

ESCUELA POLITÉCNICA NACIONAL

**DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y
ELECTRÓNICA**

**DESARROLLO DE UN MODELO DE PRONÓSTICO DE VIENTO DE
CORTO PLAZO PARA PARQUES EÓLICOS, BASADO EN SERIES
TEMPORALES, TÉCNICAS ESTADÍSTICAS Y MACHINE
LEARNING**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
MAGISTER EN ELECTRICIDAD MENCIÓN REDES ELÉCTRICAS
INTELIGENTES**

FRANKLIN HUMBERTO RUIZ VILLACÍS

DIRECTOR: Dra.- Ing. XIMENA PATRICIA GAVELA GUAMÁN

Quito, Marzo 2024

AVAL

Certifico que el presente trabajo fue desarrollado por Franklin Humberto Ruiz Villacís, bajo mi supervisión.

Dra.- Ing. Ximena Patricia Gavela Guamán
DIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARACIÓN DE AUTORÍA

Yo, Franklin Humberto Ruiz Villacís, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración dejo constancia de que la Escuela Politécnica Nacional podrá hacer uso del presente trabajo según los términos estipulados en la Ley, Reglamentos y Normas vigentes.

Franklin Humberto Ruiz Villacís

DEDICATORIA

Dedico este trabajo de titulación a mi Dios por darme las fuerzas y la sabiduría para culminar mi tesis. A mis padres por su apoyo incondicional en todo este proceso, también a mi querida esposa Gissela Chérrez que ha sido mi inspiración día a día para cumplir mis metas.

A mis hermanos y a mi sobrinita que siempre han estado presentes en este proceso para poder finalizar con mi proyecto de titulación

AGRADECIMIENTO

Agradezco a Dios por la vida que me ha regalado y que me ha dado las fuerzas para salir de los momentos difíciles.

Agradezco también a mis padres por el apoyo y el amor incondicional que me han dado durante todo este tiempo, han sido una pieza fundamental para lograr mis metas propuestas, a mis hermanos que siempre me apoyaron en los momentos difíciles.

A mi esposa Gissela Chérrez que ha me acompañado en cada experiencia vivida y que siempre me inspira para ser mejor cada día, gracias, amor por todo tu apoyo para poder terminar mi proyecto de titulación

Un agradecimiento especial a mi tutora Patricia Gavela por su paciencia, sabiduría y apoyo constante para la realización de mi proyecto de titulación. A mis profesores, por brindarme cada uno de sus conocimientos.

ÍNDICE DE CONTENIDO

AVAL.....	I
DECLARACIÓN DE AUTORÍA	II
DEDICATORIA	III
AGRADECIMIENTO	IV
ÍNDICE DE CONTENIDO.....	V
RESUMEN.....	VII
ABSTRACT.....	VIII
1. INTRODUCCIÓN.....	1
1.1 Objetivo General.....	2
1.2 Objetivos Específicos.....	2
1.3 Alcance.....	2
2. MARCO TEÓRICO.....	3
2.1. VELOCIDAD DEL VIENTO	3
2.2. GENERACIÓN EÓLICA.....	4
2.2.1. TIPOS DE PARQUES EÓLICOS.....	4
2.2.2. TURBINAS EÓLICAS.....	5
2.3. PLANIFICACIÓN OPERATIVA DE PARQUES EÓLICOS [14]	5
2.3.1. SISTEMAS DE CONVERSIÓN DE ENERGÍA EÓLICA.....	6
2.4. MODELACIÓN DEL VIENTO.....	10
2.4.1. DISTRIBUCIÓN ESTADÍSTICA DE LA VELOCIDAD DEL VIENTO [21] [4]10	
2.5. PRONÓSTICO DE VIENTO.....	11
2.5.1. Métodos Físicos.....	13
2.5.2. Métodos Estadísticos y Probabilísticos.....	15
2.5.3. Métodos de Inteligencia Artificial.....	17
2.6. VENTAJAS Y DESVENTAJAS DE LAS TÉCNICAS DE MODELACIÓN Y PRONÓSTICO DE VELOCIDAD DE VIENTO	17
2.7. SERIES TEMPORALES	19
2.8. TÉCNICAS DE MODELACIÓN DE SERIES TEMPORALES.....	20
2.8.1. Métodos Estadísticos	20
2.8.2. Métodos de Inteligencia Artificial.....	22
2.9. ÍNDICES DE DESEMPEÑO.....	24
2.10. CENTRAL EÓLICA VILLONACO [55].....	24
3. METODOLOGÍA.....	25
3.1. ADQUISICIÓN DE DATOS	26

3.2.	PROCESAMIENTO Y LIMPIEZA DE DATOS	27
3.3.	ANÁLISIS ESTADÍSTICO	28
3.4.	DISEÑO Y ENTRENAMIENTO DE SERIES TEMPORALES	29
3.4.1.	Extracción de Datos Temporales	29
3.4.2.	Creación de Dataframe basado en Temporalidad.....	30
3.4.3.	Descomposición de datos de series temporales en Tendencia y Estacionalidad	30
3.4.4.	Prueba de Dickey Fuller Aumentada (ADF)	31
3.4.5.	Creación de modelos de pronósticos.....	35
4.	RESULTADOS Y DISCUSIÓN	41
4.1.	PROCESAMIENTO Y LIMPIEZA DE DATOS	42
4.2.	ANÁLISIS ESTADÍSTICO	43
4.3.	ENTRENAMIENTO DE SERIES TEMPORALES	47
4.3.1.	Modelo SARIMA	47
4.3.2.	Modelo Autoregresivo Recursivo	54
4.3.3.	Modelo Autoregresivo Recursivo Con Predictores Modificados	58
4.3.4.	Comparativa de modelos basados en un pronóstico de 24 horas	61
4.3.5.	Cálculo de índices de rendimiento de modelos basados en un pronóstico de 72 horas	62
5.	CONCLUSIONES Y RECOMENDACIONES	65
5.1.	CONCLUSIONES	65
5.2.	RECOMENDACIONES	66
	REFERENCIAS BIBLIOGRÁFICAS	67
	ANEXOS.....	73
	ANEXO A.....	73
	ANEXO B.....	80

RESUMEN

En este trabajo de titulación se presenta el desarrollo de modelos de pronóstico de viento basados en SARIMA y modelos autorregresivos con aplicación a parques eólicos. El modelo se validó para la central de generación eólica de Villonaco, considerando entre sus datos más importantes, los históricos de la velocidad del viento, temperatura y potencia activa. Para el procesamiento de la información y en general para el desarrollo del trabajo se empleó el entorno de programación de Python. Inicialmente, se realiza la extracción del conjunto de datos hacia el entorno de Python, seguidamente se realiza un análisis de las características del conjunto de datos, con el fin conocer la presencia de datos anómalos y del tipo de datos de cada variable. Complementariamente, se realiza un análisis de tipo estadístico, que permite conocer el comportamiento de los datos y su respectiva distribución. Para el desarrollo de los modelos de pronóstico, se realiza un análisis de descomposición de series y el uso de la prueba de Dickey-Fuller, de forma que se tiene un panorama inicial de las características de la serie temporal en relación a los datos base, para posteriormente efectuar el entrenamiento de las series temporales basadas en el modelo SARIMA y las autorregresivas, para finalmente realizar las respectivas pruebas de entrenamiento y pronóstico, además de definir su precisión en base al índice de error cuadrático.

PALABRAS CLAVE: Pronósticos de viento, series temporales, sarima, modelos autorregresivos.

ABSTRACT

This degree work presents the development of wind forecasting models based on SARIMA and autoregressive models with application to wind farms. The model was validated for the Villonaco wind power plant, considering among its most important data, the historical data of wind speed, active power and temperature. The Python programming environment was used to process the information and in general for the development of the work. Initially, the data set was extracted to the Python environment, followed by an analysis of the characteristics of the data set, in order to determine the presence of anomalous data and the type of data for each variable. Complementarily, a statistical analysis is performed, which allows to know the behavior of the data and its respective distribution. For the development of the forecasting models, an analysis of series decomposition and the use of the Dickey-Fuller test is carried out, so that an initial panorama of the characteristics of the time series in relation to the base data is obtained, to later carry out the training of the time series based on the SARIMA model and the autoregressive ones, to finally carry out the respective training and forecasting tests, besides defining their accuracy based on the quadratic error index.

KEYWORDS: Wind forecasts, time series, SARIMA, autoregressive models.

1. INTRODUCCIÓN

A nivel mundial existe una clara tendencia hacia la transición energética que busca el reemplazo de las energías convencionales por energías más limpias y amigables con el medio ambiente. El Ecuador no ha sido ajeno a esta tendencia, y hace más de una década planteó su objetivo de producción de energía eléctrica diversificando su matriz energética a partir de la incorporación de energías alternativas no convencionales. Dentro de las opciones de generación sustentable, está la eólica, que, si bien no ha logrado desarrollarse a gran escala, tiene un gran potencial en provincias como Loja, Azuay, El Oro y Pichincha, según consta en el Atlas Eólico del Ecuador [1], por lo que se espera que a futuro se desarrollen más proyectos eólicos en el país, lo cual incrementa oportunidades de producción energética con este tipo de tecnología, así como los desafíos en cuanto a la predicción de energía eléctrica que pueden producir los parques eólicos.

La situación geográfica del Ecuador viabiliza la posibilidad de producir electricidad basada en recursos no convencionales, ya que el país cuenta con índices de radiación solar altos y por la diferencia de temperatura que existe entre regiones se pueden encontrar índices de viento con buenas características energéticas, pero muchas veces existen limitantes relacionadas con el hecho de que no existe suficiente conocimiento o desarrollo de métodos para la correcta evaluación de estos recursos, y no solo para la etapa de planificación de la expansión, sino que además para la planificación operativa de corto plazo.

En el contexto de las Redes Eléctricas Inteligentes (Smart Grids), uno de los principales problemas que existen con relación al abastecimiento de la demanda con electricidad producida a partir de energías renovables, es la alta variabilidad y la impredecibilidad de las fuentes solares y del viento [2]. La evaluación del desempeño de un sistema eólico está relacionada directamente con el comportamiento de su componente principal que es el viento, específicamente de la velocidad. La estocasticidad del viento natural, dificulta hacer una predicción precisa de la producción energética que pueden alcanzar los parques, pues al ser un parámetro muy variable, tiene necesariamente que analizarse de forma estadística, con el fin de obtener estimaciones aproximadas del comportamiento del viento en lugares de medición específicos, para establecer por un lado, el máximo potencial de energía eléctrica que puede ser instalado de acuerdo al lugar de estudio y número de aerogeneradores disponibles [1], en la etapa de evaluación de factibilidad del proyecto, así como la producción diaria posible una vez instalado el parque eólico en su etapa operativa.

Bajo lo expuesto, se presenta una propuesta de desarrollo de un modelo de pronóstico de viento basado en series temporales, mediante el método de SARIMA, y de técnicas Autorregresivas (AR) basadas en machine learning, para el pronóstico de viento de corto plazo para centrales eólicas implantadas en el Ecuador.

1.1 Objetivo General

Desarrollar un modelo de pronóstico de viento para la estimación de la producción energética de centrales de generación eólica en el corto plazo.

1.2 Objetivos Específicos

- Realizar un estudio del arte de los modelos de pronóstico de viento de corto y largo plazo para la estimación de producción energética de parques eólicos.
- Realizar un análisis de los principales parámetros, variables y aspectos del viento que influyen en la producción energética de un parque eólico, así como los criterios que deberían considerarse para el desarrollo de un modelo de pronóstico en el corto plazo.
- Desarrollar un modelo para el pronóstico de viento de corto plazo para una central eólica y validar el modelo con datos de viento de un parque eólico del Ecuador.

1.3 Alcance

En el presente trabajo de titulación se propone el desarrollo de un modelo de pronóstico de viento para el corto plazo, con el fin de estimar la producción energética de parques eólicos en su etapa operativa. Para el modelo se propone el uso de la Técnica de Modelo Estacionario de Auto Regresión Integrado con Media Móvil (SARIMA) basado en métodos estadísticos y Modelos Autorregresivos (ARM). La metodología considera el análisis estadístico de datos del viento, considerando parámetros como de la dirección y velocidad del viento. El modelo de la serie temporal para el pronóstico se validará usando datos históricos de velocidad de vientos disponibles para la central Eólica Villonaco.

2. MARCO TEÓRICO

En este capítulo se revisarán todos los aspectos teóricos y conceptuales relacionados con: la operación y funcionamiento de los parques de generación eólica, considerando su fundamentación en el comportamiento y caracterización del viento, así como los aspectos característicos de los modelos para los pronósticos de viento.

2.1. VELOCIDAD DEL VIENTO

Característicamente, el viento es aire en movimiento, el cual tiene como origen a la diferencia de temperatura que se encuentra en diferentes puntos de la tropósfera. La energía solar calienta el aire de manera no uniforme, y hace que las masas de aire caliente asciendan y el aire frío descienda en un movimiento circundante, este fenómeno se dará de forma suave o fuerte y dependerá de la velocidad [3].

La velocidad del viento es una magnitud vectorial tridimensional, la cual experimenta fluctuaciones aleatorias de pequeña escala en el espacio y en el tiempo, de manera que se superponen a un flujo organizado de mayor escala [3].

La velocidad del viento es un parámetro estocástico, el cual varía de acuerdo con el día, hora y año. Es un fenómeno muy variable, ya que hay días o épocas en las cuales la velocidad del viento será relativamente constante y fuerte, mientras que en otras ocasiones el viento puede ser suave o moderado.

Para el monitoreo de la velocidad del viento y la consecuente toma de mediciones, se utilizan estaciones meteorológicas. Para el registro de la velocidad del viento las estaciones consideran otras variables relacionadas con el sitio en la que se encuentra y la altura en donde se realicen las mediciones, además de otras variables meteorológicas tales como la presión atmosférica, temperatura, humedad y la densidad del aire. De esta manera queda definido que, en base a las características geográficas y meteorológicas, se presentarán ciertos perfiles de viento en una determinada hora y/lugar [4].

La información acerca de los modelos del viento puede ser de utilidad frente a la optimización del diseño de aerogeneradores, así como también de su producción y economía. La modelación más comúnmente usada es basada en la función de distribución estadística.

2.2. GENERACIÓN EÓLICA

La generación eólica es un tipo de energía limpia de emisión cero, consecuente del desarrollo tecnológico de los últimos años; su principal aprovechamiento está asociado con la producción de energía eléctrica [5] [6].

El proceso de conversión de la energía eólica consiste en la conversión de la energía física del viento en energía eléctrica; para lo cual, se considera la energía cinética que posee el viento. Esta energía cinética se convierte en energía mecánica mediante el uso de un aerogenerador, y posteriormente esta energía mecánica pasa a un proceso de conversión en energía eléctrica [7] [8].

2.2.1. TIPOS DE PARQUES EÓLICOS

En general existen modelos de parques eólicos, caracterizados en función del lugar en dónde se genera la energía eléctrica, pudiendo ser estos: un modelo de producción de energía eólica en tierra u onshore, y modelos de producción de energía eólica instalados en el mar u offshore [9].

- **Parques eólicos Terrestres u Onshore** [10]: se trata de parques eólicos formados por conjuntos de aerogeneradores de gran potencia que se encuentran conectados a la red eléctrica. Estas instalaciones se implantan en lugares que suelen tener una cantidad suficiente de influencia eólica, tales como grandes estepas o regiones costeras. Dentro de este tipo de generación, se dispone de las denominadas generadoras minieólicas, que por lo general se destina al uso doméstico, teniendo en cuenta que pueden ser construcciones a pequeña escala, capaces de ser instaladas en zonas aisladas y alejadas de la red eléctrica.
- **Parques eólicos Marinos u Offshore** [11]: Tienen una producción similar a los parques eólicos terrestres; la única diferencia que se presenta es que los aerogeneradores se instala en medio del océano, de manera que se aprovecha la capacidad eólica que se puede producir en alta mar.

2.2.2. TURBINAS EÓLICAS

De acuerdo con lo mencionado anteriormente, se conoce que los aerogeneradores generan energía a partir de la intensidad moderada del viento para impulsar su generador. Básicamente, los aerogeneradores se consideran como un avance de los antiguos molinos de viento, aunque estructuralmente en la actualidad presentan tres bordes de corte, los cuales giran alrededor de una forma plana [12].

La mayor parte de las turbinas eólicas comienzan a generar energía a velocidades de giro entre 3-4 m/s, y tienen la capacidad de generar un mayor volumen de energía en una velocidad de alrededor de 15 m/s, y una velocidad de parada frente a un posible daño por vientos fuertes de 25 m/s. En general, la operación de una turbina eólica se basa en los siguientes conceptos [13]:

- Casi toda la energía de la Tierra se la puede generar con cualquier motor giratorio.
- Se da un cambio del desarrollo de la corriente rectilínea al giro del polo mediante superficies aerodinámicas pivotantes [14].

2.3. PLANIFICACIÓN OPERATIVA DE PARQUES EÓLICOS

Un parque eólico o también denominado granja eólica, se trata de una agrupación de aerogeneradores empleados para la producción de energía eléctrica. La disposición de los aerogeneradores en los parques eólicos dependerá de la superficie disponible y de las características del viento en el sitio de emplazamiento.

En el desarrollo de los parques eólicos, uno de los aspectos más importantes es el de la operación y mantenimiento. En cierta manera, existen problemas a la hora de operar los parques eólicos, lo cual puede limitar la vida útil de los equipos y sus diferentes componentes, de forma que su afectación se puede reflejar en la generación de pérdidas, y disminución de la eficiencia.

En general, la operación de los parques eólicos se relaciona con la forma de operación del aerogenerador, consecuente de aquello se puede estimar la cantidad de energía que produce en relación de su operación. Las curvas características de los aerogeneradores definen el funcionamiento de las turbinas sin que se considere la forma de operación de estas, es decir, si se tiene una velocidad constante o velocidad variable. En este sentido, como parámetros requeridos dentro de la planificación de los parques eólicos está la

definición de la curva de potencia de los aerogeneradores, lo cual está dentro de las competencias del diseño y del fabricante.

2.3.1. SISTEMAS DE CONVERSIÓN DE ENERGÍA EÓLICA

Dentro de la operación de un sistema de generación eólica, se tiene como punto de partida el concepto de los denominados Sistemas de Conversión de Energía Eólica (WECS, por sus siglas en inglés); el cual se trata de un conjunto de dispositivos electromecánicos que se destinan a la producción de energía eléctrica a través del uso de una fuente primaria de energía, como lo es la energía cinética del viento [15].

Los sistemas de generación eólica aprovechan la energía cinética de las masas de aire, las cuales al chocar con los alabes de una turbina, producen el movimiento giratorio, con lo que parte de esa energía se convierte en energía mecánica rotatoria. A tal turbina se la denomina turbina eólica, la cual presenta un eje acoplado a un generador eléctrico, ya sea de forma directa o mediante una caja de engranajes, de esta manera la energía mecánica se convierte en energía eléctrica. Dependiendo del generador que se use, y dada la intermitencia del viento, se puede usar además de un controlador electrónico para que se mejoren las condiciones de operación del conjunto turbina-generador [15] La conexión hacia la red se la realiza a través de un convertidor de potencia. A través de este convertidor de potencia se puede hacer control de la velocidad del generador, y de esta manera, las fluctuaciones de potencia que son causadas por las variaciones de viento, las mismas que son absorbidas por cambios en la velocidad del rotor. Un esquema general del proceso de conversión de energía eólica se muestra en la Figura 2.1.

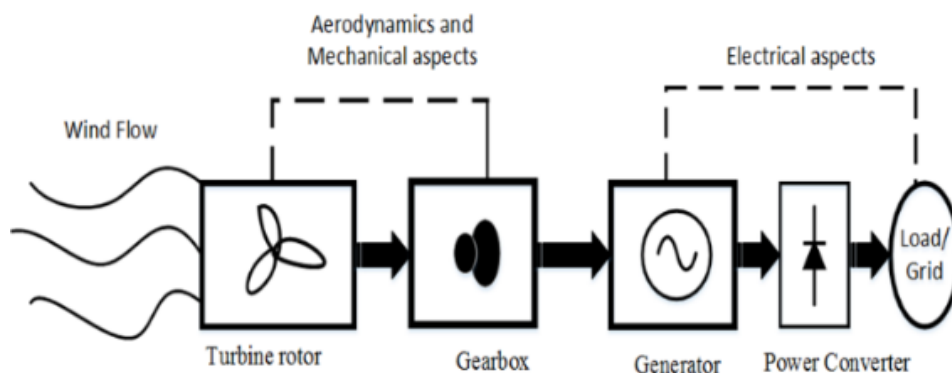


Figura 2.1. Esquema general del Sistema de Conversión de Energía Eólica [16].

2.3.2. Curva de Potencia y Operación de Parques Eólicos

La energía cinética que se encuentra disponible en el viento y su transformación en energía cinética, tienen dependencia en gran medida con la forma en la cual se realiza la interacción del rotor del generador eólico y la corriente de aire. En este sentido, se puede definir una relación entre la energía eléctrica generada y la energía cinética de la velocidad relacionada con la velocidad de viento [17] a través de una curva.

La curva de potencia de un aerogenerador determina la potencia eléctrica disponible para un aerogenerador frente a diferentes velocidades del viento [18]. De tal manera que, se tiene como objetivo, aprovechar al máximo la velocidad del viento en la producción de potencia por parte del aerogenerador.

Este comportamiento del aerogenerador queda expresado a través de la Ecuación 2.1 para el cálculo de la potencia mecánica máxima.

$$P_{m_max} = \frac{1}{2} \rho \pi c_{p_max}(\lambda, \beta) \left(\frac{R^5}{\lambda_{opt}^3} \right) \omega_m^3 \quad (2.1)$$

Donde:

ρ : densidad del aire.

c_{p_max} : coeficiente de potencia del rotor.

λ : relación de velocidad de punta.

β : ángulo de inclinación de pala.

R : radio de la turbina.

ω_m : velocidad del rotor.

De la Ecuación 2.1. se tiene que, para una velocidad de viento dada, la potencia de generación mecánica máxima se obtiene si c_p es maximizado. El valor máximo de $c_p \rightarrow c_{p_max}$, ocurre en un valor definido en $\lambda \rightarrow \lambda_{opt}$. Esto implica que para una velocidad de viento variable, es necesario que la velocidad del rotor se ajuste de manera proporcional para que se alcance siempre el valor de λ_{opt} para obtener una salida de potencia máxima de la turbina, de manera que siga la curva de potencia máxima, tal como se observa en la Figura 2.2.

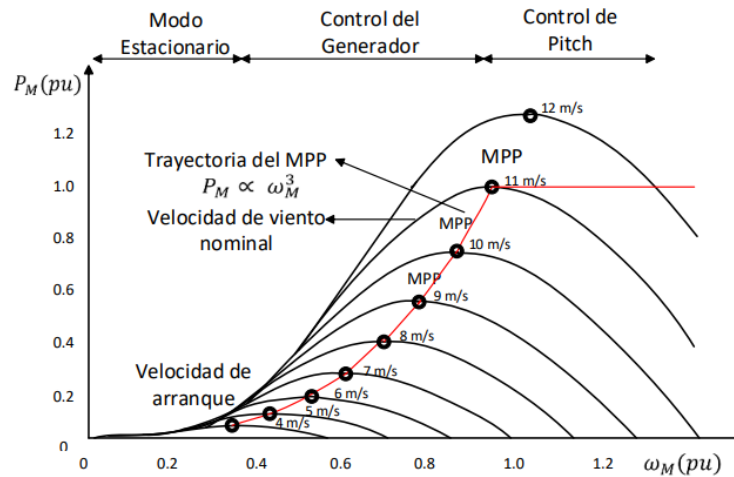


Figura 2.2. Característica Velocidad-Potencia de una turbina eólica [15].

Otro parámetro importante a considerar es el par producido T_m de la potencia mecánica, la cual está dada por la Ecuación 2.2.

$$T_m = \frac{P_m}{\omega_m} \quad (2.2)$$

2.3.3. Relación Costo-Operación de Aerogeneradores

Para el caso de los aerogeneradores modernos, en general, se diseña para un trabajo de alrededor de 120000 horas de operación a lo largo de su tiempo de vida de diseño de 20 años [19].

En cuanto a los costos de mantenimiento son en forma general muy bajos en relación con las turbinas nuevas, y va aumentando conforme la turbina va envejeciendo. En relación con las turbinas más antiguas, tienen costes de reparación y mantenimiento de una media de alrededor del 3 por ciento de la inversión inicial de la turbina. Las turbinas más nuevas que en general son más grandes, se tiene un rango estimado de entre el 1.5 al 2 por ciento de forma anual de la inversión inicial de la turbina. Particularmente, los costes de mantenimiento-operación y en su necesidad están relacionados con el desgaste y rotura en la turbina, que en general tienden a incrementarse en consecuencia del aumento de la producción energética, de tal manera que tener una buena gestión en cuanto a la relación de la producción energética con la disponibilidad del recurso eólico es de gran importancia, frente a la planificación del mantenimiento y costos de operación [19].

2.3.4. Planificación e Implicaciones Económicas y Operativas

Los parques eólicos no tienen una producción del 100% de su energía disponible, lo cual tiene relación con los factores que producen una baja capacidad de captura de energía, que están en contraste con un mal emplazamiento, variabilidad del viento y turbulencia. Teniendo en cuenta la operación de las turbinas eólicas en la actualidad, se tiene que la máxima capacidad de generación se da cuando el viento alcanza velocidades de entre 12 m/s (viento moderado) y 25 m/s (un temporal). Pero como limitante se tiene que, aunque se encuentre en una de las mejores ubicaciones, la velocidad del viento es menor que 12 m/s durante una gran parte del tiempo. En este sentido, el factor de planta alrededor del mundo es de aproximadamente 20%, aunque se tiene datos de factor de planta de entre 18% (Italia) hasta 33% (Estados Unidos) [20].

2.3.5. Características y condiciones de la Información del viento para Proyectos de evaluación de potencial eólico y emplazamientos de Parques Eólicos

Comprender de forma adecuada las características del viento es de gran importancia para el desarrollo de estudios de capacidad eólica, los cuales buscan determinar el sitio adecuado para la implementación de un parque eólico de acuerdo con las necesidades de generación que se determinan en los correspondientes estudios.

El viento se caracteriza por tener una velocidad y dirección, que tiene una afectación directa por factores como de la zona geográfica, clima y la topología del terreno; para su medición se toman en cuenta los siguientes criterios:

- *Ubicación del anemómetro:* debe estar ubicado a 10 metros de altura con relación al suelo.
- *Resistencia de los postes:* estructuralmente debe presentar resistencia a vientos fuertes.
- *Ubicación de veleta:* la veleta debe tener una ubicación correcta mediante el uso de una brújula.
- *Tiempo de muestreo:* el tiempo de muestreo recomendado es de 10 minutos, considerando su uso en pronósticos de corto plazo [21] [4].

2.4. MODELACIÓN DEL VIENTO

2.4.1. DISTRIBUCIÓN ESTADÍSTICA DE LA VELOCIDAD DEL VIENTO

La distribución estadística de la velocidad del viento tiene un comportamiento variable, que depende del lugar en el cual se esté analizando el viento, de forma que dependerá de las condiciones climáticas locales, del paisaje y de su superficie. La importancia de la distribución estadística del viento basada en la velocidad promedio del viento está en la cuantificación del recurso eólico disponible en un determinado sector [22].

Dentro de la elaboración de un anteproyecto, el diseño y la implementación de un parque eólico, se realiza la instalación de una estación meteorológica, la cual mide la velocidad del viento y otras variables. Este es un aspecto importante para decidir la viabilidad del proyecto y su potencial rentabilidad. Dentro de este análisis aparece una herramienta de modelización de la velocidad del viento que es el uso de la distribución estadística de Weibull.

Esta función de distribución de datos es la más utilizada en la modelación del viento, esto debido a que proporciona una descripción aproximada al comportamiento de la velocidad del viento a lo largo de un determinado periodo de tiempo. La expresión de la distribución estadística de Weibull está dada por la Ecuación 2.3.

$$p(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^k} \quad (2.3)$$

Donde:

v : velocidad del viento.

k : factor de forma.

c : factor de escala.

$p(v)$: frecuencia o probabilidad de ocurrencia de la velocidad de viento.

De la Ecuación 2.3 se tiene que, a un mayor valor de c , la velocidad del viento es más alta. En cuanto al valor de k , indica la estabilidad del viento, de manera que mientras mayor es k , mayor estabilidad tiene la velocidad del viento. En este sentido, mediante la Figura 2.3 se muestra una representación de la distribución de Weibull.

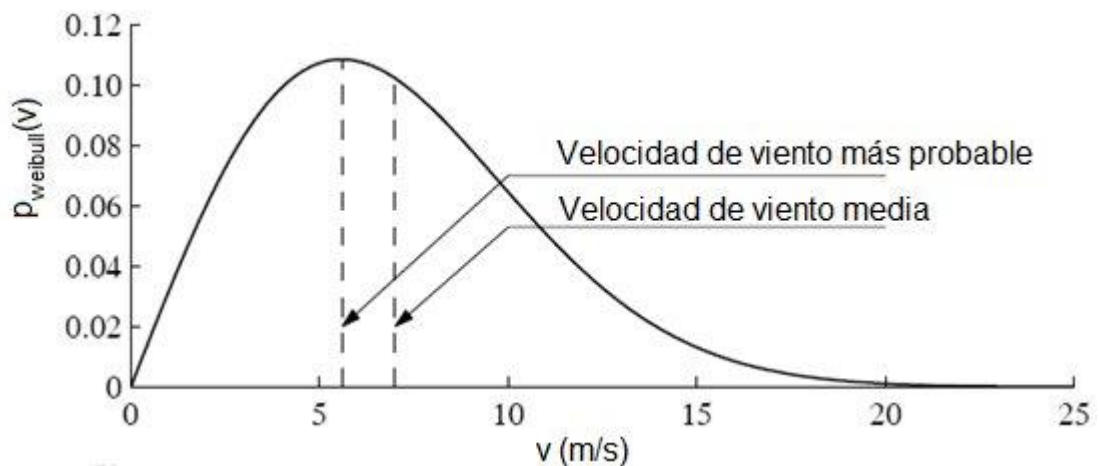


Figura 2.3. Distribución de Weibull para velocidad del viento [23].

