Modelado_Miguel.pdf

Universidad Peruana Union

Detalles del documento

Identificador de la entrega

trn:oid:::29566:438452697

Fecha de entrega

11 mar 2025, 9:56 a.m. GMT-5

Fecha de descarga

11 mar 2025, 11:41 a.m. GMT-5

Nombre de archivo

Modelado_Miguel_ApplSci (1).pdf

Tamaño de archivo

1.3 MB

11 Páginas

6,569 Palabras

30,288 Caracteres

14% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Exclusiones

- N.º de fuentes excluidas

Grupos de coincidencias

- **66** Sin cita o referencia 12%
  Coincidencias sin una citación ni comillas en el texto
- **16** Faltan citas 3%
  Coincidencias que siguen siendo muy similar al material fuente
- **0** Falta referencia 0%
  Las coincidencias tienen comillas, pero no una citación correcta en el texto
- **0** Con comillas y referencia 0%
  Coincidencias de citación en el texto, pero sin comillas

Fuentes principales

10%	Fuentes de Internet
10%	Publicaciones
12%	Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Grupos de coincidencias

66 Sin cita o referencia 12%
Coincidencias sin una citación ni comillas en el texto

16 Faltan citas 3%
Coincidencias que siguen siendo muy similar al material fuente

0 Falta referencia 0%
Las coincidencias tienen comillas, pero no una citación correcta en el texto

0 Con comillas y referencia 0%
Coincidencias de citación en el texto, pero sin comillas

Fuentes principales

10% Fuentes de Internet
10% Publicaciones
12% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	kipdf.com	<1%
2	Internet	opentransportationjournal.com	<1%
3	Internet	ses.library.usyd.edu.au	<1%
4	Internet	www.cedtech.net	<1%
5	Internet	www.wjgnet.com	<1%
6	Publicación	Ahmad Z. Naser, M.Z. Naser. "SPINEX-TimeSeries: Similarity-based predictions wit...	<1%
7	Trabajos entregados	Cranfield University on 2009-02-05	<1%
8	Publicación	Jorge Sánchez-Garcés, Juan J. Soria, Josué E. Turpo-Chaparro, Himer Avila-George, ...	<1%
9	Publicación	Narayanan, Hari. "Classifying Ragams in Carnatic Music with Machine Learning M...	<1%
10	Trabajos entregados	University of Brighton on 2016-05-03	<1%

11 Trabajos entregados
University of Macau on 2016-07-14 <1%

12 Trabajos entregados
University of Sheffield on 2016-05-18 <1%

13 Internet
futur.upc.edu <1%

14 Internet
hugepdf.com <1%

15 Internet
repository.vtc.edu.hk <1%

16 Internet
webbut2.unitbv.ro <1%

17 Internet
1library.org <1%

18 Publicación
Gabriel Huerta. "Time-varying models for extreme values", Environmental and Ec... <1%

19 Publicación
Nadia Cardenas-Escobar, German Granados, Sandra García-Bustos, María Nela Pa... <1%

20 Trabajos entregados
University of Sheffield on 2020-10-14 <1%

21 Publicación
Yutao Zhuang, Mingrui Hao. "Chapter 17 Aerodynamic Parameter Online Identific... <1%

22 Internet
fdocument.org <1%

23 Internet
iris.uniroma1.it <1%

24 Internet
www.arxiv-vanity.com <1%

25 Internet

www.mitpressjournals.org <1%

26 Trabajos entregados

Glasgow Caledonian University on 2021-08-15 <1%

27 Publicación

Madina Shayne, Leonardo A. Molina, Bin Hu, Taylor Chomiak. "Implementing Gait... <1%

28 Publicación

Sachit Mahajan, Ling-Jyh Chen, Tzu-Chieh Tsai. "An empirical study of PM2.5 forec... <1%

29 Publicación

Shuying Leng, Xizhang Gao, Tao Pei, Guoyou Zhang et al. "The Geographical Scien... <1%

30 Trabajos entregados

The Hong Kong Polytechnic University on 2012-04-08 <1%

31 Trabajos entregados

The University of Manchester on 2025-01-24 <1%

32 Trabajos entregados

University of Sydney on 2017-03-02 <1%

33 Trabajos entregados

University of Warwick on 2017-09-14 <1%

34 Internet

agupubs.onlinelibrary.wiley.com <1%

35 Publicación

"Performance Analysis of Machine Learning Singular Spectrum Analysis for Forec... <1%

36 Internet

123dok.org <1%

37 Publicación

Javier Linkolk López-Gonzales, Reinaldo Castro Souza, Felipe Leite Coelho da Silva,... <1%

38 Publicación

Pritha Bhattacharya, Anirudh Agarwal, Deepak Mishra. "Prediction Over Estimati... <1%

