

PREDICTIBILIDAD DE LOS VIENTOS EN LA REGIÓN MLT

Sobre Lima y Piura

C.Mauricio^{1 5}, J. Suclupe², M. Milla³, C. López de Castilla¹,
K. Kuyeng⁴, D. Scipion⁴, R. Rodriguez⁵

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

² Leibniz Institute of Atmospheric Physics, Rostock University, Kühlungsborn, Germany.

³ Pontificia Universidad Católica del Perú, Sección de Electricidad y Electrónica, Lima – Peru

⁴ Instituto Geofísico del Perú, Radio Observatorio de Jicamarca, Lima, Peru.

⁵ Estación Científica Ramón Mujica, Universidad de Piura, Piura, Peru





CONTENIDO

1. Objetivo y descripción de investigación
2. Datos
3. Metodología
4. Resultados
5. Conclusiones



1. Objetivo y descripción de investigación

Objetivo:

Construir modelo de predicción de vientos en la región MLT usando Machine learning.

Datos a usar:

Vientos zonales y meridionales entre 80 y 95 km de altura.

Técnicas usadas:

Imputación de datos, descomposición de series temporales, red neuronal, optimizador de hiperpámetros.

Fuentes de datos:

Radares SIMONE ubicados en Jicamarca y Piura

Software:

Python y R





2. DATOS

	Lima	Piura
Descripción	Velocidad del viento: componente zonal y meridional	
Alturas	80.5, 85.5, 90.5 y 95.5 km	
Análisis del modelo	Setiembre 2020 - Noviembre 2021 número de registros: 21888	Octubre 2021 - Diciembre 2022 número de registros: 21888
Comparación extra	2023	

Análisis del modelo





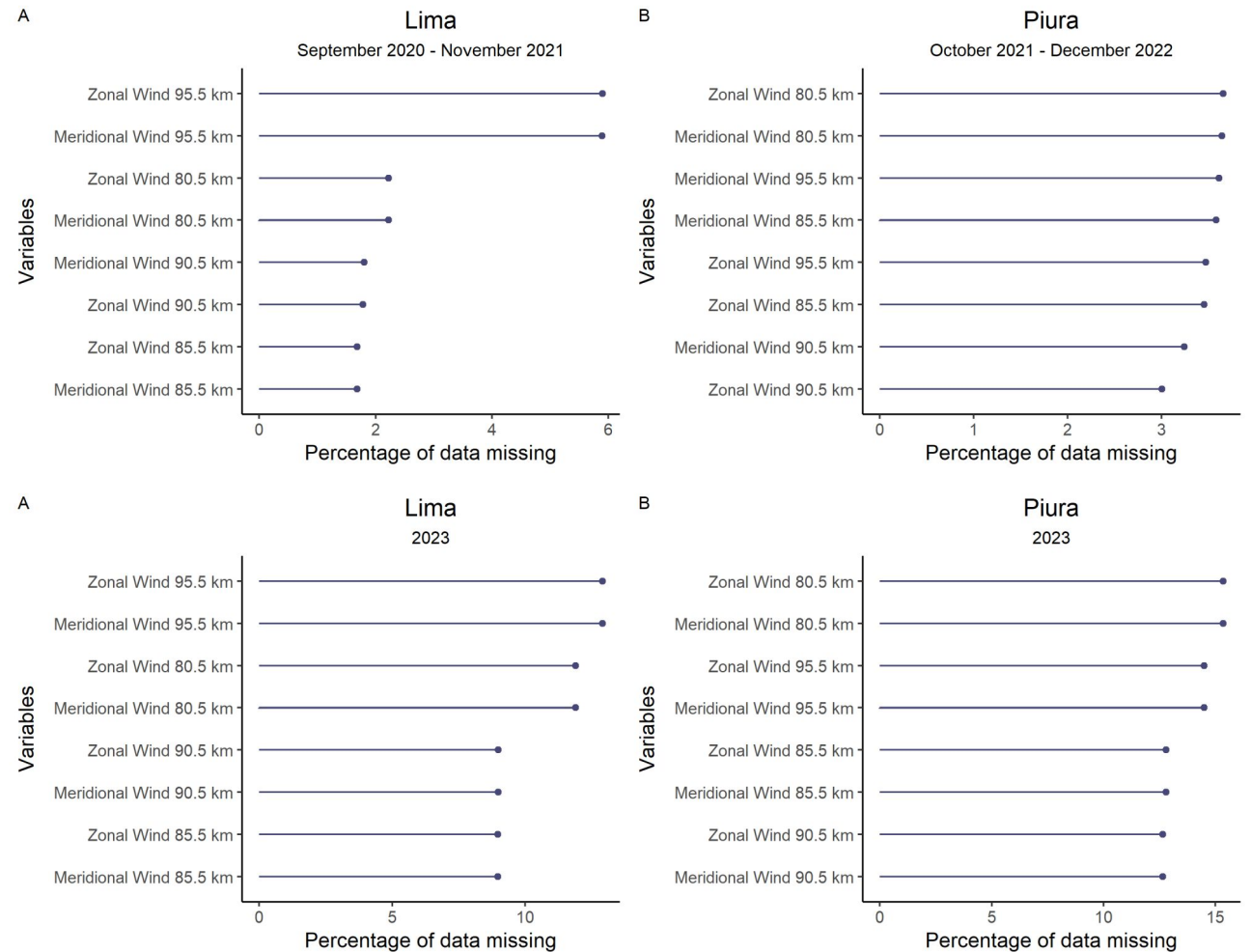
3. METODOLOGÍA

3.1 Imputación de datos faltantes con el algoritmo esperanza-maximización (EM)

El algoritmo EM es un método iterativo para estimar parámetros con datos incompletos, que siguen una distribución del tipo Gaussiano.

