



OPEN ACCESS

EDITED BY

Daniel Okoh,
The National Space Research and
Development Agency (NASRDA), Nigeria

REVIEWED BY

Xiukuan Zhao,
Chinese Academy of Sciences (CAS), China
Taiwo Ojo,
University of Toronto, Canada

*CORRESPONDENCE

Christian Mauricio,
✉ 20220334@lamolina.edu.pe

RECEIVED 01 June 2024

ACCEPTED 09 August 2024

PUBLISHED 24 September 2024

CITATION

Mauricio C, Suclupe J, Milla M, López de
Castilla C, Kuyeng K, Scipion D and
Rodriguez R (2024) Short-term prediction of
horizontal winds in the mesosphere and
lower thermosphere over coastal Peru using a
hybrid model.
Front. Astron. Space Sci. 11:1442315.
doi: 10.3389/fspas.2024.1442315

COPYRIGHT

© 2024 Mauricio, Suclupe, Milla, López de
Castilla, Kuyeng, Scipion and Rodriguez. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Short-term prediction of horizontal winds in the mesosphere and lower thermosphere over coastal Peru using a hybrid model

Christian Mauricio^{1*}, Jose Suclupe², Marco Milla³, Carlos López de Castilla¹, Karim Kuyeng⁴, Danny Scipion⁴ and Rodolfo Rodriguez⁵

¹Departamento de Estadística e Informática, Universidad Nacional Agraria La Molina, Lima, Perú,

²Leibniz Institute of Atmospheric Physics, Rostock University, Kühlungsborn, Germany, ³Instituto de Radioastronomía, Pontificia Universidad Católica del Perú, Lima, Perú, ⁴Radio Observatorio de Jicamarca, Instituto Geofísico del Perú, Lima, Perú, ⁵Estación Científica Ramón Mugica, Universidad de Piura, Piura, Perú

The mesosphere and lower thermosphere (MLT) are transitional regions between the lower and upper atmosphere. The MLT dynamics can be investigated using wind measurements conducted with meteor radars. Predicting MLT winds could help forecast ionospheric parameters, which has many implications for global communications and geo-location applications. Several literature sources have developed and compared predictive models for wind speed estimation. However, in recent years, hybrid models have been developed that significantly improve the accuracy of the estimates. These integrate time series decomposition and machine learning techniques to achieve more accurate short-term predictions. This research evaluates a hybrid model that is capable of making a short-term prediction of the horizontal winds between 80 and 95 km altitudes on the coast of Peru at two locations: Lima (12°S, 77°W) and Piura (5°S, 80°W). The model takes a window of 56 data points as input (corresponding to 7 days) and predicts 16 data points as output (corresponding to 2 days). First, the missing data problem was analyzed using the Expectation Maximization algorithm (EM). Then, variational mode decomposition (VMD) separates the components that dominate the winds. Each resulting component is processed separately in a Long short-term memory (LSTM) neural network whose hyperparameters were optimized using the Optuna tool. Then, the final prediction is the sum of the predicted components. The efficiency of the hybrid model is evaluated at different altitudes using the root mean square error (RMSE) and Spearman's correlation (r). The hybrid model performed better compared to two other models: the persistence model and the dominant harmonics model. The RMSE ranged from 10.79 to 27.04 ms^{-1} , and the correlation ranged from 0.55 to 0.94. In addition, it is observed that the prediction quality decreases as the prediction time increases. The RMSE at the first step reached 6.04 ms^{-1} with a correlation of 0.99, while at the sixteenth step, the RMSE increased up to 30.84 ms^{-1} with a correlation of 0.5.

KEYWORDS

MLT, EM, VMD, LSTM, OPTUNA

1 Introduction

The mesosphere and lower thermosphere (MLT) is the region of coupling between the lower and upper atmosphere. It is a region of complex chemical processes and dynamics (Liu et al., 2021). Understanding of this region is still in progress and it is of interest in atmospheric science and space traffic management.

The MLT dynamics is characterized by waves of different scales generated by other sources. For instance, on planetary scales, it is characterized by solar tides and planetary waves. The solar tides present periods of subharmonics of solar days and are generated mainly by the solar radiation absorption of tropospheric water vapor and stratospheric ozone (Forbes, 1995). On the other hand, the planetary waves have periods of days, e.g., the quasi-two-day waves with periods of 2 days generated *in situ* by baroclinic instabilities (McCormack et al., 2014). Moreover, the mesoscale gravity waves have periods of minutes to hours and can be generated by orographic sources and deep convection (Piani et al., 2000).

The MLT dynamics has usually been investigated using global circulation models (e.g., Liu et al., 2018), rockets (e.g., Staszak et al., 2021), satellites (e.g., Gasperini et al., 2023), lidars (e.g., Emmert et al., 2021), and radars (e.g., Chau et al., 2021). On the central and northern coast of Peru, two multi-static meteor radar networks, SIMONe Jicamarca (12°S, 77°W) in Lima and SIMONE Piura (5°S, 80°W) in Piura, allow us to measure winds between 75 and 105 km altitude since 2019 and 2021, respectively (Chau et al., 2021). Recently, diverse investigations have been conducted in the low-latitude Peruvian sector. For example, Suclupe et al. (2023) studied the climatology of large-scale dynamics, and Conte et al. (2024) studied the mesoscale dynamics.

The MLT region at low latitudes is significant for studying the effect of the lower atmospheric forces on the ionosphere (Immel et al., 2006; Vincent, 2015), a region with important implications for global communications and geo-location applications.

From another point of view, Yang et al. (2023) mention that near space (between 20 and 100 km altitude) is frequented by various aerospace vehicles. Like other research, they describe that there are complex dynamic processes, but emphasize that neutral atmospheric wind is a critical atmospheric parameter that influences the design and construction of aerospace vehicles. They argue that accurate wind prediction at these altitudes is essential for aerospace research. Similarly, Dhady et al. (2023) mention that the upper atmosphere (between 85 and 500 km altitude) has complex dynamics and that the behavior of the climate in this region directly impacts communication and navigation technologies that are important to humanity. They add that, due to our increasing dependence on these space technologies, predicting the dynamics in the upper atmosphere will become increasingly important.

In wind predictive models, several investigations have been carried out mainly at the tropospheric level (e.g., Hussin et al., 2021; Hanifi et al., 2022). These investigations usually use two types of models: statistical and machine learning models. Hussin et al. (2021) mention that there are statistical models such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive Integrated Moving Average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), and ARIMA-GARCH. However, statistical models

require the assumption of constant variance, and the original wind speed data do not meet this assumption. Hanifi et al. (2022) highlight that machine learning models are more appropriate methods for this data and that their success depends on an adequate selection of hyperparameters. They argue that the integrated use of Long Short-Term Memory (LSTM) neural networks and the hyperparameter optimizer Optuna accelerates the optimal choice of hyperparameters and gives more accurate estimates.

Recently, a time series decomposition technique called variational mode decomposition (VMD) has been introduced. This technique is applied before statistical modeling or modeling with machine learning and helps to deal with the problem of high variability. Ali et al. (2018) evaluate the application of this technique with two models for 1-step, 5-step, and 10-step forecasting horizons. In the first method, they use VMD with the ARIMA model. The second method uses VMD with artificial neural networks (ANN). These prediction methods are compared with other hybrid models such as empirical mode decomposition (EMD) with ARIMA, EMD with ANN, ensemble empirical mode decomposition (EEMD) with ARIMA, EEMD with ANN, complete ensemble empirical mode decomposition with Adaptive Noise (CEEMDAN) with ARIMA, and CEEMDAN with ANN. They conclude that the integration of VMD with ARIMA and VMD with ANN significantly outperforms existing hybrid models, for all prediction horizons.

In the prediction of wind speeds in the MLT region, Yang et al. (2023) also propose to use VMD before inputting the data into the prediction model. They used the hybrid VMD- PSO-LSTM model, which can decompose the wind time series for more accurate predictions. To do this, they used the VMD technique, which decomposes the original time series into principal components that dominate the signal. Then, each component is fed into the LSTM neural network, which finds the best hyperparameter values with the help of the particle swarm optimization (PSO) algorithm. This methodology is compared with the seasonal auto-regressive integrated moving average (SARIMA) statistical model and the hybrid models EMD-PSO-LSTM, EEMD-PSO-LSTM, CEEMDAN-PSO-LSTM and VMD-PSO-LSTM. Time horizons of 1 step, 3 steps, and 5 steps are predicted. They conclude that the proposed method has better efficiency and stability in wind speed prediction in all their comparisons.

This research studies the predictability of the mesospheric and lower thermospheric winds. In Peru, this prediction analysis has already been developed by Mauricio et al. (2023), whose model is based on the methodology used by Yang et al. (2023). The techniques used are missing data imputation, time series decomposition, and deep learning. These can identify the main components of the winds and then perform better predictions. Mauricio et al. (2023) modified the methodology of Yang et al. (2023). They added a missing data imputation analysis and used the Optuna hyperparameter optimizer in the modeling because it offers a better way to analyze the model quality.

The present research is a continuation of the work carried out by Mauricio et al. (2023). A new decomposition of the time series and a new hyperparameter search with Optuna have been performed. The main objective is to compare the efficiency of the hybrid model with a model of dominant harmonics based on the MLT climatology over Peru. The intention is to demonstrate that the

TABLE 1 Original data set.