39 Internet
api.intechopen.com <1%

40 Internet
etd.aau.edu.et <1%

41 Internet
kobra.uni-kassel.de <1%

42 Internet
socialsciencechronicle.com <1%

43 Trabajos entregados
AUT University on 2024-06-02 <1%

44 Trabajos entregados
Universidad de La Laguna on 2024-11-11 <1%

45 Internet
journals.asmarya.edu.ly <1%

46 Internet
ojs.unitama.ac.id <1%

47 Internet
www.tynxb.org.cn <1%

48 Publicación
Guang Yang, Jun Chen, Zhifan Gao, Heye Zhang, Hao Ni, Elsa Angelini, Raad Mohi... <1%

49 Trabajos entregados
UC, Boulder on 2012-01-20 <1%

50 Internet
www.actuarialsociety.org.za <1%

51 Publicación
"Proceedings of the 7th International Conference of Transportation Research Gro... <1%

52 Trabajos entregados
Charotar University of Science And Technology on 2020-06-24 <1%

53 Publicación
Nicole Jeldes, Germán Ibacache-Pulgar, Carolina Marchant, Javier Linkolk López-G... <1%

54 Trabajos entregados
University of Pittsburgh on 2023-01-27 <1%

55 Trabajos entregados
University of South Florida on 2015-04-09 <1%

56 Internet
cran.r-universe.dev <1%

57 Internet
psasir.upm.edu.my <1%

58 Trabajos entregados
Higher Education Commission Pakistan on 2014-06-12 <1%

59 Publicación
Katsuhiko Takahashi, Shusaku Hiraki, Michio Soshiroda. "Estimating inventory va... <1%

60 Publicación
Maia, A.L.S.. "Forecasting models for interval-valued time series", Neurocomputin... <1%

61 Trabajos entregados
University College London on 2018-09-05 <1%

62 Trabajos entregados
University of Gloucestershire on 2024-01-15 <1%

63 Trabajos entregados
University of the Philippines Los Banos on 2016-11-20 <1%

64 Internet
dadun.unav.edu <1%

65 Internet
www.hindawi.com <1%

66 Internet
www.semanticscholar.org <1%

67 Trabajos entregados
Agricultural University of Iceland on 2023-11-30 <1%

68 Publicación
Hasnain Iftikhar, Justyna Zywiołek, Javier Linkolk López-Gonzales, Olayan Albalaw... <1%

69 Trabajos entregados
University of Birmingham on 2023-12-11 <1%

70 Trabajos entregados
University of Southampton on 2018-09-07 <1%

71 Trabajos entregados
University of York on 2013-09-06 <1%

72 Publicación
Vitor G. Azevedo, Lucila M.S. Campos. "Combination of forecasts for the price of c... <1%

73 Publicación
Zhang, Wang, Liang, Li, Duan. "Short-Term Wind Power Prediction Using GA-BP ... <1%

74 Internet
idoc.tips <1%

75 Publicación
Su-Mei Wang, You-Wu Wang, Yi-Qing Ni, Yang Lu. "Real-Time Malfunction Detecti... <1%

76 Trabajos entregados
University of Leeds on 2018-09-20 <1%

77 Trabajos entregados
University of Mauritius on 2022-07-28 <1%

78 Internet
doi.org <1%

Article

Modeling based on artificial neural networks: Long Short-Term Memory for pollution in Metropolitan Lima

Miguel Angel Solis Teran [1], Felipe Leite Coelho da Silva [2] and Javier Linkolk López-Gonzales [1,*]

[1] Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Lima, Peru
[2] Mathematics Department, Federal Rural University of Rio de Janeiro, RJ, Seropédica, Brazil
* Correspondence: javierlinkolk@gmail.com

Abstract: Particulate matter (PM) is a mixture of fine dust and tiny droplets of liquid suspended in the air. PM_{10} are pollutant particles with a diameter of less than 10 micrometers. These particles are harmful to the respiratory system. The air quality in the region and capital Lima in the Republic of Peru has been investigated in recent years. In this context, statistical analyses of PM_{10} data with forecast models can contribute to planning actions that can improve air quality. The objective of this work is to perform a statistical analysis of the available PM_{10} data and evaluate the quality of time series classical models and neural networks for short-term forecasting. The Box-Jenkins models showed the best performance for short-term forecasting compared to the neural network models considered.

Keywords: neural network; modeling; artificial intelligence

1. Introduction

Currently, the generation of pollutants in the air has experienced a significant increase, predominantly in the form of gases and suspended particles of sizes dangerously small for human health. This situation represents a worrying motivation for environmental health. According to Sánchez [1], it is crucial to understand this phenomenon from a spatiotemporal framework, by studying elements that provide measurable data to explain the interaction between matter and energy in the environment. Similarly, Mahmud [2] suggests the importance of examining different elements within a dynamic field, which changes rapidly due to the variability of its spatial and temporal components, thus allowing objective, explainable and understandable research in the environmental field.

According to [3,4], one of the methods that allows the modeling of suspended xenobiotics is the spatial and temporal analysis of the effects of PM_{10}, as well as the influence of meteorological variables such as temperature and wind on air quality. In [5] they highlight that the understanding of these particles, due to their size, is worrying, especially due to their levels of permanence and invasion in the respiratory tract. In this context, [6] made favorable predictions through singular spectral analysis. This approach is considered a reliable stochastic process to draw a spatiotemporal visualization of PM_{10}, a pollutant with a great impact on exposed populations.

Singular spectral analysis (SSA) has proven to be an effective tool in the prediction of air pollutants such as PM_{10}, due to its ability to decompose complex time series into simpler components, thereby facilitating the identification of underlying patterns and trends in air quality data. According to the study [7], the combination of SSA with machine learning techniques significantly improves the accuracy in prediction of airborne pollutants. Furthermore, the integrated SSA-ARIMA approach, described in [8], enables multi-day forecasts with high confidence, applicable also to particulate matter such as PM_{10}. For its part, the study [9] presents hybrid deep learning models that, combined with SSA, offer high performance in the long-term prediction of pollutants such as PM_{10} and $PM_{2.5}$, highlighting the value of advanced techniques in the evaluation and control of air quality.

Mach tries to take the sensory perception of pollutants towards a current, agile and dynamic quantification system, capable of being modeled and understood through statistical processes. On the other hand, Matthews (2020) states that there are frequent variability measures within the spatial components, subject to certain meteorological factors that can generate white noise. However, Rojas et al (2010) establishes that the time series comprise a dynamic system, defined as a linear combination of various oscillators, which through series projection in principal components developed in the time-frequency domain, which allows establishing dominant oscillation models and associating them with various events compromised with PM_{10} particulates. Aguilar (2017) Singular spectral analysis (SSA) is a fundamental part of the probabilistic prediction of the energy process, autoregressive models to predict modular behaviors of matter. For Li et al (2021), the usefulness of the SSA method accompanied by others that helped to decompose the data improved the spectral analysis of Nx emission, considering its proportions and effects on the environment.

Finally, it is worth pointing out that few studies have been carried out in Peru related to suspended particulate pollutants. The last study related to pollutants smaller than 10 microns by Espinoza (2018) analyzed the spatial and temporal distribution of PM_{10} concentration in Metropolitan Lima in the period 2015-2017 with results above quality standards, however, no specific spectral studies have been carried out in order to trace sufficient frequencies to understand pollution within atmospheric patterns. Plocoste (2017) It is important to note that concentrations have non-linear behavior and fluctuate strongly on space-time scales that allows to visualize and describe widespread pollution and trace its permanence. Armstrong et al (2017) Air pollution is directly related to respirable oxygen being contaminated by compounds such as CO_2, so Goudarzi (2018) concludes that human impact on the planet is notable and the attempt to reduce it is increasingly urgent. In that sense, the objective of this research is to model and forecast the concentration of PM_{10} per hour based on artificial neural networks and traditional models.

