

# Predicción de la Velocidad del Viento en La Serena, Chile, mediante Redes Neuronales Convolucionales Temporales y Modelos SARIMAX

Raúl Frugone Zaror, Carlos Osses Álfaro & Benjamín Saéz Alcayaga

9 de julio de 2025

## Abstract

This research focused on analyzing and forecasting wind speed in La Serena, Chile, for the period between June 23 and July 23, 2025. Historical meteorological time series were processed, and relevant exogenous variables were modeled using Temporal Convolutional Networks (TCN), which are capable of capturing long-term temporal dependencies and complex cyclic patterns. Two approaches were implemented for wind speed prediction: a classical statistical SARIMAX model with exogenous variables and an advanced multivariate TCN architecture.

The results show that both models effectively fit and predict the time series with reasonable accuracy; however, the TCN model achieved significantly lower error metrics ( $MAE \approx 0.3$ ,  $RMSE \approx 0.52$ ) compared to SARIMAX(2,0,3)(2,0,3,24) ( $MAE \approx 1.49$ ,  $RMSE \approx 1.87$ ), demonstrating its superior ability to capture the variability and temporal trends of the wind. Additionally, the TCN maintained stability and coherence in its predictions when compared to historical data from the previous year, a key aspect for operational applications in wind energy generation.

This study confirms the growing utility of deep learning models in local meteorological forecasting, especially in regions with complex topographic and atmospheric conditions. The combination of physical-statistical methods and artificial intelligence techniques emerges as a promising strategy to improve accuracy and reliability in wind forecasting, thereby supporting energy planning and sustainable management of renewable resources.

**Keywords:** Wind forecasting, Temporal Convolutional Networks (TCN), SARIMAX model, Meteorological time series, La Serena, Renewable energy, Statistical models.

## Resumen Ejecutivo

Esta investigación se centró en el análisis y predicción de la velocidad del viento en La Serena, Chile, para el periodo comprendido entre el 23 de junio y el 23 de julio de 2025. Se procesaron series temporales meteorológicas históricas y se modelaron variables exógenas relevantes mediante redes neuronales convolucionales temporales (TCN), las cuales permiten capturar dependencias temporales a largo plazo y patrones cíclicos complejos. Para la predicción de la velocidad del viento se implementaron dos enfoques: un modelo estadístico clásico SARIMAX con variables exógenas y una arquitectura avanzada basada en TCN multivariada.

Los resultados muestran que ambos modelos son capaces de ajustar y predecir la serie temporal con buena precisión, aunque la red TCN presenta métricas de error ( $MAE \approx 0.3$ ,  $RMSE \approx 0.52$ ) son inferiores a las obtenidas por SARIMAX(2,0,3)(2,0,3,24) ( $MAE \approx 1.49$ ,  $RMSE \approx 1.87$ ), evidenciando su superior capacidad para capturar la variabilidad y tendencias temporales del viento. Además, la TCN demostró tener estabilidad y coherencia en sus predicciones al compararlas con datos históricos del año anterior, aspecto clave para aplicaciones operativas en generación eólica.

Este estudio confirma la creciente utilidad de los modelos de aprendizaje profundo en el pronóstico meteorológico local, especialmente en regiones con condiciones topográficas y atmosféricas complejas. La combinación de métodos físicos-estadísticos y técnicas de inteligencia artificial emerge como una estrategia pro-

metedora para mejorar la precisión y confiabilidad del pronóstico eólico, contribuyendo así a la planificación energética y la gestión sostenible de recursos renovables.

**Palabras clave:** Predicción de viento, Redes convolucionales temporales (TCN), Modelo SARIMAX, Series temporales meteorológicas, La Serena, Modelos estadísticos.

## Introducción

La investigación reciente en Sudamérica sobre pronóstico de velocidad de viento se concentra principalmente en Brasil, Perú, Ecuador, Colombia, Chile y Argentina, todas estas regiones son influidas por el Anticiclón del Pacífico Sur (APS) y distintos chorros de nivel superior (jet streams), estos jet streams corresponden bandas de viento fuerte y estrechas que se encuentran en la atmósfera terrestre, generalmente a altitudes elevadas (entre 7 y 16 kilómetros) y que se extienden a lo largo de miles de kilómetros, con transiciones topográficas irregulares entre cordillera, valle central y llanura costera. Siguiendo la idea anterior, aunque cada estudio de los buscados empleo conjuntos de datos específicos, en general comparten la combinación de observaciones de torres anemométricas (TA) (Dematte, Gandolfo Raso, y Huespe, 2023; Lazzús y Salfate, 2017) con productos de re-análisis atmosférico (por ejemplo ERA5 o MERRA-2) para extender sus series cuando las campañas locales son demasiado cortas o presentan discontinuidades (Miguel, Fadigas, y Sauer, 2019). En estudios que disponen de mayor capacidad de cómputo, se añaden salidas del Modelo de Investigación y Pronóstico del Tiempo (WRF, Weather Research and Forecasting) como variables explicativas, lo que refuerza la coherencia físico-estadística, especialmente en zonas costeras y mar adentro, como se observa en la Cuenca de Campos (Soares da Silva, Dragaud, y Jacinto, 2024).

En cuanto a la vertiente metodológica, todos los artículos utilizan alguna versión de modelo de persistencia o Modelo Autorregresivo Integrado de Media Móvil (ARIMA) como línea base. Sin embargo, estos enfoques lineales exhiben errores entre 15 % y 30 % superiores a los métodos modernos (Costa, Ruiz-Cárdenas, y Mineti, 2021). La mejora más notable proviene del Aprendizaje Automático Tradicional (AAT), donde algoritmos como la Máquina de Incremento por Gradiente (GBM, Gradient Boosting Machine) y los Random Forest (RF) reducen el Error Cuadrático Medio (RMSE) alrededor de un 20 % en el litoral de Osório, logrando un Error Porcentual Absoluto Medio (MAPE) cercano al 7 % (Khosravi, Machado, y Nunes, 2018). Asimismo, la utilización de Redes Neuronales Artificiales (ANN, Artificial Neural Networks) optimizadas con el Algoritmo de Enjambre de Partículas (PSO, Particle Swarm Optimization) en La Serena permitió mejorar el RMSE en torno al 15 % frente a una ANN convencional (Lazzús y Salfate, 2017).

La tendencia a mezclar aproximaciones lineales y no lineales se extiende al nordeste brasileño, donde la inserción de residuos ARIMA dentro de una ANN alcanzó errores mensuales inferiores al 4 % (Camelo, Lucio, LealJunior, dos Santos, y de Carvalho, 2018). A escala continental, Almeida y colaboradores comprobaron en múltiples emplazamientos que las ANN, cuando se alimentan con conjuntos de datos homogéneos, reducen el RMSE aproximadamente un 10 % con respecto a la persistencia hasta un horizonte de seis horas (Almeida, Bastos, y Santos, 2021).

La irrupción del aprendizaje profundo ha permitido capturar de forma explícita la dependencia espacio-temporal en el campo eólico. El modelo U-Convolutional (CNN, Convolutional Neural Network) (Bastos, Cyrino Oliveira, y Milidiú, 2021b) y la red ComPonentNet (CPNet) (Bastos, Cyrino Oliveira, y Milidiú, 2021a) integran no solo la magnitud de la velocidad sino también sus componentes vectoriales  $U$  y  $V$ , aumentando la coherencia direccional y mejorando la Media Absoluta del Error (MAE) entre un 8 % y un 15 %. Estos resultados se potencian cuando se entrena con conjuntos depurados; de acuerdo con Goulart y Camargo, filtrar episodios extremos o ventanas meteorológicamente heterogéneas reduce el MAE alrededor de un 10 % (Goulart y de Camargo, 2021).

El rendimiento de las arquitecturas profundas sigue dependiendo de la topografía y de los forzantes sinópticos locales, los forzantes sinópticos se refieren a las condiciones atmosféricas a gran escala que influyen en el clima y el tiempo a nivel local. Estas condiciones son típicamente patrones meteorológicos que se extienden sobre cientos o miles de kilómetros, como sistemas de alta y baja presión, frentes y corrientes en chorro. Los forzantes sinópticos son importantes porque pueden modificar la temperatura, la precipitación, la humedad y otros elementos del clima en una región específica. En particular, las Capas Límite Atmosféricas Estables (ACLE) que se forman durante la noche en sotavento de la Cordillera de la Costa chilena o del Planalto brasileño provocan subestimaciones en vientos débiles y en rachas fuertes. La nueva parametrización

estable propuesta por Maroneze, et al. (2023), demuestra que una mejor representación de la ACLE dentro de WRF puede reducir estos sesgos hasta en un 30 % (Maroneze, Costa, y Acevedo, 2023).

Al sintetizar los hallazgos, es evidente que las redes profundas especialmente cuando se combinan con modelos físicos y técnicas de hibridación evolutiva ofrecen una mayor precisión para horizontes operativos de 1 a 24 horas. No obstante, casi todas las investigaciones enfatizan la necesidad de al menos tres años de medición con TA. Miguel, et al. muestran que extender las campañas de dos a cinco años dentro del método Medir–Correlacionar–Predecir (MCP) recorta la incertidumbre de energía anual cerca de un 40 % (Miguel y cols., 2019). Sin embargo, los eventos extremos siguen siendo un reto, por lo que se recomienda aplicar corrección de sesgo y funciones de pérdida asimétricas para mejorar la captura de rachas.

Finalmente, la literatura es todavía escasa para las zonas altoandinas y la Patagonia chilena. Extender estas metodologías—integrando salidas WRF de alta resolución con ensambles CNN/LSTM y validando mediante Validación Cruzada Bloqueada (BCV, Blocked Cross-Validation) podría disminuir la incertidumbre operativa en futuros parques eólicos.

Cuadro 1: Variables meteorológicas y auxiliares utilizadas por los estudios revisados.

Estudio	Variables empleadas
Lazzús y Salfate (2017)	Velocidad y dirección del viento a 10 m; fecha-hora.
Khosravi y cols. (2018)	Velocidad y dirección a 80 m; temperatura del aire; humedad relativa.
Costa y cols. (2021)	Velocidad del viento a 60 m; temperatura del aire.
Camelo y cols. (2018)	Velocidad media mensual; presión a nivel del mar.
Bastos y cols. (2021b)	Componentes $U/V$ del viento (reanálisis) + registros TA; elevación de la estación.
Bastos y cols. (2021a)	Componentes $U/V$ del viento; temperatura de superficie del mar.
Almeida y cols. (2021)	Velocidad del viento a 60 m; potencia eólica generada; temperatura del aire.
J. R. Silva, Dias, y Moreira (2022)	Velocidad y dirección (hub-height); humedad relativa.
Soares da Silva y cols. (2024)	Salidas WRF (viento a 10 m y 100 m; presión; temperatura del mar).
Maroneze y cols. (2023)	Flujos de calor sensible/latente; temperatura del suelo; velocidad del viento.
Dematte y cols. (2023)	Velocidad del viento a 50 m; rugosidad del terreno; altitud.
Goulart y de Camargo (2021)	Velocidad del viento; temperatura; presión superficial.
Miguel y cols. (2019)	Velocidad del viento (TA); correlaciones con reanálisis ERA5.
D. Silva y Lima (2018)	Velocidad y dirección en chorro NE; frecuencia diaria del jet.