	Lima	Piura
Description	Wind speeds in two components: zonal and meridional	
Heights	80.5, 85.5, 90.5 and 95.5 km	
Development of the model	September 2020 -November 2021 number of rows in the dataset: 21,888	October 2021 -December 2022 number of rows in the dataset: 21,888
Analysis 2023	Data of 2023 number of rows in the dataset: 17,519	

hybrid model gives more accurate predictions. This will be evaluated with the root mean square error (RMSE) and nonparametric Spearman correlation (r).

2 Materials and methods

2.1 Original data set

The zonal and meridional winds were estimated using the homogeneous method (e.g., [Chau et al., 2021](#)) recorded by SIMONe (Spread-spectrum Interferometric Multistatic meteor radar Observing Network) radars in Lima (12°S, 77°W) and Piura (5°S, 80°W). These systems have a horizontal coverage of approximately 400 km in diameter ([Chau et al., 2021](#)). The horizontal winds have a resolution of 1-h, and 2-km and were estimated every 30-min and 1-km (sampling) at heights of 80.5 km, 85.5 km, 90.5 km, and 95.5 km.

The available data for the Lima and Piura stations is described on [Table 1](#). The period is different because the stations started operating in different years. The data sets are divided into two parts. The first part is used for model development, covering the training, validation and testing stages. In the case of Lima, the period runs from September 2020 to November 2021. In the case of Piura, the period runs from October 2021 to December 2022. In both cases, those periods were chosen since they have the same amount of records so that the modeling would be similar, although they differ on the amount of missing data ([Figure 1](#)).

In the second part, the performance of the proposed model is evaluated in both stations using data from the year 2023. In this section, the number of records and the period coincide, but they differ on the amount of missing data ([Figure 2](#)).

2.2 Missing data imputation

Missing data imputation was performed using the Expectation-maximization (EM) algorithm, which applies to data following a Gaussian distribution. This algorithm is used to estimate the parameters of a probability distribution from incomplete data by iteratively maximizing the likelihood of the available data. In the context of multivariate Gaussian data, the probability distribution can be characterized by the vector of means and the variance-covariance matrix ([Schneider, 2001](#)).

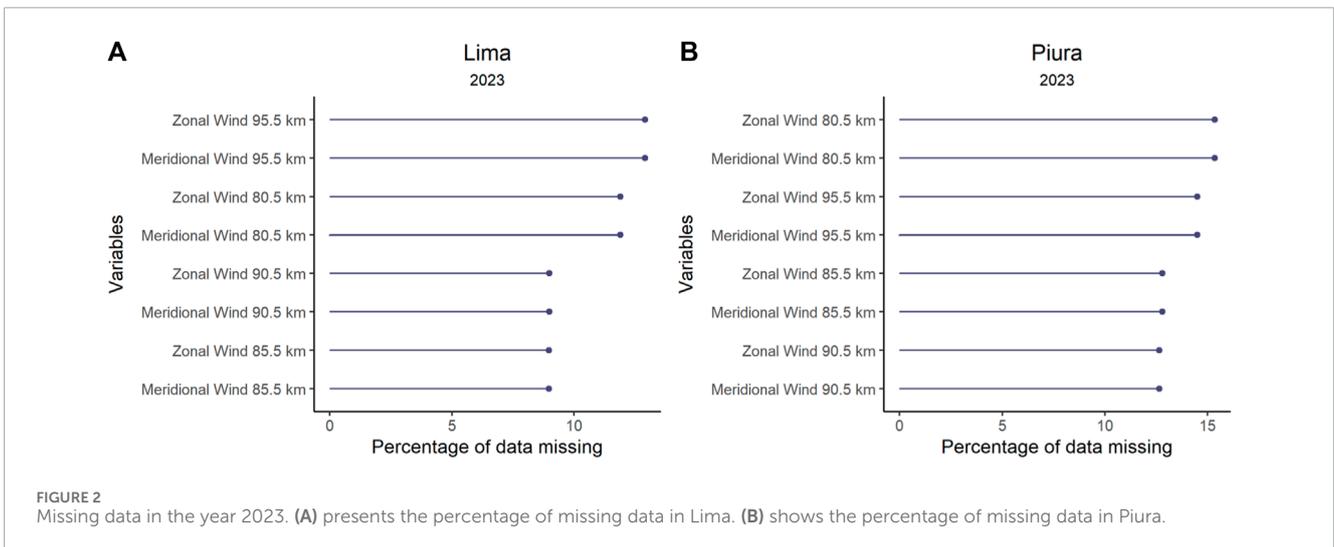
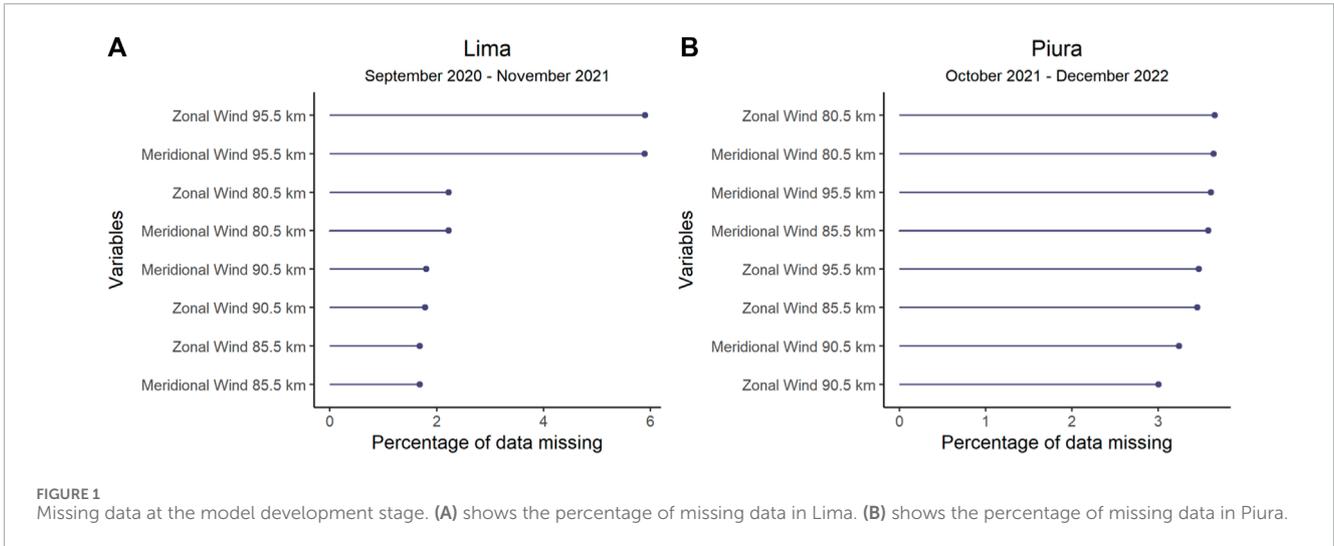
The EM algorithm has two stages: the E-step, which assumes the population mean and variance are known, and the M-step, which uses these values to estimate the population mean and covariance matrix. This iteration process continues until the parameter estimates of interest no longer change significantly. Maximum likelihood methods for incomplete multivariate data, particularly in the case of normally distributed data, focus on estimating the observed data parameters, such as the vector of means and the variance-covariance matrix. If the data follow a multivariate normal distribution, one can apply known properties to estimate those unknown parameters ([Pigott, 2001](#)).

The imputation analysis consisted of two parts. First, the imputation procedure described by [Mauricio et al. \(2023\)](#) was replicated. Additionally, the descriptive statistical indicators were presented to comprehend the process better. The second part involved exclusively imputing data from the year 2023, because these were obtained at a later point in time. [Figure 2](#) shows the percentages of missing data for the year 2023, for the Lima and Piura stations. The absence of data was mainly due to two reasons. First, during the first months of the year, the equipment was affected by the presence of heavy rains in the region, caused by cyclone Yaku and the El Niño Costero phenomenon. Secondly, during the year, there were some radar hardware failures. These unexpected situations forced the radars to be inoperative at certain intervals.

In case of high percentages of missing data, the time series is partitioned in such a way as to omit the time intervals with accumulated absences. The maximum percentage limit allowed in each subseries is 6%, which is an approximate value of that detected in the article of [Mauricio et al. \(2023\)](#). Then, for each subseries, the percentage is reanalyzed. If the percentage is less than 6%, we proceed with the estimation of the missing data. Otherwise, the time subseries are further separated until their percentages are lower than 6%.

2.3 Data preprocessing

Once the missing data are imputed, they are averaged every 3 h. Data not belonging to the year 2023 are divided chronologically into train (70%), validation (10%) and test (20%). Training and validation data are used for the building and optimization of the model, while the test data is used specifically for the metric evaluation. The intention is that the test data is not involved in the modeling and optimization, to avoid overfitting ([Mauricio et al.,](#)



2023). Finally, each data block is normalized based on the training data, before moving on to the time series decomposition stage.