De acuerdo al gráfico de distribución normal de Weibull de la Figura 2.3, se tiene que el área bajo la curva siempre vale exactamente 1, considerando que la probabilidad de que el viento sople a cualquiera de las velocidades, incluyendo el cero, tiene que ser el 100 por ciento. La velocidad del viento medio se trata del promedio de las observaciones de la velocidad del viento. Además, de la gráfica se puede observar que la distribución de velocidades del viento es sesgada, es decir, no es simétrica. En ciertas ocasiones tendrá velocidades muy altas, pero son muy raras.

Además de esto, hay que considerar como punto importante, que la distribución estadística de las velocidades del viento varía de un lugar a otro, ~~que son~~ como consecuencia de las condiciones climáticas locales, del paisaje y de su superficie, de manera que la distribución de Weibull puede variar tanto en forma como de su valor medio.

2.5. PRONÓSTICO DE VIENTO

En función del horizonte de tiempo para el pronóstico, los modelos de pronósticos de velocidad del viento se pueden clasificar en: pronósticos a largo plazo (a meses de distancia o estacional), pronósticos a mediano plazo (días o semanas) y pronósticos a corto plazo (a una hora o minutos de distancia).

- *Modelos de corto plazo:* basados en un enfoque estadístico de series de tiempo, los cuales son capaces de predecir de manera eficiente la energía eólica para horas.
- *Modelos de mediano plazo:* de la misma manera que los modelos de corto plazo, se puede basar en modelos de series temporales, de forma que se puedan elaborar pronósticos de días hasta semanas.

- *Modelos de largo plazo:* Permite pronósticos de energía eólica considerando ventanas de tiempo de meses hasta años.

En la Tabla 2.1 se detalla las aplicaciones de cada uno de los modelos de pronósticos en consideración al horizonte de tiempo [24]:

Tabla 2.1. Horizonte de Tiempo para Modelos de Pronóstico de Viento

Escala de Tiempo	Aplicaciones
Corto Plazo	<ul style="list-style-type: none"> • Planificación del mantenimiento. • Gestión Operativa. • Optimización de la capacidad de transmisión • Control de la orientación de las palas
Mediano Plazo	<ul style="list-style-type: none"> • Planificación Económica del Despacho de carga. • Toma de decisiones. • Seguridad operativa en el mercado eléctrico.
Largo Plazo	<ul style="list-style-type: none"> • Operaciones de la red en tiempo real. • Acciones de control. • Costo Óptimo de operación.

Entre los parámetros más relevantes para la implementación de un modelo de pronóstico se tiene a las siguientes [4]:

- *Variable de predicción:* relacionada con la salida del modelo de pronóstico, la cual puede ser la potencia o la velocidad del viento.
- *Entradas del modelo:* son las variables de entrada hacia el modelo de pronóstico.
- *Tipo de modelo:* relacionado con el tipo de modelo de pronóstico a utilizarse.
- *Horizonte y Tiempo de muestreo:* se definen la ventana de tiempo relacionado con el tipo de pronóstico a desarrollar (corto, mediano y largo plazo).
- *Criterio para evaluar los resultados:* se refieren a los índices que permiten la evaluación del desempeño del modelo.

Dentro del enfoque de modelación de velocidad de viento, se tienen los siguientes métodos generales básicos:

- **Método físico (base numérica):** es un modelo numérico de pronóstico (NWP, por sus siglas en inglés), se usa para la elaboración de pronósticos de fenómenos meteorológicos, los cuales se representan por ecuaciones físicas a través del uso de métodos numéricos [25].
- **Métodos Estadísticos:** basado en el cálculo de correlación entre las mediciones de la velocidad del viento y la potencia media del parque eólico para derivar los pronósticos de la potencia eólica [25] [26].
- **Métodos de Inteligencia Artificial (AI):** como aspecto importante al usar este tipo de modelación es que la predicción de series temporales se las puede desarrollar sin la necesidad de un modelo matemático [27] [28]. Para aquello se emplea una construcción basada en el uso de algoritmos de aprendizaje automático, conocimiento experto o en algunos casos se emplea una mezcla de ambos. Una limitante para el uso de este tipo de modelos es la capacidad de procesamiento computacional que requiere [29].

A continuación, se realizará un breve estudio del estado del arte de diversas aplicaciones enfocadas al pronóstico de velocidad de viento, definiendo las distintas metodologías empleadas, las mismas que en su mayoría se tratan de desarrollos que sugieren nuevas técnicas de modelación y entrenamiento.

2.5.1. Métodos Físicos

2.5.1.1. Weather Research and Forecasting (WRF)

Consiste en un sistema numérico de mesoescala, el cual se emplea en el ámbito de investigación y de predicción operativa atmosférica para mediano y largo plazo. Se compone por dos núcleos dinámicos: ARW (Advanced Research WRF) y el NMM (Nonhydrostatic Meso-scale Model), un sistema de asimilación de datos y una arquitectura de software que facilita una operación en paralelo y la extensibilidad del sistema [30].

Basado en este modelo, se tiene como primer estudio de referencia al presentado en [31], el cual desarrolla ocho modelos de pronóstico, cuatro de ellos se tratan de un modelo numérico de mesoescala híbrido y regresión lineal (WRF-RL) y los otros cuatro en base a un modelo numérico de mesoescala y redes neuronales artificiales (WRF-RNA). Estos modelos se emplean para el pronóstico de mediano y corto plazo. El análisis se desarrolla en un ciclo diario, con una herramienta que permite suministrar información relacionada

con el despacho óptimo de energía en el sistema eléctrico. En general el WRF es un modelo de software libre de código abierto de gran respaldo de la comunidad científica.

Para el caso de WRF-RL, se tiene que la potencia pronosticada dependerá de la densidad del aire y la velocidad del viento pronosticada a la altura del eje del aerogenerador. Para el caso del WRF-RNA, considera como variables de entrada al modelo de pronóstico a la densidad del aire, velocidad del viento y la dirección del viento medidas por parte de la estación meteorológica, además, se añade la declinación solar de manera que se incorpora la estacionalidad anual como información de entrada y una lógica difusa que permite discriminar la estabilidad e inestabilidad atmosférica. Estos valores provienen de datos históricos de la estación con una resolución de 30 minutos, cuya evaluación de rendimiento lo realizan para un escenario de 24 horas.

Considerando un aprendizaje supervisado, en la citada referencia se propone un método de retropropagación para el entrenamiento de la red neuronal, empleando dos capas (una capa oculta y una de salida), para el caso de la capa oculta se manejan 10 neuronas y la capa de salida maneja una neurona. El trabajo concluye que el modelo que mejor desempeño obtuvo es el híbrido del tipo WRF-RNA. El desempeño se evalúa a través del uso del error medio absoluto.

WRF fue desarrollado en conjunto por distintos institutos, entre los cuales está el National Center for Environmental Prediction (NCEP), del cual en el estudio realizado en [32], a partir de la implementación de módulos de pre-procesamiento (WPS), emplea los datos de los modelos globales que provee el NCEP, los mismos que considera como condiciones iniciales y de borde para los pronósticos locales de viento.

En adición a la literatura citada, existen varios centros de investigación e institutos que realizan investigación dedicada al desarrollo de modelos para la predicción de fenómenos meteorológicos, entre ellos está ARMINES que es una Unidad de investigación conjunta con la *Ecole des Mines de Paris*, quienes han desarrollado modelos de pronóstico de energía para parques eólicos basados en el uso de SCADA online y de las predicciones meteorológicas numéricas, cuyo sistema de previsión puede aceptar de manera genérica como entrada a diferentes tipos modelos de pronóstico meteorológicos [33].

2.5.2. Métodos Estadísticos y Probabilísticos

2.5.2.1. Pronóstico Probabilístico a Corto Plazo del Viento

Los modelos probabilísticos se refieren a la forma en la cual se puede tomar un conjunto de dato obtenido de muestreo a partir de un comportamiento que se supone es de tipo aleatorio. Se considera como un modelo matemático que emplea probabilidad, el cual incluye una serie de asunciones relacionados con la generación de algunos de los datos muestreables, de forma que tenga semejanza con los datos de una población mayor. Considera dentro del desarrollo del modelo un conjunto de ecuaciones las cuales relacionan diversas variables aleatorias, y además de variables en donde pueden aparecer otras variables no aleatorias [34].

Se tiene como trabajo de referencia al desarrollado en [35], que para el pronóstico de velocidad de viento de corto plazo (días) y muy corto plazo (horas) se basa en una técnica de persistencia Lagrangiana local, la cual se usa de manera amplia en pronósticos probabilísticos de precipitación mediante el uso de sistemas de radar a escala continental. Para la creación de un pronóstico probabilístico primero se requiere la predicción de las distribuciones de la velocidad del viento.

Las densidades predictivas de la velocidad del viento se basan en vectores de viento, los mismos que llegan al modelo del aerogenerador en el momento en el cual el pronóstico sea válido. Uno de los enfoques para la generación probabilística es mediante la búsqueda de los valores pronosticados en los puntos próximos de punto de interés. De esta manera, la base del pronóstico probabilístico a corto plazo está relacionada con la nube de puntos u observaciones que se encuentran dentro de la ventana espacio-temporal definida [36].

Por otra parte, del estudio desarrollado en [37] para un modelo de pronóstico de corto plazo, considera que la elección de las funciones de densidad debe ser consistente con los patrones del viento a través de mediciones meteorológicas en una ventana de tiempo de un año; para el ejemplo se tomó una estadística de viento desde el año 2002 al 2003 en la localidad de Oregon en Estados Unidos, de tal manera que se realiza un modelo enfocado al despacho económico. Los datos de velocidad de viento tienen una periodicidad de una hora considerando valores medios. Las velocidades de viento no presentan valores no negativos y en general están sesgadas a la derecha por la baja probabilidad de valores altos. No definen el número mínimo de datos ni de la ventana de tiempo en cuanto a los datos históricos necesarios para el desarrollo del modelo, pero consideran una resolución de hasta 15 min para definir pronósticos de 24 horas. Los índices para la evaluación del

desempeño que utilizan son el error cuadrático, el error absoluto y el porcentaje de error absoluto [38].

2.5.2.2. Métodos Estadísticos

Los modelos estadísticos se caracterizan por no considerar procesos físicos, de forma que se encargan de parametrizar las relaciones que existen entre los pronósticos meteorológicos y la producción del parque de forma que se logre producir periodos de tiempo corto, de minutos a horas [39].

Para el caso de pronósticos de viento basados en métodos estadísticos se tiene el estudio de [40], en donde se realiza un desglose de diversas técnicas empeladas para este fin.

- *Línea Base: Pronóstico de persistencia:* tiene como parámetro de referencia el hecho de que la velocidad del viento se mantendrá constante en el horizonte de pronóstico. Se trata de un método de alta eficacia para pronóstico de viento en la escala de tiempo subhoraria, de manera que su precisión aumenta a medida que disminuye el horizonte temporal.

- *Pronóstico de persistencia de segundo orden:* aumenta la persistencia al realizar una extrapolación lineal del cambio en la velocidad del viento entre los dos pasos de tiempo anteriores.

- *Regresión rápida de múltiples entradas:* es un tipo de regresión lineal, de manera que puede ser interpretado mediante la inspección de sus coeficientes. Tiene un parámetro de regularización en su función objetivo que se usa en casos de que existe una alta colinealidad entre las variables de entrada.

- *Regresión rápida con cambio de régimen:* en este modelo los datos de entrada se separan en intervalos por la dirección del viento en el sitio objetivo, es decir, se entrena para cada casilla de dirección del viento. Se tiene una alta sensibilidad en los resultados de acuerdo al tamaño de los intervalos.

- *Series Temporales:* El objetivo principal del análisis basado en Series temporales es de la construcción de un modelo que permita la descripción de una determinada estructura de datos, para posteriormente desarrollar predicciones de los valores futuros a corto, mediano y largo plazo de las series temporales

- *Multilayer Perceptron con Múltiples Entradas:* se trata de una red neuronal feed-forward densamente conectada con una o varias capas ocultas. Para el entrenamiento de

los pesos del modelo se usa un descenso gradual a través de retropropagación. Se trata en general de un modelo de aprendizaje automático no lineal. Este modelo captura las correlaciones espaciales no lineales en el campo de viento que los modelos de regresión rápida no los podía capturar.

2.5.3. Métodos de Inteligencia Artificial

2.5.3.1. Redes Neuronales

En la literatura existen algunos modelos de pronóstico de viento basados en redes neuronales, considerando características predefinidas, como, por ejemplo, en el estudio realizado en [41], quienes emplean una red neuronal conectada en cascada que se basa en características de entradas mixtas. Entre las principales características que presenta este tipo de configuraciones está el de aprender de una manera efectiva información útil de una gran cantidad de características de entrada, de forma que se evite problemas de sobreajuste del modelo; de manera que entre los objetivos se tiene el de evaluar el impacto de varias variables meteorológicas de diferentes estaciones vecinas en el desempeño de los modelos de pronóstico basados en redes neuronales.

En el estudio propuesto en [42] se realiza la predicción a corto plazo a partir de series de tiempo incompletas, lo cual se desarrolla a partir del uso de observaciones pasadas de la variable a pronosticar (velocidad del viento). El referido trabajo utiliza una red neuronal artificial (RNA) como un enfoque de predicción y algoritmos genéticos (AG) para definir la topología de la RNA. Posterior al entrenamiento de la RNA, se realiza la reconstrucción de la serie de tiempo de velocidad del viento.

2.6. VENTAJAS Y DESVENTAJAS DE LAS TÉCNICAS DE MODELACIÓN Y PRONÓSTICO DE VELOCIDAD DE VIENTO

Con base al análisis realizado, a continuación, se presenta un cuadro comparativo con las principales ventajas y desventajas de los métodos para el pronóstico de viento:

Tabla 2.2. Tabla comparativa de Modelos de Pronóstico.

TIPO	MÉTODO	VENTAJAS	DESVENTAJAS
Probabilístico	Weather Research and Forecasting (WRF)	<ul style="list-style-type: none"> • Se trata de un modelo de fuente abierta. • Considera los efectos de la superficie terrestre. • Útil para eventos de corta y larga duración. 	<ul style="list-style-type: none"> • Requiere de ventanas de tiempo altas y de mayor número de evaluaciones de casos de estudio para tener una gran precisión de pronósticos.

			<ul style="list-style-type: none"> • Requieren de altos recursos computacionales.
	Pronóstico probabilístico de corto plazo	<ul style="list-style-type: none"> • Considera dentro de su modelo la presencia de incertidumbres asociados a los pronósticos. 	<ul style="list-style-type: none"> • Se requiere de técnicas de modelado numérico y estadísticos para la estimación de la incertidumbre. • Requieren de altos recursos computacionales.
Estadísticos	ARIMA (Autoregressive Integrated Moving Average)	<ul style="list-style-type: none"> • Útiles para la predicción en el corto plazo, debido a su capacidad de aprendizaje de los cambios en la dinámica de la serie. • No requieren de altos recursos computacionales. 	<ul style="list-style-type: none"> • Limitación frente a pronósticos en momentos de cambios estructurales y puntos de giro. • Horizonte de pronóstico en ocasiones muy corto.
	SARIMA (Seasonal ARIMA-Autoregressive Integrated Moving Average)	<ul style="list-style-type: none"> • Es eficiente frente a la presencia de componentes estacionales, es decir, componentes que se repiten en todas las observaciones. La misma que no puede ser abordada por el modelo ARIMA. • Considera parámetros no estacionales y estacionales de los modelos. • No requieren de altos recursos computacionales. 	<ul style="list-style-type: none"> • No está destinado al uso en pronósticos multivariantes. • No se cuenta con otras variables explicativas, emplea tan solo rezagos de la variable dependiente.
	Modelos Autorregresivos (ARMs)	<ul style="list-style-type: none"> • Se basan en modelos de redes neuronales y métodos probabilísticos-estadísticos. • Se optimiza el tiempo computacional considerando el proceso que se implementa. • Su uso está destinado a los pronósticos de corto plazo, considerando la cantidad de datos que maneja y sus métodos probabilísticos y estadísticos inmersos. • Se emplea en procesos aleatorios (estocásticos). 	<ul style="list-style-type: none"> • Paralización unidireccional: enfoque de evaluación o enfoque muestral. • Se requiere de una alta precisión de los valores históricos.
Inteligencia Artificial	Redes Neuronales Artificiales	<ul style="list-style-type: none"> • Capturan de una forma óptima y efectiva características complejas. • El procesado de la información es local. • Las neuronas pueden reconocer patrones que no han sido aprendidos. 	<ul style="list-style-type: none"> • Necesitan un mayor preprocesamiento de los datos. • Suelen necesitar mayor volumen de datos para el entrenamiento del modelo. • Tiempo de aprendizaje elevado. • No son fácilmente explicables.

De acuerdo a lo descrito en la Tabla 2.2, se considera que los modelos que tienen una mejor practicidad frente a la aplicación para la elaboración de pronósticos de viento son el modelo SARIMA y los modelos Autorregresivos, teniendo en cuenta la principal característica de aplicación a modelos de pronóstico de viento a corto plazo, además de presentar un menor esfuerzo computacional en donde se optimiza la capacidad computacional dentro de su elaboración y puesta en prueba.

2.7. SERIES TEMPORALES

Una serie temporal X_t se la representa como una secuencia basada en observaciones que se miden de forma uniforme en el tiempo, y que se ordenan de forma cronológica, de forma que pueden representar una dependencia entre sí [43].

En general, una serie temporal se la puede definir como la unión de tres componentes, los cuales se representan a través de la siguiente ecuación [44]:

$$X_t = T_t + E_t + I_t \quad (2.4)$$

T_t : componente de tendencia

E_t : componente estacional

I_t : Componente aleatoria

- *Componente de Tendencia*: definida como el cambio que se da a largo plazo en cuanto al valor de la media.
- *Componente Estacional*: esta componente no presenta ningún tipo de patrón de comportamiento, sino que se convierte en el resultado de factores fortuitos o aleatorios, los cuales inciden de forma aislada en una serie de tiempo.

De acuerdo a las características que pueden presentar las series temporales, se tiene la siguiente clasificación [43]:

- *Estacionarias*: en este tipo de series temporales, la varianza y la media son constantes en el tiempo. De forma visual se la puede determinar observando si la serie oscila alrededor de una media constante, además, la variabilidad con respecto a ese valor también lo es.
- *No Estacionarias*: en este caso la tendencia y/o variabilidad cambian en el tiempo. Por otra parte, los cambios en la media determinan una tendencia creciente o

decreciente a largo plazo, de forma que la serie no oscila alrededor de un valor constante.

Para que una serie temporal sea estacionaria debe cumplir las siguientes propiedades:

$$\text{Media } E(X_t) = E(X_{t+k}) = \mu \quad (2.5)$$

$$\text{Varianza } V(X_t) = V(X_{t+k}) = \sigma^2 \quad (2.6)$$

$$\text{Covarianza } \gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)] \quad (2.7)$$

De forma general, se conoce que, si una serie de tiempo es de tipo estacionaria, su media, su varianza y su covarianza en diferentes instantes de tiempo permanecen iguales sin tener en cuenta el momento en el cual sean medidos, es decir, son invariantes en el tiempo.

2.8. TÉCNICAS DE MODELACIÓN DE SERIES TEMPORALES

2.8.1. Métodos Estadísticos

2.8.1.1. ARIMA (Autoregressive Integrated Moving Average)

Se la usa para la previsión a corto plazo. Se la define en como la combinación de dos modelos autorregresivos (AR), los cuales se integran con el modelo de media móvil (MA). La notación se da como ARIMA (p,d,q), de donde se tiene que p es el grado del proceso de AR, d es el orden de diferenciación y q es el grado del proceso de MA [45].

El modelo autorregresivo con orden AR(p) del modelo de ARIMA (p,0 ,0) se da de la siguiente manera [46]:

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + e_t \quad (2.8)$$

Donde:

Y_t : serie de tiempo estacionaria.

θ_0 : constante

θ_p : parámetro del modelo autorregresivo

e_t : tiempo residual (t)

En cuanto al modelo de media móvil con orden de MA(q) o ARIMA (0, 0, q) se la representa de la siguiente manera:

$$Y_t = \phi_0 - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_q Y_{t-q} + e_t \quad (2.9)$$

Donde:

Y_t : serie de tiempo estacionaria.

ϕ_0 : constante

ϕ_q : coeficiente del modelo que muestra las ponderaciones de la media móvil

e_t : tiempo residual (t)

2.8.1.2. SARIMA (Seasonal ARIMA- Autoregressive Integrated Moving Average)

El modelo de ARIMA (p, d, q) está compuesto por tres partes: d, es el nivel de diferenciación, p, es el orden autorregresivo y q, es el orden de promedio móvil [47].

Inicialmente, se tiene que el modelo de ARIMA es:

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.10)$$

Donde, z_t es el nivel de diferenciación, δ es un valor constante, ϕ es el operador autorregresivo, a es un valor aleatorio correspondiente al periodo de tiempo t, y θ es un operador de media móvil.

El Modelo de SARIMA se usa para series de tiempo que presentan una variación estacional, que es justamente la característica del comportamiento del viento. En adición al modelo precedente, incluye una notación autorregresiva estacional (P) y una notación de media móvil estacional (Q), las mismas que forman el proceso multiplicativo de SARIMA como (p, d, q) (P, D, Q).

La forma generalizada del modelo SARIMA (p, P, q, Q) es:

$$\phi_p(B)\phi_p(B^S)z_t = \delta + \theta_q(B)\theta_q(B^S)a_t \quad (2.11)$$

El desarrollo del modelo SARIMA se realiza en base al procedimiento definido en [48] y detallado a continuación:

Paso 1: Identificación

A través de este paso se efectúa la selección del orden de diferenciación regular (d), diferenciación estacional (D), el orden no estacional de autorregresivo (p), el orden estacional de autorregresivo (P), el orden no estacional de media móvil (q), y el orden estacional autorregresivo (Q).

La identificación del número de orden se realiza a través de la observación de las autocorrelaciones de muestra (SAC) y de las autocorrelaciones parciales de muestra (SPAC).

Paso 2: Estimación

Los datos de tipo histórico se usan para la estimación de los parámetros del modelo tentativo del Paso 1.

Paso 3: Comprobación de diagnóstico

Este tipo de comprobación de usa para la comprobación de idoneidad del modelo tentativo.

Paso 4: Pronóstico

El modelo precedente del Paso 3 se usa para el pronóstico de los valores a estimar.

2.8.2. Métodos de Inteligencia Artificial

2.8.2.1. Modelos Autorregresivos (ARMs)

Los modelos de distribución autorregresivos (ARM) realizan una descomposición de una distribución conjunta de alta dimensión en distribuciones condicionales unidimensionales, basado en el uso de la regla de la cadena de la teoría de la probabilidad. En la actualidad los modelos ARM están basados en redes neuronales [49].

El entrenamiento de estos modelos se efectúa mediante la evaluación de probabilidad y se debe considerar que el procesamiento computacional empírico de este proceso es mucho más alto que el del muestreo. De tal manera que, los ARM se diseñan para una evaluación rápida, pero en el caso del muestreo de estos modelos es lento [50].

Partiendo de que se tiene a la variable $x \in X$, en donde X es un espacio discreto. Se tiene que un modelo autorregresivo considera a x como una secuencia de variables unidimensionales $(x_i)_{i=1}^d$, de manera que se tiene el siguiente modelo universal de probabilidad [49]:

$$P_{ARM}(X) = \prod_{i=1}^d P_{ARM}(x_i | X_{<i}) \quad (2.12)$$

La denominada universalidad viene derivada de la regla de la cadena de la teoría de la probabilidad, y $X_{<i}$ denota valores $(x)_{j=1}^{i-1}$.

La función de pronóstico F_i puede ser formalizada de la siguiente manera:

- Se tiene que las variables $X_{<i}$ son muestras válidas desde el ARM.
- Sea \tilde{x}_i un pronóstico para la variable x_i .

De esta manera se tiene que una función de pronóstico F_i , que tiene como objetivo de inferir sobre la secuencia futura más probable a partir de una posición i dada toda la información que se dispone:

$$\tilde{x}_i, \tilde{x}_{i+1}, \dots = F_i(x) \quad (2.13)$$

Utilizando cada una de las entradas del muestreo, el ARM puede realizar el cálculo de una secuencia de salida para pasos de tiempo futuros: x'_i, x'_{i+1}, \dots en paralelo. La salida del ARM x'_i es válida: si es una muestra de la distribución del modelo verdadero, considerando el condicionamiento $X_{<i}$ que no contiene pronósticos.

En cada paso consecutivo t , en donde el pronóstico \tilde{x}_{i+t} es igual a la salida del ARM x'_{i+t} , la salida posterior x'_{i+t+1} es válida también.

En el caso en el que la función de pronóstico no concuerde con el ARM, se escribe la última salida válida en el vector de entrada x , y se pasa a la siguiente iteración del muestreo de pronóstico; se tiene un proceso iterativo hasta que se haya muestreado todas las variables.

En la Figura 2.4 se muestra un detalle del muestreo predictivo. En donde se tiene una secuencia de muestras ARM x_0, \dots, x_{i-1} la cual se extiende con pronósticos $\tilde{x}_i, \tilde{x}_{i+1}, \dots$ que son dadas como entrada al ARM.

Como se observa, el ARM presenta una estricta dependencia triangular, en donde su primera salida x'_i es válida y el acondicionamiento consta únicamente de muestras ARM.

Si el pronóstico \tilde{x}_i es igual a x'_i , la siguiente salida x'_{i+1} también será válida.

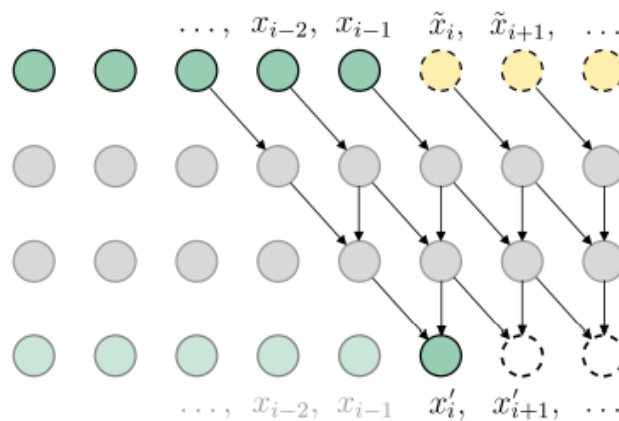


Figura 2.4. Esquema general del muestreo predictivo [51].

2.9. ÍNDICES DE DESEMPEÑO

Una manera de garantizar la precisión y la reducción del nivel de error de los modelos de series temporales entrenados es mediante el uso y medición de índices como:

- Error cuadrático Medio (MSE): proporciona el promedio de la diferencia cuadrática que relaciona a la predicción del modelo y el valor de destino [52].
- Error medio absoluto (MAE): se relaciona con el promedio de la diferencia absoluta entre el resultado de la predicción del modelo y el valor objetivo [53].
- R2 Score: Es una medida estadística que representa la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Se usa comúnmente para evaluar la bondad de ajuste de un modelo de regresión [54].

Tabla 2.3. Índices de desempeño [54].

Criterio	Fórmula
Error cuadrático Medio (MSE)	$\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h$
Error medio absoluto (MAE)	$\sum_{t=T+1}^{T+h} \hat{y}_t - y_t / h$
R2 Score	$1 - \frac{\sum_{t=1}^h (y_t - \hat{y}_t)^2}{\sum_{t=1}^h (y_t - \bar{y})^2}$

2.10. CENTRAL EÓLICA VILLONACO

La Central Eólica Villonaco de 16.5 MW de potencia, es una central que se ubica en la provincia de Loja, en el cantón de Loja, y se considera como la primera central eólica del Ecuador continental. La central cuenta con 11 aerogeneradores de 1.5 MW cada uno, los cuales entre sus características están de presentar una velocidad promedio anual de 12.7 m/s a una altitud de 2700 msnm [55].

Los datos históricos de viento los manejan en la Central, pero como limitante, no disponen de una herramienta de pronóstico, ya que se rigen a los pronósticos emitidos por parte del INHAMI (Instituto Nacional de Meteorología e Hidrología).

3. METODOLOGÍA

En este capítulo se presenta la metodología propuesta para la construcción del modelo de pronóstico de viento de corto plazo, considerando los métodos de series temporales de SARIMA y Autoregresivos.

El esquema general se presenta en la Figura 3.1.

