



---

---

UNIVERSIDAD AUTÓNOMA DE YUCATÁN

FACULTAD DE INGENIERÍA

UNIDAD DE POSGRADO E INVESTIGACIÓN

“PREDICCIÓN DE GENERACIÓN DE POTENCIA EN  
INSTALACIONES EÓLICAS CONECTADAS A LA RED  
MEDIANTE REDES NEURONALES ARTIFICIALES”

TESIS

PRESENTADA POR:

ING. JESÚS EDUARDO CASTRO SUASTE

EN OPCIÓN AL GRADO DE

MAESTRO EN INGENIERÍA

OPCIÓN ENERGÍAS RENOVABLES

MÉRIDA, YUCATÁN, MÉXICO

2019

“Aunque este trabajo hubiere servido para el Examen de Grado y hubiere sido aprobado por el Sínodo, sólo el autor es responsable de las doctrinas emitidas en él”

# Contenido

Índice de Figuras .....	2
Índice de Tablas.....	3
1 Introducción .....	4
1.1 Antecedentes .....	4
1.2 Planteamiento del Problema .....	9
1.3 Objetivos .....	9
1.4 Justificación .....	9
2. Marco Teórico .....	11
2.1 Energía Eólica .....	11
2.2 Series de Tiempo .....	12
2.3 Modelos de Predicción.....	14
2.4 Redes Neuronales .....	16
2.4.1 Perceptrón Multicapa .....	26
2.5 Determinación del Orden del Sistema .....	27
2.5.1 Sistema dinámico determinista.....	27
2.5.2 Espacio de fase de un sistema dinámico.....	27
2.5.3 Atractor extraño.....	28
2.5.4 Teoría de embebimiento o inserción .....	28
2.5.5 Selección de tiempo de retardo .....	31
2.5.6 Método basado en la función de autocorrelación.....	31
2.5.7 Ejemplo de la determinación del orden del sistema.....	32
2.6 Aprendizaje en Redes Neuronales .....	34
2.7 Filtro de Kalman .....	36
2.8 Filtro Extendido de Kalman .....	39
2.9 Análisis Wavelet .....	41
2.9.1 Transformada Wavelet Continua (CWT) .....	42
2.9.2 Transformada Wavelet Discreta (DWT) .....	42
2.10 Redes Neuronales Wavelet (WNNs).....	43
2.11 Modelos Estadísticos.....	45
Algoritmo ARIMA.....	45
Modelo AR.....	45

Modelo MA.....	46
Modelo ARMA.....	46
Modelo ARIMA.....	47
3. Metodología.....	50
3.1 Procesamiento de Datos.....	50
3.2 Determinación del Orden del Sistema.....	52
3.3 Prueba Red Neuronal NARX.....	55
3.4 Calculo de redes neuronales y método estadístico.....	57
3.5 Criterios Estadísticos.....	59
4 Resultados.....	61
4.1 Red Neuronal NARX.....	62
4.2 Predicción Perceptrón.....	64
4. Red Neuronal Wavelet con Filtro Extendido de Kalman (NWFEK).....	66
4.3 Métodos Estadísticos ARMA.....	72
4.1 Análisis Comparativo de Resultados.....	73
5 Conclusiones.....	75
Referencias.....	77

## Índice de Figuras

Figura 1. Resultado de predicción de energía de viento a 24 hrs con BP.....	6
Figura 2. Resultado predicción de carga con filtro wavelet.....	8
Figura 3. Resultado predicción de velocidad de viento con filtro wavelet.....	8
Figura 4. Crecimiento de producción mundial de energía eólica.....	11
Figura 5. Modelo de una neurona.....	21
Figura 6. Función de activación escalón.....	21
Figura 7. Función de activación sigmoide logística.....	22
Figura 8. Función de activación lineal a tramos.....	23
Figura 9. Función de activación sigmoide tangente hiperbólica.....	23
Figura 10. Red neuronal monocapa.....	25
Figura 11. Red neuronal multicapa.....	25
Figura 12. Red neuronal recurrente.....	26
Figura 13. Ejemplo de espacio de fase formado por 3 estados de un sistema dinámico.....	27
Figura 14. Datos de irradiancia.....	32
Figura 15. Función de autocorrelación.....	33
Figura 16. Dimensión de embedding de irradiancia.....	33
Figura 17. Teoría de aprendizaje.....	35
Figura 18. Ciclo del Filtro de Kalman.....	36