En conjunto, la mayoría de los estudios ya incluyen variables básicas de viento y temperatura, mientras que variables de obtención más difícil como los cambios en la velocidad o dirección del viento en la atmósfera de manera vertical, la altura de la capa límite y los flujos turbulentos siguen poco explorados y representan una oportunidad de mejora para la investigación.

Esta investigación busca responder ¿Cuál será el comportamiento esperado de la velocidad del viento en La Serena, Chile, durante el periodo comprendido entre el 23 de junio y el 23 de julio de 2025? mediante una serie de objetivos:

- Procesar las series temporales meteorológicas históricas de La Serena, Chile, incluyendo la velocidad del viento y sus covariables asociadas, para el período de entrenamiento previo al 23 de junio de 2025.
- Modelar las covariables meteorológicas relevantes mediante redes neuronales convolucionales temporales (TCN), obteniendo predicciones que aseguren su disponibilidad para modelos de predicción de velocidad del viento.
- Implementar dos enfoques de modelado para la velocidad del viento, un modelo clásico ARIMAX que integre las covariables como variables exógenas y una arquitectura de red neuronal TCN multivariada.
- Evaluar el desempeño predictivo de ambos modelos (ARIMAX y TCN) utilizando métricas cuantitativas como el Error Absoluto Medio (MAE) y la Raíz del Error Cuadrático Medio (RMSE).
- Analizar el comportamiento proyectado de la velocidad del viento en La Serena durante el mes en estudio.

## Datos y metodología

La presente investigación adopta un enfoque cuantitativo, basado en el análisis de series temporales de velocidad del viento registradas en diversos puntos de medición en Chile. El estudio se enmarca dentro de un

tipo descriptivo y predictivo, cuyo objetivo principal es analizar la variabilidad espacio-temporal del viento y construir modelos de pronóstico para distintos horizontes temporales (1, 7 y 30 días).

El diseño metodológico es no experimental y longitudinal, dado que no se manipulan variables, sino que se observan series históricas a lo largo del tiempo. Esta estructura permite identificar patrones, estacionalidades y fluctuaciones en la variable de interés.

La recopilación de datos se llevó a cabo utilizando diversas bases de datos que entrega la Dirección Meteorológica de Chile para la estación "La Florida, La Serena Ad" la cual tiene el código de estación 290004. Los datos proporcionados por la Dirección Meteorológica de Chile, sin procesos de control de calidad, es decir, corresponden directamente a los registros capturados por los sensores. Esta característica contribuye a otorgar solidez y validez a los resultados obtenidos.

Durante el proceso de recopilación, se detectaron valores faltantes en prácticamente todas las variables (dirección del viento, velocidad del viento (en nudos), presión QFF, presión QFE, temperatura). Las presiones QFE y QFF representan, respectivamente, la presión corregida al nivel medio del mar considerando condiciones reales de temperatura, y la presión a nivel de la estación o aeródromo (Hung, 2025). Para abordar esta problemática, se implementó una estrategia de imputación de datos mediante el algoritmo k-nearest neighbors (kNN), programado en el lenguaje R (James, Witten, Hastie, y Tibshirani, 2013).

En particular, para las variables presión QFE, presión QFF (medidas en hPa) y velocidad del viento (medida en nudos), se imputaron los datos faltantes calculando el promedio de las horas más cercanas según el algoritmo kNN. En el caso de la temperatura (medida en grados Celsius), se utilizó una modificación del método: por cada valor perdido, se estimó un promedio correspondiente al mismo horario, considerando un rango de  $\pm 5$  días alrededor de la fecha faltante. Este procedimiento permitió una imputación coherente con los patrones térmicos horarios observados.

El análisis de los datos se desarrolló en dos etapas principales. Primero, se aplicó un análisis estadístico exploratorio a las series temporales para caracterizar su comportamiento. Posteriormente, se implementaron diversos modelos de predicción, que incluyeron tanto métodos clásicos como ARIMA con covariables y enfoques automatizados basados en redes convolucionales temporales. Cada modelo fue evaluado en términos de su precisión predictiva a corto, mediano y largo plazo (1, 7 y 30 días), utilizando métricas como MAE, RMSE y MAPE.

Finalmente, para asegurar una adecuada representación espacial de las condiciones del viento, se estableció una correspondencia entre estaciones meteorológicas y parques eólicos utilizando la distancia de Haversine. Según lo descrito por (Korn y Korn, 2000) en el apéndice B9 de su obra, esta función permite calcular de forma precisa distancias geodésicas a partir de coordenadas geográficas, especialmente útil en contextos de navegación y geodesia. En esta investigación, la función de Haversine se empleó para vincular cada parque eólico con la estación meteorológica más cercana, reduciendo así la variabilidad asociada a la ubicación y minimizando posibles sesgos en los datos. Esta estrategia garantiza que las condiciones meteorológicas empleadas en los modelos sean representativas del entorno inmediato de cada parque.

## Implementación Red Convolucional Temporal para predecir covariables

El objetivo de esta sección consiste en pronosticar simultáneamente las próximas 720 horas de las cinco variables meteorológicas utilizadas para predecir la velocidad del viento (dirección y rachas de viento, presión QFE, presión QFF y temperatura) mediante una TCN condicionada por su propia historia y cuatro codificadores cíclicos (seno/coseno de hora y día del año). El proceso que sigue este algoritmo completo es: datos, seguido del preprocesamiento, escalado de los datos, modelado y arquitectura red TCN, entrenamiento y exportación de resultados.

La arquitectura de esta red consiste principalmente en lo siguiente:

- Entrada: matriz de entrada de 672 observaciones y 9 variables (batch, 672, 9).
- Bloques residuales (7 en total):
  - Cada bloque contiene  $2 \times \text{Conv1D}(32, k = 3, \text{causal}, d) + \text{Dropout}(0.2)$ .
  - Dilaciones  $d \in \{1, 2, 4, 8, 16, 32, 64\}$ ; conexión residual con  $1 \times 1$  convolución si procede.
- Flatten el aplanamiento deja solo las 5 variables meteorológicas y 720 observaciones a futuro  $\rightarrow \text{Dense}(720 \times 5) \rightarrow \text{Reshape}(\text{batch}, 720, 5)$ .

Cuadro 2: Resumen de pasos de preprocesamiento

Paso	Detalle técnico
Variables	5 series físicas; 4 variables cíclicas, totalizando 9 predictores.
Marca temporal	Conversión de la variable momento a formato datetime64 (dd-mm-aaaa).
Codificación cíclica	Las variables cíclicas se definen como: $\sin\left(\frac{2\pi h}{24}\right), \cos\left(\frac{2\pi h}{24}\right), \sin\left(\frac{2\pi \text{doy}}{365}\right), \cos\left(\frac{2\pi \text{doy}}{365}\right).$
Ventanas	WINDOW de 672 horas anteriores y HORIZON de 720 horas.
Partición	80 % entrenamiento / 20 % prueba (en orden temporal).
Escalado	StandardScaler: ajustado con entrenamiento; aplicado a prueba y predicción futura.

- Función de pérdida: Huber ( $\delta = 0,25$ ). Optimizador: Adam ( $\eta = 10^{-3}$ ).
- Parámetros totales:  $\approx 7,4$  Millones.

Cabe destacar que un bloque residual consiste en un conjunto de varias capas ejecutadas en secuencia + un salto (skip-connection) que suma la entrada al final. Básicamente, su función es facilitar el flujo del gradiente (evitando el desvanecimiento o explosión del gradiente) y permitir que la red profundice sin “olvidar” la señal original.

El campo receptivo de la TCN implementada es:

$$R = (k - 1) \sum_{l=0}^6 d_l + 1 = (3 - 1)(1 + 2 + 4 + 8 + 16 + 32 + 64) + 1 \approx 1350 \text{ h} > H$$

El cual resulta suficiente para capturar periodicidad mensual.

En cuanto al entrenamiento y el test se realizarán mediante los siguientes parámetros

Cuadro 3: Hiperparámetros de entrenamiento

Parámetro	Valor
Épocas máximas	150
Tamaño de lote	32
shuffle	False (respeta el orden temporal)
Validación	20 % <i>hold-out</i> ; monitor <code>val_loss</code> .
Callback	EarlyStopping (paciencia = 10).

Cuadro 4: Hiperparámetros y coeficientes del modelo.

Símbolo / valor	Ubicación	Justificación
$k = 3$	Kernel en Conv. 1D	Captura correlaciones locales con coste bajo.
$F = 32$	Nº filtros	Capacidad suficiente sin sobre-parametrizar.
$d = 1-64$	Dilaciones	Campo receptivo $\approx 1350$ pasos.
$p = 0,2$	Dropout	Regularización ligera contra sobre-ajuste.
$\eta = 10^{-3}$	LR (Adam)	Valor estándar estable con Huber.
$\delta = 0,25$	Huber loss	Robustez a outliers tras estandarizar.

La justificación de la elección de estos parámetros consiste en que una ventana de 28 días cubre ciclos referidos a 4 semanas anteriores, el horizonte consiste en el plazo de 30 días a pronósticar considerando proponer modelos que entreguen una contribución al sector de generación de energía renovable, la incorporación de función seno y coseno para los días y las horas, contribuye a generar el ciclo que esta implícito en estas variables, que numéricamente no es posible, la elección de la función de pérdida de Huber es principalmente

h

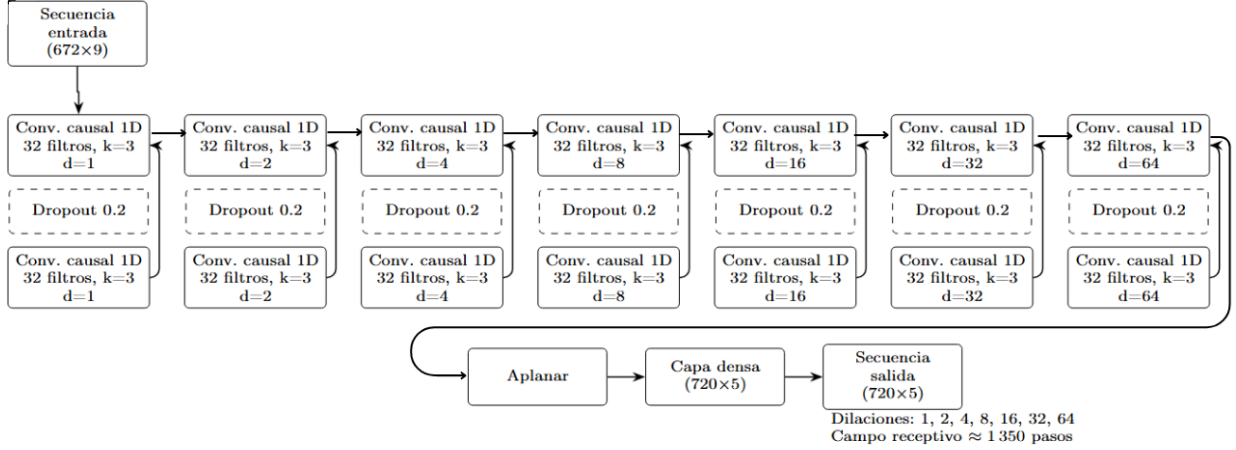


Figura 1: Diagrama Red Neuronal Convolutiva Temporal

debido a su naturaleza híbrida capacitada para reducir el impacto de los valores atípicos que pueden generar los sensores meteorológicos, el Dropout de 0.2 y Early Stopping fueron utilizados para evitar el sobre-ajuste en el entrenamiento y por último se seleccionaron dilataciones exponenciales para aumentar el campo receptivo sin implicar un coste extra alto.