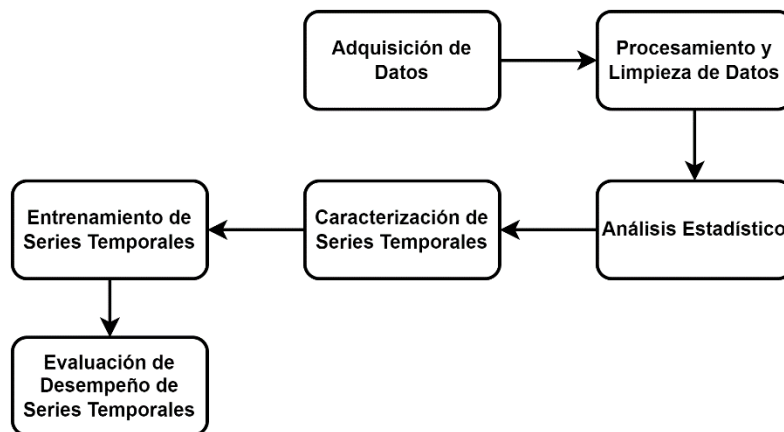


Figura 3.1. Esquema general de la metodología propuesta

La metodología básicamente consiste en un proceso sistemático de seis etapas que parten de la adquisición y procesamiento de datos hasta la caracterización de la serie y su entrenamiento para el pronóstico.

En la Figura 3.1, se observa que el proceso inicia con la adquisición y procesamiento de datos obtenidos desde la estación meteorológica ubicada en campo. Los datos incluyen: estampa de tiempo, velocidad del viento con una ventana de tiempo de cinco años, potencia activa media, rendimiento energético de los aerogeneradores, temperatura ambiente promedio y la dirección del viento.

Con base a esta información se realiza un análisis de estadística descriptiva para determinar el comportamiento de los datos y establecer así parámetros característicos de los mismos, como: la correlación de variables, datos atípicos, valores máximos y mínimos, y la distribución de datos, de manera que se puede tener un panorama de cómo organizar los datos en el dataset, además de definir la correlación existente entre dichas variables.

Una vez determinado el dataset, además de haber definido qué variables se van a considerar dentro de la elaboración de los modelos, en este caso, la velocidad del viento, se procede al desarrollo de las series temporales que sirven de modelos de predicción, los cuales se basan en los métodos de SARIMA y Autorregresivos (AR).

Con relación a los modelos de pronóstico (SARIMA y Autorregresivos), se inicia con la caracterización de los modelos para posteriormente definir la metodología de entrenamiento y evaluación de desempeño. Es decir, para cada uno de los modelos se considera la parametrización necesaria en consecuencia de los datos disponibles, además de la elaboración de los respectivos algoritmos.

Para el entrenamiento y validación del modelo de pronóstico de velocidad del viento se utilizarán datos de la Central Eólica Villonaco cuyas características se describen en la sección 2.10.

A continuación, cada una de estas seis etapas son descritas con un mayor detalle:

3.1. ADQUISICIÓN DE DATOS

El proceso de adquisición de datos permite conocer las características de la información extraída de la estación meteorológica. El proceso inicia con la definición de aspectos tales como las variables que contiene el conjunto de datos, la precisión que presentan los mismos, además de la ventana de tiempo propia del muestreo de datos. Para la adquisición de datos para el modelo es importante en primera instancia definir el horizonte de tiempo y la precisión deseada en relación con los valores de la velocidad del viento. Considerando que los modelos de pronóstico a entrenar son de corto plazo, el horizonte de tiempo que se ha considerado es de 10 minutos para un pronóstico 24 de horas .

Como un ejemplo del proceso, en la Tabla 3.1, se presentan los datos de la Unidad U1 de la central eólica Villonaco.

La información obtenida corresponde a datos de velocidades de viento, temperatura ambiente y potencia activa teórica promedios, las cuales son medidas en un intervalo de tiempo de 10 minutos para un periodo de 5 años, constituyendo un total de 241500 datos hasta noviembre del 2022. Para entrenar el modelo se tomará los datos del último año entre el 29 de febrero de 2020 al 16 de noviembre de 2022 disponible en el dataset con un intervalo de 10 minutos para la predicción de 24 horas y hasta 72 horas, ya que el objetivo del entrenamiento es crear un modelo de pronóstico a corto plazo.

Tabla 3.1. Conjunto de datos referencial del dataset de la Unidad U1 correspondientes al mes de noviembre del 2022

No.	Time	Avg Wind Speed(m/s)	Avg Theory active power(kW)	Avg Active Power (kW)	Wind Turbine Energy yield(kWh)	Avg Ambient Temp(°C)	Avg ABS Wind Direction(deg)
1	2022-11-09 00:00	1.33	4.03	-0.81	0.00	13.23	193.39
2	2022-11-09 00:10	1.12	3.13	-1.51	0.00	13.62	193.21
3	2022-11-09 00:20	1.94	6.99	-0.52	0.00	13.69	196.48
4	2022-11-09 00:30	1.32	4.41	-0.96	0.00	13.73	204.57
5	2022-11-09 00:40	1.31	6.59	-0.97	0.00	13.54	217.48
6	2022-11-09 00:50	0.92	2.39	-0.31	0.00	13.65	201.23
7	2022-11-09 01:00	2.03	5.19	-2.03	0.00	13.68	201.02
8	2022-11-09 01:10	2.70	9.38	-0.64	0.00	13.50	189.72
9	2022-11-09 01:20	3.30	27.78	-1.18	0.00	13.52	195.49
10	2022-11-09 01:30	2.60	9.86	-2.01	0.00	13.57	200.68

En el ANEXO B se muestra el dataset utilizado en el presente trabajo en cuanto a la información correspondiente la central eólica. Se ha considerado limitar el número de datos presentados, dado que la gran cantidad de información disponible dificulta su observación.

3.2. PROCESAMIENTO Y LIMPIEZA DE DATOS

Para iniciar con el procesamiento de la información de los conjuntos de datos, se requiere el uso del software de programación de Python, con su IDLE de Jupyter, bajo el procedimiento mostrado en el diagrama de bloques de la Figura 3.2.

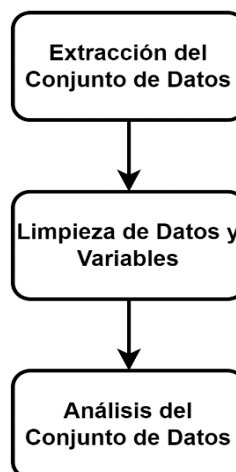


Figura 3.2. Proceso de procesamiento y limpieza de datos

El proceso básicamente sigue 3 etapas:

- *Extracción del conjunto de datos:* En el entorno de Jupyter se extrae cada conjunto de datos asociados a las unidades de la central eólica, de forma que se construyen los datasets a través del módulo de “pandas”.
- *Limpieza de datos y variables:* Se identifican datos anómalos que contiene cada dataset, con base en un proceso de minería de datos que permite establecer la calidad de datos de un dataset.
- *Análisis de conjuntos de datos:* En esta fase se analizan aspectos como el tipo de datos y número de datos que contiene cada dataset, de forma que se tiene un mejor panorama de cómo se dispone los datos, además de definir qué variables contienen información no necesaria para el desarrollo de los modelos.

3.3. ANÁLISIS ESTADÍSTICO

El análisis estadístico arrojará parámetros que constituirán indicadores del comportamiento de los datos por cada una de las variables del conjunto de datos. Para el efecto, se considera el uso de la función “*describe()*” que se encuentra en el módulo de pandas de Python.

A través de la función “*describe()*”, se tiene una breve descripción de la forma estadística de los datos presentes en los datasets; de forma que se dispondrá del siguiente resumen estadístico [56]:

- *Count:* retorna el número de valores no nulos.
- *Mean:* retorna los valores medios de cada variable.
- *Std:* retorna la desviación estándar de cada variable.
- *Min:* retorna los valores mínimos de cada variable.
- *25%:* retorna el percentil 25% de cada variable.
- *50%:* retorna el percentil 50% de cada variable.
- *75%:* retorna el percentil 75% de cada variable.
- *Max:* retorna el valor máximo de cada variable.

Además, se propone el uso de histogramas para conocer la distribución que presenta cada una de las variables, en relación a los valores máximos y mínimos, a través del uso de la función de pandas:

Df.hist()

Donde Df se considera al conjunto de datos o DataFrame.

Los valores máximos y mínimos permiten conocer cómo varía la serie temporal en cuanto a los posibles límites entre los cuales pueden variar los datos.

Un aspecto importante previo a la definición de las variables necesarias para la construcción de los modelos es el cálculo de correlación. En este sentido, el modelo propone un análisis de correlación entre las variables de velocidad de viento, temperatura y potencia, a través de los coeficientes de Pearson, Spearman y Kendall, para lo cual se usarán los siguientes extractos de script:

```
print('Correlación Pearson: ', data_in['Avg Wind Speed(m/s)'].corr(data_in['Avg Ambient Temp(°C)'], method='pearson'))
print('Correlación spearman: ', data_in['Avg Wind Speed(m/s)'].corr(data_in['Avg Ambient Temp(°C)'], method='spearman'))
print('Correlación kendall: ', data_in['Avg Wind Speed(m/s)'].corr(data_in['Avg Ambient Temp(°C)'], method='kendall'))

print('Correlación Pearson: ', data_in['Avg Wind Speed(m/s)'].corr(data_in['Avg Theory active power(kW)'], method='pearson'))
print('Correlación spearman: ', data_in['Avg Wind Speed(m/s)'].corr(data_in['Avg Theory active power(kW)'], method='spearman'))
print('Correlación kendall: ', data_in['Avg Wind Speed(m/s)'].corr(data_in['Avg Theory active power(kW)'], method='kendall'))
```

Figura 3.3. Script de cálculo de coeficientes de correlación

3.4. DISEÑO Y ENTRENAMIENTO DE SERIES TEMPORALES

En esta sección se detalla el procedimiento para el diseño y entrenamiento de las series temporales (SARIMA y AR) que se utilizarán para el pronóstico de velocidades de viento, el cual, parte del tratamiento del conjunto de datos y su procesamiento en series temporales, además de la definición de las características para su implementación, cuyo proceso se explica a detalle a continuación:

3.4.1. Extracción de Datos Temporales

En esta etapa se realiza la transformación de formato de los datos de tiempo, de forma que pueda ser definido como datos temporales que serán usados en la construcción de las series temporales.

Inicialmente se consideran dentro del DataFrame a los valores de tiempo, teniendo en cuenta el formato de *"datetime64[ns]"*, el cual se trata de un formato basado en el módulo de numpy, que proporciona un formato de fecha en una matriz de tipo numérica, basado

en el formato de año-mes-día-hora, considerando una resolución de nanosegundos [ns] de precisión [57].

En este sentido, se considera el siguiente extracto del script de Python para el presente proceso:

```
data_in['date_time'] = data_in['Time'].astype('datetime64[ns]')
```

Figura 3.4. Acceso a datos de tiempo en formato datetime64[ns]

De la Figura 3.4 se tiene la siguiente estructura de datos:

- data_in['date_time']: dataframe con los datos convertidos en estructura de tiempo.
- Data_in['Time']: dataframe con los datos obtenidos en el histórico.
- astype('datetime64[ns]'): función de conversión de tipo de dato.

3.4.2. Creación de Dataframe basado en Temporalidad

La creación del conjunto de datos que tenga como índice la variable de tiempo, se realiza adjuntando a la variable como una columna del dataset y definiendo a los datos temporales como índice. Para esto, se ha utilizado el siguiente extracto de script:

```
data_viento=pd.DataFrame()  
data_viento['Date']=data_in['date_time']  
data_viento['Velocidad']=data_in['Avg Wind Speed(m/s)']
```

```
data_viento=data_viento.set_index(data_viento['Date'])
```

Figura 3.5. Creación de dataframe basado en temporalidad

3.4.3. Descomposición de datos de series temporales en Tendencia y Estacionalidad

Un proceso importante del desarrollo de modelos de pronóstico es el de dividir a una serie de tiempo en componentes, de manera que se pueda definir el mejor ajuste de los modelos, de acuerdo a las características de los modelos, en contraste con los datos empleados en su desarrollo, las cuales pueden ser de tipo sistemáticos y no sistemáticos.

- *Componentes sistemáticos*: se caracterizan por presentar consistencia o recurrencia, además de poder ser descritos y modelados.
- *Componentes no sistemáticos*: estos no pueden ser modelados de forma directa.

En general, una serie de tiempo presenta tres componentes sistemáticos los cuales son: nivel, tendencia y estacionalidad, y un componente no sistemático denominado como ruido:

- *Nivel*: el cual se relaciona con el valor promedio en la serie.
- *Tendencia*: se trata de un valor creciente o decreciente en la serie.
- *Estacionalidad*: se caracteriza por ser un ciclo de corto plazo que tiende a repetirse en la serie.
- *Ruido*: se trata de una variación aleatoria en la serie.

Para el modelo propuesto, la descomposición de la serie temporal se realizará a través del siguiente extracto de script:

```
decomposition = sm.tsa.seasonal_decompose(y, model='additive')
fig = decomposition.plot()
plt.show()
```

Figura 3.6. Descomposición de datos de una serie temporal

3.4.4. Prueba de Dickey Fuller Aumentada (ADF)

Se trata de una prueba de raíz unitaria enfocada a la definición de la estacionalidad de una serie temporal. Una raíz unitaria se trata de una tendencia de tipo estocástica dentro de una serie de tiempo. Se considera que las raíces unitarias pueden causar resultados impredecibles dentro del análisis de series de tiempo.

Dentro de la denominada prueba se tienen las siguientes hipótesis:

- *Hipótesis nula*: la cual se efectúa para definir la existencia de una raíz unitaria.
- *Hipótesis alternativa*: en este tipo, la alternativa básica se considera que la serie de tiempo sea de tipo estacionaria o que presente una tendencia estacionaria.

En el desarrollo de los modelos de las series temporales dentro del presente trabajo, se emplea la Hipótesis nula, para definir la presencia de raíces unitarias dentro de las series temporales.

Previo a la ejecución de la prueba de ADF, es importante que se realice una inspección de los datos. Por ejemplo, si se tiene que la media es distinta de cero, es un indicativo de que la regresión presenta un término constante


```

def tsplot(y, lags=None, figsize=(12, 7), style='bmh'):

    if not isinstance(y, pd.Series):
        y = pd.Series(y)

    with plt.style.context(style):
        fig = plt.figure(figsize=figsize)
        layout = (2, 2)
        ts_ax = plt.subplot2grid(layout, (0, 0), colspan=2)
        acf_ax = plt.subplot2grid(layout, (1, 0))
        pacf_ax = plt.subplot2grid(layout, (1, 1))

        y.plot(ax=ts_ax)
        p_value = sm.tsa.stattools.adfuller(y)[1]
        ts_ax.set_title('Time Series Analysis Plots\nDickey-Fuller: p={0:.5f}'.format(p_value))
        smt.graphics.plot_acf(y, lags=lags, ax=acf_ax)
        smt.graphics.plot_pacf(y, lags=lags, ax=pacf_ax)
        plt.tight_layout()

from statsmodels.tsa.stattools import adfuller
dfctest = adfuller(df2.Velocidad, autolag = 'AIC')
print("1. ADF : ",dfctest[0])
print("2. P-Value : ", dfctest[1])
print("3. Num Of Lags : ", dfctest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calculation :", dfctest[3])
print("5. Critical Values :")
for key, val in dfctest[4].items():
    print("\t",key, ": ", val)

```

Figura 3.7. Extracto de Script de Despliegue de resultados de la prueba de ADF

En la Figura 3.7 se muestra el extracto de código para el análisis de la prueba de ADF, en donde inicialmente, a través de la función `tsplot`, se realiza un análisis gráfico de la serie de tiempo, de manera que despliega, y complementariamente un ejemplo de sus gráficas se muestran en el ejemplo de la Figura 3.10:

- *Gráfico de serie temporal:* se realiza un gráfico de la serie temporal original para tener una vista previa de las características de los datos originales.
- *Gráfico de función de autocorrelación (ACF):* con base a los datos originales, se busca un patrón que indique que la media de los datos no es de tipo estacionaria; de forma muy común se pueden encontrar picos grandes a través de los retrasos, los cuales tienden a extinguirse de forma lenta.
- *Gráfico de función autocorrelación parcial (PACF):* dentro de los datos estacionarios, se busca patrones que indiquen la presencia de términos autorregresivos en un modelo ARIMA, de forma que, si los datos originales no son de tipo estacionarios, se puede usar la gráfica PACF de la serie diferenciada, de forma que se puedan buscar términos candidatos dentro del modelo ARIMA. Una serie diferenciada se la usa para eliminar la tendencia y la estacionalidad de una serie con tendencia estocástica.

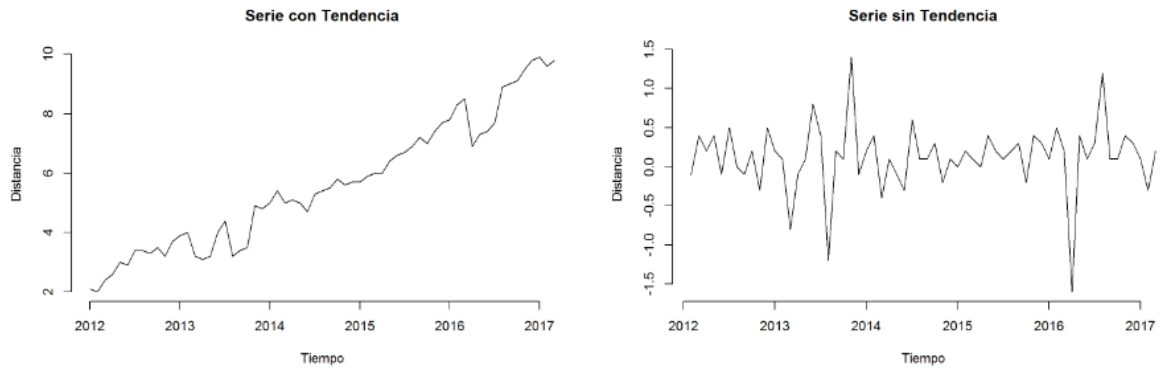


Figura 3.8. Ejemplo de series temporales con Tendencia y sin tendencia

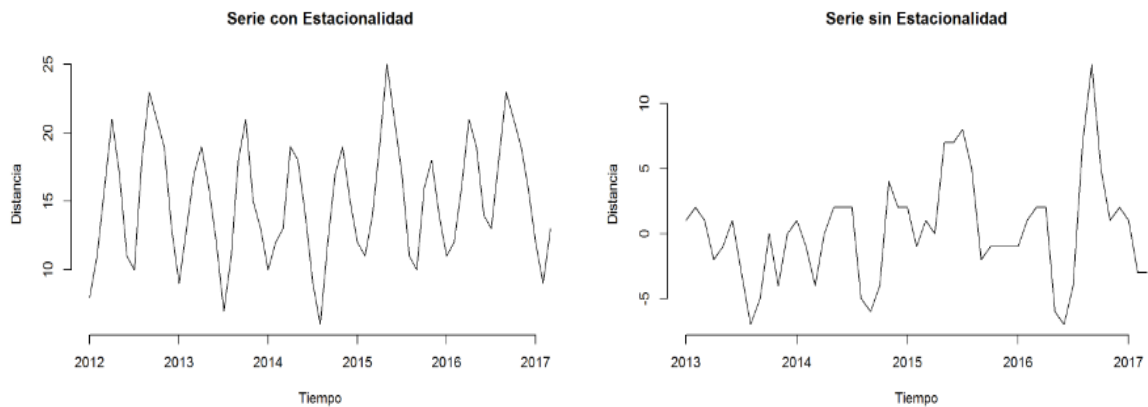


Figura 3.9. Ejemplo de series temporales con estacionalidad y sin estacionalidad

Hay que tener en cuenta los siguientes conceptos: tendencia, se trata de un patrón que indica una media no estacionaria, y por otra parte la diferenciación, es un proceso que permite hacer que la media sea estacionaria.

Dentro de la hipótesis nula hay que considerar lo siguiente:

- $p - value \leq nivel\ de\ significación\ o\ estadística\ de\ prueba \leq valor\ crítico$: en este caso se rechaza la hipótesis nula, esto, debido a que los datos denotan evidencia de que los datos son estacionarios, de forma que se recomienda realizar el análisis sin diferencia alguna.
- $p - value > nivel\ de\ significación\ o\ estadística\ de\ prueba > valor\ crítico$: en este caso no se rechaza a la hipótesis nula, teniendo en cuenta de que los datos no proporcionan evidencia de que los datos son estacionarios, de forma que se recomienda realizar el análisis sin una diferenciación que hará que la media de los datos sea estacionaria.

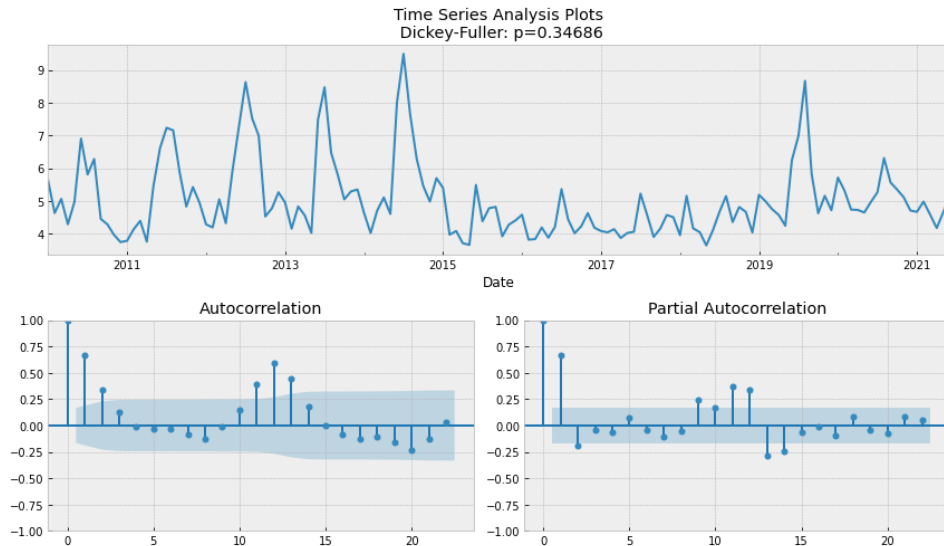


Figura 3.10. Ejemplo de análisis gráfico de la prueba de ADF

El proceso de diferenciación de series se realiza a través del siguiente extracto de script (Figura 3.11):

```
ts_t_adj = y - y.shift(1)
ts_t_adj = ts_t_adj.dropna()
ts_t_adj.plot()

print(adf_test(y))
```

FIGURA 3.11. Script para la diferenciación de series

Considerando el despliegue de resultados de la prueba de ADF del script de la Figura 3.11, se tiene la siguiente información [58]:

- *ADF*: valor resultante de la estadística de prueba.
- *P-value*: valor de p aproximado
- *Número de lags*: el número de retrasos utilizados.
- *Número de observaciones*: se trata del número de observaciones utilizadas para la regresión ADF y el cálculo de los valores críticos.
- *Valores críticos*: valores críticos para la estadística de prueba en los niveles de 1, 5 y 10 %.

3.4.5. Creación de modelos de pronósticos

Posterior a la definición de las características estacionales y tendencia de la serie temporal de los datos originales del conjunto de datos, se procede a definir el proceso y características de los modelos de las series temporales de SARIMA y AR.

3.4.5.1. MODELO SARIMA

Inicialmente se realiza una selección de los parámetros (p, d, q) para el modelo de la serie temporal. Al realizar la búsqueda del ajuste de los datos de las series temporales basado en un modelo ARIMA estacional (SARIMA), como primer objetivo, se tiene el de encontrar los valores que permitan optimizar una determinada métrica de interés.

En este sentido, existen diversas pautas y prácticas para que se logre dicho objetivo, de forma que una adecuada parametrización de los modelos ARIMA se puede convertir en un proceso manual y laborioso, el cual conlleva en cierta forma de tiempo de procesamiento.

Para el presente trabajo se propone el uso de la denominada búsqueda en cuadrícula, la cual permite la exploración de forma iterativa de diferentes combinaciones de parámetros. Cada una de estas combinaciones se ajusta con un nuevo modelo SARIMA con la función del módulo y se evalúa su calidad en general. Posterior a que se haya hecho una exploración de todo el panorama de parámetros, el conjunto óptimo de parámetros se determinará en base al mejor rendimiento en función a los criterios de interés definidos.

A través extracto de script de la Figura 3.12 se realiza la resolución de dicho problema, en donde se selecciona los valores de los parámetros óptimos para el modelo de la serie temporal: p, d, q.

```
# Define the p, d and q parameters to take any value between 0 and 2
p = d = q = range(0, 2)

# Generate all different combinations of p, q and q triplets
pdq = list(itertools.product(p, d, q))

# Generate all different combinations of seasonal p, q and q triplets
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]

print('Examples of parameter combinations for Seasonal ARIMA...')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[4]))
```

Figura 3.12. Extracto de script de selección de parámetros de modelo ARIMA

A continuación, se realiza una iteración a través de las combinaciones de parámetros, de forma que se va ajustando el modelo ARIMA. En el argumento “order” se especifican los

parámetros (p, d, q), mientras que en el argumento “seasonal_order” se especifica al componente estacional del modelo SARIMA.

Como parámetro de evaluación se usa el criterio de información de Akaike (AIC), el cual define una medida del ajuste del modelo de datos. De forma que si un modelo tiene un buen ajuste (obtención de los mejores parámetros) a los datos mientras se usa muchas características, se le asigna una puntuación AIC mayor que un modelo en el cual se usa menores características para que se logre la misma relación de ajuste. En este sentido, el objetivo será buscar al modelo que arroje el AIC más bajo.

Luego del ajuste del modelo, se procede a imprimir los parámetros de (p, d, q).

```
warnings.filterwarnings("ignore") # specify to ignore warning messages

for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.seasonality.statespace.SARIMAX(y,
                                                    order=param,
                                                    seasonal_order=param_seasonal,
                                                    enforce_stationarity=False,
                                                    enforce_invertibility=False)

            results = mod.fit()

            print('ARIMA{}x{}12 - AIC:{}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

Figura 3.13. Extracto de script para el ajuste de modelo a ARIMA

Posteriormente, se realiza una evaluación gráfica del modelo, de forma que se utiliza el siguiente extracto de código:

```
results.plot_diagnostics(figsize=(15, 12))
plt.show()
```

Figura 3.14. Gráfico de diagnóstico de modelo SARIMA

Los gráficos de análisis que la función “plot_diagnostics” arroja son [59], y se muestran en la Figura 3.15:

- *Gráfico de residuos estandarizados:* se trata de una medida de la diferencia entre los valores observados y esperados. En general, se trata de una medida de la significancia de los datos en relación con el valor de chi-cuadrado. Al realizar la comparación de datos, el residuo estandarizado permite definir qué datos tienen mayor o menor contribución a los valores.

Para interpretar al gráfico de residuos estandarizados se tiene los siguientes criterios:

1. Si el residuo es menor que -2, la frecuencia observada del dato es menor que la frecuencia esperada.
 2. Si el residuo es mayor que 2, la frecuencia observada es mayor que la frecuencia esperada.
- *Gráfico de Histograma de densidad estimada de residuos normalizados*: la gráfica de densidad revela la distribución de datos relacionados con los residuos con una media de cero.
 - *Gráfica normalizada Q-Q (cuantil-cuantil)*: los datos distribuidos de forma normal aparecen como una línea recta, aunque de forma general, los extremos del diagrama Q-Q en cierta forma comienzan a desviarse de la línea recta. Los puntos en el gráfico Q-Q dan un indicativo de la normalidad univariante del conjunto de datos, de forma que, si los datos se distribuyen normalmente, los puntos tienden a caer en la línea de referencia de 45 grados, por otra parte, si no presentan una distribución normal, los puntos se desvían de la línea de referencia.
 - *Correlograma*: se trata de un gráfico de coeficientes de autocorrelación de orden k frente al retardo k.

Frente a lo indicado, hay que recalcar que se debe de definir un horizonte de tiempo y una precisión de los datos, de manera que, por ejemplo, para el caso de la Figura 3.13 se tiene un horizonte de datos de 8 años con una precisión de 15 minutos, para un pronóstico de corto plazo, como el propuesto en este trabajo.

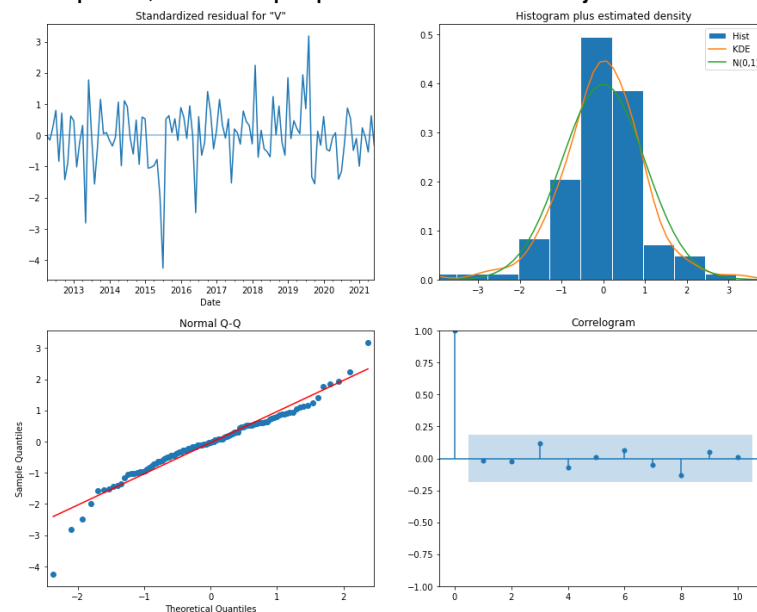


Figura 3.15. Ejemplo de gráficos de diagnóstico de modelo SARIMA

A través de la función “get_prediction()”, se tiene la instancia de resultados de predicción, la cual contiene la predicción y la varianza de la misma, de forma que también puede calcular intervalos de confianza y tablas de resumen bajo demanda para la predicción de la media y de nuevas observaciones. En el extracto de código de la Figura 3.16 se muestra la función de obtención de pronósticos.

