



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E
INFORMÁTICA
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO
MAESTRÍA EN CIENCIAS COMPUTACIONALES Y
TECNOLOGÍAS DE LA INFORMACIÓN

PREDICCIÓN DE GENERACIÓN DE ENERGÍA ELÉCTRICA
CON UN DÍA DE ANTICIPO MEDIANTE MODELOS DE
APRENDIZAJE SUPERVISADO A PARTIR DE DATOS
EÓLICOS

TESIS

QUE PARA OBTENER EL GRADO DE
MAESTRA EN CIENCIAS COMPUTACIONALES Y
TECNOLOGÍAS DE LA INFORMACIÓN

P R E S E N T A:

I.S.C. SARA EDITH PINZÓN PINEDA

ASESORES:

DR. MARTÍN HERIBERTO CRUZ ROSALES

DR. GUILLERMO SANTAMARÍA BONFÍL



Facultad de Contaduría,
Administración e Informática

CUERNAVACA, MORELOS
MAYO 2019

JURADO ASIGNADO

_____ : DR. JOSÉ ALBERTO HERNÁNDEZ AGUILAR

_____ : DR. GUILLERMO SANTAMARÍA BONFÍL

_____ : DR. FEDERICO ALONSO PECINA

_____ : DR. MARTÍN HERIBERTO CRUZ ROSALES

_____ : DR. JOSÉ CRISPÍN ZAVALA DÍAZ

**PREDICCIÓN DE GENERACIÓN DE
ENERGÍA ELÉCTRICA CON UN DÍA DE
ANTICIPO MEDIANTE MODELOS DE
APRENDIZAJE SUPERVISADO A
PARTIR DE DATOS EÓLICOS**

DEDICATORIA

Este presente trabajo lo dedico a:

Dios, por su amor infinito, por sostenerme siempre y por estar conmigo a lo largo de mi vida.

A mi hijo Gael, por todo su apoyo, por ser mi motor de vida y ser un hijo ejemplar, comprensivo e inteligente que Dios envió a mi vida.

A mi esposo, por todo su amor, palabras de aliento y comprensión.

A mi madre, por su ejemplo de luchadora incansable, porque te admiro y respeto muchísimo.

A mis hermanos, por creer en mí, por su ejemplo y por su amor.

A mi tío Chavis, Tey, Mamá Lolita y Papá Luis, quienes han aportado en mi vida un gran ejemplo de vida, lleno de amor y me inspiran a salir adelante.

AGRADECIMIENTOS

Agradezco a:

La UAEM, por haberme permitido realizar mis estudios de maestría, la tesis y con ella la obtención del grado.

A los Dres. que me acompañaron y asesoraron en este valioso proceso, al Dr. Alberto Hernández, Dr. Guillermo Santamaría, Dr. Gustavo Arroyo, Dr. Martín Heriberto Cruz y revisores Dr. Federico Alonso Pecina, Dr. Martín Heriberto Cruz Rosales y Dr. José Crispín Zavala Díaz quienes hicieron posible este importante proceso en mi formación profesional.

Al Ing. Jaime Agredano Díaz, Dr. Salvador González Castro y Dr. Alberto Reyes Ballesteros por su valioso apoyo, lo cual me permitió desarrollar este trabajo.

RESÚMEN

Actualmente existe una preocupación ambiental considerable a nivel mundial, que ha dado origen una iniciativa importante sobre el uso de las energías limpias para diversos sectores. Un tercio del total de la contaminación generada a nivel mundial proviene del proceso de producción de electricidad; el cual depende en gran parte del uso de combustibles de alto contenido de CO₂. En la búsqueda de un futuro con bajas emisiones de CO₂ se propone como alternativa, el uso de Energías Limpias, lo que representa un desafío interesante sobre el tema de generación, transmisión y distribución de energía en la red eléctrica.

El mayor desafío para el caso de la generación de energía eólica, es que es totalmente dependiente del clima, lo que origina un alto grado de intermitencia en varias escalas del tiempo, así como también, no existe un mecanismo eficaz para almacenar a gran escala la energía generada para disponer de esta en momentos de gran demanda dentro de la red eléctrica. Lo que origina problemas de estabilidad en la red eléctrica.

En este trabajo, se discuten los modelos estadísticos auto regresivos clásicos Box and Jenkins aplicados a problemas de predicción y se analiza su aplicación en el sector energético. Posteriormente, se propone una metodología basada en la aplicación de estos modelos para la predicción de generación de energía eléctrica a partir de datos urbanos a una altura de 40 m. obtenidos de la Ciudad de San Francisco los cuales son medidos por estaciones meteorológicas instaladas por la SFPUC (Comisión de Servicios Públicos de San Francisco) que monitorean las características del viento en San Francisco, y datos convencionales a una altura de hasta 60 m, en un parque eólico denominado CERTE de la Ventosa de una estación anemométrica denominada la Ventosa2, ubicada en la Ciudad de Juchitán de Zaragoza, Oaxaca, México. Finalmente, se presenta el modelo y los resultados de la predicción obtenidos para la zona urbana de San Francisco y la convencional del CERTE basados en el desarrollo de la presente tesis.

Palabras clave: Modelos auto regresivos, Box and Jenkins, Datos eólicos, Predicción, Generación eólica, Energía eléctrica.

CONTENIDO

<i>DEDICATORIA</i>	4
<i>AGRADECIMIENTOS</i>	5
<i>RESÚMEN</i>	6
<i>INDICE DE TABLAS</i>	9
<i>INDICE DE FIGURAS</i>	10
<i>INDICE DE ECUACIONES</i>	11
<i>CAPÍTULO 1</i>	12
<i>Introducción</i>	12
1.1. <i>Planteamiento del problema y justificación</i>	13
1.2. <i>Hipótesis</i>	14
1.3. <i>Objetivos del trabajo de investigación</i>	15
1.3.1 <i>Objetivo general</i>	15
1.3.2 <i>Objetivos específicos</i>	15
1.4. <i>Estructura de la tesis</i>	16
<i>CAPÍTULO 2</i>	17
<i>Marco teórico</i>	17
2.1. <i>Antecedentes</i>	17
2.2. <i>Potencial de producción eólica en base al origen de los vientos</i>	18
2.3. <i>Generación Convencional</i>	19
2.4. <i>Generación Urbana</i>	21
2.5. <i>Ciencia de Datos</i>	22
2.6. <i>Exploración de Metodologías de Predicción de Generación de Energía Eléctrica a partir de Datos Eólicos</i>	24
2.6.1. <i>Metodología de modelos autoregresivos Box and Jenkins</i>	28

CAPÍTULO 3.....	34
<i>Metodología propuesta</i>	34
3.1. <i>RMSE y MAE</i>	36
CAPÍTULO 4.....	39
<i>Experimentación</i>	39
4.1. <i>Recopilación de datos urbanos de San Francisco</i>	39
4.2. <i>Pre procesamiento de los datos de San Francisco</i>	40
4.2.1. <i>Limpieza y normalización sobre los datos de San Francisco</i>	40
4.2.2. <i>Análisis exploratorio de los datos</i>	42
4.2.3. <i>Interpolación de la serie de tiempo con los datos de San Francisco</i>	43
4.3. <i>Selección de los parámetros y entrenamiento del modelo sobre los datos de San Francisco</i>	46
4.4. <i>Pronóstico con comprobación y pruebas con los mejores 5 modelos sobre los datos de San Francisco en base a su menor RMSE y MAE</i>	48
4.5. <i>Recopilación de los datos de un Parque eólico denominado CERTE</i>	52
4.6. <i>Pre procesamiento de los datos del CERTE</i>	53
4.6.1. <i>Limpieza y normalización de los datos del CERTE</i>	53
4.6.2. <i>Análisis exploratorio de los datos</i>	54
4.6.3. <i>Interpolación de la serie de tiempo</i>	55
4.7. <i>Selección de los parámetros y entrenamiento del modelo sobre los datos del CERTE</i> ..	58
4.8. <i>Pronóstico con comprobación y pruebas con los 5 mejores modelos sobre los datos del CERTE para recalcular su RMSE y MAE</i>	61
4.9. <i>Análisis y resultados</i>	65
CAPÍTULO 5.....	67
<i>Conclusiones y trabajo futuro</i>	67
REFERENCIAS.....	69
APÉNDICES.....	72

INDICE DE TABLAS

Tabla 1. Datos de San Francisco (sfgov.org, 2019).....	40
Tabla 2. Momentos estadísticos de 501 Twin Peaks Blvd. San Francisco – Velocidad del viento ...	41
Tabla 3. Interpolación con Datos de San Francisco (WS-4B & 4C/ 1-2 kW).....	45
Tabla 4. División de los datos para el experimento sobre la Serie de Tiempo	46
Tabla 5. 72 combinaciones de ARIMA y sus RMSE con datos de San Francisco. $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$	47
Tabla 6. Medidas de error de las mejores 5 torres de las 72 combinaciones de San Francisco	48
Tabla 7. Medidas de error de las mejores torres de San Francisco en potencia (W)	49
Tabla 8. Momentos estadísticos para los datos del CERTE – Velocidad del viento (m/s).....	54
Tabla 9. Especificaciones de la torre denominada T300-28	55
Tabla 10. Interpolación con Datos del CERTE en Kw (T300-28).....	57
Tabla 11. División de los datos para el experimento sobre la Serie de Tiempo	58
Tabla 12. 72 combinaciones de ARIMA y sus RMSE con datos del CERTE con $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$	59
Tabla 13. Medidas de error - RMSE de los 5 mejores parámetros del modelo para el CERTE.....	60
Tabla 14. Medidas de error de las mejores torres del CERTE con una muestra de 31,564 datos en kW.....	61
Tabla 15. Momentos estadísticos con datos de San Francisco y CERTE sobre la velocidad del viento (m/s).....	65
Tabla 16. Mejores modelos para ambos casos en kW	66
Tabla 17. Mejores modelos en porcentaje en kW para San Francisco y CERTE	66

INDICE DE FIGURAS

Figura 1. Distribución geográfica del recurso eólico potencial (Fuente: CFE, EnergíaUNAM) __	20
Figura 2. Proceso de Ciencia de Datos (Data Science Process) (Schutt, 2013) _____	23
Figura 3. Métodos de predicción de Generación de Velocidad del Viento (Box, 1976) (Lei, 2009) (Lawan, 2014) _____	25
Figura 4. Ubicación de las 26 torres de la SFPUC (Comisión de Servicios Públicos de San Francisco) _____	39
Figura 5. Mapeo de torres de San Francisco con base en su media _____	41
Figura 6. Rosa del viento con datos de Tween Peaks (Región urbana) – Cd. Y Condado de San Francisco, California _____	42
Figura 7. Oy Windside WS-12 references (WS-12/ 8 kW) _____	44
Figura 8. Función arima y predict de la librería stats en R para San Francisco _____	49
Figura 9. Ejercicio de validación con el modelo ARIMA (0,0,1) de San Francisco _____	50
Figura 10. Ejercicio de validación con el modelo ARIMA (0,0,2) de San Francisco _____	50
Figura 11. Ejercicio de validación con el modelo ARIMA (0,0,3) de San Francisco _____	51
Figura 12. Ejercicio de validación con el modelo ARIMA (0,0,4) de San Francisco _____	51
Figura 13. Ejercicio de validación con el modelo ARIMA (0,0,5) de San Francisco _____	52
Figura 14. Ubicación del CERTE la Ventosa 2 en la Ciudad de Juchitán de Zaragoza, Oaxaca. México _____	53
Figura 15. Rosa del viento con datos del CERTE (Región convencional) – Cd. de Juchitán de Zaragoza, Oaxaca, México. _____	54
Figura 16. Curva de potencia de T300-28 (Turbowinds) _____	56
Figura 17. Función arima y predict de la librería stats en R para CERTE _____	61
Figura 18. Ejercicio de validación con el modelo ARIMA (0,0,1) del CERTE _____	62
Figura 19. Ejercicio de validación con el modelo ARIMA (0,0,2) del CERTE _____	63
Figura 20. Ejercicio de validación con el modelo ARIMA (0,0,3) del CERTE _____	63
Figura 21. Ejercicio de validación con el modelo ARIMA (0,0,4) del CERTE _____	64
Figura 22. Ejercicio de validación con el modelo ARIMA (0,0,5) del CERTE _____	64

INDICE DE ECUACIONES

<i>Ecuación 1. Modelo AR(Autorregresivo)</i>	29
<i>Ecuación 2. Modelo MA (Medias Móviles)</i>	30
<i>Ecuación 3. Modelo ARMA (Autorregresivo de Medias Móviles)</i>	31
<i>Ecuación 4. Modelo ARIMA (Autorregresivo Integrado de Medias Móviles)</i>	31
<i>Ecuación 5. RMSE (Root Mean Square Error)</i>	37
<i>Ecuación 6. MAE (Mean Absolute Error)</i>	37
<i>Ecuación 7. Interpolación lineal</i>	43
<i>Ecuación 8. Interpolación para CERTE</i>	55

CAPÍTULO 1

Introducción

Actualmente la generación de energía eléctrica posee una gran dependencia de los combustibles de alto contenido en CO₂, en donde la mayor parte de la electricidad es generada utilizando gas, combustóleo y carbón. En estos últimos años, esta situación, ha despertado una notable preocupación ambiental a nivel mundial en la búsqueda de un futuro con baja intensidad de carbono. Por lo que, debido a estas consideraciones ambientales, se ha producido una iniciativa importante en el uso de las energías limpias para la generación de electricidad (Moselle, 2010).

Sin embargo, aún en el empleo de energías limpias se tienen que resolver algunos desafíos sobre el tema de generación de energía en la red eléctrica. Para el caso de la generación de energía eólica, existe un alto grado de intermitencia en el tiempo dado que son totalmente dependientes del clima (Foley, 2010). La intermitencia causada por la generación a partir de energías limpias, provoca problemas de control y estabilidad de la red eléctrica.

Otro de los grandes problemas es el almacenamiento de energía, ya que no se cuenta aún con un mecanismo eficaz para almacenar la energía generada para disponer de esta en momentos de gran demanda dentro de la red eléctrica (Posso, 2002).

La energía eólica ha tenido un crecimiento considerable en los últimos años a nivel mundial, las instalaciones de energía eólica en el mundo, crecieron un 12,4% en 2016, hasta situarse en 486.749 MW, según datos del Global Wind Energy Council (GWEC). China, Estados Unidos, Alemania, India y España son los primeros productores mundiales (Eólica, 2017).

El rápido crecimiento de la energía eólica ha llevado a la necesidad de emplear diversos métodos de pronóstico o predicción que den origen a un sistema de planificación y gestión en el mercado eléctrico para su equilibrio y estabilidad. El pronóstico preciso de la energía

eólica, reduce el riesgo de incertidumbre y permite una mejor integración en la red eléctrica (Foley, 2010).

La presente tesis tiene como intención demostrar el uso de técnicas auto regresivas Box and Jenkins para la predicción de generación a corto plazo de energía eléctrica a través de energía eólica, bajo el estudio y desarrollo de metodologías estadísticas sobre herramientas computacionales capaces de predecir la generación de energía eólica, que han sido utilizadas de manera satisfactoria en la predicción de generación del viento a lo largo del tiempo como se describe en (Jiménez, 2007).

1.1. Planteamiento del problema y justificación

La acción global frente al cambio climático es ineludible e impostergable. De acuerdo con la comunidad científica internacional, es necesario que todos los países reduzcan, de manera conjunta y decidida, las emisiones de gases y compuestos de efecto invernadero para evitar que las consecuencias se agraven. Para México, este desafío conlleva problemas sociales, económicos y ambientales que ya afectan a su población, infraestructura, sistemas productivos y ecosistemas, lo que impone a la necesidad de planear a corto plazo y actuar de inmediato para adaptarnos ante los impactos potencialmente adversos, así como reducir las emisiones de gases y compuestos de efecto invernadero. Debido a esta problemática, el 30 de mayo de 2013 se expide la Estrategia Nacional de Cambio Climático (ENCC) la cual plantea que el país crecerá de manera sostenible y promoverá el manejo sustentable y equitativo de sus recursos naturales, así como el uso de energías limpias que le permitan un desarrollo con bajas emisiones de gases y compuestos de efecto invernadero. Asimismo, se aspira a que México se vuelva un país próspero, competitivo, socialmente incluyente y con responsabilidad global que genere empleos suficientes y bien remunerados para toda su población, en particular para la más vulnerable. Propone una nación socialmente equitativa, con una economía verde, con ecosistemas y poblaciones resilientes al cambio climático y con ciudades sustentables (México, 2013).

En la actualidad, debido a la naturaleza intermitente del viento, es difícil predecir con exactitud la cantidad de producción total de generación de energía eléctrica a través energía eólica y tampoco se cuenta con los mecanismos eficaces para el almacenamiento de energía eléctrica a gran escala, lo que origina problemas de estabilidad en la red eléctrica, que se ve reflejada en grandes pérdidas económicas sobre el sector eléctrico.

En la búsqueda de una óptima solución a nuestro problema de conocer la cantidad de producción de generación de energía eléctrica con un día de anticipo mediante energía eólica, se plantea como alternativa un escenario de estudio de series de tiempo para regiones urbanas y convencionales desarrollando los métodos de predicción Box and Jenkins, con la intención de brindar una herramienta de apoyo a los tomadores de decisiones sobre las acciones adecuadas sobre la red eléctrica.

Para resolver esta problemática, actualmente en la literatura, existen diversas técnicas o métodos de predicción de Generación de Energía Eléctrica a partir de Datos eólicos, destacando principalmente: Métodos físicos, matemáticos, estadísticos, de inteligencia artificial, etc. No obstante, en esta tesis se tiene como finalidad demostrar que es posible la predicción de generación de energía eléctrica a partir de datos eólicos de zonas urbanas y convencionales, a través de los métodos estadísticos clásicos denominados Box & Jenkins (AR, ARMA, ARIMA) (Box, 1976).

1.2. Hipótesis

Se plantean dos hipótesis de prueba en este trabajo:

H1 Es posible realizar la predicción de generación de energía eléctrica con un día de anticipación a través de datos eólicos mediante métodos estadísticos clásicos Box-Jenkins.

H0 No es posible la realizar predicción de generación de energía eléctrica con un día de anticipación a través de datos eólicos mediante métodos estadísticos clásicos Box-Jenkins.

1.3. Objetivos del trabajo de investigación

1.3.1. Objetivo general

El objetivo de este trabajo es proponer una metodología de predicción de generación de energía eléctrica a partir de datos eólicos con un día de anticipo utilizando métodos Box and Jenkins y realizar un estudio comparativo de acuerdo a los resultados reportados en la literatura.

1.3.2. Objetivos específicos

Para llevar a cabo el objetivo propuesto se plantean dos escenarios de prueba:

- Datos eólicos urbanos provenientes de San Francisco.
- Datos eólicos convencionales provenientes de un parque denominado CERTE de la Ventosa en Oaxaca.

En ambos escenarios se llevarán a cabo los siguientes objetivos específicos:

- Realizar el ajuste de los datos a través de las particularidades de las series de tiempo.
- Desarrollar modelos de predicción de generación de energía eléctrica con un día de anticipo a partir de datos eólicos, empleando la metodología Box and Jenkins.
- Realizar una comparación de las metodologías utilizadas en la literatura contra los resultados de las pruebas desarrolladas de acuerdo a la propuesta.

1.4. Estructura de la tesis

En el capítulo 2, se presenta un breve resumen del trabajo relacionado con el potencial de producción eólica en base al origen de los vientos, se describe el proceso de Ciencia de datos y la metodología Box and Jenkins.

En el capítulo 3, se describe la metodología propuesta, así como la empleada para calcular y reducir el índice de error en la predicción.

En el capítulo 4, se lleva a cabo la experimentación de acuerdo a los datos obtenidos para las zonas convencionales y urbanizadas. Se llevan a cabo la recopilación de datos, el pre-procesamiento, la selección y entrenamiento del modelo para llevar a cabo los ejercicios de predicción con diversas pruebas hasta reducir el error.

Finalmente, se presentan las conclusiones y trabajos futuros de la investigación, y las referencias.

CAPÍTULO 2

Marco teórico

2.1. Antecedentes

La energía eólica, considerada como una alternativa potente y rentable para la generación de energía eléctrica y el cuidado del medio ambiente, ha tenido un crecimiento impulsado por el gobierno sumamente importante a nivel mundial del 35% para el 2017 y que tiene como objetivo futuro el uso de energía limpia, para el sector energético del 40% para 2035 y 50% para 2050 (GWEC, 2017).

Este tipo de generación de energía eléctrica, es una de las fuentes de mayor competitividad dentro de las energías limpias, pero posee una serie de desventajas, de acuerdo a (Guardado, 2007), descritas a continuación:

- La potencia del viento es considerada como una fuente de potencia intermitente debido a que el viento no tiene fluidez constante en el tiempo.
- Otro de los factores es que los sitios de los vientos superiores están generalmente ubicados en áreas distantes, por lo que se requiere la infraestructura adecuada para llevar la potencia generada por el viento a los centros de carga y actualmente no existen métodos confiables para esto.
- Gran parte de los generadores son de inducción; los cuales consumen potencia reactiva durante su operación, lo que puede provocar caídas de tensión en los sistemas de potencia.
- Los patrones de flujo de potencia y las características dinámicas del sistema deben cambiarse al integrarse parques eólicos a gran escala al sistema eléctrico.
- Deben realizarse estudios de flujo, tanto para condiciones normales de operación como ante situaciones de contingencia, para analizar la sobrecarga de potencial de las líneas de transmisión, así como revisar los niveles del cortocircuito, los cuales

deberán ser analizados antes y después de la introducción de parques eólicos en la red, con el objetivo de realizar mejoras en el mismo.

Cabe mencionar que los errores anteriores se relacionan directamente con la Calidad de la Energía Eléctrica (CEL) asociada a los parques eólicos (PE). Donde por lo general, existen cuatro variables que definen la CEL (insertar cita): Amplitud de señal, Frecuencia de la señal, Forma de la señal y Continuidad del servicio.

El conocimiento anticipado del comportamiento del viento es indispensable para múltiples aplicaciones como: Diseñar campos de generación de energía eléctrica, determinar el comportamiento de los ciclones tropicales, determinar la periodicidad del viento, determinar el sitio de evaluación del recurso eólico (Hernández-Escobedo, 2012).