Al final se reportan MAE y RMSE en escala original (grados celcius para la temperatura, hPa para la presión, dirección en grados sexagesimales).

El diagrama de la red convolutiva utilizada se ve a continuación:

En el diagrama anterior se visualiza como es el funcionamiento de la red TCN aplicada para predecir las covariables, se ve cómo la entrada, compuesta por 672 pasos temporales y 9 variables, procede a una pila de 7 bloques convolucionales organizados de forma secuencial. Cada bloque consta de una convolución causal 1D con 32 filtros, un kernel de tamaño 3 y una dilación creciente (1, 2, 4, 8, 16, 32, 64), lo que permite expandir el campo receptivo sin pérdida de resolución temporal. Además, cada bloque incorpora una capa de dropout con tasa 0.2 para evitar sobreajuste, y una conexión residual que permite el aprendizaje profundo estable. Al finalizar las convoluciones, la salida se aplanar y es pasada a una capa densa que entrega una secuencia de salida de 720 pasos y 5 variables por paso. Esta arquitectura, conocida como Temporal Convolutional Network (TCN), permite modelar dependencias a largo plazo en series temporales, alcanzando un campo receptivo aproximado de 1.350 pasos.

En resumen, lo que busca esta red es:

1. Extraer la última ventana histórica (672 h), escalar y predecir 720 valores, obteniendo finalmente una matriz de predicciones para las covariables ( $720 \times 5$ ).
2. Des-escalar y reconstruir codificadores cíclicos para el horizonte.
3. Exportar a Excel: historico, covariables para 720 horas siguientes y los escalares utilizados.

## Implementación de modelo SARIMAX para predecir la velocidad del viento

El objetivo consiste en modelar y predecir la velocidad del viento a largo plazo (30 días) mediante un modelo SARIMAX, incorporando como variables exógenas cinco covariables meteorológicas previamente pronosticadas mediante una red convolutiva temporal explicadas anteriormente, aparte de el componente estacional, la cual ayuda en la detección de patrones en un determinado tiempo

Estas covariables son las predecidas por la red TCN implementada anteriormente:

- Dirección del viento (**dd\_Valor**)
- Variabilidad del viento (**VRB\_Valor**)
- Presión a nivel de estación (**Presion\_QFE**)
- Presión reducida al nivel del mar (**Presion\_QFF**)
- Temperatura del aire (**Temperatura**)

Se espera que su inclusión permita capturar relaciones lineales entre estos fenómenos y la variable objetivo, mejorando así la capacidad predictiva del modelo que se presentara a continuación.

$$y_t = \phi^\top \mathbf{y}_{t-p} + \Phi^\top \mathbf{y}_{t-Ps} + \theta^\top \varepsilon_{t-q} + \Theta^\top \varepsilon_{t-Qs} + \beta^\top \mathbf{x}_t + \varepsilon_t \quad (1)$$

### Coefficientes del modelo

- $y_t$ : valor de la serie temporal en el tiempo  $t$ .
- $\mathbf{y}_{t-p} = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]^\top$ : rezagos no estacionales.
- $\phi$ : coeficientes autorregresivos (AR).
- $\mathbf{y}_{t-Ps} = [y_{t-s}, y_{t-2s}, \dots, y_{t-Ps}]^\top$ : rezagos estacionales.
- $\Phi$ : coeficientes autorregresivos estacionales (SAR).
- $\varepsilon_{t-q} = [\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}]^\top$ : errores pasados (no estacionales).
- $\theta$ : coeficientes de media móvil (MA).
- $\varepsilon_{t-Qs} = [\varepsilon_{t-s}, \varepsilon_{t-2s}, \dots, \varepsilon_{t-Qs}]^\top$ : errores estacionales.
- $\Theta$ : coeficientes de media móvil estacional (SMA).
- $\mathbf{x}_t$ : vector de variables exógenas en el tiempo  $t$ .
- $\beta$ : coeficientes de regresión asociados a las variables exógenas.
- $\varepsilon_t$ : término de error aleatorio (ruido blanco).

El modelo implementado corresponde a un **SARIMAX(2,0,3)(2,0,3,24)**, lo cual implica que la serie original de velocidad del viento ya estacionaria, luego modelada como una combinación lineal de sus dos retardos anteriores, tanto del apartado estacional como no estacional y con 3 retardo de errores de la misma forma y ademas del uso de las cinco covariables exógenas, para perfeccionar el modelo.

A continuación se presenta el flujo general del proceso de modelado SARIMA(2,0,3)(2,0,3,24) para la predicción de la velocidad del viento. Cada etapa corresponde a un paso clave desde la adquisición de datos históricos hasta la evaluación de las predicciones.

## Esquema general del proceso de modelado SARIMAX

### ■ Datos históricos

- Se fusionan variables meteorológicas relevantes con información de estacionalidad diaria y anual, permitiendo al modelo capturar tanto patrones físicos como cíclicos del comportamiento del viento.
- Este enfoque favorece una modelización más robusta, al contemplar tanto la variabilidad natural como las influencias atmosféricas.

### ■ Preprocesamiento

- Se utiliza una metodología automatizada para seleccionar la estructura óptima del modelo ARI-MAX, reduciendo la subjetividad y acelerando el proceso de ajuste.
- Este enfoque permite identificar de forma eficiente la combinación adecuada de parámetros que minimiza criterios de información como AIC y BIC.

### ■ Modelo basado en variables exógenas

- El uso de covariables exógenas amplía la capacidad del modelo autoregresivo tradicional, permitiendo incorporar información externa que mejora el poder predictivo.
- Esto resulta especialmente útil en entornos meteorológicos, donde variables interrelacionadas pueden anticipar comportamientos de la variable objetivo.

### ■ Validación estadística de supuestos

- Se implementan pruebas formales para evaluar si el modelo cumple con los supuestos estadísticos requeridos, como independencia y normalidad de los residuos.
- Este paso es crucial para garantizar que las inferencias derivadas del modelo sean válidas y que las predicciones no estén sesgadas por errores sistemáticos.

### ■ Preparación para predicción futura

- Una vez validado, el modelo puede ser aplicado a escenarios futuros, utilizando predicciones de las variables exógenas como insumos.
- Esto permite construir sistemas de pronóstico operativos que combinan técnicas estadísticas y modelos de aprendizaje automático.

## Implementación de red TCN para predecir la velocidad del viento

Cuadro 5: Resumen del flujo de datos y pre-procesamiento

Paso	Detalle técnico y justificación
Variables base	dd_Valor, VRB_Valor, Presion_QFE, Temperatura, Presion_QFF: descriptores físicos inmediatos.
Variables cíclicas	hora_sin, hora_cos, doy_sin, doy_cos para periodicidad diaria y anual.
Lag anual	velocidad del año anterior (rezago 8760 horas (1 año)) para memoria estacional (Bai, Koltun, y Koltun, 2018).
Ventanas	Ventana de entrada $W = 720$ h y horizonte $H = 720$ h (30 d).
Escalado	StandardScaler entrenado sólo en el conjunto <i>train</i> .
División temporal	80 % entrenamiento / 20 % prueba, manteniendo orden cronológico.

Las secuencias se generan con un barrido deslizante:

$$\underbrace{x_{t-719}, \dots, x_t}_{\text{ventana}} \longrightarrow \underbrace{y_{t+1}, \dots, y_{t+720}}_{\text{horizonte}}$$



## Arquitectura de la red TCN

- Seis bloques TCN con filtros  $F = 32$ , kernel  $k = 3$  y dilaciones  $d = \{1, 2, 4, 8, 16, 32\}$ .
- Cada bloque: Conv1D  $\rightarrow$  Dropout(0.2)  $\rightarrow$  Conv1D  $\rightarrow$  Dropout(0.2)  $\rightarrow$  residual con función de activación ReLU.
- Causalidad estricta (padding=causal) evita data leakage (fuga de datos).
- Flatten  $\rightarrow$  Dense( $H$ )  $\rightarrow$  Reshape( $H, 1$ ) produce pronóstico multistep directo.
- aproximadamente 16.625.000 parámetros entrenables.

## Entrenamiento

- Pérdida Huber ( $\delta = 0,25$ ), optimizador Adam (lr= $10^{-3}$ ).
- Tamaño de batch de 256 (número de ejemplos que se introducen en la red para que entrene en cada epoch) *shuffle* desactivado para mantener estructura temporal.
- 150 épocas máx. con ".EarlyStopping"(paciencia=10) y restauración del mejor modelo.
- Semillas fijadas (Python, NumPy, TensorFlow) y determinismo activado para reproducibilidad.

## Predicción y exportación

1. Se extrae la última ventana histórica hasta el **23 jun. 2025 08:00** CLT y se escala.
2. El modelo produce  $\hat{y}_{1:H}$ ; se des-escala y se aplica  $\max(0, \cdot)$ .
3. Se genera un índice horario del **23 jun. 2025 09:00** al **23 jul. 2025 08:00**.
4. Se escriben tres hojas Excel: histórico, pronóstico y parámetros del escalador.

## Adecuación de la arquitectura al problema

- Convoluciones causales con dilaciones cubren dependencias de largo plazo con complejidad.
- Residuales preservan gradientes; Dropout de 0.2 actúa como regularizador bayesiano.
- Multistep directo evita acumulación de error propia de esquemas autoregresivos.
- El rezago anual explícito aporta información estacional sin ampliar innecesariamente el campo receptivo.
- Código y complejidad (cantidad de parámetros) compatible con entorno "GPU T4" gratuito en Google Colab.