```

pred = results.get_prediction(start=pd.to_datetime('2019-12-01'), dynamic=True)
pred_ci = pred.conf_int()

ax = y['2021:'].plot(label='observed')
pred.predicted_mean.plot(ax=ax, label='One-step ahead Forecast', alpha=.7, figsize=(14, 7))
ax.fill_between(pred_ci.index,
                pred_ci.iloc[:, 0],
                pred_ci.iloc[:, 1], color='k', alpha=.2)
ax.set_xlabel('Date')
ax.set_ylabel('Velocidad del viento')
plt.legend()
plt.show()

```

Figura 3.16. Extracto de script para la elaboración de pronósticos

Para evaluar el rendimiento del modelo se utiliza el indicador del error cuadrático medio, el cual se encuentra implementado en el extracto de código de la Figura 3.17.

```

error_mse = mean_squared_error(
    y_true = y['2019-12-01:'],
    y_pred = pred.predicted_mean
)
print(f"Error de test (mse): {error_mse}")

```

Figura 3.17. Extracto de script para el cálculo del error cuadrático medio

3.4.5.2. MODELO AUTOREGRESIVO RECURSIVO

En esta sección se define el procedimiento para el desarrollo de la serie temporal basada en el modelo Autoregresivo, de forma que se detallan cada uno de los pasos necesarios para su parametrización, entrenamiento y pruebas.

Inicialmente, se definen las variables de entrenamiento y de prueba, como se muestra en la Figura 3.18. Otro aspecto que se define es el parámetro “step”, el cual se relaciona con cada paso de predicción.

```

steps = 15
datos_train = y[:-steps]
datos_test = y[-steps:]

```

Figura 3.18. Definición de variables de entrenamiento y de prueba

Se crea y se entrena el modelo basado en la función “ForecasterAutoreg” a partir de un regresor del tipo “RandomForestRegressor”, y una ventana temporal de n lags. El número de lags definirá la ventana temporal utilizada como predictores.

```
forecaster_rf = ForecasterAutoreg(regressor=RandomForestRegressor(random_state=123),lags=45)
forecaster_rf.fit(y=datos_train)
forecaster_rf
```

Figura 3.19. Creación y entrenamiento del modelo de pronóstico autorregresivo recursivo

Una vez que se dispone del modelo entrenado, se realiza la predicción a partir de los datos de prueba, considerando un número n de steps (número de pasos a futuro), como se muestra en el extracto de script de la Figura 3.20.

```
steps = 15
predicciones = forecaster_rf.predict(steps=steps)

predicciones = pd.Series(data=predicciones, index=datos_test.index)
predicciones.head()

fig, ax = plt.subplots(figsize=(9, 4))
datos_train.plot(ax=ax, label='train')
datos_test.plot(ax=ax, label='test')
predicciones.plot(ax=ax, label='predicciones')
ax.legend();
```

Figura 3.20. Extracto de script para el cálculo de pronósticos del modelo autorregresivo recursivo

Finalmente, se define la función del cálculo de error, basado en el error cuadrático medio, como se muestra en el extracto de script de la Figura 3.21.

```
error_mse = mean_squared_error(
    y_true = datos_test,
    y_pred = predicciones
)
print(f"Error de test (mse): {error_mse}")
```

Figura 3.21. Extracto de script para el cálculo del error del modelo de pronósticos autorregresivo recursivo

3.4.5.3. MODELO AUTOREGRESIVO RECURSIVO CON PREDICTORES MODIFICADOS

Considerando determinados escenarios (por minutos, horas o días), se puede hacer la incorporación de otras características de la serie temporal a las ya disponibles como los lags; por ejemplo, se puede definir la media móvil de los últimos n valores que puede ser de utilidad para la captura de la tendencia de la serie. El uso de esta característica está en que se puede usar de forma recurrente en el comportamiento del viento en la Central Eólica, de forma que se busca una mejor aproximación en cuanto al modelo de pronóstico.

Para esta finalidad se utiliza la función “ForecasterAutoregCustom”, la cual tiene similitud a la función “ForecasterAutoreg”, pero con la diferencia de que se puede definir la función a emplearse en la creación de los predictores.

En el extracto de script de la Figura 3.22 se muestra inicialmente la definición de los datos de entrenamiento y prueba, además del número de steps a utilizarse.

La creación y el entrenamiento del modelo se realiza a través del uso de un regresor “RandomForestRegressor”. Este regresor se trata de un metaestimador, el cual ajusta una serie de árboles de decisión de clasificación en diferentes submuestras del conjunto de datos, además, utiliza el promedio para que se mejore la precisión predictiva y se controle el sobreajuste [60]. Por otra parte, para la creación de los predictores, se emplea la función “create_predictors()”, la cual calcula los primeros n-número de lags y la media móvil de los últimos m-número de valores.

Al crear el modelo de pronóstico, el argumento window_size debe ser un valor, como mínimo, tan grande en relación con la ventana que se usa en la función de creación de los predictores.

```
steps = 15
datos_train = y[:-steps]
datos_test = y[-steps:]

def create_predictors(y):
    """
    Crear los primeros 10 lags.
    Calcular la media móvil de los últimos 20 valores.
    """
    X_train = pd.DataFrame({'y':y.copy()})
    for i in range(0, 40):
        X_train[f'lag_{i+1}'] = X_train['y'].shift(i)
    X_train['moving_avg'] = X_train['y'].rolling(20).mean()
    X_train = X_train.drop(columns='y').tail(1).to_numpy()
    return X_train

forecaster_rf = ForecasterAutoregCustom(
    regressor = RandomForestRegressor(random_state=123),
    fun_predictors = create_predictors,
    window_size = 40
)

forecaster_rf.fit(y=datos_train)
forecaster_rf
```

Figura 3.22. Creación y entrenamiento de modelo de pronóstico autorregresivo recursivo con predictores modificados

En el extracto de script de la Figura 3.23 se muestra cómo se realiza las pruebas de pronóstico del modelo entrenado y su respectiva función de gráfica.

```

steps = 15
predicciones = forecaster_rf.predict(steps=steps)
# Se añade el índice temporal a las predicciones
predicciones = pd.Series(data=predicciones, index=datos_test.index)

```

```

fig, ax = plt.subplots(figsize=(9, 4))
datos_train.plot(ax=ax, label='train')
datos_test.plot(ax=ax, label='test')
predicciones.plot(ax=ax, label='predicciones')
ax.legend();

```

Figura 3.23. Pruebas de modelo de pronóstico autorregresivo recursivo con predictores modificados

Finalmente, se dispone del cálculo de los errores del modelo de pronóstico, para lo cual se usa el error cuadrático medio, de forma que se considera el extracto de código de la Figura 3.24.

```

error_mse = mean_squared_error(
    y_true = datos_test,
    y_pred = predicciones
)
print(f"Error de test (mse): {error_mse}")

```

Figura 3.24. Cálculo del error cuadrático medio del modelo de pronóstico autorregresivo recursivo con predictores modificados

4. RESULTADOS Y DISCUSIÓN

En este capítulo se desarrolla el modelo de pronóstico del viento propuesto bajo la metodología descrita en la Sección 3. Para la validación del modelo se ha considerado como caso de estudio a la Central Eólica Villonaco ubicada en la provincia de Loja. La central tiene una potencia de 16.5 MW, con 11 aerogeneradores de 1.5 MW cada uno. La localidad en la cual se encuentra instalada la central presenta una velocidad promedio anual de 12.7 m/s a una altitud de 2700 msnm.

Los criterios y condiciones que se han considerado para el pronóstico son los siguientes:

- Datos históricos meteorológicos de la central eólica Villonaco.
- Procesamiento y limpieza de la información dentro del conjunto de datos.
- Ventana de tiempo de un año de datos históricos y precisión de datos de 10 minutos para el entrenamiento y validación de los modelos de pronóstico.
- Cálculo de Índices de rendimiento basados en el error.
- Entrenamiento y la validación de los modelos para una ventana de tiempo de hasta 72 horas con una precisión mínima de 10 minutos y hasta 180 minutos

Los resultados se analizan para los siguientes modelos de pronóstico que en la Sección 3 se proponen como parte de la metodología:

- SARIMA
- Autoregresivo
- Autoregresivo con predictores modificados

Una vez definida la fuente de datos y los modelos de pronóstico empleados, a continuación, se van detallando y analizando cada una de las etapas de la Metodología propuesta en la Sección 3:

4.1. PROCESAMIENTO Y LIMPIEZA DE DATOS

Para este trabajo se entregaron datos meteorológicos de la Central Eólica Villonaco comprendidos entre el 29 de febrero de 2020 al 16 de noviembre de 2022; de estas fechas se tomarán 87,600 datos correspondientes a un año, con una precisión de 10 minutos, para el entrenamiento de los modelos descritos en la Sección 3.

Los datos que se usarán corresponden a datos de velocidad del viento, temperatura y potencia activa de la Unidad 1 de la central. Teniendo en cuenta que el modelo de pronóstico está orientado a la velocidad del viento, se realizará un análisis de correlación directa entre estas variables, para verificar su dependencia dentro del modelo. Es importante considerar que se realizarán pronósticos a corto plazo que sean menores o igual a la ventana de tiempo considerados en el entrenamiento de las series temporales

El conjunto de datos se importa al entorno de Jupyter. En la Figura 4.1 se muestran los primeros cinco valores por cada una de las variables del conjunto de datos.

No.	Time	Avg Wind Speed(m/s)	Avg Theory active power(kW)	Avg Active Power (kW)	Wind Turbine Energy yield(kWh)	Avg Ambient Temp(°C)	Avg ABS Wind Direction(deg)	
0	240433	2022-11-09 00:00:00	1.33	4.03	-0.81	0	13.23	193.39
1	240434	2022-11-09 00:10:00	1.12	3.13	-1.51	0	13.62	193.21
2	240435	2022-11-09 00:20:00	1.94	6.99	-0.52	0	13.69	196.48
3	240436	2022-11-09 00:30:00	1.32	4.41	-0.96	0	13.73	204.57
4	240437	2022-11-09 00:40:00	1.31	6.59	-0.97	0	13.54	217.48

Figura 4.1. Detalle de extracción de datos de la U1 en el entorno de Jupyter

Posteriormente, se procede a identificar las características de cada una de las variables disponibles en el conjunto de datos, de manera que, entre los aspectos más importantes a tomar en cuenta, es la presencia de datos nulos.

Como se observa en la Figura 4.2, cada una de las variables no presentan valores nulos, además de tener un tipo de datos específica, diferenciándose aquellos que son de tipo temporal, como son los datos de las fechas

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 864 entries, 0 to 863
Data columns (total 8 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   No.                                         864 non-null    int64
1   Time                                       864 non-null    datetime64[ns]
2   Avg Wind Speed(m/s)                       864 non-null    float64
3   Avg Theory active power(kW)               864 non-null    float64
4   Avg Active Power (kW)                     864 non-null    float64
5   Wind Turbine Energy yield(kWh)            864 non-null    int64
6   Avg Ambient Temp(°C)                       864 non-null    float64
7   Avg ABS Wind Direction(deg)                864 non-null    float64
dtypes: datetime64[ns](1), float64(5), int64(2)
memory usage: 54.1 KB
```

Figura 4.2. Información detallada del conjunto de datos de la U1

Una vez procesada la información, es posible la extracción de cada una de las variables que van a ser consideradas dentro del análisis estadístico y de correlación con la finalidad de determinar las variables que van a ser consideradas como entradas a los modelos de pronóstico. Un ejemplo del proceso mencionado se presenta en la Figura 4.3.

```
Velocidad del viento: 0      1.33
1      1.12
2      1.94
3      1.32
4      1.31
...
859    7.58
860    6.70
861    8.53
862    8.33
863    8.55
Name: Avg Wind Speed(m/s), Length: 864, dtype: float64
```

Figura 4.3. Despliegue de los datos de velocidad del viento

4.2. ANÁLISIS ESTADÍSTICO

Para el análisis estadístico se considera las variables de velocidad del viento, potencia activa y temperatura ambiente, extraídos de forma individual del conjunto de datos. La finalidad de este análisis es determinar las características de los datos en función de sus valores estadísticos, para posteriormente realizar un análisis de correlación entre las variables y determinar la posible dependencia entre variables, de forma que se pueda descartar o considerar tales variables como datos de entrada a los modelos de pronóstico.

Para el caso de la velocidad del viento, de la Figura 4.4 se observa que su valor medio es de 4.631 m/s, teniendo una desviación estándar de 2.501m/s, un valor mínimo de 0.13m/s y máximo de 10.93 m/s.

```

count      864.000000
mean       4.631030
std        2.501799
min        0.130000
25%        2.710000
50%        4.170000
75%        6.700000
max        10.930000
Name: Avg Wind Speed(m/s), dtype: float64

```

Figura 4.4. Resultados del análisis estadístico de la velocidad del viento

La Figura 4.5 muestra que la potencia activa tiene un valor medio de 174 kW, con una desviación estándar de 223.45 kW, un valor mínimo de 0 y máximo de 1028.44 kW.

```

count      864.000000
mean       174.620671
std        223.450762
min        0.000000
25%        12.782500
50%        58.545000
75%        289.057500
max        1028.440000
Name: Avg Theory active power(kW), dtype: float64

```

Figura 4.5. Resultados de análisis estadístico de la potencia generada

Finalmente, para el caso de la Temperatura, se tiene un valor medio de 14.52 °C, con una desviación estándar de 2.234 °C, un valor mínimo de 11.07 y máximo de 21 °C, tal como se observa en la Figura 4.6.

```

count      864.000000
mean       14.520509
std        2.234381
min        11.070000
25%        12.830000
50%        13.835000
75%        16.092500
max        21.000000
Name: Avg Ambient Temp(°C), dtype: float64

```

Figura 4.6. Resultados de análisis estadístico de la temperatura ambiente

El histograma de la Figura 4.7 muestra que la distribución de datos se asemeja a la distribución de Weibull alrededor de su valor medio (14.52 m/s), teniendo valores cercanos al valor mínimo con una notable frecuencia, y con valores mayores al valor máximo (12.5 m/s) que sucede de forma casi no tan frecuente. Tal distribución es congruente con el

comportamiento estadístico del viento utilizado para estudios de planificación y de operación de parques eólicos.

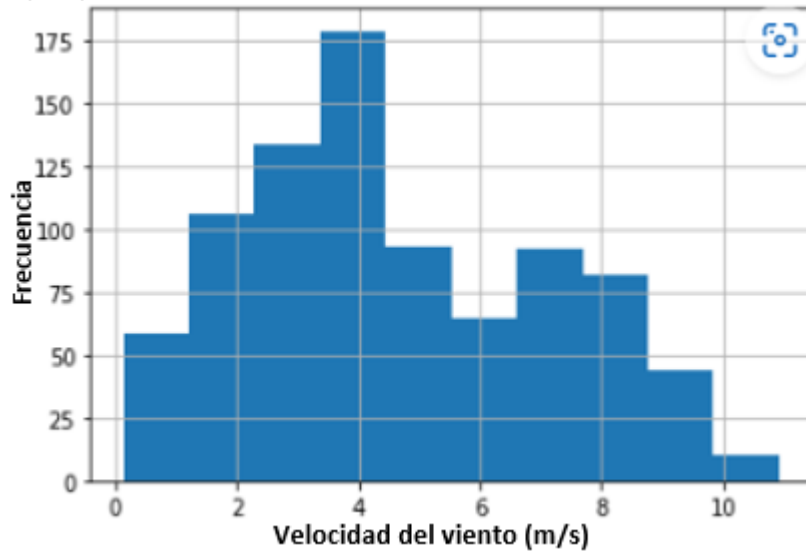


Figura 4.7. Histograma de datos de velocidad del viento en m/s

En el caso del histograma de la Figura 4.8, relacionado con la potencia activa, se observa que en gran medida se tiene valores cercanos a 0, teniendo una menor frecuencia hacia valores que superen o lleguen al valor máximo.

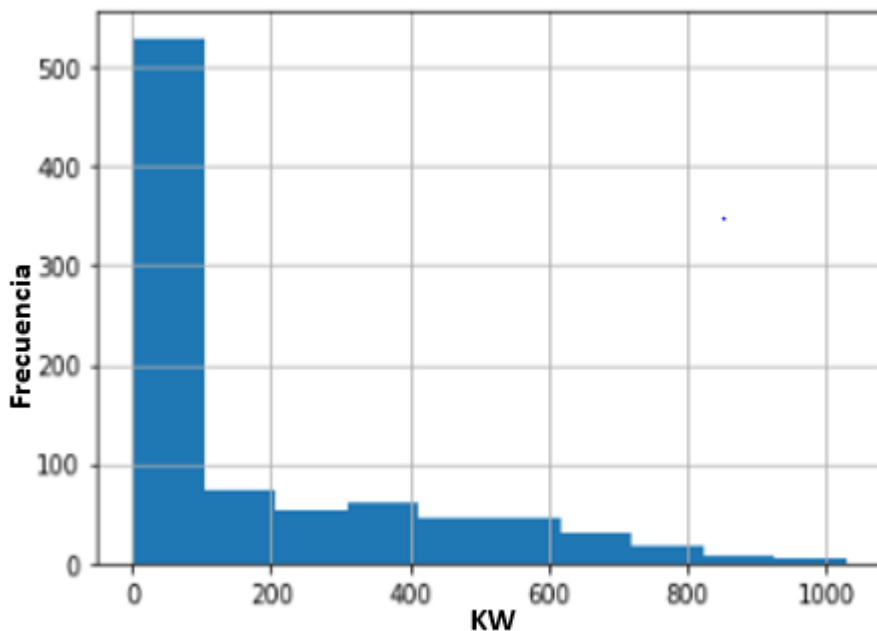


Figura 4.8. Histograma de datos de potencia generada en kW

El histograma de la Figura 4.9, muestra que en gran medida los valores de temperatura se encuentran entre el valor medio y mínimo, aunque, por otra parte, en forma uniforme, se distribuye el comportamiento de la temperatura hacia el valor máximo.

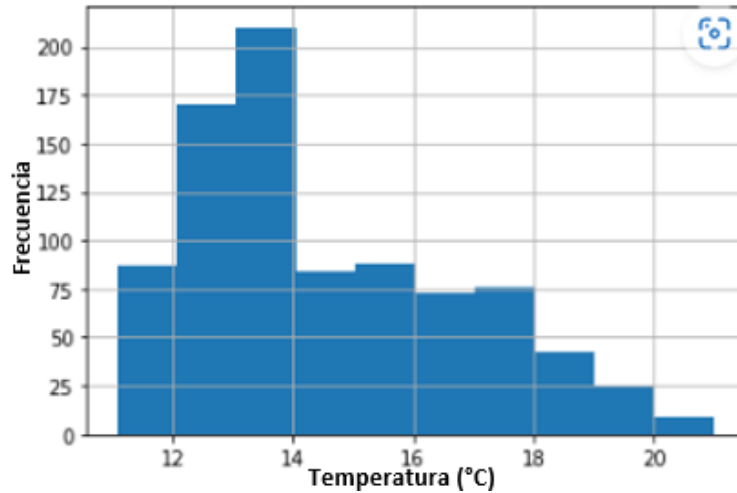


Figura 4.9. Histograma de datos de temperatura en °C

Para definir la relación y dependencia entre las variables se ha considerado el cálculo de los coeficientes de correlación.

Para el caso de la velocidad del viento y la potencia activa, se observa que se tiene una alta correlación, lo cual es de esperarse considerando la naturaleza de la central.

```
Correlación Pearson: 0.9269815233378679
Correlación spearman: 0.9968577858427271
Correlación kendall: 0.9618699630642817
```

Figura 4.10. Índices de correlación entre velocidad del viento y potencia activa

Para el caso de la velocidad del viento y la temperatura, se puede observar que se presenta una baja correlación, muy cercana a cero, lo cual es un indicativo de que no existe una dependencia entre estas variables, de manera que se puede hacer una modelación sin considerar estas variables en conjunto.

```
Correlación Pearson: 0.3040051541615946
Correlación spearman: 0.29619140564720536
Correlación kendall: 0.18932106497282594
```

Figura 4.11. Índices de correlación entre velocidad del viento y temperatura

En conclusión, la variable de temperatura no define el comportamiento de la velocidad de viento y consecuentemente no tiene influencia en la producción del parque, de forma que se la descarta dentro del modelo de datos para el entrenamiento del modelo de pronóstico.

4.3. ENTRENAMIENTO DE SERIES TEMPORALES

En esta sección se describe el proceso para el entrenamiento de los modelos de series temporales SARIMA y modelos Autorregresivos, teniendo en cuenta los datos proporcionados por la Central Eólica Villonaco, además de los procedimientos definidos en la sección 3.

El conjunto de datos de velocidad de viento comprende datos desde el 29 de febrero de 2020 al 16 de noviembre de 2022, de estos datos se tomarán 87,600 datos correspondientes a un año, con una precisión de 10 minutos. El entrenamiento y la validación de los modelos de pronóstico de las series temporales se realizará utilizando una ventana de tiempo de hasta 72 horas con una precisión mínima de 10 minutos y hasta 180 minutos, de manera que se emplea un remuestreo de los datos. El entrenamiento tiene como objetivo tomar la mayor parte de características de las series temporales tales como: valores pico o del comportamiento propio de la velocidad del viento a lo largo del tiempo.

Vale recalcar que el objetivo de esta propuesta es el entrenamiento de los modelos de pronóstico de viento a corto plazo con capacidad de pronóstico de un horizonte de 24 a 72 horas, el cual se validará a través de los casos de estudio en cada modelo.

Además, se ha considerado al error cuadrático medio como índice de evaluación del desempeño, con el fin de definir cuál es el modelo que mejor ajuste tiene de acuerdo con las características de los datos de velocidad de viento.

4.3.1. Modelo SARIMA

Para la disposición de datos de tipo temporal, se consideran los valores horarios presentes en el conjunto de datos, de manera que se pueda considerar su formato relacionado con la estampa de tiempo, tal como se muestra en la Figura 4.12, en donde se ha determinado el tipo de dato que contienen los valores de temporalidad.


```

DatetimeIndex(['2022-11-09 00:00:00', '2022-11-09 03:00:00',
              '2022-11-09 06:00:00', '2022-11-09 09:00:00',
              '2022-11-09 12:00:00', '2022-11-09 15:00:00',
              '2022-11-09 18:00:00', '2022-11-09 21:00:00',
              '2022-11-10 00:00:00', '2022-11-10 03:00:00',
              '2022-11-10 06:00:00', '2022-11-10 09:00:00',
              '2022-11-10 12:00:00', '2022-11-10 15:00:00',
              '2022-11-10 18:00:00', '2022-11-10 21:00:00',
              '2022-11-11 00:00:00', '2022-11-11 03:00:00',
              '2022-11-11 06:00:00', '2022-11-11 09:00:00',
              '2022-11-11 12:00:00', '2022-11-11 15:00:00',
              '2022-11-11 18:00:00', '2022-11-11 21:00:00',
              '2022-11-12 00:00:00', '2022-11-12 03:00:00',
              '2022-11-12 06:00:00', '2022-11-12 09:00:00',
              '2022-11-12 12:00:00', '2022-11-12 15:00:00',
              '2022-11-12 18:00:00', '2022-11-12 21:00:00',
              '2022-11-13 00:00:00', '2022-11-13 03:00:00',
              '2022-11-13 06:00:00', '2022-11-13 09:00:00',
              '2022-11-13 12:00:00', '2022-11-13 15:00:00',
              '2022-11-13 18:00:00', '2022-11-13 21:00:00',
              '2022-11-14 00:00:00', '2022-11-14 03:00:00',
              '2022-11-14 06:00:00', '2022-11-14 09:00:00',
              '2022-11-14 12:00:00', '2022-11-14 15:00:00',
              '2022-11-14 18:00:00', '2022-11-14 21:00:00'],
              dtype='datetime64[ns]', name='Date', freq='180T')

```

Figura 4.12. Conversión de conjunto de datos con temporalidad

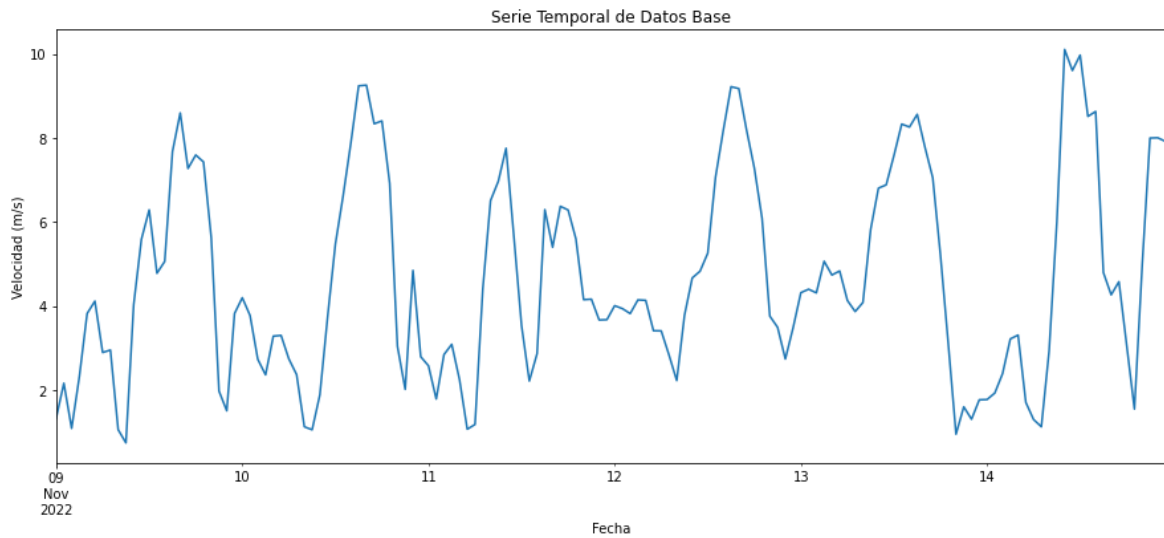


Figura 4.13. Comportamiento Histórico de los valores del Dataset

Teniendo en cuenta la variable anterior, se procede a la creación de la serie temporal de los datos históricos, para lo cual se procedió a realizar un remuestreo de datos, cambiando a una precisión de 180 minutos, de manera que se tenga una breve visión del comportamiento de los valores en el tiempo, tal como se muestra en la Figura 4.13.

En la Figura 4.14 se despliega los resultados gráficos de la prueba de Dickey Fuller. Lo cual como se explicó en la sección 3.4.4 permite definir la estacionalidad de la serie

temporal. Inicialmente se obtiene la vista previa de la serie temporal de los datos base (Figura 4.14-a).

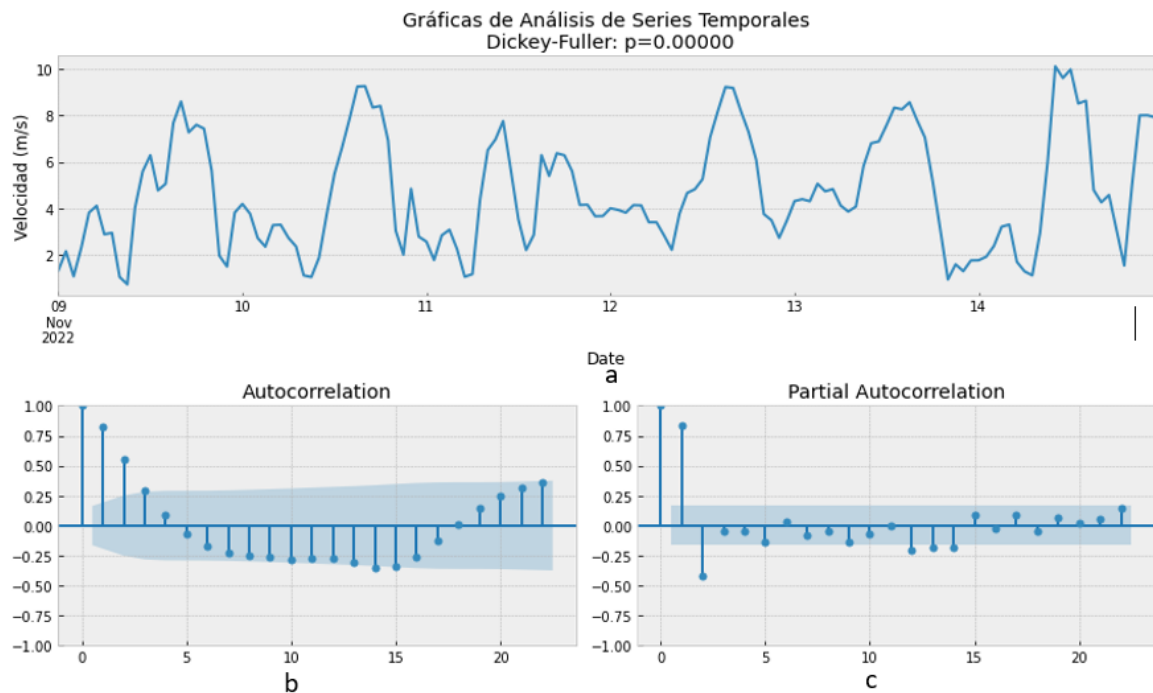


Figura 4.14. a. Prueba de Dickey-Fuller para el conjunto de datos de la U1; **b.** gráfico de autocorrelación del conjunto de datos; **c.** gráfico de autocorrelación parcial del conjunto de datos

Un aspecto importante dentro de este análisis es la definición de los lags, los cuales se convierten en un parámetro empleado en el entrenamiento del modelo de pronóstico de SARIMA. Los lags se determina definiendo los últimos N periodos de la serie temporal, de tal manera que hay que considerar el horizonte y precisión de tiempo que se está considerando en la aplicación.