Para predecir la potencia eólica, se debe estimar la producción de energía eléctrica de un aerogenerador en un tiempo futuro. El pronóstico se puede realizar en diferentes horizontes de tiempo. La predicción a corto plazo, puede llevarse a cabo desde milisegundos a pocos minutos y se emplea generalmente durante el control activo del aerogenerador. La predicción que se lleva a cabo en escala de horas hasta 3 días, son de mediano plazo y se utilizan para la gestión del sistema de energía y el comercio de la misma. Este tipo de predicción generalmente se utiliza para decidir el uso de tipo de las plantas convencionales y la optimización del despacho de estas.

Las predicciones que se realizan de 5 a 7 días son llamados de largo plazo y se emplean para la planeación, mantenimiento y toma de decisiones (Héctor, 2018).

2.2. Potencial de producción eólica en base al origen de los vientos

La producción de energía eólica depende del calentamiento dispar de la superficie terrestre por la acción de la radiación solar. En las zonas ecuatoriales se produce una gran absorción de radiación solar, en comparación con las zonas polares; el aire caliente que se eleva en los trópicos es reemplazado por masas de aire frío superficial que proviene de los

polos, donde el ciclo finaliza con el desplazamiento del aire en la alta atmósfera hacia los polos. Este fenómeno no podría ser apreciado visualmente, debido al movimiento rotatorio de la tierra, que influye en la generación de zonas de vientos dominantes que responden a patrones definidos.

Durante todo el año, las variaciones estacionales de la radiación solar incidente provocan variaciones en la intensidad y dirección de los vientos dominantes en cada uno de los puntos de la corteza terrestre. A su vez, el calentamiento desigual de las masas de aire y los fenómenos de características locales originan estructuras particulares de los vientos. Tal es el caso de las brisas de tierra y mar, motivadas por el calentamiento desigual de las masas de aire que durante el día se generan a lo largo de la costa vientos desde el mar hacia la tierra, revirtiendo el proceso en horas nocturnas. Así como también un fenómeno similar sucede en zonas montañosas donde las brisas de la montaña y de valle tienen su origen en el calentamiento del aire en contacto con las laderas, generando corrientes ascendentes durante las horas de sol y descendentes durante la noche (Moragues, 2003).

Cabe destacar que la velocidad del viento varía fundamentalmente de acuerdo a la altura y la naturaleza del terreno sobre el cual se desplazan las masas de aire. Por esto, para el uso de la energía eólica se debe tomar en cuenta velocidades medias, ráfagas, direcciones dominantes y algunos obstáculos eventuales para la selección de los lugares de emplazamiento como las características de construcción (altura de la torre, velocidades máximas soportadas, velocidad de marcha, etc.) de las máquinas a instalar (Moragues, 2003).

2.3. Generación Convencional

En México, diversas instituciones federales, han efectuado el reconocimiento de pequeñas redes anemométricas y exploraciones de regiones con vientos técnicamente aprovechables y económicamente viables. Lo cual ha servido para establecer un parque o central eólica en las siguientes regiones (González-Ávila, 2006):

- a) Sur del Istmo de Tehuantepec. La cual comprende un área 1000 km², capaz de asimilar una capacidad instalada de 2,000 a 3000 MW, con un factor de planta medio de 0.45.

Las zonas más propicias para generar electricidad son las aledañas a la Venta, Oaxaca, donde en 1994 se instaló una planta Eolo eléctrica con una capacidad de 1575KW, conformada por siete aerogeneradores de 225KW denominada La Venta (González-Ávila, 2006).



Figura 1. Distribución geográfica del recurso eólico potencial (González-Ávila, 2006)

Otra de las áreas con gran potencial eólico de ese mismo estado se denomina La Ventosa, que en los últimos años ha tenido un desarrollo económico y promoción gubernamental (González-Ávila, 2006).

- b) Península de Baja California. Se considera una zona con una barrera eólica natural perpendicular a los vientos occidentales. Sus principales áreas son las cercanas a los poblados de la rumorosa y zonas aledañas, el paso entre la Sierra de Juárez y la

Sierra de San Pedro Mártir. Así como el Cardón es una de las zonas más adecuada para el desarrollo de proyectos eólicos (González-Ávila, 2006).

- c) Península de Yucatán. Se considera un área de vientos alisios en primavera y verano, lo que origina que zonas como Cabo Cantoche, la costa de Quintana Roo y el oriente de Cozumel se consideren zonas de gran potencial eólico (González-Ávila, 2006).
- d) Altiplano Norte. Desde la región central del estado de Zacatecas hasta la frontera con Estados Unidos de Norte América (González-Ávila, 2006).
- e) Región Central. Prevalecen vientos alisios de verano, desde Tlaxcala a Guanajuato. La complejidad orográfica de esta zona genera como consecuencia que algunos pasos y mesetas sean adecuados para su explotación energética (González-Ávila, 2006) y se han obtenido como resultado pequeños proyectos eólicos e híbridos que electrifican comunidades rurales remotas (González-Ávila, 2006).

Dentro del litoral mexicano y sus islas, existen áreas con grandes posibilidades de producción de energía eólica a pequeña escala. Tal es el caso de Laguna Verde en Veracruz y la Isla del Carmen ubicada en Laguna de Términos, Campeche.

2.4. Generación Urbana

En la actualidad la instalación y uso de turbinas eólicas en edificios es reciente y son casi nulos los casos de edificios que consideren desde su diseño la integración de turbinas eólicas. Los colectores solares y las celdas fotovoltaicas son los sistemas de energías renovables más comunes integrados en los edificios debido a que son los más consolidados. Esto se debe a que los modelos de turbina con mejor eficiencia para usos urbanos hace algunos años aún eran prototipos y por otro lado que existen nuevas consideraciones estructurales y aerodinámicas que deben tomarse en cuenta por el diseñador del edificio (Arriaga, 2018).

Existe tecnología de microgeneración eólica, constituidas por turbinas eólicas de pequeña escala dentro de un entorno urbano. Un potencial importante de estas tecnologías de microgeneración es la reducción de las emisiones de CO₂ relacionadas con el entorno

construido junto con reducciones de costos de electricidad de los consumidores (Bahaj, 2007).

De acuerdo a estudios realizados por la Asociación Europea de Energía Eólica comenta que el Reino Unido posee el mejor recurso eólico de Europa (European Wind Energy Association., 2008). Pero, aun así, se debe estudiar bien la zona para conocer su potencial de energía contenida del viento (Arriaga, 2018).

Para un análisis eólico de zonas urbanas, se debe evaluar y determinar las zonas donde la influencia de las formas y dimensiones de las estructuras que producen aceleraciones de flujo de aire, no generen velocidades desfavorables.

Existen estudios que demuestran que existe producción de vientos importantes a nivel del suelo, debido a gradientes horizontales de presión generadas por los espacios entre edificaciones de cierta altura lo cual generan campos de velocidades desfavorables en áreas de circulación peatonal y en zonas abiertas que suelen comprometer la propia seguridad del peatón o el funcionamiento de los elementos dentro de esta. Para prevenir estos problemas, diversos investigadores han definido criterios de aceptación para los niveles de velocidad permisibles en áreas públicas (Marighetti, 2010) (Anssen, 2013) (Sparks, 2012).

Para apoyar el estudio de los efectos del movimiento del aire sobre o alrededor de objetos sólidos existen herramientas como el conocido Túnel del viento, que fue una contribución importante para la ciencia de la aerodinámica, desarrollada por los hermanos Wright, la cual proporciona modelos de simulación de las condiciones que experimentaría el objeto con datos valiosos a cierta escala para una situación real (Dodson, January, 2005).

2.5. Ciencia de Datos

Es considerada como un nuevo paradigma de la investigación que tiene por objetivo convertir los datos en ideas (Conway, 2010). La ciencia de datos, permite llevar a cabo procesos de Descubrimiento del Conocimiento (Knowledge Discovery) ya que estudia la

extracción del conocimiento a partir de los datos. En este nuevo paradigma, los investigadores suelen obtener ayuda de diversas herramientas para hacer frente a grandes cantidades de datos, así como también a una selección inmensa de ecuaciones, modelos, algoritmos de estimación de datos, evaluación y explicación de resultados complejos (Liu, 2017).

De acuerdo a (Hayashi, 1998) Data Science o Ciencia de Datos consta de tres fases: Diseño de datos, Recopilación de datos y Análisis en datos que deben tratarse con el concepto de unificación sobre la filosofía fundamental de la ciencia. En estas fases, los métodos que se ajustan al objeto y son válidos, deben estudiarse con una buena perspectiva.

Por otro lado, en la figura 1, de acuerdo a (Schutt, 2013), se describe el proceso de Ciencia de los datos en cada una de sus etapas.

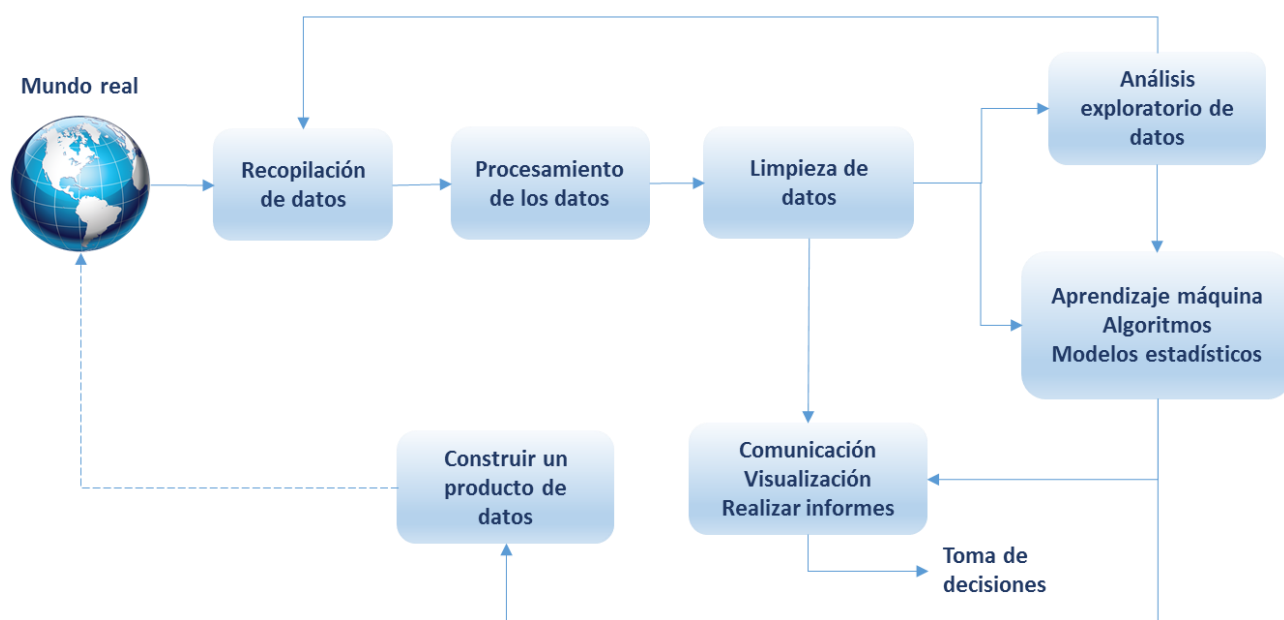


Figura 2. Proceso de Ciencia de Datos (Data Science Process) (Schutt, 2013)

De acuerdo a las fases propuestas por (Schutt, 2013), durante la primera fase se lleva a cabo la Recopilación de los datos, bajo el cual se identifica un problema del mundo real que permitirá la extracción de múltiples datos valiosos, por ejemplo, algunos provenientes del uso de páginas web, buscadores, concentrados de información, etc., de alguna fuente de

información donde exista la posibilidad de obtener grandes cantidades de datos sin procesar.

Antes de iniciar el análisis se debe realizar un pre procesamiento de los datos a través de una limpieza y formato uniforme para eliminar datos incompletos, vacíos, erróneos, etc., para extraer información valiosa. Y así poder revelar los datos con las menores inconsistencias posibles.

En las fases de Pre procesamiento y Limpieza de los datos, se procede al análisis y eliminación de datos no válidos, se completan ciclos y se utilizan durante los posteriores ejercicios.

En la fase del análisis exploratorio de los datos, se eligen los datos más efectivos que podrán ser utilizados con el menor margen de error posible. En este importante paso, se pueden aplicar métodos estadísticos, matemáticos, etc., que permitan elegir los de las mejores zonas.

En la fase de Aprendizaje máquina se suele utilizar el aprendizaje máquina en conjunto con algoritmos o modelos estadísticos para llegar a un resultado final.

2.6. Exploración de Metodologías de Predicción de Generación de Energía Eléctrica a partir de Datos Eólicos

De acuerdo a la literatura descrita en (Lei, 2009) (Box, 1976) (Lawan, 2014) donde nos mencionan que en la actualidad existen diversos modelos para realizar predicción o pronóstico de generación de energía eléctrica (w). Existen modelos físicos o matemáticos NWP (Numerical Weather Prediction o de Predicción Numérica del tiempo), estadísticos convencionales, de persistencia, modelos de correlación espacial, de inteligencia artificial (Vemuri, 1994) y nuevos modelos. Ver figura 3. A continuación se describen brevemente cada uno de ellos:

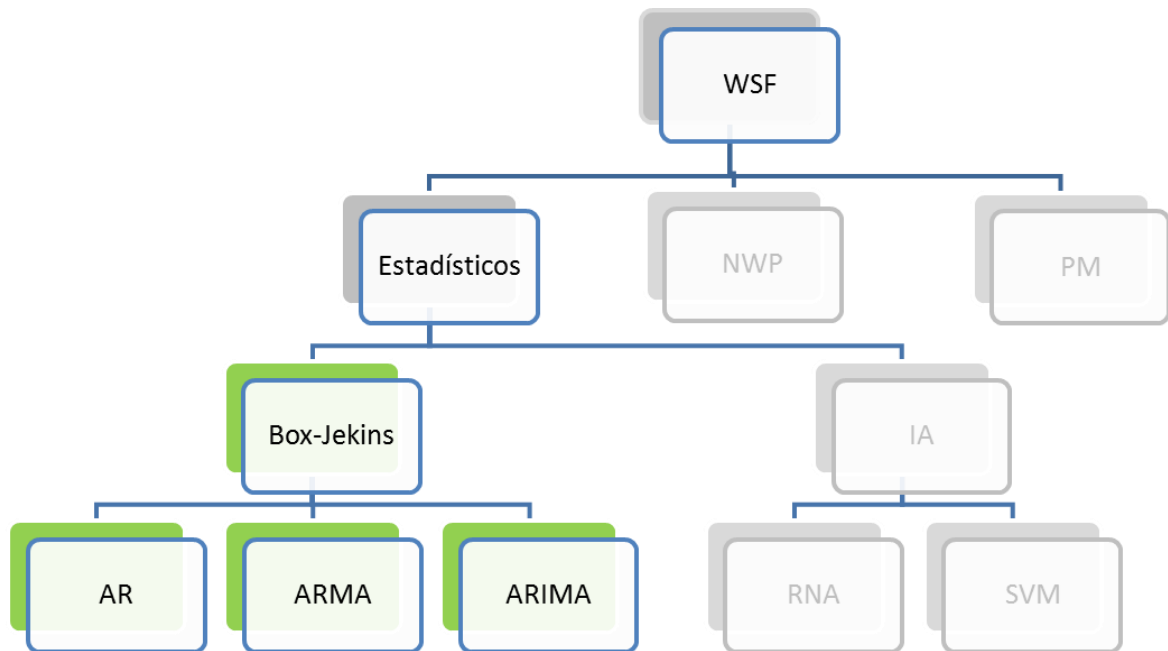


Figura 3. Métodos de predicción de Generación de Velocidad del Viento (Box, 1976) (Lei, 2009) (Lawan, 2014)

Para determinar el método o modelo de predicción más adecuado, como primer paso, lo más importante es elegir las variables de entrada apropiadas para construir un modelo de pronóstico o predicción eficiente, es decir; se necesitan diferentes variables para los diversos modelos. A su vez, otro factor que permitirá elegir un buen modelo, es conocer el horizonte de predicción de los datos, que ayude a determinar si el modelo a elegir será uno de corto, mediano o largo plazo (Box, 1976). Los diferentes modelos se describen a continuación:

El modelo de Persistencia (PM o *Persistence model*), asume que existe una correlación de alta resistencia entre los valores presentes y futuros de la velocidad del viento. Esta técnica utiliza ecuaciones lineales simples para realizar la predicción donde asume que el instante $t+x$ será el mismo que el momento t . Este enfoque es utilizado principalmente por los meteorólogos como una herramienta de comparación para completar la predicción en términos numéricos. Pero la debilidad de este método es que la precisión del modelo se reduce rápidamente al aumentar el tiempo de predicción (Lawan, 2014) (Persistence Model, 2018).

En un modelo físico, se utilizan consideraciones físicas como el terreno, los obstáculos, la presión, la temperatura, consideraciones meteorológicas, la descripción orográfica, la rugosidad, la humedad, entre otros obstáculos, etc. considerados como variables de entradas auxiliares de otros modelos estadísticos y suelen ser el primer paso para estimar o pronosticar la velocidad del viento en el futuro (Lei, 2009).

El modelo NWP (*Numerical Weather Prediction* o Predicción Numérica del tiempo) fue desarrollado por meteorólogos para la predicción del clima a gran escala, sin embargo, este modelo no genera resultados precisos en la predicción a corto plazo. Con este modelo, se resuelven numéricamente algunas ecuaciones de conservación del sitio específico donde al final, con algún modelo de salida estadístico (MOS) se reducirá el error restante. Este método se utiliza principalmente con datos meteorológicos actuales como variables de alimentación de complejos modelos físico - matemáticos para predecir la evolución meteorológica.

Por consiguiente, el método físico sirve como un análisis preliminar que hace que el resultado del método estadístico sea más eficiente y aunque no siempre es el caso, las velocidades históricas del viento del mismo sitio, como de los vecinos, son necesarias para el modelo de correlación espacial. También, la curva de potencia de los aerogeneradores, es necesaria para efectuar la predicción de la energía eólica o velocidad del viento (Lei, 2009).

Con el desarrollo de la inteligencia artificial y otros métodos de pronóstico o predicción, varios de los modelos para predecir la velocidad del viento y la predicción de energía están proliferando. Algunos de los métodos desarrollados incluyen Redes Neuronales Artificiales (ANN o Artificial Neuronal Network), métodos de lógica difusa (Fuzzy Logic), Máquinas de Soporte Vectorial (Support Vector Machine) y algunos métodos híbridos (Lei, 2009).

De acuerdo a este mismo autor las Redes Neuronales Artificiales (ANN) son uno de las técnicas más utilizadas en las últimas décadas (Lei, 2009). Este método consta del desarrollo de varias capas, una de entrada, otra de salida y algunas más ocultas. En cada capa, existen diversas neuronas conectadas a la capa anterior pero independientes una de la otra. Cada conexión tiene su propio peso y cada neurona tiene su función de

transferencia. El proceso de entrenamiento es el procedimiento para obtener los pesos de cada conexión y el umbral de valor de las neuronas. Existen desarrollados algunos algoritmos de entrenamiento y ellos incluyen un algoritmo conocido como Algoritmo de propagación (BP o Back-propagation) o el algoritmo Levenberg Marquardt (LM) que tiene por objetivo reducir el valor mínimo de error en la Red.

Los modelos ANN son poderosos sobre datos no lineales. Una ANN aprende de ejemplos de muestra dados, construyendo un mapeo de entrada-salida para realizar predicciones de muestras futuras. Esta técnica es también una de las más adecuadas para aplicaciones de predicción de velocidad/energía del viento, ya que consta de muchas unidades de procesamiento simples idénticas interconectadas. Las técnicas consumen menos tiempo en comparación con otros métodos convencionales (Lawan, 2014).

Otro de los modelos son los de lógica difusa (Fuzzy logic) que utiliza cantidades de valores en el intervalo $[0,1]$ y variables difusas como largo, medio y corto para explicar los valores. Este modelo, generalmente se utiliza en sistemas difíciles de modelar (Lawan, 2014).

También el modelo de Máquinas de Soporte Vectorial (Support Vector Machine) es un enfoque novedoso que puede superar algunas de las desventajas de las Redes Neuronales como el punto mínimo local, o sobre el aprendizaje (Lawan, 2014).

Los modelos Estadísticos Convencionales (Conventional statistical models) son idénticos a los Modelos Aleatorios de Series de Tiempo (Lei, 2009). Se basan en una serie de datos históricos, utilizan la identificación de patrones, estimación de parámetros y para la verificación del modelo, son utilizados los métodos matemáticos. De acuerdo con los métodos propuestos por Box and Jenkins, estos modelos se dividen de la siguiente manera: Modelos Autoregresivos (AR), Modelos de Medias Móviles (MA), Modelo Autoregresivo de Medias Móviles (ARMA) y Modelo Autoregresivo Integrado de Medias Móviles (Guoyang, 2005).

Existen también los métodos híbridos que han sido ampliamente utilizados por diversos autores para predicción de velocidad del viento en bases de datos históricos. Este método implica la combinación de técnicas físicas y estadísticas o la combinación de diferentes

modelos en diferentes horizontes o la combinación de modelos estadísticos alternativos. Sus principales objetivos del modelo híbrido son la capacidad de probar la función de rendimiento del modelo basados en resultados observados y simulados entre los dos modelos, a diferencia de los enfoques estadísticos. Cabe señalar que el rendimiento del modelo se puede juzgar utilizando diferentes medidas de ajuste de bondad como: error cuadrático medio (RMSE), error absoluto medio (MAE), error de prejuicio medio (MBA) y el error de porcentaje absoluto medio (MAPE) (Lawan, 2014).

2.6.1. Metodología de modelos autoregresivos

Box and Jenkins

De los métodos estadísticos convencionales utilizados para identificar, estimar y diagnosticar modelos dinámicos para series temporales univariantes y estacionarias en los que la variable tiempo juega un papel fundamental son los llamados Box and Jenkins, en honor a sus creadores George E. P. Box, profesor de Estadística de la Universidad de Wisconsin and Gwilym M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster en 1970 (Box, 1976).