Primero estima los valores faltantes basándose en los datos observados y los parámetros actuales. Luego, actualiza los parámetros del modelo usando estas estimaciones.

Este proceso se repite hasta que las estimaciones de los parámetros convergen.



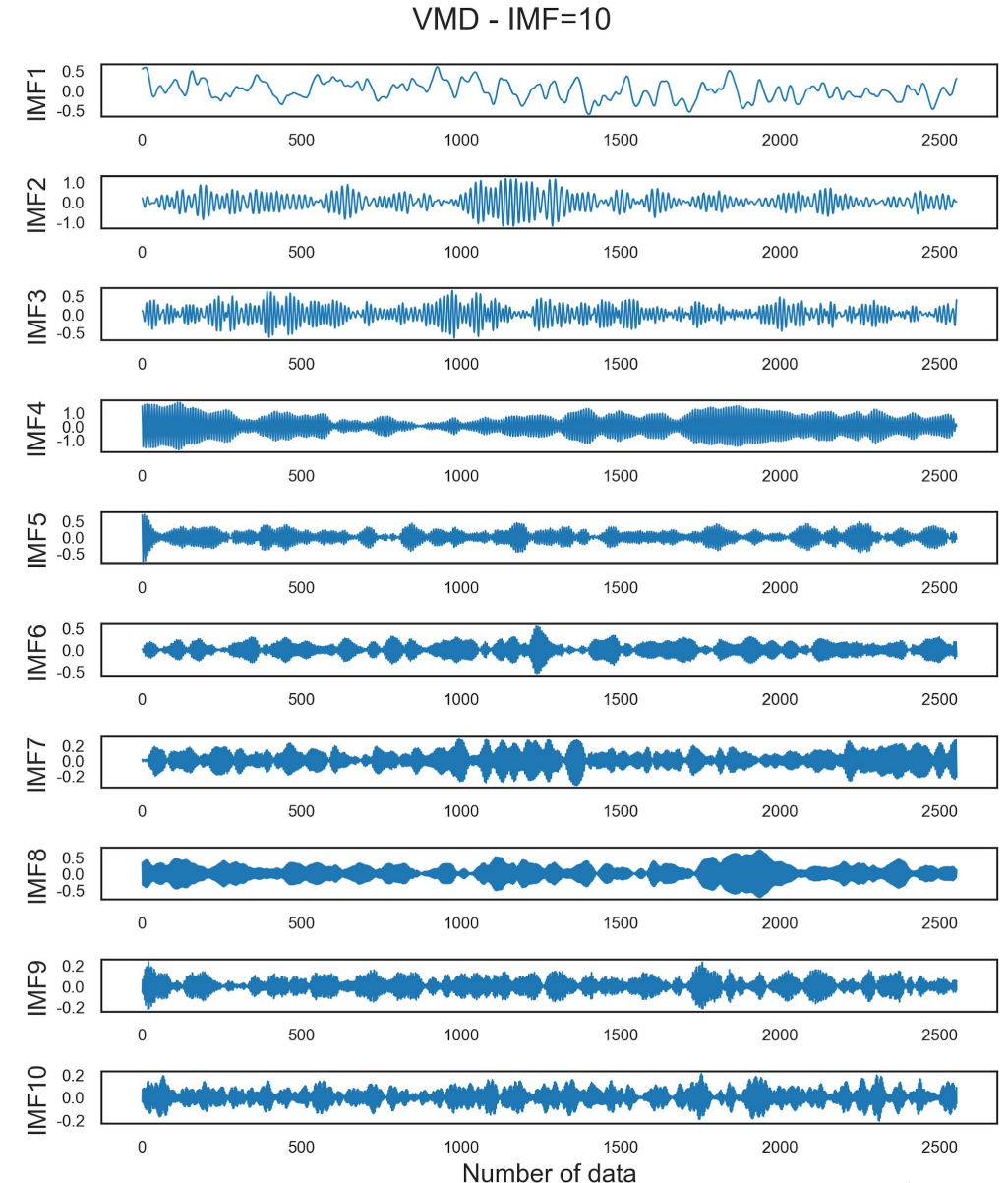


3. METODOLOGÍA

3.2 Descomposición VMD

Este método descompone las señales complejas en partes más simples denominadas modos intrínsecos (IMF). Esta técnica utiliza funciones oscilatorias para captar distintas partes de la señal (con mucha variabilidad) en términos de tiempo y frecuencia.

