: a deep learning approach. Urban Clim. 36, 100800.
Tagle, M., Reyes, F., Vasquez, Y., Carbone, S., Saarikoski, S., Timonen, H., Gramsch, E., Oyola, P., 2018. Spatiotemporal variation in composition of submicron particles in Santiago Metropolitan Region, Chile. Front. Environ. Sci. 6, 27. https://doi.org/10.3389/fenvs.2018.00027.
Thomas, S., Jacko, R.B., 2007. Model for forecasting expressway fine particulate matter and carbon monoxide concentration: application of regression and neural network models. Air & Waste Manage. Assoc. 57, 480–488.
Toro, R., Kvakic, M., Klaic, Z., Koracin, D., Morales, R., Leiva, M., 2019. Exploring atmospheric stagnation during a severe particulate matter air pollution episode over complex terrain in Santiago, Chile. Environ. Pollut. 244, 705–714.
USACH, 2014. Update of Emissions Inventory of Atmospheric Pollutants in the Metropolitan Region of Santiago, Chile. http://metadatos.mma.gob.cl/sinia/articles56914_Inf_Inventarios_FINAL.pdf.
WHO, 2021. Global Urban Ambient Air Pollution Database (update 2016). https://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_summary_results_2016_v02.pdf.
Zhou, G., Xu, J., Xie, Y., Chang, L., Gao, W., Gu, Y., Zhou, J., 2017. Numerical air quality forecasting over eastern China: an operational application of WRF-Chem. Atmos. Environ. 153, 94–108.
Zinkevich, M., 2003. Online convex programming and generalized infinitesimal gradient ascent. In: Twentieth International Conference on Machine Learning.