2.4 Time serie decomposition

Variational mode decomposition (VMD) is a method used in signal processing. This method was proposed in 2014 to overcome the limitations of techniques such as wavelet analysis and Empirical Mode Decomposition (EMD). The VMD technique decomposes a sequence, such as wind time series, into multiple subsequences known as intrinsic modal function (IMF) components. The importance of VMD is its ability to optimally adapt the center frequency and bandwidth of each IMF according to the signal characteristics, which makes it effective in dealing with non-smoothness in data series, such as wind speed data (Yang et al., 2023). The VMD algorithm has as input a signal or time series $x(t)$, which is decomposed into subseries called modes or harmonic

signals $u_k(t)$, where k is the number of total modes.

$$x(t) = \sum_{k=1}^K u_k(t)$$

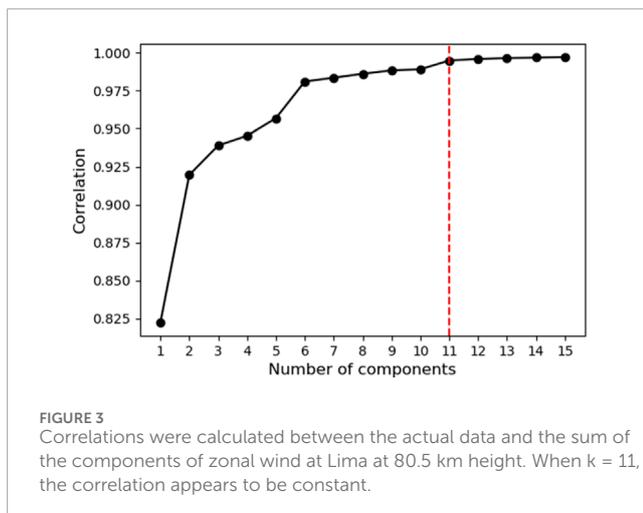
$$u_k = A_k(t) \cos(\phi_k(t))$$

Each mode has an instantaneous amplitude $A_k(t)$, an instantaneous phase $\phi_k(t)$, a limited spectral bandwidth $B(u_k(t))$ and an instantaneous center frequency ω_k (Ali, Khan, and Rehman, 2018), where it is assumed that each mode varies slower than the phase and is non-negative (Gan et al., 2021). The complete mathematical process that follows this decomposition and bandwidth estimation can be found in the article of Dragomiretskiy and Zosso (2014).

The VMD algorithm package in Python was obtained from the code developed by Carvalho et al. (2020). The choice of components is determined under two complementary rules. The first rule follows the methodology of Yang et al. (2023) which involves observing center frequencies. It starts by decomposing the original time series

TABLE 2 Evaluation of the optimal number of components, with center frequencies.

K	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7	IMF 8	IMF 9	IMF 10	IMF 11	IMF 12
1	0.0027											
2	0.0031	0.1944										
3	0.0030	0.1271	0.2665									
4	0.0030	0.1247	0.2430	0.3674								
5	0.0030	0.1231	0.1721	0.2520	0.3707							
6	0.0026	0.0596	0.1261	0.2079	0.2747	0.3748						
7	0.0025	0.0585	0.1250	0.1758	0.2491	0.3620	0.4436					
8	0.0025	0.0582	0.1247	0.1727	0.2455	0.3058	0.3718	0.4506				
9	0.0025	0.0570	0.1236	0.1501	0.1943	0.2496	0.3101	0.3732	0.4526			
10	0.0025	0.0560	0.1230	0.1419	0.1853	0.2467	0.3023	0.3636	0.4061	0.4621		
11	0.0016	0.0261	0.0690	0.1244	0.1533	0.1970	0.2499	0.3069	0.3669	0.4093	0.4634	
12	0.0016	0.0242	0.0659	0.1238	0.1440	0.1780	0.2185	0.2588	0.3123	0.3685	0.4108	0.4641



with different values of k , where each component has an associated central frequency. The optimal number of components is then obtained when the central frequency values are approximate.

Table 2 shows the decomposition of the normalized time series of the zonal wind at Lima at 80.5 km height, and the central frequencies are obtained for different values of k . It is observed that when the series are decomposed in $k = 10$, $k = 11$ and $k = 12$, the last central frequencies are approximately 0.4621, 0.4634, and 0.4641, respectively. Then, it can be deduced that the time series can be optimally decomposed into 12 components.

A second complementary way of determining the optimal number of components is proposed by evaluating the correlation between the initial series data and the sum of the IMF components. Being an additive decomposition, the sum of the components results

in the estimate of the initial data. Figure 3 shows the correlation between the normalized data of the zonal wind at Lima at 80.5 km height and the sum of its IMF components, for different values of k . It is observed that when the series is decomposed from 11 components to more, the correlation values are approximate and almost constant. In this instance, both rules give similar results. Then it is determined that the number of components for this series is 12. Figure 4 shows the 12 components.

2.5 Long short-term memory neural network

Long short-term memory neural network (LSTM) is an enhancement of the recurrent neural network (RNN), which has limitations such as gradient bursting and fading, lack of retention of historical information over time; and not distinguish between information that should be further processed and information that should be deleted. The LSTM network uses control gates that help solve the above problems. Within the LSTM block, there is a ring buffer and three gates named: input, forget, and output. Like RNNs, the LSTM network also has a hidden layer that processes the flow of information (Son and Jung, 2020).

Figure 5 shows the basic structure of the LSTM network. Its main features are the hidden layer h_t , the memory cell c_t and the control gates (represented by the letters inside the circles). The first gate is the forgetting gate (letter f), which evaluates the elements to be purged from the cell state c_{t-1} and outputs a resultant vector f_t . Then the cell state \tilde{c} is partially updated, processing the input x_t and the previous hidden state h_{t-1} . The second gate is the input gate (letter i) which has a resultant vector i_t . This gate has the task of evaluating which information from the partial cell state \tilde{c}_t serves to fully update the

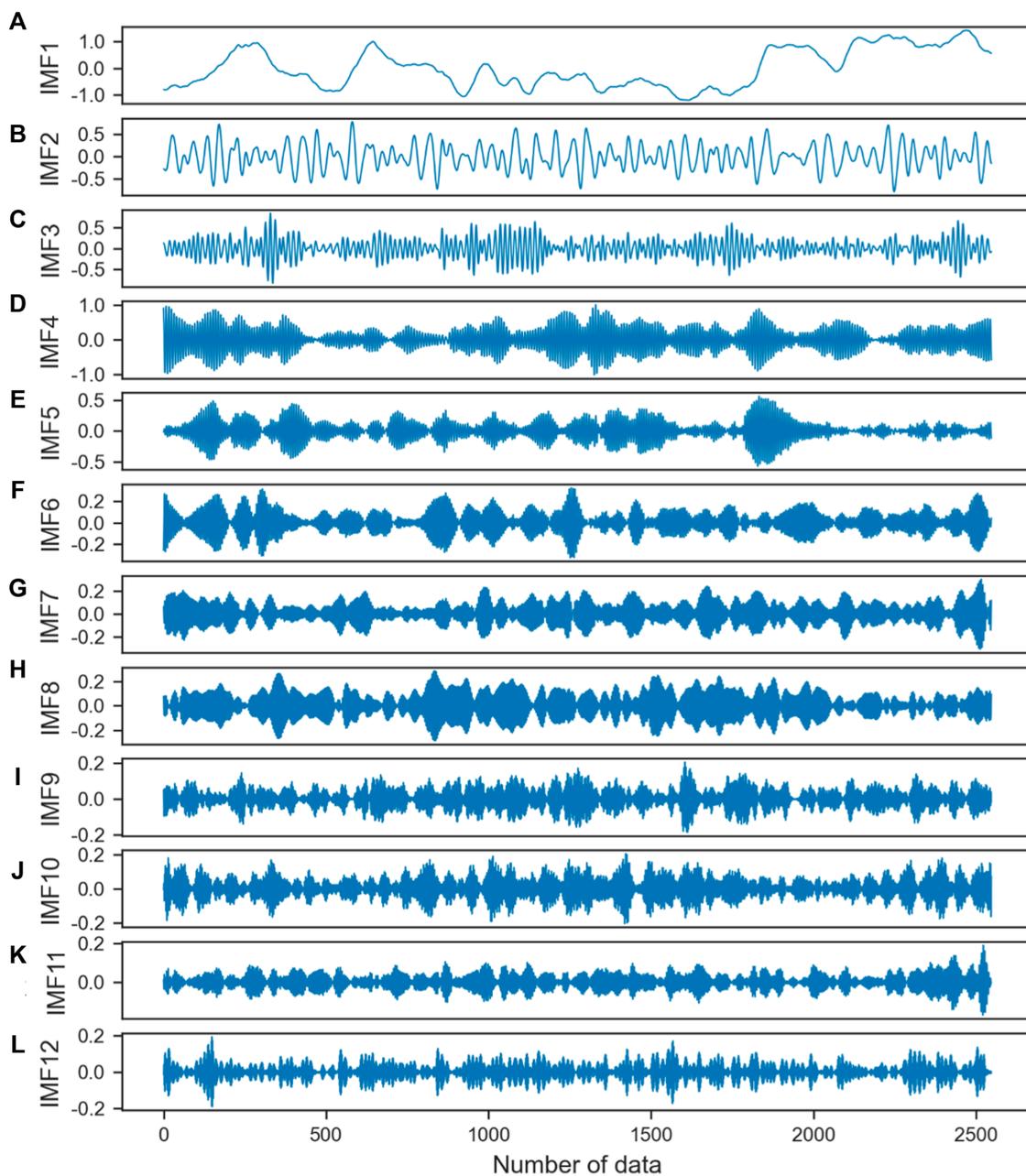


FIGURE 4
Components of the zonal wind at Lima at 80.5 km altitude. (A–L) show the 12 time series into which the original time series was decomposed.

current time cell state c_t . Finally, there is the output gate (letter o) that controls the sending of the information from the current cell state c_t to the new hidden state h_t (Kratzert et al., 2018).

LSTM neural networks can be implemented using the function of the same name from the TensorFlow library in Python. Like the methodology proposed by Yang et al. (2023) a model with LSTM was performed for each IMF of each time series. For example, if a time series is determined to have 12 modal components, then 12 models with LSTM should be performed.