K	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7	IMF 8	IMF 9	IMF 10
1	0.0550									
2	0.0541	0.2466								
3	0.0473	0.1284	0.2746							
4	0.0464	0.1250	0.2482	0.3717						
5	0.0569	0.1245	0.1792	0.2513	0.3727					
6	0.0110	0.0644	0.1266	0.2477	0.3055	0.3786				
7	0.0102	0.0620	0.1252	0.1830	0.2505	0.3703	0.4544			
8	0.0102	0.0617	0.1250	0.1778	0.2486	0.3039	0.3743	0.4617		
9	0.0101	0.0616	0.1248	0.1716	0.2469	0.2697	0.3140	0.3754	0.4645	
10	0.0100	0.0613	0.1243	0.1452	0.1915	0.2486	0.2825	0.3235	0.3770	0.4665



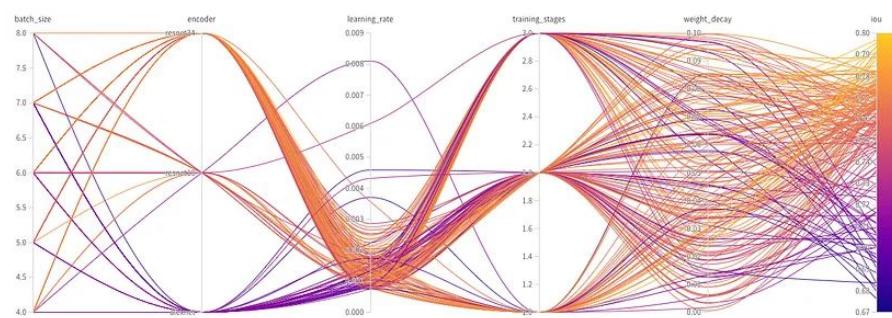


3. METODOLOGÍA

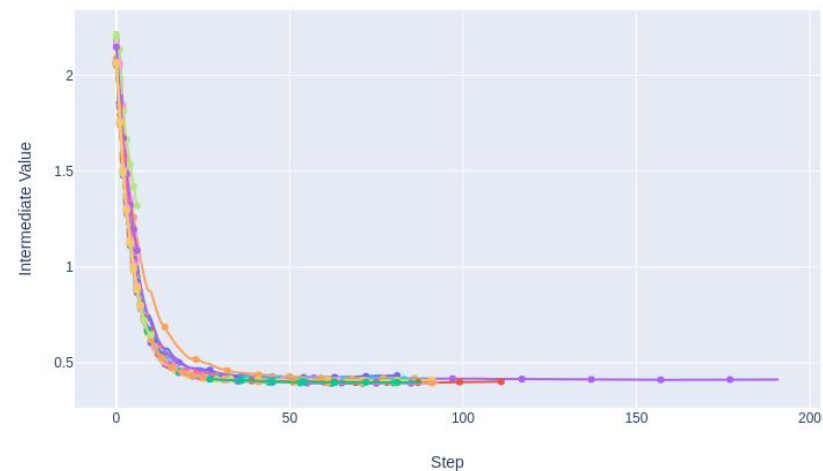
3.3 LSTM + Optuna

La integración de LSTM con Optuna optimiza la búsqueda de hiperparámetros de cada IMF. Cada componente se procesa a través de un LSTM, donde Optuna va analizando combinaciones de hiperparámetros. Este software tiene una función objetivo y un algoritmo de poda que permiten seleccionar de manera eficiente, disminuyendo tiempos de procesamiento.

Hyperparameters	
Capas	1
Neuronas	Valores entre 1 y 100
Dropout	Valores entre 0 y 0.5
Learning rate	Valores entre 0.00001 y 0.1
Epoch	100,120,140,160,180 y 200
Batchsize	16, 24, 32, 40, 48, 56, 64
Lambda (regularizador L2)	Valores entre 0.001 y 0.01



