

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

proviene del proceso de producción de electricidad”. Por lo que, es necesario el desarrollo de fuentes de energías renovables.

De acuerdo a [2]: “La energía eólica tiene su origen en la energía solar. El calentamiento de masas de aire como consecuencia de la radiación solar, contribuye a la aparición de corrientes ascendentes, siendo el espacio que liberan ocupado por otras masas adyacentes de aire más frío. La desviación del viento es proporcional a la velocidad del mismo: A mayor velocidad mayor desviación con respecto a la dirección inicial”.

Al viento se le considera un recurso energético y también como suministro. Su característica particular lo describe como una fuente con importantes variaciones temporales, tanto en su superficie como en su altura, al que se agrega una componente aleatoria que afecta en buena medida su variación total [2].

Por este motivo, se considera importante predecir la velocidad del viento para la planificación de la cantidad de energía que se produce en el tiempo [1]. De acuerdo a lo anterior, se debe evaluar y planificar la producción eólica debido a la incorporación de parques eólicos en el mercado eléctrico, por lo que es necesario realizar ofertas de energía para plazos que oscilan entre 24 y 36 horas, lo que requiere un conocimiento preciso de las condiciones de funcionamiento de los parques durante esos periodos.

La dificultad de estas predicciones radica en el grado de detalle y la cantidad de variables a analizar, y de la necesidad de contar con horizontes de tiempo [2].

Debido a lo anterior se han realizado diversas investigaciones de predicción de generación del viento, basados en modelos matemáticos, estadísticos y de inteligencia artificial [3], [4].

Problema de investigación

Aplicar la Metodología Box-Jenkins para la predicción de la generación de energía eléctrica a partir de datos obtenidos de un parque eólico localizado en la Ventosa, Oaxaca, México.

Para este propósito, esta investigación se ha dividido en cuatro secciones: en la primera sección, se describe la Introducción y se señala el problema de investigación, en la segunda sección, se presenta la metodología empleada en la cual se utilizan como base los modelos auto regresivos clásicos basados en la metodología Box-Jenkins utilizando el software SPSS v.22. En la tercera sección, se presentan y discuten los resultados obtenidos. Finalmente, se presentan las conclusiones y trabajos futuros.

Los datos, fueron obtenidos de 3 periodos históricos proporcionado por el Centro Regional de Tecnología Eólica en el estado de Oaxaca, que es considerado como el quinto más grande del país, está ubicado en la parte sur de México, con una superficie de aproximadamente 95.364 km². Cuenta con una orografía de zonas montañosas, mesetas planas, valles y costas, con una gran variedad de climas, tropical en toda la costa y templado en el interior. La lluvia aparece generalmente a finales de abril y continúa hasta finales de octubre. La temperatura media oscila entre 26 y 28 1C en toda la costa [5]. Los datos con los que se realiza el análisis se tomaron de una muestra de 3 periodos históricos de una torre anemométrica a 80m. Los periodos específicos son del 2007 (enero a diciembre), 2012 (febrero a diciembre) 2013 (enero) y 2015 (enero a diciembre) del CERTE en el estado de Oaxaca. Esta zona se caracteriza por la complejidad de su terreno que es: montañosa en el Norte a 5 Km y plano en el Este, Sur y Oeste.

2. Metodología

Para llevar a cabo el desarrollo de esta metodología, se utilizaron datos del Centro Regional de Tecnología Eólica – CERTE, localizado en la localidad de la Ventosa del Istmo de Tehuantepec del estado de Oaxaca, que es considerado como el quinto más grande del país, está ubicado en la parte sur de México, con una superficie de aproximadamente 95.364 km². Cuenta con una orografía de zonas montañosas, mesetas planas, valles y costas, con una gran variedad de climas, tropical en toda la costa y templado en el interior. La lluvia aparece generalmente a finales de abril y continúa hasta finales de octubre. La temperatura media oscila entre 26 y 28 °C en toda la costa [5].