Before inputting the data to the neural network, the dimensions of the input and output data must be specified. In this case,

we have experimented with a window of 56 consecutive data, corresponding to a period of 1 week. While the output data are 16 steps in the future, corresponding to 2-day records. Subsequently, hyperparameters such as dropout, number of layers and neurons, Adam optimizer, learning rate, and batch size were added (Mauricio et al., 2023).

Additionally, the Ridge regularization was introduced, which is a penalty applied to the loss function of the proposed model. This regularization was also used in the paper by Rosa et al. (2020), who performed a similar analysis with VMD and LSTM for the Piura River flow time series in Peru.

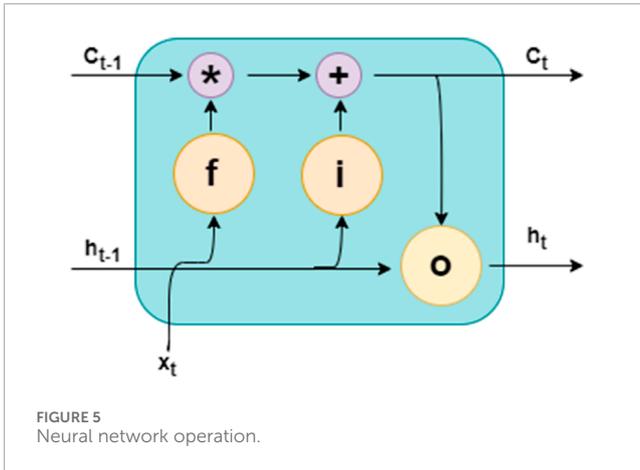


FIGURE 5
Neural network operation.

2.6 Optuna

Akiba et al. (2019) presented Optuna, a framework designed for hyperparameter optimization in the context of machine learning and deep learning. Optuna is based on a sequential optimization approach that uses the adaptive search tree-based optimization (TPE) algorithm to efficiently explore the hyperparameter space. Optuna is characterized by its ability to dynamically adapt to the hyperparameter space, allowing it to converge more quickly to optimal solutions. This is achieved by systematically exploring hyperparameter combinations, where decisions are based on previous observations to steer the search towards promising regions of the search space. The framework offers a simple user interface and seamless integration with machine learning and deep learning libraries, facilitating its use in a variety of applications. In addition, Optuna provides advanced tools such as experiment management and result visualization to facilitate the organization and analysis of optimization results. Optuna's scalability stands out, as it is designed to handle large datasets and complex search spaces efficiently. This makes it suitable for both small-scale applications and large-scale research projects requiring complex model optimization. In summary, Optuna represents a significant breakthrough in automated hyperparameter optimization, offering a powerful and efficient solution for improving model performance. Its ability to dynamically adapt to the hyperparameter space and its seamless integration with popular libraries make it a valuable tool for researchers and practitioners in the field of machine learning.

Table 3 shows the set of hyperparameters used in this study. During optimization, Optuna provides tools that help to plot the loss curve that allows to evaluation of the learning performance of the models. Figure 6 shows the loss curves for the 12 components of the Piura zonal wind at 80.5 km altitude. Optuna has by default that the X-axis represents the number of epochs and the Y-axis represents the mean square error (MSE). However, since the data are normalized, the MSE values have no units of measurement.

2.7 Hybrid model

In summary of what has been explained in the previous paragraphs, the hybrid model applied in combines the VMD, LSTM

TABLE 3 Hyperparameter options.

Hyperparameters	
Layers	1
Neurons in layer	Values between 1 and 100
Dropout	Values between 0.0 and 0.5
Learning rate	Values between 0.00001 and 0.1
Epoch	100,120,140,160,180 years 200
Batch size	16, 24, 32, 40, 48, 56, 64
Lambda (regularizer L2)	Values between 0.001 and 0.01

and Optuna techniques (Figure 7). First, time series of normalized and complete data are obtained, which are then decomposed into their optimal components using the VMD algorithm. These components are reorganized into windows of 56 data, equivalent to 7 days. Each of these blocks is fed into an individual LSTM neural network, where the hyperparameters are optimized with Optuna. The output of each LSTM provides windows of 16 steps, corresponding to 2 days. Finally, the final prediction is the denormalization of the sum of the estimates calculated for each component.

2.8 Persistence model

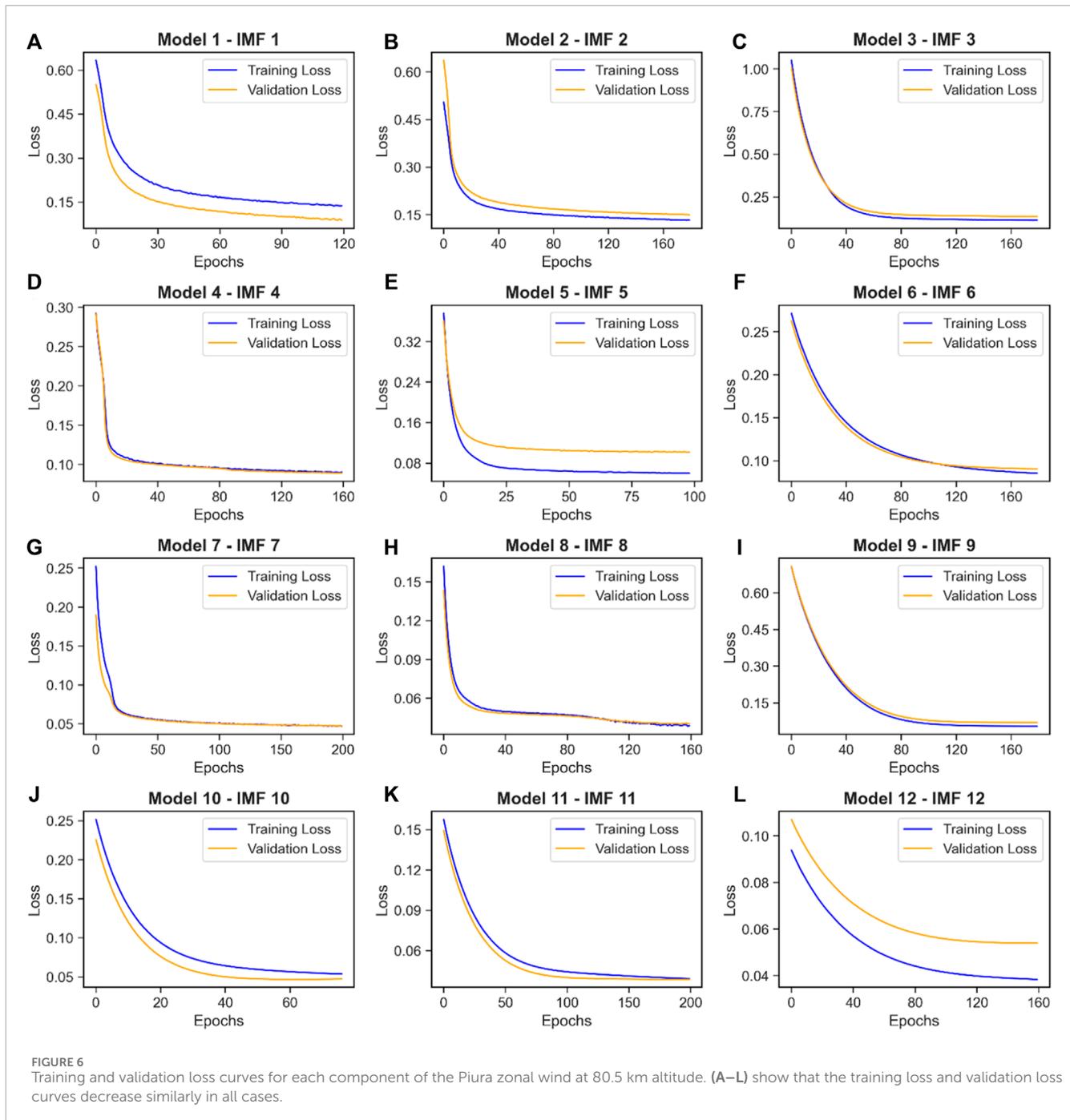
Mauricio et al. (2023) used a persistence model that consists of the wind speed values of two previous days being repeated in the following 2 days. Through this model, a time series is reconstructed and compared with the hybrid model estimates. The intention is to demonstrate that the hybrid model is superior to the simplicity of the persistence model.

2.9 Model of dominant harmonics

The model of dominant harmonics was built from a sum of sinusoidal series with specified periods. In this model, the mean winds (an average of meridional wind and an average of zonal wind), as well as the amplitudes and phases of specific waves, were fitted using the least squares method, with a 7-day window. The selected periods correspond to the dominant wave periods obtained by the MLT climatology over the Peruvian sector (see Suclupe et al., 2023 for more details), which are 48, 24, 12, 8, and 6 h. Finally, the model was interpolated in a window of 2 days to compare it with the proposed hybrid model.

2.10 Evaluation of the hybrid model

Mauricio et al. (2023) reconstructed the time series by making predictions with a 16-step horizon. This approach involved the generation of sequential predictions of consecutive blocks of



16 steps within the time series. Specifically, predictions were initially applied to steps 1 through 16. Then, the process was repeated for steps 17 through 32, and so on, ensuring progressive coverage of the entire time series in 16-step intervals. This method allowed the evaluation of the predictive capability of the proposed model versus a 2-day persistence model, for multiple segments (training, validation, testing and February 2023 data), ensuring a complete evaluation of its performance over time. The metric used was RMSE and the proposed model was shown to be better.