Del gráfico de autocorrelación (Figura 4.14-b), se observa que los primeros y últimos lags tienen una correlación positiva, de forma que se puede determinar qué valores altos de velocidad corresponden con valores altos en el lag especificado. Por otra parte, los lags con valores de correlación negativa, indica que los valores altos de velocidad son correspondientes con valores pequeños en el lag especificado. Teniendo en cuenta lo indicado, se puede encontrar dentro de los patrones de la serie temporal que se encontrarán valores picos de velocidad con una tendencia lenta de extinción.

En el caso de la gráfica de autocorrelación parcial (Figura 4.14-c), se muestra que se tiene correlaciones significativas en el primer y segundo lag, y se observa correlaciones que en

cierta manera no son significativas, lo que da un indicativo de que existe un término autoregresivo en los datos.

En la Figura 4.15, se muestra la descomposición de la serie temporal, se observa que de acuerdo a los datos obtenidos, la gráfica de estacionalidad es nula. En cuanto a los valores de residuo presentan valores menores a 0.01, lo cual muestra la presencia reducida de valores aleatorios en la serie. En cuanto a la gráfica de tendencia, se tiene un comportamiento consecuente de la serie temporal original.

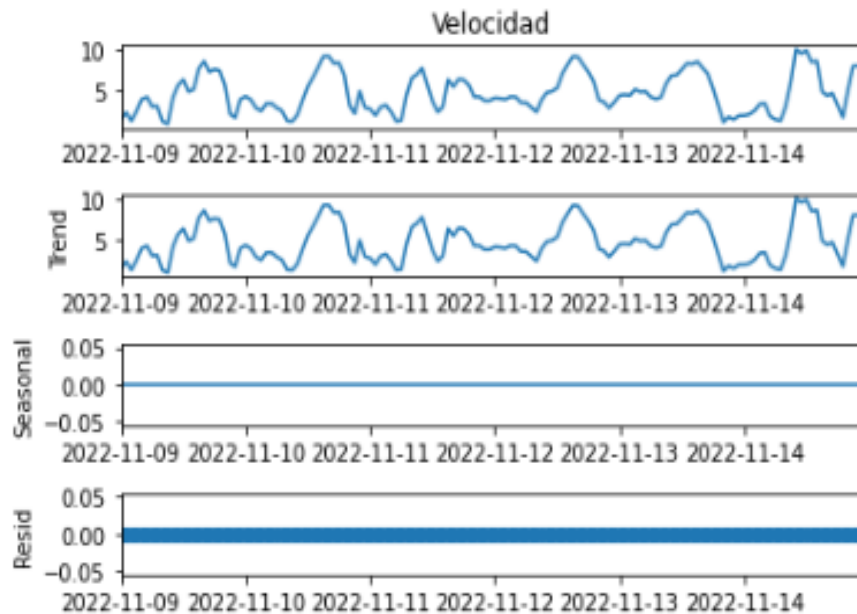


Figura 4.15. Descomposición de datos de la serie temporal del conjunto de datos original

De la Figura 4.16, se tiene que el valor de la prueba estadística es -5.695 aproximadamente, el p-value es 0, lo cual indica que la hipótesis nula es falsa, de forma que no se puede descartar de que la serie temporal sea estacionaria.

Por otra parte, considerando que el valor de prueba es menor que los valores críticos, se tiene que en la serie temporal no existe una raíz unitaria.

```
Results of Dickey-Fuller Test:
Test Statistic          -5.695712e+00
p-value                 7.883120e-07
#Lags Used              1.300000e+01
Number of Observations Used  1.300000e+02
Critical Value (1%)     -3.481682e+00
Critical Value (5%)    -2.884042e+00
Critical Value (10%)   -2.578770e+00
dtype: float64
None
```

Figura 4.16. Resultados de la prueba de Dickey-Fuller del conjunto de datos

En la Figura 4.17 se muestra los resultados de las combinaciones realizadas para la parametrización del modelo SARIMA.

```

Examples of parameter combinations for Seasonal ARIMA...
SARIMAX: (0, 0, 1) x (0, 0, 1, 12)
SARIMAX: (0, 0, 1) x (0, 1, 0, 12)
SARIMAX: (0, 1, 0) x (0, 1, 1, 12)
SARIMAX: (0, 1, 0) x (1, 0, 0, 12)

```

Figura 4.17. Resultados de las combinaciones de parámetros p, d y q, para el modelo SARIMA

Del modelo de SARIMA desarrollado, se tienen los siguientes valores:

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.1486	0.306	0.486	0.627	-0.451	0.748
ma.L1	0.1588	0.320	0.496	0.620	-0.468	0.786
ar.S.L12	-0.3167	0.155	-2.047	0.041	-0.620	-0.013
ma.S.L12	-0.4555	0.173	-2.634	0.008	-0.794	-0.117
sigma2	1.8869	0.179	10.553	0.000	1.536	2.237

Figura 4.18. Análisis de principales coeficientes del modelo SARIMA

Para el caso de los valores de $P > |z|$, se muestra que la serie temporal es estacional.

En la Figura 4.19-a de residuos estandarizados, se observa que los valores en mayor parte oscilan dentro de ± 2 , teniendo valores que se aproximan a valores de 3, en este sentido se tienen frecuencias observadas de esos datos que son menores que la frecuencia esperada, mayores que la misma, de forma respectiva, lo cual se complementa y verifica con el histograma de residuos (Figura 4.19-b). En el caso de la Figura 4.19-c correspondiente al gráfico QQ, se observa que los datos tienen una distribución normal, en cuanto al correlograma (Figura 4.19-d) da indicios de que los valores grandes se relacionan con los valores pequeños del lag especificado.

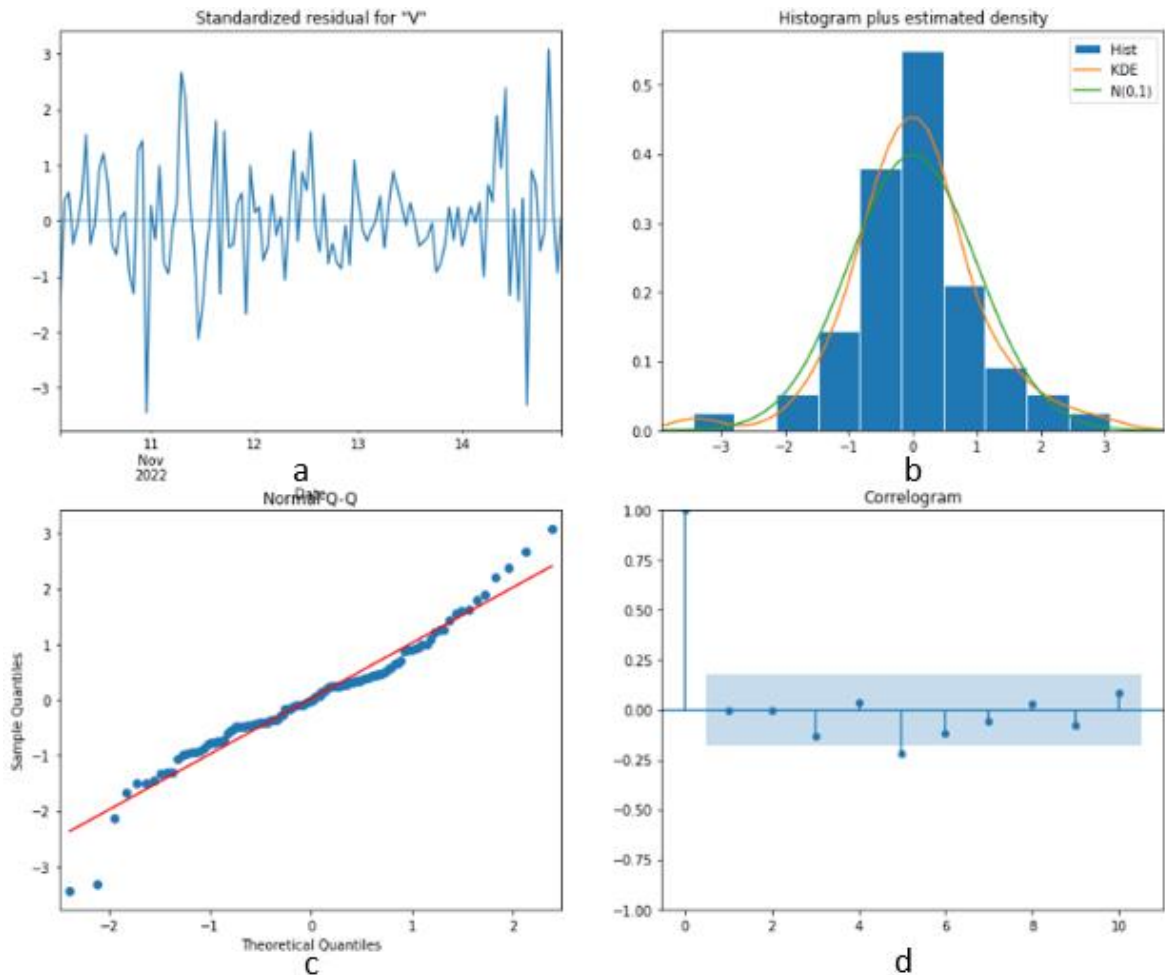


Figura 4.19. Gráfico de Diagnóstico de Modelo SARIMA. **a.** Gráfico de residuos estandarizados; **b.** histograma de densidad estimada; **c.** gráfico normal Q-Q; **d.** gráfico de correlograma

4.3.1.1. Validación del Modelo

Posterior al entrenamiento del modelo de SARIMA se procede a la validación de este. Para el proceso de validación del modelo se considera el cálculo de los índices de rendimiento MSE, MAE y R2 Score para definir la precisión del modelo en función de la comparativa de los resultados de los pronósticos obtenidos con los datos históricos base. En este sentido, la comparativa de los datos históricos con los resultados del modelo de pronóstico se realiza con una ventana de 72 horas (3 días) con una precisión de 30 minutos (dentro del rango propuesto), de tal manera que se obtiene el pronóstico mostrado en la Figura 4.20, en donde se puede observar que el pronóstico de SARIMA tiene una significativa aproximación hacia los datos históricos, lo cual se va a corroborar con el cálculo del índice MAE.

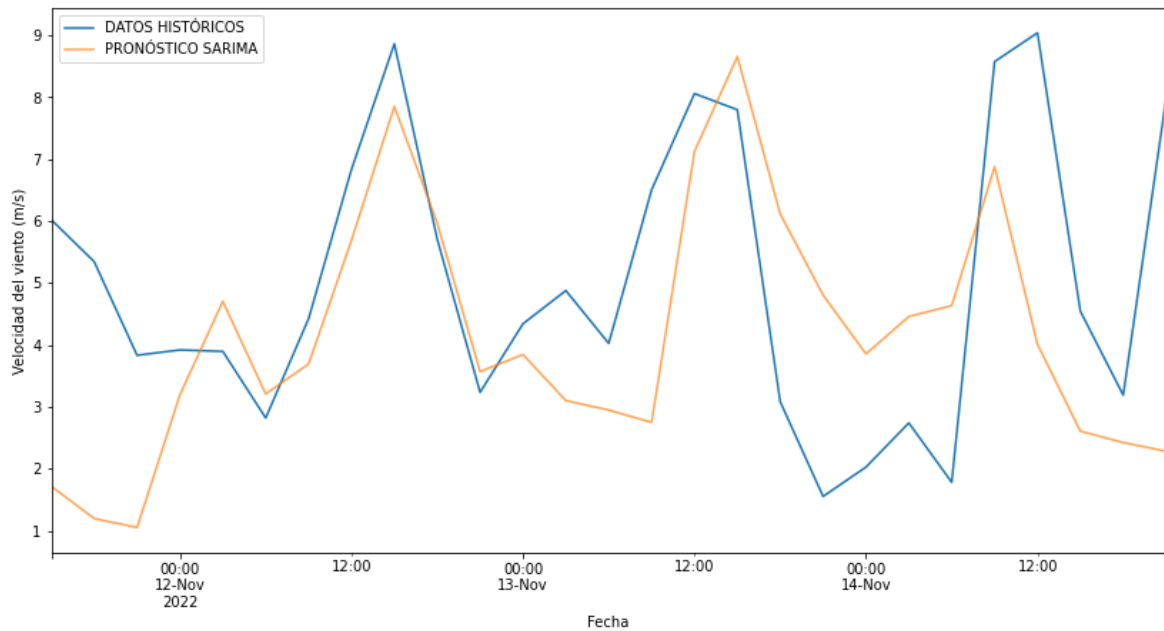


Figura 4.20. Escenario de Validación de Modelo SARIMA

Se procede a realizar un cálculo del error obtenido entre los datos de históricos y del pronóstico obtenido. Como se observa en la Figura 4.21, se tiene un MSE de 5.43, lo cual es un valor no del todo aceptable, aunque será un valor referencial a tener en cuenta, para relacionarlo con el cálculo de los índices de los modelos posteriores. Para el caso del MAE, se tiene un valor de 1.87, el cual indica que se tiene una gran aproximación entre los valores del pronóstico y los valores históricos. Por otra parte, para el caso del índice de r^2 se tiene un valor de -0.07, lo que indica que hay una baja relación de precisión por parte del modelo pronóstico considerando como referencia los valores históricos.

```

Error de test (mse): 5.433522326772365
Error de test (mae): 1.8718189283866298
Error de test (r2_score): -0.072250580049394

```

Figura 4.21. Cálculo de índices de rendimiento. Modelo SARIMA

4.3.1.2. Caso de Estudio

En la Figura 4.21, se observa un escenario de prueba considerando una ventana de tiempo de 24 horas con una precisión de 10 minutos, de forma que se puede verificar el comportamiento de los resultados del pronóstico. Como se puede observar, la serie temporal no sigue una tendencia marcada, es decir, tiene un comportamiento variable, en concordancia con las características variables propias de la velocidad del viento. Hay que considerar que, en promedio, de acuerdo con el índice calculado, se tendrá un MSE de 4.738 aproximadamente. Para el caso del MAE, se tiene un valor de 1.654,

y el valor de R2 de 0.058. Considerando cada uno de estos índices, se observa que el modelo presenta una leve dispersión en cuanto a la comparativa con los datos históricos, teniendo un error promedio absoluto de 1.654 y R2 positivo, aunque muy cercano a cero, que es indicativo de no presentar un modelo muy preciso.

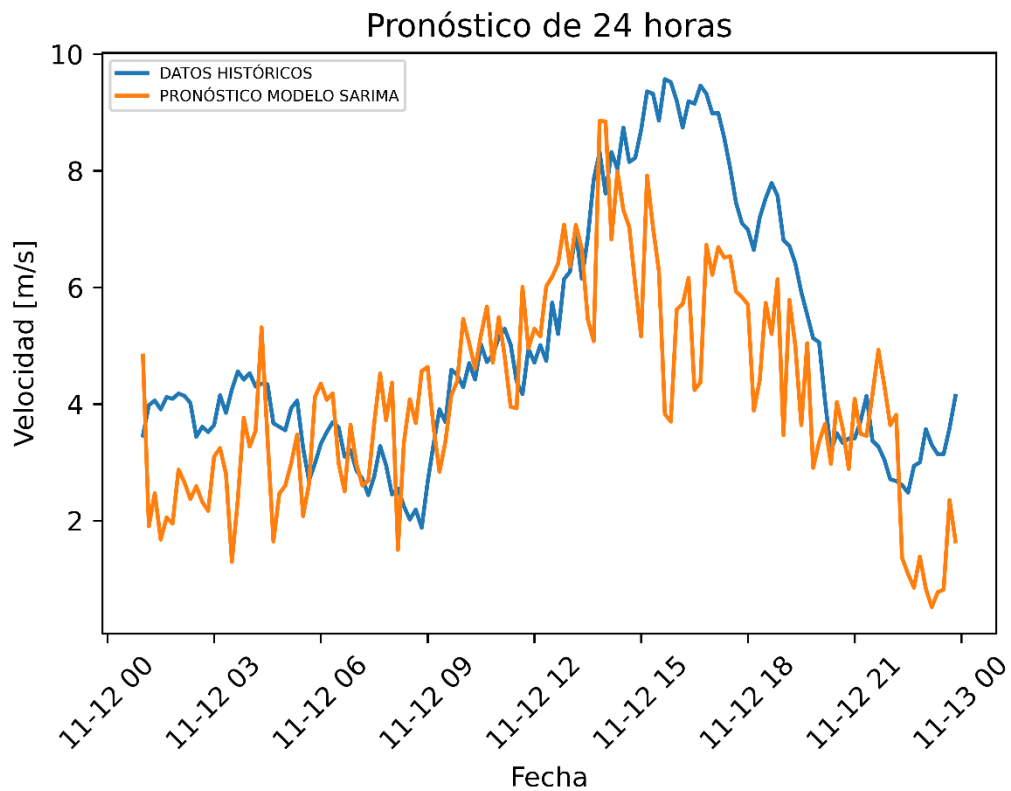


Figura 4.21. Escenario de prueba de pronóstico del modelo de SARIMA

Tabla 4.1. Cálculo de índices de rendimiento. Modelo SARIMA. Caso de Estudio

MODELO	MSE	MAE	R2
SARIMA	4.738	1.654	0.058

4.3.2. Modelo Autoregresivo Recursivo

Inicialmente, se considera la definición de los datos de entrenamiento y de los datos de prueba. En este sentido, se ha considerado los primeros valores -n de la serie como datos de entrenamiento, y los 20 últimos valores (n) como datos de prueba, como se muestra en la Figura 4.22.

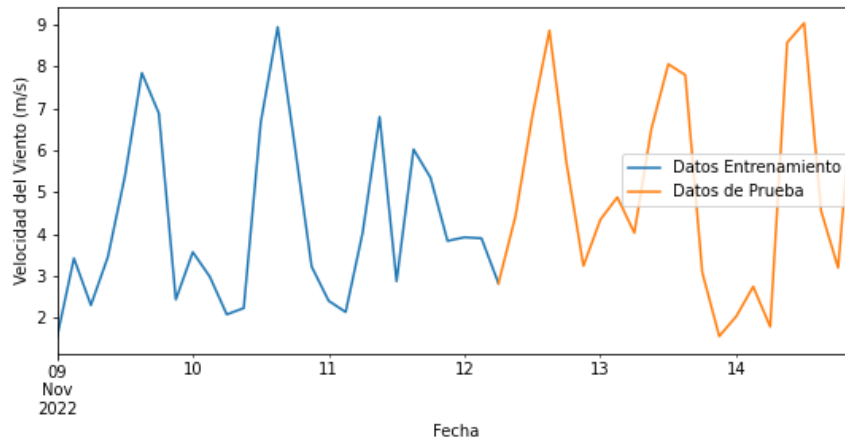


Figura 4.22. Selección de datos de entrenamiento y prueba. Modelo AR

En la Figura 4.23, se despliegan los parámetros del modelo entrenado, en donde se muestra el número de lags considerados (16), el tipo de regresor utilizado, la ventana de tiempo utilizada en el entrenamiento, la frecuencia de datos (por hora), además de la parametrización del regresor. En este sentido, los datos de entrada al modelo de pronóstico serán:

- Tipo de regresor: “*Random Forest*”
- Lags: últimos 16 periodos de la serie temporal original
- Ventana de tiempo de datos de entrenamiento
- Frecuencia de los datos de entrenamiento: 180 minutos

```

=====
ForecasterAutoreg
=====
Regressor: RandomForestRegressor(random_state=123)
Lags: [ 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16]
Transformer for y: None
Transformer for exog: None
Window size: 16
Weight function included: False
Exogenous included: False
Type of exogenous variable: None
Exogenous variables names: None
Training range: [Timestamp('2022-11-09 00:00:00'), Timestamp('2022-11-12 06:00:00')]
Training index type: DatetimeIndex
Training index frequency: 180T
Regressor parameters: {'bootstrap': True, 'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, 'max_samples': None, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'n_estimators': 100, 'n_jobs': None, 'oob_score': False, 'random_state': 123, 'verbose': 0, 'warm_start': False}
Creation date: 2023-06-28 10:28:18
Last fit date: 2023-06-28 10:28:18
Skforecast version: 0.6.0
Python version: 3.8.15

```

Figura 4.23. Parametrización del modelo de pronóstico

4.3.2.1. Validación del Modelo

La validación del modelo se realiza considerando una ventana de tiempo de 72 horas (3 días) con una precisión de 30 minutos (dentro del rango propuesto), relacionando los valores históricos y de los obtenidos en el modelo de pronóstico, cuya comparativa se muestra en la Figura 4.24.

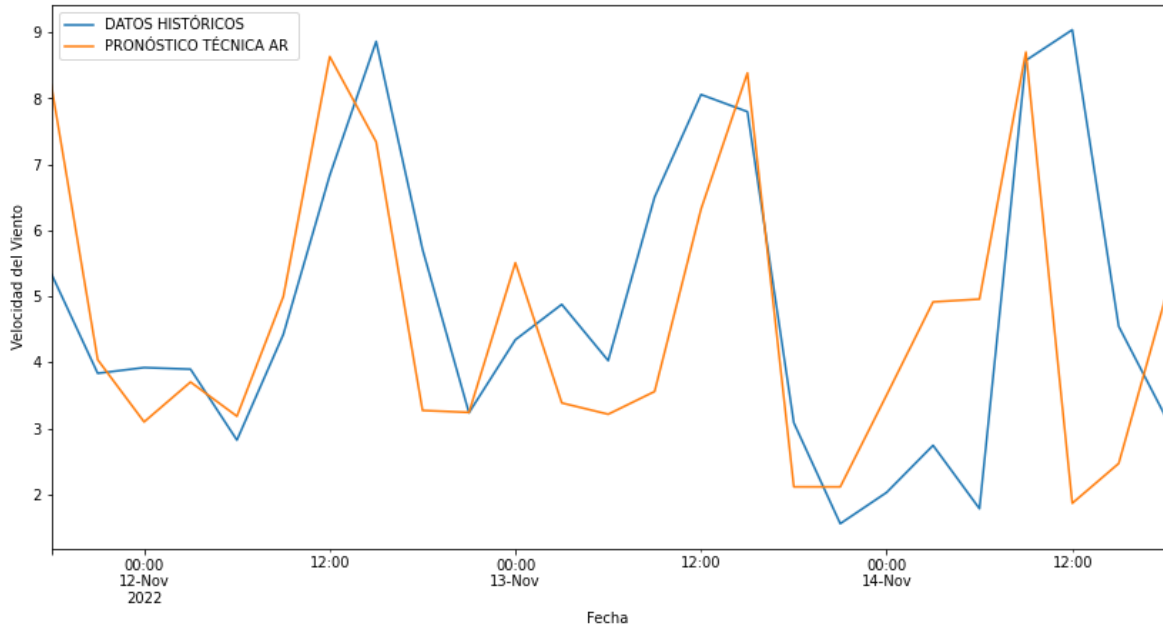


Figura 4.24. Escenario de Validación de Modelo AR

```
Error de test (mse): 3.8560765492802465
Error de test (mae): 1.4218722222222224
Error de test (r2_score): 0.21906673765352525
```

Figura 4.25. Cálculo de índices de rendimiento. Modelo AR

De la Figura 4.24 se puede observar que el modelo de pronóstico tiene una aproximación al comportamiento de los datos históricos, lo cual se podrá cuantificar con los índices MSE, MAE y R2.

En relación con el MSE, se tiene un valor de 3.856, el cual es mucho menor al valor obtenido en el modelo de SARIMA. En cuanto a los índices MAE y R2, se obtiene valores de 3.856 y 1.422, lo cual indica una mejor precisión obtenida frente el modelo de SARIMA, considerando una menor dispersión de los datos en cuanto a la comparativa de los datos históricos y de los obtenidos en el modelo de pronóstico.

4.3.2.2. Caso de Estudio

Posterior a la validación del modelo, se puede realizar una prueba del entrenamiento del modelo de pronóstico, con una ventana de 24 horas con una precisión de 10 minutos, de forma que se observa cómo se comporta el modelo frente a los datos de entrenamiento y de prueba, como se observa en la Figura 4.26, en donde se ha considerado los datos históricos y de los resultados del pronóstico. De tal manera que, como se puede verificar, la serie temporal tiene una aproximación significativa en relación a los datos históricos, lo cual va a ser cuantificado a través del índice MSE, MAE y R2. Además, la serie temporal no presenta una tendencia de forma determinista, lo cual es aceptable en relación al comportamiento de la velocidad del viento.

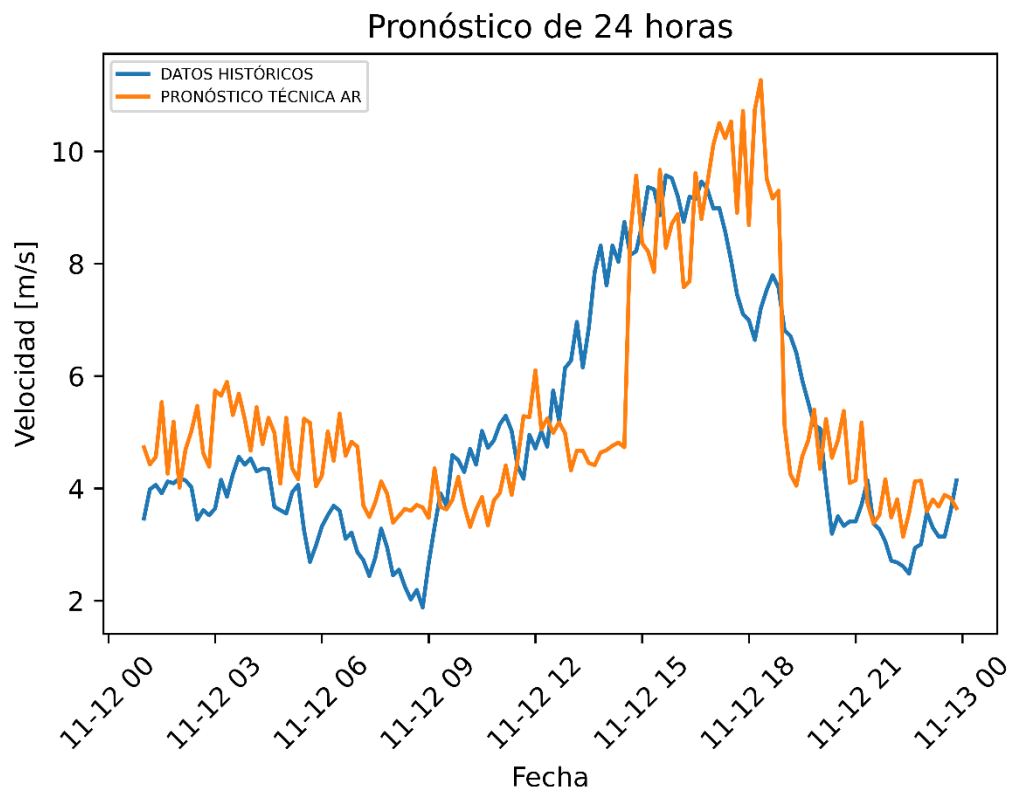


Figura 4.26. Escenario de prueba de modelo de pronóstico

Para cuantificar la precisión del modelo se calcula el MSE, cuyo valor obtenido es de 3.573 aproximadamente, relativamente bajo en relación con el modelo de SARIMA, además de valores de MAE y R2 de 1.358 y 0.276, con lo cual se puede observar una mejor precisión en cuanto a los resultados de los pronósticos obtenidos, minimizando la dispersión de los datos dentro de la obtención de los pronósticos.

Tabla 4.2. Cálculo de índices de rendimiento. Modelo AR. Caso de Estudio

MODELO	MSE	MAE	R2
MODELO AUTOREGRESIVO RECURSIVO	3.573	1.358	0.276

4.3.3. Modelo Autoregresivo Recursivo Con Predictores Modificados

Inicialmente, se realiza la selección de los datos de entrenamiento y de prueba, teniendo en cuenta los datos históricos como referencia y ventana de tiempo seleccionada, tal como se observa en la Figura 4.27. En la Figura 4.28 se dispone del despliegue de los parámetros del modelo de pronóstico. De la misma manera como en el caso anterior, se dispone del tipo de regresor, el número de lags, el intervalo de tiempo en el entrenamiento y la periodicidad de los datos utilizados, además de la parametrización del regresor.