De acuerdo a (Box, 1976), este modelo originalmente propone tres etapas principales:

1. **Identificación y selección del modelo.** En esta fase se debe identificar la estacionalidad (Villavicencio, 2010), es decir; si la serie de tiempo presenta los mismos patrones de repetición en un periodo de tiempo (mensual, semestral, trimestral, etc.) y asegurar la estacionariedad de las variables de la serie de tiempo. Para verificar la estacionariedad (La media o varianza de la Serie de Tiempo permanecen constantes en el tiempo. Es decir; cuando la Serie de Tiempo no presenta crecimiento, declinación o tendencia, se considera estacionaria) (Gras, 2001). En esta fase se determina cual componente se debe usar en el modelo, el promedio Autoregresivo (AR) o un promedio de Medias Móviles (MA).
2. **Estimación de parámetros del modelo.** En esta fase se determina los coeficientes que de mejor ajuste para el modelo.

3. **Comprobación del modelo.** Se realizan comprobaciones mediante el ensayo y se verifica respecto a los residuos que son los resultados después del ajuste del modelo, para muchos modelos de Series de tiempo, los residuos son equivalentes a la diferencia entre las observaciones y los valores ajustados correspondientes. Los residuos son útiles para verificar si el modelo es el adecuado de acuerdo a los datos y pronóstico sobre la serie de tiempo (Hyndman R. J., 2018)
- Para que esto sea exitoso, los residuos deben ser independientes y la media y la varianza deben ser constantes en el tiempo. Si la estimación no es la mejor, se deben realizar ajustes al modelo, hasta encontrar el mejor.

Los Modelos Box and Jenkins (Box, 1976) clásicos son:

Modelo Autoregresivo (AR o Auto Regressive Model). Son un conjunto de observaciones en el tiempo, donde el valor de la variable en un instante t es explicado por un conjunto de valores de ella misma X_t añadiéndose un término de error. Los modelos autoregresivos se expresan por AR(p). Donde p = es el orden del modelo, que expresa el número de observaciones retrasadas en la serie temporal analizada. De acuerdo a (Pandit, 1983) y (Hyndman R. J., 2018) su ecuación es:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_q X_{t-q} + \varepsilon_t$$

Ecuación 1. Modelo AR(Autorregresivo)

X_t = Valores de variables actuales de correlación o dependientes de la serie de tiempo

$X_{t-1}, X_{t-2}, \dots, X_{t-q}$ = Valores pasados o anteriores

$\Phi_1, \Phi_2, \dots, \Phi_q$ = Constante definida por un tiempo

ε_t = Residuo desconocido o perturbaciones aleatorias en el periodo t que no pueden ser explicadas por el modelo (Ruido blanco)

Modelo de Medias Móviles (MA o Moving Average).

Explica el valor de una determinada variable en un periodo t en función de un término independiente y una sucesión de errores correspondientes a valores precedentes, ponderados convenientemente.

En lugar de utilizar los valores actuales de la variable de pronóstico en una regresión, utiliza los errores de pronóstico pasados en un modelo, similar a los de regresión. Donde ε_t es ruido blanco, que es una variable aleatoria con media cero, varianza constante, que no muestran autocorrelación entre los valores pasados de la serie (Parra, 2007). Este modelo se conoce como MA(q), el cual puede ser utilizado para pronosticar valores futuros (Hyndman R. J., 2014).

$$X_t = C + \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2} + \dots + \Phi \varepsilon_{t-q} + \varepsilon_t$$

Ecuación 2. Modelo MA (Medias Móviles)

C = Constante definida por un tiempo

$\Phi_1, \Phi_2, \dots, \Phi_n$ = Constante definida por un tiempo o valores de correlación de la serie de tiempo

$\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \dots, \varepsilon_{t-q}$ = Valor pasado o valores anteriores del error

ε_t = Residuo desconocido o error en el periodo t

Donde cada valor de X_t se puede considerar como un promedio ponderado de los últimos errores de pronóstico. No se debe confundir los modelos de promedio móvil con el de suavizado de promedio móvil. Un modelo de promedio móvil se utiliza para pronosticar valores futuros mientras que el suavizado de promedio móvil se utiliza para estimar el ciclo de tendencia de los valores pasados.

Modelo Autorregresivo de Medias Móviles (ARMA o Auto Regresive Moving Average).

Este modelo involucra un término autorregresivo y uno de medias móviles, es decir; es una combinación entre AR(p) y MA(q).

De acuerdo a (Pandit, 1983) su ecuación se define por:

$$X_t = \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Ecuación 3. Modelo ARMA (Autorregresivo de Medias Móviles)

$\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_n$ = Constante definida por un tiempo o valores de correlación de la serie de tiempo.

$X_{t-1}, X_{t-2}, \dots, X_{t-q}$ = Valores u observaciones de la Serie de tiempo

$\theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ = Residuos pasados de orden $\varepsilon_{t-1} \dots$ (Residuo que se puede conocer)

ε_t = Residuo desconocido o ruido blanco

El Modelo Box and Jenkins avanzado más común es:

Modelo Autorregresivo Integrado de Medias Móviles (ARIMA o Auto Regressive Integrated Moving Average).

Los modelos ARIMA(p,d,q) proporcionan otro enfoque en la predicción o pronóstico de series temporales.

Si se combina un modelo Autorregresivo AR(p) con diferenciación y el modelo de Media móvil MA(q) se obtiene un modelo ARIMA no estacional. ARIMA (AutoRegressive Integrated Moving Average) donde la “integración” en este contexto es lo contrario a la diferenciación.

El modelo ARIMA(p,d,q) se puede describir mediante la siguiente ecuación (Hyndman R. J., 2014):

$$Y'_t = C + \underbrace{\Phi_1 Y'_{t-1} + \dots + \Phi_p Y'_{t-p}}_{\text{AR(P)}} + \underbrace{\theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}}_{\text{MA}} + \varepsilon_t$$

Ecuación 4. Modelo ARIMA (Autorregresivo Integrado de Medias Móviles)

Donde,

Y'_t = Es la serie diferenciada

$Y'_{t-1} + \dots + Y'_{t-p}$ = Valores pasados de las observaciones de la serie de tiempo

$\mathcal{E}_{t-1} + \dots + \mathcal{E}_{t-q}$ = Valores pasados del error

\mathcal{E}_t = Residuo desconocido o ruido blanco

Los parámetros de la serie diferenciada (la cual puede ser diferenciada en diversas ocasiones) en ARIMA(p,d,q) son:

Donde,

p = Orden de la parte autorregresiva (términos autorregresivos)

d = Grado de diferenciación involucrada

q = Orden de media móvil (moving average)

En este modelo se aplican las mismas condiciones de estacionariedad e invertibilidad que se utilizan para los modelos autorregresivos y de media móvil o promedio móvil.

Para desarrollar un modelo ARIMA(p,d,q) de acuerdo a Box and Jenkins se sugiere un mínimo de 50 observaciones, así como también, son útiles para series que presentan patrones estacionales. Pero los métodos Box and Jenkins pueden aplicarse tanto a series estacionarias (es aquella cuya media, varianza y función de autocorrelación permanecen constantes en el tiempo) y no estacionarias (Pankratz, 2012).

Algunos otros autores como (Aguirre, 1994) describen más a detalle proceso de la metodología de los modelos ARIMA (Box, 1976) se resume en las siguientes etapas principalmente:

1. Determinación de la estacionariedad (Engineering Statistics Handbook, 2019) en la serie de tiempo. En esta etapa se debe determinar si la Serie de tiempo es estacionaria, en caso contrario, se debe inducir a la estacionariedad con la finalidad de estabilizar la serie. El proceso estacionario, está descrito por una secuencia de

datos o valores que no presentan cambios en la media, ni en la varianza; es decir, se mantiene estable.

2. Identificación del modelo. Una vez que se asegura la estabilidad en la serie de tiempo, se procede a identificar qué modelo rige el proceso de la serie de tiempo
3. Estimación de los parámetros del modelo. En esta fase se estiman los coeficientes del modelo escogido tentativamente en la etapa anterior. Existen diversos criterios para el cálculo de estos. Uno de los principales criterios para el cálculo de los parámetros es Estimación de la máxima verosimilitud (Maximum Likelihood o ML) que es un método muy preciso, pero el cálculo es demasiado robusto, lo cual podría demorar bastante tiempo, aún con el uso de computadora (Pankratz, 2012). Debido a esta razón (Pankratz, 2012) (Aguirre, 1994) sugieren el uso del criterio de mínimos cuadrados (least squares o LS)
4. Evaluación del modelo. En esta etapa se verifica si el modelo estadísticamente es el más adecuado (Se debe verificar que los residuales son independientes entre sí, es decir; completamente aleatorios). En caso de no serlo, se repite el proceso desde la fase 2. Identificación del modelo.
5. Predicción. En esta fase se predicen los valores futuros de la serie de tiempo y verificar si el modelo del pronóstico es estadísticamente adecuado. En caso de no ser adecuado, repetir a partir de la fase 2. Identificación del modelo.

Los modelos ARIMA son utilizados principalmente en series de tiempo no estacionarias (Box, 1976), muestran otro enfoque para el pronóstico de las series de tiempo y pretenden describir la autocorrelación de los datos sobre la misma (Hyndman R. J., 2018).

CAPÍTULO 3

Metodología propuesta

De acuerdo a los hallazgos en (Box, 1976) (Hyndman R. J., 2014) (Brockwell, 1991), se desarrollaron diversos ejercicios con la metodología Box and Jenkins a través de las fases descritas en el proceso de Ciencia de Datos, aplicados a la Predicción de Generación de Potencia de los datos de velocidad del viento recopilados de las torres entre rangos de 40 y 80 metros de altura de una región urbana y otra convencional, de acuerdo a (Pineda, 2017):

- **Zona urbana con 246,612 registros y 5 de las mejores torres seleccionadas de San Francisco**, a una altura de 40 mts. se eligieron las mejores torres de acuerdo a su mejor media y desviación estándar. Así como también las que tuvieran mayor completos en 4 periodos equivalentes a 4 años (cada año con sus 12 meses) de datos completos, con un horizonte de tiempo de medición cada 5 min (Pineda S. E.-B., 2017).
- **Zona convencional del CERTE de México con 157,824 registros a 40, 60 y 80 metros de altura**. Se concentraron archivos de datos de 3 periodos completos equivalentes a 3 años (cada año con sus 12 meses) de datos completos en un formato abierto .csv con un horizonte de tiempo de medidas de velocidad de cada 10 min (Pineda, 2017).

Una vez realizado el proceso de recopilación de datos, se dio paso a la fase de Pre Procesamiento de los datos (Hayashi, 1998):

De acuerdo a la figura 2, la primera etapa que se llevó a cabo es la limpieza y el análisis exploratorio de los datos.

Posteriormente, se procedió a desarrollar el proceso de interpolación de los datos tanto para San Francisco como para el CERTE.

En la segunda etapa de Entrenamiento/Aprendizaje del modelo se llevaron a cabo diversos ejercicios con los datos de las torres seleccionadas.

Los pasos que comprenden la metodología son:

- 1. Recopilación de datos.** Se recopilaron dos tipos de datos, uno para las torres de San Francisco (Datos urbanos) y otros obtenidos del CERTE de la región de la Ventosa II (Datos convencionales).
- 2. Pre procesamiento de los datos**
 - a) Limpieza de los datos.** Una vez obtenida la colección de datos, se procede a revisar en base a los periodos, si son completos, si no existen datos vacíos, errores en algunos de ellos, etc., para posteriormente evaluar si se procede a eliminarlos o se completan en base a alguna metodología. Limpieza de los datos (Excel SPSS -> Datos incompletos). Se realizó la recopilación de la información en un archivo en formato .CSV.
 - b) Normalización de datos.** Posterior a la limpieza, se deben normalizar los datos, completando los faltantes. Para este caso, se completaron los datos en base al modelo persistente, mediante el cual se elegía el dato anterior para simularlo.
 - c) Análisis exploratorio (Mejores medias, rosas del viento, mejores zonas para las 5 mejores torres de San Francisco).** En ese paso, se realizan; el cálculo de las mejores medias de San Francisco y CERTE. Para esto, se desarrolló una función basada en el lenguaje de programación de R (Ver Apéndice B) que permite la modelación de las rosas del viento (Es una visualización circular donde se observa la distribución de la velocidad y la dirección del viento de un lugar determinado durante un periodo determinado) (Departament of Agriculture, 2019). Para poder observar el origen del mejor viento en ambos casos, se utilizaron las librerías ggplot, CircStats, MASS y boot de RStudio Versión 1.0.143, la finalidad de este paso es analizar el origen del viento y sus momentos estadísticos (Medida cuantitativa específica de la forma de una función) (Kenney, 2013) como su media, desviación

estándar, curtosis, mediana, etc. (Código de consulta en Apéndice B) y así poder determinar que torre es la más completa para su posterior estudio.

- 3. Selección de los parámetros y entrenamiento del modelo.** Se desarrolló una función en lenguaje de programación R (Ver Apéndice C), en el cual sobre el 50% de los datos de Entrenamiento del modelo, se entrenan los modelos de acuerdo a las posibles combinaciones ARIMA $p = \{0, \dots, 5\}$; $d = \{0, 1\}$; $q = \{0, \dots, 5\}$. De esos modelos se calculan los de menor RMSE y MAE. Para pasar a realizar la comprobación sobre el otro 50% de los datos de validación.

Sobre el otro 50% de los datos de validación se procede a realizar la comprobación del modelo y se realiza la primera predicción sobre los mejores modelos, eligiendo su menor RMSE y MAE.

- 4. Pronóstico con comprobación (Cálculo de RMSE y MAE).** Una vez realizado el entrenamiento, para el desarrollo de la predicción, se desarrolló una función que realizaba los ajustes correspondientes sobre la serie de tiempo y su predicción (Ver Apéndice D), posteriormente se eligen el de menor RMSE y MAE y se procede a realizar con los 5 mejores modelos resultantes la operación sobre los datos de prueba final (20% del total de los datos) y se calcula su RMSE y MAE como se describe en la siguiente sección. En este paso es posible determinar el mejor modelo de todos sobre la serie de tiempo.

3.1. RMSE y MAE

Tanto el Root Mean Square Error (RMSE) (Raíz del error cuadrático medio) o el Mean Absolute Error (MAE) (Error absoluto medio) se utilizan regularmente en estudios de evaluación de algún modelo (Chai, 2014).

La raíz del error cuadrático medio (RMSE) se ha utilizado como una métrica estadística estándar para medir el rendimiento del modelo en meteorología, calidad del aire y estudios de investigación del clima. RMSE es una regla de puntuación cuadrática que también mide

la magnitud promedio del error. Es la raíz cuadrada del promedio de las diferencias al cuadrado entre la predicción y la observación real.

Por otro lado, el error absoluto medio (MAE) es otra medida útil ampliamente utilizada en evaluaciones de diversos modelos y es comúnmente empleada para medir el error en los pronósticos (Gonzalo, 2008).

Cuando se calculan ambas métricas, MAE tiende a ser mucho más pequeño que el RMSE, porque el RMSE penaliza grandes errores, mientras que MAE les da el mismo peso a todos los errores (Chai, 2014). La ecuación para RMSE es la siguiente (Gonzalo, 2008):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\theta_i^s - \theta_i^a)^2}$$

Ecuación 5. RMSE (Root Mean Square Error)

Donde

θ_i^s = Valor real de θ_i

θ_i^a = Valor pronosticado de θ_i

n = Número de periodos

Mediante el RMSE se cuantifica la cantidad de error que hay en un conjunto de valores pronosticados y valores observados. Con el RMSE se puede medir la dispersión de la variable simulada en el curso del tiempo (Gonzalo, 2008).

Cuando un valor RMSE es menor, la aproximación es mejor entre los valores que se predicen y los observados (Chai, 2014).

La ecuación para MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\theta_i^s - \theta_i^a|$$

Ecuación 6. MAE (Mean Absolute Error)

θ_i^s = Valor real de θ_i

θ_i^a = Valor pronosticado de θ_i

n = Número de periodos

El MAE es adecuado para describir la distribución de los errores, mide la magnitud promedio de los errores en un conjunto de pronósticos, sin considerar su dirección. Mide la precisión de las variables continuas. Es decir; Es el promedio sobre la sobre la verificación simple de valores absolutos de las diferencias entre el pronóstico y las observaciones correspondientes (EUMETRAIN, 2019).

El MSE en la predicción, para cada punto, calcula la diferencia cuadrada entre las predicciones y el objetivo para posteriormente promediarlas. Entre más alto sea el valor, el modelo es peor (Sitio BigData, 2019).

Tanto el RMSE como el MAE expresan error promedio de predicción en el modelo en unidades de la variable de interés. Son puntuaciones orientadas negativamente, lo que significa que los valores más bajos son mejores.

CAPÍTULO 4

Experimentación

4.1. Recopilación de datos urbanos de San Francisco

Se obtuvieron datos de generación del viento de 26 torres en una página que contiene un diccionario abierto de datos del Gobierno de San Francisco (sfgov.org, 2019), los cuales son medidos por estaciones meteorológicas instaladas por la SFPUC (Comisión de Servicios Públicos de San Francisco) que tienen monitoreo del viento en San Francisco y en tierras propiedad de SFPUC. Los parámetros registrados incluyen: velocidad del viento, dirección del viento, temperatura ambiente e irradiancia solar (a través de esta se conoce cuánta energía solar incide sobre un área en un tiempo determinado) (sfgov.org, 2019) (Hernández Mora, 2014), de las cuales se eligió la torre de mejor media para el desarrollo del ejercicio de predicción. Ver figura 4.

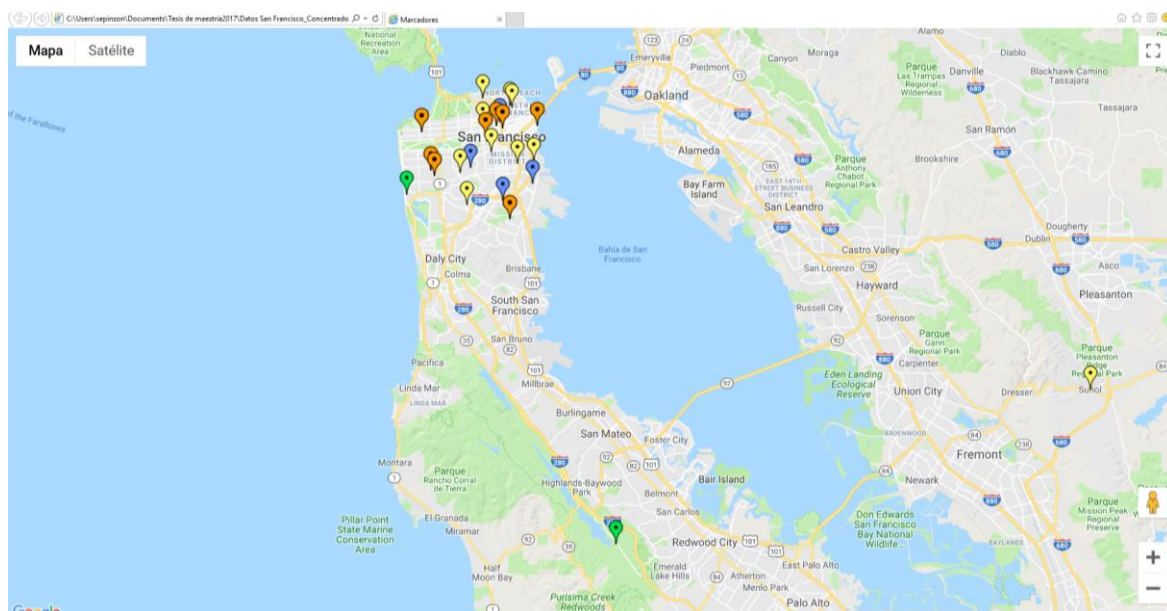


Figura 4. Ubicación de las 26 torres de la SFPUC (Comisión de Servicios Públicos de San Francisco)

4.2. Pre procesamiento de los datos de San Francisco

Posteriormente, se llevó a cabo la limpieza, normalización, el análisis exploratorio de los datos y la interpolación de los datos para poder realizar las predicciones sobre los datos de San Francisco, así como se describe a continuación.

Limpieza y normalización sobre los datos de San Francisco

Para realizar la normalización de los datos se procedió a evaluar en base a sus momentos estadísticos (media y desviación estándar) para posteriormente seleccionar las torres que tuvieran mejor media e información de al menos 5 periodos (Cada periodo es equivalente a un año) completos. En esta fase, también se procedió a generar la información faltante en base a la teoría de los Modelos persistentes (Lamoureux, 1990) y así dar paso a la normalización de los valores en la Serie de Tiempo.