In this paper, the RMSE was again used and the nonparametric Spearman correlation metric (r) was added. This coefficient is used to measure the strength and direction of the association between two variables, regardless of the distribution of the data. Assuming that the results show a mostly positive correlation, a correlation is considered to be very weak if its value is between 0.00 and 0.20, weak between 0.21 and 0.40, moderate between 0.41 and 0.60, strong between 0.61 and 0.80, and very strong between 0.81 and 1.00.

It was decided to evaluate the hybrid model in three stages. The first stage is related to evaluating that the estimates of the

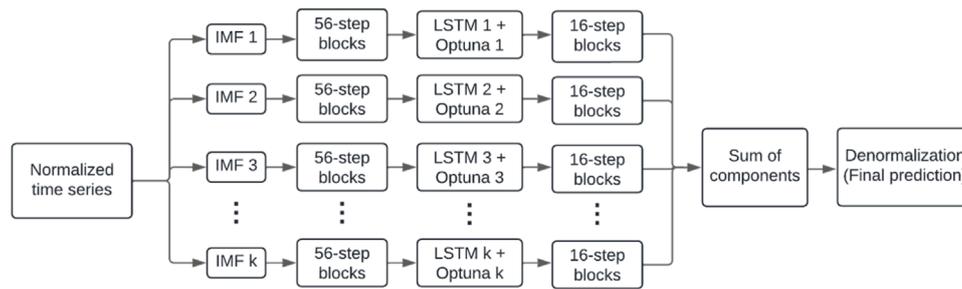


FIGURE 7
Diagram of the hybrid model.

proposed model do not present overfitting and that it is superior to a simple model such as the persistence model. Similar to the analysis performed by [Mauricio et al. \(2023\)](#), the entire time series is estimated using 16-step predictions. That is, the 16-step blocks are joined and form a complete time series. It is confirmed that there is no overfitting when the calculated metrics of the hybrid model are similar in the training, validation and test data sets. On the other hand, it is confirmed that the hybrid model is better than the persistence model when the metrics of the hybrid model are better than those of the persistence model, using the same data sets.

The second stage was the step-by-step analysis. This method involves generating multiple predictions using continuous input windows. Predictions are made using the first 56 input steps, then the input window is shifted one position and new predictions are made using the next 56 steps, and so on. This process is repeated until the entire time series is covered. This approach facilitates obtaining vectors of estimates of all 1-steps, all 2-steps and so on up to the vector of 16-steps. This analysis was only performed on the test data, evaluating the predictive model performance at each time step.

In the third stage, the original data for the year 2023 (January to December) are available. The objective is to evaluate whether the hybrid model is better than the model of dominant harmonic for the year 2023. As in the first stage, a time series is reconstructed through the estimated consecutive blocks of 16 steps. It is evaluated which model is better, observing which one has better metrics.

3 Results

3.1 Data imputation

[Table 4](#) shows the comparison of the descriptive statistics values before and after imputation for data previous to 2023. This was mentioned by [Mauricio et al. \(2023\)](#), but the values were not shown there.

In the second stage, the 2023 data were analyzed and a summary of the information on the amount of missing data was obtained. The result shows that the percentages were in the range between 8% and 15% ([Table 5](#)). As these percentages of missing data are higher than 6%, the time series was subdivided into blocks with lower percentages and the imputation process was replicated.

3.2 VMD decomposition

The VMD decomposition analysis to determine the number of components of each time series is shown in [Table 6](#). The values show convergence between 11 and 12 components. Above these values, the method is not optimal.

3.3 Evaluation of the hybrid model

3.3.1 Comparison with persistence model

The metrics in [Tables 7, 8](#) are analyzed with the training, validation and test data sets for the Lima and Piura locations, respectively. The metrics show that the hybrid model is better than the persistence model in both locations.

In the case of Lima, the hybrid model has RMSE values that vary between 12.38 m s^{-1} and 22.5 m s^{-1} , while the correlation values vary between 0.81 and 0.94. In comparison, the persistence model has higher RMSE with values that vary between 26.36 m s^{-1} and 47.52 m s^{-1} , while the correlation values are lower and vary between 0.24 and 0.64.

In the case of Piura, the RMSE values vary between 10.79 m s^{-1} and 27.04 m s^{-1} , while the correlation values vary between 0.67 and 0.93. In comparison, the persistence model has higher RMSE with values that vary between 29.01 m s^{-1} and 43.78 m s^{-1} , while the correlation values are lower than 0.48.

In addition, the values of the hybrid model metrics in the three data sets (training, validation and test), do not have large differences. This is an indicator that the model does not have overfitting.

3.3.2 Step analysis

As noted in the methodology, this analysis was only performed on the test data set. The metrics calculated for each time step show an increase in RMSE ([Figure 8](#)) and a decrease in correlation ([Figure 9](#)). This indicates that within the windows of 16 prediction steps, the first steps fit better than the last steps.

In the case of Lima, the RMSE values increase from 6.46 m s^{-1} to 30.84 m s^{-1} , while the correlation values decrease from 0.99 to 0.66. For example, the meridional wind at 85.5 km of altitude, has the best 1-step with an RMSE of 6.46 m s^{-1} and a

TABLE 4 Comparison of descriptive statistics before and after imputation.

Height		Before data imputation				After data imputation			
		80.5 km	85.5 km	90.5 km	95.5 km	80.5 km	85.5 km	90.5 km	95.5 km
Zonal wind over Lima	Mean	1.44	-0.48	-5.10	-7.47	1.45	-0.48	-5.00	-8.30
	Standard deviation	31.09	33.36	37.98	40.62	31.11	33.35	38.00	40.85
	Min	-109.97	-146.21	-134.82	-149.52	-109.97	-146.21	-134.82	-149.52
	Q1	-21.00	-23.85	-31.07	-34.92	-20.97	-23.85	-31.00	-36.02
	Median	2.33	-0.10	-4.16	-7.37	2.28	-0.12	-4.11	-8.06
	Q3	25.12	23.03	21.44	19.79	25.09	22.99	21.56	19.15
	Max	87.58	107.67	145.81	145.40	102.64	123.87	145.81	145.40
	Skewness	-0.16	-0.07	-0.09	0.01	-0.15	-0.07	-0.09	0.01
	Kurtosis	-0.48	-0.20	-0.26	-0.01	-0.47	-0.20	-0.26	-0.01
Meridional wind over Lima	Mean	-2.97	-4.22	-4.46	-1.73	-2.99	-4.22	-4.43	-1.31
	Standard deviation	31.65	39.33	45.22	49.21	31.67	39.37	45.27	49.16
	Min	-105.98	-124.41	-129.33	-139.68	-113.33	-166.54	-150.39	-174.41
	Q1	-25.72	-33.56	-37.46	-38.43	-25.72	-33.56	-37.41	-37.62
	Median	-3.27	-5.43	-4.03	-1.13	-3.29	-5.40	-3.90	-0.58
	Q3	19.45	24.00	28.17	33.67	19.40	23.99	28.18	33.93
	Max	112.53	120.89	139.03	149.94	112.53	120.89	139.03	195.99
	Skewness	0.05	0.16	0.00	0.06	0.05	0.15	0.00	0.06
	Kurtosis	-0.36	-0.39	-0.50	-0.52	-0.35	-0.37	-0.49	-0.48
Zonal wind over Piura	Mean	-2.99	-6.22	-9.32	-11.83	-2.97	-6.20	-9.25	-11.73
	Standard deviation	27.84	30.44	34.19	38.77	27.96	30.52	34.20	38.79
	Min	-130.96	-132.43	-145.58	-146.29	-130.96	-145.27	-145.58	-146.29
	Q1	-21.04	-26.76	-32.01	-38.80	-21.08	-26.76	-31.97	-38.66
	Median	-2.33	-5.47	-8.31	-10.60	-2.30	-5.46	-8.27	-10.58
	Q3	15.41	14.20	13.85	15.75	15.48	14.28	13.91	15.82
	Max	147.98	143.47	138.13	137.72	147.98	143.47	138.13	137.72
	Skewness	-0.04	-0.07	-0.11	-0.10	-0.05	-0.07	-0.11	-0.09
	Kurtosis	0.38	-0.06	-0.10	-0.19	0.38	-0.03	-0.10	-0.19

(Continued on the following page)

TABLE 4 (Continued) Comparison of descriptive statistics before and after imputation.

Height		Before data imputation				After data imputation			
		80.5 km	85.5 km	90.5 km	95.5 km	80.5 km	85.5 km	90.5 km	95.5 km
Meridional wind over Piura	Mean	-2.69	-2.77	-4.53	-5.25	-2.60	-2.71	-4.59	-5.31
	Standard deviation	31.36	34.00	36.41	41.20	31.41	34.02	36.47	41.28
	Min	-149.54	-142.49	-144.62	-148.40	-149.54	-142.49	-149.39	-148.40
	Q1	-23.94	-26.94	-30.37	-34.36	-23.90	-26.83	-30.41	-34.45
	Median	-3.45	-4.05	-5.08	-5.44	-3.30	-3.93	-5.21	-5.56
	Q3	17.76	20.32	20.21	22.58	17.89	20.41	20.20	22.55
	Max	149.34	123.16	135.49	148.12	149.34	123.16	135.49	148.12
	Skewness	0.14	0.14	0.13	0.08	0.13	0.13	0.12	0.08
	Kurtosis	0.12	-0.16	-0.16	-0.20	0.12	-0.15	-0.14	-0.19

TABLE 5 Missing data in 2023.