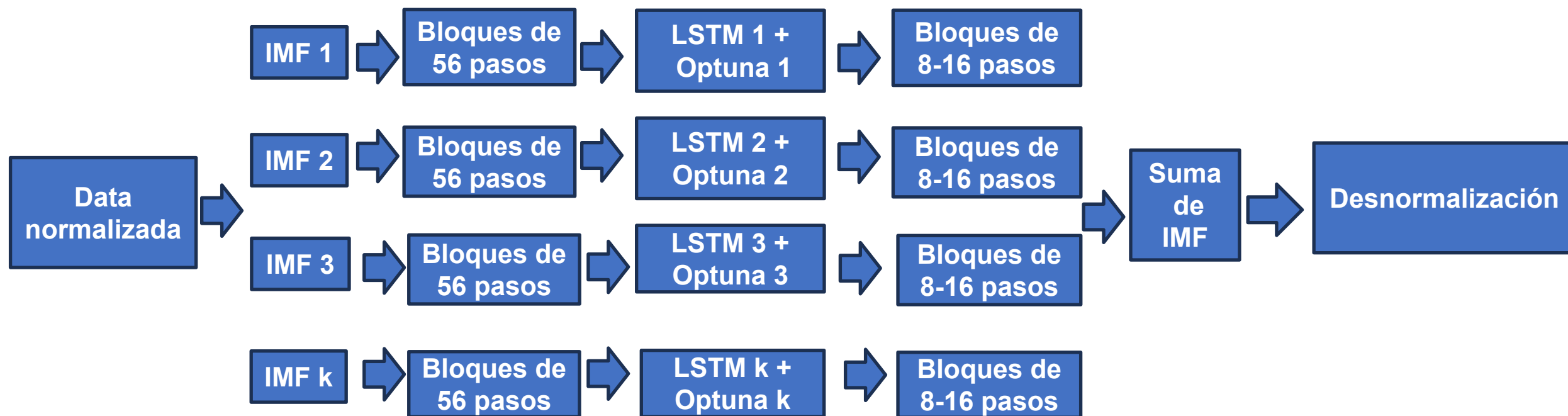
Intermediate Values Plot





3. METODOLOGÍA

3.4 Modelo Híbrido





3. METODOLOGÍA

3.4 Evaluación de modelo

Evaluación con serie reconstruida

	Serie temporal reconstruida																							
Predicción 1	X1	X2	X3	X4	X5	X6	X7	X8																
Predicción 2								X9	X10	X11	X12	X13	X14	X15	X16									
Predicción 3															X17	X18	X19	X20	X21	X22	X23	X24		
⋮																								...
	bloque 1 (8 pasos)							bloque 2 (8 pasos)								bloque 3 (8 pasos)								...

Evaluación por pasos

Serie de tiempo	X1	X2	X3	...	X8	X9	X10	...
Predicción 1	Paso 1	Paso 2	Paso 3	...	Paso 8			
Predicción 2		Paso 1	Paso 2	Paso 3	...	Paso 8		
Predicción 3			Paso 1	Paso 2	Paso 3	...	Paso 8	
⋮				...	...	...	...	...