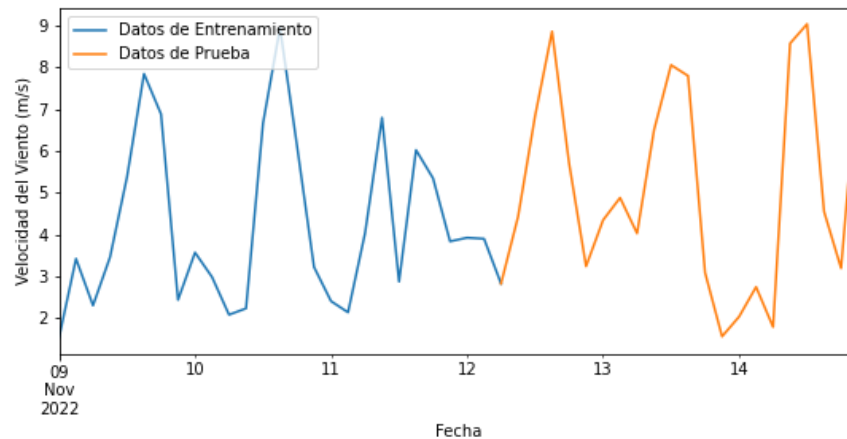


Figura 4.27. Selección de datos de entrenamiento y prueba. Modelo AR con Predictores modificados

```

=====
ForecasterAutoregCustom
=====
Regressor: RandomForestRegressor(random_state=123)
Predictors created with function: create_predictors
Transformer for y: None
Transformer for exog: None
Window size: 10
Weight function included: False
Exogenous included: False
Type of exogenous variable: None
Exogenous variables names: None
Training range: [Timestamp('2022-11-09 00:00:00'), Timestamp('2022-11-11 03:00:00')]
Training index type: DatetimeIndex
Training index frequency: 180T
Regressor parameters: {'bootstrap': True, 'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, 'max_samples': None, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'n_estimators': 100, 'n_jobs': None, 'oob_score': False, 'random_state': 123, 'verbose': 0, 'warm_start': False}
Creation date: 2023-06-28 11:31:11
Last fit date: 2023-06-28 11:31:11
Skforecast version: 0.6.0
Python version: 3.8.15
    
```

Figura 4.28. Despliegue de parámetros del modelo de pronóstico

4.3.3.1. Validación del Modelo

La validación del modelo se realiza comparando los valores históricos con los resultados del modelo de pronóstico para una ventana de 72 horas (3 días) con una precisión de 30 minutos (dentro del rango propuesto), y a través del cálculo de los índices MSE, MAE y R2, de forma que inicialmente se tiene la Figura 4.29.

De la Figura 4.29, se observa como es el comportamiento de los valores pronosticados, considerando la presencia de un margen de error en relación con los valores históricos. El valor obtenido de MSE es de 3.749, el MAE de 1.568 y R2 de 0.241. En este sentido, se verifica una mejor precisión y una baja dispersión de los valores resultantes, en relación con los modelos de SARIMA y AR.

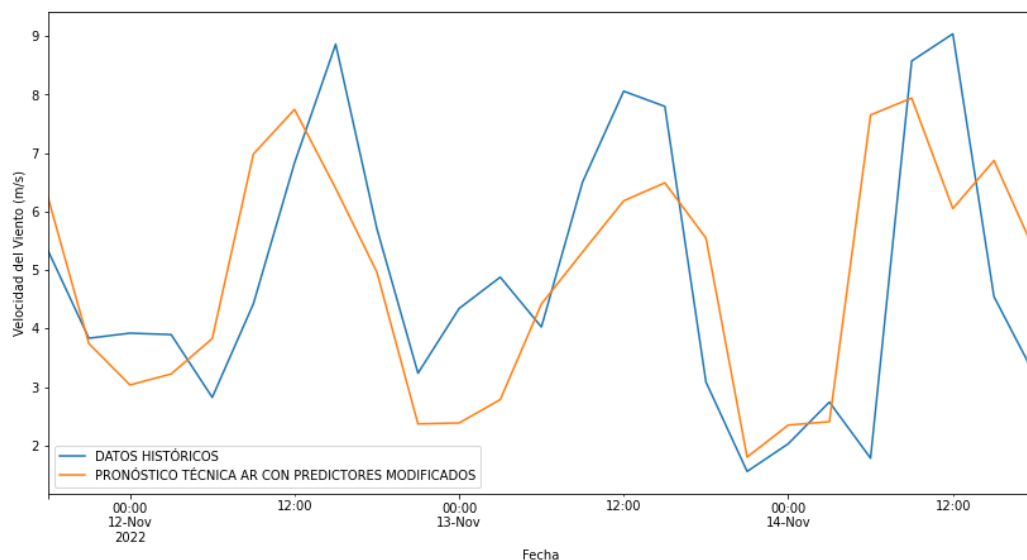


Figura 4.29. Escenario de Validación de Modelo AR con Predictores modificados

```
Error de test (mse): 3.7494680192586483
Error de test (mae): 1.5680311847619048
Error de test (r2_score): 0.24065711483607077
```

Figura 4.30. Cálculo de índices de rendimiento. Modelo AR con Predictores modificados

4.3.3.2. Caso de Estudio

Posterior a la parametrización del modelo, se considera una prueba del pronóstico, con una ventana de tiempo de 24 horas con una precisión de 10 minutos. En la Figura 4.31. Se puede observar que el modelo obtenido tiene una mejor aproximación con relación a los datos históricos lo cual se corroborará con el cálculo de los índices MSE, MAE y R2, además, la serie temporal, al igual que los modelos anteriores, no presenta una tendencia marcada en cuanto a los resultados de los pronósticos, de forma que se tiene un

comportamiento adecuado en relación con la variabilidad de la velocidad del viento. De la Figura indicada, se puede denotar que la desviación que se presenta en mayor parte se da por la variación de los valores pico, que en relación con el MAE nos da un indicativo que no es un valor que supera el valor de 1.378 de la velocidad del viento.

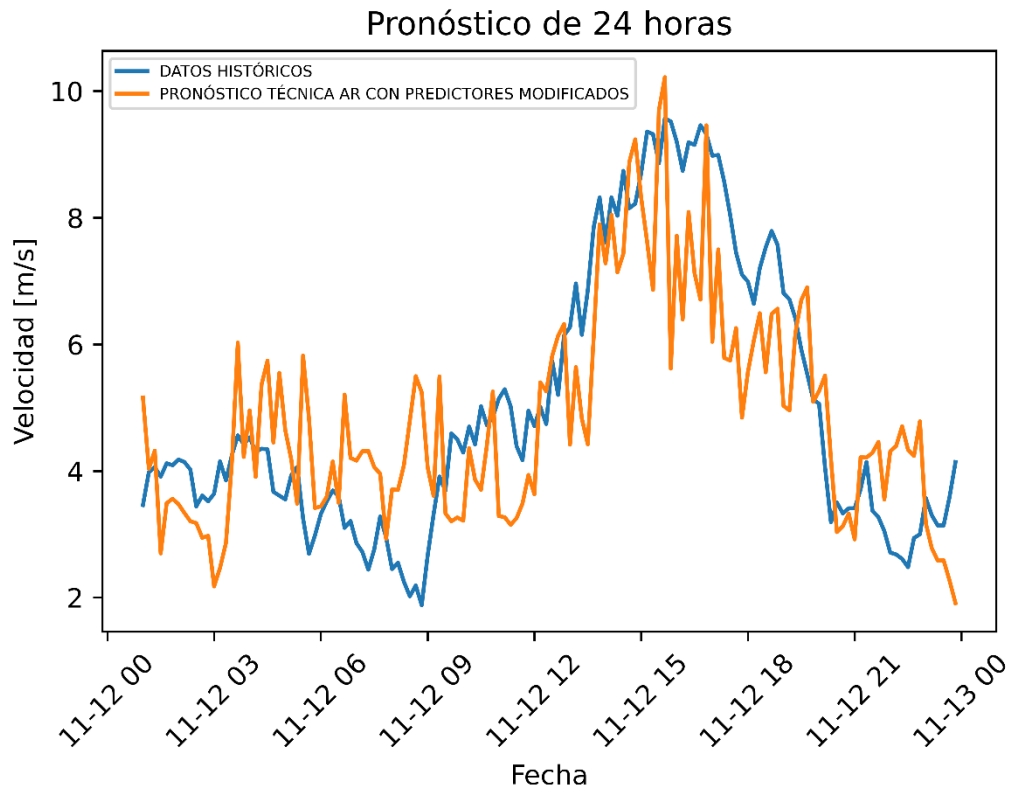


Figura 4.31. Escenario de prueba de modelo de pronóstico

Tabla 4.3. Cálculo del MSE, MAE, R2 del modelo entrenado

MODELO	MSE	MAE	R2
MODELO AUTOREGRESIVO RECURSIVO CON PREDICTORES MODIFICADOS	3.292	1.378	0.294

El valor obtenido de MSE para este modelo es de 3.292, el cual con relación a los modelos anteriores es un indicativo de una mejor aproximación del modelo en consideración con los datos del modelo histórico. En relación con el ajuste del modelo, los valores de MAE y R2, son de 1.378 y 0.294 respectivamente presentando un mejor ajuste de los resultados del modelo de pronóstico.

4.3.4. Comparativa de modelos basados en un pronóstico de 24 horas

Para el presente caso de estudio se ha considerado la elaboración de una prueba de pronóstico con una ventana de 24 horas con precisión de 10 minutos.

Para tener una mejor visión de cómo se aproximan cada uno de los modelos de pronóstico a los datos históricos, se realiza una gráfica comparativa de los mismos con referencia a los datos históricos de velocidad de viento; con esto se puede conocer cómo se comportan cada uno de los modelos frente a la presencia de valores picos, los cuales se vuelven críticos al momento de desarrollar los pronósticos.

En la Figura 4.32 se puede observar el comportamiento de los modelos de pronóstico frente a una ventana de corto plazo. Como se puede visualizar, existe una gran proximidad en cuanto a cada uno de los modelos en relación con los datos históricos de la velocidad del viento. Aunque para el caso de la técnica AR, se tiene un pronunciado sobre pico en uno de los tramos de la ventana de tiempo. Y, por otra parte, a breve rasgos, se puede visualizar como los pronósticos del modelo SARIMA no alcanza a tener un buen rendimiento en el tramo de las velocidades altas.

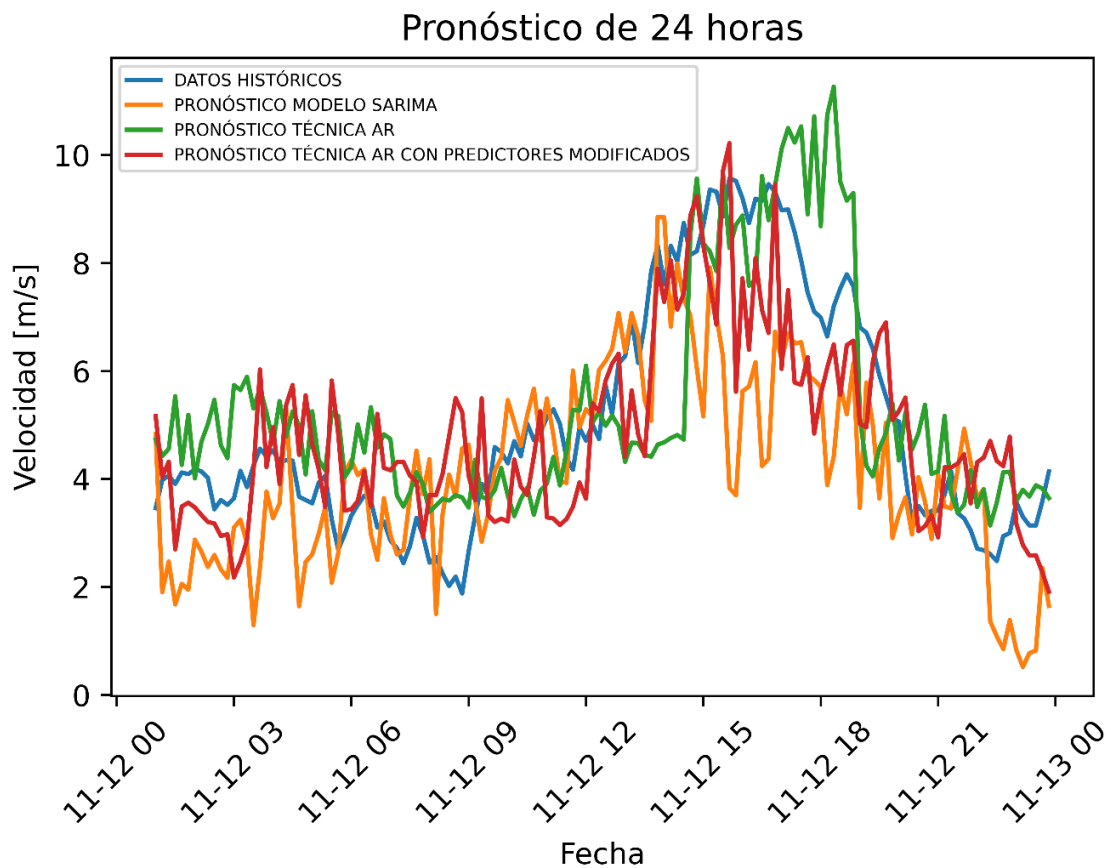


Figura 4.32. Pronóstico de 24 horas de los modelos entrenados

Para corroborar las hipótesis asumidas en cuanto al rendimiento de los modelos de pronóstico, se tiene el cálculo de los índices de rendimiento de la Tabla 4.4; Se observa que el modelo SARIMA presenta un alto MSE, y el modelo que presenta un mejor MSE es el AR recursivo modificado con un valor de 3.29, lo cual sucede de igual forma para el caso del índice MAE.

El modelo que tiene un mejor rendimiento en función del error, es el modelo AR con predictores modificados. Para el caso del índice R2, de igual forma, el modelo AR con predictores modificados tiene una mejor correspondencia entre los valores pronosticados y de los datos históricos.

Tabla 4.4. Tabla de comparación de índices de rendimiento calculados

MODELO	MSE	MAE	R2
SARIMA	4.738	1.654	0.058
MODELO AUTOREGRESIVO RECURSIVO	3.573	1.358	0.276
MODELO AUTOREGRESIVO RECURSIVO CON PREDICTORES MODIFICADOS	3.292	1.378	0.294

Para el caso de los modelos AR, se tiene un bajo MAE, en consecuencia, aunque se observa que estos modelos tienen la ventaja de tener un buen ajuste y poder acercarse de forma muy aproximada en los valores transitorios, que como se puede observar tienen como consecuencia la presencia de un alto índice de error, lo cual ocurre en el caso del modelo de SARIMA.

Cada uno de los modelos propuestos presentan características que se adaptan a la información disponible para el entrenamiento, de manera que, para el objetivo del manejo de pronósticos a corto plazo, el modelo que mejor se adapta a los objetivos operativos de la central eólica es el Modelo Autoregresivo Recursivo con predictores modificados.

4.3.5. Cálculo de índices de rendimiento de modelos basados en un pronóstico de 72 horas

En la Tabla 4.5 se muestra el detalle comparativo de los índices de rendimiento de cada uno de los modelos considerados, de forma que se pueda definir ciertos criterios.

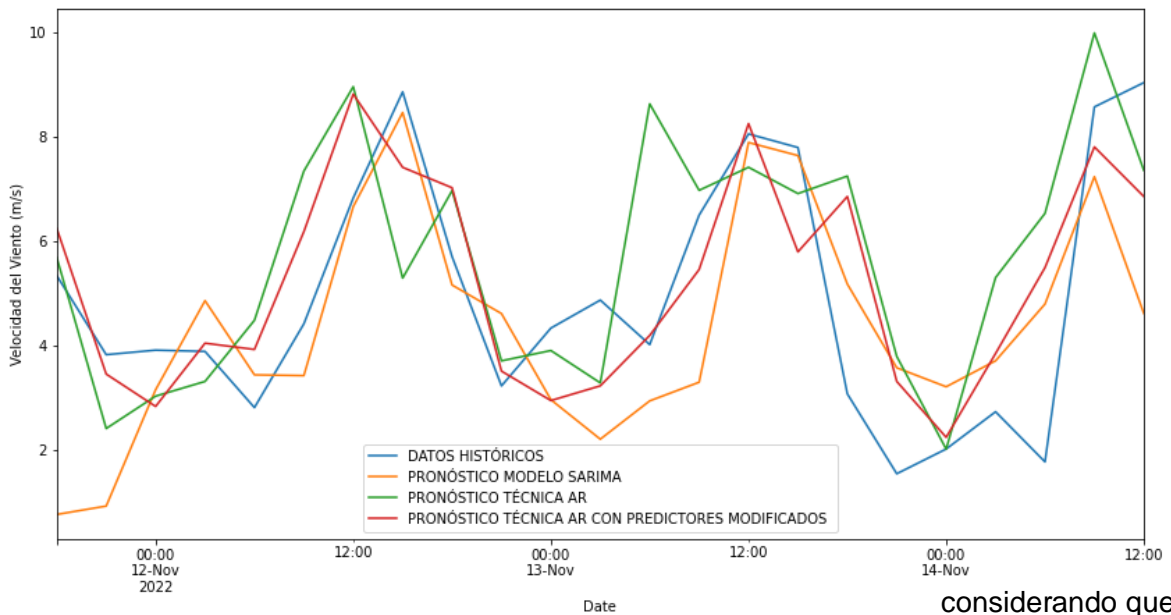
Tabla 4.5. Tabla de comparación de los índices de rendimiento calculados

MODELO	MSE	MAE	R2
SARIMA	5,43	1.87	-0.07
MODELO AUTOREGRESIVO RECURSIVO	3.86	1.42	0.21
MODELO AUTOREGRESIVO RECURSIVO CON PREDICTORES MODIFICADOS	3.74	1.56	0.24

Se observa que el modelo SARIMA con un valor de MSE de 5.43 es el que relativamente tiene un menor ajuste frente al entrenamiento de pronósticos a corto plazo, mientras que el Modelo Autoregresivo Recursivo con predictores modificados tiene un mayor ajuste frente al pronóstico a corto plazo. Para verificar el comportamiento de las series temporales, se ha considerado la gráfica de los tres modelos, teniendo en cuenta una ventana a de aproximadamente 72 horas (3 días), de forma que se pueda verificar de forma más clara el comportamiento en cuanto a los valores transitorios, es decir en los valores pico y valores mínimos, tal como se observa en la Figura 4.33.

Figura 4.33. Análisis comparativo de los modelos implementados

Cada uno de los modelos como se puede observar tienen sus distintos desempeños,



considerando que El comportamiento de la velocidad del viento es muy variante en el tiempo, de manera que la presencia de los errores propios de los modelos de pronóstico, se presentan por dicho comportamiento.

Para efectos de considerar un análisis de sensibilidad con respecto a la precisión y de las mínimas ventanas de tiempo que se pueden manejar con los modelos entrenados, se tiene el despliegue de resultados del ANEXO B.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- En el presente trabajo de titulación se ha desarrollado un modelo de pronóstico de viento para centrales de generación eólica en el corto plazo considerando técnicas basadas en series temporales, comprobando que el modelo con mejor adaptación a las condiciones y características de la Central Eólica Villonaco y con un bajo MSE de 3.292 es el Modelo Autoregresivo Recursivo con predictores modificados.
- Como punto de partida en el desarrollo del trabajo de titulación, se ha realizado un estudio del estado del arte de los diferentes modelos de pronóstico de viento de corto y largo plazo en centrales eólicas, como base conceptual y teórica para la definición de la metodología propuesta. De la revisión bibliográfica realizada se tiene que los modelos más utilizados y con mejor practicidad frente a la aplicación para la elaboración de pronósticos de viento son el modelo SARIMA y los modelos Autorregresivos, teniendo en cuenta la principal característica de aplicación a modelos de pronóstico de viento a corto plazo. La modelación de series temporales permite definir un modelo en donde se pueda determinar el comportamiento de un determinado fenómeno y su variable asociada en el tiempo, de manera que se pueda analizar y proponer diferentes medidas frente a su tratamiento de datos.
- Se realizó un análisis de la información y de los principales parámetros, variables y aspectos relacionados con la velocidad del viento, los cuales influyen en el desarrollo de los modelos de pronóstico de a corto plazo, lo cual sirve como información base para el desarrollo de la aplicación metodológica.
- El modelo de pronóstico de viento propuesto en este trabajo de titulación se realizó como métodos de series temporales empleando técnicas estadísticas (SARIMA), Autoregresivas y Autoregresivas con predictores modificados, considerando los datos de velocidad de viento de la Central Eólica Villonaco. La finalidad del desarrollo de estos tres modelos es de determinar el mejor rendimiento frente al conjunto de datos y comportamiento de viento de la central Villonaco.
- De acuerdo con el análisis de correlaciones realizado en este trabajo se pudo determinar que no existe un parámetro que afecte directamente al comportamiento del viento en el sitio de implantación de la central. Que en promedio los vientos son

de 10.93m/s, máximo y mínimo de 0,13 m/s, y que el modelo de pronóstico tiene resultados

- En conclusión, al analizar los resultados del modelo SARIMA para un pronóstico a corto plazo de 24 horas presenta un MSE de 4.738, mientras que el modelo Autoregresivo y Autoregresivo con predictores modificados presentan un MSE de 3,573 y 3,292 respectivamente, evidenciando que el modelo con mejor adaptabilidad a los objetivos operativos de la central eólica es el Modelo Autoregresivo Recursivo con predictores modificados.

5.2. RECOMENDACIONES

- Frente a las limitaciones de procesamiento de datos de los computadores, y considerando la gran cantidad de información que puede disponer un determinado conjunto de datos, es importante la opción de manejar un muestreo de la información, de forma que se pueda minimizar el impacto computacional para el procesamiento y el cálculo de los modelos sin la posibilidad de pérdida de información.
- En el procesamiento de la información, es importante que se realice un análisis de la disposición de los datos, teniendo en cuenta características tales como presencia de valores anómalos o nulo, además de la identificación de parámetros que den una perspectiva del tipo de datos que se maneja en los conjuntos de datos. La realización de un análisis previo tal como de un análisis estadístico, permite tener una mejor visión de forma amplia de cómo es el comportamiento de los datos de acuerdo con cada una de las variables, de forma que se puede definir aspectos tales como de los límites en los cuales opera o se comporta determinada variable, además de la distribución de los datos. Otro aspecto importante que se considera es el de uso del cálculo de los índices de correlación, que de cierta manera da un indicativo de la independencia de los datos frente al desarrollo de los modelos o análisis a desarrollare.
- Hay que considerar y tener definido los conjuntos de datos, teniendo en cuenta de que tengan la posibilidad de tener un marco de temporalidad necesarios para la creación de las series temporales.

REFERENCIAS BIBLIOGRÁFICAS

- [1] D. R. Vallejo Ortega, «Estimación del Recurso Eólico dentro del Área de Concesión de la Empresa Eléctrica Quito para una posible Instalación de una Central de Generación Eléctrica-Análisis de Alternativas de Implementación y Estimación del Precio de la Energía Producida,» Quito, 2020.
- [2] G. Galán Casillas, «Desarrollo de modelos de predicción de la generación de energía eólica y solar para la optimización de Centros de Datos en un entorno de Smart Grid,» Madrid, 2017.
- [3] F. Eraso Checa, E. Escobar Rosero, D. F. Paz y C. Morales, «Metodología para la determinación de características del viento y evaluación del potencial de energía eólica en Túquerres - Nariño,» *Revista Científica*, vol. 31, nº 1, pp. 19-31, 2017.
- [4] I. Erbetta Mattig, «Predicción de la Potencia para la Operación de Parques Eólicos,» Santiago de Chile, 2010.
- [5] A. D. Sahin, «Progress and recent trends in wind energy,» *Progress in Energy Combustion Sci.*, vol. 30, nº 5, pp. 501-543, 2004.
- [6] M. Stiebler, «Wind Energy Systems for Electric Power Generation,» Springer-Verlag Berlin Heidelberg, Berlin, 2008.
- [7] K. Dai, A. Bergot, C. Liang, W.-N. Xiang y Z. Huang, «Environmental issues associated with wind energy –A review,» *Renewable Energy*, vol. 75, pp. 911-921, 2015.
- [8] R. Saidur, M. R. Islam, N. A. Rahim y K. H. Solangi, «A review on global wind energy policy,» *Renewable Sustainable Energy Rev.*, vol. 14, nº 7, pp. 1744-1762, 2010.
- [9] REPSOL, «¿Qué es la energía eólica?,» REPSOL, 2022. [En línea]. Available: <https://www.repsol.com/es/conocenos/que-hacemos/desarrollo-energias-renovables/energia-eolica/index.cshtml>.
- [10] J. Cifuentes, «Energía eólica residencial como uso eficiente de la energía renovable y adaptación al cambio climático,» 2016.
- [11] V. Díaz Lagares, «LOS RETOS DE LA ENERGÍA EÓLICA MARINA EN ESPAÑA: EL PAPEL DE LAS C.C.A.A. Y LA ORDENACIÓN DE LOS ESPACIOS MARINOS ANTE LA DIRECTIVA 2014/89/UE,» *Actualidad Jurídica Ambiental*, nº 56, 2016.
- [12] A. Kumar, M. Z. Ullah Khan y B. Pandey, «Wind Energy: A Review Paper,» *Gyancity Journal of Engineering and Technology*, vol. 4, nº 2, pp. 29-37, 2018.

- [13] A. Kalmikov y K. Dykes, «Wind Power Fundamentals,» MIT Mechanical Engineering.
- [14] L. Betancourt, «Aspectos Básicos para la Operación, Control y Mantenimiento de un Parque Eólico,» Bogotá, 2004.
- [15] A. Valenzuela Holguin, «Modelo de un Sistema de Generación Eólica basada en PMSG para el simulador de estabilidad ESTABI,» Torreón, 2018.
- [16] Devashish y A. Thakur, «A Comprehensive Review on Wind Energy System for Electric Power Generation: Current Situation and Improved Technologies to Realize Future Development,» *INTERNATIONAL JOURNAL of RENEWABLE ENERGY RESEARCH*, vol. 7, nº 4, pp. 1786-1805, 2017.
- [17] A. Falces de Andrés, «Planificación de Parques Eólicos mediante sistemas de información geográfica y Algoritmos genéticos,» Logroño, 2015.
- [18] «Curva de Potencia de un Aerogenerador,» Danish Wind Industry Association, 07 mayo 2003. [En línea]. Available: <http://xn--drmstrre-64ad.dk/wp-content/wind/miller/windpower%20web/es/tour/wres/pwr.htm>.
- [19] «Costes de operación y de mantenimiento en aerogeneradores,» Danish Wind Industry Association, 12 mayo 2003. [En línea]. Available: <http://xn--drmstrre-64ad.dk/wp-content/wind/miller/windpower%20web/es/tour/econ/oandm.htm>.
- [20] I. Núñez, «Generación eólica y solar fotovoltaica: ¿Qué tan variables son?,» *Breves de Energía*, 2015.
- [21] L. A. Enríquez García y H. O. Moreno Avilés, «Características del viento y potencia eólica disponible en la región andina de la provincia de Chimborazo,» *Revista Científica Dominio de las Ciencias*, vol. 6, nº 3, pp. 1077-1093, 2020.
- [22] A. Hernández y O. Ortega, «Análisis estadístico del viento como recurso energético,» México, 2014.
- [23] A. González González y E. Zulueta, «Estrategias de Control de Aerogeneradores,» *XVI Congreso de Confiabilidad*, 2014.
- [24] X. Wang, P. Guo y X. Huang, «A Review of Wind Power Forecasting Models,» *Energy Procedia*, vol. 12, pp. 770-778, 2011.
- [25] N. Chen, Z. Qian, T. Nabney y X. Meng, «Wind power forecasts Using Gaussian Processes and Numerical Weather prediction,» *IEEE Transaction on Power Systems*, vol. 29, nº 2, pp. 656-665, 2014.
- [26] R. G. Kavasseri y K. Seetharaman, «Day-ahead wind speed forecasting using f-ARIMA models,» *ELSEVIER Renewable Energy* 34, pp. 1388-1393, 2009.

- [27] M. Bilgili, B. Sahin y A. Yasar, «Application of artificial neural networks for the wind speed prediction of target station using reference stations data,» *ELSEVIER Renewable Energy* 32, pp. 2350-2360, 2007.
- [28] Kusiak y Z. Zhang, «Short- Horizon Prediction of Wind Power: A Data driven approach,» *IEEE Transactions on Energy Conversion*, vol. 25, nº 4, pp. 1112-1122, 2010.
- [29] P. H. Ibarguengoytia-González, A. Reyes-Ballesteros, M. Borunda-Pacheco y U. A. García-López, «Predicción de potencia eólica utilizando técnicas modernas de Inteligencia Artificial,» *Ingeniería, investigación y tecnología*, vol. 19, nº 4, 2018.
- [30] Z. Janjic, T. Black, M. Pyle, B. Ferrier y H. Chuang, «NMM version 3 modelling system user's guide Developmental Testbed Center,» National Centers for Environmental Prediction, 2014.
- [31] E. de Almeida, A. Gutiérrez y S. Camargo, «Pronóstico de Energía Eólica en Uruguay para horizontes Temporales de Corto Plazo en base a modelo Numérico de Mesoescala y Redes Neuronales Artificiales,» *enerLAC*, vol. IV, nº 1, pp. 32-43, 2020.
- [32] C. Waimann, «Desarrollo de un sistema de pronóstico estocástico-dinámico de producción de energía eólica basado en el modelo WRF/CIMA,» Buenos Aires, 2016.
- [33] G. Giebel, «The State-Of-The-Art in Short-Term Prediction of Wind Power. A Literature Overview,» 2003.
- [34] M. Arquimedes Ferrera, «Introducción a los Modelos Probabilísticos de la Función de Confiabilidad (R(t)),» Predictiva21, 2021. [En línea]. Available: <https://predictiva21.com/modelos-probabilisticos-funcion-confiabilidad/>.
- [35] «Very short-term probabilistic forecasting of wind power based on dual-Doppler radar measurements in the North Sea,» *Journal of Physics: Conference Series*, pp. 1-10, 2018.
- [36] U. Germann y I. Zawadzki, «*Journal of Applied Meteorology*,» vol. 43, pp. 74-89, 2004.
- [37] X. Zhu y M. G. Genton, «Short-Term Wind Speed Forecasting for Power System Operations,» *International Statistical Review*, vol. 80, nº 1, pp. 2-23, 2012.
- [38] M. G. Genton y A. S. Hering, «Blowing in the Wind,» *Significance*, vol. 4, pp. 11-14, 2007.
- [39] J. W. Gouri Ren, «Analysis of wind power intermittency based on historical wind power data,» *Energy*, vol. 150, pp. 482-492, 2018.