Figura 19. Diagrama de bloques Filtro de Kalman.....	37
Figura 20. Ejemplo de estructura de red neuronal wavelet. ....	44
Figura 21. Procedimiento Modelo ARIMA .....	48
Figura 22. Secuencia de la Metodología .....	50
Figura 23. Ubicación del Campo Eólico.....	51
Figura 24. Campo eólico en Baja California Sur .....	51
Figura 25. Datos de potencia eólica. ....	52
Figura 26. Función de autocorrelación, mes de febrero. ....	53
Figura 27. Convergencia de la dimensión de embedding por el método de Cao para el mes de febrero. ....	54
Figura 28. Diagrama de flujo de la prueba en red Neuronal NARX .....	55
Figura 29. Estructura de red neuronal para el filtro de Kalman .....	58
Figura 30. Estructura de red Neuronal NARX. ....	58
Figura 31. Resultado red NARX (mes de febrero) .....	62
Figura 32. Legendas de gráfica red NARX .....	62
Figura 33. Testing con red NARX.....	63
Figura 34. Descenso del error vs épocas.....	63
Figura 35. Perceptrón, 5 épocas.....	64
Figura 36. Perceptrón, 10 épocas. ....	65
Figura 37. Perceptrón, 1000 épocas.....	65
Figura 38. Evolución de entrenamiento. ....	66
Figura 39. Entrenamiento de red neuronal wavelet con Filtro extendido de Kalman .....	67
Figura 40. Error de estimación para el entrenamiento NWFEK.....	67
Figura 41. Error relativo absoluto para NWFEK.....	68
Figura 42 Evolución error medio cuadrático. ....	68
Figura 43. Testing (marzo) con NWFEK.....	70
Figura 44. Zoom Testing .....	70
Figura 45 Error relativo absoluto para etapa de testing NWFEK.....	71
Figura 46. Resultado modelo ARMA.....	72
Figura 47. Zoom resultados modelo ARMA .....	73

## Índice de Tablas

Tabla I. Evolución de redes neuronales. ....	17
Tabla II. Características de los datos empleados .....	51
Tabla III. Rezago óptimo de embedimiento.....	53
Tabla IV. Embedding para los datos de potencia.....	54
Tabla V. Respuesta del Modelo Neuronal empleando diferentes retardos.....	56
Tabla VI. Parámetros del modelo de la Red Neuronal.....	56
Tabla VII. Resultados en etapa de entrenamiento NWFEK .....	69
Tabla VIII. Resultados en etapa de entrenamiento NWFEK .....	71
Tabla IX. Resultados método ARMA.....	73
Tabla X. Comparación de Resultados .....	74

# 1 Introducción

## 1.1 Antecedentes

Los combustibles fósiles son la base para satisfacer nuestra necesidad energética actual, por lo que han sido explotados de manera excesiva. En la actualidad casi todo proceso de transformación se realiza teniendo disponibilidad de materiales y su correspondiente cuota de energía. Los problemas comienzan al no tener los recursos para generar dicha energía. Varios estudios dan como probable el año 2050 para la falta de petróleo. Sin embargo, esto podría adelantarse aún más, si se incorporan al “consumo” a los más de 1.400 millones de seres humanos que hoy viven bajo la línea de pobreza, estimando que el agotamiento podría adelantarse varios años (García & Salcedo, 2010). La economía humana basada en el capitalismo ha crecido en gran parte guiada por el costo de producción y en las ganancias que podría generar, sin considerar los daños ocasionados al ambiente (o el agotamiento de recursos) como prioritario al momento de generar la energía requerida.

Un simple análisis sugeriría que se deben realizar cambios tan pronto sea posible para minimizar el riesgo de no poder satisfacer la demanda energética de la raza humana, pues la demanda global por energía continúa creciendo rápidamente, amenazando la estabilidad en los costos y causando preocupaciones acerca de la seguridad del suministro: es claro que se requiere un fuerte despliegue en energías alternativas (Arana, 2015).

Para la realización de la sociedad segura, tenemos que reducir los sistemas de energía convencionales, tales como plantas de energía de combustibles fósiles o centrales nucleares. Una de las tecnologías clave para construir una sociedad con la seguridad ecológica es introducir el sistema de generación de energía renovable. A partir de estos puntos de vista, los sistemas de redes inteligentes que son los sistemas de próxima generación en producción de energía, incluyendo los generadores de energías renovables prestan mucha atención recientemente. Sin embargo, el sistema

de generación de fuentes renovables tiene problemas indeseables para poner en práctica en los sistemas de suministro de energía; la energía eléctrica producida es siempre cambiante por las condiciones del entorno. Por lo tanto, tenemos que introducir un método sofisticado para utilizar los sistemas de generación de energía (Kimura & Kimura, 2013) y poder predecir su comportamiento con la mayor exactitud posible. Entre estos métodos están la utilización de redes neuronales, generalmente alimentadas y entrenadas por medio de series de tiempo. Estas redes neuronales han sido implementadas dentro de la rama de las energías para diversidad de aplicaciones, como control de sistemas híbridos (Gamez, Sanchez, & Ricalde, 2011), predicción de velocidad del viento (Balluff, Bendfeld, & Krauter, 2015), producción de potencia (Jiménes), etc.