2. Methodology

In this study, PM_{10} data from Peru were used. This analysis considered five weather stations located in the region of the province of Lima. The study area comprises the capital of the Peru Republic, the province of Lima (Figure 1). Lima is located at 77^0 W and 12^0 S of the South American continent. The data used covers the period 2017-2018.

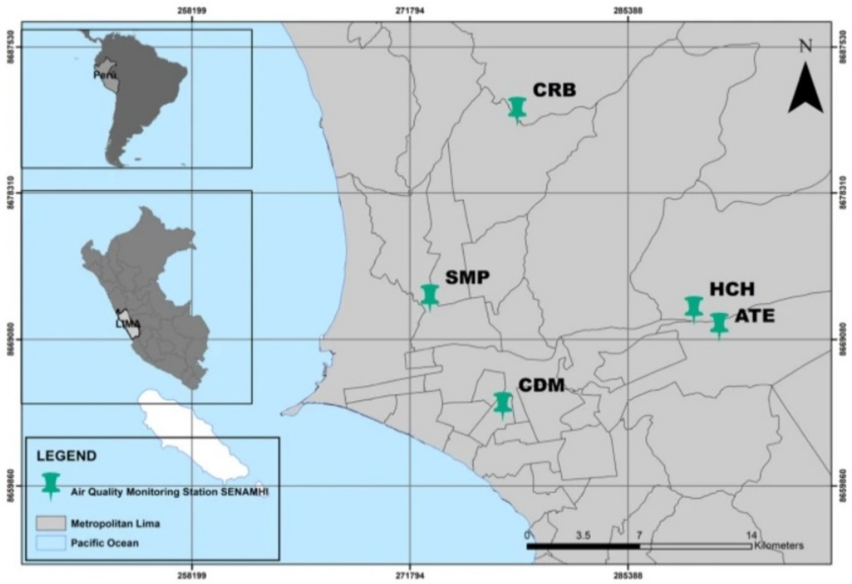

Figure 1. Map with study area and locations of the Lima air quality monitoring stations: Ate (ATE), Campo de Marte (CDM), Carabayllo (CRB), Huachipa (HCH) and San Martin de Porres (SMP).

53

To evaluate the prediction accuracy of the models, the PM_{10} time series was divided into 10 training and test sets. From the first defined training set, a forecast of days seven ahead was carried out. Then, the training set received one more observation, and again the forecast seven days ahead. Following this methodology, the steps mentioned above were carried out 10 times.

In this work, the classical time series models were used, such as the Box-Jenkins and exponential smoothing models. Also, Autoregressive neural network models (NNAR), multilayer perceptrons (MLP) and long short term memory (LSTM) were used. Furthermore, the naive models, TBATS, dynamic linear model (DLM), and singular spectrum analysis (SSA) were adopted.

2.1. Box-Jenkins Models

The Box-Jenkins [10] methodology is widely used in analyzing parametric time series models. This methodology includes fitting integrated autoregressive moving average models, ARIMA(p,d,q), to a data set using the autocorrelation functions between observations.

The model used was seasonal ARIMA (SARIMA), which incorporated the seasonality component into the data. The structure of the seasonal ARIMA model of order (p,d,q) x $(P,D,Q)_s$ is given by

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D Z_t = \theta(B)\Theta(B^s)a_t \tag{1}$$

where a_t is white noise; $\phi(B)$ is the autoregressive operator of order p; $\theta(B)$ is the moving average operator of order q; $\Phi(B^s)$ is the seasonal autoregressive operator of order P; $\Theta(B^s)$ is the seasonal moving average operator of order Q; Δ^d is the simple difference operator and; Δ_s^D is the seasonal difference operator; s is the number of observations per year (period). The Box-Jenkins model was obtained using the algorithm proposed by Hyndman and Khandakar [11].

2.2. Exponential Smoothing Method

The exponential smoothing method was proposed in the 1950s by [12], [13] and [14]. In this method, forecasts are generated from a weighted average of past observations, with the weight of observations decreasing exponentially as the observations age. Hyndman et al. [15] proposed a classification that depends on the time series' error, trend, and seasonality. The error can be additive (A) or multiplicative (M), the trend can be additive (A), additive with damping (A_d), or in the case of non-existence, none (N), in turn, seasonality, can be additive (A), multiplicative (M) or none (N). Therefore, each exponential smoothing model used in this work can be classified as ETS (Error, Trend, Seasonality). Furthermore, the ETS algorithm proposed by Hyndman [16] was used to adjust the exponential smoothing models.

2.3. Neural network autoregression

The NNAR model is capable to capture the time series components of level, trend and seasonality. This paper has used the algorithm proposed by [17].

According to [18], the NNAR model can be defined by

$$\begin{aligned} z_j &= f\left(\alpha_j + \sum_{i=1}^{p} w_{i,j} y_{t-i} + \sum_{i=1}^{P} w_{i,j} y_{t-im}\right), \\ \hat{y} &= g(\alpha_0 + \sum_{j=1}^{k} w_j z_j), \end{aligned} \tag{2}$$

such that z_j represents the j-th hidden layer neuron, p represents the number of input layer neurons of the non-seasonal time series part, P represents the number of input layer of the seasonal part, α_j represents the intercept of the j-th hidden neuron, α_0 is the intercept of the output neuron, $w_{i,j}$ denotes the weights assigned to the connection between the input and the hidden layer, w_j are the weights of the connection between the hidden and the output layer, y_{t-i} and y_{t-im} are the observations (covariates or neurons) of the input layer $(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, y_{t-m}, y_{t-2m}, \ldots, y_{t-Pm})$, $f(\cdot)$ is the activation function introduced in this work by logistic function and $g(\cdot)$ is a function that permits a final transformation of the output.

Thus, the $NNAR(p, P, k)_m$ model considers a feedforward network with one hidden layer, p inputs, k nodes in the hidden layer, P seasonal lags and m periods of time series [18].