El diagrama de como se implementa esta red TCN es:

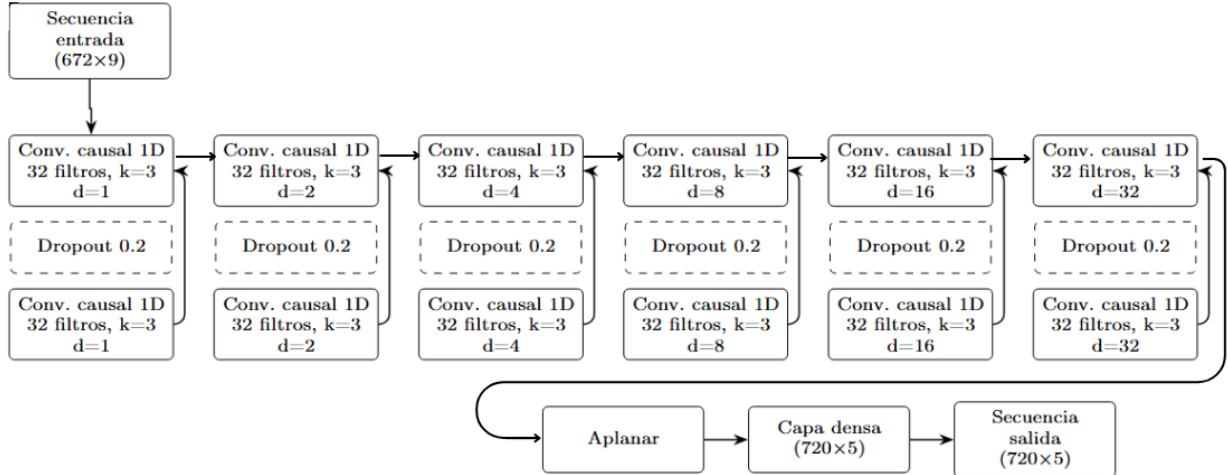


Figura 2: Diagrama Red TCN para velocidad del viento

En el diagrama anterior podemos ver cómo funciona en resumen la TCN que busca ser implementada de manera resumida. La entrada es una secuencia de 672 pasos temporales con 9 variables. Esta secuencia pasa por una serie de bloques convolucionales causales 1D, cada uno con 32 filtros, un tamaño de kernel 3, y dilataciones que van duplicándose en cada bloque ( $d=1, 2, 4, 8, 16, 32$ ), lo que permite capturar dependencias a largo plazo en la serie temporal.

Cada bloque convolucional está compuesto por dos capas de convolución causal 1D con un dropout de 0.2 entre ellas para evitar sobreajuste. Las conexiones de salto (skip connections) también se aprecian, mejorando el flujo de información y la capacidad de aprendizaje.

Finalmente, la salida de la última capa convolucional se aplanar y pasa por una capa densa, que produce una secuencia de salida de tamaño 720 con 5 variables predichas. Este diseño permite modelar series temporales con dependencia de largo alcance y producir predicciones secuenciales multivariadas.

## Resultados

### Análisis exploratorio

El análisis exploratorio de datos se realizará sobre la base datos de la estación meteorológica nacional de código 290004, esta estación registra variables climáticas a lo largo del tiempo, lo que permite estudiar patrones como la variación de temperatura, presión atmosférica, dirección y velocidad del viento, entre otros.

El EDA tiene como objetivo:

1. Comprender la estructura general de los datos
2. Evaluar la distribución de variables relevantes
3. Identificar posibles inconsistencias o valores nulos
4. Visualizar comportamientos temporales y estacionales
5. Explorar relaciones entre variables que puedan alimentar futuros modelos

Esta parte del estudio cumple un rol fundamental como paso previo a la construcción de un modelo SARIMAX, el cual se utiliza para predecir series temporales considerando el efecto de variables exógenas. Por ende, el EDA no solo busca caracterizar los datos, sino también identificar que variables podrían ser útiles en el modelamiento predictivo, asegurando que cumplan con los supuestos necesarios y aporten información significativa. Para empezar se definen las variables brevemente.

Variable	Descripción
momento	Fecha y hora del registro
dd_Valor	Dirección del viento en grados (0–360°)
ff_Valor	Velocidad del viento promedio (m/s)
VRB_Valor	Indicador de dirección de viento indefinida (1: sí, 0: no)
Presion_QFE	Presión atmosférica en superficie (hPa)
Presion_QFF	Presión atmosférica reducida al nivel medio del mar (hPa)
Temperatura	Temperatura del aire en grados Celsius (°C)
Año	Año en que se realizó el registro
Estacion	Código de la estación meteorológica
CodigoNacional	Código nacional del punto de observación

Cuadro 6: Descripción de variables contenidas en la base de datos de la estación 290004

La base de datos posee 39.731 registros y 10 variables, todas estas variables tiene el tipo de datos adecuado y no hay problemas de estructura que impidan el análisis.

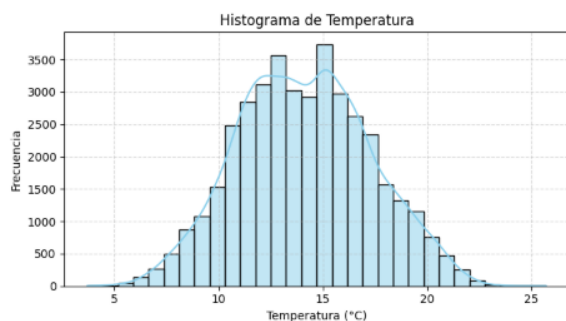
Se realizó un resumen estadístico de las variables y pudimos ver que la temperatura tiene una media 14,1°C con un rango de 3,7°C a 25,7°C con variabilidad estacional evidente. Las presiones QFE y QFF presentan una baja dispersión, lo que sugiere condiciones atmosféricas estables. Las velocidades del viento se encuentran entre 0 y 21 m/s, con media cercana a 6 m/s. Por ende, se observan eventos de viento fuerte. También la dirección del viento abarca todo el rango 0° a 360°, concentrada en el suroeste, es 216° en promedio.

También se comprueban valores nulos y debido al análisis previo no existen, por lo tanto no es necesario hacer limpieza por omisión o imputación. Esto nos quiere decir que la base es apta para continuar con visualizaciones y modelos.

Una vez realizadas estas comprobaciones y observaciones previas, se procede a realizar un análisis univariado, que contribuye examinando el comportamiento individual de cada variable meteorológica registrada en la estación 290004. Este análisis proporciona información importante sobre la distribución, centralización y dispersión de cada variable, lo que permite una evaluación inicial de la calidad y el comportamiento de los datos.

En este caso, como se observa en el diccionario de variables existen seis variables que analizar específicamente, que son: temperatura, presión atmosférica QFE y QFF, velocidad del viento, dirección del viento y el indicador de dirección del viento indefinido.

Utilizando histogramas individuales para cada variable, se evalúa la frecuencia de cada variable para identificar asimetrías, concentraciones de valores y posibles valores atípicos. Este paso es esencial antes de examinar relaciones más complejas entre variables.



La temperatura se asemeja a una distribución normal con una ligera asimetría hacia la derecha. El rango de temperaturas va desde los 3,7°C hasta aproximadamente 25,7°C. La mayoría de los valores están concentrados entre 12°C y 16°C. Esto indica un clima moderado predominante en la estación analizada, sin temperaturas extremas.

Figura 3: Distribución univariada de la variable Temperatura.

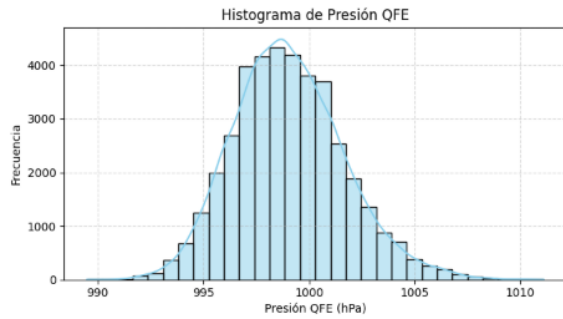


Figura 4: Distribución univariada de la variable Presión QFE.

La presión QFE se concentra fuertemente entre 996 y 1002 hPa, con una leve asimetría negativa hacia la izquierda. Esto sugiere que la presión atmosférica en esta estación suele mantenerse estable, con pequeñas variaciones, lo que puede estar asociado a condiciones meteorológicas relativamente constantes.

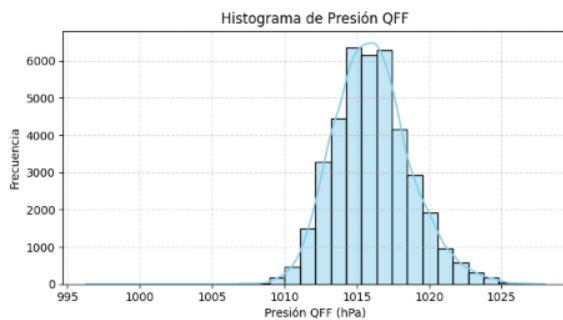


Figura 5: Distribución univariada de la variable Presión QFF.

La presión QFF tiene una distribución un poco más simétrica que QFE, con un claro "peak" alrededor de los 1016 hPa. Esta variable representa la presión corregida a nivel del mar, por lo que la estabilidad en sus valores refuerza la idea de que esta región no presenta eventos extremos frecuentes a nivel atmosférico.

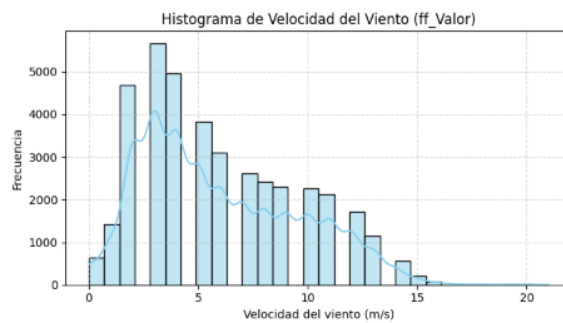


Figura 6: Distribución univariada de la velocidad del viento.

La mayoría de las observaciones de velocidad del viento se sitúan entre 0 y 10 m/s con una caída rápida a medida que aumenta la velocidad. Hay eventos de hasta 21 m/s pero son poco frecuentes. Esto indica que el viento suele ser moderado con algunas ráfagas ocasionales más intensas.

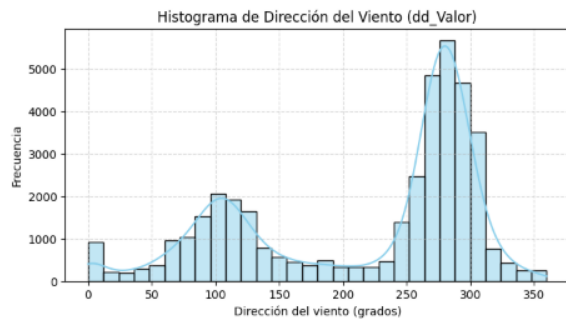


Figura 7: Distribución univariada de la dirección del viento.