Tabla 1. Datos de San Francisco (sfgov.org, 2019)

Estació	Nombre de la estación	Dirección	2008	2009	2010	2011	2012	2013	2014	2015	2016	NA	No NA	Total	Media	sd
SF34	Twin Peaks	501 Twin Peaks Blvd										363	246249	246612	5.480218	3.590461
SF33	Tesla	Tesla										0	62092	62092	5.460152	3.391418
SF11	Hillcrest Elementary School	810 Silver										0	496003	496003	4.081996	2.728818
SF10	Southeast Water Treatment Plant	750 Phelps										1	36263	36264	3.217029	1.502994
SF05	Ida B. Wells High School	1099 Hayes St										0	42061	42061	2.996863	2.044959
SF06	525 Golden Gate	525 Golden Gate										0	452	452	2.971575	1.734463
SF01	Washington High School	600 32nd										0	195071	195071	2.737852	1.560446
SF32	SF Judiciary Bldg	400 McAllister St										9	369851	369860	2.684898	1.619587
SF18	Pier 40	Pier 40										0	790778	790778	2.342873	1.674132
SF35	Sunset Reservoir	Sunset Reservoir										0	26245	26245	2.262233	2.267334
SF04	Lincoln High School	2162 24th Ave										946	286953	287899	2.261137	1.267364
SF14	Visitation Valley Middle School	450 Raymond										0	590753	590753	2.140347	2.39633
SF02	Marina Middle School	3500 Fillmore										0	304188	304188	1.904749	1.428511
SF23	Chinatown Health Center	1410 Mason										0	264313	264313	1.877827	1.37191
SF08	Everett Middle School	450 Church										0	237119	237119	1.834582	1.610884
SF15	Warmerville Switchyard	Oakdale Warmerville										0	801493	801493	1.760859	1.320953
SF19	MUNI Building - 1080 Minnesota	1080 Minnesota										0	497995	497995	1.623716	1.659392
SF22	City College Science Bldg	50 Phelan Ave										0	189378	189378	1.586147	0.9846948
SF24	Chinatown Library	1135 Powell										0	324146	324146	1.562127	1.089545
SF09	SF General Hospital	10001 Portrero										6	239648	239654	1.465151	1.268488
SF20	Moccasin Peaks	Moccasin Peaks										0	469058	469058	1.446537	1.381293
SF07	Laguna Honda Hospital	375 Laguna Honda										0	596199	596199	1.28215	1.062648
SF21	Maxine Hall Health Center	1301 Pierce										0	289999	289999	1.19551	1.035812
SF12	Sunol	Sunol										0	506726	506726	1.094118	0.924393
SF16	SF Zoo	1 Zoo Rd										0	488711	488711	0.5946513	1.240419
SF17	Pulgas Reservoir, San Mateo	Canada Road										0	795186	795186	0.5862563	0.8693864

Con la información de la tabla anterior, se determinó que la estación denominada Twin Peaks con ubicación en 501 Twin Peaks Blvd en San Francisco es la que tiene los 5 periodos más completos (cada periodo de datos comprende un año a partir del 2010, 2011, 2012, 2013 y 2014) y presenta la mejor media de 5.48 m/s así como los siguientes momentos estadísticos:

Tabla 2. Momentos estadísticos de 501 Twin Peaks Blvd. San Francisco – Velocidad del viento

San Francisco m/s	media	mediana	sd	kurtosis	sesgo	min	max
	5.48	5.055	3.593	-0.101	0.436	0	58.41

La serie de tiempo de esta estación cuenta con 246,612 registros y se localiza en el centro del siguiente mapa con una leyenda sobresaliendo de esta:

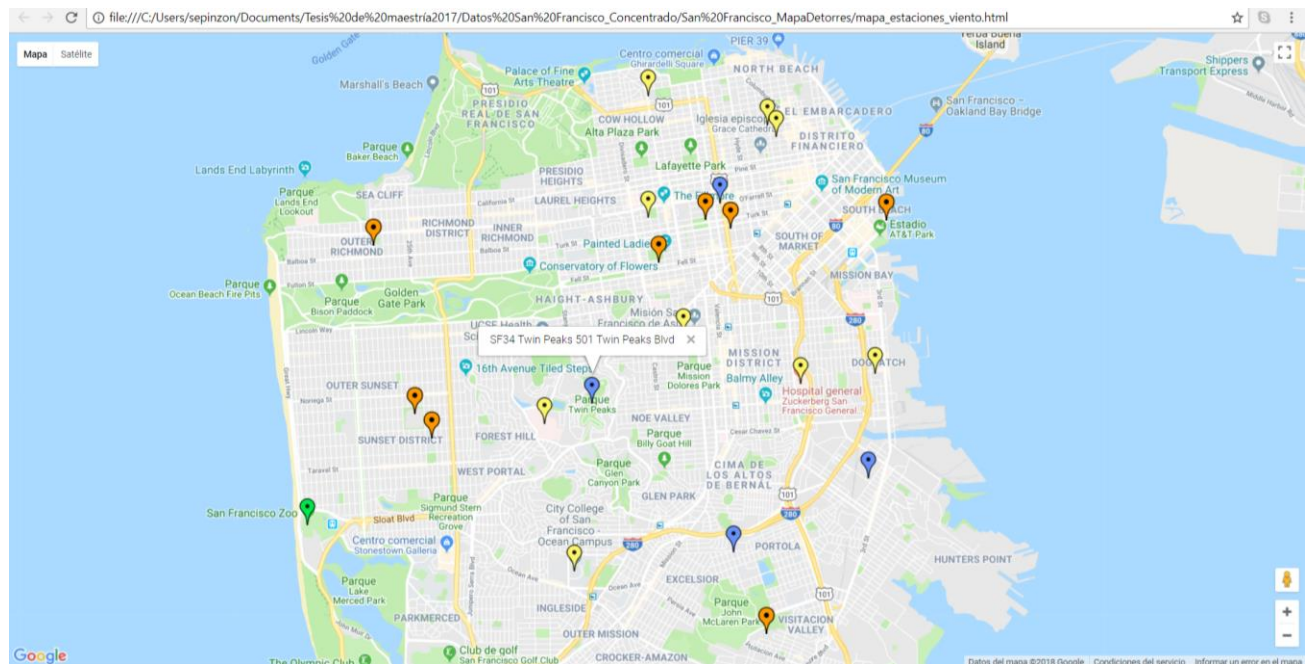


Figura 5. Mapeo de torres de San Francisco con base en su media (sfgov.org, 2019)

Para llevar a cabo los ejercicios de predicción, se procedió a interpolar los datos de velocidad del viento (v) de la mejor torre a potencia (W) sobre un catálogo de fabricantes europeos de Turbinas eólicas. Procedimiento descrito a detalle en la sección 4.2.3 Interpolación de las Series de Tiempo.

Análisis exploratorio de los datos

En la etapa de análisis exploratorio de los datos, se desarrollaron las rosas del viento para las mejores torres, mostrando a continuación la que fue seleccionada para el ejercicio.

Para el análisis de datos de San Francisco, la torre urbana más prometedora en base a su mejor media y completos en sus periodos (5 periodos completos) fue Tween Peaks de la Ciudad y Condado de San Francisco California. A continuación, en la figura 6 se muestran gráficamente las direcciones del viento:

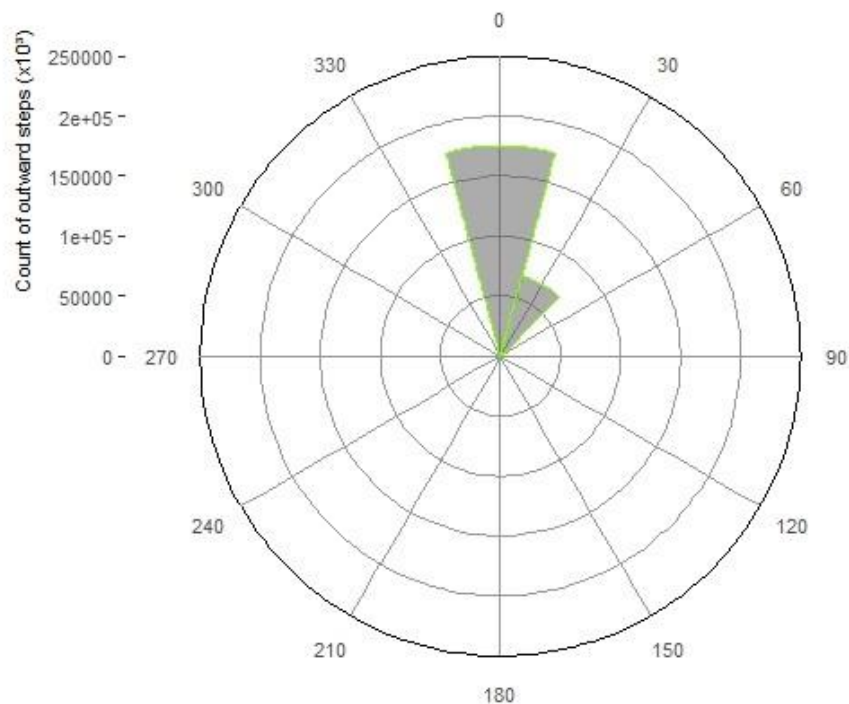


Figura 6. Rosa del viento con datos de Tween Peaks (Región urbana) – Cd. Y Condado de San Francisco, California

Para el cálculo de la rosa de los vientos en R (lenguaje de programación basado en librerías y paquetes), se utilizaron algunas funciones de las librerías: ggplot2, MASS, boot y CircStats y se desarrolló un código que permite calcular y graficar los valores detectados en la serie de tiempo, tal como se muestra en el Apéndice B:

Del análisis, se puede apreciar que el origen del viento favorable y significativo para esta región es proveniente del Noreste.

Interpolación de la serie de tiempo con los datos de San Francisco

Para transformar a potencia (W) los valores de la velocidad del viento (m/s), se llevó a cabo la interpolación utilizando un catálogo de fabricantes europeo de turbinas eólicas (programme, European Commission under the Intelligent Energy - Europe, 2017). De las turbinas eólicas que generan mayor cantidad de potencia, se eligió la torre: OY Windside Production Ltd - VAWT from 1 kW to 8 kW para desarrollar los ejercicios (cita).

Para la interpolación lineal se utilizó la siguiente ecuación (Friedman, 1962):

$$y = y_a + (x - x_a) \frac{(y_b - y_a)}{(x_b - x_a)}$$

Ecuación 7. Interpolación lineal

Donde La interpolación lineal consiste en trazar una recta que pasa por (X_a, Y_a) y (X_b, Y_b) , $y = r(x)$ y calcular los valores intermedios según esta recta en lugar de la función $y = f(x)$:

$$y_a = f(X_a) ; y_b = f(X_b) ; y = f(X) \text{ Para: } (X_a < X < X_b)$$

Sustituyendo los valores sobre la serie de tiempo de San Francisco, los valores sustituyen de la siguiente forma:

$$y_a = W_{\text{Inferior}}$$

$$y_b = \text{Velocidad}$$

$$x = W_{\text{Superior}}$$

$Xa = \text{LimInferior}$

La curva de potencia de la turbina seleccionada, así como sus especificaciones técnicas, se muestran en la siguiente figura:

Technical information

POWER		Unit
1) Rated power	8	kW
2) Rated wind speed	20	m/s
3) Cut-in wind speed	2	m/s
4) Cut-out wind speed	None	m/s
5) Maximum wind speed the turbine can withstand	216	Km/h
DIMENSIONS		
6) Rotor weight	3000	kg
7) Rotor diameter	2	m
8) Rotor height (for VAWT only)	6	m
9) Swept area	12	m ²
10) Height of the mast	Not relevant	m
OTHER INFORMATION		
11) Maximum rpm	100 - 300	At rated wind speed
12) Gear box type	No gear box	
13) Brake system	Disk brake	
14) Number of blades	2	
15) Blades material	Aluminium	
16) Output voltage	0 - 200	V
17) Minimum operation temperature	- 60	°C
18) Maximum operation temperature	+ 80	°C
19) Acoustic levels at a distance of 20 m ? at nacelle ? (wind = 5 m/s)	0	DB
20) Lifetime	100	Years
21) Is the machine self-starting	Yes	
22) Use of an asynchronous generator	Yes	
23) Yaw control system	Not needed	
24) Upwind or downwind		

Calculated power curve

Wind speed (m/s)	Power (W)
1	0
2	30
3	60
4	120
5	211
6	300
7	450
8	630
9	900
10	1 200
11	1 650
12	2 100
13	2 700
14	3 600

In battery charging the electricity production will be changing according to the voltage level chosen by the client.

Power curve

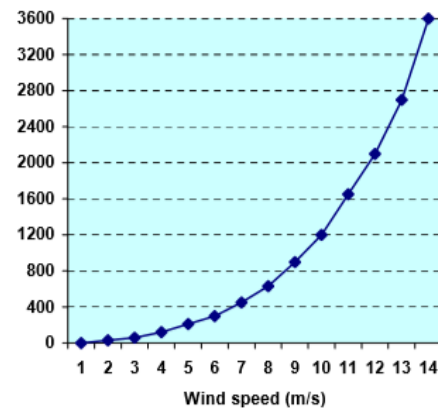


Figura 7. Oy Windside WS-12 references (WS-12/ 8 kW)

Donde los valores de Velocidad del Viento (m/s) presentados, utilizando la curva de potencia, permite estimar la producción de potencia (W) sobre los datos de San Francisco seleccionados, lo que se utilizó para desarrollar el ejercicio de Predicción de Generación de Energía Eléctrica. El resultado de la interpolación de los datos, se muestra en la tabla 3, donde se pueden observar algunos de los valores como: Velocidad del viento, los límites

inferior y superior de la curva de potencia, y la potencia generada a partir de la velocidad del viento si se utilizara la torre OY Windside Production Ltd - VAWT de 1 kW to 8 kW.

Tabla 3. Interpolación con Datos de San Francisco (WS-4B & 4C/ 1-2 kW)

Ind_Time	Wind_Veloc	Lim Inf	Lim Sup	Winferior	Wsuperior	W
/2010 17:15	10.37	10	11	1200	1650	1366.5
/2010 17:20	11.61	11	12	1650	2100	1924.5
/2010 17:25	12.88	12	13	2100	2700	2628
/2010 17:30	12.14	12	13	2100	2700	2184
/2010 17:35	12.62	12	13	2100	2700	2472
/2010 17:40	12.16	12	13	2100	2700	2196
/2010 17:45	12.21	12	13	2100	2700	2226
/2010 17:50	11.73	11	12	1650	2100	1978.5
/2010 17:55	11.76	11	12	1650	2100	1992
/2010 18:00	11.77	11	12	1650	2100	1996.5
/2010 18:05	12.42	12	13	2100	2700	2352
/2010 18:10	12.75	12	13	2100	2700	2550
/2010 18:15	12.19	12	13	2100	2700	2214
/2010 18:20	11.99	11	12	1650	2100	2095.5
/2010 18:25	11.9	11	12	1650	2100	2055
/2010 18:30	12.02	12	13	2100	2700	2112
/2010 18:35	11.7	11	12	1650	2100	1965
/2010 18:40	12.07	12	13	2100	2700	2142
/2010 18:45	12.38	12	13	2100	2700	2328
/2010 18:50	11.62	11	12	1650	2100	1929
/2010 18:55	11.19	11	12	1650	2100	1735.5
/2010 19:00	12.26	12	13	2100	2700	2256
/2010 19:05	11.89	11	12	1650	2100	2050.5
/2010 19:10	11.91	11	12	1650	2100	2059.5
/2010 19:15	12.26	12	13	2100	2700	2256
/2010 19:20	12.55	12	13	2100	2700	2430
/2010 19:25	12.53	12	13	2100	2700	2418
/2010 19:30	12.72	12	13	2100	2700	2532
/2010 19:35	11.96	11	12	1650	2100	2082
/2010 19:40	11.77	11	12	1650	2100	1996.5
/2010 19:45	11.51	11	12	1650	2100	1879.5
/2010 19:50	12.09	12	13	2100	2700	2154
/2010 19:55	11.6	11	12	1650	2100	1920
/2010 20:00	11.19	11	12	1650	2100	1735.5
/2010 20:05	11.15	11	12	1650	2100	1717.5
/2010 20:10	11.48	11	12	1650	2100	1866
/2010 20:15	11.37	11	12	1650	2100	1816.5
/2010 20:20	10.25	10	11	1200	1650	1312.5
/2010 20:25	9.98	9	10	900	1200	1194
/2010 20:30	9.32	9	10	900	1200	996
/2010 20:35	9.44	9	10	900	1200	1032
/2010 20:40	9.99	9	10	900	1200	1197
/2010 20:45	9.52	9	10	900	1200	1056
/2010 20:50	9.91	9	10	900	1200	1173
/2010 20:55	9.92	9	10	900	1200	1176
/2010 21:00	10.19	10	11	1200	1650	1285.5
/2010 21:05	10.36	10	11	1200	1650	1362
/2010 21:10	9.81	9	10	900	1200	1143
/2010 21:15	10.19	10	11	1200	1650	1285.5
/2010 21:20	9.96	9	10	900	1200	1188
/2010 21:25	9.14	9	10	900	1200	942
/2010 21:30	10.21	10	11	1200	1650	1294.5
/2010 21:35	10.37	10	11	1200	1650	1366.5
/2010 21:40	10.05	10	11	1200	1650	1222.5
/2010 21:45	9.87	9	10	900	1200	1161
/2010 21:50	10.06	10	11	1200	1650	1227

4.3. Selección de los parámetros y entrenamiento del modelo sobre los datos de San Francisco

Para llevar a cabo el experimento de la selección de los parámetros, entrenamiento y comprobación del modelo, se dividió la serie de tiempo con los datos urbanos de San Francisco, como se describe en la siguiente tabla:

Tabla 4. División de los datos para el experimento sobre la Serie de Tiempo

Datos de San Francisco - 100%		
246,612		
Selección y entrenamiento del modelo - 80%		Comprobación - 20%
197,290		49,323
Selección de parámetros - 50%	Entrenamiento - 50%	Dividido en muestras
98,645	98,645	10,000

De 246,612 (100%) de los datos se dividió en 197,290 (80%) para la selección de los parámetros y entrenamiento del modelo. El otro 49,232 (20%) restante se dividió de 10,000 en 10,000 para realizar la comprobación del modelo.

Del 80% de los datos se dividió en 98,645 (50%) para la selección de los parámetros del modelo y 98,645 (50%) para el entrenamiento para el desarrollo del ejercicio sobre la serie de tiempo.

Posteriormente, se desarrolló un código de programación sobre el lenguaje de R studio (Fox, 2018), en el que se propone una combinación de parámetros para un conjunto de órdenes de $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$ y se realizaron las pruebas de entrenamiento y predicción con todas las combinaciones posibles con la herramienta $arima(p,d,q)$ y $predict([Serie de Tiempo], \dots)$ del paquete de librerías de stats y forecast del paquete *forecast* (Hyndman R. J., 2014).

Para determinar los parámetros del modelo, se realizaron los cálculos de medidas de error estándar sobre las 72 combinaciones posibles de los órdenes de $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$ como se muestra en la siguiente tabla:

Tabla 5. 72 combinaciones de ARIMA y sus RMSE con datos de San Francisco. $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$

Modelo	RMSE	Modelo	RMSE
ARIMA(0,0,0)	867.9560438	ARIMA(2,0,3)	1003.032298
ARIMA(1,0,0)	1004.162707	ARIMA(3,0,3)	989.3242976
ARIMA(2,0,0)	1005.043074	ARIMA(4,0,3)	991.9601811
ARIMA(3,0,0)	1004.631496	ARIMA(5,0,3)	1004.987008
ARIMA(4,0,0)	1003.95289	ARIMA(0,1,3)	1009.42402
ARIMA(5,0,0)	1003.5181	ARIMA(1,1,3)	1009.46244
ARIMA(0,1,0)	1017.510606	ARIMA(2,1,3)	1009.18133
ARIMA(1,1,0)	1014.240572	ARIMA(3,1,3)	1009.023087
ARIMA(2,1,0)	1012.375046	ARIMA(4,1,3)	1009.233159
ARIMA(3,1,0)	1011.078855	ARIMA(5,1,3)	1009.014318
ARIMA(4,1,0)	1010.317288	ARIMA(0,0,4)	905.3572359
ARIMA(5,1,0)	1009.938112	ARIMA(1,0,4)	1002.961158
ARIMA(0,0,1)	764.1899802	ARIMA(2,0,4)	1003.106948
ARIMA(1,0,1)	1003.871318	ARIMA(3,0,4)	996.9195914
ARIMA(2,0,1)	1002.932879	ARIMA(4,0,4)	1001.598591
ARIMA(3,0,1)	1003.054254	ARIMA(5,0,4)	986.1939253
ARIMA(4,0,1)	1002.933588	ARIMA(0,1,4)	1009.240404
ARIMA(5,0,1)	1003.727233	ARIMA(1,1,4)	1009.261956
ARIMA(0,1,1)	1010.879307	ARIMA(2,1,4)	1009.583676
ARIMA(1,1,1)	1009.316577	ARIMA(3,1,4)	1010.128543
ARIMA(2,1,1)	1009.308994	ARIMA(4,1,4)	1011.884632
ARIMA(3,1,1)	1009.036311	ARIMA(5,1,4)	1011.188145
ARIMA(4,1,1)	1007.399669	ARIMA(0,0,5)	919.6118636
ARIMA(5,1,1)	1009.405813	ARIMA(1,0,5)	1002.936538
ARIMA(0,0,2)	833.2729991	ARIMA(2,0,5)	1003.039037
ARIMA(1,0,2)	1003.052612	ARIMA(3,0,5)	997.8250827
ARIMA(2,0,2)	1001.807321	ARIMA(4,0,5)	995.88353
ARIMA(3,0,2)	1002.977817	ARIMA(5,0,5)	1003.522921
ARIMA(4,0,2)	1003.001075	ARIMA(0,1,5)	1009.115465
ARIMA(5,0,2)	1002.931064	ARIMA(1,1,5)	1009.135139
ARIMA(0,1,2)	1009.555717	ARIMA(2,1,5)	1001.85183
ARIMA(1,1,2)	1009.547398	ARIMA(3,1,5)	1008.894766
ARIMA(2,1,2)	1009.351308	ARIMA(4,1,5)	1005.825384
ARIMA(3,1,2)	1009.154721	ARIMA(5,1,5)	995.3099747
ARIMA(4,1,2)	1009.032172		
ARIMA(5,1,2)	1007.277618		
ARIMA(0,0,3)	877.9328357		
ARIMA(1,0,3)	1003.043415		

Cabe mencionar que los resultados del cálculo del RMSE mostrados en la Tabla 5 y posteriores, son más lejanos a 0, debido a que los cálculos se desarrollaron sobre la serie de tiempo que recibió la interpolación de velocidad (m/s) a potencia (W) pudiendo alcanzar hasta un máximo de 3600 Watts, destacando que la selección de los mejores modelos se basó en la aplicación del mismo criterio de medición del RMSE y MAE los cuales arrojaron los menores resultados, más próximos a 0.

Para estimar el día de anticipo, en el código de simulación en R, se asignó un valor al parámetro `n.ahead` (de la librería `stats`) = 288 intervalos de tiempo equivalentes a un día de anticipo (sobre el horizonte de lectura de datos de cada 5 minutos), para realizar la predicción.