El parque eólico CERTE de acuerdo a [6]: fue construido con el apoyo económico del Fondo Mundial para el Medio Ambiente (GEF por sus siglas en inglés) a través del Programa de las Naciones Unidas para el Desarrollo (PNUD), como parte de las metas del proyecto “Plan de Acción para Eliminar Barreras para el Desarrollo de la Generación Eólica eléctrica en México”. El 1 de julio de 2010 al haber cumplido con los requisitos y pruebas establecidas por la Comisión Federal de Electricidad (CFE), entró en operación normal el CERTE perteneciente de la Gerencia de Energías No Convencionales del anterior Instituto de Investigaciones Eléctricas - IIE que actualmente fue renombrado Instituto Nacional de Electricidad y Energías Limpias – INEEL.

Es un centro de investigación y desarrollo tecnológico que tiene como objetivos además de buscar dar solución a los problemas de adecuación, instalación y mantenimiento, el desarrollo de sus propias tecnologías para fortalecer y potencializar el sector eólico nacional.

A continuación se describen los modelos auto regresivos clásicos, posteriormente se presentan los pasos empleados para el procesamiento de los datos obtenidos del parque eólico para la predicción de la generación de la energía eléctrica.

2.1. Modelos para predicción o pronóstico

Los métodos para predecir la generación de energía eólica pueden ser categorizados en métodos físicos, métodos estadísticos, métodos basados en redes neuronales, y métodos híbridos [7].

2.2. Modelos auto regresivos clásicos para predicción o pronóstico

En 1970, Box and Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental [1].

De acuerdo a [8]:

“Este modelo implica identificar un proceso ARIMA adecuado realizando el ajuste de los datos y luego realizando la predicción. El procedimiento original del modelo Box-Jenkins implicó un proceso iterativo de tres etapas:

1. Identificación o selección del modelo,
2. Estimación de los parámetros,

3. Verificación del modelo.

Explicaciones recientes del proceso (por ejemplo, Makridakis, Wheelwright y Hyndman, 1998) [9] añaden una etapa de pre procesamiento de datos y una representación final de la aplicación del modelo o pronóstico.

El modelado de Box-Jenkins implica identificar un proceso ARIMA adecuado, ajustándolo a los datos, y luego utilizando el modelo de previsión. Una de las características atractivas del enfoque de Box-Jenkins para la predicción es que los procesos ARIMA son una clase muy rica de posibles modelos y suele ser posible encontrar un proceso que proporcione una descripción adecuada de los datos”.

Pasos para el desarrollo del modelo según [8] [9] se deberá efectuar una serie de etapas que se describen a continuación:

Preparación de los datos (Pre procesamiento de los datos)

En esta etapa, se realiza la limpieza de los datos y se puede hacer un análisis exploratorio de los mismos, también de manera gráfica se puede observar si la media y la varianza son constantes en el tiempo para determinar a grandes rasgos si la serie de tiempo es estacionaria.

En caso de no ser estacionaria, se pueden realizar transformaciones o diferenciaciones de los datos. Las transformaciones de los datos pueden ser a través de raíces cuadradas o logarítmicas y se utilizan para estabilizar la varianza en las series de tiempo.

La diferenciación de los datos se realiza tomando observaciones consecutivas o de años anteriores, generalmente para casos donde la varianza no permanece constante en el tiempo.

Box-Jenkins de acuerdo a [9] y [10] propone tres pasos para el desarrollo del modelo:

- 1. Identificación y selección del modelo.** Este modelo utiliza varios gráficos basados en transformaciones y diferenciaciones para tratar de identificar posibles procesos ARIMA los cuales podrían proporcionar un buen ajuste a los datos de los desarrollos posteriores. Existen algunas herramientas de selección de modelos como el Criterio de Información de Akaike o los criterios ACF y PACF.
- 2. Estimación de parámetros.** Se deben encontrar los valores para los coeficientes del modelo que proporcionen un mejor ajuste de los datos.
- 3. Comprobación del modelo mediante ensayo.** Consiste en probar las suposiciones del modelo para identificar cualquier área donde el modelo es inadecuado. Si el modelo se considera inadecuado, es necesario volver al Paso 2 y tratar de identificar un modelo mejor.

Realización del Pronóstico.