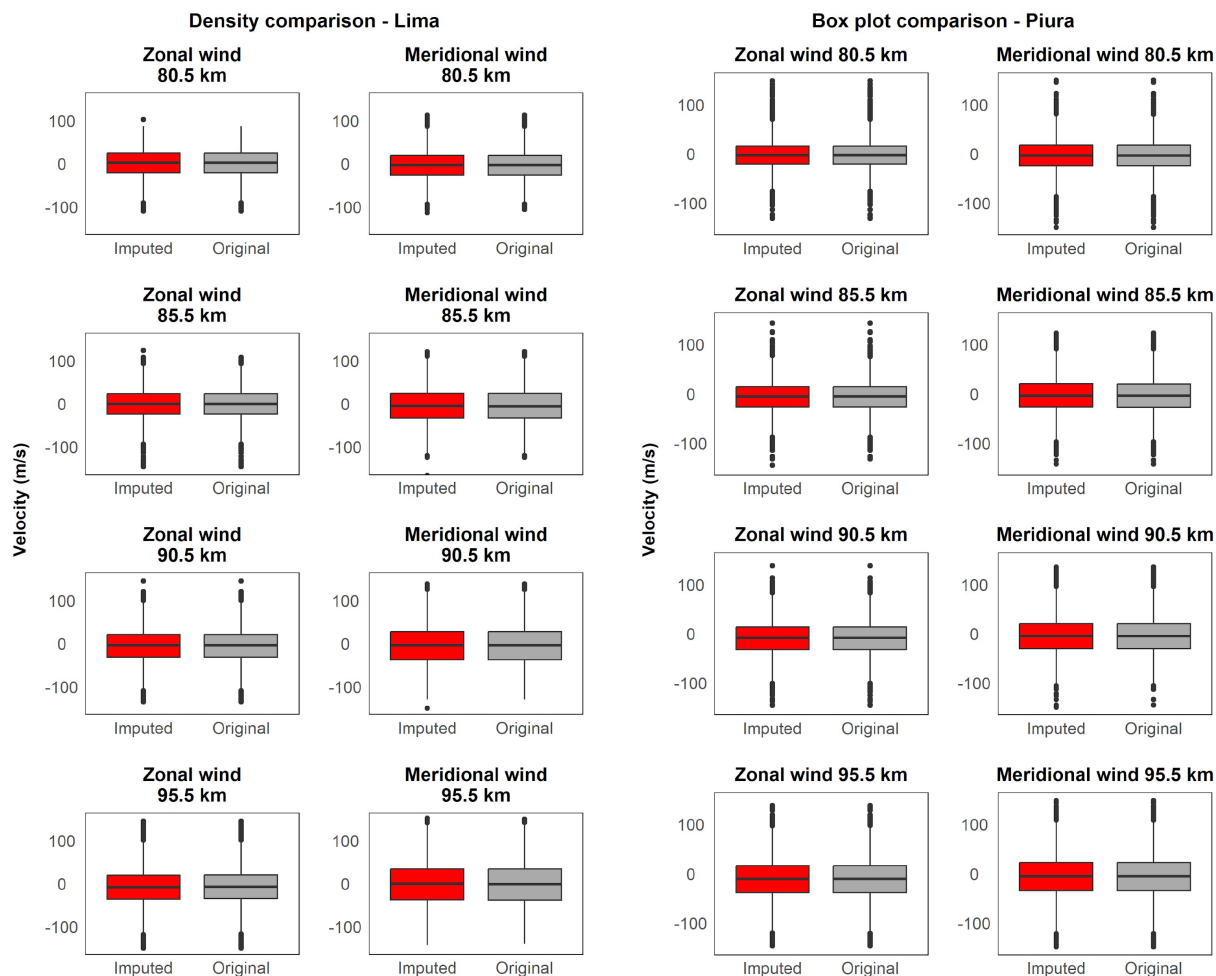
Métricas

- Correlación
- Raíz del error cuadrático medio (RMSE)



4. Resultados

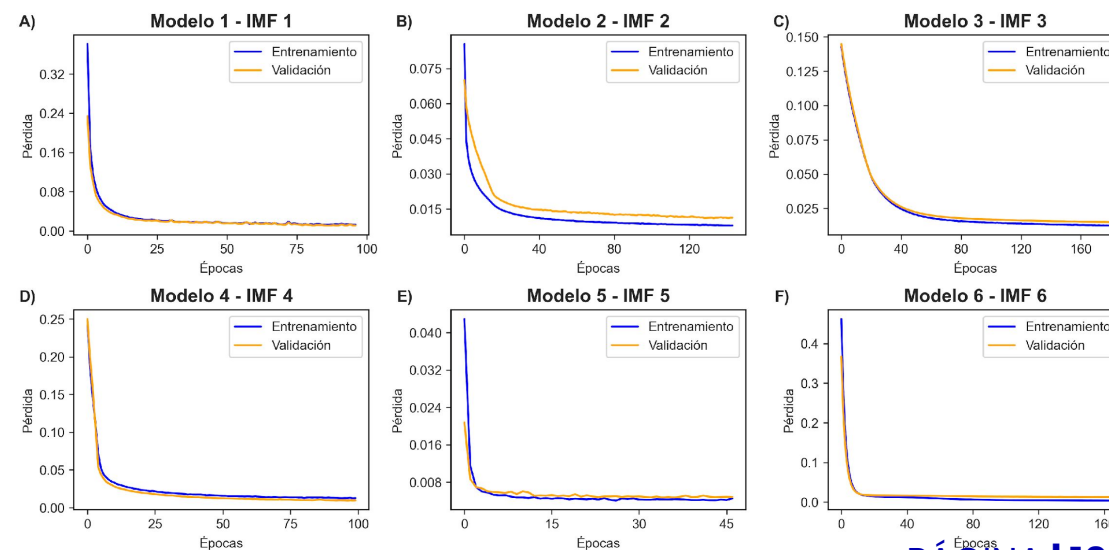
4.1 Imputación de datos



4.2 Componentes óptimas

Alturas	Jicamarca		Piura	
	Viento zonal	Viento meridional	Viento zonal	Viento meridional
80.5 km	11	11	11	11
85.5 km	11	10	11	10
90.5 km	11	11	10	11
95.5 km	11	11	11	11