2.4. *Multi-layer perceptron*

The MLP model or feedforward neural network is a mathematical function mapping a sort of input values to output values. According to [19], the main objective of a feedforward neural network is to approximate any function f^*, defining a mapping $Y = f(x; \theta)$ and learn the value of the parameter θ that makes the better function approximation.

The MLP model proposed by [20] is defined by

$$\widehat{Y} = v_0 + \sum_{j=1}^{N} v_j g(w_j^T x') \tag{3}$$

where x' is the input vector x, augmented with 1, i.e., $x' = (1, x^T)^T$, w_j is the weight vector for the jth hidden node, v_0 is the output layer intercept, $v_1, \ldots, v_N$ are the weights for the output node, $\widehat{Y}$ is the network output, and g is the activation function used to allow a possible nonlinearity at the hidden layer. We used the logistic activation function. In this work, the algorithm used was developed by [21].

2.5. *Long short-term memory*

The LSTM network is a special kind of recurrent neural network (RNN) and was proposed by [22] to avoid the long-term dependency problem. RNN are useful for time series data considering that each neuron can use its internal ”memory" to maintain the information of the previous input. The fundamental issue in a simple RNN is the vanishing gradient problem, the incapability to capture the long-term dependencies in a series, i.e., the recurrent connections allow a ”memory" of previous inputs to persist in the network's internal state, influencing the network output [23]. This problem is solved using a LSTM network, which considers long-term dependencies in a time series.

In this work, we introduced the LSTM model following the description given by [24] and [25]. Thus, the LSTM structure follows the equations:

$$\begin{aligned}
\mathbf{i}_t &= \sigma(W_{ix}\mathbf{x}_t + W_{ih}\mathbf{h}_{t-1} + \mathbf{b}_i)\\
\mathbf{f}_t &= \sigma(W_{fx}\mathbf{x}_t + W_{fh}\mathbf{h}_{t-1} + \mathbf{b}_f)\\
\mathbf{g}_t &= \tanh(W_{gx}\mathbf{x}_t + W_{gh}\mathbf{h}_{t-1} + \mathbf{b}_g)\\
\mathbf{o}_t &= \sigma(W_{ox}\mathbf{x}_t + W_{oh}\mathbf{h}_{t-1} + \mathbf{b}_o)\\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t\\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(c_t)
\end{aligned} \tag{4}$$

where $\mathbf{h}_t$ is the hidden state at time t, c_t is the cell state at time t, $\mathbf{x}_t$ is the input at time t, $\mathbf{h}_{t-1}$ is the hidden state of layer at time $t-1$ or the initial state at time 0, and $\mathbf{i}_t$, $\mathbf{f}_t$, $\mathbf{g}_t$, $\mathbf{o}_t$ are input gates, forget, cell and output. σ is the activation function (sigmoid), and $\odot$ is the Hadamard product. The W terms denote weight matrices and $\mathbf{b}$ are bias vectors. In this study, it was used the algorithm developed by [26] and [27].

The LSTM model was defined with 50 hidden layer nodes, 100 epochs on training, and an ADAM optimizer with a learning rate equal to 0.001. This configuration was proposed following [28] and [29]. To obtain the optimal parameters, the grid search method was used [29].

turnitin Página 13 of 19 - Engrega de integridad Identificador de la entrega trn:oid:::29566:438452697

2.6. *Dynamic linear model*

The DLM or state space model define a very general class of time series models [30]. DLM consider terms to model trends, seasonality, covariates, and autoregressive components. From a state-space structure, the general form of the DLM can be written using two main equations. The so-called observational equation (Equation 5) and the system evolution equation (Equation 6) [30] are given by

$$Y_t = F_t'\theta_t + v_t, \quad v_t \sim N(0, V_t) \tag{5}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \tag{6}$$

where θ_t is the state vector in time t, F_t is a known vector, and G_t is a known matrix of the evolution of the system. It is assumed that V_t is the variance associated with the error of the observational equation and W_t is a covariance matrix associated with the errors of the evolution equation of the system.

As mentioned, F_t and G_t are known and used to build the structure of the model according to the components found in the series to be analyzed - level, trend, and seasonality. In this work, it was adopted a linear growth model with a monthly seasonal component using standard specifications of the package DLM [31]. We performed the estimation of the observational variance V_t and the evolution covariance matrix W_t through MCMC method, using the Gibbs sampler algorithm.

3. Results

In this section, an exploratory analysis will be presented for PM_{10} data from meteorological stations in Peru. Then, the time series models will be used to obtain short-term forecasting of PM_{10}.

3.1. *Exploratory Analysis*

Figure 2 shows the PM_{10} concentration trajectory of the monitoring station in Peru. The HCH station presented the highest PM_{10} concentration followed by the ATE station. On the other hand, the CDM station had the lowest daily PM_{10} concentration rates (Figure 3). The highest daily PM_{10} concentration value occurred in April 2018 at the HCH meteorological station.

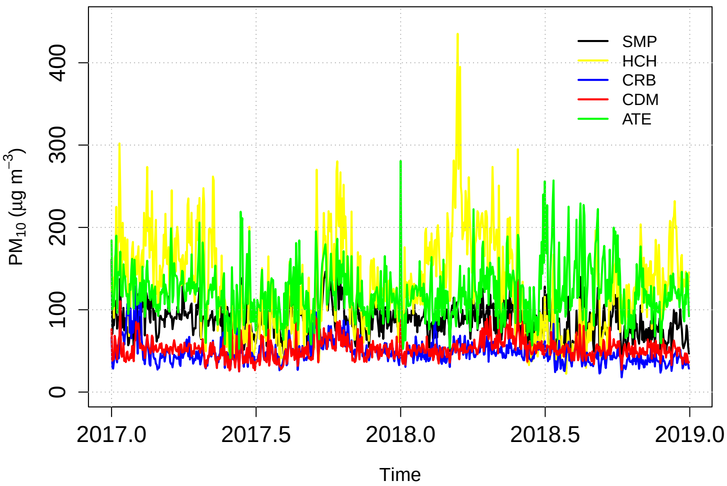

Figure 2. Average daily concentration of PM_{10} between 2017 and 2018 for monitoring station in Peru.