Las redes neuronales artificiales (RNAs) se han aplicado de forma exitosa en problemas como el control de sistemas, clasificación de patrones, reconocimiento de patrones, predicción de series de tiempo y solución de problemas de optimización (Alanís, 2009). Las redes neuronales artificiales son atractivas debido a su fácil implementación, estructura relativamente simple, robustez y la capacidad de ajustar sus parámetros en línea (Haykin, *Neural Networks*, 1999). Recientemente, se están aplicando las herramientas de inteligencia computacional en el campo de las energías renovables en problemas de: modelado dinámico, control y predicción. En Holschneider (2013) aplican las RNAs para estimar el ajuste dinámico de precios de energía eléctrica, generando un modelo de respuesta a la demanda de los consumidores a los cambios en precios. Utilizando un algoritmo de optimización se puede identificar el ajuste dinámico de precios óptimo para inducir la respuesta deseada con respecto a una función objetivo predefinida y las restricciones dadas por los consumidores. El ajuste dinámico de precios y los métodos de predicción ayudan a una mejor gestión de los recursos energéticos y reducción de costos de electricidad. Otra aplicación se describe en Feng & Lin (2013) donde se utilizan las RNAs para predecir la velocidad del viento con el fin de reducir la fluctuación de la potencia debido a la naturaleza fluctuante de la energía eólica.

Se han realizado trabajos donde se usan diversos modelos de RNA para predicción de potencial eléctrico, uno de los modelos más usados es el de retro propagación o back propagation (BP). En Wen-hui, Jin, & Zheng-zhong (2014) al analizar el resultado de un parque eólico que es de 49.5 MW de potencia en 6 meses, se demuestra que la red neural de BP puede ser muy buena para predecir la potencia de salida del parque eólico pronosticada día a día. Utilizando las velocidades del viento, dirección del viento seno, dirección del viento coseno y temperatura como variables de entrada y la estructura de red 4-9-1, el error absoluto promedio de 24 h de anticipación es de 4.169 MW; Su error cuadrático medio cuadrático es de 12.26%.

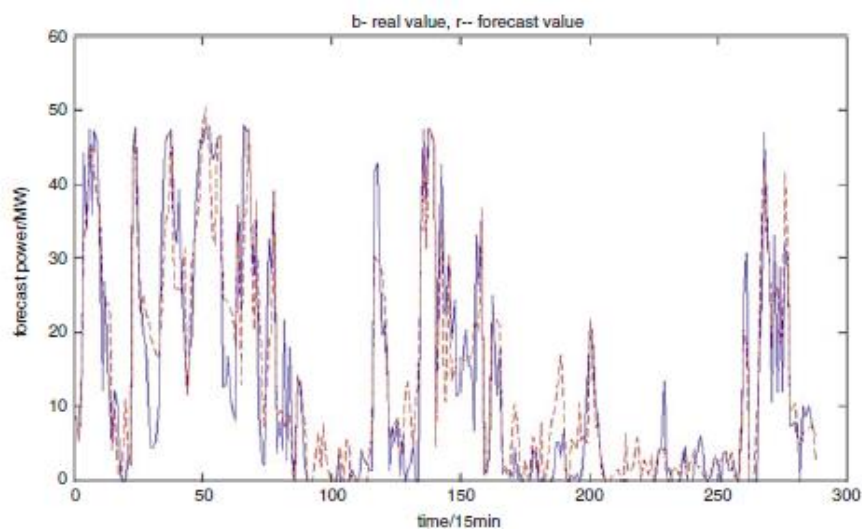


Figura 1. Resultado de predicción de energía de viento a 24 hrs con BP.

Otro elemento que se ha utilizado recientemente en conjunto con las redes neuronales es la transformada Wavelet, esta ha ganado mucha atención recientemente, se ha adoptado en muchas aplicaciones, debido a su impresionante contenido y la frecuencia de aislamiento de características temporales se convierten en buenos incentivos para que los investigadores la utilicen en el área de análisis de sistemas de potencia. La transformada puede ser utilizada para suavizar los datos de entrenamiento que se utilizan para la red neuronal artificial (RNA) para predecir la carga, la radiación solar y velocidad del viento. La primera etapa es el uso de la descomposición wavelet para



convertir los datos en aproximaciones y detalles con menores niveles de resolución (Ghareeb & Mohammed, 2013).