4.3 LSTM + Optuna

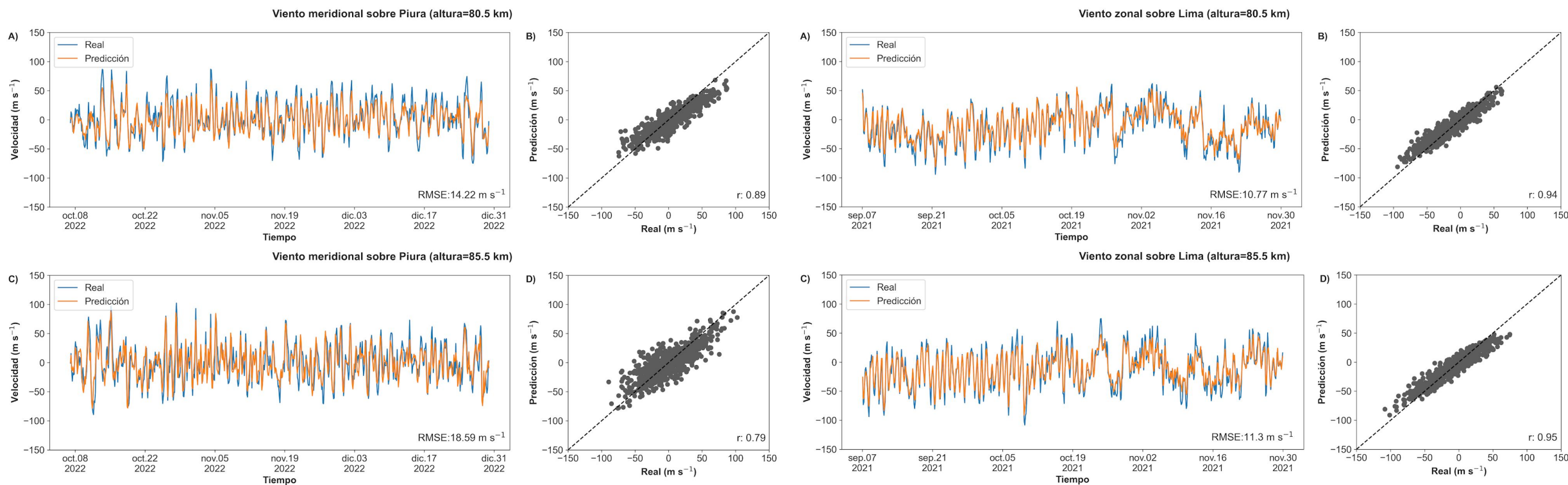




4. Resultados

4.4 Evaluación con series reconstruidas

En la etapa de modelamiento las series temporales reconstruidas con el modelo híbrido, presentan valores de correlación superiores a 0,54, y valores de RMSE inferiores a 28 m/s.

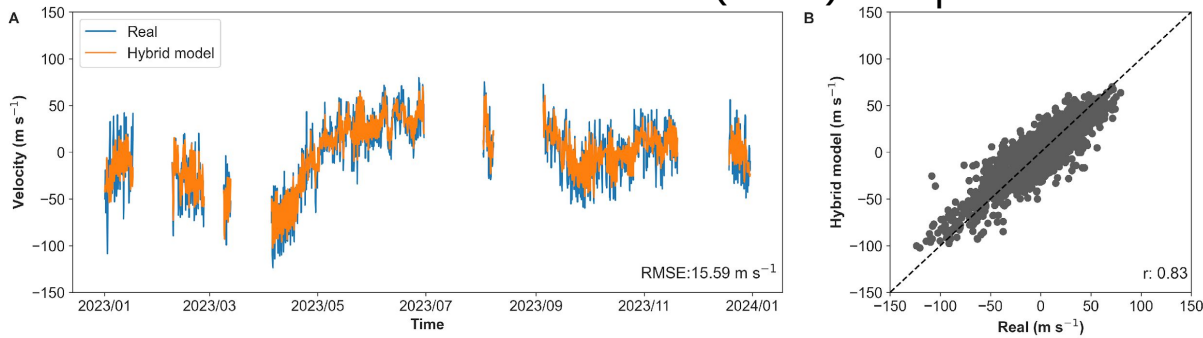




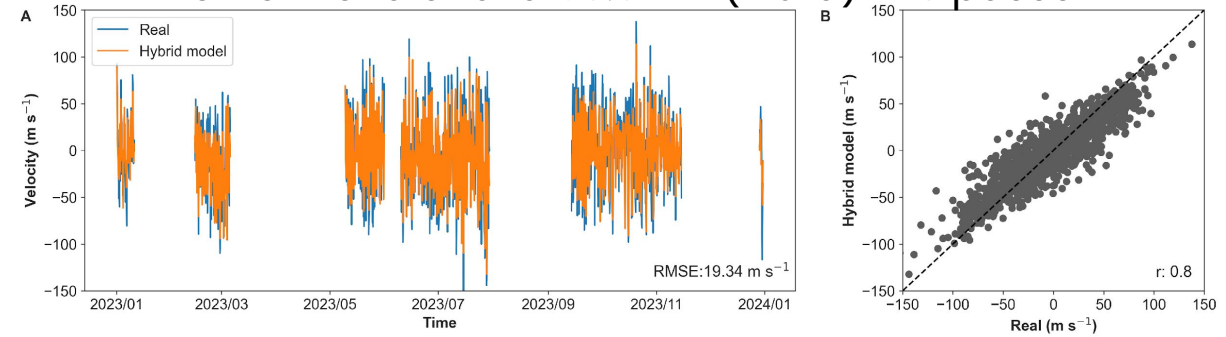
4. Resultados

4.4 Predicción 2023

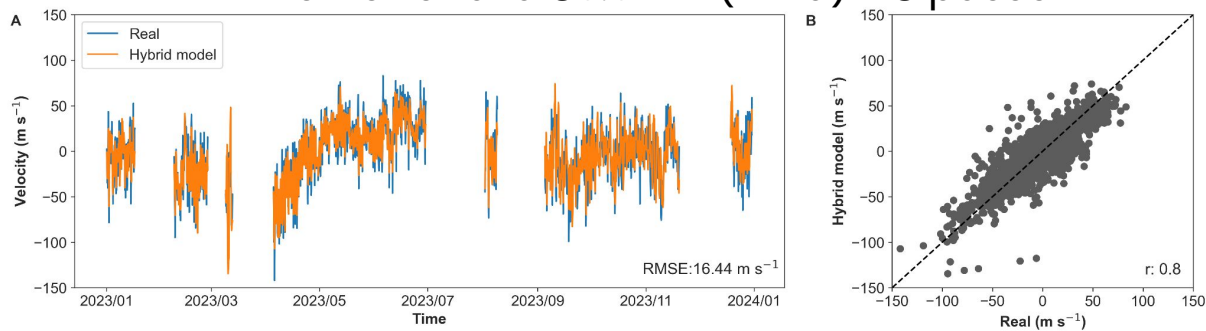
Viento zonal a 80.5 km (Lima) – 16 pasos