Table 1 provides the main statistical measures of daily PM_{10} concentration for meteorological stations considered in this study. It can be seen in this table that the HCH station

turnitin Página 13 of 19 - Engrega de integridad Identificador de la entrega trn:oid:::29566:438452697

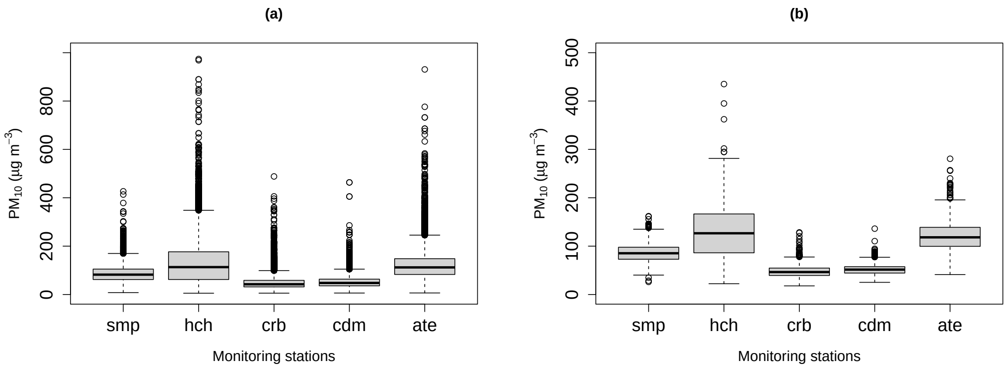

Figure 3. Hourly (a) and daily (b) PM_{10} boxplots for the monitoring stations located in Peru.

has the highest average concentration of PM_{10} and also greater variability. And the ATE station with the lowest average PM_{10}.

Table 1. Statistical measures for daily PM_{10} concentration from weather stations.

Measures	SMP	HCH	CRB	CDM	ATE
Minimum	26.01	22.22	17.84	25.12	41.26
1st Qu.	72.89	86.21	39.38	44.35	99.59
Median	85.29	126.61	46.40	51.46	118.19
Mean	86.05	130.03	48.69	52.30	121.56
3rd Qu.	97.69	166.56	54.63	57.51	138.72
Maximum	161.97	435.15	128.44	136.16	280.72
Standard deviation	20.88	58.63	14.50	12.25	33.29

Hourly temperature and PM_{10} data at each station analyzed did not present a correlation (Figure 4). The HCH station presented the highest correlation between stations, being approximately 0.4, but still a low value.

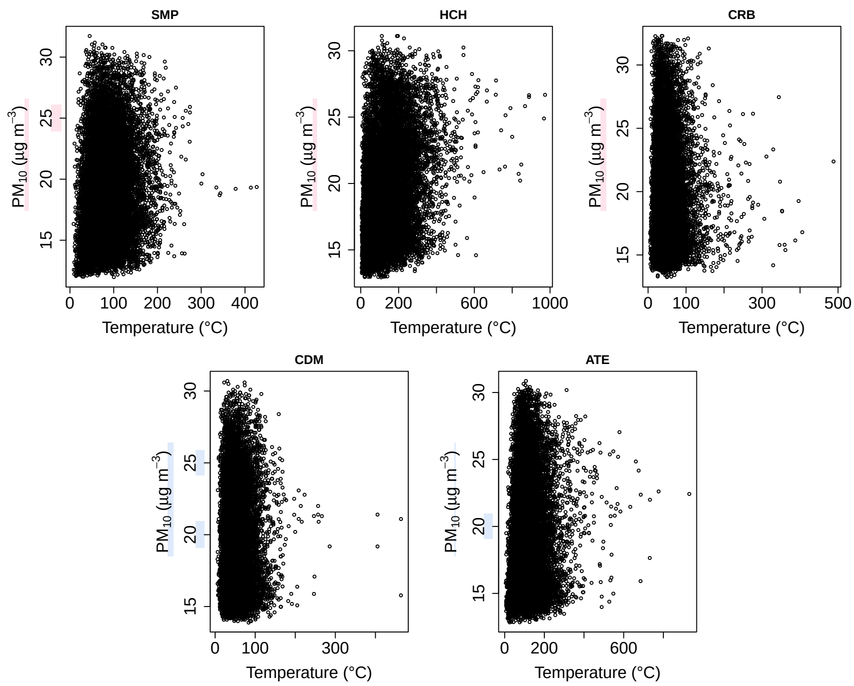

Figure 4. Spatial distribution of the correlation between temperature and PM_{10} at meteorological stations (SMP, HCH, CRB, CDM, ATE) in Peru.

turnitin Página 15 of 19 - Engrega de integridad Identificador de la entrega trn:oid:::29566:438452697

Figure 5(a) presents the spatial correlation between weather stations considered in Peru. This figure shows that the highest correlation (0.63) occurred between the SMP and ATE stations. On the other hand, the spatial correlation between monitoring stations for the temperature variable is greater than 0.70 (Figure 5(b)). Therefore, there is a low spatial correlation between some air quality monitoring stations for variable PM_{10} compared with temperature variable.

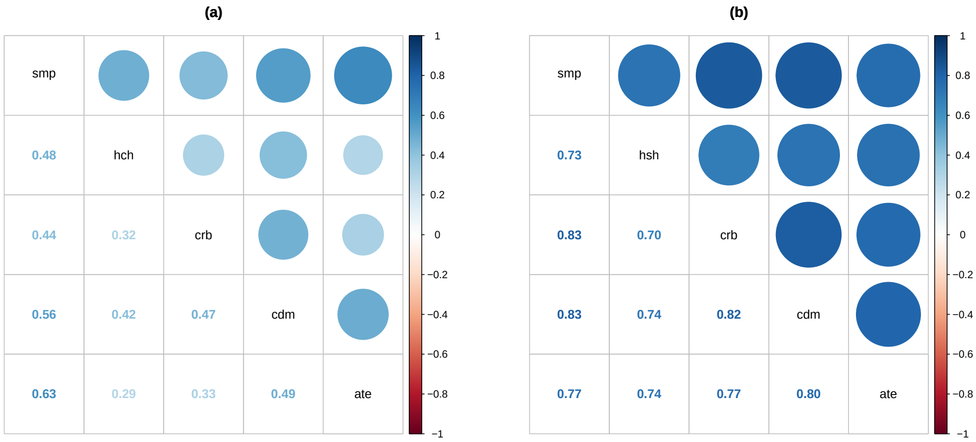

Figure 5. Spatial correlation between weather stations for PM_{10} (a) and temperature (b) in Peru.