De los ejercicios de las 72 combinaciones mostrados en la tabla 5, se eligieron las mejores 5 con base en su menor RMSE, así como se muestra a continuación:

Tabla 6. Medidas de error de las mejores 5 torres de las 72 combinaciones de San Francisco

Modelo	RMSE (W)
ARIMA(0,0,1)	764.189
ARIMA(0,0,2)	833.272
ARIMA(0,0,3)	877.932
ARIMA(0,0,4)	905.357
ARIMA(0,0,5)	919.611

Cabe destacar que en estos ejercicios, el mejor modelo fue el ARIMA(0,0,1) equivalente a un modelo MA(1).

4.4. Pronóstico con comprobación y pruebas con los mejores 5 modelos sobre los datos de San Francisco en base a su menor RMSE y MAE

Para las pruebas de validación o test, de los resultados del ejercicio anterior, se repitió el ejercicio anterior, pero únicamente con las 5 mejores combinaciones que presentaron el menor RMSE, como se muestra en la tabla 6.

Nuevamente, se procedió a ejecutar la función que se desarrolló en lenguaje R para el cálculo de los parámetros de los modelos ARIMA, pero ahora para desarrollar la predicción sobre parámetros conocidos con la función `predict(object,...)`, la cual es una función de la librería `stats`, genérica para desarrollar predicciones a partir del resultado de los ajustes en el modelo, utilizando R studio:

```
tmp.fitT <- arima(tmp.trainT, order = c(reer$p[j],reer$d[j],reer$q[j]), include.mean = T)
tmp.yhatT <- predict(tmp.fitT, n.ahead = 288, newdata = tmp.localT)
```

Figura 8. Función arima y predict de la librería stats en R para San Francisco

Para la fase de comprobación del modelo, los ejercicios nuevamente se desarrollaron como en el paso anterior, pero ahora con el 20% restante de los datos de validación. Los resultados concentrados del análisis de los modelos ARIMA(p,d,q) se muestran a continuación en forma de tabla.

Tabla 7. Medidas de error de las mejores torres de San Francisco en potencia (W)

Modelo	RMSE (W)	MAE (W)
ARIMA(0,0,1)	390.589	302.451
ARIMA(0,0,2)	365.007	257.333
ARIMA(0,0,3)	361.331	242.165
ARIMA(0,0,4)	362.226	234.602
ARIMA(0,0,5)	363.390	231.133

Medidas de error

En tabla 7, se muestra que el mejor modelo que tiene el menor RMSE (variabilidad) y MAE (precisión) sobre los datos de validación del 20% (49,323) con el ejercicio de 10,000 en 10,000 de los datos de San Francisco, es el **ARIMA(0,0,5) equivalente a MA(5)**.

Los resultados obtenidos, se muestran gráficamente, en donde las líneas de azul muestran las observaciones sobre la serie de tiempo y las de rojo indican los valores de la predicción sobre esta misma.

Para ARIMA (0,0,1) con la muestra de 10,000 datos para la prueba, sus RMSE = 390,589 W y MAE = 302.451 W.

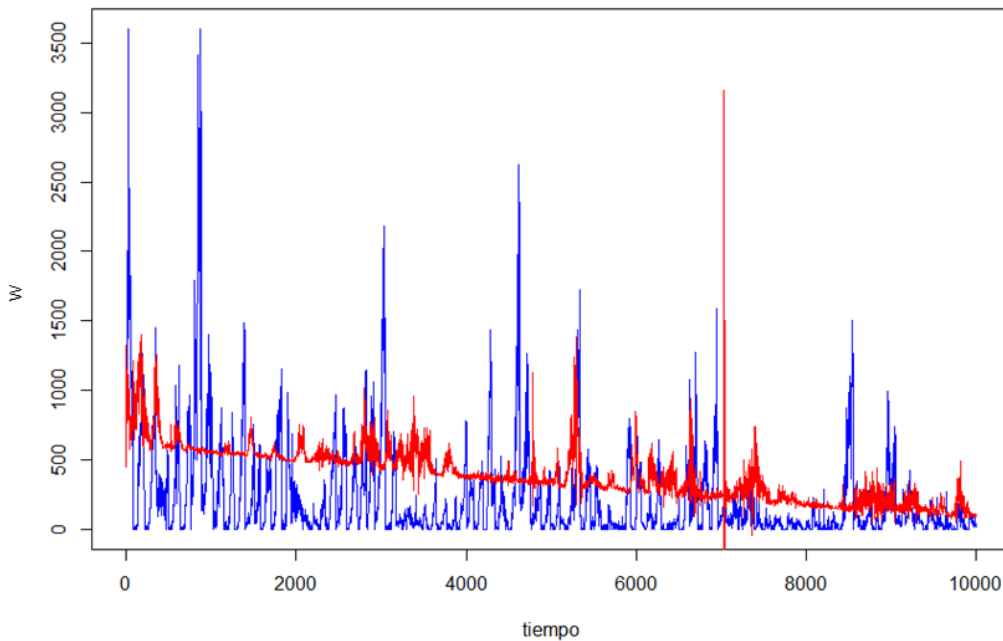


Figura 9. Ejercicio de validación con el modelo ARIMA (0,0,1) de San Francisco

Para ARIMA(0,0,2) se obtuvieron los siguientes valores: RMSE = 365.007 W y MAE = 257.333 W.

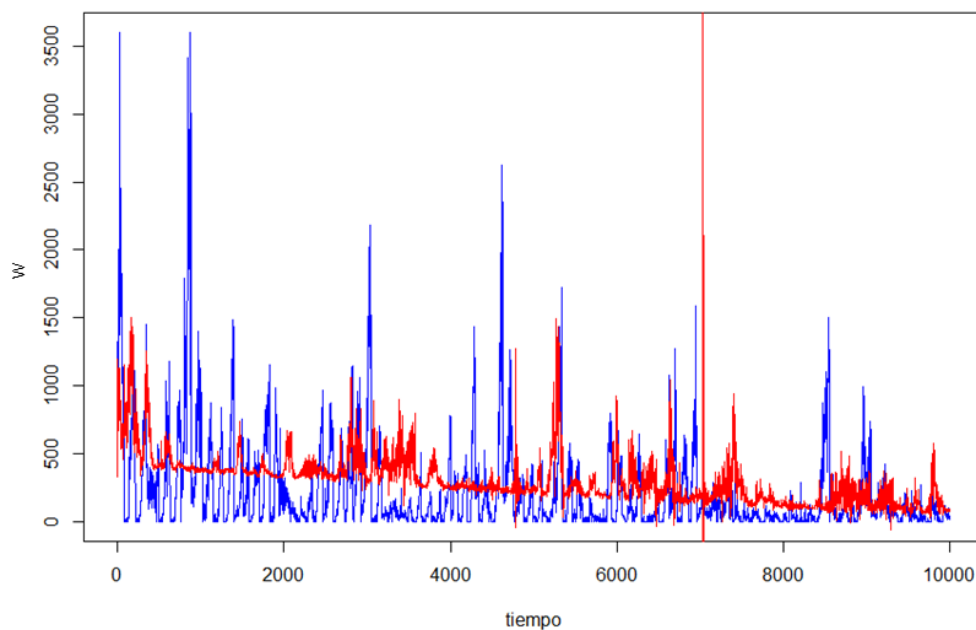


Figura 10. Ejercicio de validación con el modelo ARIMA (0,0,2) de San Francisco

Para ARIMA(0,0,3) con 10000 datos, se obtuvieron los siguientes RMSE = 361.331 W y MAE = 242.165 W.

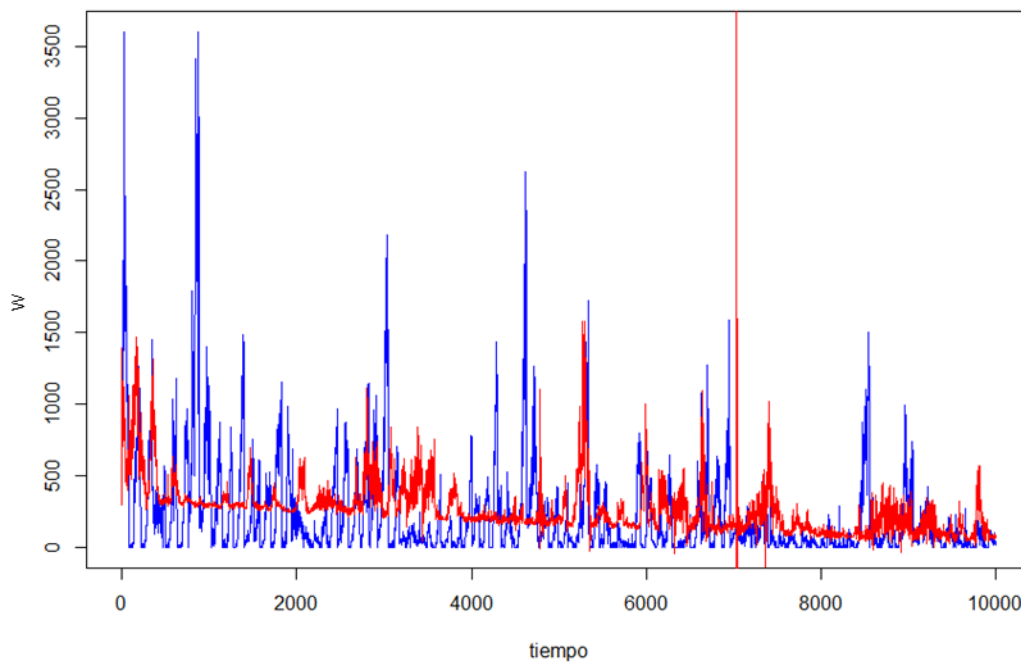


Figura 11. Ejercicio de validación con el modelo ARIMA (0,0,3) de San Francisco

Para ARIMA(0,0,4) con 10000 datos, se obtuvieron los siguientes RMSE = 362.226 W y MAE = 234.602 W.

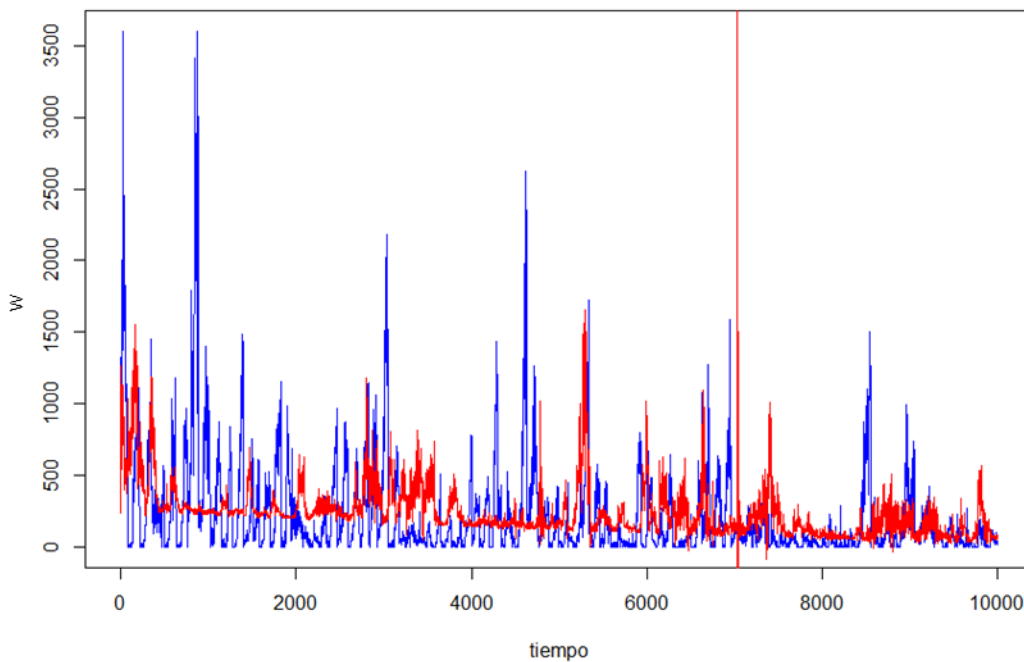


Figura 12. Ejercicio de validación con el modelo ARIMA (0,0,4) de San Francisco

Para ARIMA(0,0,5) con 10000 datos, se obtuvieron los siguientes RMSE = 363.390 W y MAE = 231.133 W.

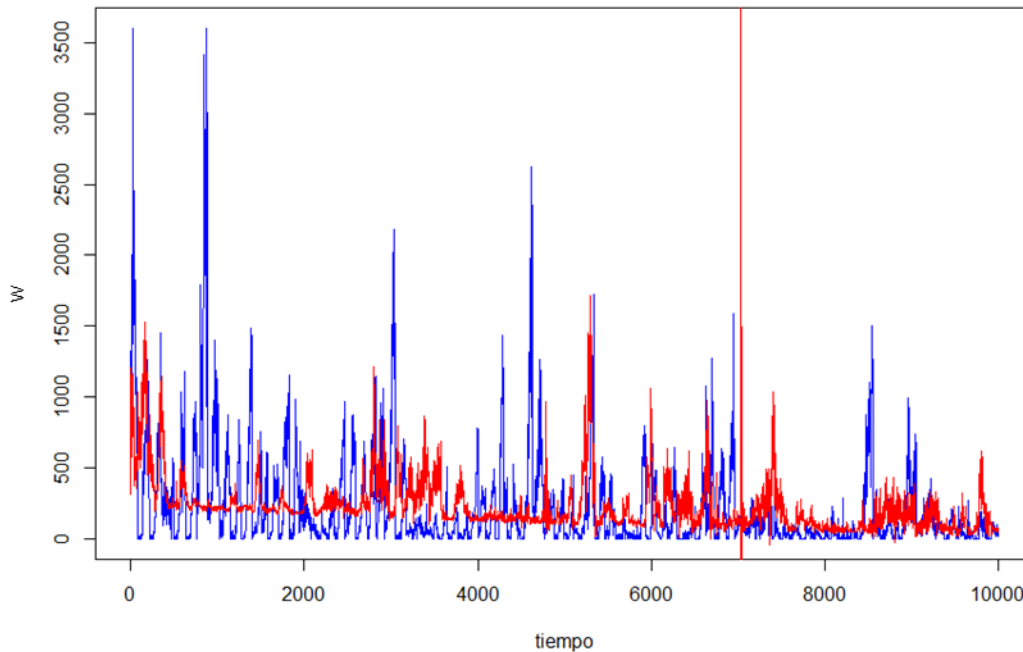


Figura 13. Ejercicio de validación con el modelo ARIMA (0,0,5) de San Francisco

4.5. Recopilación de los datos de un Parque eólico denominado CERTE

Para el ejercicio del CERTE (Centro Regional de Tecnología Eólica), se obtuvieron datos de tres periodos completos de una estación anemométrica denominada la Ventosa2, ubicada en la Ciudad de Juchitán de Zaragoza, Oaxaca. México, en una zona montañosa en el Norte de 5 km y plano en el Este, Sur y Oeste. De una torre de altura de 80 metros, con alturas de medición de 40, 60 y 80 metros en un horizonte de tiempo de 10 min. (La estación monitorea todas las variables cada 2 segundos. Calcula y registra las velocidades promedio del viento, las velocidades máximas del viento, las direcciones del viento en ambos niveles cada 10 minutos). Los parámetros registrados son: Velocidad, Dirección y Desviación Estándar del viento.

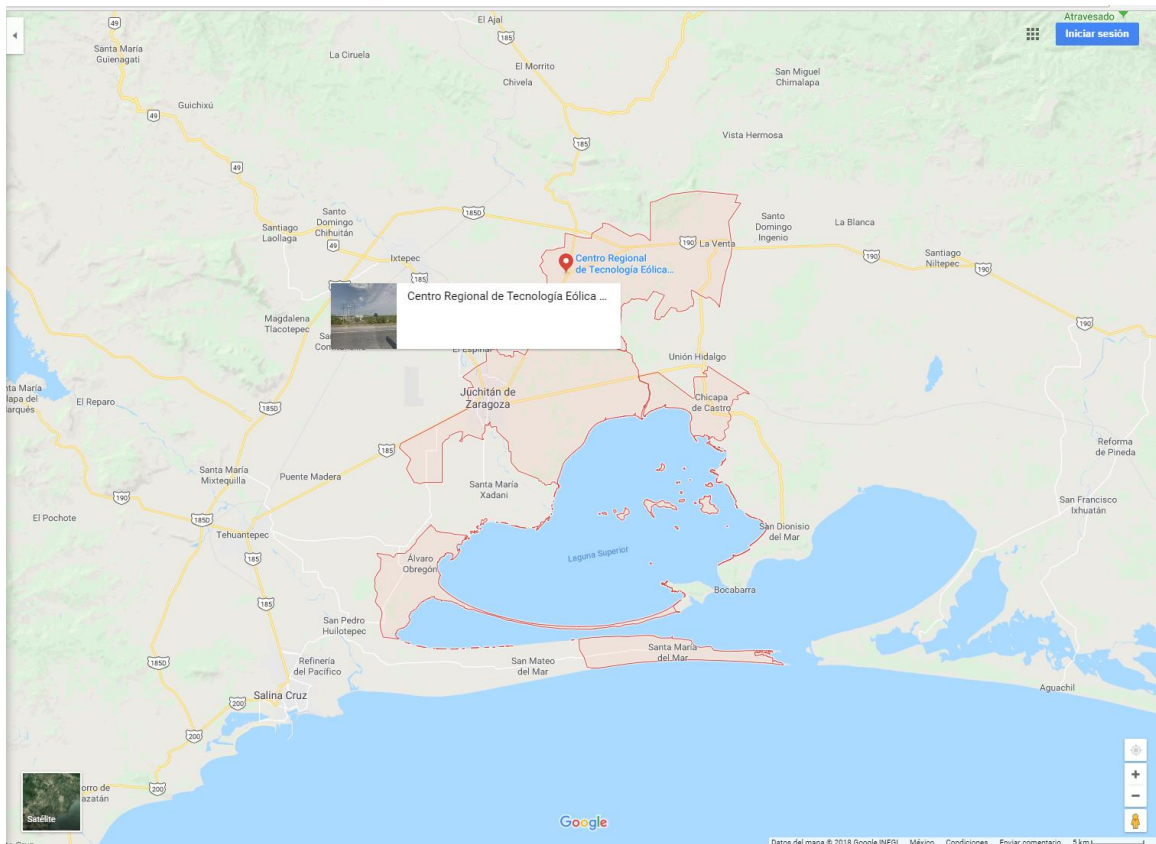


Figura 14. Ubicación del CERTE la Ventosa 2 en la Ciudad de Juchitán de Zaragoza, Oaxaca, México (citar la fuente)

4.6. Pre procesamiento de los datos del CERTE

En esta etapa, se lleva a cabo el pre procesamiento con los datos del CERTE y se repitió el ejercicio completo que se llevó a cabo como el ejercicio de San Francisco.

Limpieza y normalización de los datos del CERTE

Para los datos del CERTE, se cuenta con información de 3 periodos completos de los años 2007, 2012 a 2013 y 2015, donde cada periodo es equivalente a 12 meses de mediciones cada 10 minutos de una torre, para la cual se realizó la limpieza de los datos y se calculó su media = 140.1921 y su desviación estándar = 117.4326.

Tabla 8. Momentos estadísticos para los datos del CERTE – Velocidad del viento (m/s)

CERTE m/s	media	mediana	sd	kurtosis	sesgo	min	max
	9.359	9.01	5.270	2.944	0.475	0	32.5

Análisis exploratorio de los datos

En la etapa de análisis exploratorio de los datos, se desarrolló la rosa del viento para los datos convencionales, de una torre de la región del CERTE de la Ciudad de Juchitán de Zaragoza de Oaxaca, México (De la cual se contaba con la información de 3 periodos completos para el ejercicio).

Para desarrollar la rosa del viento del CERTE, se aplicó el mismo procedimiento de codificación que en San Francisco.

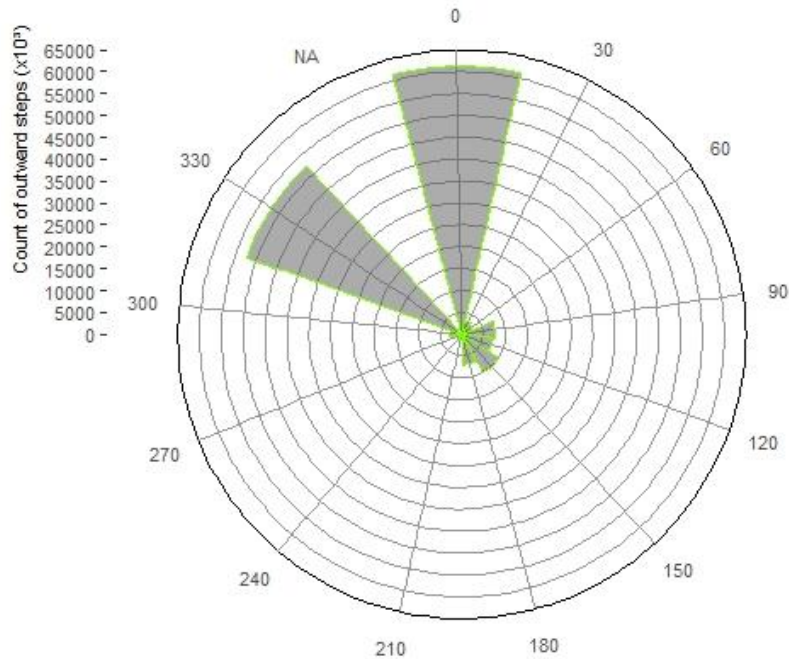


Figura 15. Rosa del viento con datos del CERTE (Región convencional) – Cd. de Juchitán de Zaragoza, Oaxaca, México.

Del análisis, se puede apreciar que el origen del viento favorable y significativo para la región del CERTE es proveniente del Noroeste.

Interpolación de la serie de tiempo

Para los datos del CERTE, se llevó a cabo la interpolación de los datos de velocidad del viento a potencia sobre un catálogo de fabricantes europeo de turbinas eólicas (programme, European Commission under the Intelligent Energy - Europe, 2017). Es decir; de las turbinas eólicas que arrojaran mayor cantidad de potencia se eligió la de mayor completitud de información.