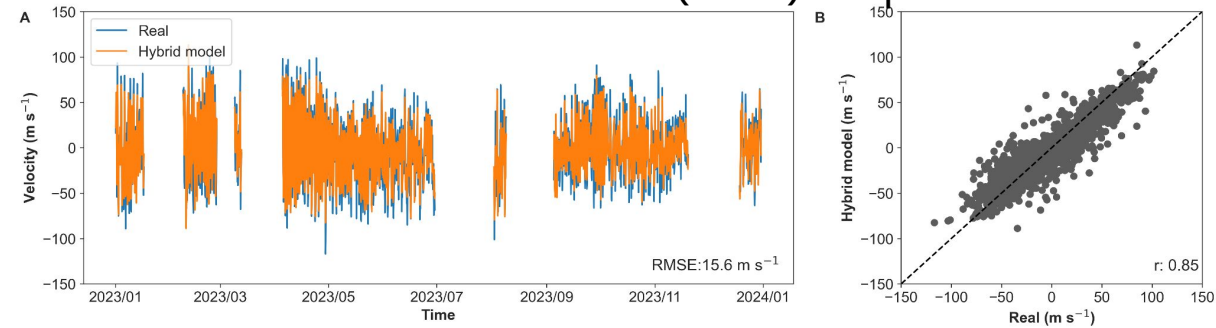
Viento meridional a 95.5 km (Piura) – 16 pasos



Viento zonal a 85.5 km (Lima) – 8 pasos



Viento meridional 85.5 km (Lima) – 8 pasos



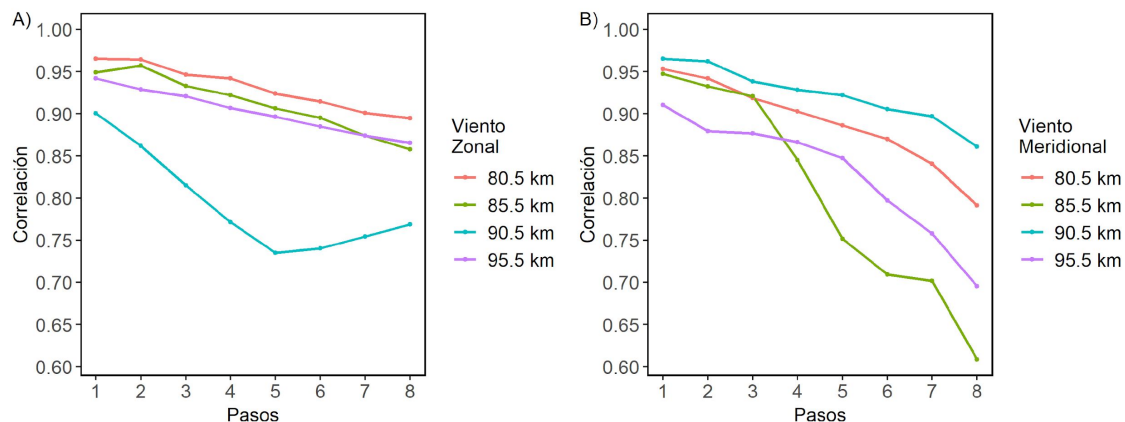


4. Resultados

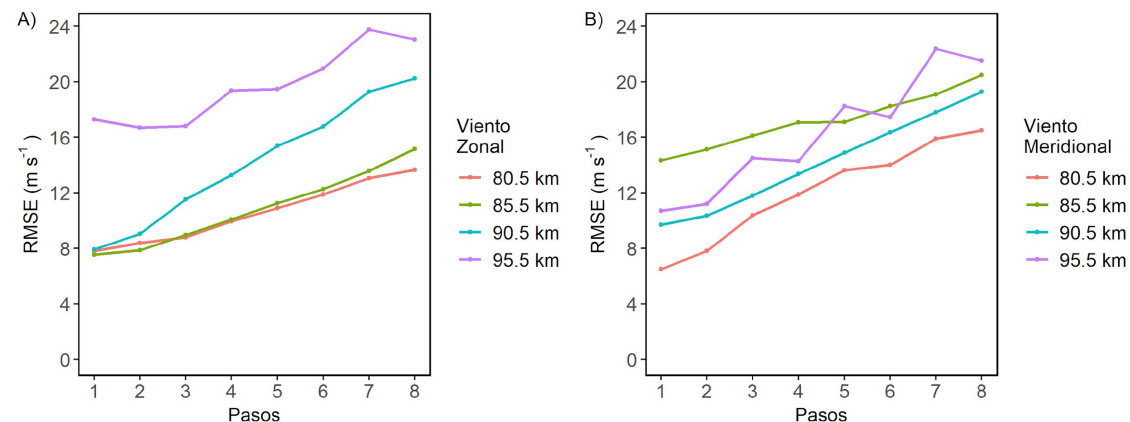
4.4 Evaluación con series reconstruidas

En la etapa del test, las predicciones de 8 pasos tienen correlaciones que disminuyen desde 0.99 hasta 0.61. Mientras que el RMSE puede aumentar desde 6.74 m/s hasta 25.60 m/s.