La dirección del viento está distribuida entre  $0^\circ$  y  $360^\circ$  aunque muestra mayor frecuencia en el rango entre  $200^\circ$  y  $300^\circ$ , esto indica que los vientos predominantes provienen del suroeste. Dicha observación puede estar relacionada con características geográficas o patrones climáticos típicos de la zona.

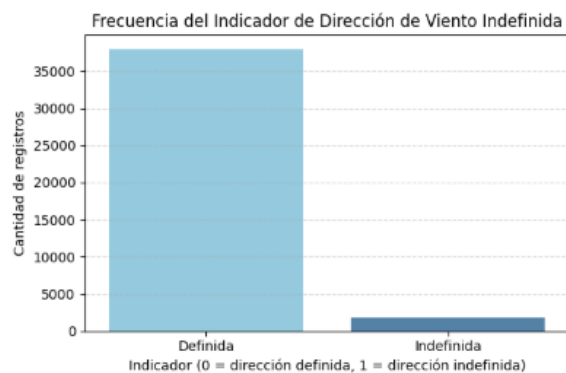


Figura 8: Frecuencia del indicador de dirección de viento indefinida.

La gran mayoría de las observaciones están definidas, esto indica que el viento tiene una dirección definida en la mayor parte del tiempo. Y solo una fracción muy pequeña de los registros tiene una dirección indefinida, por ende, solo en ocasiones el viento presentaba una dirección indefinida o cambiante. Este fenómeno puede asociarse a condiciones atmosféricas inestables, turbulencias locales o velocidades de viento muy bajas que impiden detectar una dirección dominante.

Para el análisis bivariado, se empleara principalmente gráficos de dispersión temporal, líneas de tendencia y boxplots para observar la evolución y variabilidad de las variables meteorológicas a lo largo del tiempo. Se evaluaron las relaciones entre presión QFE, presión QFF, velocidad del viento, dirección del viento y su variación temporal, tanto mensual y anual. Este análisis proporciona evidencia visual de comportamientos estacionales, anomalías climáticas o patrones que podrían ser útiles para modelamientos posteriores.

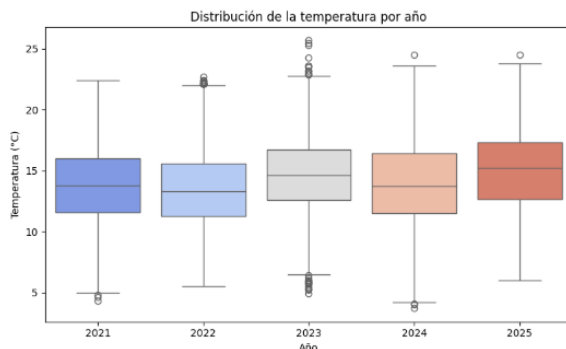


Figura 9: Distribución de la temperatura por año.

Como se puede ver existe una leve tendencia creciente en la mediana de la temperatura a lo largo de los años. También se observan valores atípicos que podrían indicar eventos de calor extremo o registros puntuales.

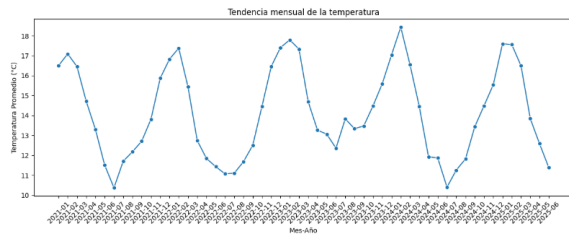


Figura 10: Distribución de la temperatura por meses.

Con el análisis mensual, se observa una variación estacional clara con subidas y bajadas periódicas, lo que refleja ciclos anuales de temperatura. Además, se puede identificar una tendencia ascendente a largo plazo, especialmente en los años más recientes.

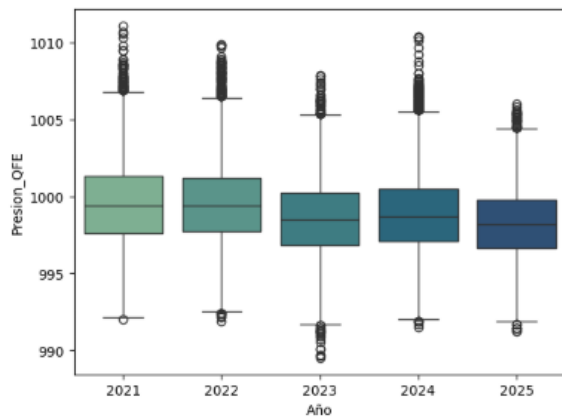


Figura 11: Distribución de la Presión QFE por años.

Este gráfico muestra que la presión QFE sigue una distribución relativamente estable a lo largo de los años, con ligeras variaciones en los valores atípicos. No se observan tendencias marcadas, lo que sugiere condiciones barométricas similares entre años.

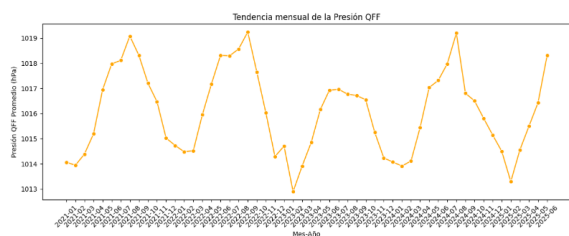
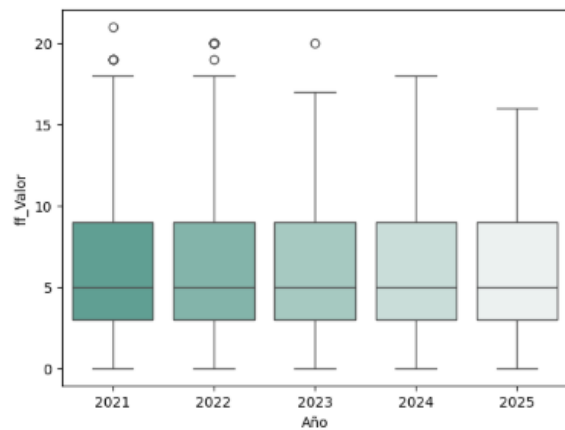


Figura 12: Distribución de la presión QFF por meses.

La presión QFF tiene oscilaciones a lo largo de los meses, pero en general presenta una tendencia estable. Las caídas o alzas puntuales pueden deberse a fenómenos meteorológicos específicos.



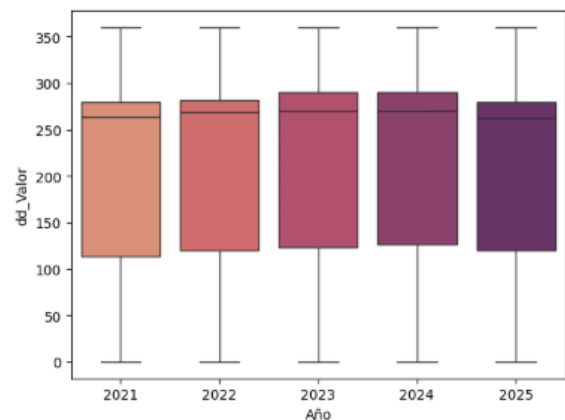
La velocidad del viento tiene una distribución relativamente homogénea entre los años. Se aprecian valores atípicos, especialmente en los primeros años analizados, lo cual puede estar asociado a eventos de viento fuertes o extremos.

Figura 13: Distribución de la velocidad del viento por año.



Se observa un patrón estacional en la velocidad del viento que tiende a aumentar en ciertos meses del año, lo que sugiere la influencia de fenómenos climáticos cíclicos.

Figura 14: Distribución de la velocidad del viento por meses.

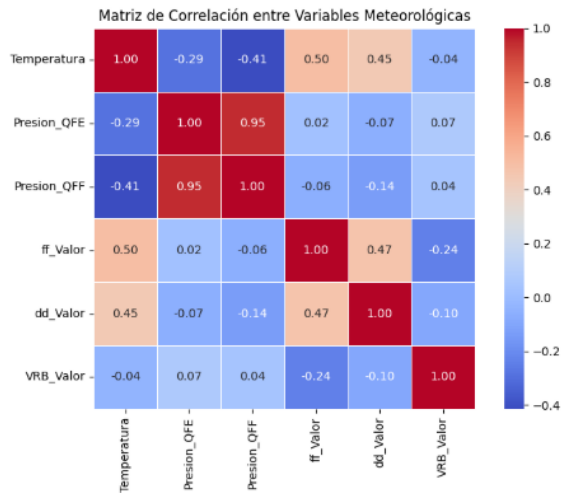


La dirección del viento muestra una gran dispersión durante todos los años, lo cual es esperable ya que el viento puede provenir de cualquier dirección. No se identifica una dirección dominante en este nivel de análisis.

Figura 15: Distribución de la dirección del viento por año.

Para terminar el análisis exploratorio se muestra una matriz de correlación para evaluar la relación lineal entre las variables numéricas disponibles, permitiendo identificar posibles asociaciones significativas que podrían influir en el comportamiento de la variable objetivo. Comprender estas relaciones es fundamental tanto para el análisis exploratorio como para etapas posteriores de modelamiento.

En específico, este análisis resulta útil al momento de seleccionar variables exógenas relevantes que podrían incorporarse en modelos de predicción multivariados como SARIMAX. A través de la matriz de correlación y representaciones visuales complementarias, se explorará el grado y la dirección de asociación entre variables como la temperatura, presión atmosférica, velocidad y dirección del viento.



Existe una alta correlación positiva entre Presión QFE y Presión QFF, lo cual es esperable ya que ambas representan tipos de presión. La temperatura muestra baja correlación con otras variables, lo que sugiere que podría tener un comportamiento más independiente en el modelo. Con respecto al indicador de dirección de viento indefinida, al ser una variable indicadora binaria, tiene correlaciones muy bajas, como es habitual en variables categóricas binarias.

Figura 16: Matriz de Correlación entre Variables Meteorológicas.

Como conclusión de este análisis exploratorio, este permite comprender en profundidad la estructura, comportamiento y relaciones de las variables registradas. Con el análisis univariado, se identifican las posibles distribuciones, rangos y posibles outliers para variables como temperatura, presión, dirección y velocidad del viento. En el análisis bivariado, se examinaron relaciones entre pares de variables, así como su evolución temporal, permitiendo detectar estacionalidades y patrones relevantes. Posteriormente, el análisis de correlación entregó evidencia cuantitativa sobre las asociaciones lineales entre las variables. Este paso fue especialmente útil para orientar la selección de variables exógenas en el modelo SARIMAX, descartando aquellas con redundancia o escasa relación con la variable objetivo. También se identificaron combinaciones altamente correlacionadas que podrían simplificarse sin pérdida de información, lo que favorece la eficiencia y estabilidad del modelo.