Utilizando diferentes modelos de redes neuronales con filtro wavelet, se han realizado diversos estudios para probar su eficiencia. Por ejemplo, citando nuevamente el trabajo de Ghareeb & Mohammed (2013), la radiación solar, la velocidad del viento y la carga se pronosticaron con exactitud utilizando filtrado wavelet y RNA. La transformada wavelet se utilizó para suavizar los datos de entrenamiento, los valores pronosticados se comparan con los reales y el error se utilizó para adaptar la red continuamente. Se demostró que esta combinación mejoraba la previsibilidad. Estas previsiones se integraron al sistema de gestión de energía para lograr una alimentación segura y continua a la carga de las fuentes de energía renovables y la red de AC (sólo si es necesario). El experimento llevado a cabo tiene una configuración que consiste en PV y emuladores de viento conectado a la red de AC a través de un inversor controlado.

Los resultados de la predicción para la carga y velocidad de viento se muestran en la Figura 2 y Figura 3, La Figura 2 muestra la carga horaria pronosticada versus la carga real durante una semana. Se muestra que la red neuronal puede rastrear los cambios bruscos de carga, en algunos puntos la flexibilidad de la red no es suficiente para predecir exactamente el cambio de carga. Esto se debe a la pendiente pronunciada de la carga en algunos puntos. La velocidad prevista del viento y la velocidad real del viento se muestran en la Figura 3. De la figura, se observa que la red neural reacciona rápidamente a los cambios repentinos del viento. La curva de viento pronosticada está muy cerca de la curva de viento real, lo que confirma la alta previsibilidad del sistema.

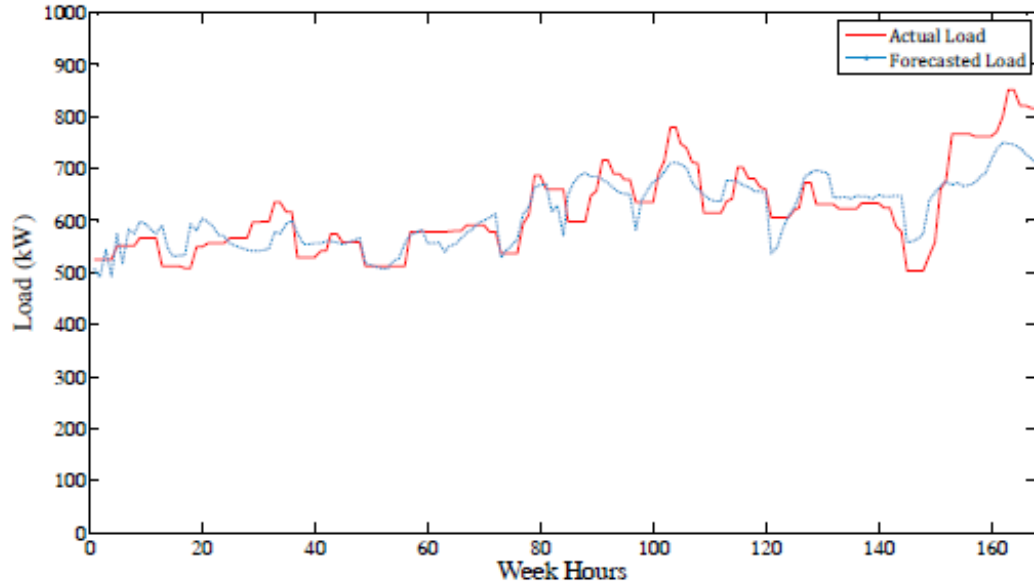


Figura 2. Resultado predicción de carga con filtro wavelet.

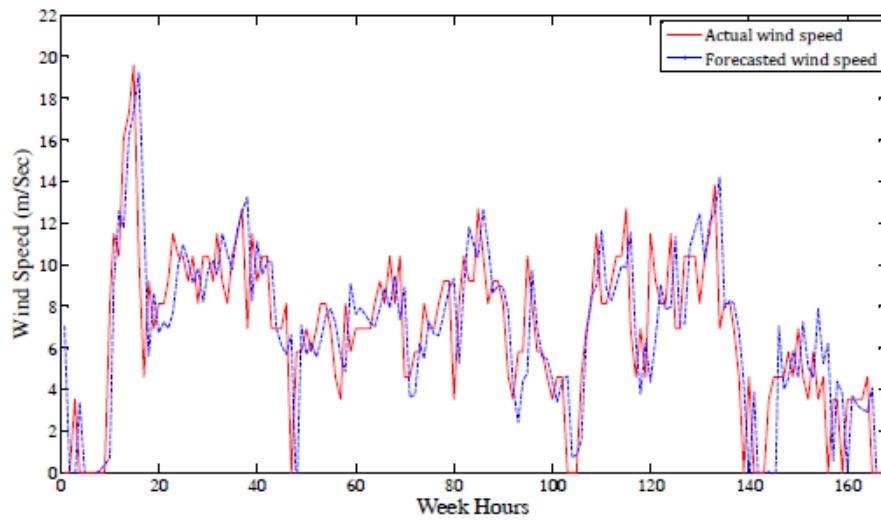


Figura 3. Resultado predicción de velocidad de viento con filtro wavelet.