3.2. Application of forecasting models.

Table 2 presents the RMSE and sMAPE accuracy measures for the forecast models in each test set and the average. Based on the average values of the RMSE and sMAPE metrics, the Box-Jenkins model presented the best prediction accuracy for PM_{10} data of SMP station.

Table 2. Forecasting performance of the 9 models applied to the PM_{10} data of the SMP station via RMSE and sMAPE metrics.

Metric	Sets	ETS	ARIMA	TBATS	NNAR	MLP	DLM	SSA	NAÏVE	LSTM
RMSE	1	8.07	12.13	9.99	8.86	10.51	39.73	8.08	15.61	9.89
	2	11.86	13.69	12.28	13.87	13.46	41.70	13.23	12.47	13.19
	3	15.95	15.12	14.93	19.77	17.94	13.32	18.41	15.57	13.62
	4	18.13	16.36	16.95	22.17	20.87	13.70	21.70	17.22	13.18
	5	20.41	15.96	17.77	25.40	23.17	38.76	23.76	22.96	14.92
	6	24.39	15.51	18.87	28.31	26.59	82.10	24.87	37.53	21.54
	7	23.80	15.34	18.06	26.94	24.40	15.91	25.82	21.82	13.98
	8	21.76	15.00	17.47	24.88	23.64	27.77	26.22	15.38	15.65
	9	16.92	12.88	14.75	18.96	20.19	66.44	25.32	10.75	15.07
	10	13.83	14.05	15.02	14.69	20.08	48.24	20.81	14.05	22.23
	Average	17.51	14.60	15.61	20.39	20.08	38.77	20.82	18.34	15.33
sMAPE	1	3.75	5.59	4.06	4.60	5.57	18.37	3.70	8.30	5.07
	2	5.41	7.27	5.91	6.18	6.18	34.34	6.22	5.95	6.67
	3	8.27	8.42	7.97	9.89	8.88	7.39	9.27	8.15	7.91
	4	10.47	10.09	10.07	12.52	11.39	8.37	11.66	10.24	8.20
	5	12.47	10.16	11.08	15.15	13.80	20.65	14.06	13.78	9.82
	6	15.12	10.05	12.17	17.73	16.72	36.88	15.23	22.30	14.22
	7	14.95	10.08	11.45	16.88	15.19	9.89	16.01	13.70	9.29
	8	13.84	9.96	11.30	16.03	15.15	20.17	16.00	10.17	10.32
	9	11.12	8.83	9.96	12.46	13.01	67.75	15.83	6.14	10.05
	10	9.46	9.10	9.82	9.79	12.54	60.07	13.47	8.94	14.46
	Average	10.49	8.95	9.38	12.12	11.84	28.39	12.14	10.77	9.60

Finally, it is worth pointing out that few studies have been carried out in Peru related to suspended particulate pollutants. The last study related to pollutants smaller than

turnitin Página 15 of 19 - Engrega de integridad Identificador de la entrega trn:oid:::29566:438452697

Table 3. Forecasting performance of the nine models applied to the PM_{10} data of the HCH station via RMSE and sMAPE metrics.

Metric	Sets	ETS	ARIMA	TBATS	NNAR	MLP	DLM	SSA	NAÏVE	LSTM
RMSE	1	75.22	41.61	37.84	39.03	38.77	41.33	47.85	56.38	11.12
	2	61.50	42.79	39.08	46.53	40.50	62.16	58.98	36.13	11.68
	3	43.87	37.37	34.54	36.79	30.84	82.41	57.58	22.63	16.23
	4	19.68	23.66	20.93	36.60	15.35	167.44	43.67	33.82	14.78
	5	24.71	26.95	25.17	41.81	17.40	18.76	31.74	16.42	14.91
	6	24.45	22.96	22.12	50.73	16.93	17.26	18.21	17.17	15.01
	7	42.72	30.08	28.90	49.70	21.15	98.88	14.40	32.39	17.06
	8	53.26	33.73	32.87	65.84	29.55	77.75	11.16	39.01	18.88
	9	26.48	22.54	21.04	21.81	15.45	107.38	9.19	12.68	15.71
	10	25.84	24.07	23.39	36.61	18.89	11.32	13.32	12.40	20.76
	Average	39.77	30.58	28.59	42.54	24.48	68.47	30.61	27.90	15.62
sMAPE	1	19.91	11.52	10.43	10.99	10.41	10.80	13.12	15.47	5.72
	2	17.44	12.35	11.22	13.24	11.41	23.61	16.88	10.20	5.91
	3	12.89	11.17	10.24	10.92	9.03	38.54	16.74	7.15	9.17
	4	5.92	6.21	5.88	9.33	4.63	79.16	12.93	11.51	9.01
	5	7.74	7.86	7.47	11.22	5.82	5.95	9.44	4.65	9.81
	6	8.14	6.79	7.05	13.80	5.05	5.10	5.23	4.84	10.08
	7	12.89	9.10	9.01	15.09	6.78	24.55	4.71	10.06	10.99
	8	16.74	11.09	10.91	19.68	9.86	22.33	2.94	12.92	12.02
	9	7.88	7.19	6.77	6.14	5.31	62.35	2.37	3.97	10.39
	10	8.04	7.73	7.46	10.76	5.97	3.55	3.50	4.35	13.48
	Average	11.76	9.10	8.64	12.12	7.43	27.59	8.79	8.51	9.66

Table 4. Forecasting performance of the nine models applied to the PM_{10} data of the CRB station via RMSE and sMAPE metrics.