Entre las técnicas univariantes existen algunas sencillas, como por ejemplo los modelos autoregresivos de primer orden, o los modelos de tendencia lineal o exponencial, mientras que otras técnicas resultan más complejas. [11]

Existen diversas técnicas para hacer pronóstico de series de tiempo de las cuales algunos autores afirman que la metodología Box and Jenkins es muy efectiva con el pronóstico a corto plazo. [11]

En la figura 1 se detalla de manera general el comportamiento del modelo:

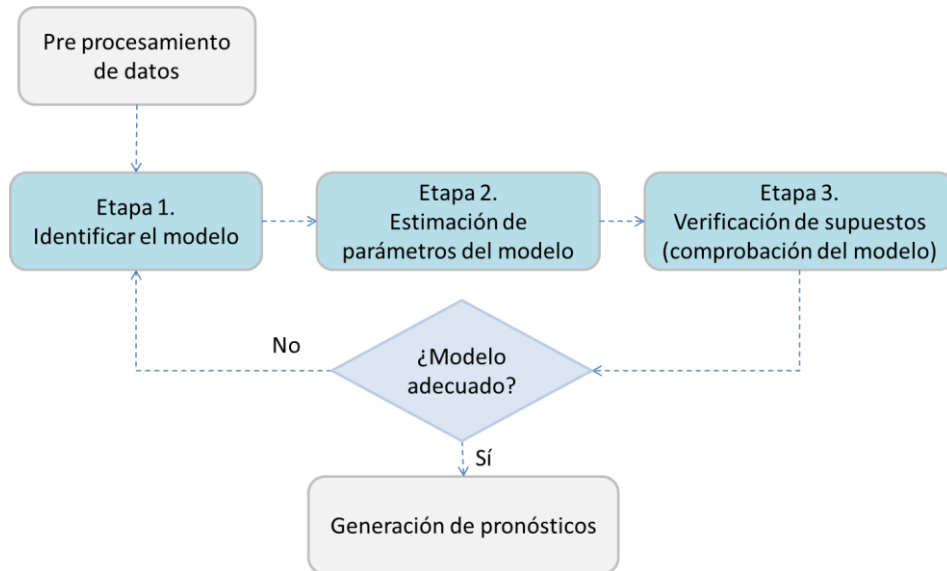


Fig. 1. Etapas de la Metodología Box - Jenkins.

En la Figura 2 se muestran los modelos clásicos Box – Jenkins para la predicción del viento:

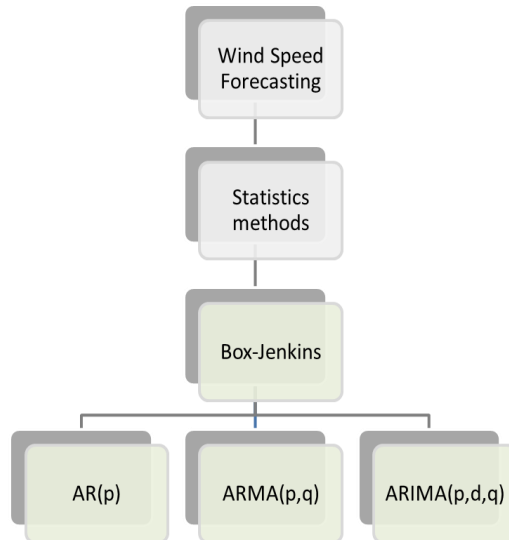


Fig. 2. Modelos clásicos Box-Jenkins para predicción o pronóstico del viento.

A continuación, de acuerdo a [12] y [8] se describen las ecuaciones los modelos Estadísticos Box and Jenkins:

Modelo Auto regresivo de primer orden AR(p). Sus valores expresan la dependencia de X_t en sus valores pasados. Definido mediante la siguiente ecuación:

$$X_t = \Phi X_{t-1} + \varepsilon_t, \quad (1)$$

dónde:

- X_t = Valores de correlación de la serie de tiempo,
- X_{t-1} = Valor pasado (Variable determinística fija que al no ser fija se convierte en modelo dinámico),
- Φ = Constante definida por un tiempo,
- ε_t = Residuo desconocido.