## 1.2 Planteamiento del Problema

Esta investigación pretende lograr el diseño e implementación de un algoritmo para predecir en series de tiempo la potencia eléctrica generada a partir del viento, aplicable a campos eólicos de producción, desarrollándolo en el entorno de simulación Matlab utilizando arquitecturas de redes neuronales artificiales. Primero se procesarán los datos y luego se insertarán a los modelos de las redes neuronales para analizar el error de la estimación de la predicción alcanzada en el sistema y así obtener una comparación estadística.

## 1.3 Objetivos

El objetivo fundamental del proyecto es el desarrollo de un nuevo modelo (basado en redes neuronales) para la predicción de la producción eléctrica en parques eólicos.

Objetivos secundarios:

- Determinar la producción eléctrica lograda de los parques eólicos mediante los datos obtenidos con el modelo de predicción.
- Predecir la potencia que será introducida a la red eléctrica por parte de las plantas eólicas.

## 1.4 Justificación

Es indispensable tener una buena estimación del potencial eléctrico para poder integrar de manera efectiva la energía generada por medio de este recurso renovable en el sistema de suministro eléctrico. La combinación de la predicción de la velocidad del viento y la energía de salida es una buena manera de mejorar la planeación de la gestión de energía en un parque eólico (Jie, Gao, Sizhe, & Jun, 2009). La integración del recurso eólico a los sistemas de transmisión de electricidad actuales tiene como

consecuencia directa beneficios ambientales al reducir las emisiones de dióxido de carbono derivadas de la generación de electricidad por medios tradicionales, contribuyendo por tanto a la preservación de nuestro planeta. Otras razones por las cuales es importante dicha estimación son (El-Fouly, El-Saadany, & Salama, 2008):

- Planeación del mantenimiento de las unidades en los parques eólicos.
- Flujo de energía óptimo entre las unidades convencionales y los parques eólicos.
- Oferta de comercialización de la electricidad.
- Programación de los sistemas generadores de energía.
- Planeación y programación de las reservas y almacenamientos de energía.

## 2. Marco Teórico

### 2.1 Energía Eólica

Las nuevas energías se refieren generalmente a fuentes de energía no convencionales, tales como energía eólica, energía solar, energía del océano, energía hidroeléctrica, energía de la biomasa, energía geotérmica, y así sucesivamente. En los últimos años, el desarrollo y la utilización de nuevas energías se ha convertido en uno de los enfoques más importantes para resolver la presión sobre los recursos y el deterioro del medio ambiente. La energía eólica es la fuente de energía renovable de más rápido crecimiento en el mundo, su avance exponencial se puede observar en la Figura 4. En los últimos cinco años, Estados Unidos ha sido testigo de un crecimiento exponencial de la capacidad de generación de energía eólica con más de 65.88 GW generado en 2014 por sí solas. Esta capacidad no sólo depende de la eficiencia de las turbinas de viento, sino también en sus ubicaciones. La influencia de los sistemas de conversión de energía eólica sobre la estabilidad y la fiabilidad del sistema de energía aumentó significativamente con el crecimiento de la penetración de los sistemas de energía renovable. Como resultado, se han introducido requisitos más estrictos para la calidad de la energía procedente de las turbinas de viento y parques eólicos en los últimos años (Shahat, Haddad, & Kalaani, 2015).

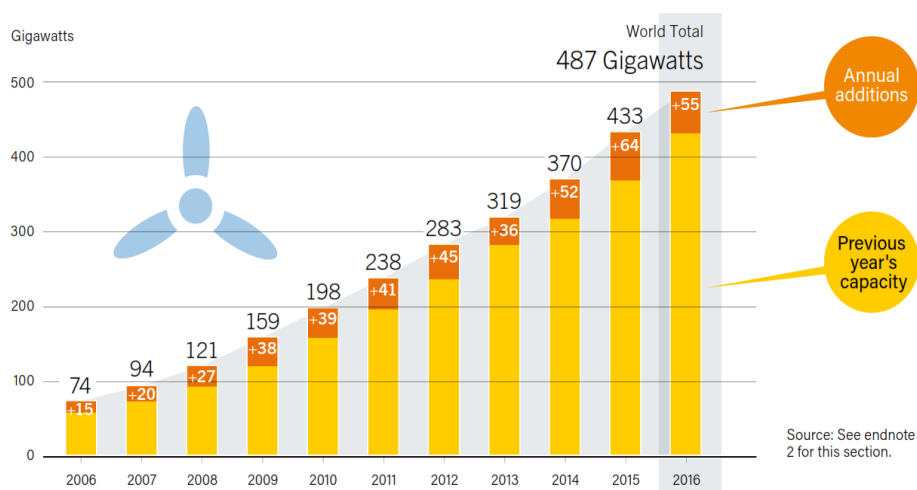


Figura 4. Crecimiento de producción mundial de energía eólica.