Para llevar a cabo el desarrollo de la interpolación lineal se utilizó la misma ecuación y procedimiento que se utilizó en el ejercicio de San Francisco:

$$y = y_a + (x - x_a) \frac{(y_b - y_a)}{(x_b - x_a)}$$

Ecuación 8. Interpolación para CERTE

Para el ejercicio del CERTE se eligió una torre denominada: T300-28, con una curva de potencia con las siguientes especificaciones (The WindPower - Wind Energy Market Intelligence, 2018):

Tabla 9. Especificaciones de la torre denominada T300-28 (Cita)

Main data	Rotor	Wind speeds
Rated power: 300 kW Rotor diameter: 28 m Old model Offshore model: no	Number of blades: 3 Type: Pitch Swept area: 615.8 m ² Power density: 2.05 m ² /kW Maximum speed: 43 rd/min	Cut-in wind speed: 4,5 m/s Rated wind speed: 14 m/s Cut-off wind speed: 25 m/s
Generator	Tower	Gearbox
Type: ASYNC Number: 1 Maximum speed: 1500 rd/min Voltage: 690 V	Minimum hub height: 30 m Maximum hub height: 30 m	Gearbox: yes Stages: 3

De los valores de la torre T300-28, se muestra la curva de potencia que alcanza hasta 300 kW con base en la velocidad del viento, así como observa en la siguiente figura:

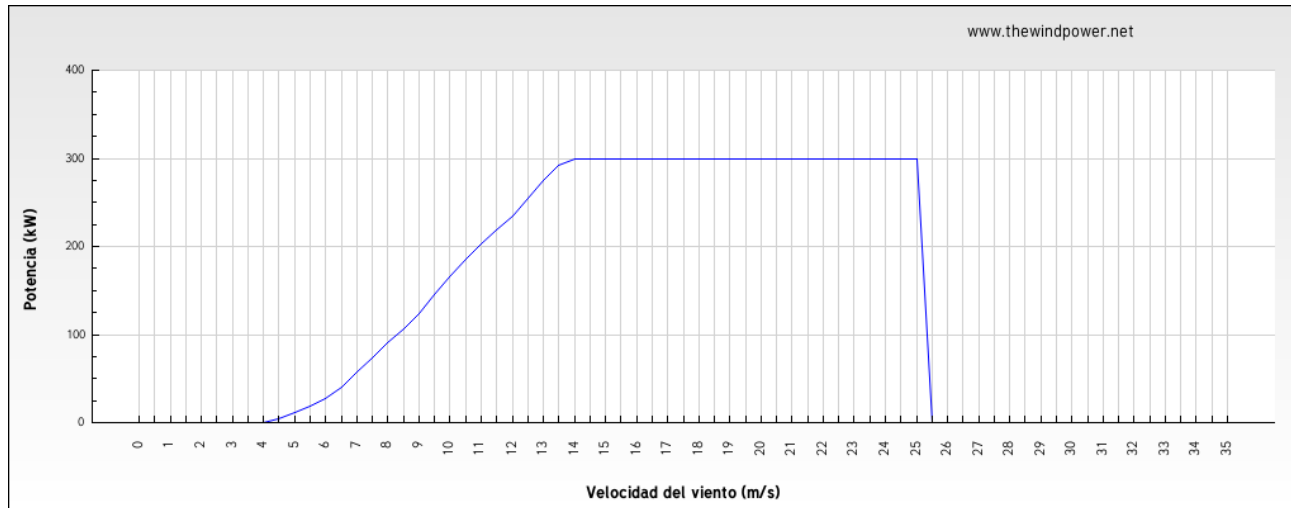


Figura 16. Curva de potencia de T300-28 (Turbowinds, año)

Para la interpolación de los datos de Velocidad del viento del CERTE sobre esta curva de potencia, se pueden apreciar los resultados en kilo Watts (kW) de la columna 5 de la Tabla 10, donde es posible también observar que la medición de los datos se efectúa cada 10 minutos y asimismo la dirección del viento.

Horaymin	Orden	Direccion80	Velocidad80	Kw
10	1	329.6	16.03	300
20	2	329.3	16.84	300
30	3	327.3	15.99	300
40	4	327.3	15.9	300
50	5	327	15.18	300
100	6	327.5	14.72	300
110	7	327	14.53	300
120	8	329.3	15.71	300
130	9	326.6	15.89	300
140	10	329.2	15.31	300
150	11	328.5	13.81	295.25
200	12	332.8	15.18	300
210	13	329.4	14.56	300
220	14	328.6	15.96	300
230	15	328	15.75	300
240	16	327.7	14.98	300
250	17	330.4	16.5	300
300	18	328.8	15.09	300
310	19	330.3	16.81	300
320	20	328.8	15.33	300
330	21	329.8	15.78	300
340	22	328.2	14.93	300
350	23	329.5	15.37	300
400	24	329.4	14.74	300
410	25	327.4	14.87	300
420	26	327.1	15.32	300
430	27	326.8	15.55	300
440	28	328.5	16.03	300
450	29	324.2	15.26	300
500	30	324.7	15.44	300
510	31	326.3	14.65	300
520	32	325.1	15.1	300
530	33	324.8	14.86	300
540	34	322.6	15.48	300
550	35	322.3	13.26	281.5
600	36	323.7	13.35	283.75
610	37	322.3	12.37	251.375
620	38	321.6	12.37	251.375
630	39	323.4	12.31	249.125
640	40	325.4	14.43	300
650	41	327.3	15.34	300
700	42	327.2	15.12	300
710	43	328.9	15.29	300
720	44	327.4	15.79	300
730	45	324.2	15.92	300
740	46	326	15.88	300
750	47	325.5	15.83	300

Tabla 10. Interpolación con Datos del CERTE en Kw (T300-28)

4.7. Selección de los parámetros y entrenamiento del modelo sobre los datos del CERTE

Para el experimento de la selección de los parámetros, entrenamiento y comprobación del modelo, se dividió la serie de tiempo con los datos convencionales del CERTE, así como se describe en la siguiente tabla:

Tabla 11. División de los datos para el experimento sobre la Serie de Tiempo

Datos del CERTE - 100%		
157,824		
Selección y entrenamiento del modelo - 80%		Comprobación - 20%
126,260		31,564
Selección de parámetros - 50%	Entrenamiento - 50%	Dividido en muestras
63,130	63,130	10,000

De 157,824 (100%) de los datos del CERTE se dividió en 126,260 (80%) para la selección de los parámetros y entrenamiento del modelo. El otro 31,564 (20%) restante dividido en muestras de 10,000 en 10,000 se utilizó para la comprobación.

Del 80% de los datos se dividió en 63,130 (50%) para la selección de los parámetros del modelo y 63,130 (50%) para el entrenamiento para el desarrollo del ejercicio sobre la serie de tiempo.

Así como el ejercicio anterior, para el CERTE, se desarrolló un código de programación sobre el lenguaje de R studio (Fox, 2018), en el que se propone una combinación de parámetros para un conjunto de órdenes de $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$ y se realizaron las pruebas de entrenamiento y predicción con todas las combinaciones posibles con la herramienta `arima(p,d,q)` y `predict([Serie de Tiempo],...)` del paquete de librerías de `stats` y `forecast` del paquete `forecast` (Hyndman R. J., 2014).

El entrenamiento del modelo, se lleva a cabo para determinar los mejores parámetros del modelo, bajo el cual se realizaron los cálculos de medidas de error estándar sobre las 72 combinaciones posibles de los órdenes de $p = [0,5]$, $d = [0,1]$, $q = [0,5]$ como se muestra a continuación:

Tabla 12. 72 combinaciones de ARIMA y sus RMSE con datos del CERTE con $p = \{0, \dots, 5\}$, $d = \{0, 1\}$, $q = \{0, \dots, 5\}$

Modelo	RMSE	Modelo	RMSE
ARIMA(0, 0, 0)	133.9157645	ARIMA(3, 0, 3)	114.2280018
ARIMA(1, 0, 0)	114.3776173	ARIMA(4, 0, 3)	117.2258769
ARIMA(2, 0, 0)	114.7676021	ARIMA(5, 0, 3)	115.077566
ARIMA(3, 0, 0)	114.9322955	ARIMA(0, 1, 3)	116.0736281
ARIMA(4, 0, 0)	115.0319363	ARIMA(1, 1, 3)	116.0738764
ARIMA(5, 0, 0)	115.0831792	ARIMA(2, 1, 3)	116.0971727
ARIMA(0, 1, 0)	116.1564029	ARIMA(3, 1, 3)	116.0723305
ARIMA(1, 1, 0)	116.1339377	ARIMA(4, 1, 3)	116.0867838
ARIMA(2, 1, 0)	116.119494	ARIMA(5, 1, 3)	115.8902586
ARIMA(3, 1, 0)	116.1010123	ARIMA(0, 0, 4)	104.9359761
ARIMA(4, 1, 0)	116.0886488	ARIMA(1, 0, 4)	115.1093243
ARIMA(5, 1, 0)	116.0966656	ARIMA(2, 0, 4)	115.1177353
ARIMA(0, 0, 1)	104.4009305	ARIMA(3, 0, 4)	115.1061943
ARIMA(1, 0, 1)	114.9805472	ARIMA(4, 0, 4)	115.097985
ARIMA(2, 0, 1)	115.1165706	ARIMA(5, 0, 4)	115.0803079
ARIMA(3, 0, 1)	115.1205661	ARIMA(0, 1, 4)	116.0732043
ARIMA(4, 0, 1)	115.0822083	ARIMA(1, 1, 4)	116.0712345
ARIMA(5, 0, 1)	115.0744701	ARIMA(2, 1, 4)	116.075201
ARIMA(0, 1, 1)	116.1149171	ARIMA(3, 1, 4)	116.0789565
ARIMA(1, 1, 1)	116.0705972	ARIMA(4, 1, 4)	116.0840979
ARIMA(2, 1, 1)	116.0663593	ARIMA(5, 1, 4)	116.4051468
ARIMA(3, 1, 1)	116.0846998	ARIMA(0, 0, 5)	105.9199697
ARIMA(4, 1, 1)	116.0916473	ARIMA(1, 0, 5)	115.0784473
ARIMA(5, 1, 1)	116.096309	ARIMA(2, 0, 5)	115.0825543
ARIMA(0, 0, 2)	102.0878476	ARIMA(3, 0, 5)	115.0927098
ARIMA(1, 0, 2)	115.0821578	ARIMA(4, 0, 5)	115.0942763
ARIMA(2, 0, 2)	116.0820119	ARIMA(5, 0, 5)	114.8552541
ARIMA(3, 0, 2)	114.9168998	ARIMA(0, 1, 5)	116.0827606
ARIMA(4, 0, 2)	115.0991555	ARIMA(1, 1, 5)	116.0830554
ARIMA(5, 0, 2)	115.0478288	ARIMA(2, 1, 5)	116.0815338
ARIMA(0, 1, 2)	116.0881199	ARIMA(3, 1, 5)	116.0798898
ARIMA(1, 1, 2)	116.0658139	ARIMA(4, 1, 5)	115.6912569
ARIMA(2, 1, 2)	115.8233032	ARIMA(5, 1, 5)	112.0846362
ARIMA(3, 1, 2)	116.0761655		
ARIMA(4, 1, 2)	116.1013926		
ARIMA(5, 1, 2)	116.3074108		
ARIMA(0, 0, 3)	103.6082868		
ARIMA(1, 0, 3)	115.1129028		

Para el caso del CERTE, los resultados del cálculo del RMSE mostrados en la Tabla 12 y posteriores, son un poco más cercanos a 0, debido a que los cálculos se desarrollaron sobre la serie de tiempo que recibió la interpolación de velocidad (m/s) a potencia (kW) pudiendo alcanzar hasta un máximo de 300 kW, donde la selección de los mejores modelos se basó en la aplicación del mismo criterio de medición del RMSE y MAE los cuales arrojaron los menores resultados, más próximos a 0.

Para estimar el día de anticipo, en el código de simulación en R, se asignó un valor al parámetro n.ahead (de la librería stats) = 144 intervalos de tiempo (Figura 17) equivalentes a un día de anticipo (sobre el horizonte de lectura de datos de cada 10 minutos), para realizar la predicción.

De los ejercicios de las 72 combinaciones mostrados en la tabla 5, se eligieron las mejores 5 con base en su menor RMSE, así como se muestra a continuación:

Tabla 13. Medidas de error - RMSE de los 5 mejores parámetros del modelo para el CERTE

Modelo	RMSE (kW)
ARIMA(0,0,1)	104.400
ARIMA(0,0,2)	102.087
ARIMA(0,0,3)	103.608
ARIMA(0,0,4)	104.935
ARIMA(0,0,5)	105.919

Se puede observar en la tabla 13 que para este caso el mejor modelo es ARIMA(0,0,2) debido a que es el que presenta menor RMSE.

4.8. Pronóstico con comprobación y pruebas con los 5 mejores modelos sobre los datos del CERTE para recalcular su RMSE y MAE

Para las pruebas de validación o test, de los resultados del ejercicio anterior, se repitió el ejercicio anterior, pero únicamente con las mejores 5 combinaciones que presentaron menor RMSE, como se muestra en la tabla 13.

Posteriormente, se procedió a desarrollar el modelo de Predicción con la función `predict(object,...)`, la cual es una función de la librería `stats`, genérica para desarrollar predicciones a partir del resultado de funciones de ajuste del modelo en R studio:

```
tmp.fitT <- arima(tmp.trainT, order = c(reer$p[j],reer$d[j],reer$q[j]), include.mean = T)
tmp.yhatT <- predict(tmp.fitT, n.ahead = 144, newdata = tmp.localT)
```

Figura 17. Función arima y predict de la librería stats en R para CERTE

En la etapa de comprobación del modelo del ejercicio del CERTE, los ejercicios nuevamente se desarrollaron como en el paso anterior, pero ahora con el 20% restante de los datos de validación. Los resultados concentrados del análisis de los modelos ARIMA(p,d,q) se muestran a continuación en forma de tabla:

Tabla 14. Medidas de error de las mejores torres del CERTE con una muestra de 31,564 datos en kW

Modelo	RMSE (kW)	MAE (kW)
ARIMA(0,0,1)	121.450	98.048
ARIMA(0,0,2)	127.962	100.590
ARIMA(0,0,3)	131.609	102.425
ARIMA(0,0,4)	133.946	103.814
ARIMA(0,0,5)	135.373	104.597

Medidas de error

En tabla 14, se aprecia que el mejor modelo que arroja el menor RMSE y menor MAE sobre los datos de validación del CERTE es el modelo **ARIMA(0,0,1)** y que el término de error o de medias móviles es el más pequeño.

Los resultados obtenidos, se muestran a continuación gráficamente, en donde las líneas de azul muestran las observaciones sobre la serie de tiempo y las de rojo indican los valores de la predicción sobre la misma.

Para el caso de ARIMA (0,0,1) con 10000 datos para una muestra de la prueba, sus RMSE = 121.450 kW y MAE = 98.048 kW.

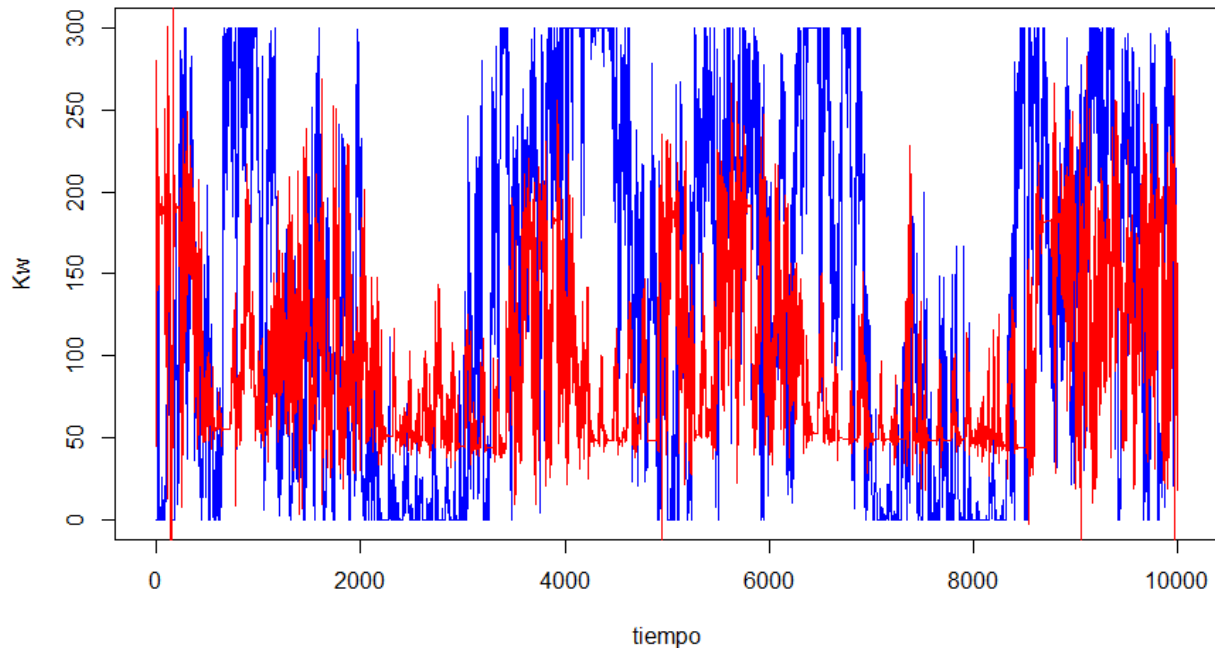


Figura 18. Ejercicio de validación con el modelo ARIMA (0,0,1) del CERTE

Para ARIMA (0,0,2) con 10000 datos para la prueba, sus RMSE = 127.962 kW y MAE = 100.590 kW.

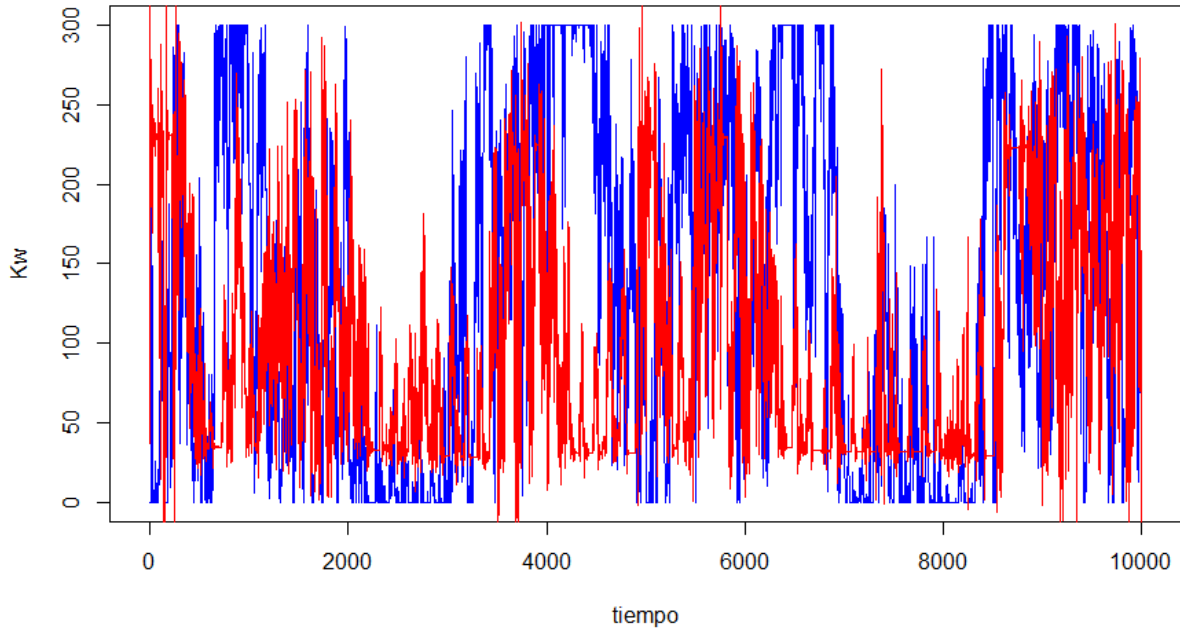


Figura 19. Ejercicio de validación con el modelo ARIMA (0,0,2) del CERTE

Para ARIMA (0,0,3) con 10000 datos para la prueba, sus RMSE = 131.609 kW y MAE = 102.425 kW.

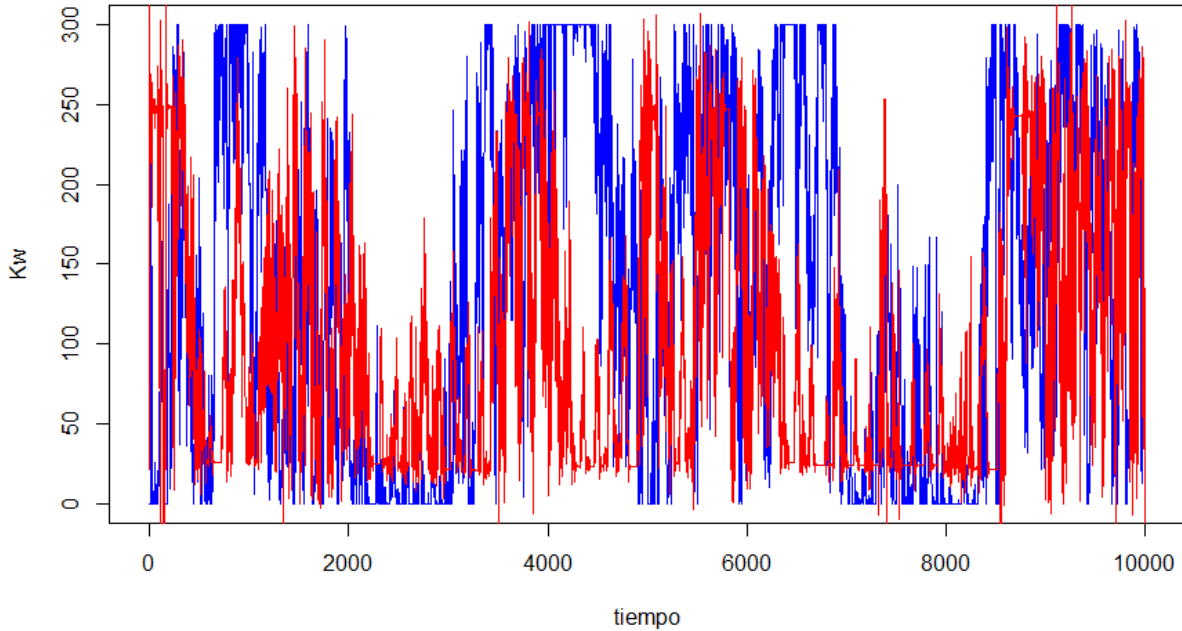


Figura 20. Ejercicio de validación con el modelo ARIMA (0,0,3) del CERTE

Para ARIMA (0,0,4) con 10000 datos para la prueba, sus RMSE = 133.946 kW y MAE = 103.814 kW.