## Predicciones de covariables mediante red TCN implementada

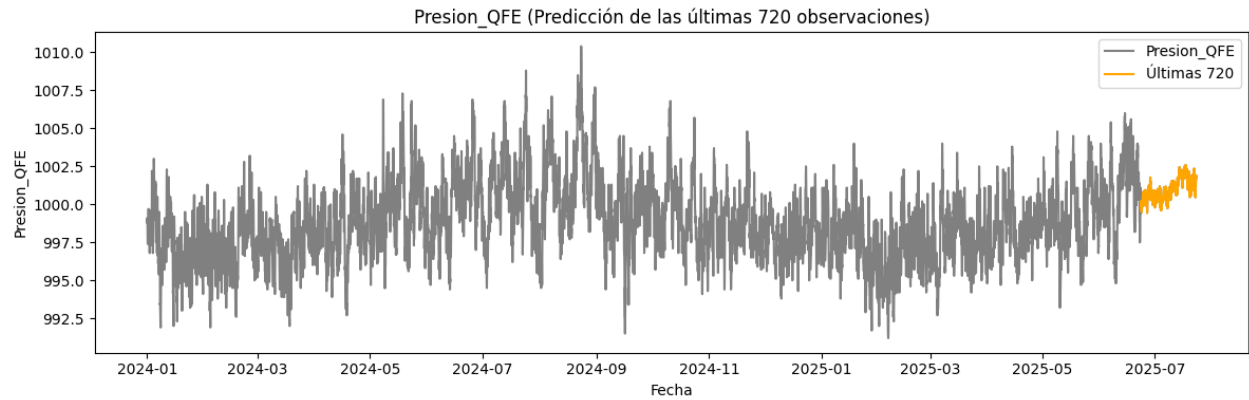


Figura 17: Predicción para la Presión QFE

En la Figura 17 se puede ver la predicción para la presión QFE, la cual será utilizada para la posterior predicción de la velocidad del viento.

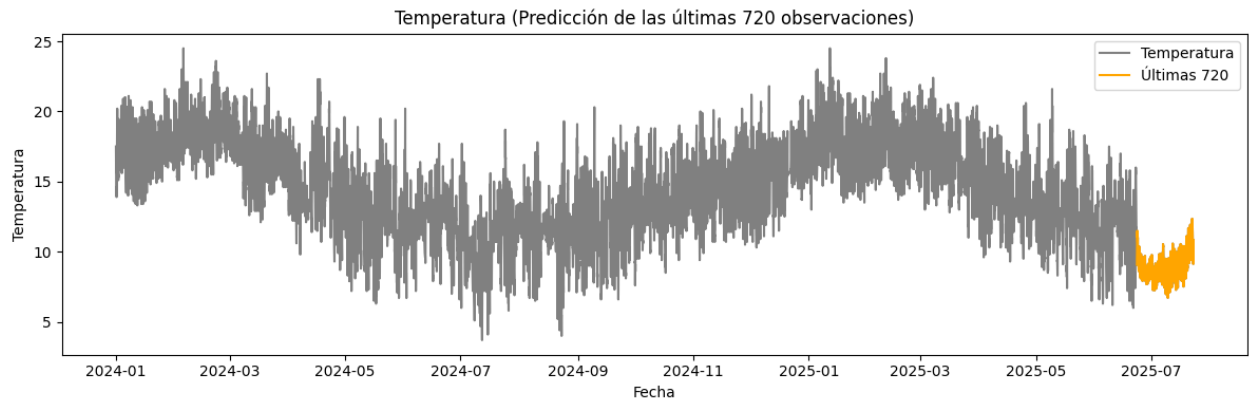


Figura 18: Predicción Temperatura

En la Figura 18 se puede ver la predicción para la temperatura, la cual será utilizada para la posterior predicción de la velocidad del viento.

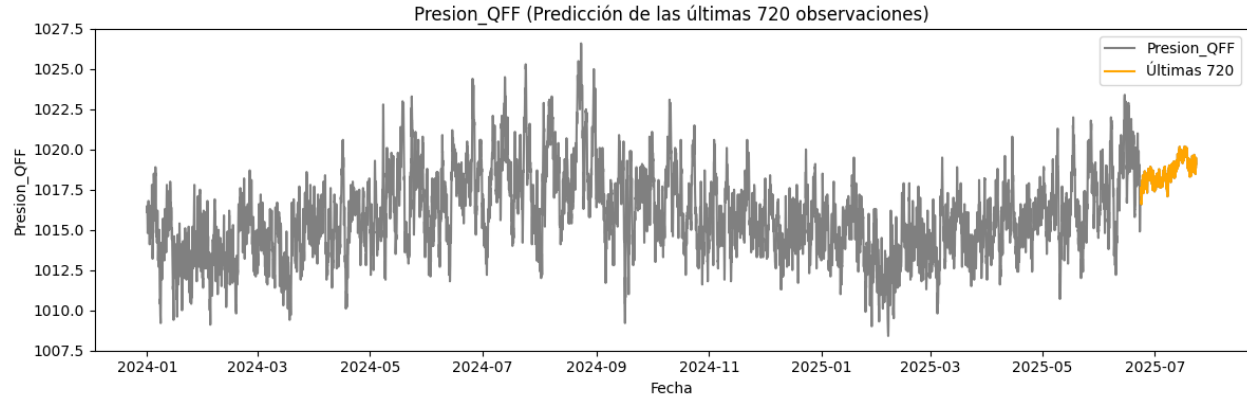


Figura 19: Predicción Presión QFF

En la Figura 19 se puede ver la predicción para la presión QFF, la cual será utilizada para la posterior predicción de la velocidad del viento.

## Modelo Sarima

### Resultados del Modelo SARIMAX(2,0,3)(2,0,3)<sub>24</sub>

Se ajustó un modelo SARIMAX(2,0,3)(2,0,3)<sub>24</sub> para modelar y predecir la variable dependiente  $ff\_Valor$ , correspondiente a la velocidad del viento. El modelo incorpora tanto componentes autorregresivos y de medias móviles no estacionales como estacionales con periodicidad 24 haciendo referencia a ciclos diarios, además de múltiples variables exógenas.

#### 0.0.1. Evaluación del Modelo SARIMAX

El modelo SARIMAX ajustado fue evaluado tanto en términos de su capacidad de ajuste como de su desempeño predictivo. Desde el punto de vista del ajuste, el modelo presentó una log-verosimilitud de  $-6137,32$ , un criterio de información de Akaike (AIC) de  $12318,64$  y un criterio bayesiano de Schwarz (BIC) de  $12450,78$ . Estas métricas indican un ajuste razonablemente bueno considerando la complejidad del modelo y la inclusión de términos estacionales y de medias móviles. Específicamente, valores relativamente bajos de AIC y BIC reflejan una adecuada relación entre la bondad de ajuste y la penalización por complejidad, lo cual sugiere que el modelo logra capturar eficientemente la estructura de la serie sin sobreajustarla.

Desde la perspectiva predictiva, se calculan dos métricas ampliamente utilizadas: el Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE). En este caso, el modelo obtuvo un RMSE de  $1,8710$  y un MAE de  $1,4871$ .

El MAE representa el error medio absoluto entre las predicciones del modelo y los valores observados, permitiendo una interpretación directa en las mismas unidades que la serie original. Un MAE de  $1,4871$  implica que, en promedio, el modelo se equivoca por menos de dos unidades en cada observación, lo cual es una señal de precisión aceptable para una serie con alta variabilidad horaria.

Por su parte, el RMSE penaliza con mayor severidad los errores grandes al elevar al cuadrado las desviaciones antes de promediarlas. Un valor de  $1,8710$  cercano al MAE sugiere que no existen errores extremos significativamente altos, lo cual indica un comportamiento consistente del modelo a lo largo del tiempo.

En conjunto, estas métricas muestran que el modelo SARIMAX no solo logra un ajuste estructural adecuado a los datos históricos, sino que también es capaz de generar predicciones con un nivel de error aceptablemente bajo. Esto refuerza la validez del modelo como herramienta para el análisis y pronóstico de la serie temporal considerada.

#### Estimaciones de los coeficientes del modelo SARIMAX(2,0,3)

Cuadro 7: Coeficientes estimados del modelo SARIMAX(2,0,3)(2,0,3)<sub>24</sub>

Parámetro	Coef.	Error Std.	z	p-valor	[0.025, 0.975]
dd_Valor	0.0035	0.000	8.185	0.000	[0.003, 0.004]
VRB_Valor	-3.5326	0.194	-18.171	0.000	[-3.914, -3.152]
Presion_QFE	-0.1618	0.070	-2.299	0.021	[-0.300, -0.024]
Temperatura	0.3163	0.035	9.007	0.000	[0.247, 0.385]
Presion_QFF	0.1598	0.069	2.314	0.021	[0.024, 0.295]
hora	0.0683	0.016	4.331	0.000	[0.037, 0.099]
doy	-0.0079	0.048	-0.162	0.871	[-0.103, 0.087]
hora_sin	-1.5057	0.198	-7.620	0.000	[-1.893, -1.118]
hora_cos	0.3742	0.148	2.523	0.012	[0.084, 0.665]
doy_sin	-0.5661	1.233	-0.459	0.646	[-2.983, 1.851]
doy_cos	-1.0389	2.995	-0.347	0.729	[-6.909, 4.831]
ar.L1	0.3393	2.396	0.142	0.887	[-4.357, 5.036]
ar.L2	0.0498	1.191	0.042	0.967	[-2.285, 2.384]
ma.L1	0.1033	2.396	0.043	0.966	[-4.592, 4.799]
ma.L2	0.0323	0.149	0.217	0.828	[-0.260, 0.324]
ma.L3	0.0175	0.047	0.376	0.707	[-0.074, 0.109]
ar.S.L24	0.4286	4.458	0.096	0.923	[-8.309, 9.166]
ar.S.L48	0.1924	3.074	0.063	0.950	[-5.833, 6.218]
ma.S.L24	-0.2921	4.458	-0.066	0.948	[-9.030, 8.446]
ma.S.L48	-0.1209	2.467	-0.049	0.961	[-4.957, 4.715]
ma.S.L72	0.0038	0.160	0.023	0.981	[-0.310, 0.318]
sigma2	3.6560	0.100	36.602	0.000	[3.460, 3.852]

A continuación, se interpretan los principales coeficientes del modelo, enfocándonos en aquellas variables exógenas que resultaron estadísticamente significativas ( $p < 0,05$ ):

- **dd\_Valor (Dirección del viento)** muestra una relación positiva y significativa con la velocidad del viento (*ff\_Valor*). Esto sugiere que a medida que cambia la dirección del viento (medida en grados), la velocidad tiende a aumentar levemente, posiblemente asociada a la entrada de vientos predominantes más intensos en ciertas direcciones.
- **VRB\_Valor (Viento variable)** presenta un coeficiente negativo y altamente significativo. Esto indica que cuando se registra una condición de viento variable (sin dirección fija), la velocidad del viento tiende a ser menor. Es coherente con la naturaleza de estos vientos, que suelen ser suaves e irregulares.
- **Presion\_QFE (Presión atmosférica al nivel de la estación)** tiene un efecto negativo y significativo sobre la velocidad del viento. Esto implica que, a mayor presión medida localmente, la velocidad del viento disminuye, lo cual puede reflejar situaciones de estabilidad atmosférica.
- **Presion\_QFF (Presión atmosférica reducida al nivel del mar)** muestra un efecto positivo y también significativo. Esto sugiere que cuando aumenta la presión corregida al nivel del mar, la velocidad del viento también tiende a incrementarse, posiblemente asociada con sistemas meteorológicos de gran escala.
- **Temperatura (°C)** influye positivamente en la velocidad del viento. Es decir, a medida que aumenta la temperatura, también lo hace la velocidad del viento. Este efecto podría deberse a la formación de gradientes térmicos que favorecen el movimiento del aire.
- **hora (hora del día)** tiene un coeficiente positivo y significativo, lo que indica que el viento tiende a ser más intenso en ciertas horas del día, particularmente durante las horas centrales, debido a la convección y otros factores térmicos diurnos.