Uno de los problemas más comunes relacionados con los sistemas de energía eólica es la naturaleza impredecible del nivel de potencia de salida. Variaciones de la velocidad del viento pueden causar cambios repentinos en la cantidad de energía generada por las turbinas de viento y en consecuencia todo el parque eólico. Para los grandes parques eólicos, estos cambios pueden causar inestabilidades incluso en todo el sistema de alimentación (Novakovic, Pashaie, & Nasiri, 2014).

Para optimizar la selección de la ubicación y generación de potencia en la turbina, las mediciones a gran escala de velocidades del viento tienen que ser recogida durante un largo período de tiempo que en muchos casos no es factible. Por lo tanto, modelos predictivos precisos son necesarios para optimizar los potenciales de estas fuentes. En los últimos años, la generación de energía eólica ha sido testigo de un renovado interés en el modelado y control de sistemas híbridos. La formulación matemática de tales modelos es tediosa, por tanto, varios investigadores han desarrollado diferentes técnicas para la simulación por ordenador que representan muy amplia gama de aplicaciones de turbinas eólicas. En el presente trabajo se propone un modelo de predicción de la generación de energía eólica utilizando mediciones de muestras como datos de entrenamiento para los algoritmos de predicción. En este modelo, las redes neuronales artificiales (RNA) serán utilizadas para predecir la potencia eléctrica generada a partir de datos previos en series de tiempo.

## 2.2 Series de Tiempo

A lo largo de la historia, la predicción del comportamiento de los fenómenos de interés ha sido importante. Este hecho se refleja en la diversidad de aplicaciones de predicción en prácticamente todas las áreas de conocimiento. Cuando la información cuantitativa en relación con el comportamiento de un fenómeno particular con respecto al tiempo está disponible, entonces uno tiene una serie de tiempo. Las previsiones de series de tiempo se generan comúnmente a través de técnicas estadísticas tradicionales. El conjunto de conocimientos que abarca estas técnicas se llama análisis de series temporales (Aguilar, Rodriguez, & Ríos, 2006)

Una serie de tiempo es un conjunto de observaciones realizadas de forma secuencial a través del tiempo. Los ejemplos incluyen la temperatura del aire medida por hora, precios de las acciones y medidas a diario los niveles de monóxido de carbono medidos anualmente. En la predicción de series de tiempo, el enfoque general es mirar los datos del pasado, adaptarse a un modelo apropiado y luego proyectar valores futuros utilizando el modelo.

La idea general primero a destacar es que la previsión de series de tiempo va en contra de todos los demás trabajos estadísticos. Cuando los alumnos aprenden técnicas como la regresión que se les enseña habitualmente que no es prudente utilizar un modelo ajustado para hacer predicciones fuera del rango de los valores sobre los que el modelo ha sido equipado. En otras palabras, la extrapolación es generalmente imprudente. Sin embargo, en la predicción de series de tiempo, eso es exactamente lo que el analista tiene que hacer. Aunque a menudo inevitable, esto significa que la predicción de series de tiempo es potencialmente desastrosa cuando el futuro no es como el pasado (Chatfield, 2005)

Las series de tiempo continuas son grabadas instantánea y continuamente, como un oscilógrafo graba las oscilaciones armónicas de un amplificador de audio. La mayoría de las mediciones en las ciencias sociales son hechas en intervalos regulares, y los datos de esta serie de tiempo son discretos. Las acumulaciones de lluvia medidas discretamente a un intervalo regular pueden ser un ejemplo. Otras pueden ser agrupadas de observaciones individuales para concentrar un resumen estadístico, medido en intervalos regulares sobre el tiempo. Algunas series que no están ordenadas cronológicamente pueden ser sujetas al análisis de series de tiempo. Idealmente, las series usadas consisten en observaciones que son equidistantes una de otra en tiempo y no contienen observaciones faltantes.

Si los valores futuros pueden ser predichos exactamente de valores pasados, se dice que la serie es determinista. Por otro lado, la mayoría de las series son estocásticas o aleatorias, por lo que el futuro es solamente parcialmente determinado por valores pasados (Arana, 2015).

## 2.3 Modelos de Predicción

Las previsiones son generalmente las bases para la toma de decisiones a todos los niveles que van desde las decisiones operativas del día a día a los más estratégicos a largo plazo en muchas organizaciones. Dada su importancia, no es de extrañar que la predicción se ha convertido en un área de investigación muy activa.