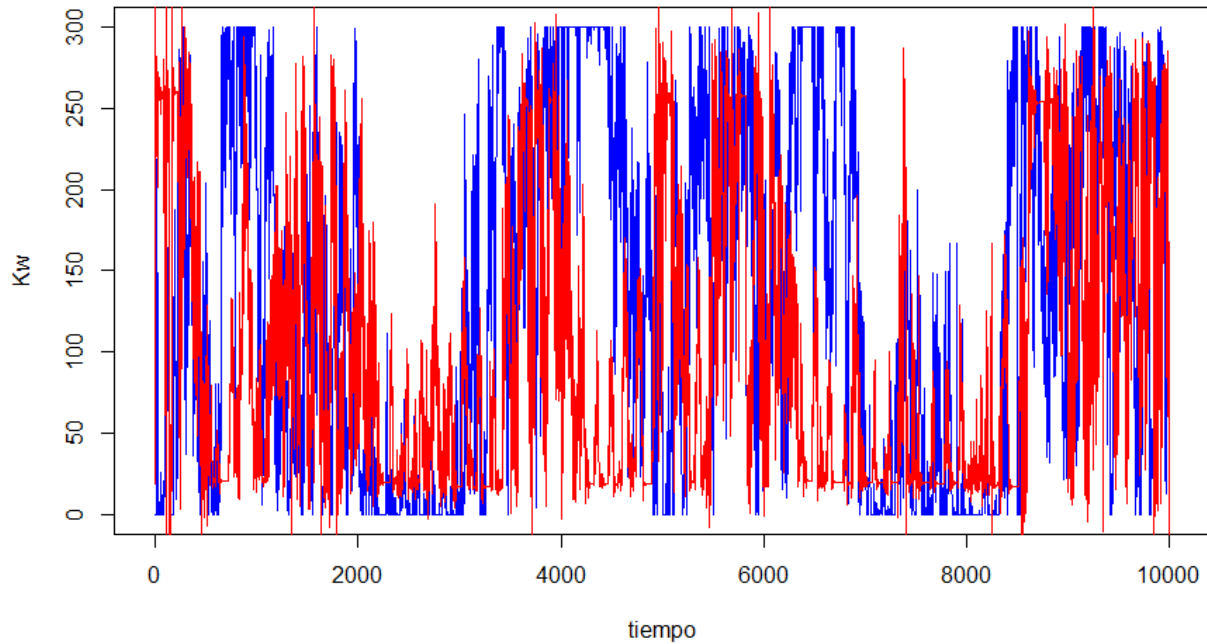


Figura 21. Ejercicio de validación con el modelo ARIMA (0,0,4) del CERTE

Para ARIMA (0,0,5) con 10000 datos para la prueba, sus RMSE = 135.373 kW y MAE = 104.597 kW.

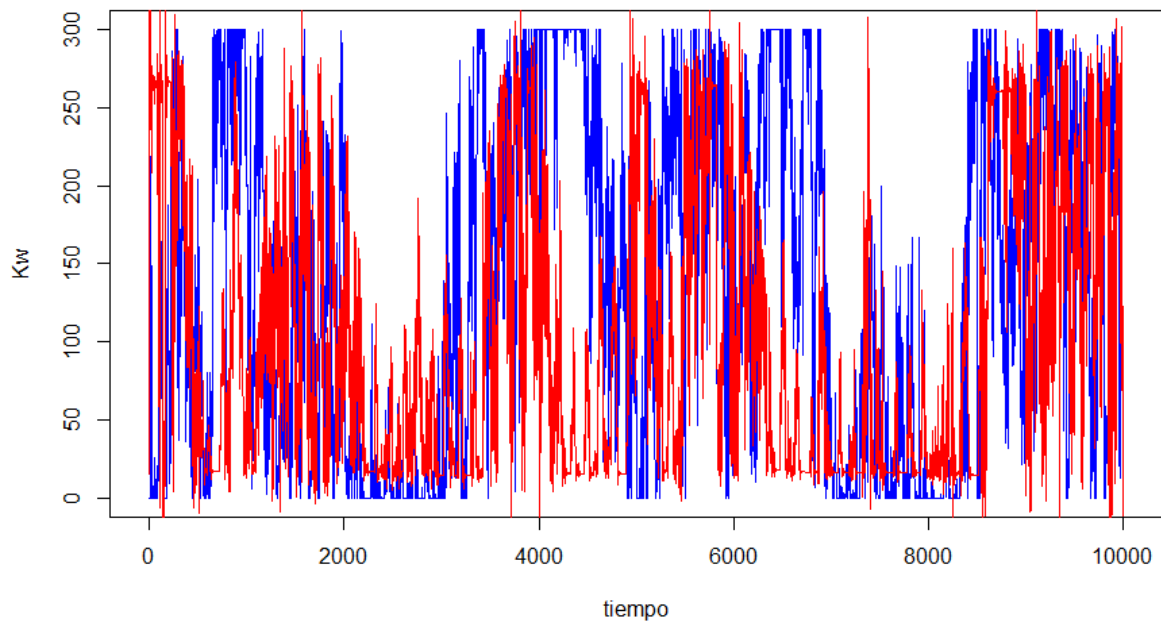


Figura 22. Ejercicio de validación con el modelo ARIMA (0,0,5) del CERTE

4.9. Análisis y resultados

Mediante estos experimentos, se llevó a cabo el análisis completo de los datos de San Francisco (urbanos) y del CERTE (convencionales), para los cuales se observan sus siguientes momentos estadísticos sobre sus datos de velocidad del viento:

Tabla 15. Momentos estadísticos con datos de San Francisco y CERTE sobre la velocidad del viento (m/s)

Velocidad del viento m/s	media	mediana	sd	kurtosis	sesgo	min	max
San Francisco	5.472	5.055	3.593	-0.101	0.436	0	58.41
CERTE	9.359	9.01	5.270	2.944	0.475	0	32.5

Con base en sus momentos estadísticos, se puede apreciar que la torre del CERTE ubicada a una altura mayor, entre 40,60 y 80 mts, de acuerdo a su mejor media, cuenta con un mayor recurso eólico sobre los datos de San Francisco que se encuentra ubicada a una altura de 40 mts.

Posteriormente, se llevó a cabo el análisis de las rosas del viento, para los dos casos, en las cuáles se puede apreciar que los vientos predominantes para ambos tienen su origen en la región norte. Para San Francisco del noreste y para el CERTE del noroeste.

En la fase experimental, sobre los resultados mostrados en las tablas 5 y 12, se aprecia considerablemente que los mejores modelos de acuerdo a sus menores RMSE y MAE, para San Francisco (zonas urbanas) son el **ARIMA(0,0,5)** equivalente a **MA(5)** con un **RMSE** de **363.390 W** y un **MAE** de **231.133 W**, y para el CERTE es **ARIMA(0,0,1)** equivalente a **MA(1)** con un **RMSE** de **121.450 kW** y un **MAE** de **98.048 kW**. Cabe mencionar que la torre de prueba de San Francisco muestra la interpolación en W (Watts) y la del CERTE en kW (kilo Watts) por lo que se realizó una transformación de San Francisco a kilo Watts para su comparación, dando como resultado lo siguiente:

Tabla 16. Mejores modelos para ambos casos en kW

	Modelo	RMSE (kW)	MAE (kW)
San Francisco	ARIMA(0,0,5)	0.363	0.231
CERTE	ARIMA(0,0,1)	121.450	98.048

En el resultado, se puede apreciar el mismo modelo para ambos, ya que no cuentan con parámetro autoregresivo, ni de integración, pero si cuentan con el parámetro de Media Móvil (*Moving Average*) o término de error sobre el modelo ARIMA(p,d,q) o MA(q).

Con base en la capacidad nominal de generación de Potencia de los aerogeneradores, se procedió a dividir el error obtenido entre el potencial de capacidad nominal de producción de la simulación con los aerogeneradores (3.2 kW y 300 kW) para obtener el mejor modelo en cuanto a su capacidad de predicción de acuerdo a su índice de variabilidad, es el de los datos de San Francisco, el cual tiene un menor RMSE = 10% y MAE = 6% sobre los del CERTE, los cuales cuentan con un RMSE = 40% y MAE = 32%.

Tabla 17. Mejores modelos en porcentaje en kW para San Francisco y CERTE

	Modelo	RMSE (%)	MAE (%)
San Francisco	ARIMA(0,0,5)	10%	6%
CERTE	ARIMA(0,0,1)	40%	32%

Así mismo, para el caso de San Francisco, durante la fase de predicción sobre los datos de prueba o *test*, se puede observar en las figuras 8-12 un comportamiento constante en su tendencia cuando el parámetro de Media Móvil (*Moving Average*) es mayor, y para el CERTE durante la fase de predicción sobre los datos de prueba o *test*, en las gráficas 16-20 se observa que el comportamiento de la tendencia se mantiene constante en el tiempo y el término de Media Móvil (*Moving Average*) para el mejor modelo es menor.

En la literatura existen diversos modelos estocásticos para realizar la Predicción de Generación de Energía Eléctrica a través de Eólica, dentro de ellos, se encuentran los de análisis en las series de datos como: Series de tiempo (ALEASOFT, AEOLIS, CASANDRA, CENER, GARRAD, HASSAN, METEOROLOGIA, METEOSIM o METEOTEMP (Enke,

1997)), Estadísticos (AR, ARMA, ARIMA) (Zhang, 2014). De Inteligencia Artificial: Redes Neuronales Artificiales (ANN), Redes Bayesianas (Pearl, 2014), lógica difusa y Máquinas de Soporte Vectorial (SVM) (Vapnik, 1998). Por su parte, los modelos físicos, se basan en consideraciones físicas como: Rugosidad, orografía, obstáculos y la atmósfera, mientras que, en los estocásticos, la predicción se basa en el análisis de la serie de tiempo, técnicas estadísticas de inteligencia artificial y técnicas de Redes Bayesianas para la predicción de potencia eólica. Cabe mencionar que de los modelos encontrados en la literatura, varios utilizan los modelos ARIMA(p,d,q) de Box & Jenkins para realizar sus ajustes sobre sus series de tiempo (Ibargüengoytia-González, 2018), para posteriormente dar paso a la predicción, similar a la metodología que se desarrolló en la presente tesis.

CAPÍTULO 5

Conclusiones y trabajo futuro

De acuerdo a los experimentos realizados, se demuestra que es posible realizar la Predicción de Generación de Energía Eléctrica con un día de anticipo, a partir de Datos eólicos utilizando los modelos autoregresivos Box-Jenkins.

Así mismo, se llevaron a cabo varios experimentos entre los que se eligieron para su presentación dos de ellos, uno para zonas urbanas de San Francisco y otro para zonas convencionales del CERTE en México, mediante los cuales se puede observar que para ambos casos de predicción utilizando modelos autoregresivos Box-Jenkins, a partir de datos eólicos, ambos modelos presentan características similares: ARIMA(0,0,5) o MA(5) y ARIMA(0,0,1) o MA(1), con una significativa diferencia en sus parámetros de Medias Móviles (MA).

Sobre el ejercicio, el modelo que presenta menor variabilidad de RMSE = 10% y MAE= 6% en la predicción, es el de San Francisco, ARIMA(0,0,5) o MA(5), mientras que el modelo del CERTE, ARIMA(0,0,1) o MA(1) presenta un 32% de variabilidad en el error.

Con los ejercicios realizados, se alcanzó el objetivo para proponer y desarrollar una metodología de Predicción de Generación de Energía Eléctrica a partir de Energía Eólica utilizando modelos autoregresivos Box and Jenkins en base a los hallazgos publicados en la literatura.

Como resultado, se determina que ARIMA es un método para analizar series de tiempo univariantes estacionarias y dentro de sus tres etapas principales se mencionan: la identificación, estimación de parámetros y verificación de los modelos. De los cuales la etapa de clasificación es considerada como la más crucial en la definición de este modelo, por lo que de acuerdo a la literatura existen otros métodos de predicción efectivos bajo los cuales se puede llevar a cabo los procedimientos específicos para la predicción sobre series de tiempo.

Como trabajo futuro, se pretende desarrollar diversos modelos con diversos tipos de datos utilizando Redes Neuronales (ANN) y Maquinas de Soporte Vectorial (SVM), Árboles de Regresión y Redes Bayesianas, que de acuerdo a la literatura son los métodos más usados en pronóstico de Generación de Energía eléctrica.

REFERENCIAS

- Aguirre, A. (1994). *Introducción al tratamiento de series temporales: aplicación a las ciencias de la salud*. España: Díaz de Santos.
- Anssen, W. B. (2013). "Pedestrian wind comfort around buildings: Comparison of wind comfort criteria based on whole-flow field data for a complex case study". *Building and Environment*, Vol. 59, pp. 547-562.
- Arriaga. (2018). L. C. EVALUACIÓN DEL RECURSO EÓLICO PARA GENERACIÓN DE ENERGÍA ELÉCTRICA EN ZONAS URBANAS.
- Bahaj, A. S. (2007). Urban energy generation: Influence of micro-wind turbine output on electricity consumption in buildings. *Energy and buildings*, 39(2), 154-165.
- Box, G. E. (1976). *Time series analysis: forecasting and control*. Oakland, California: Holden-Day.
- Brockwell, P. J. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.
- Chai, T. &. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3), 1247-1250.
- Cincunegui, M. E. (2007). Las rosas de los vientos, mundos cardinales en la búsqueda de la identidad en Finisterre de María Rosa Lojo. *Jornadas de Literatura Argentina, Instituto literario y cultural hispánico*, 217-225.
- Conway, D. (2010). The data science venn diagram. 10.
- de Saussure, L. (1923). *L'origine de la rose des vents et l'invention de la boussole*.
- Departament of Agriculture. (26 de Mayo de 2019). *Gallatin National Forest Avalanche Center*. Obtenido de Gallatin National Forest Avalanche Center: <https://www.mtavalanche.com/weather/windrose>
- Diccionario Náutico*. (24 de Mayo de 2019). Obtenido de <https://diccionario-nautico.com.ar/rosa-de-los-vientos/>
- Dodson, M. G. (January, 2005). An Historical and Applied Aerodynamic Study of the Wright Brothers' Wind Tunnel Test Program and Application to Successful Manned Flight. *ASME 2005 Fluids Engineering Division Summer Meeting*, 269-278.
- Enke, W. &. (1997). Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, 8(3), 195-207.
- Eólica, A. E. (15 de Agosto de 2017). *ae* (Asociación Empresarial Eólica). Obtenido de <https://www.aeeolica.org/es/sobre-la-eolica/la-eolica-en-el-mundo/>
- EUMETRAIN*. (15 de 04 de 2019). Obtenido de International training project sponsored by EUMETSAT to support and increase the use of meteorological satellite data: http://www.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm
- Foley, A. M. (2010). Wind power forecasting & prediction methods. In *Environment and Electrical Engineering (EEEIC), 2010 9th International Conference*. IEEE, 61-64.
- Fox, J. &. (2018). *An R companion to applied regression*. Sage Publications.
- Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association*, 57(300), 729-757.
- González-Ávila, M. E.-M.-D.-R. (2006). Potencial de aprovechamiento de la energía eólica para la generación de energía eléctrica en zonas rurales de México. *Interciencia*, 31(4), 240-245.
- Gonzalo, R. (2008). Series de Tiempo. *Universidad De Chile Facultad De Ciencias Físicas y Matemáticas*.
- Gras, J. A. (2001). *Diseños de series temporales: técnicas de análisis* (Vol. Vol. 46). Barcelona, España: Universitat Barcelona.

- Guardado, E. C. (2007). Uso de la energía eólica dentro del concepto de generación distribuida y la calidad de la energía. *Ingeniería Energética*, 28 (2), 5.
- Guoyang, W. Y. (2005). Discussion about short-term forecast of wind speed on wind farm. *Jilin Electric Power*, 181(5), 21-24.
- GWEC, G. W. (2017). *GWEC - Global Wind Report*. Obtenido de Global wind report. 2017. Brussels: GWEC.: <http://files.gwec.net/files/GWR2017.pdf>
- Hayashi, C. (1998). What is data science? Fundamental concepts and a heuristic example. In Data Science, Classification, and Related Methods . En C. Hayashi, *Data Science, Classification, and Related Methods* (págs. pp. 40-51). Springer, Tokyo.
- Héctor, I. G. (2018). Predicción de potencia eólica utilizando técnicas modernas de Inteligencia Artificial. *Ingeniería, investigación y tecnología*, 1(11), 19(4).
- Hernández Mora, J. A. (2014). Modelamiento de la irradiancia y la temperatura ambiente utilizando funciones de probabilidad. *Tecnura*, 18(39).
- Hernández-Escobedo, Q. E.-A.-F.-B. (2012). Evaluación del potencial eólico para la generación de energía eléctrica en el estado de Veracruz, México. *Dyna*, 79 (171).
- Hyndman, R. J. (2014). *Forecasting: principles and practice*. OTexts. OText.
- Hyndman, R. J. (2018). *Forecasting: principles and practice*. OTexts. O Text.
- Ibargüengoytia-González, P. H.-B.-P.-L. (2018). Predicción de potencia eólica utilizando técnicas modernas de Inteligencia Artificial. *Ingeniería, investigación y tecnología*, 19(4).
- Jiménez, L. A. (2007). Modelos avanzados para la predicción a corto plazo de la producción eléctrica en parques eólicos. *Universidad de la Rioja*.
- Kenney, J. F. (2013). *Mathematics of statistics*. D. Van Nostrand Company Inc; Toronto; Princeton; New Jersey; London; New York: Affiliated East-West Press Pvt-Ltd; New Delhi.
- Lamoureux, C. G. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2), 225-234.
- Lawan, S. M. (2014). Different models of wind speed prediction; a comprehensive review. *International Journal of Scientific and Engineering Research*, 5(1), 1760-1768.
- Lei, M. S. (2009). A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4), 915-920.
- Liu, D. A. (12 de Octubre de 2017). *IBM Analytics*. Obtenido de <http://www.researchmethods.org/DataScienceDataScientists.pdf>
- Marighetti, J. O. (2010). Análisis de técnicas de medición en estudios de confort en túnel de viento sobre estructuras de dimensiones y formas distintas con categoría de terreno diferentes. *XXXIV Jornadas Sudamericanas de Ingeniería Estructural*. San Juan, Argentina.
- México, S. d. (30 de Mayo de 2013). *Acuerdo por el que se expide la Estrategia Nacional de Cambio Climático*. Obtenido de D.O. de la Federación: http://www.dof.gob.mx/nota_detalle_popup.php?codigo=5301093
- Moragues, J. &. (2003). Energía eólica. *Instituto Argentino de la Energía "General Mosconi"*, 3.
- Moselle, B. P. (2010). *Electricidad verde: energías renovables y sistema eléctrico*. Madrid, Barcelona, Buenos Aires: Marcial Pons.
- Pandit, S. M.-M. (1983). *Time series and system analysis with applications*. New York: John Wiley & Sons, Inc.
- Pankratz, A. (2012). *Forecasting with dynamic regression models* (Vol. 935). John Wiley & Sons.
- Parra, O. J. (2007). Análisis comparativo de las técnicas de series de tiempo ARIMA y ANFIS para pronosticar tráfico Wimax. *Ingeniería*, 12(2), 73-79.
- Pearl, J. (2014). Probabilistic reasoning in intelligent systems: networks of plausible inference. *Elsevier*.
- Pineda, S. E.-F. (2017). Aplicación de modelos auto regresivos para la predicción de generación de energía eléctrica a partir de datos eólicos. *Science*, 139, 59-70. *Research in Computing*.

- Posso, F. (2002). Energía y ambiente: pasado, presente y futuro. Parte dos: Sistema energético basado en energías alternativas. *Geoenseñanza*, 54-73.
- programme, European Commision under the Intelligent Energy - Europe. (25 de 05 de 2017). *Catalogue of European Urban Wind Turbine Manufacturers*. Obtenido de Catalogue of European Urban Wind Turbine Manufacturers: http://www.urbanwind.net/pdf/CATALOGUE_V2.pdf
- Ramos Maldonado, S. I. (2013). La Naturalis Historia de Plinio el Viejo: lectura en clave humanística de un clásico. *Ágora: Estudios clásicos em debate*, 51-94.
- Sara Edith Pinzón-Pineda, G. S.-B.-A. (2017). Ciencia de los datos aplicada a la predicción de la generación de energía eléctrica en zonas urbanas. (S. M. Artificial, Ed.) *Komputer Sapiens.*, 32-36.
- Schutt, R. &. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media. *ScienceDirect*. (10 de 05 de 2018). Obtenido de Persistence Model: <https://www.sciencedirect.com/topics/engineering/persistence-model>
- sfgov.org*. (26 de 03 de 2019). Obtenido de Ciudad y Condado de San Francisco: <https://datasf.org/opendata/>
- Simon, J. M. (1999). *L'art de Ramon Llull I la Teoria Escolàstica de la Ciència*. Barcelona. *Sitio BigData*. (Mayo de 2019). Obtenido de <http://sitiobigdata.com/index.php/2018/08/27/machine-learning-metricas-regresion-mse/#>
- Sparks, P. E. (2012). "A comparison of pedestrian comfort criteria applied to a city center". *Wind Engineering 1983 3C Proceedings of the Sixth international Conference on Wind Engineering, Gold Coast, Australia, Auckland, New Sealand, March 21-25*, 123.
- The WindPower - Wind Energy Market Intelligence*. (27 de agosto de 2018). Obtenido de https://www.thewindpower.net/turbine_es_276_turbowinds_t300-28.php
- Vapnik, V. N. (1998). *Statistical Learning Theory* Wiley-Interscience.
- Vemuri, R. V. (1994). *Artificial neural networks. forecasting time series*. Los Alamitos, California: IEEE Computer Society Press.
- Villavicencio, J. (2010). *Instituto de Estadísticas de Puerto Rico*. Obtenido de <http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx>
- Zhang, Y. W. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32, 255-270.