Height(km)	Lima				Piura			
	Zonal Wind		Meridional Wind		Zonal Wind		Meridional Wind	
	Amount	Percentage	Amount	Percentage	Amount	Percentage	Amount	Percentage
80.5	2084	11.90	2084	11.90	2,689	15.30	2,689	15.30
85.5	1,572	8.97	1,572	8.97	2,241	12.80	2,241	12.80
90.5	1,574	8.98	1,574	8.98	2,216	12.60	2,216	12.60
95.5	2,262	12.90	2,262	12.90	2,539	14.50	2,539	14.50

TABLE 6 Number of IMF for each time series.

Height(km)	Lima		Piura	
	Zonal Wind	Meridional Wind	Zonal Wind	Meridional Wind
80.5	12	12	12	11
85.5	11	12	12	12
90.5	12	11	12	12
95.5	12	12	12	12

correlation of 0.99, while the 16-step has an RMSE of 23.05 and a correlation of 0.8.

In the case of Piura, the RMSE values increase from 6.04 m s⁻¹ to 30.41 m s⁻¹, while the correlation values decrease from 0.99 to 0.5. For example, the zonal wind at 85.5 km of altitude, has the best 1-step with an RMSE of 6.04 m s⁻¹ and a correlation of 0.98, while the 16-step has an RMSE of 17.42 m s⁻¹ and a correlation of 0.79.

3.3.3 Comparison with the model of dominant harmonics

The comparison of metrics between the hybrid model and the model of dominant harmonics with data from 2023 is shown in Table 9 for Lima and Table 10 for Piura.

In the case of Lima, the hybrid model has RMSE values ranging from 14.92 m s⁻¹ to 26.95 m s⁻¹, while the correlation values

TABLE 7 Evaluation of the hybrid model for Lima data.

		Zonal Wind				Meridional Wind			
		Real vs Hybrid model		Real vs Persistence		Real vs Hybrid model		Real vs Persistence	
Height	Data set	RMSE ($m s^{-1}$)	r	RMSE ($m s^{-1}$)	r	RMSE ($m s^{-1}$)	r	RMSE ($m s^{-1}$)	r
80.5 km	Train	12.38	0.91	26.36	0.61	13.30	0.90	29.90	0.49
	Validation	13.42	0.82	28.39	0.25	14.39	0.84	30.32	0.38
	Test	16.84	0.83	35.69	0.30	14.37	0.88	30.51	0.51
85.5 km	Train	20.92	0.92	32.16	0.44	13.60	0.94	34.52	0.54
	Validation	22.06	0.92	37.13	0.33	18.58	0.88	38.81	0.46
	Test	21.64	0.92	40.99	0.24	15.92	0.91	33.41	0.64
90.5 km	Train	14.22	0.91	36.95	0.40	19.07	0.91	38.76	0.58
	Validation	15.40	0.91	43.38	0.27	21.87	0.88	43.75	0.52
	Test	14.73	0.93	42.99	0.39	21.63	0.85	39.76	0.54
95.5 km	Train	19.31	0.84	41.09	0.34	19.33	0.91	42.06	0.56
	Validation	20.84	0.81	44.30	0.24	22.37	0.87	45.09	0.49
	Test	22.50	0.84	47.52	0.34	19.07	0.91	43.68	0.53

TABLE 8 Evaluation of the hybrid model for Piura data.

		Zonal Wind				Meridional Wind			
		Real vs Hybrid model		Real vs Persistence		Real vs Hybrid model		Real vs Persistence	
Height	Data set	RMSE ($m s^{-1}$)	r	RMSE ($m s^{-1}$)	r	RMSE ($m s^{-1}$)	r	RMSE ($m s^{-1}$)	r
80.5 km	Train	11.93	0.89	29.01	0.38	14.59	0.85	30.18	0.41
	Validation	11.92	0.86	33.71	-0.03	20.09	0.77	36.55	0.34
	Test	11.98	0.88	30.40	0.30	18.48	0.78	33.63	0.40
85.5 km	Train	10.79	0.93	32.49	0.36	15.13	0.87	32.30	0.44
	Validation	12.87	0.92	38.92	0.08	16.88	0.87	35.66	0.43
	Test	12.01	0.90	35.63	0.21	18.11	0.83	35.16	0.41
90.5 km	Train	13.90	0.90	36.38	0.34	18.07	0.84	35.27	0.44
	Validation	15.63	0.91	39.95	0.39	27.04	0.67	38.45	0.37
	Test	15.50	0.87	39.18	0.22	21.33	0.77	38.82	0.30
95.5 km	Train	17.91	0.87	40.58	0.34	16.97	0.90	40.31	0.43
	Validation	22.80	0.84	43.09	0.48	17.95	0.88	42.32	0.38
	Test	17.03	0.89	42.57	0.26	20.29	0.80	43.78	0.22

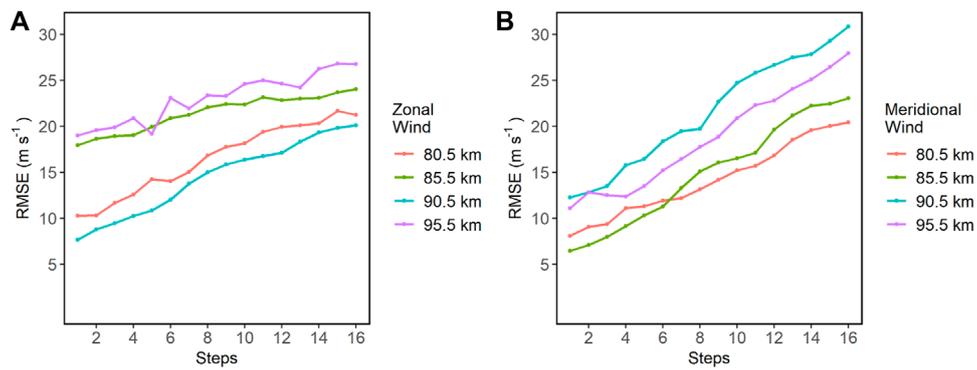


FIGURE 8 RMSE values for each time step in Lima. (A) shows the RMSE values for the zonal wind. (B) shows the RMSE values for the meridional wind.

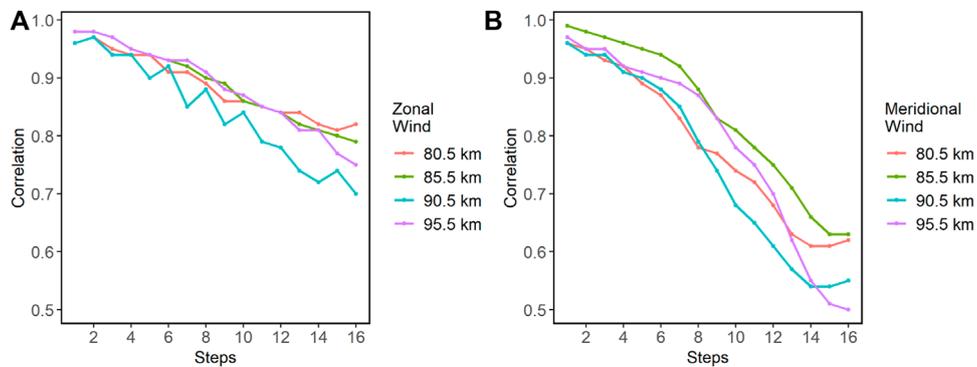


FIGURE 9 Correlation values for each time step in Piura. (A) shows the correlation values for the zonal wind. (B) shows the correlation values for the meridional wind.

TABLE 9 Comparison metrics between the hybrid model and interpolation in 2023 (Lima).

Height	Zonal Wind				Meridional Wind			
	Real vs Hybrid model		Real vs Dominant harmonics model		Real vs Hybrid model		Real vs Dominant harmonics model	
	RMSE (m s^{-1})	r	RMSE (m s^{-1})	r	RMSE (m s^{-1})	r	RMSE (m s^{-1})	r
80.5	15.53	0.83	23.68	0.63	14.92	0.80	26.23	0.44
85.5	21.85	0.69	27.01	0.51	15.28	0.86	31.24	0.51
90.5	23.40	0.71	31.55	0.41	23.21	0.81	37.76	0.51
95.5	26.95	0.70	36.03	0.43	23.99	0.84	41.21	0.58