Piura - correlación - 8 pasos



Lima - RMSE - 8 pasos



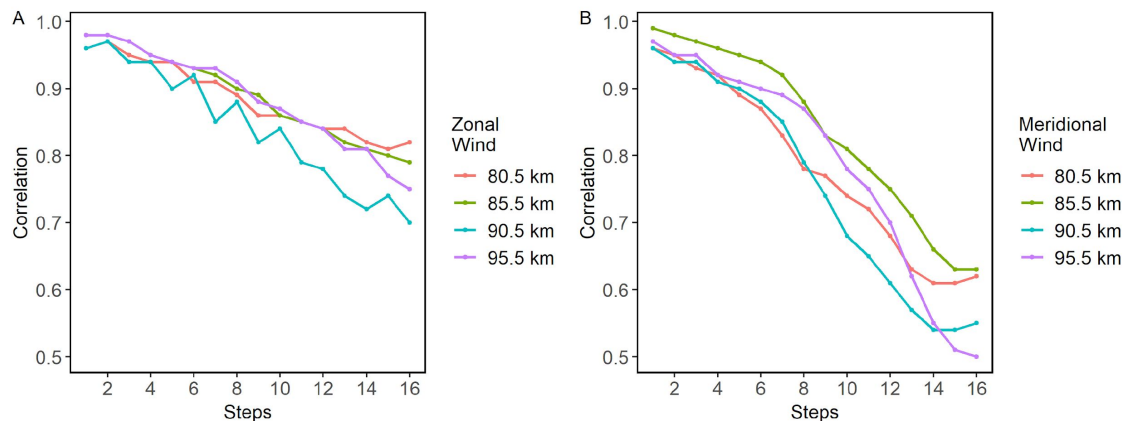


4. Resultados

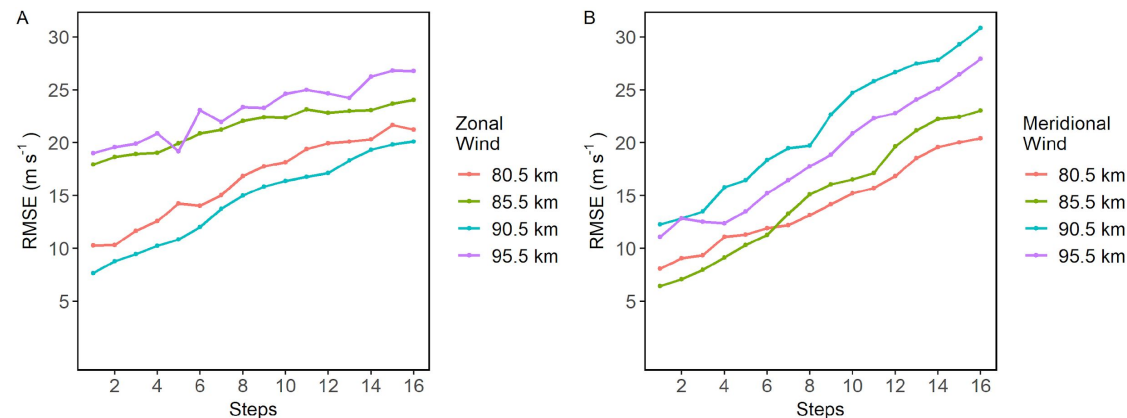
4.4 Evaluación con series reconstruidas

En la etapa del test, las predicciones de 16 pasos tienen correlaciones que disminuyen desde 0.99 hasta 0.66. Mientras que el RMSE puede aumentar desde 6.46 m/s hasta 30.84 m/s.

Piura - correlación - 16 pasos



Lima - RMSE - 16 pasos





5. Conclusiones

- El algoritmo EM consigue una imputación adecuada de los valores perdidos. En específico el algoritmo funciona bien con porcentajes máximo de 6% de datos faltantes.
- La descomposición VMD resulta fundamental para optimizar el modelado de cada componente. Se detectó que el número de componentes óptimos no supera los 12 IMF.
- Optuna facilita la evaluación del rendimiento del modelo y mejora la eficiencia en la búsqueda de los mejores hiperparámetros..
- El análisis por reconstrucción de series temporales demuestra que las predicciones del modelo híbrido se ajustan a los datos reales.
- El análisis por pasos demuestra que cuanto más largo es el horizonte de predicción, más difícil es conseguir buenas estimaciones. El error aumenta.
- Las predicciones de 2023 demuestran la capacidad del modelo para generar predicciones aceptables en la región MLT sobre Lima y Piura. El modelo capta la variabilidad y las modulaciones día a día, incluso bajo la influencia de ondas planetarias cuasi-diurnas y la presencia del evento El Niño Costero.



Organizado por:



IGP

Instituto
Geofísico
del Perú



PUCP



UNIVERSIDAD
DE PIURA

Con el auspicio de:



Convenio No 075-2021-FONDECYT



*Ciencia para **protegernos***
*Ciencia para **avanzar***