- **hora\_sin (Componente senoidal de la hora del día)** presenta un efecto negativo considerable y altamente significativo, lo que refleja una dinámica cíclica diaria donde la velocidad del viento tiende a disminuir en determinados momentos del ciclo horario.
- **hora\_cos (Componente cosenoidal de la hora del día)** también es significativa y positiva, lo que refuerza la idea de que el viento tiene un comportamiento periódico a lo largo del día, influenciado por factores ambientales regulares como la radiación solar y la estabilidad atmosférica.

**Variables exógenas no significativas ( $p > 0,05$ ):**

- *doy* (día del año), *doy\_sin* y *doy\_cos* no presentaron significancia estadística, lo que sugiere que los patrones anuales no tienen un efecto claro en la velocidad del viento en esta muestra.

**Componentes ARIMA:**

- Ninguno de los coeficientes autorregresivos (AR) ni de medias móviles (MA), tanto estacionales como no estacionales, resultaron estadísticamente significativos ( $p > 0,05$ ), lo que sugiere que gran parte de la estructura temporal está explicada por las variables exógenas.

**Varianza del error:** El valor estimado de  $\sigma^2$  fue de 3.656, con alta significancia estadística ( $p < 0,001$ ), lo que representa la varianza del término de error blanco.

Resultando así, un modelo sarimax con los siguientes componentes.

$$\begin{aligned}
y_t = & 0,3393 y_{t-1} + 0,0498 y_{t-2} + 0,4286 y_{t-24} + 0,1924 y_{t-48} \\
& + 0,1033 \varepsilon_{t-1} + 0,0323 \varepsilon_{t-2} + 0,0175 \varepsilon_{t-3} \\
& - 0,2921 \varepsilon_{t-24} - 0,1209 \varepsilon_{t-48} + 0,0038 \varepsilon_{t-72} \\
& + \begin{bmatrix} 0,0035 \\ -3,5326 \\ -0,1618 \\ 0,3163 \\ 0,1598 \\ 0,0683 \end{bmatrix}^T \begin{bmatrix} \text{dd\_Valor}_t \\ \text{VRB\_Valor}_t \\ \text{Presion\_QFE}_t \\ \text{Temperatura}_t \\ \text{Presion\_QFF}_t \\ \text{Hora}_t \end{bmatrix} + \varepsilon_t
\end{aligned} \tag{2}$$

## Diagnóstico del Modelo

Los resultados de las pruebas estadísticas aplicadas a los residuos muestran:

- **Prueba de Ljung-Box:** No se detecta autocorrelación en los residuos ( $Q(1) = 0.12$ ,  $p = 0,73$ ), lo cual indica que el modelo capta adecuadamente la dependencia temporal de la serie.
- **Prueba de Jarque-Bera:** Sugiere que los residuos no siguen una distribución normal perfecta ( $p < 0,01$ ).
- **Heterocedasticidad:** Se detecta presencia de heterocedasticidad significativa ( $H = 1.33$ ,  $p < 0,01$ ), lo que podría indicar variabilidad no constante en los errores del modelo.

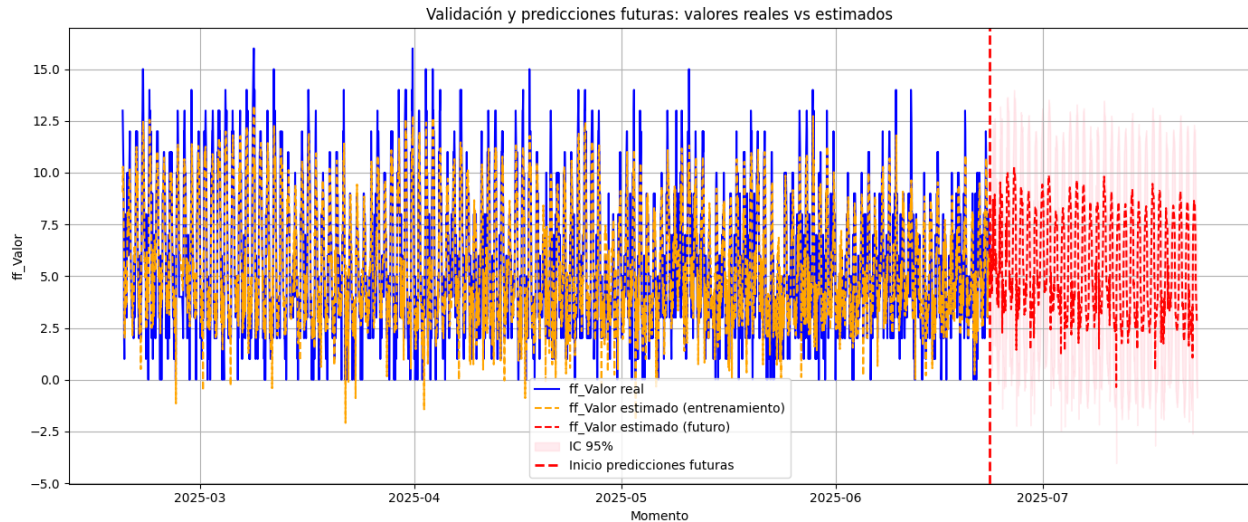


Figura 20: modelo y predicciones

**Evolución del Modelo y Proyecciones** El gráfico anterior muestra cómo el modelo SARIMAX ha sido entrenado para ajustarse al comportamiento de la serie temporal. A lo largo del periodo de entrenamiento, es posible observar que el modelo sigue de manera bastante cercana las fluctuaciones de los valores reales, lo que sugiere que logra capturar las dinámicas principales presentes en los datos. Esta etapa es fundamental, ya que permite calibrar el modelo y preparar las condiciones necesarias para realizar predicciones confiables.

Una vez finalizado el entrenamiento, y marcado por la línea vertical roja, comienzan las proyecciones hacia el futuro. En este caso, se presentan las predicciones para el mes de junio del año 2025. Estas estimaciones están acompañadas de bandas que representan el intervalo de confianza al 95 %, el cual indica el rango dentro del cual se espera que se encuentren los valores reales con alta probabilidad.

Desde el punto de vista visual, se observa que las predicciones futuras no presentan cambios abruptos ni comportamientos inesperados, sino que mantienen una línea coherente con la evolución previa de la serie. Además, el intervalo de confianza, aunque permite cierto margen de variación, se mantiene relativamente estable en su amplitud, lo que sugiere que el modelo no anticipa una mayor incertidumbre en el corto plazo. En conjunto, esta visualización permite evaluar tanto la calidad del ajuste como el grado de confianza asociado a las proyecciones realizadas por el modelo.

Para las predicciones a corto, mediano y largo plazo, el modelo SARIMA(2,0,3)(2,0,3,24) implementado arroja estos resultados:

Para el día 24 de junio a las 8:00 A.M. el modelo SARIMA estima que la velocidad será de 4.49 nudos. Para el día 30 de junio a la misma hora, se espera que la velocidad del viento sea de 3.78 nudos. Finalmente, para el día 23 de julio de 2025 a las 8:00 A.M., el modelo predice que la velocidad del viento será de 2.87 nudos aproximadamente.

## Predicción para el viento utilizando red TCN

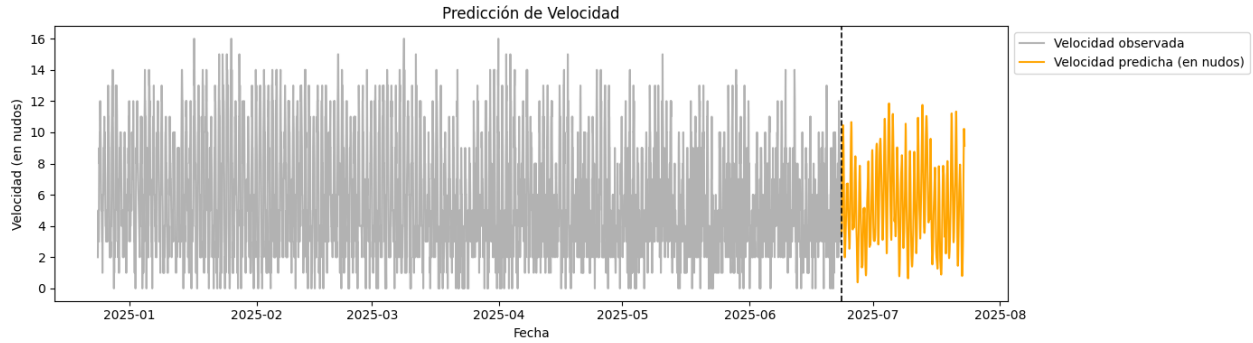


Figura 21: Predicción para la velocidad mediante TCN

En la Figura 21 se puede observar como la predicción del viento tiende a tomar valores cercanos a lo que se visualiza en la parte anterior, se puede observar que la variabilidad de la predicción sigue la tendencia de las observaciones hasta la fecha.

Las métricas de la variabilidad de los resultados obtenidos por la red TCN implementada son:

Cuadro 8: Métricas del modelo	
Métrica	Valor
MAE (Testeo)	0.302
RMSE (Testeo)	0.527
MAE (Entrenamiento)	0.307
RMSE (Entrenamiento)	0.529
Desviación estándar verdadera	2.9894
Desviación estándar predicciones	2.9405

Se observa que el promedio de las diferencias absolutas entre los valores predichos y los valores reales (MAE) es de 0.302 nudos. Mientras que para el conjunto de entrenamiento el promedio de las diferencias es de 0.307 nudos lo que indica que el modelo aprende de manera satisfactoria y entrega predicciones cercanas a la realidad.