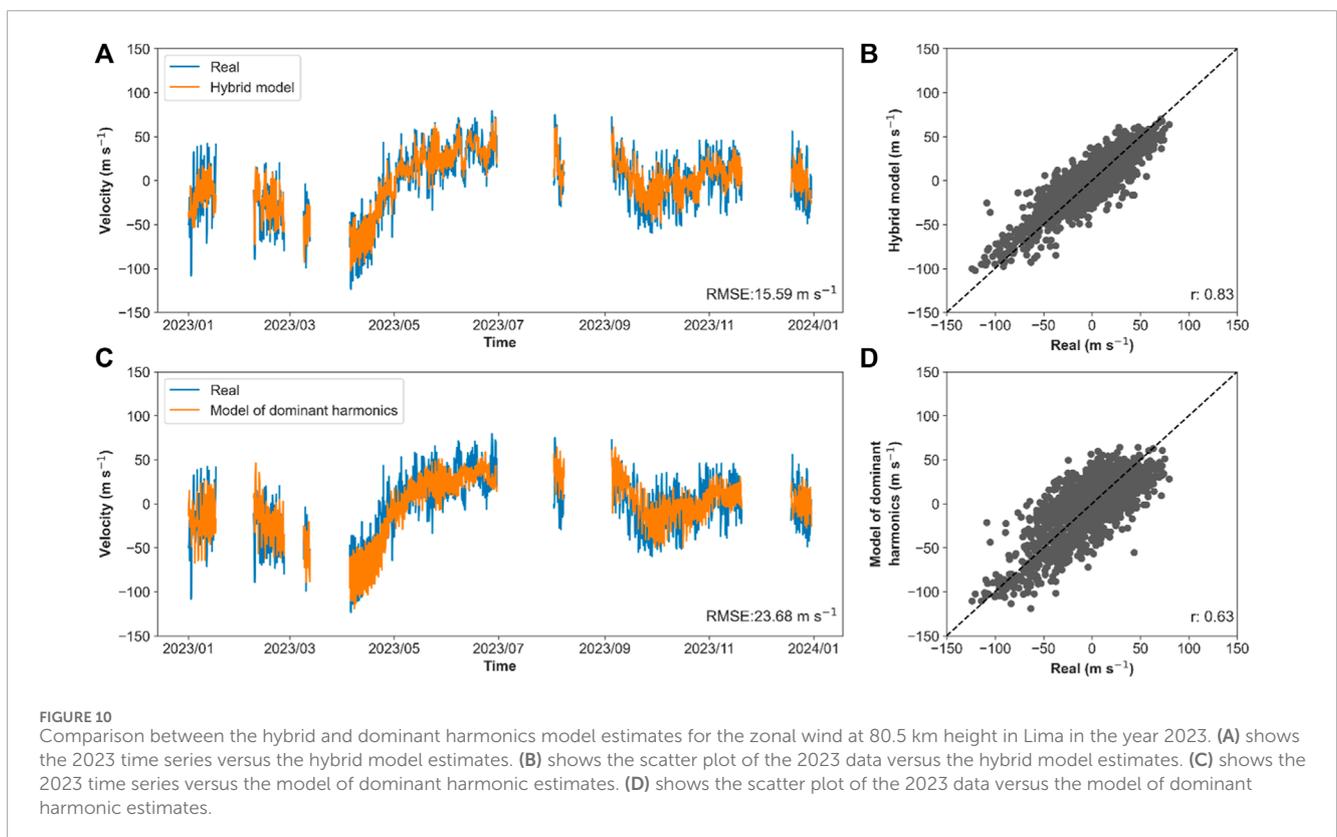
range from 0.69 to 0.86. The model of dominant harmonics has RMSE values ranging between 23.68 m s^{-1} and 41.21 m s^{-1} , while correlation values vary between 0.41 and 0.63. Figure 10 shows the time series predictions of the zonal wind in Lima at 80.5 km altitude, made by both models, with their respective scatter diagrams. The hybrid model has an RMSE of 15.53 m s^{-1} and a correlation of 0.83, while the model of dominant harmonics has an RMSE of 23.68 m s^{-1}

and a correlation of 0.63. It is observed that the hybrid model has a lower RMSE value and a higher correlation value. Additionally, in the scatter plot, the predictions of the hybrid model are better fitted to the data in Lima.

In the case of Piura, the hybrid model has RMSE values ranging from 15.47 m s^{-1} to 26.80 m s^{-1} , while the correlation values range from 0.55 to 0.80. The model of dominant

TABLE 10 Comparison metrics between the hybrid model and interpolation in 2023 (Piura).

Height	Zonal Wind				Meridional Wind			
	Real vs Hybrid model		Real vs Dominant harmonics model		Real vs Hybrid model		Real vs Dominant harmonics model	
	RMSE (m s^{-1})	r	RMSE (m s^{-1})	r	RMSE (m s^{-1})	r	RMSE (m s^{-1})	r
80.5	15.47	0.75	21.87	0.50	16.31	0.68	25.39	0.34
85.5	15.80	0.75	24.99	0.43	16.78	0.73	29.97	0.26
90.5	26.80	0.55	28.90	0.34	21.40	0.74	37.50	0.28
95.5	25.20	0.62	32.89	0.35	19.34	0.80	41.57	0.40

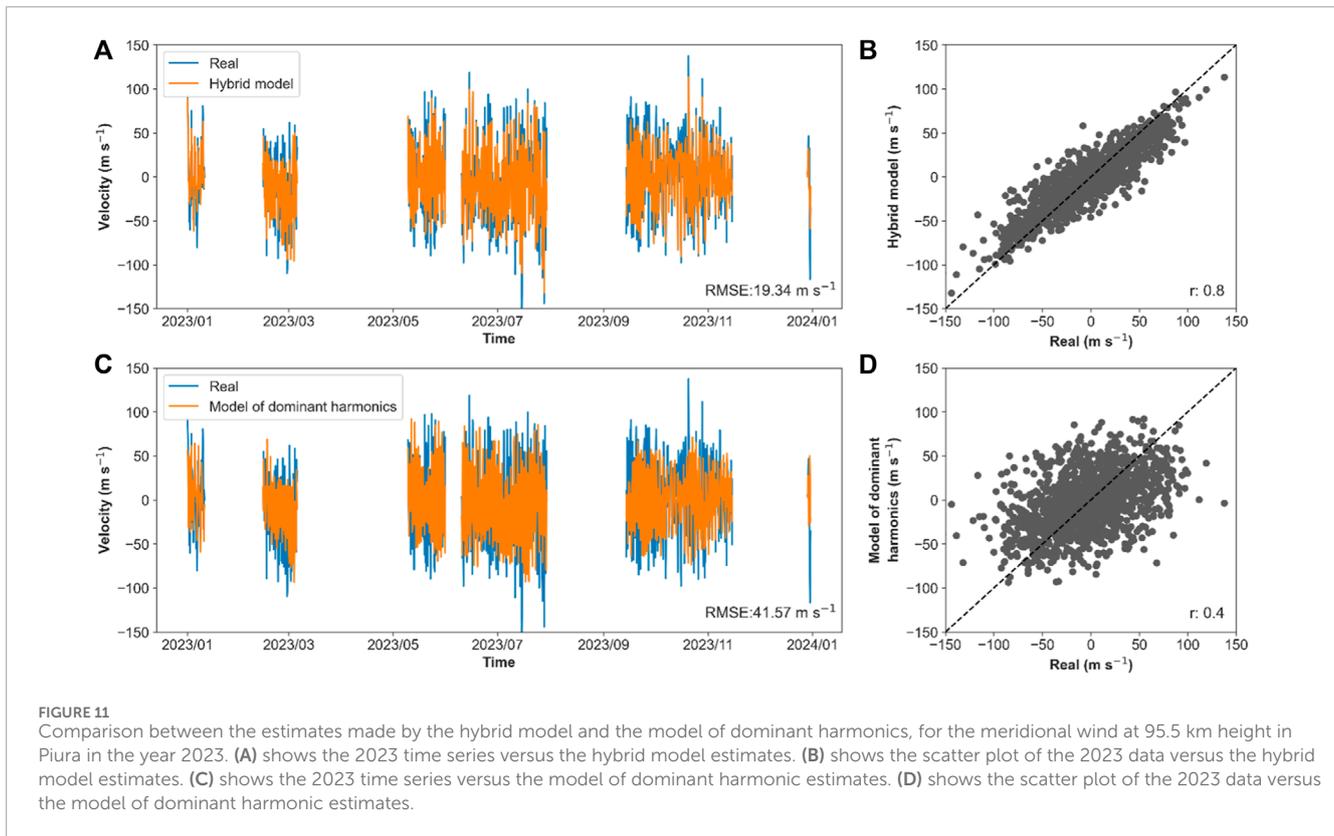


harmonics has RMSE values ranging between 21.87 m s^{-1} and 41.57 m s^{-1} , while correlation values vary between 0.26 and 0.5. Figure 11 shows the time series predictions of the meridional wind in Piura at 95.5 km altitude, made by both models, with their respective scatter diagrams. The hybrid model has an RMSE of 19.34 m s^{-1} and a correlation of 0.80, while the model of dominant harmonics has an RMSE of 41.57 m s^{-1} and a correlation of 0.40. Similar to the results for Lima, the hybrid model is found to have a lower RMSE, a higher correlation, and less dispersion with respect to the model of dominant harmonics. Also, their scatter plot indicates that the predictions of the hybrid model better fit the 2023 data collected in Piura.

4 Discussion and conclusions

The imputation of missing data for 2023 was performed under the assumption that the percentage limit of missing data is 6%. However, it is necessary to have sufficient data to allow this analysis to be carried out. If there are only a few data points available, the distribution may not be symmetrical because of certain values that dominate that period. In addition, it is essential to have at least 72 continuous data to be able to make predictions and calculate metrics (56 input data points and 16 output data points).

In determining the number of components of the time series, convergence is observed between the values of 11 and 12 components. Exceeding these values can lead to poor estimates



of the original series. Having these limits defined helps to avoid making unnecessary models of the components, avoiding execution times and the use of computational resources. Additionally, for future research, associations between these components and other meteorological time measurements could be evaluated.

During the modeling phase, the inclusion of the regularizer (L2) resulted in a marked improvement in the visualization of the loss curves for both the training and validation sets, which helped to avoid overfitting the models. As for Optuna, only 30 search iterations per component were performed. Although a more exhaustive search could provide even more accurate estimates, the complexity of the model, with 11 or 12 components, implies lengthy processing and higher consumption of computational resources. Similarly, increasing the number of epochs and the number of layers could improve the search, but the processing would be much longer.