		ETS	ARIMA	TBATS	NNAR	MLP	DLM	SSA	Ingênuo	Hybrid
RMSE	1	5.61	5.29	5.30	7.40	7.90	20.10	5.67	9.63	5.14
	2	6.30	6.64	6.50	7.81	7.57	40.28	6.65	8.91	6.27
	3	7.12	7.07	6.88	9.43	9.10	8.13	7.75	7.10	7.07
	4	7.28	7.04	6.84	10.42	9.34	12.49	7.75	7.52	7.15
	5	7.74	7.26	6.89	11.30	10.02	12.19	8.29	8.21	7.50
	6	10.31	8.87	8.57	13.27	12.65	38.32	8.85	16.02	9.58
	7	9.57	7.97	6.43	11.93	11.46	8.86	9.18	8.11	8.75
	8	9.18	8.19	7.85	12.11	11.76	6.97	9.50	8.04	8.68
	9	4.33	4.18	3.50	6.96	6.87	45.43	8.21	8.40	4.18
	10	4.38	5.41	5.09	9.29	8.76	11.74	6.93	5.54	4.80
	Average	7.18	6.79	6.38	9.99	9.54	20.45	7.88	8.75	6.91
SMAPE	1	5.70	4.26	4.22	7.72	8.03	18.74	5.31	9.49	4.38
	2	4.89	6.87	6.42	6.61	6.17	70.78	5.98	10.51	5.38
	3	6.80	7.38	7.21	8.98	8.29	8.71	7.15	7.76	7.08
	4	7.06	7.05	6.95	10.56	9.23	11.87	7.30	6.96	7.06
	5	7.91	7.64	7.31	12.32	10.58	12.79	8.37	8.49	7.77
	6	11.50	10.09	9.71	15.07	14.33	32.63	9.49	17.73	10.80
	7	10.66	8.64	7.19	13.39	12.84	11.08	9.91	8.69	9.59
	8	10.51	9.29	8.61	14.26	13.90	7.69	10.10	9.01	9.91
	9	4.89	4.45	3.94	7.87	7.67	79.69	9.08	11.49	4.65
	10	5.20	6.11	5.80	9.99	8.78	18.29	7.91	6.57	5.67
	Average	7.51	7.18	6.74	10.68	9.98	27.23	8.06	9.67	7.23

10 microns by Espinoza (2018) analyzed the spatial and temporal distribution of PM_{10} concentration in Metropolitan Lima in the period 2015-2017 with results above quality standards, however, no specific spectral studies have been carried out in order to trace sufficient frequencies to understand pollution within atmospheric patterns. Plocoste (2017) It is important to note that concentrations have non-linear behavior and fluctuate strongly on space-time scales that allows to visualize and describe widespread pollution and trace

	ETS	ARIMA	TBATS	NNAR	MLP	DLM	SSA	Ingênuo	Hybrid
1	7.44	6.70	6.38	7.59	8.47	4.66	5.56	7.16	7.06
2	8.61	8.63	8.35	9.61	9.54	24.27	7.37	7.11	8.61
3	9.70	9.56	9.17	10.77	10.98	26.21	8.04	11.70	9.63
4	10.18	10.03	9.62	12.22	11.31	6.82	8.76	9.09	10.10
5	10.75	10.53	10.13	11.87	11.97	20.89	8.84	12.91	10.64
6	10.96	10.78	10.36	13.26	12.14	14.37	10.30	6.40	10.86
7	8.92	9.49	9.07	12.05	10.16	20.62	8.86	4.57	9.19
8	7.72	8.73	8.36	11.59	9.51	8.32	8.41	4.66	8.20
9	5.45	7.18	6.65	8.61	7.75	29.59	8.52	8.71	6.23
10	5.80	8.13	7.74	10.48	9.55	11.52	8.87	4.75	6.87
11	8.55	8.98	8.58	10.80	10.14	16.73	8.36	7.71	8.74

its permanence. Armstrong et al (2017) Air pollution is directly related to respirable oxygen being contaminated by compounds such as CO_2, so Goudarzi (2018) concludes that human impact on the planet is notable and the attempt to reduce it is increasingly urgent. In that sense, the objective of this research is to model and forecast the concentration of PM_{10} per hour based on artificial neural networks and traditional models.

	ETS	ARIMA	TBATS	NNAR	MLP	DLM	SSA	Ingênuo	Hybrid
1	6.47	6.04	5.83	6.23	7.12	3.91	5.18	6.31	6.25
2	7.58	7.76	7.58	8.70	8.27	39.34	6.60	6.70	7.67
3	9.02	9.02	8.71	9.69	9.99	20.76	7.68	10.64	9.02
4	10.18	10.09	9.74	11.74	11.06	6.98	8.90	9.24	10.13
5	10.91	10.74	10.33	11.69	12.08	19.51	9.04	13.03	10.82
6	11.42	11.16	10.71	13.42	12.46	17.33	10.72	6.57	11.29
7	8.78	9.39	8.86	11.84	10.08	28.30	8.88	4.40	9.09
8	7.24	8.75	8.32	11.99	9.67	8.93	8.34	4.62	8.00
9	5.36	6.49	5.64	7.45	6.57	54.97	8.15	9.70	5.64
10	6.00	7.80	7.29	10.35	8.85	10.21	8.47	5.21	6.69
11	8.30	8.72	8.30	10.31	9.61	21.02	8.20	7.64	8.46

Analysis of PM_{10} concentrations at different monitoring stations in Metropolitan Lima reveals significant spatial heterogeneity, with Huachipa (HCH) recording the highest levels. This finding may be related to specific emission sources, such as industries and heavy traffic, underlining the need for targeted local policies. Furthermore, the observed temporal fluctuations, influenced by factors such as seasons and meteorological changes, indicate the importance of considering spatiotemporal dynamics when planning interventions. Previous studies conducted in other cities have found significant correlations between PM_{10} and factors such as wind and humidity, but the low correlation found in Lima suggests that local dynamics could respond to different factors, such as topography and urbanization.

The results of this study partially align with international research highlighting the usefulness of advanced tools such as SSA and hybrid models to predict pollutants. However, the lower effectiveness of neural networks in this case highlights the importance of adapting models to local particularities. This includes not only adjusting the techniques used, but also expanding the database to improve predictive capacity. Likewise, a more robust integration of factors such as environmental policies, urban growth and human behavior could enrich the models and improve mitigation strategies. In this sense, the study raises the need for an interdisciplinary approach to effectively address the challenges of air pollution in Metropolitan Lima and other similar cities.

4. Conclusions

70

This study demonstrates that classical time series models, such as ARIMA, offer better performance in predicting PM_{10} concentrations in Metropolitan Lima compared to neural networks, specifically LSTMs. This finding suggests that for data with seasonal patterns and relatively short time series, traditional models may be more efficient and robust. Although neural networks have the potential to capture more complex relationships and long-term dependencies, their performance may be limited by hyperparameter settings and intrinsic data characteristics.