Modelo Auto regresivo de Medias Móviles ARMA(p,q). Es un modelo donde se involucran un término de orden auto regresivo y un término de media móvil. Es una combinación entre AR y MA (Medias móviles). Su ecuación se define por:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \quad (2)$$

dónde:

- X_t = Suma de la parte independiente no correlacionadas entre sí y sus pasados inmediatos anteriores,
- Φ_1, Φ_2 = Constantes definidas en el tiempo y su valor pasado a esa constante,
- X_{t-2} = Valor precedente al X_{t-1} ,
- $\theta_1 \varepsilon_{t-1}$ = Residuo de orden a_{t-1} (Residuo que se puede conocer),
- ε_t = Residuo desconocido.

Modelo Auto regresivo Integrado de Medias Móviles ARIMA(p,d,q) - No estacional. Si combinamos la diferenciación con auto regresión y un modelo de media móvil, obtenemos un modelo ARIMA no estacional. ARIMA es un acrónimo para el modelo *AutoRegressive Integrated Moving Average* ("integración" en este contexto es el inverso de la diferenciación). El modelo completo puede escribirse como [9]:

$$Y'_t = C + \Phi_1 Y'_{t-1} + \dots + \Phi_p Y'_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t, \quad (3)$$

dónde Y'_t es la serie diferenciada (Puede haber sido diferenciada más de una vez). Los "Predictores" en el lado derecho incluyen tanto los valores de retardo de Y_t y los errores de retardo. Según [9] y [13] el modelo es llamado ARIMA(p,d,q) cuando:

- p = El orden de la parte auto regresiva (El número de términos auto regresivos).
- d = Grado de la diferenciación involucrada (Número de diferencias no estacionales necesarias para generar la estacionariedad).
- q = El orden de la parte de media móvil (El número de errores de predicción de retardo en la ecuación de predicción).

3. Diseño experimental

A continuación se describen los pasos utilizados de acuerdo a la metodología para analizar y procesar los registros obtenidos de tres periodos completos (del 2007 al 2010) del parque eólico.

3.1. Pre procesamiento de los datos

El método propuesto está basado en los trabajos de [14] [15], Del primer utiliza el enfoque para el análisis de los datos y pronóstico de fenómenos físicos mediante modelos auto regresivos y del segundo se toma la metodología KDD (Knowledge Discovery from Data – Descubrimiento de Conocimiento de los Datos) para la obtención, pre procesamiento, limpieza y almacenamiento de los datos, y visualización e identificación de patrones. Cabe señalar que para esta investigación se utilizó el software SPSS Versión 22 de IBM y R en su versión 3.2.

En la figura 1. Se muestra la curva de potencia del aerogenerador del modelo T300-28 del fabricante Turbowinds (Bélgica), de potencia máxima de 300 kw [16], modelo que se encuentra ubicado dentro del parque eólico y bajo la cual se realizó la interpolación de datos para simular la potencia generada.

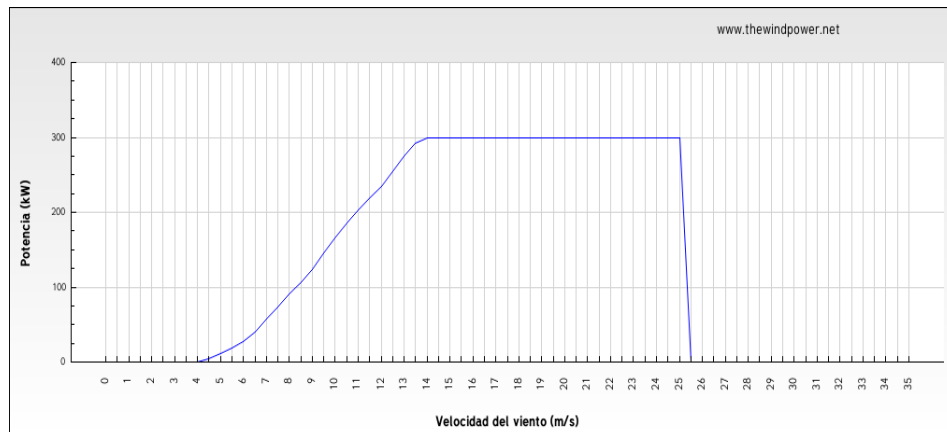


Fig. 3. Curva de potencia del modelo T300-28.

3.2. Identificación del modelo

1. Selección de la muestra

La muestra consistió de 157824 registros correspondientes a las lecturas diarias tomadas con una frecuencia de 10 minutos durante el periodo.