On the other hand, the amount of data used in the modeling corresponds to a period of approximately 1 year. Since the SIMONE radars are still active, this amount of data could increase. This would allow a better recognition of temporal patterns to make better estimates.

As for the input and output data, there is also a possibility for improvement. Two paths can be followed. First, keep the same input data window to predict fewer output steps. For example, one can predict 8 steps corresponding to 1 day of records. Second, one can increase the size of the input window to 112, corresponding to 2 weeks of records, while keeping the 16 output steps.

Detailed analysis of the predictions reveals a significant deterioration in the metrics towards the final steps. This observation

justifies the exploration of predictions with a shorter time horizon than the 16 steps, keeping the same input window. When reconstructing the time series using blocks of 16 estimates, it was observed that certain parts of these series did not correctly match the actual data. This mismatch was mainly attributed to the low precision of the estimates for the last steps. Despite the challenges, the proposed hybrid model has shown better performance than the model of dominant harmonics when comparing the predictions to the actual 2023 data using RMSE and correlation metrics. In other words, the hybrid model more accurately matched the actual data. Possibly, the hybrid model can capture other significant waves of planetary scale and mesoscale. This speculation will be evaluated in future works.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

CM: Writing—original draft, Writing—review and editing. JS: Writing—original draft, Writing—review and editing. MM: Writing—review and editing. CL: Writing—review and editing. KK: Writing—review and editing. DS: Writing—review and editing. RR: Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported and financed by PROCENCIA (Peru), under contract No 075-2021-FONDECYT.

Acknowledgments

We also wish to thank the Leibniz Institute of Atmospheric Physics for providing MLT winds from SIMONE radar networks. We also thank Jorge Chau for his suggestions that contributed to the development of this work. Additionally, we acknowledge the collaboration of the Estación Ramón Mujica of the Universidad of Piura and the Radio Observatorio de Jicamarca of the Instituto Geofísico del Perú for providing support in the operation of the SIMONE radar system. Finally, we extend our gratitude to the Pontificia Universidad Católica

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: a next-generation hyperparameter optimization framework," in In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery), 2623–2631.
- Ali, M., Khan, A., and Rehman, N. (2018). Hybrid multiscale wind speed forecasting based on variational mode decomposition. *Int. Trans. Electr. Energy Syst.* 28 (1), e2466. doi:10.1002/etep.2466
- Carvalho, V. R., Moraes, M. F. D., Braga, A. P., and Mendes, E. M. A. M. (2020). Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification. *Biomed. Signal Process. Control* 62, 102073. doi:10.1016/j.bspc.2020.102073
- Chau, J. L., Urco, J. M., Vierinen, J., Harding, B. J., Clahsen, M., Pfeffer, N., et al. (2021). Multistatic specular meteor radar network in Peru: system description and initial results. *Earth Space Sci.* 8 (1), 1–22. doi:10.1029/2020EA001293
- Conte, J. F., Chau, J. L., Yiğit, E., Suclupe, J., and Rodríguez, R. (2024). Investigation of mesosphere and lower thermosphere dynamics over central and northern Peru using SIMONE systems. *J. Atmos. Sci.* 81 (1), 93–104. doi:10.1175/JAS-D-23-0030.1
- Dhadly, M., Sassi, F., Emmert, J., Drob, D., Conde, M., Wu, Q., et al. (2023). Neutral winds from mesosphere to the thermosphere—past, present, and future outlook. *Front. Astronomy Space Sci.* 9. doi:10.3389/fspas.2022.1050586
- Dragomiretskiy, K., and Zosso, D. (2014). Variational mode decomposition. *IEEE Trans. Signal Process.* 62 (3), 531–544. doi:10.1109/TSP.2013.2288675
- Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Mlynarczyk, M. G., et al. (2021). NRLMSIS 2.0: a whole-atmosphere empirical model of temperature and neutral species densities. *Earth Space Sci.* 8 (3). doi:10.1029/2020EA001321
- Forbes, J. M. (1995). Tidal and planetary waves'. In *Geophysical monograph series*, edited by R. M. Johnson, and T. L. Killeen, 67–87. Washington, D. C.: American Geophysical Union. doi:10.1029/GM087p0067
- Gan, M., Pan, H., Chen, Y., and Pan, S. (2021). Application of the variational mode decomposition (VMD) method to river tides. *Estuar. Coast. Shelf Sci.* 261, 107570. doi:10.1016/j.ecss.2021.107570
- Gasparini, F., Jones, M. A., Harding, B. J., and Immel, T. J. (2023). Direct observational evidence of altered mesosphere lower thermosphere mean circulation from a major sudden stratospheric warming. *Geophys. Res. Lett.* 50 (7). doi:10.1029/2022GL102579
- Hanifi, S., Lotfian, S., Zare-Behtash, H., and Cammarano, A. (2022). Offshore wind power forecasting—a new hyperparameter optimisation algorithm for deep learning models. *Energies* 15 (19), 6919. doi:10.3390/en15196919
- Hussin, N. H., Yusof, F., Jamaludin, A. R., and Norrulashikin, S. M. (2021). Forecasting wind speed in peninsular Malaysia: an application of arima and arima-garch models. *Pertanika J. Sci. Technol.* 29 (1), 31–58. doi:10.47836/pjst.29.1.02
- Immel, T. J., Sagawa, E., England, S. L., Henderson, S. B., Hagan, M. E., Mende, S. B., et al. (2006). Control of equatorial ionospheric morphology by atmospheric tides. *Geophys. Res. Lett.* 33 (15). doi:10.1029/2006GL026161
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). Rainfall-runoff modelling using Long short-term memory (LSTM) networks. *Hydrology Earth Syst. Sci.* 22 (11), 6005–6022. doi:10.5194/hess-22-6005-2018
- Liu, H., Yamazaki, Y., and Lei, J. (2021). Day-to-Day variability of the thermosphere and ionosphere. *Geophys. Monogr. Ser.*, 275–300. doi:10.1002/9781119815631.ch15
- Liu, H., Li, Bardeen, C. G., Foster, B. T., Lauritzen, P., Liu, J., Lu, G., et al. (2018). Development and validation of the whole atmosphere community climate model with thermosphere and ionosphere extension (WACCM-X 2.0). *J. Adv. Model. Earth Syst.* 10 (2), 381–402. doi:10.1002/2017MS001232
- Mauricio, C., Suclupe, J., Milla, M., López de Castilla, C., Karim, K., Rodríguez, R., et al. (2023). "Short-term prediction of wind speed in the mesosphere and lower thermosphere over Peru's coastal north and central," in 2023 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Valdivia, Chile, 05-07 December 2023 (IEEE), 1–6.
- McCormack, J. P., Coy, L., and Singer, W. (2014). Intraseasonal and interannual variability of the quasi 2 Day wave in the northern hemisphere summer mesosphere. *J. Geophys. Res. Atmos.* 119 (6), 2928–2946. doi:10.1002/2013JD020199
- Piani, C., Durran, D., Alexander, M. J., and Holton, J. R. (2000). A numerical study of three-dimensional gravity waves triggered by deep tropical convection and their role in the dynamics of the QBO. *J. Atmos. Sci.* 57 (22), 3689–3702. doi:10.1175/1520-0469(2000)057<3689:ANSOTD>2.0.CO;2
- Pigott, T. D. (2001). A review of methods for missing data. *Educ. Res. Eval.* 7 (4), 353–383. doi:10.1076/edre.7.4.353.8937
- Rosa, L., La, G., and Sanchez, I. (2020). "Hybrid models based on mode decomposition and recurrent neural networks for streamflow forecasting in the chira river in Peru," in Proceedings of the 2020 IEEE Engineering International Research Conference, EIRCON 2020, Lima, Peru, 21-23 October 2020 (Institute of Electrical and Electronics Engineers Inc.).
- Schneider, T. (2001). Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* 14 (5), 853–871. doi:10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2
- Son, N., and Jung, M. (2020). Analysis of meteorological factor multivariate models for medium- and long-term photovoltaic solar power forecasting using Long short-term memory. *Appl. Sci.* 11 (1), 316. doi:10.3390/app11010316
- Staszak, T., Strelnikov, B., Latteck, R., Renkwitz, T., Friedrich, M., Baumgarten, G., et al. (2021). Turbulence generated small-scale structures as PMWE formation mechanism: results from a rocket campaign. *J. Atmos. Solar-Terrestrial Phys.* 217, 105559. doi:10.1016/j.jastp.2021.105559
- Suclupe, J., Chau, J. L., Federico Conte, J., Milla, M., Pedatella, N. M., and Kuyung, K. (2023). Climatology of mesosphere and lower thermosphere diurnal tides over Jicamarca (12°S, 77°W): observations and simulations. *Earth, Planets Space. Springer Sci. Bus. Media Deutschl. GmbH* 75, 186. doi:10.1186/s40623-023-01935-z
- Vincent, R. A. (2015). The dynamics of the mesosphere and lower thermosphere: a brief review. *Prog. Earth Planet. Sci.* 2 (1), 4. doi:10.1186/s40645-015-0035-8
- Yang, S., Yang, H., Li, N., and Ding, Z. (2023). Short-term prediction of 80–88 Km wind speed in near space based on VMD-PSO-LSTM. *Atmosphere* 14 (2), 315. doi:10.3390/atmos14020315

del Perú y la Universidad Nacional Agraria La Molina for guiding the methodology of data processing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.