Data analysis indicates that stations such as Huachipa (HCH) have significantly higher PM_{10} concentrations, which could be associated with local factors such as industrial activity and vehicular traffic. These results have important implications for public health, since PM_{10} particles are directly linked to respiratory and cardiovascular diseases. In addition, the levels recorded at several stations exceed international air quality standards, highlighting the urgent need to implement pollution mitigation policies in Metropolitan Lima.

The use of advanced tools such as singular spectral analysis (SSA) and neural networks, together with classical models, provides valuable insight for air pollutant prediction. However, the results underline the importance of adapting these approaches to local conditions and improving data availability to optimize forecast accuracy. Future research could explore hybrid models that combine the best of classical and modern approaches, as well as consider the influence of meteorological variables and human activities on pollutant dynamics. This will strengthen environmental management strategies and reduce the impact of pollution on community health.

References

1. Sánchez-Balseca, J.; Pérez-Foguet, A. Spatio-temporal air pollution modelling using a compositional approach. *Heliyon* **2020**, *6*.
2. Mahmud, H.; Shobnom, K.; Ali, M.R.; Muntakim, N.; Kulsum, U.; Baroi, D.S.; Ahmed, Z.; Rahman, M.M.; Hassan, M.Z. Micro-environmental dynamics of particulate (PM2. 5 and PM10) air pollution in Rajshahi City: a spatiotemporal analysis. *Management of Environmental Quality: An International Journal* **2024**.
3. Akdi, Y.; Gölveren, E.; Ünlü, K.D.; Yücel, M.E. Modeling and forecasting of monthly PM 2.5 emission of Paris by periodogram-based time series methodology. *Environmental Monitoring and Assessment* **2021**, *193*, 1–15.
4. Leng, S.; Gao, X.; Pei, T.; Zhang, G.; Chen, L.; Chen, X.; He, C.; He, D.; Li, X.; Lin, C.; et al. Tempo-Spatial Processes and Modelling of Environmental Pollutants. *The Geographical Sciences During 1986—2015: From the Classics To the Frontiers* **2017**, pp. 367–390.
5. Thompson, J.E. Airborne particulate matter: human exposure and health effects. *Journal of occupational and environmental medicine* **2018**, *60*, 392–423.
6. Cekim, H.O. Forecasting PM10 concentrations using time series models: a case of the most polluted cities in Turkey. *Environmental Science and Pollution Research* **2020**, *27*, 25612–25624.
7. Performance Analysis of Machine Learning Singular Spectrum Analysis for Forecasting Air Contamination, 2023. https://doi.org/10.1109/ICSCAN58655.2023.10395309.
8. Kumar, U.; et al. An integrated SSA-ARIMA approach to make multiple day ahead forecasts for the daily maximum ambient O3 concentration. *Aerosol and Air Quality Research* **2015**, *15*, 208–219.
9. Cai, J.; Gu, C.; Fang, K.; Wang, L.; Lv, M. Long-term PM2.5 Concentration Prediction Using Hybrid Deep Learning Model, 2023. https://doi.org/10.1109/CAIT59945.2023.10469109.
10. Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*; Holden-Day series in time series analysis and digital processing, Holden-Day, 1970.
11. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* **2008**, *26*, 1–22. https://doi.org/10.18637/jss.v027.i03.
12. Brown, R.G. Statistical forecasting for inventory control, 1959.
13. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages (Office of Naval Research, Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. Reprinted in the International Journal of Forecasting, 2004, 1957. https://doi.org/https://doi.org/10.1016/j.ijforecast.2003.09.015".
14. Winters, P.R. Forecasting seasonals and trends by exponentially weighted moving averages. *Management Science* **1960**, *6*, 324–342. https://doi.org/https://doi.org/10.1287/mnsc.6.3.324.
15. Hyndman, R.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Australia, 2018.
16. Hyndman, R.; Koehler, A.; Ord, K.; Snyder, R. Forecasting with exponential smoothing. The state space approach, 2008. https://doi.org/10.1007/978-3-540-71918-2.
17. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software* **2008**, *27*, 1–22. https://doi.org/10.18637/jss.v027.i03.
18. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: principles and practice*; OTexts, 2018.
19. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep learning*; MIT press, 2016.
20. Ahmed, N.; Atiya, A.; Gayar, N.; El-Shishiny, H. An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews* **2010**, *29*, 594–621. https://doi.org/10.1080/07474938.2010.481556.

21. Kourentzes, N. *Time Series Forecasting with Neural Networks*, 2022. R package version 0.9.8.
22. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural computation* **1997**, *9*, 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
23. Graves, A. *Supervised Sequence Labelling with Recurrent Neural Networks*, 1 ed.; Springer-Verlag Berlin Heidelberg, 2012. https://doi.org/http://doi.org/10.1007/978-3-642-24797-2.
24. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE transactions on smart grid* **2017**, *10*, 841–851. https://doi.org/10.1109/TSG.2017.2753802.
25. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation* **2019**, *31*, 1235–1270, [https://direct.mit.edu/neco/article-pdf/31/7/1235/1053200/neco_a_01199.pdf]. https://doi.org/10.1162/neco_a_01199.
26. Allaire, J.; Chollet, F. *keras: R Interface to 'Keras'*, 2023. R package version 2.11.1.
27. Allaire, J.; Tang, Y. *tensorflow: R Interface to 'TensorFlow'*, 2022. R package version 2.11.0.
28. Karevan, Z.; Suykens, J.A. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks* **2020**, *125*, 1–9. https://doi.org/https://doi.org/10.1016/j.neunet.2019.12.030.
29. Abbasimehr, H.; Shabani, M.; Yousefi, M. An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering* **2020**, *143*, 106435. https://doi.org/https://doi.org/10.1016/j.cie.2020.106435.
30. West, M.; Harrison, J. *Bayesian forecasting and dynamic models*; Springer Science and Business Media, 2006.
31. Petris, G. An R Package for Dynamic Linear Models. *Journal of Statistical Software* **2010**, *36*, 1–16.