2. Visualización de los datos

En esta etapa se analiza el comportamiento y tendencias de los datos, se evalúa si existen comportamientos estacionarios dentro de la serie de tiempo.

Para graficar la Información se utilizó el módulo de predicciones en su opción gráficos de secuencia de SPSS versión 22 y R en su versión 3.2 [17].

Los resultados de este análisis preliminar se muestran en las Figuras 4, 5 y 6. En la figura 5 se muestra la dirección del viento medida a partir de los datos eólicos. En la figura 6 se muestra la rosa del viento en donde se aprecia que el viento proviene preponderantemente del norte y noroeste. De acuerdo al análisis exploratorio de los datos, se concluye que la serie de tiempo es estacionaria ya su tendencia se mantiene estable (Priestley, 1988).

3.3. Identificación de los parámetros del modelo

Para la identificación de los parámetros del modelo se utilizó el módulo de Predicción de SPSS versión 22 a partir del cual se creó el modelo utilizando la opción: Crear Modelos de la figura 7.

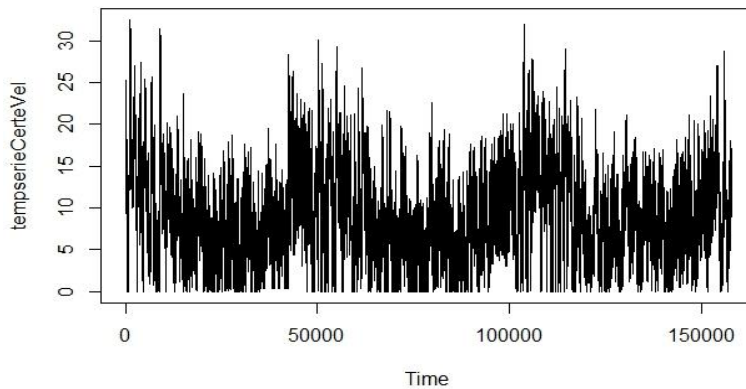


Fig. 4. Grafica preliminar de los datos a partir de la velocidad del viento a 80 m (fuente propia) de los datos procesados en R versión 3.2.

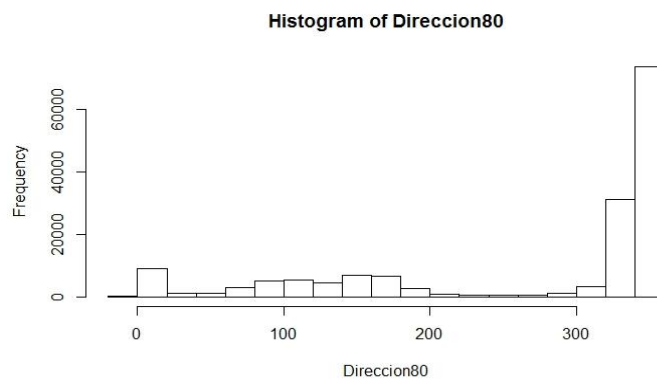


Fig. 5. Histograma de los datos de la dirección del viento a 80 m (fuente propia) procesados en R Versión 3.2.

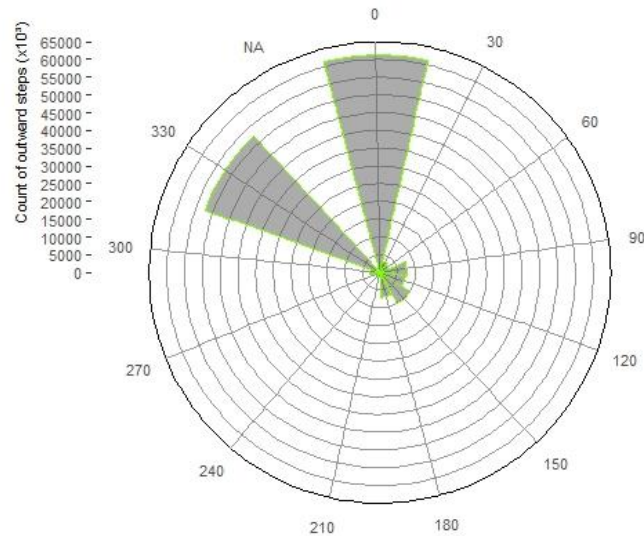


Fig. 6. Rosa de viento a partir de los datos a una altura de 80 m (fuente propia) de la dirección en R Versión 3.2.