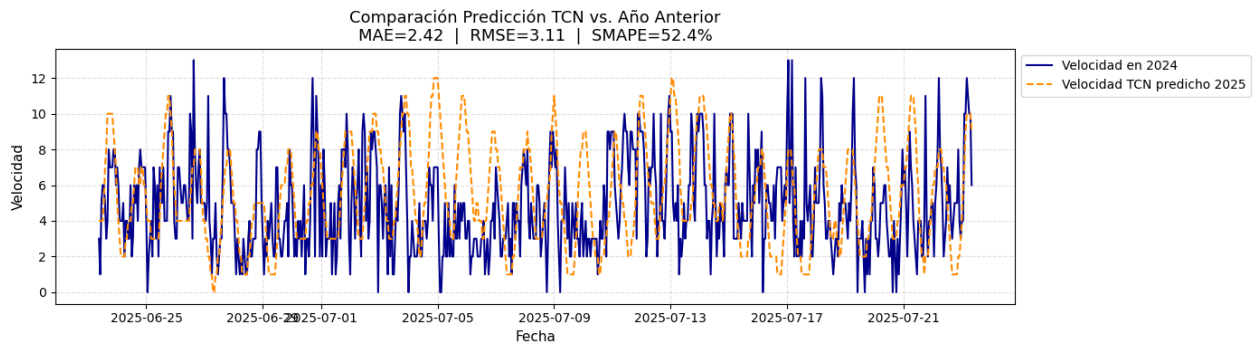


Figura 22: Comparación de la predicción con sus respectiva fecha en el año anterior

Se puede visualizar que los resultados de 2025 entregan valores que tienen una variabilidad similar a la

variabilidad presentada en el mismo periodo de tiempo el año anterior. Con esto se confirma en cierta medida que el modelo se ajusta correctamente en el contexto.

A partir de la predicción con la red TCN se obtiene que para las siguientes 24 horas posterior a la última observación recolectada se espera una velocidad de 3.23 nudos específicamente a las 8:00 A.M. del día 24 de junio de 2025, mientras que para la semana próxima se espera una velocidad de 5.143 nudos a la misma hora del día 30 de junio. Finalmente, para la observación del próximo mes, la red TCN prevee que la velocidad del viento será de 8.42 nudos a las 8:00 A.M del 23 de julio de 2025.

## Discusión y conclusiones

Los resultados obtenidos en esta investigación, donde se evaluaron dos enfoques para la predicción de la velocidad del viento en La Serena, Chile, durante el periodo del 23 de junio al 23 de julio de 2025, muestran una concordancia y complementariedad con la literatura actual en el campo del pronóstico eólico en Sudamérica.

Por un lado, el modelo SARIMAX(2,0,3)(2,0,3,24), que incorpora tanto componentes autorregresivos y de medias móviles como términos estacionales y variables exógenas, evidenció un ajuste estructural adecuado a los datos históricos, con métricas como RMSE y MAE en rangos aceptables (RMSE aproximado 1.87 y MAE 1.49). Esto es consistente con los resultados reportados por Lazzús y Salfate (2017) y Costa y cols. (2021), quienes también enfatizan que los modelos ARIMA y sus extensiones estacionales son un punto de partida robusto para modelar series temporales meteorológicas, especialmente cuando se incluyen variables exógenas relevantes.

No obstante, como se observa en la literatura, los modelos lineales como SARIMAX suelen presentar limitaciones al capturar la complejidad y no linealidad inherentes a la dinámica atmosférica, principalmente en presencia de fenómenos como las Capas Límite Atmosféricas Estables (ACLE) y forzantes sinópticos variables (Maroneze y cols., 2023). En este sentido, la implementación de redes neuronales convolucionales temporales (TCN) para modelar la velocidad del viento y sus covariables ofreció un desempeño predictivo competitivo y, en algunos aspectos, superior al modelo clásico. Los resultados de la TCN mostraron un MAE cercano a 0.3 y RMSE alrededor de 0.52 en datos de test, cifras que sugieren una mayor precisión en la captura de la variabilidad diaria y las tendencias temporales, alineándose con los avances reportados por Bastos y cols. (2021b) y Almeida y cols. (2021) en el uso de arquitecturas profundas para la predicción eólica.

Además, la capacidad de la TCN para incorporar la dependencia temporal a largo plazo y manejar múltiples variables simultáneamente representa una ventaja frente a los modelos lineales, los cuales pueden no captar eficientemente interacciones complejas entre variables meteorológicas (Goulart y de Camargo, 2021). La comparación de las predicciones con las observaciones del año anterior (Figura 22) y la evaluación de métricas indican que la TCN logra mantener una coherencia y estabilidad en la variabilidad de sus proyecciones, lo que es crucial para aplicaciones operativas y planificación energética.

Sin embargo, es importante destacar que, tal como lo evidencian varios autores, la robustez de los modelos de predicción está altamente influenciada por la calidad, cantidad y representatividad de los datos de entrenamiento (Miguel y cols., 2019). En esta investigación, la disponibilidad de datos meteorológicos históricos y la incorporación de covariables relevantes permitieron un entrenamiento adecuado, pero la validación futura en condiciones reales será clave para confirmar la aplicabilidad del modelo TCN en entornos locales con topografía compleja y variabilidad sinóptica marcada.

Finalmente, la combinación de enfoques, como SARIMAX con técnicas de aprendizaje profundo representa una línea prometedora para mejorar la precisión y confiabilidad del pronóstico de viento, particularmente en regiones costeras y montañosas como La Serena. La integración de variables exógenas derivadas de modelos numéricos atmosféricos de alta resolución y la aplicación de técnicas de corrección de sesgo podrían potenciar aún más los resultados como sugieren Soares da Silva y cols. (2024) y Maroneze y cols. (2023).

En resumen, los resultados presentados coinciden en que los modelos TCN multivariados ofrecen una herramienta poderosa para la predicción de variables meteorológicas críticas, superando en varios aspectos a los enfoques tradicionales lineales, y aportando un soporte sólido para la toma de decisiones en la planificación energética y gestión de recursos renovables en la región de estudio.

## Anexo

En el siguiente link se encuentra el repositorio en GitHub con todos los codigos, referencias y archivos utilizados.

<https://github.com/RaulFrugone/Velocidad-del-Viento-en-La-Serena>

## Referencias

- Almeida, R. C., Bastos, B. Q., y Santos, D. (2021). Practical application of artificial neural networks to wind-power prediction in brazilian territory. doi: 10.1109/IJCNN52387.2021.9533425
- Bai, S., Kolter, J. Z., y Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, *abs/1803.01271*, 1–17. doi: 10.48550/arXiv.1803.01271
- Bastos, B. Q., Cyrino Oliveira, F. L., y Milidiú, R. L. (2021a). Componentnet: Processing u- and v-components for spatio-temporal wind speed forecasting. *Electric Power Systems Research*, *192*, 106922. doi: 10.1016/j.epsr.2020.106922
- Bastos, B. Q., Cyrino Oliveira, F. L., y Milidiú, R. L. (2021b). U-convolutional model for spatio-temporal wind speed forecasting. *International Journal of Forecasting*, *37*(2), 949–970. doi: 10.1016/j.ijforecast.2020.10.007
- Camelo, H. d. N., Lucio, P. S., LealJunior, J. B. V., dos Santos, D. v. G., y de Carvalho, P. C. M. (2018). Innovative hybrid modeling of wind speed prediction involving time-series models and artificial neural networks. *Atmosphere*, *9*(2), 77. doi: 10.3390/atmos9020077
- Costa, M., Ruiz-Cárdenas, R., y Mineti, L. B. (2021). Dynamic time-scan forecasting for multi-step wind speed prediction. *Renewable Energy*, *177*, 584–595. doi: 10.1016/j.renene.2021.05.160
- Dematte, R., Gandolfo Raso, E., y Huespe, J. (2023). Wind statistical characterization in mendoza province, argentina. *WSEAS Transactions on Environment and Development*, *19*, 1392–1405. doi: 10.37394/232015.2023.19.126
- Goulart, A. J. H., y de Camargo, R. (2021). On data selection for training wind forecasting neural networks. *Computers and Geosciences*, *155*, 104825. doi: 10.1016/j.cageo.2021.104825
- Hung, F.-y. (2025). *Introduction to air pressure (part ii)*. Descargado 2025-07-04, de <https://www.weather.gov.hk/en/education/weather/meteorology-basics/00117-introduction-to-air-pressure-part-ii.html>
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (2013). *An introduction to statistical learning with applications in r*. New York: Springer.
- Khosravi, A., Machado, L., y Nunes, R. O. (2018). Time-series prediction of wind speed using machine learning algorithms: A case study osório wind farm, brazil. *Applied Energy*, *224*, 550–566. doi: 10.1016/j.apenergy.2018.05.043
- Korn, G. A., y Korn, T. M. (2000). *Mathematical handbook for scientists and engineers: Definitions, theorems, and formulas for reference and review* (3.<sup>a</sup> ed.). New York: Dover Publications. (Apéndice B9: Plane and Spherical Trigonometry)
- Lazzús, J. A., y Salfate, I. (2017). Long-term prediction of wind speed in la serena city (chile) using hybrid neural network–particle swarm algorithm. *Earth Sciences Research Journal*, *21*(1), 29–35. doi: 10.15446/esrj.v21n1.50337
- Maroneze, R., Costa, F. D., y Acevedo, O. C. (2023). A new stable boundary-layer parameterization for weather and forecasting models: A heat-flux budget approach. *Boundary-Layer Meteorology*, *188*, 209–235. doi: 10.1007/s10546-023-00810-4
- Miguel, J. V. P., Fadigas, E. A., y Sauer, I. L. (2019). The influence of the wind measurement campaign duration on a measure-correlate-predict (mcp)-based wind resource assessment. *Energies*, *12*(19), 3606. doi: 10.3390/en12193606
- Silva, D., y Lima, C. (2018). Brazilian northeast jet stream: Frequency, wind speed and direction. *Meteorological Applications*, *25*, 589–599. doi: 10.1002/met.1688
- Silva, J. R., Dias, F. G., y Moreira, C. A. (2022). Wind speed prediction study: Practical application of



- ann to energy production in brazilian territory. En S. Ahuja (Ed.), *Advances in intelligent systems and computing*. Springer. doi: 10.1007/978-3-030-96302-6\_59
- Soares da Silva, M., Dragaud, I. C. D. V., y Jacinto, L. d. F. R. (2024). Operational wind forecast over the offshore campos basin and adjacent coastal region in brazil: Support for the oil and gas industry toward renewable energy transition. *Journal of Ocean Engineering and Marine Energy*, 10, 589–610. doi: 10.1007/s40722-024-00332-2