- [40] «Short-term wind forecasting using statistical models with a fully observable wind flow,» *Journal of Physics: Conference Series*, vol. 1452, pp. 1-12, 2019.
- [41] Q. Chen y K. A. Folly, «Short-Term Wind Power Forecasting Using Mixed Input Feature-Based Cascade-connected Artificial Neural Networks,» *Front. Energy Res.*, 2021.
- [42] H. Rodríguez, N. García, J. Flores, L. Morales y G. Manjarrez, «Pronóstico a corto plazo de velocidad del viento a partir de datos incompletos,» *Research in Computing Science*, pp. 81-95, 2017.
- [43] J. A. Pérez Arenas, «Pronóstico de Demanda de Energía Eléctrica de Colombia Utilizando un Modelo Estadístico a partir de la Metodología de Box-Jenkins,» Bogotá, 2020.
- [44] G. Ríos y C. Hurtador, «Series de Tiempo,» 2008.
- [45] S. Makridakis y C. Wheelwright, «Metode Mathematical Programming, and System Dynamics Methods,» *International Journal of Forecasting*, vol. 12, pp. 329-343, 1999.
- [46] G. Box y G. Jenkins, «Time Series Analysis: Forecasting and Control,» Revised ed. Holden-Day, 1976.
- [47] B. L. Bowerman y R. T. O'Connell, «Forecasting and Time Series An Applied Approach,» *Duxbury Thomson Learning*, nº 3, 1993.
- [48] A. E. Permanasari, I. Hidayah y I. A. Bustoni, «SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast The Number of Malaria Incidence,» *ICITEE 2013*, pp. 203-207, 2013.
- [49] H. Larochelle y I. Murray, «The neural autoregressive distribution estimator,» *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pp. 29-37, 2011.
- [50] S. E. Reed, A. van den Oord, N. Kalchbrenner, S. G. Colmenarejo, Z. Wang, Y. Chen, D. Belov y N. de Freitas, « Parallel multiscale autoregressive density estimation,» *Proceedings of the 34th International Conference on Machine Learning, ICML*, pp. 2912-2921, 2017.
- [51] A. Wiggers y E. Hoogeboom, «Predictive Sampling with Forecasting Autoregressive Models,» *Proceedings of the 37 th International Conference on Machine Learning*, 2020.
- [52] IBM Corporation, «Error cuadrático promedio,» IBM, 14 abril 2021. [En línea]. Available: <https://www.ibm.com/docs/es/cloud-paks/cp-data/3.5.0?topic=overview-mean-squared-error>.

- [53] IBM Corporation, «Error absoluto promedio,» IBM, 14 abril 2021. [En línea]. Available: <https://www.ibm.com/docs/es/cloud-paks/cp-data/3.5.0?topic=overview-mean-absolute-error>.
- [54] «Sales forecasting newspaper with ARIMA: A case study,» *The 3rd International Conference on Industrial, Mechanical, Electrical, and Chemical Engineering*, pp. 1-10, 2018.
- [55] Ministerio de Energía y Minas, «CENTRAL EÓLICA "VILLONACO",» Ministerio de Energía y Minas, 2023. [En línea]. Available: <https://www.rekursyenergia.gob.ec/central-eolica-villonaco/>.
- [56] pandas, «pandas.DataFrame.describe,» pandas, 2023. [En línea]. Available: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>.
- [57] pandas, «pandas.to_datetime,» pandas, 2023. [En línea]. Available: https://pandas.pydata.org/docs/reference/api/pandas.to_datetime.html.
- [58] J. Perktold, S. Seabold y J. Taylor, «modelos estadísticos 0.14.0,» statsmodels-developers, 2019. [En línea]. Available: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html>.
- [59] J. Perktold, S. Seabold y J. Taylor, «statsmodels 0.14.0,» statsmodels-developers, 2019. [En línea]. Available: https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima.model.ARIMAResults.plot_diagnostics.html.
- [60] scikit-learn developers, «Regresor Random Forest,» scikit-learn, 2023. [En línea]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>.
- [61] L. Carvajal, Metodología de la Investigación Científica. Curso general y aplicado, 28 ed., Santiago de Cali: U.S.C., 2006, p. 139.
- [62] OLADE, «MEER presentó el primer Atlas Eólico del Ecuador,» Energías Renovables , 2015. [En línea]. Available: <http://expertosenred.olade.org/energiasrenovables/meer-presento-el-primer-atlas-eolico-del-ecuador/>.
- [63] CELEC, «Identificados sitios con potencial eólico,» 2015. [En línea]. Available: <https://www.celec.gob.ec/termopichincha/index.php/noticias/274-identificados-sitios-con-potencial-eolico>.
- [64] R. Stull, *Meteorology for Scientist and Enginners*, Segunda ed., pp. 181-186.
- [65] I. Rodríguez López, C. Estrada González y I. Hernández Arriaga, «Análisis comparativo entre el modelo dinámico y el modelo de la IEC para obtener

la curva de potencia de turbinas eólicas,» *Revista de Innovación Sistemática*, vol. 1, nº 2, pp. 31-38, 2017.

- [66] IEC, *IEC 61400-12-1. Section: Wind turbines-Part 12-1: Power performance measurements of electricity producing wind turbines*, IEC, 2005.
- [67] L. LI li, H. Yang y Y.-M. Wang, «A physical approach of the short-term wind power prediction based on CFD pre-calculated flow fields,» *Journal of Hydrodynamics*, vol. 25, nº 1, pp. 56-61, 2013.
- [68] G. Teneler, «Wind flow analysis on a complex terrain,» Visby-Sweden, 2011.

ANEXOS

ANEXO A

Tabla A.1. Conjunto de datos de la Unidad U1

S01 10-minute Data Query

No.	Time	Avg Wind Speed(m/s)	Avg Theory active power(kW)	Avg Active Power (kW)	Wind Turbine Energy yield(kWh)	Avg Ambient Temp(°C)	Avg ABS Wind Direction(deg)
1	2022-11-09 00:00	1.33	4.03	-0.81	0.00	13.23	193.39
2	2022-11-09 00:10	1.12	3.13	-1.51	0.00	13.62	193.21
3	2022-11-09 00:20	1.94	6.99	-0.52	0.00	13.69	196.48
4	2022-11-09 00:30	1.32	4.41	-0.96	0.00	13.73	204.57
5	2022-11-09 00:40	1.31	6.59	-0.97	0.00	13.54	217.48
6	2022-11-09 00:50	0.92	2.39	-0.31	0.00	13.65	201.23
7	2022-11-09 01:00	2.03	5.19	-2.03	0.00	13.68	201.02
8	2022-11-09 01:10	2.70	9.38	-0.64	0.00	13.50	189.72
9	2022-11-09 01:20	3.30	27.78	-1.18	0.00	13.52	195.49
10	2022-11-09 01:30	2.60	9.86	-2.01	0.00	13.57	200.68
11	2022-11-09 01:40	1.02	3.53	-1.09	0.00	13.55	210.91
12	2022-11-09 01:50	1.31	7.52	-0.31	0.00	13.26	210.87
13	2022-11-09 02:00	1.39	3.44	-0.57	0.00	13.86	216.32
14	2022-11-09 02:10	1.13	3.71	-1.05	0.00	13.91	216.68
15	2022-11-09 02:20	1.09	4.40	-1.40	0.00	13.47	216.88
16	2022-11-09 02:30	1.25	5.40	-0.30	0.00	13.24	217.04
17	2022-11-09 02:40	0.72	1.59	-0.37	0.00	13.25	217.10
18	2022-11-09 02:50	0.93	3.22	-2.33	0.00	13.21	239.94
19	2022-11-09 03:00	1.37	5.86	-0.30	0.00	13.03	245.48
20	2022-11-09 03:10	1.76	5.73	-1.60	0.00	12.99	252.91
21	2022-11-09 03:20	2.80	12.54	-2.34	0.00	13.13	98.87
22	2022-11-09 03:30	2.97	15.60	-1.90	0.00	13.04	65.53
23	2022-11-09 03:40	2.35	4.16	-0.88	0.00	13.10	62.22
24	2022-11-09 03:50	2.67	9.09	-0.95	0.00	12.83	65.79
25	2022-11-09 04:00	3.38	28.09	-1.96	0.00	12.56	43.58
26	2022-11-09 04:10	3.86	48.12	-1.19	0.00	12.50	52.81
27	2022-11-09 04:20	4.48	77.11	-0.61	0.00	12.57	54.18
28	2022-11-09 04:30	3.96	53.80	-1.12	0.00	12.58	59.21
29	2022-11-09 04:40	3.65	41.47	-1.64	0.00	12.54	71.61
30	2022-11-09 04:50	3.59	37.16	-0.53	0.00	12.54	63.75
31	2022-11-09 05:00	3.28	25.12	-0.97	0.00	12.49	59.93
32	2022-11-09 05:10	4.10	59.49	-2.37	0.00	12.42	55.54
33	2022-11-09 05:20	3.78	45.04	-0.64	0.00	12.32	50.19

34	2022-11-09 05:30	4.52	78.97	-0.78	0.00	12.21	65.44
35	2022-11-09 05:40	4.73	90.30	-2.00	0.00	12.26	55.14
36	2022-11-09 05:50	4.28	66.68	-0.36	0.00	12.28	55.18
37	2022-11-09 06:00	3.74	44.66	-0.65	0.00	12.29	55.35
38	2022-11-09 06:10	3.35	27.70	-1.99	0.00	12.36	59.91
39	2022-11-09 06:20	2.48	7.20	-1.43	0.00	12.46	55.63
40	2022-11-09 06:30	2.96	15.26	-5.28	0.00	12.42	48.96
41	2022-11-09 06:40	2.57	7.12	-2.61	0.00	12.45	53.24
42	2022-11-09 06:50	2.25	4.87	-0.86	0.00	12.79	75.41
43	2022-11-09 07:00	2.67	8.74	-0.99	0.00	12.65	63.10
44	2022-11-09 07:10	3.40	28.10	-2.09	0.00	12.68	48.47
45	2022-11-09 07:20	3.34	25.61	-0.27	0.00	12.79	57.56
46	2022-11-09 07:30	3.48	32.20	-0.21	0.00	12.75	57.73
47	2022-11-09 07:40	2.83	11.62	-1.46	0.00	12.96	57.56
48	2022-11-09 07:50	1.98	6.55	-0.65	0.00	13.18	57.26
49	2022-11-09 08:00	1.94	5.80	-0.24	0.00	13.50	57.42
50	2022-11-09 08:10	1.51	6.84	-0.38	0.00	13.63	57.24
51	2022-11-09 08:20	0.47	0.00	-0.27	0.00	14.16	55.97
52	2022-11-09 08:30	0.36	0.00	-0.09	0.00	14.05	55.23
53	2022-11-09 08:40	1.30	3.83	-0.02	0.00	14.30	55.09
54	2022-11-09 08:50	0.74	0.59	-0.24	0.00	15.34	55.16
55	2022-11-09 09:00	0.65	0.81	-0.18	0.00	15.41	54.90
56	2022-11-09 09:10	1.24	5.86	-0.33	0.00	15.23	54.88
57	2022-11-09 09:20	0.67	0.47	-0.42	0.00	16.04	70.80
58	2022-11-09 09:30	0.80	2.62	-0.21	0.00	15.98	72.11
59	2022-11-09 09:40	0.53	0.51	-0.22	0.00	16.06	71.64
60	2022-11-09 09:50	0.55	0.31	-0.14	0.00	16.26	71.42
61	2022-11-09 10:00	2.18	12.55	-0.82	0.00	15.53	57.25
62	2022-11-09 10:10	3.94	52.55	-0.95	0.00	15.74	45.34
63	2022-11-09 10:20	4.14	62.99	-0.59	0.00	15.66	43.24
64	2022-11-09 10:30	4.27	65.97	-1.37	0.00	15.48	37.15
65	2022-11-09 10:40	4.45	77.96	-5.27	0.00	15.70	47.85
66	2022-11-09 10:50	5.20	131.64	-1.10	0.00	15.87	41.23
67	2022-11-09 11:00	5.80	181.32	-0.65	0.00	15.96	36.20
68	2022-11-09 11:10	6.08	203.65	-0.94	0.00	16.06	40.44
69	2022-11-09 11:20	5.82	191.49	-1.06	0.00	15.91	34.54
70	2022-11-09 11:30	5.26	139.87	-0.12	0.00	15.94	38.50
71	2022-11-09 11:40	5.32	141.69	-0.53	0.00	16.27	37.21
72	2022-11-09 11:50	5.26	141.63	-2.06	0.00	16.04	50.13
73	2022-11-09 12:00	5.92	194.82	-1.09	0.00	15.84	54.50
74	2022-11-09 12:10	6.25	232.89	142.78	24.00	15.87	50.55
75	2022-11-09 12:20	7.24	359.76	269.17	44.00	15.84	31.64
76	2022-11-09 12:30	6.70	280.36	192.01	33.00	15.78	30.91

77	2022-11-09 12:40	5.85	183.17	119.00	20.00	15.71	30.78
78	2022-11-09 12:50	5.77	179.88	132.90	23.00	15.61	30.79
79	2022-11-09 13:00	5.24	134.94	105.88	17.00	15.55	35.17
80	2022-11-09 13:10	5.14	133.06	112.05	18.00	15.54	41.85
81	2022-11-09 13:20	5.58	159.96	123.62	20.00	15.53	30.65
82	2022-11-09 13:30	5.27	127.31	86.15	15.00	15.54	30.60
83	2022-11-09 13:40	3.52	40.81	20.73	3.00	15.52	27.84
84	2022-11-09 13:50	3.92	55.12	27.76	5.00	15.40	35.69
85	2022-11-09 14:00	3.90	55.06	32.59	6.00	15.43	34.95
86	2022-11-09 14:10	5.37	152.79	124.73	18.00	15.26	52.66
87	2022-11-09 14:20	5.53	154.31	143.28	25.00	15.18	58.71
88	2022-11-09 14:30	4.73	97.60	80.97	13.00	15.16	42.85
89	2022-11-09 14:40	5.04	116.96	82.68	14.00	15.01	33.42
90	2022-11-09 14:50	5.79	182.30	138.36	23.00	14.76	41.44
91	2022-11-09 15:00	7.14	336.54	310.59	52.00	14.42	51.14
92	2022-11-09 15:10	6.62	271.77	287.39	47.00	14.06	52.54
93	2022-11-09 15:20	7.56	411.27	465.22	78.00	13.92	52.50
94	2022-11-09 15:30	8.45	550.17	626.34	102.00	13.61	52.57
95	2022-11-09 15:40	8.15	499.85	578.37	97.00	13.35	52.62
96	2022-11-09 15:50	8.13	496.48	554.98	92.00	13.09	51.92
97	2022-11-09 16:00	8.14	503.69	583.80	98.00	12.87	51.25
98	2022-11-09 16:10	9.00	649.59	769.24	133.00	12.75	53.98
99	2022-11-09 16:20	9.76	796.43	902.16	155.00	12.46	60.31
100	2022-11-09 16:30	8.43	546.82	636.58	104.00	12.34	60.06
101	2022-11-09 16:40	8.40	549.82	596.43	101.00	12.21	60.19
102	2022-11-09 16:50	7.85	459.01	566.88	94.00	12.19	59.93
103	2022-11-09 17:00	7.85	470.21	512.36	87.00	12.05	59.89
104	2022-11-09 17:10	7.37	389.37	420.90	71.00	11.97	59.66
105	2022-11-09 17:20	7.85	462.17	495.94	80.00	11.89	59.00
106	2022-11-09 17:30	7.42	397.25	374.14	64.00	11.89	58.68
107	2022-11-09 17:40	6.49	258.62	273.96	46.00	11.83	51.35
108	2022-11-09 17:50	6.66	279.99	273.58	46.00	11.83	50.18
109	2022-11-09 18:00	7.17	352.39	363.74	60.00	11.77	54.33
110	2022-11-09 18:10	6.60	275.17	283.48	48.00	11.76	54.04
111	2022-11-09 18:20	6.87	318.27	317.13	52.00	11.73	54.44
112	2022-11-09 18:30	7.54	420.66	414.96	69.00	11.67	54.51
113	2022-11-09 18:40	8.73	617.81	690.77	115.00	11.58	54.26
114	2022-11-09 18:50	8.66	605.07	662.70	111.00	11.56	51.02
115	2022-11-09 19:00	7.94	489.85	513.23	85.00	11.58	50.20
116	2022-11-09 19:10	7.79	462.17	484.63	79.00	11.54	49.88
117	2022-11-09 19:20	7.36	389.46	432.52	69.00	11.47	49.00
118	2022-11-09 19:30	7.63	426.95	474.75	83.00	11.42	48.70
119	2022-11-09 19:40	7.06	349.37	383.74	62.00	11.25	53.30

120	2022-11-09 19:50	6.80	301.62	375.19	65.00	11.20	55.18
121	2022-11-09 20:00	6.93	319.13	327.25	54.00	11.11	55.14
122	2022-11-09 20:10	6.45	250.69	255.21	41.00	11.12	55.38
123	2022-11-09 20:20	6.30	231.10	219.34	37.00	11.10	55.20
124	2022-11-09 20:30	6.07	198.73	213.85	35.00	11.07	55.27
125	2022-11-09 20:40	5.04	113.19	114.27	19.00	11.15	44.29
126	2022-11-09 20:50	2.96	19.20	16.36	3.00	11.17	52.49
127	2022-11-09 21:00	2.38	5.51	-2.38	0.00	11.11	71.90
128	2022-11-09 21:10	2.87	15.42	11.98	3.00	11.16	43.60
129	2022-11-09 21:20	2.23	6.28	-2.73	0.00	11.13	37.95
130	2022-11-09 21:30	2.31	8.09	-0.34	0.00	11.09	68.62
131	2022-11-09 21:40	1.35	4.95	-0.92	0.00	11.45	96.08
132	2022-11-09 21:50	0.69	1.04	-0.41	0.00	12.07	95.78
133	2022-11-09 22:00	0.87	3.70	-1.93	0.00	12.10	95.25
134	2022-11-09 22:10	1.57	5.71	-0.37	0.00	12.02	94.79
135	2022-11-09 22:20	1.23	4.38	-0.58	0.00	12.00	94.88
136	2022-11-09 22:30	0.76	1.09	-1.32	0.00	12.05	94.69
137	2022-11-09 22:40	2.20	5.50	-1.00	0.00	12.09	80.30
138	2022-11-09 22:50	2.39	5.81	-1.91	0.00	12.41	60.27
139	2022-11-09 23:00	2.76	11.27	-1.77	0.00	12.44	70.69
140	2022-11-09 23:10	3.90	47.84	-1.38	0.00	12.38	62.69
141	2022-11-09 23:20	3.74	40.74	-3.35	0.00	11.89	57.82
142	2022-11-09 23:30	3.85	45.23	-1.09	0.00	11.79	59.05
143	2022-11-09 23:40	4.22	59.66	-2.19	0.00	11.79	51.32
144	2022-11-09 23:50	4.45	71.34	-1.04	0.00	11.74	51.06
145	2022-11-10 00:00	4.39	69.16	-0.61	0.00	11.75	60.56
146	2022-11-10 00:10	4.34	66.78	-2.23	0.00	11.99	58.94
147	2022-11-10 00:20	4.02	52.98	-0.34	0.00	11.99	62.39
148	2022-11-10 00:30	4.03	52.20	-1.99	0.00	11.69	57.04
149	2022-11-10 00:40	4.07	54.49	-0.88	0.00	11.61	47.86
150	2022-11-10 00:50	4.33	67.03	-0.79	0.00	11.50	47.90
151	2022-11-10 01:00	4.63	82.00	-2.32	0.00	11.33	43.27
152	2022-11-10 01:10	4.16	57.11	-1.42	0.00	11.59	47.74
153	2022-11-10 01:20	3.57	34.76	-1.62	0.00	11.55	37.00
154	2022-11-10 01:30	3.13	19.69	-2.32	0.00	11.58	46.01
155	2022-11-10 01:40	3.47	33.34	-3.26	0.00	11.80	41.67
156	2022-11-10 01:50	3.68	37.81	-2.01	0.00	12.10	39.74
157	2022-11-10 02:00	3.14	19.57	-0.47	0.00	12.10	46.75
158	2022-11-10 02:10	3.15	20.40	-2.71	0.00	12.18	56.18
159	2022-11-10 02:20	2.70	9.30	-1.52	0.00	12.05	40.33
160	2022-11-10 02:30	2.68	8.07	-0.76	0.00	11.82	48.76
161	2022-11-10 02:40	2.58	7.80	-3.34	0.00	11.85	47.77
162	2022-11-10 02:50	2.12	5.60	-1.00	0.00	12.01	61.82

163	2022-11-10 03:00	2.16	6.41	-1.41	0.00	12.10	48.19
164	2022-11-10 03:10	1.91	5.48	-1.59	0.00	12.16	53.60
165	2022-11-10 03:20	2.68	9.95	-0.53	0.00	12.29	62.04
166	2022-11-10 03:30	2.50	8.42	-2.39	0.00	12.30	83.81
167	2022-11-10 03:40	2.36	9.93	-0.39	0.00	12.30	85.78
168	2022-11-10 03:50	2.55	11.42	-1.54	0.00	12.25	76.08
169	2022-11-10 04:00	2.57	10.35	-1.04	0.00	12.18	63.19
170	2022-11-10 04:10	3.41	29.26	-0.90	0.00	12.10	58.91
171	2022-11-10 04:20	3.14	20.69	-2.53	0.00	12.04	54.36
172	2022-11-10 04:30	3.31	27.23	-0.53	0.00	11.99	36.96
173	2022-11-10 04:40	3.61	37.25	-2.28	0.00	12.00	33.46
174	2022-11-10 04:50	3.65	38.31	-1.06	0.00	12.01	39.80
175	2022-11-10 05:00	3.32	28.35	-1.89	0.00	12.03	51.45
176	2022-11-10 05:10	3.65	39.19	-2.01	0.00	12.03	52.33
177	2022-11-10 05:20	3.98	51.06	-0.53	0.00	12.00	41.20
178	2022-11-10 05:30	3.10	20.96	-1.98	0.00	11.97	37.45
179	2022-11-10 05:40	2.87	13.37	-0.68	0.00	11.96	42.34
180	2022-11-10 05:50	2.87	13.91	-5.09	0.00	11.90	86.58
181	2022-11-10 06:00	2.64	8.14	-6.35	0.00	11.90	166.70
182	2022-11-10 06:10	2.52	7.44	-0.39	0.00	11.99	40.43
183	2022-11-10 06:20	2.52	5.98	-2.34	0.00	12.05	40.07
184	2022-11-10 06:30	2.75	11.22	-1.63	0.00	12.08	36.70
185	2022-11-10 06:40	2.97	15.09	-1.93	0.00	12.10	44.97
186	2022-11-10 06:50	3.04	18.51	-2.12	0.00	12.23	41.63
187	2022-11-10 07:00	3.08	18.78	-0.32	0.00	12.30	56.61
188	2022-11-10 07:10	2.90	13.04	-2.32	0.00	12.36	58.20
189	2022-11-10 07:20	2.38	3.91	-0.44	0.00	12.51	40.14
190	2022-11-10 07:30	1.82	6.52	-1.09	0.00	12.63	39.37
191	2022-11-10 07:40	2.13	4.01	-2.28	0.00	12.99	54.35
192	2022-11-10 07:50	1.89	7.05	-0.41	0.00	13.28	34.38
193	2022-11-10 08:00	1.68	6.05	-1.69	0.00	13.39	34.41
194	2022-11-10 08:10	1.61	5.80	-0.66	0.00	13.56	34.41
195	2022-11-10 08:20	1.25	6.20	-0.78	0.00	14.00	24.67
196	2022-11-10 08:30	1.13	4.18	-1.38	0.00	14.48	42.02
197	2022-11-10 08:40	0.69	2.31	-0.85	0.00	14.20	41.26
198	2022-11-10 08:50	0.38	0.23	-0.21	0.00	14.99	39.46
199	2022-11-10 09:00	0.49	0.36	-0.13	0.00	15.00	39.05
200	2022-11-10 09:10	0.75	1.20	-0.10	0.00	14.95	38.40
201	2022-11-10 09:20	1.14	5.12	-0.09	0.00	14.99	35.62
202	2022-11-10 09:30	1.47	4.60	-1.15	0.00	15.40	87.47
203	2022-11-10 09:40	1.41	5.24	-5.82	0.00	14.86	314.08
204	2022-11-10 09:50	1.05	2.80	-0.40	0.00	15.45	312.04
205	2022-11-10 10:00	0.85	1.88	-0.11	0.00	16.61	312.23

206	2022-11-10 10:10	1.00	2.91	-0.14	0.00	17.01	312.06
207	2022-11-10 10:20	1.52	4.99	-0.59	0.00	17.09	313.76
208	2022-11-10 10:30	2.23	10.17	-1.86	0.00	16.99	70.66
209	2022-11-10 10:40	3.00	25.15	-1.18	0.00	16.88	28.57
210	2022-11-10 10:50	2.71	15.28	-0.76	0.00	16.89	60.96
211	2022-11-10 11:00	3.19	25.16	-0.15	0.00	16.89	33.69
212	2022-11-10 11:10	3.41	32.45	-1.19	0.00	17.34	40.03
213	2022-11-10 11:20	4.04	55.69	-1.31	0.00	17.03	34.82
214	2022-11-10 11:30	3.71	46.09	-0.55	0.00	17.25	42.67
215	2022-11-10 11:40	4.20	70.25	-0.91	0.00	17.34	33.84
216	2022-11-10 11:50	3.92	60.93	-0.37	0.00	17.05	44.45
217	2022-11-10 12:00	4.41	80.04	-1.50	0.00	17.20	34.33
218	2022-11-10 12:10	4.73	95.79	-1.31	0.00	17.53	34.66
219	2022-11-10 12:20	5.78	191.33	-1.10	0.00	17.49	36.43
220	2022-11-10 12:30	6.08	209.90	-0.38	0.00	17.72	36.84
221	2022-11-10 12:40	6.50	272.71	-0.56	0.00	17.66	44.94
222	2022-11-10 12:50	5.34	152.75	-1.75	0.00	17.83	41.06
223	2022-11-10 13:00	4.78	135.20	-1.85	0.00	18.02	55.16
224	2022-11-10 13:10	5.84	213.66	-0.72	0.00	18.22	55.50
225	2022-11-10 13:20	7.05	386.97	175.42	28.00	17.64	35.95
226	2022-11-10 13:30	7.15	371.34	295.30	50.00	17.97	38.61
227	2022-11-10 13:40	7.43	404.56	290.11	48.00	18.22	31.52
228	2022-11-10 13:50	7.51	422.20	333.27	57.00	17.91	31.88
229	2022-11-10 14:00	6.89	336.16	228.03	37.00	17.86	30.47
230	2022-11-10 14:10	7.78	486.93	382.37	66.00	17.85	41.13
231	2022-11-10 14:20	8.30	571.48	469.64	76.00	17.88	59.42
232	2022-11-10 14:30	8.05	515.27	507.18	84.00	17.68	59.00
233	2022-11-10 14:40	8.06	523.84	476.92	80.00	17.71	51.63
234	2022-11-10 14:50	8.25	544.55	475.63	75.00	17.72	39.69
235	2022-11-10 15:00	8.94	665.00	567.71	94.00	17.32	39.34
236	2022-11-10 15:10	9.08	681.54	487.85	81.00	17.19	39.34
237	2022-11-10 15:20	9.55	771.12	611.20	103.00	17.20	39.40
238	2022-11-10 15:30	9.24	716.80	623.01	104.00	17.23	39.40
239	2022-11-10 15:40	8.87	643.13	494.95	83.00	17.06	39.43
240	2022-11-10 15:50	9.77	807.35	724.53	120.00	16.75	39.39
241	2022-11-10 16:00	9.59	773.03	759.89	128.00	16.57	42.93
242	2022-11-10 16:10	10.09	873.63	749.20	124.00	16.09	43.54
243	2022-11-10 16:20	9.17	694.15	639.53	108.00	15.81	43.61
244	2022-11-10 16:30	9.26	688.95	719.29	116.00	15.51	43.20
245	2022-11-10 16:40	8.80	619.95	654.98	110.00	15.25	43.70
246	2022-11-10 16:50	8.64	594.77	644.96	105.00	15.14	43.37
247	2022-11-10 17:00	8.15	517.67	564.08	97.00	14.89	53.91
248	2022-11-10 17:10	8.17	525.50	516.46	83.00	14.66	58.76