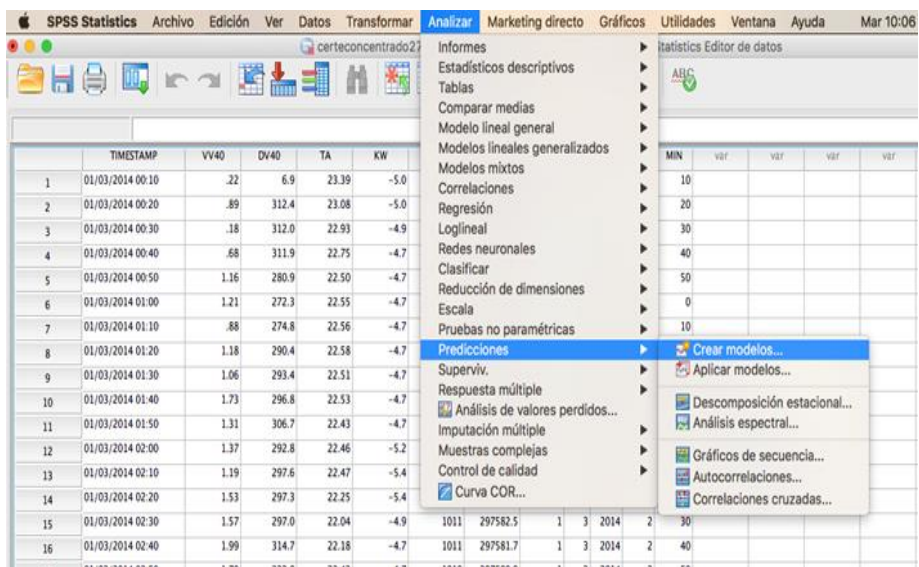


Fig. 7. Módulo para la creación de modelos predictivos.

El generador de modelos que se utilizó fue el experto, el cual permite seleccionar la variable dependiente y las variables independientes o predictoras. Se probó con las variables existentes Velocidad del Viento (VV80), dirección del viento (DV-80), y se encontró que la velocidad del viento es la principal variable predictora.

4. Resultados

4.1. Verificación o comprobación del modelo

Para la prueba del modelo se seleccionaron tres muestras, la primera de 50000 al 32%, la segunda 126580 que corresponde al 81% y la tercera de 154824 del 100% de los datos. Para ello se utilizó el módulo “Aplicar modelos”, disponible en el menú de Analizar -> Predicciones -> Aplicar modelos.

Los resultados se presentan en la Tabla 2, en él se resumen la R^2 estacionaria y el RMSE.

Tabla 1. Comparativo para los diferentes modelos de predicción generados.

Experimento	n=50000	n=126580	n=154824
Variable predictora	Velocidad del Viento y Dirección del Viento	Velocidad del viento	Velocidad del viento
R cuadrado	0.966	0.974	0.972
RMSE	21.654	18.862	19.792
Modelo generado	ARIMA(5,1,5)	ARIMA(3,1,3)	ARIMA(3,1,5)

Con el comparativo anterior las variables más adecuadas para predecir la generación de energía eléctrica es la velocidad del viento de acuerdo a la mayor R cuadrada y la menor RMSE.

Tabla 2. Descripción del mejor.

Tipo de modelo		
ID modelo	W Modelo_1	ARIMA(3,1,3)

El modelo anterior nos indica que para predecir la generación de energía eléctrica a partir de la velocidad del viento para nuestros datos es un modelo ARIMA(p,d,q) con:

p = 3,

d = 1,

q = 3.

5. Conclusiones y trabajo a futuro

Los resultados demuestran la necesidad de tener una base de datos confiable para poder predecir la generación de energía eléctrica a partir de fuentes eólicas. Es indispensable que esta base de datos se valide por expertos en el área, en el sentido de que los datos representen la realidad y tengan sentido.

A partir de múltiples experimentos se pudo obtener un primer modelo predictivo ARIMA (3,1,3), con 1 sola variable predictora que corresponde a la velocidad del

viento. Cabe mencionar que aún se requiere la prueba del modelo y la validación del mismo, lo cual es un área en la que aún se está trabajando.