En la actualidad la mayoría de los sistemas maduros de previsión de energía eólica se desarrolla principalmente por Europa, los Estados Unidos y otros países desarrollados. Por ejemplo, el sistema de predicción desarrollado por el laboratorio nacional danés utiliza el llamado “modelo físico”, el sistema de predicción desarrollado por la universidad alemana de Walden utiliza el método de predicción de combinación y puede dar 2 días de predicción de la energía eólica en gran área; El proyecto ANMOS, desarrollado conjuntamente por Francia, Irlanda, España, Alemania, Grecia, Dinamarca e Inglaterra, utiliza el modelo de previsión combinada y puede aplicarse tanto a los parques eólicos terrestres como marinos; Además, hay también el sistema de EWind que es desarrollado por la compañía de TrueWind de Estados Unidos y el sistema de GH-FORECASTER que es desarrollado por la compañía de Garrad Hassan. El sistema de pronóstico que ha sido diseñado por el instituto chino de ciencias eléctricas ha sido puesto en aplicación; Su error RMS (error cuadrático medio por sus siglas en inglés) es de alrededor del 15%.

Según las diferentes escalas de tiempo, la predicción de la energía del viento se puede dividir en tres clases: una es la predicción a corto plazo (algunos minutos), aplicada principalmente en el control del generador de viento; Otra es la predicción a medio plazo (horas a días), utilizada principalmente en la red de energía eólica conectada a la red y el envío de la red; El último es predicción a largo plazo (semanas y meses), utilizado principalmente en parques eólicos y plan de mantenimiento de la red. De acuerdo con las diferentes necesidades de modo de operación de la disposición del sistema de planificación del sistema de energía, la predicción de la energía eólica se puede dividir en la predicción de día a día y la predicción en tiempo real, la previsión del día es la previsión de 24 h. La predicción en tiempo real es la previsión de balanceo para cada punto. De acuerdo con el modelo de predicción diferente, el método de



predicción de energía eólica generalmente se puede dividir en métodos físicos y métodos de estadística (Wen-hui, Jin, & Zheng-zhong, 2014).

De acuerdo con los diferentes métodos de modelado utilizados, los actuales modelos de predicción de energía eólica pueden dividirse en modelos físicos, modelos estadísticos, inteligencia artificial y modelos híbridos. Algunos factores físicos y geográficos, tales como la temperatura del aire, la presión atmosférica, la densidad atmosférica, la topografía y la rugosidad de la superficie, se aplican en modelos físicos para obtener la velocidad del viento en el cubo del ventilador axial. Por lo tanto, la predicción numérica de alta resolución del tiempo se realiza por este medio, que es especialmente adecuado para la predicción de energía eólica a largo plazo. Basados en grandes cantidades de datos históricos, los modelos estadísticos, que generalmente incluyen el modelo de persistencia, el modelo de serie temporal y el modelo de filtrado de Kalman, tienen como objetivo establecer una relación lineal entre la entrada y la salida de los modelos de predicción. En los últimos años, la tecnología inteligente artificial ha sido ampliamente utilizado en el campo de la predicción de la energía eólica. La inteligencia artificial toma muchas formas, como la red neuronal wavelet (WNN), la red neural de retropropagación (BP), la red neural de función de base radial (RBF), la máquina vectorial de soporte (SVM) y la lógica difusa. Con el fin de evitar las limitaciones de los métodos de predicción individuales, los modelos híbridos se están proponiendo cada vez más en los últimos años (Bingheng Wu, 2014).

El uso de métodos de predicción lineales, tales como medias móviles, suavización exponencial, regresión lineal y la descomposición de series de tiempo han dominado la escena durante décadas. Especialmente significativo en términos de uso es la técnica llamada autorregresivos integrados de media móvil (ARIMA en inglés), desarrollado por Box & Jenkins (1976). A pesar del uso generalizado de estos métodos lineales, la existencia de relaciones no lineales en la serie de tiempo puede limitar su aplicación en muchos casos. Relaciones no lineales son, de hecho, nada raras en absoluto en la realidad, por lo que es necesario recurrir a técnicas capaces de reflejar adecuadamente este tipo de comportamiento. Las redes neuronales artificiales (RNAs) se han sugerido como una técnica de pronóstico alternativa debido a su capacidad

para aproximar el comportamiento no lineal. De hecho, las RNA también se han demostrado para ser competitivos incluso cuando el caso es uno de linealidad (Medeiros & Pedreira, 2011); (Zhang, 2001).