APÉNDICES

A. Publicaciones

En esta sección se presentan las publicaciones derivadas del presente trabajo de investigación. El siguiente artículo publicado en la revista *Research in Computing Science*.

ISSN 1870-4069

Aplicación de modelos auto regresivos para la predicción de generación de energía eléctrica a partir de datos eólicos

Sara Edith Pinzón Pineda¹, José Alberto Hernández Aguilar²,
Gustavo Arroyo-Figueroa¹

¹ Instituto Nacional de Electricidad y Energías Limpias, Ciudad de México, México

² Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México

sepinzon@iie.org.mx, jose_hernandez@uaem.mx,
garroyo@iie.org.mx

Resumen. Se discuten los modelos estadísticos auto regresivos clásicos *Box and Jenkins* utilizados en problemas de predicción, y se analiza su aplicación en el sector energético. Posteriormente, se propone una metodología basada en la aplicación de estos modelos para la predicción de generación de energía eléctrica a partir de datos obtenidos en un parque eólico de la Ventosa, Oaxaca. Finalmente, se presenta el modelo y los resultados preliminares obtenidos destacando, que se alcanza un $R^2 = 0.974$ y un $RMSE = 18.862$.

Palabras clave: modelos auto regresivos, *Box and Jenkins*, predicción, energía eléctrica, energía eólica, datos eólicos.

Application of Autoregressive Models for the Prediction of Electric Power Generation from Wind

Abstract. We discuss the classic auto regressive statistical models *Box and Jenkins* used in prediction problems, and their application is analyzed in the energy sector. Subsequently, we propose a methodology based on the application of these models for the prediction of electric power generation, from data obtained at a wind farm in Ventosa, Oaxaca. Finally, the model and the preliminary results obtained shows an $R^2 = 0.974$ and an $RMSE = 18.862$.

Keywords: autoregressive models, *Box and Jenkins*, prediction or forecasting, electric power, wind power, wind data, wind power prediction.

1. Introducción

Actualmente se está alcanzando el límite de capacidad de carga dentro de los ecosistemas para regenerarse de la contaminación producida por la actividad humana. De acuerdo a [1]: "Un tercio del total de la contaminación generada a nivel mundial

ARTÍCULO EN REVISIÓN
Ciencia de datos

La ciencia de datos es un campo interdisciplinario que utiliza el método científico, procesos y sistemas para extraer conocimiento o

identificar datos faltantes, frecuencia de muestreo y de ser necesario interpolación. Finalmente, se realiza la limpieza de los datos para continuar con la fase de análisis

Ciencia de datos aplicada a la predicción de la generación de energía eléctrica en zonas urbanas

por Sara Edith Pinzón Pineda, Guillermo Santamaria-Bonfil, José Alberto Hernández Aguilar

información a partir de conjuntos de datos, estructurados o no, de manera similar a la minería de datos [1].

De acuerdo a Hayashi [2] el objetivo de la ciencia de datos es revelar la estructura oculta de fenómenos naturales, humanos y sociales desde un punto de vista diferente al utilizado por los métodos tradicionales. Este enfoque requiere nuevas formas de pensar multidimensionales, dinámicas y flexibles.

Una de las técnicas de la ciencia de datos es la Analítica de Datos (AD), la cual trata de identificar patrones en conjuntos de datos utilizando procesos algorítmicos o procesos mecánicos. Actualmente AD ha sido utilizada para predecir la producción y el consumo de energía eléctrica generada a partir de fuentes de energías renovables, incluyendo la energía eólica [3].

Etapas de la Ciencia de Datos

En la Figura 1 se muestran las etapas del proceso de la ciencia de datos [4]:



Figura 1. Etapas del proceso de Ciencia de datos.

- 1. Determinar el problema en el mundo real.** Se requiere identificar con claridad el problema a tratar, para determinar el tipo de fuente de datos que se utilizará en la búsqueda de las mejores alternativas.
- 2. Recopilación, Procesamiento, Limpieza y Análisis exploratorio de los datos.** Partiendo del problema bien definido, en esta etapa se deberá contar una fuente de datos confiable. Se revisan factores que pudieran afectar los resultados como:

exploratorio, que es la fase donde se determinan los datos más adecuados para la selección de un buen modelo.

- 3. Selección del modelo.** En esta fase se elige un método para generar un modelo entrenado a partir de los datos disponibles. Una vez realizado el entrenamiento del modelo, este será capaz de realizar pronósticos a partir de nuevos conjuntos de datos.
- 4. Conclusiones.** Una vez elegido el modelo, se deben mostrar las conclusiones a las que se llegó como resultado del análisis previo. Esta fase servirá de referencia para la toma de decisiones.
- 5. Construir un producto de datos.** Para finalizar, en esta etapa se determinará si la solución podría garantizar con éxito la construcción de algún producto final.

A continuación, se presenta un caso del mundo real, el cual aborda el problema de la predicción de generación de energía eléctrica en zonas urbanas, donde son aplicadas las diferentes etapas de la Ciencia de los Datos.

Problema del Mundo real

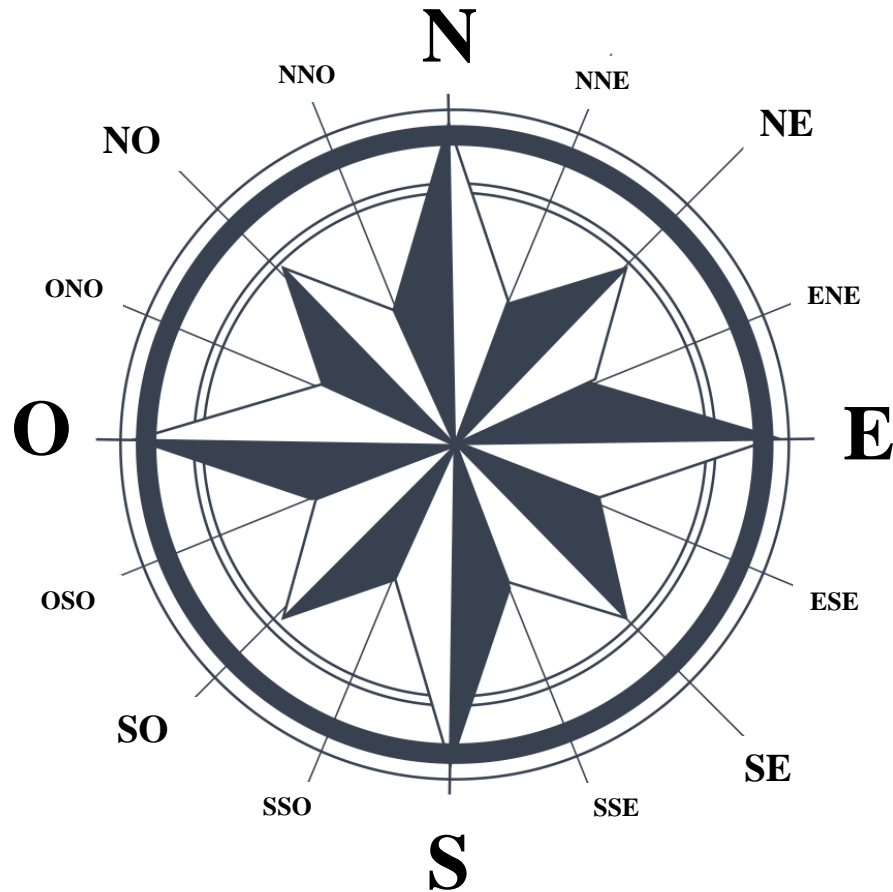
En la actualidad se está alcanzando el límite de capacidad de carga dentro de los ecosistemas para regenerarse de la contaminación producida por la actividad humana. Palit & Popovic [5] señalan que "Un tercio del total de la contaminación generada a nivel mundial proviene del proceso de producción de electricidad". Para apoyar la reducción de los niveles contaminantes en el sector eléctrico, año con año se invierte más en el desarrollo de fuentes de energías renovables [6]. Es por ello que el estudio de la energía eólica ha tomado gran auge debido a que utiliza el viento como recurso renovable, sin generar contaminación en aire, agua o suelo; con escasa ocupación exclusiva del terreno y con leve impacto en el medio ambiente [5].

B. Rosa de los vientos

Una rosa de los vientos o rosa náutica es una figura en forma de círculo que tiene trazados alrededor los rumbos en los que se divide la vuelta del horizonte. La rosa de los vientos ha sido crucial en la historia de la navegación, su invención se le atribuye al mallorquín Ramon Llull (Simon, 1999), aunque su descripción pareciera tener referencia en (Ramos Maldonado, 2013) ya que gracias a esta los navegantes que cruzaban los mares eran capaces de reconocer de dónde provenía el viento, norte, sur, este, o este, etc. (Cincunegui, 2007).

Las orientaciones fundamentales de la rosa de los vientos son cuatro: Norte, Sur, Este y Oeste, a partir de éstos el horizonte queda dividido en cuatro partes de 90° cada una. La línea que une norte y sur se llama meridiana o línea norte-sur, mientras que la que une los otros puntos se llama línea este-oeste (de Saussure, 1923). De las bisectrices de cada uno de los ángulos rectos formados por las líneas meridiana y este-oeste resultan ocho nuevas orientaciones o rumbos llamados laterales que son: noreste, sureste, suroeste y noroeste. Si volvemos a dividir los rumbos laterales y los principales tendremos ocho nuevos rumbos llamados colaterales que son: nor-noreste, este-noreste, este-sureste, sur-sureste, sur-suroeste, oeste-suroeste, oeste-noroeste y norte-noroeste.

En general, la rosa de los vientos es representada con las anteriores orientaciones, sin embargo, nuevos rumbos pueden señalarse con el mismo procedimiento, pero por lo general se emplean a partir de las divisiones señaladas los ángulos, partiendo del Norte (0°) con dirección positiva siguiendo las manecillas de reloj, es decir hacia el Este (Diccionario Náutico, 2019). Un ejemplo de la rosa de los vientos se muestra a continuación (Cincunegui, 2007):



Para el desarrollo de la rosa del viento de San Francisco y CERTE se utilizó el siguiente código:

```

thm <- theme_bw() +
  theme(axis.text.x = element_text(size=8, face = "plain"),
        axis.text.y = element_text(size=8, face = "plain"),
        axis.title.x = element_blank(),
        axis.title.y = element_text(size=8, face = "plain", hjust = 0.9, vjust = 1.3),
        panel.border = element_blank(),
        panel.grid = element_blank())

summary(dir.binned)

# initialise the plot
plt.dirrose <- ggplot() +
  # since the plot background is blank we'll add a series of horizontal lines, at 5000 count intervals, up to 25000
  # Contadores de 5000 hasta 25000
  geom_hline(yintercept = seq(0, 150000, by = 20000), colour = "grey60", size = 0.3) +
  # Now we add a darker horizontal line as the top border at 30000.
  geom_hline(yintercept = 150000, colour = "black", size = 0.5) +
  # We want 12 vertical lines representing the centers of the 30° ranges.
  geom_vline(xintercept = c(seq(1,12,1)), colour = "grey60", size = 0.3) +
  # on top of everything we place the histogram bars.
  geom_bar(data = data.df, aes(x = dir.binned), width = 1, colour="chartreuse", size = 0.3, alpha=0.5) +
  # Add the x-axis labels
  scale_x_discrete(drop = FALSE, labels = c(0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330)) +
  # Add the y-axis labels
  scale_y_continuous(limits = c(0, 150000), expand = c(0, 0),
                    breaks = seq(0, 150000, by = 20000),
                    #labels = seq(0, 300, by = 5)) +
                    labels = seq(0, 150000, by = 20000)) +
  # Add the axis titles
  labs(x = 'outward step bearing (°)', y = 'Count of outward steps (x10³)') +
  # If you only use the plot code up till here you will get a histogram.
  # the next line wraps the histogram into a windrose
  coord_polar(start = -(deg/2)*(pi/180)) +
  # apply theme
  thm

plt.dirrose

```

C. Cálculo de las 75 combinaciones para los modelos ARIMA(p,d,q) para San Francisco y CERTE

Función de las posibles combinaciones de los modelos ARIMA $p = \{0, \dots, 5\}$; $d = \{0, 1\}$; $q = \{0, \dots, 5\}$ desarrollado con el lenguaje de programación de R.

```
for(j in 1:75){ # Se realiza el cálculo para los 75 modelos

  ytrueT <- numeric()
  yhatT <- numeric()
  constT <- 10000
  idx1T <- (length(st_datos80_50_2)-constT)+1

  for(i in 1:10000)
  {
    if(i==1)
    {
      tmp.trainT <- st_datos80_50_1[(idx1T):length(st_datos80_50_1)]
      initT <- i
      end1T <- p
      tmp.localT <- tmp.validate[initT:end1T]
      ytrueT <- c(ytrueT,tmp.validate[end1T+1])
    }else
    {
      initT <- i
      end1T <- (i+p)-1
      tmp.localT <- tmp.validate[initT:end1T]
      ytrueT <- c(ytrueT,tmp.validate[end1T+1])

      if(i < constT){
        tmp.trainT <- append(st_datos80_50_1[seq(idx1T+i,length(st_datos80_50_1), by = 1)],st_datos80_50_1[1:i])
      }
    }
    arimaFit1 <- tryCatch(arima(tmp.trainT, order = c(reer$p[j],reer$d[j],reer$q[j]), include.mean = T),
      error=function( err ) FALSE,
      warning=function( err ) FALSE )

    if( !is.logical( arimaFit1 ) ) {
      tmp.fitT <- arima(tmp.trainT, order = c(reer$p[j],reer$d[j],reer$q[j]), include.mean = T)
      tmp.yhatT <- predict(tmp.fitT, n.ahead = 1, newdata = tmp.localT)
      yhatT <- c(yhatT,tmp.yhatT$pred[1])
      mod.valormse[j]<- paste(reer$p[j],reer$d[j],reer$q[j])
    } else {
```

D. Ajuste del modelo y Predicción

Se desarrolló un modelo que recorre la serie de tiempo y predice sobre los parámetros estimados, como se muestra a continuación:

```
for(i in 1:10000)
{
  if(i==1)
  {
    tmp.trainT <- st_datos80_50_1[(idx1T):length(st_datos80_50_1)]
    initT <- i
    end1T <- p
    tmp.localT <- tmp.validate[initT:end1T]
    ytrueT <- c(ytrueT,tmp.validate[end1T+1])
  }else
  {
    initT <- i
    end1T <- (i+p)-1
    tmp.localT <- tmp.validate[initT:end1T]
    ytrueT <- c(ytrueT,tmp.validate[end1T+1])

    if(i < constT){
      tmp.trainT <- append(st_datos80_50_1[seq(idx1T+i,length(st_datos80_50_1), by = 1)],st_datos80_50_1[1:i])
    }
  }
  arimaFit1 <- tryCatch(arima(tmp.trainT, order = c(reer$p[j],reer$d[j],reer$q[j]), include.mean = T),
    error=function( err ) FALSE,
    warning=function( err ) FALSE )

  if( !is.logical( arimaFit1 ) ) {
    tmp.fitT <- arima(tmp.trainT, order = c(reer$p[j],reer$d[j],reer$q[j]), include.mean = T)
    tmp.yhatT <- predict(tmp.fitT, n.ahead = 288, newdata = tmp.localT)
    #tmp.yhatT <- forecast(tmp.fitT,h=constT)
    yhatT <- c(yhatT,tmp.yhatT$pred[1])
    mod.valorrmse[j]<- paste(reer$p[j],reer$d[j],reer$q[j])
  } else {
    tmp.fitT <- arima(tmp.trainT, order = c(p,d,q), include.mean = T)
    tmp.yhatT <- predict(tmp.fitT, n.ahead = 288, newdata = tmp.localT)
```



Cuernavaca, Morelos a 14 de mayo del 2019.

DRA. LAURA PATRICIA CEBALLOS GILES
DIRECTORA DE LA FCAeI
PRESENTE

En mi carácter de revisor de Tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado **Maestría en Ciencias Computacionales y Tecnologías de la Información**, de la estudiante **Sara Edith Pinzón Pineda**, con matrícula **90201501**, con el título **PREDICCIÓN DE GENERACIÓN DE ENERGÍA ELÉCTRICA CON UN DÍA DE ANTICIPO MEDIANTE MODELOS DE APRENDIZAJE SUPERVISADO A PARTIR DE DATOS EÓLICOS** por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, concluyo que el trabajo se caracteriza por el establecimiento de objetivos académicos pertinentes y una metodología adecuada para su logro. Además construye una estructura coherente y bien documentada, por lo cual considero que los resultados obtenidos contribuyen al conocimiento del tema tratado.

Con base en los argumentos precedentes me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que la estudiante continúe con los trámites que esta Secretaría de Investigación y Posgrado tenga establecidos para obtener el grado mencionado.

Atentamente

Dr. Martín Heriberto Cruz Rosales
Profesor - Investigador
Facultad de Contaduría, Administración e Informática
Universidad Autónoma del Estado de Morelos



Cuernavaca, Morelos a 14 de mayo del 2019.

DRA. LAURA PATRICIA CEBALLOS GILES
DIRECTORA DE LA FCAeI
PRESENTE

En mi carácter de revisor de Tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado **Maestría en Ciencias Computacionales y Tecnologías de la Información**, de la estudiante **Sara Edith Pinzón Pineda**, con matrícula **90201501**, con el título **PREDICCIÓN DE GENERACIÓN DE ENERGÍA ELÉCTRICA CON UN DÍA DE ANTICIPO MEDIANTE MODELOS DE APRENDIZAJE SUPERVISADO A PARTIR DE DATOS EÓLICOS** por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, concluyo que el trabajo se caracteriza por el establecimiento de objetivos académicos pertinentes y una metodología adecuada para su logro. Además construye una estructura coherente y bien documentada, por lo cual considero que los resultados obtenidos contribuyen al conocimiento del tema tratado.

Con base en los argumentos precedentes me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que la estudiante continúe con los trámites que esta Secretaría de Investigación y Posgrado tenga establecidos para obtener el grado mencionado.

Atentamente

Dr. Guillermo Santamaria Bonfil
Profesor - Investigador
Gerencia de Tecnologías de la Información
Instituto Nacional de Electricidad y Energías Limpias



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA

Cuernavaca, Morelos a 14 de mayo del 2019.

DRA. LAURA PATRICIA CEBALLOS GILES
DIRECTORA DE LA FCAeI
PRESENTE

En mi carácter de revisor de Tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado **Maestría en Ciencias Computacionales y Tecnologías de la Información**, de la estudiante **Sara Edith Pinzón Pineda**, con matrícula **90201501**, con el título **PREDICCIÓN DE GENERACIÓN DE ENERGÍA ELÉCTRICA CON UN DÍA DE ANTICIPO MEDIANTE MODELOS DE APRENDIZAJE SUPERVISADO A PARTIR DE DATOS EÓLICOS** por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, concluyo que el trabajo se caracteriza por el establecimiento de objetivos académicos pertinentes y una metodología adecuada para su logro. Además construye una estructura coherente y bien documentada, por lo cual considero que los resultados obtenidos contribuyen al conocimiento del tema tratado.

Con base en los argumentos precedentes me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que la estudiante continúe con los trámites que esta Secretaría de Investigación y Posgrado tenga establecidos para obtener el grado mencionado.

Atentamente

Dr. José Alberto Hernández Aguilar
Profesor - Investigador
Facultad de Contaduría, Administración e Informática
Universidad Autónoma del Estado de Morelos



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA

Cuernavaca, Morelos a 14 de mayo del 2019.

DRA. LAURA PATRICIA CEBALLOS GILES
DIRECTORA DE LA FCAeI
PRESENTE

En mi carácter de revisor de Tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado **Maestría en Ciencias Computacionales y Tecnologías de la Información**, de la estudiante **Sara Edith Pinzón Pineda**, con matrícula **90201501**, con el título **PREDICCIÓN DE GENERACIÓN DE ENERGÍA ELÉCTRICA CON UN DÍA DE ANTICIPO MEDIANTE MODELOS DE APRENDIZAJE SUPERVISADO A PARTIR DE DATOS EÓLICOS** por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, concluyo que el trabajo se caracteriza por el establecimiento de objetivos académicos pertinentes y una metodología adecuada para su logro. Además construye una estructura coherente y bien documentada, por lo cual considero que los resultados obtenidos contribuyen al conocimiento del tema tratado.

Con base en los argumentos precedentes me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que la estudiante continúe con los trámites que esta Secretaría de Investigación y Posgrado tenga establecidos para obtener el grado mencionado.

Atentamente

Dr. Federico Alonso Pecina
Profesor - Investigador
Facultad de Contaduría, Administración e Informática
Universidad Autónoma del Estado de Morelos



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA

Cuernavaca, Morelos a 14 de mayo del 2019.

DRA. LAURA PATRICIA CEBALLOS GILES
DIRECTORA DE LA FCAeI
PRESENTE

En mi carácter de revisor de Tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado **Maestría en Ciencias Computacionales y Tecnologías de la Información**, de la estudiante **Sara Edith Pinzón Pineda**, con matrícula **90201501**, con el título **PREDICCIÓN DE GENERACIÓN DE ENERGÍA ELÉCTRICA CON UN DÍA DE ANTICIPO MEDIANTE MODELOS DE APRENDIZAJE SUPERVISADO A PARTIR DE DATOS EÓLICOS** por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, concluyo que el trabajo se caracteriza por el establecimiento de objetivos académicos pertinentes y una metodología adecuada para su logro. Además construye una estructura coherente y bien documentada, por lo cual considero que los resultados obtenidos contribuyen al conocimiento del tema tratado.

Con base en los argumentos precedentes me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que la estudiante continúe con los trámites que esta Secretaría de Investigación y Posgrado tenga establecidos para obtener el grado mencionado.

Atentamente

Dr. José Crispín Zavala Díaz
Profesor - Investigador
Facultad de Contaduría, Administración e Informática
Universidad Autónoma del Estado de Morelos