Los resultados aquí presentados no son de ninguna manera los finales, representan tan solo los primeros pasos para la generación de un modelo robusto que nos permita la Predicción de la Generación de energía eléctrica. Este ejercicio está en línea con lo publicado en la literatura, en los que se sabe que la predicción del viento se encuentra correlacionada con la velocidad del mismo.

La relevancia de este trabajo radica en una primera aproximación para la comprensión de la predicción de la generación de energía eléctrica a partir de modelos auto regresivos.

Parte del trabajo futuro consiste en implementar los modelos en lenguaje R para disminuir los tiempos de ejecución que se llevaron a cabo con SPSS.

Referencias

1. De Arce, R., Mahía, R.: Modelos Arima. Programa CITUS: Técnicas de Variables Financieras (2003)
2. Díez, R. P., Fernández, J. P., Priore, P., Gómez, A. G.: Pronóstico de la Velocidad y Dirección del Viento mediante Redes Neuronales Artificiales. In: VIII Congreso de Ingeniería de Organización, pp. 905–914 (2004)
3. Lawan, S. M., Wawzawy, C., Baharun, A., Masri, T.: Different models of wind speed prediction: A comprehensive review. *International Journal of Scientific and Engineering Research*, pp. 1760–1768 (2014)
4. Wang, X., Guo, P., Huang, X.: A review of wind power forecasting models. In: *Energy Procedia, The Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011)*, 12, pp. 770–778 (2011)
5. Cadenas, E., Rivera, W.: Wind speed forecasting in the south coast of Oaxaca, Mexico. *Renewable energy*, 32(12), pp. 2116–2128 (2007)
6. López, R. A. A.: Transferencia Tecnológica y Creación de Capacidades Tecnológicas, en el Enfrentamiento al Cambio Climático: El caso de Proyectos MDL en Energía Eólica en México.
7. Liu, Z., Gao, W., Wan, Y. H., Muljadi, E.: Wind power plant prediction by using neural networks. In: *Energy Conversion Congress and Exposition (ECCE), IEEE*, pp. 3154–3160 (2012)
8. Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M.: *Time series analysis: forecasting and control*. John Wiley & Sons (2015)
9. Hyndman, R. J., Athanasopoulos, G.: *Forecasting: principles and practice*. OTexts (2014)
10. Pankratz, A.: *Forecasting with univariate Box-Jenkins models: Concepts and cases*. John Wiley & Sons, Vol. 224 (2009)
11. Guerrero, J. F. J., Fernández, R. S., Abad, J. C. G.: La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turístico. *Revista Europea de Dirección y Economía de la Empresa*, 15(3), pp. 185–198 (2006)
12. *Statistical forecasting: Notes on regression and time series analysis*, 26 02 2016. Disponible en: <http://people.duke.edu/~rnau/411home.htm>
13. Santamaría-Bonfil, G., Reyes-Ballesteros, A., Gershenson, C.: Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85, pp. 790–809 (2016)
14. Han, J., Pei, J., Kamber, M.: *Data mining: concepts and techniques*. Elsevier (2011)
15. *Wind Power: The Wind Power*. Disponible en: http://www.thewindpower.net/turbine_es_276_turbowinds_300.php (2017)

16. IBM: IBM SPSS Software. Disponible en: <http://www.ibm.com/analytics/us/en/technology/spss> (2017)
17. Priestley, M. B.: Non-linear and non-stationary time series analysis (1988)
18. Jiménez, L. A. F.: Modelos avanzados para la predicción a corto plazo de la producción eléctrica en parques eólicos. Doctoral dissertation, Universidad de La Rioja (2007)
19. Torgo, L., Torgo, L.: Data mining with R: learning with case studies. Boca Raton, FL: Chapman & Hall/CRC (2011)
20. Mattig, I. E.: Predicción de la potencia para la operación de parques eólicos. Memoria para optar al título de Ingeniero civil electricista. Universidad de Chile (2011)
21. CERTE: Centro Regional de Tecnología Eólica (CERTE 2014). Disponible en: <https://www.ineel.mx/detalle-sede-oaxaca.html>