249	2022-11-10 17:20	8.38	551.21	570.48	93.00	14.41	58.63
250	2022-11-10 17:30	8.66	605.68	570.59	94.00	14.12	57.43
251	2022-11-10 17:40	8.12	510.30	548.65	88.00	13.97	42.62
252	2022-11-10 17:50	8.55	575.59	696.12	118.00	13.60	50.00
253	2022-11-10 18:00	8.06	495.37	568.59	95.00	13.41	64.18
254	2022-11-10 18:10	8.49	565.32	573.58	97.00	13.19	53.45
255	2022-11-10 18:20	8.47	574.20	558.97	91.00	12.97	58.52
256	2022-11-10 18:30	8.79	624.01	577.49	97.00	12.81	58.07
257	2022-11-10 18:40	8.26	534.31	485.69	81.00	12.63	58.27
258	2022-11-10 18:50	8.37	548.75	571.80	96.00	12.54	57.97
259	2022-11-10 19:00	8.33	538.09	628.86	104.00	12.38	57.84
260	2022-11-10 19:10	8.08	502.78	572.60	100.00	12.28	57.68
261	2022-11-10 19:20	7.74	452.77	492.31	81.00	12.22	56.98
262	2022-11-10 19:30	7.03	344.05	345.37	59.00	12.15	52.30
263	2022-11-10 19:40	5.55	177.86	179.92	33.00	12.17	48.15
264	2022-11-10 19:50	4.78	106.08	92.71	17.00	12.16	61.16
265	2022-11-10 20:00	3.71	48.68	45.85	7.00	12.22	60.75
266	2022-11-10 20:10	3.93	51.85	52.56	9.00	12.12	50.20
267	2022-11-10 20:20	3.60	36.66	32.40	5.00	12.26	56.76
268	2022-11-10 20:30	2.13	7.54	-2.87	0.00	12.68	125.81
269	2022-11-10 20:40	1.35	4.33	-1.33	0.00	12.70	245.98
270	2022-11-10 20:50	3.51	38.21	10.81	2.00	12.34	170.57
271	2022-11-10 21:00	2.46	13.35	5.71	2.00	12.36	69.14
272	2022-11-10 21:10	1.40	4.26	-3.89	0.00	12.60	93.46
273	2022-11-10 21:20	2.33	7.09	-1.20	0.00	13.05	72.61
274	2022-11-10 21:30	2.24	4.84	-0.44	0.00	13.07	80.86
275	2022-11-10 21:40	2.10	4.77	-0.56	0.00	13.03	92.77
276	2022-11-10 21:50	1.56	7.37	-0.60	0.00	13.02	94.56
277	2022-11-10 22:00	3.24	26.17	-2.42	0.00	12.75	74.26
278	2022-11-10 22:10	4.19	56.70	-0.39	0.00	12.47	90.92
279	2022-11-10 22:20	4.77	88.52	-0.76	0.00	12.44	68.70
280	2022-11-10 22:30	5.66	156.28	-1.97	0.00	12.33	62.14
281	2022-11-10 22:40	5.82	164.59	-0.55	0.00	12.27	53.29
282	2022-11-10 22:50	5.40	129.44	-0.55	0.00	12.17	67.27

ANEXO B

ANÁLISIS DE SENSIBILIDAD DE LA TÉCNICA DE SARIMA

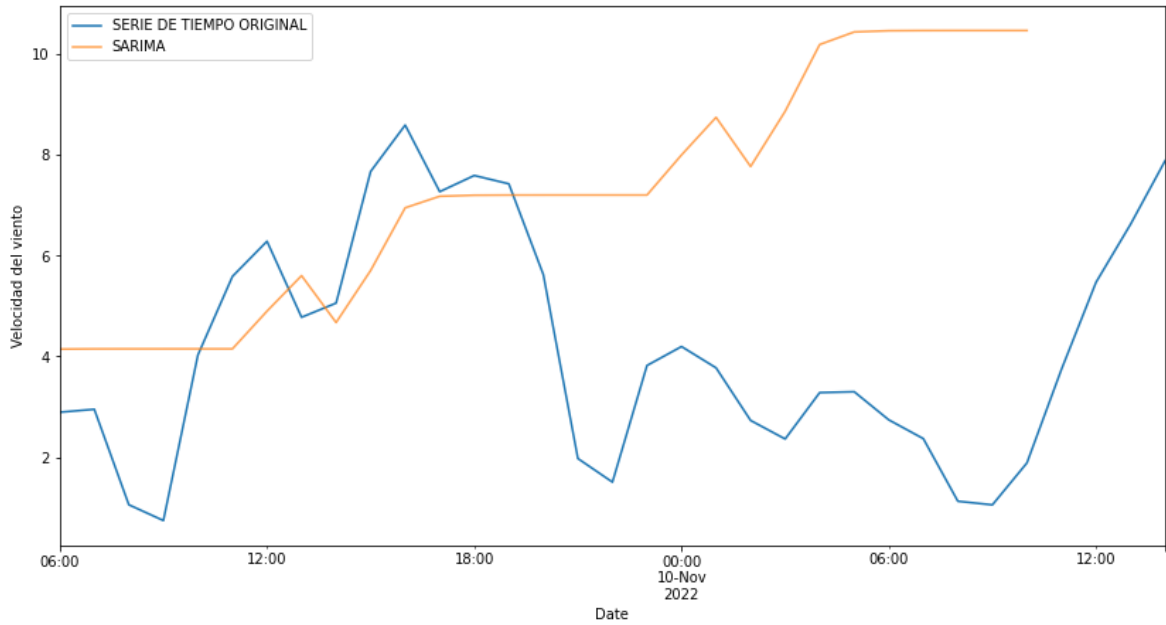


Figura A1. Desarrollo de serie temporal con ventana de tiempo: '2022-11-09 06:00:00' al '2022-11-10 11:00:00'

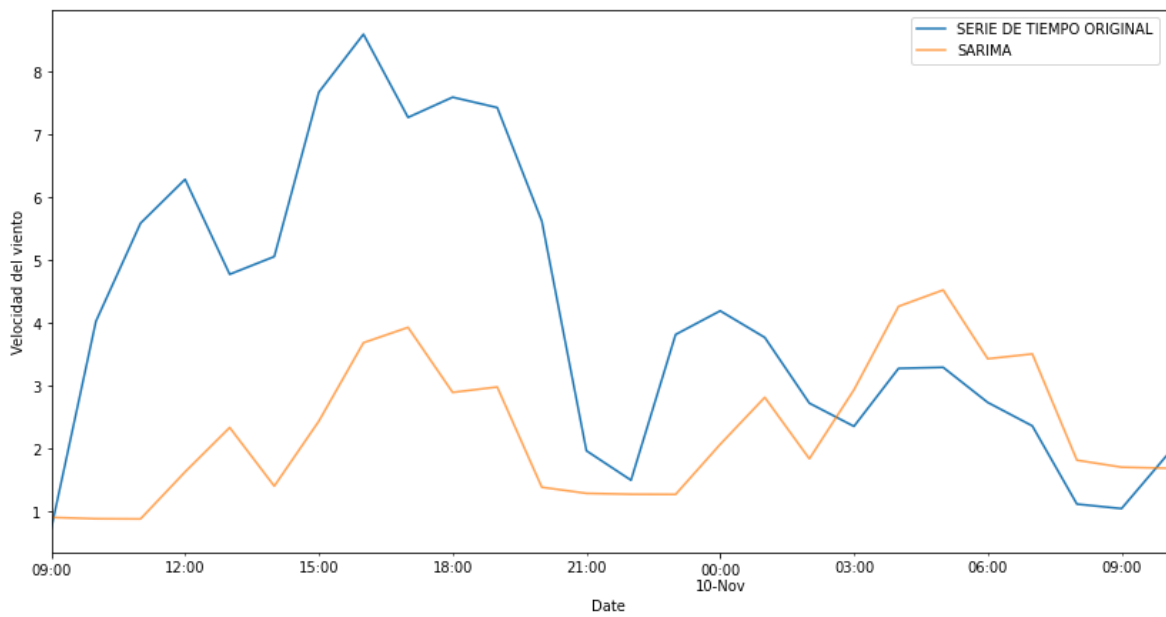


Figura A2. Desarrollo de serie temporal con ventana de tiempo: '2022-11-09 09:00:00' al '2022-11-09 10:00:00'

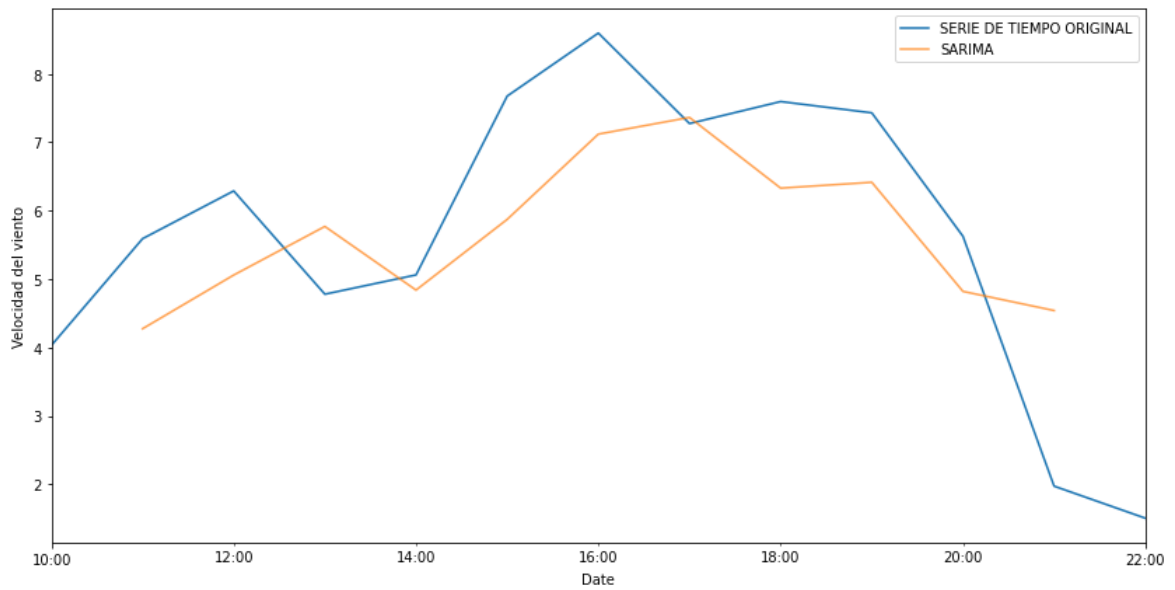


Figura A3. Desarrollo de serie temporal con ventana de tiempo: '2022-11-09 11:00:00' al '2022-11-09 21:00:00'

ANÁLISIS DE SENSIBILIDAD DE LA TÉCNICA DE AUTOREGRESIVA

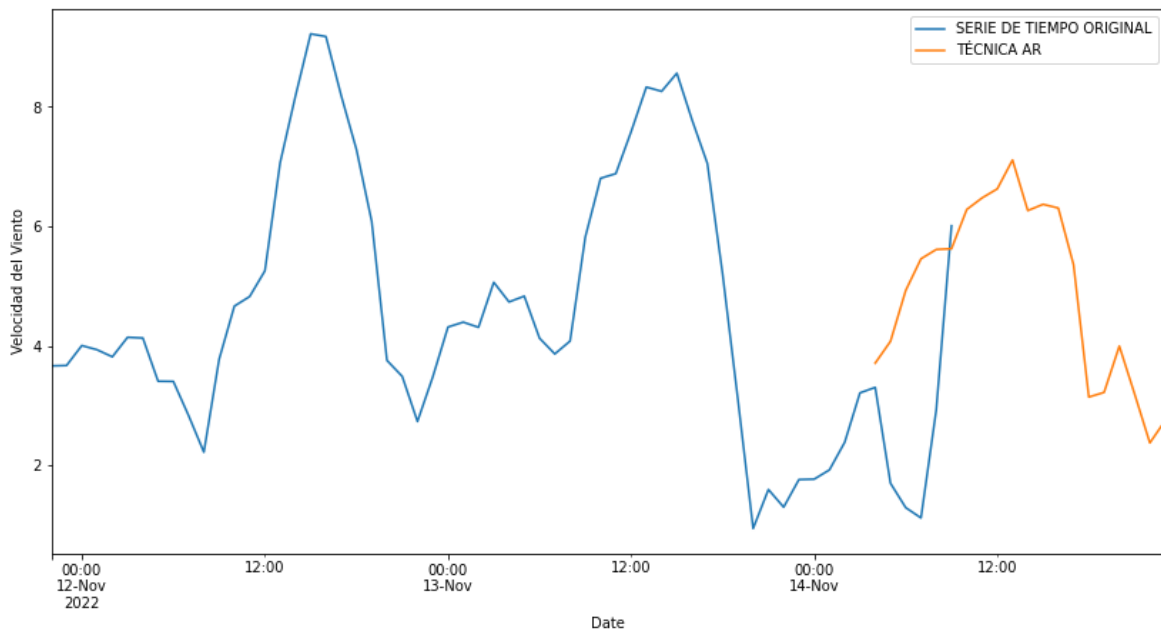


Figura A4. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-14 18:00:00'

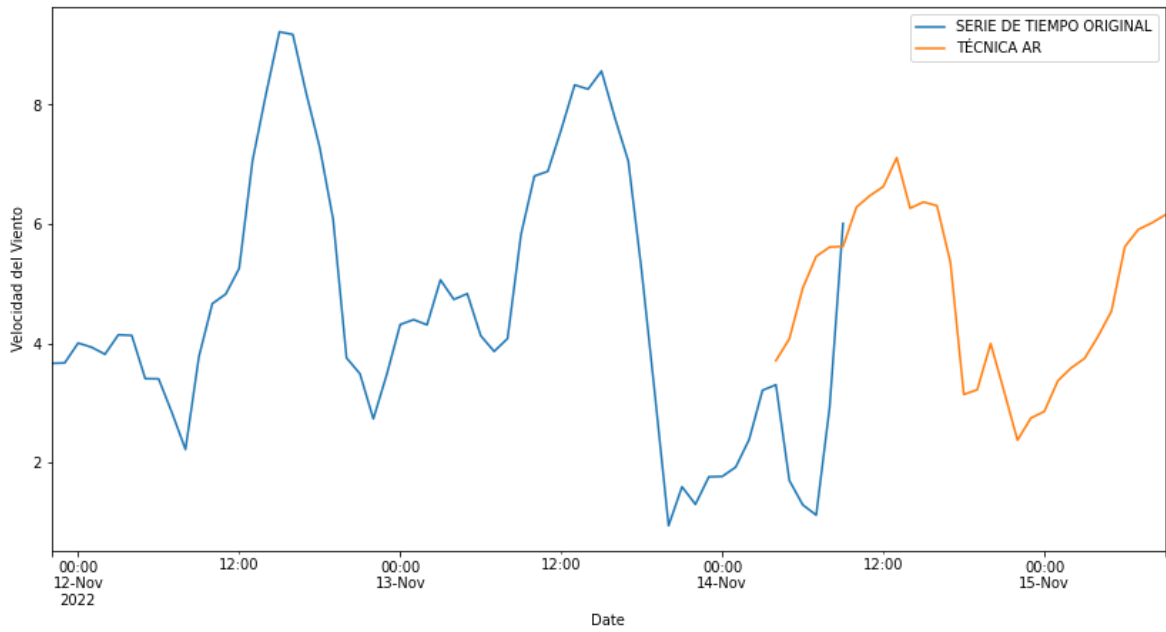


Figura A5. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-15 06:00:00'

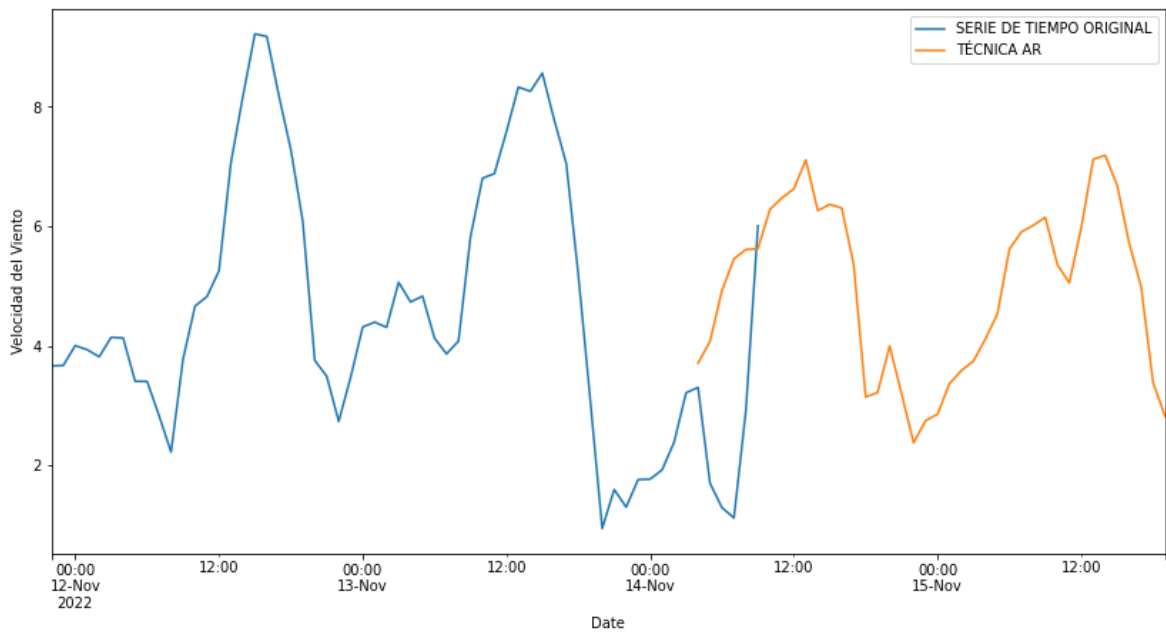


Figura A6. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-14 14:00:00'

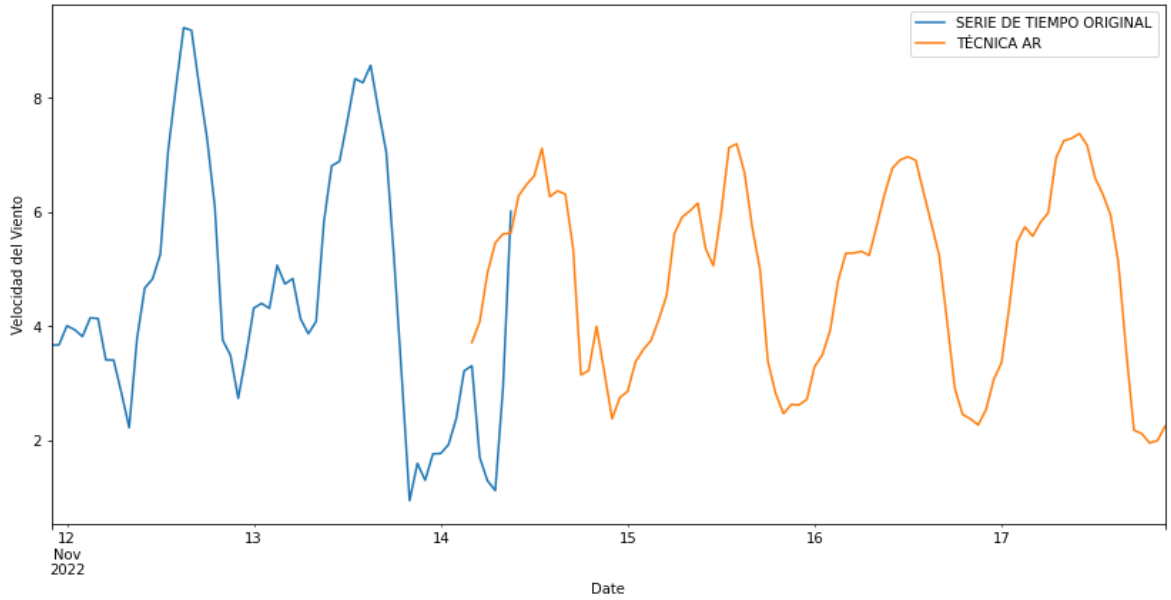


Figura A7. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-17 16:00:00'

ANÁLISIS DE SENSIBILIDAD DE LA TÉCNICA DE AUTOREGRESIVA

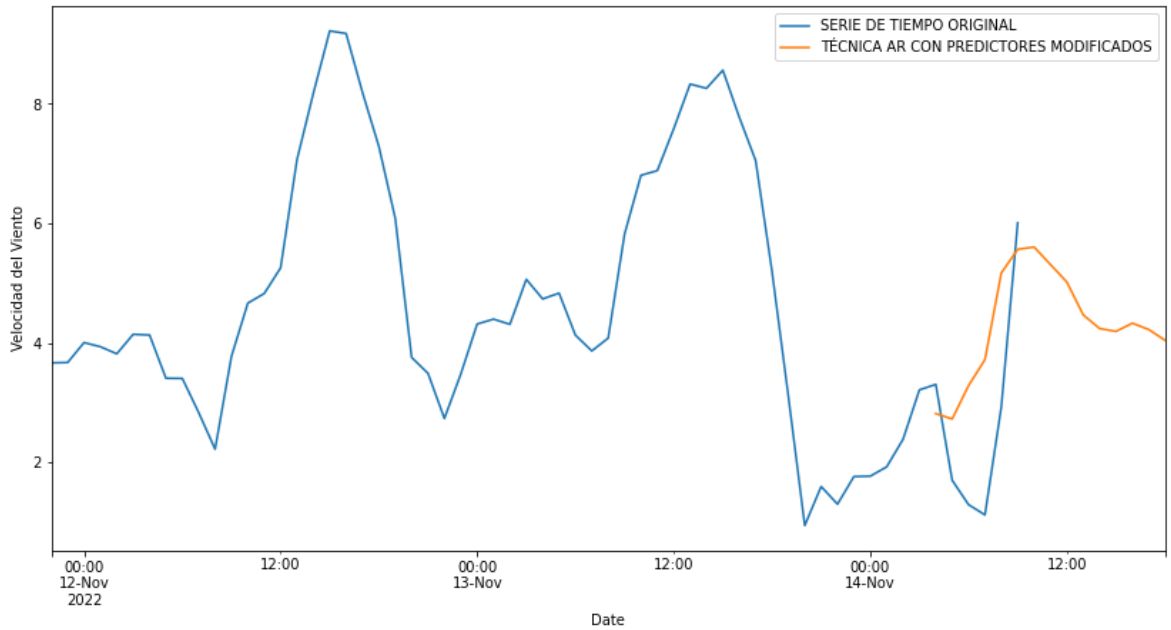


Figura A8. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-14 14:00:00'

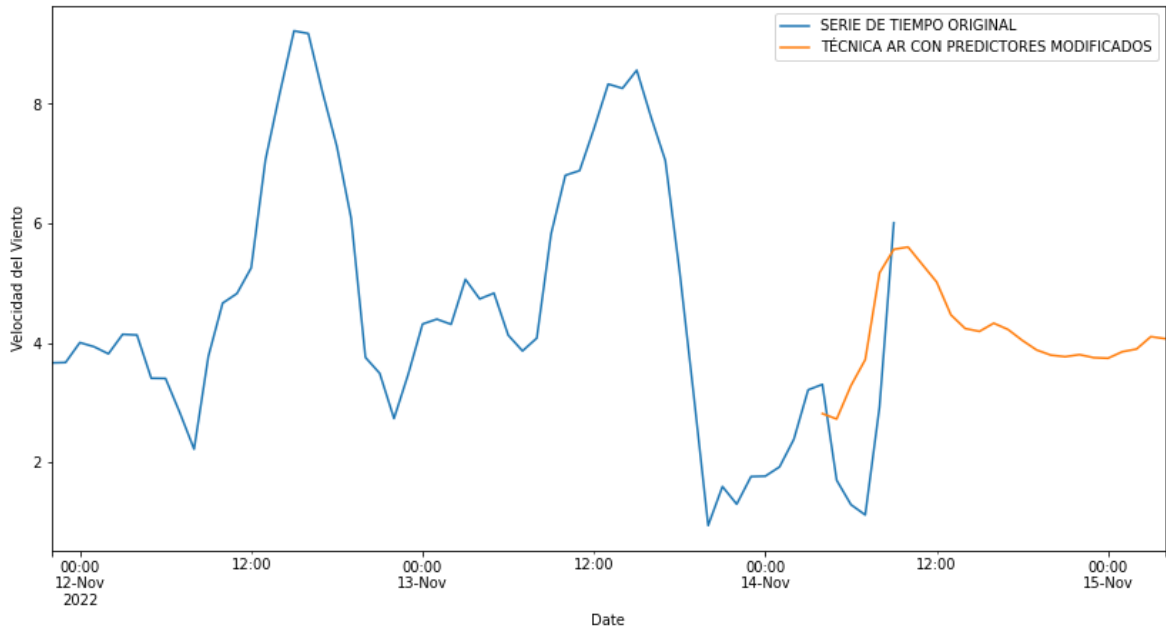


Figura A9. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-15 06:00:00'

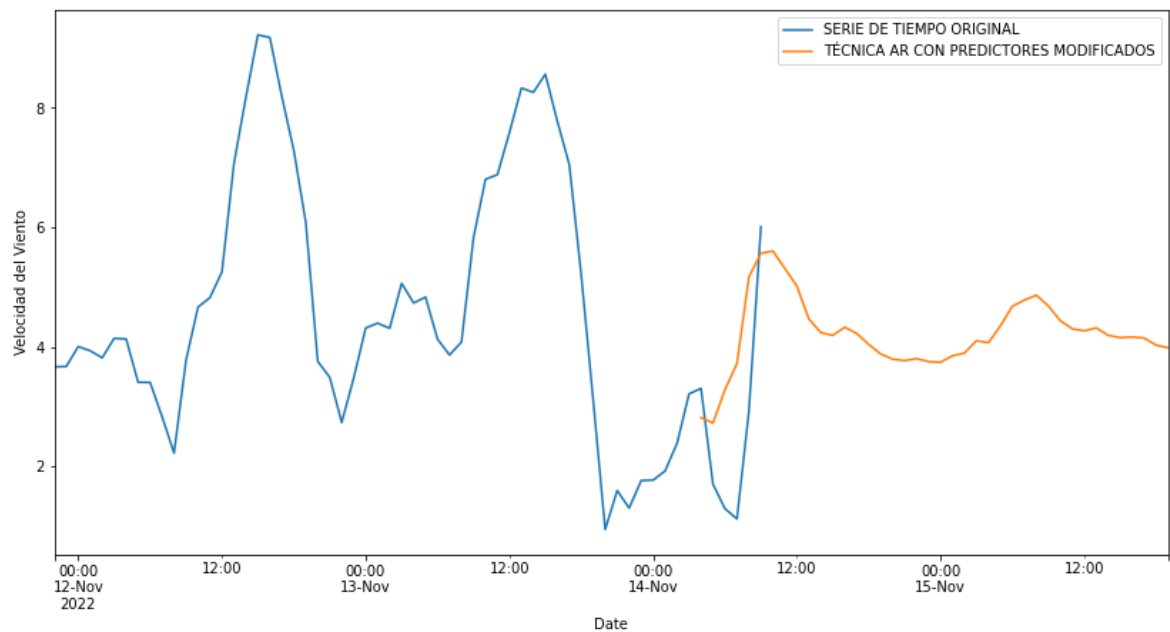


Figura A10. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-15 14:00:00'

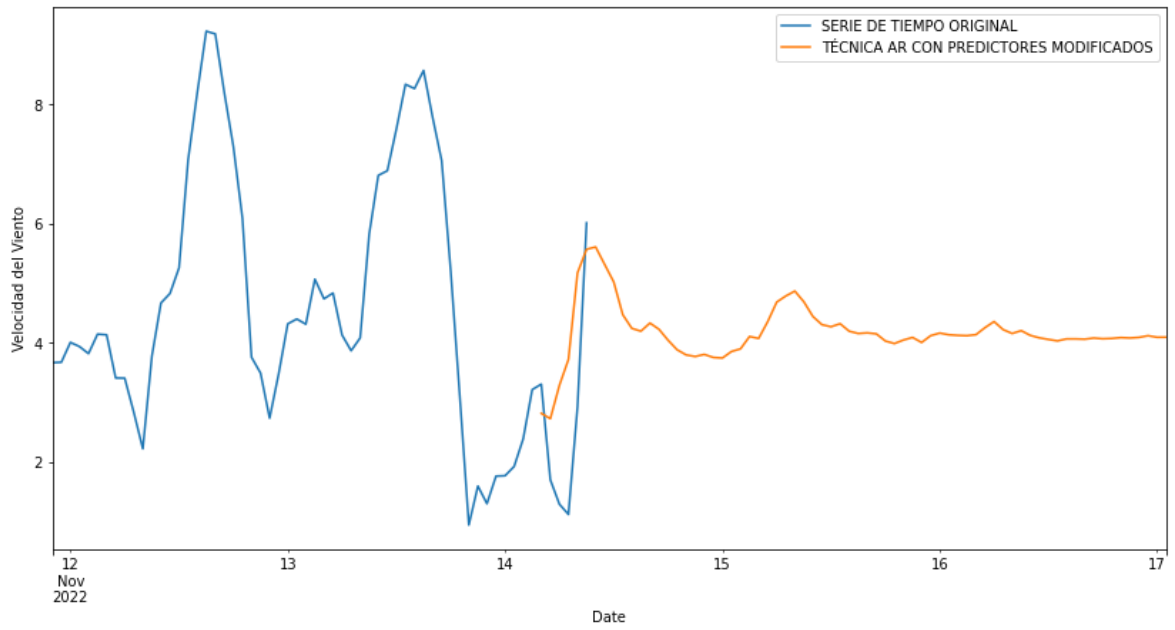


Figura A11. Desarrollo de serie temporal con ventana de tiempo: '2022-11-14 06:00:00' al '2022-11-17 02:00:00'

ORDEN DE EMPASTADO