Las redes neuronales artificiales (RNAs) se han aplicado de forma exitosa en problemas como el control de sistemas, clasificación de patrones, reconocimiento de patrones, predicción de series de tiempo y solución de problemas de optimización (Alanís, 2009). Las redes neuronales artificiales son atractivas debido a su fácil implementación, estructura relativamente simple, robustez y la capacidad de ajustar sus parámetros en línea (Haykin, Neural Networks, 1999). Recientemente, se están aplicando las herramientas de inteligencia computacional en el campo de las energías renovables en problemas de: modelado dinámico, control y predicción. En Holschneider (2013) aplican las RNAs para estimar el ajuste dinámico de precios de energía eléctrica, generando un modelo de respuesta a la demanda de los consumidores a los cambios en precios. Utilizando un algoritmo de optimización se puede identificar el ajuste dinámico de precios óptimo para inducir la respuesta deseada con respecto a una función objetivo predefinida y las restricciones dadas por los consumidores. El ajuste dinámico de precios y los métodos de predicción ayudan a una mejor gestión de los recursos energéticos y reducción de costos de electricidad. Otra aplicación se describe en Feng & Lin (2013) donde se utilizan las RNAs para predecir la velocidad del viento con el fin de reducir la fluctuación de la potencia debido a la naturaleza fluctuante de la energía eólica.

## 2.4 Redes Neuronales

Una red neuronal artificial es un procesador paralelo distribuido formado por unidades de procesamiento (neuronas) y cuya función es almacenar conocimiento adquirido de manera experimental y generalizar para estímulos similares. El conocimiento es adquirido por la red a través de un proceso de aprendizaje por ejemplos y las fuerzas de conexión entre las neuronas, conocidas como pesos sinápticos, son usadas para almacenar dicho conocimiento. El procedimiento usado para realizar el proceso de

aprendizaje es llamado algoritmo de entrenamiento (Cámara, 2016). Este algoritmo de entrenamiento es el que es alimentado por nuestros datos, que para el caso serían mediciones meteorológicas.

A continuación, en la Tabla I se muestran las redes neuronales más conocidas, sus aplicaciones importantes y desarrolladores (Espadas G. , 2002)

Tabla I. Evolución de redes neuronales.

RNA	AÑO	APLICACIONES	DESARROLLADOR
Perceptron	1957	Reconocimiento de caracteres impresos	Fran Rosenblatt
Adaline/Madaline	1960	Filtrado de señales. Ecuilizador adaptativo	Berned Widrow
Avalancha	1967	Reconocimiento de habla continua. Control brazos robot.	Stephen Grossberg
Cerebellatron	1969	Control de brazos robóticos.	David Marr, James Albus, Andres Pellionez.
Back-propagation	1974-86	Síntesis de voz desde texto. Control de robots. Predicción. Reconocimiento de patrones.	Paul Werbos, David Parker, David Rumelhart.
Brain-Estate-in-a-Box	1977	Extracción de conocimiento de base de datos.	James Anderson

Neocognitron	1978-84	Reconocimiento de caracteres manuscritos.	K. Fukushima
Self Organizing-Map. Opology Preserving Map.	1980-84	Reconocimiento de patrones. Codificación de datos. Optimización	Teuvo Kohonen.
Hopfield	1982	Reconstrucción de patrones y optimización	John Hopfield
Memoria Asociativa bidireccional	1985	Memoria heteroasociativa de acceso por contenido.	Bart Kosko
Máquinas de Boltzmann y Cauchy	1985-86	Reconocimiento de patrones. Optimización	Jeffrey Hinton, Terry Sejnowski, Harold Szu.
Teoría Resonancia Adaptativa	1986	Reconocimiento de patrones.	Gail Carpenter, Stephen Grossberg.
Counter propagation	1986	Compresión de imágenes	Robert Hecht-Nielsen.

Las redes neuronales, o redes neuronales artificiales para ser más precisos, representan una tecnología que está basado en muchas disciplinas: neurociencia, matemáticas, estadística, física, ciencias de la computación e ingeniería. Trata de emular la estructura del cerebro, que tiene la habilidad de construir sus propias reglas basadas en sus experiencias previas. Su uso ofrece las siguientes propiedades y capacidades útiles:

- No linealidad. Una neurona artificial puede ser lineal o no lineal. Es básicamente no lineal.
- Mapeo de entrada-salida. Modificación de los pesos sinápticos de la red neuronal al aplicar lo que se llama “entrenamiento”.
- Adaptabilidad. Adapta a los pesos sinápticos a cambios en el ambiente que lo rodea. Una red neuronal entrenada a operar en un ambiente específico puede ser fácilmente re-entrenada.
- Respuesta evidencial. En el contexto de reconocimiento de patrones, no solo da información de un patrón específico, sino la confianza en la decisión hecha.
- Información contextual. Cada neurona de la red es potencialmente afectada por la actividad global de las otras neuronas.
- Tolerancia a fallas. Su rendimiento se degrada levemente en condiciones de operación adversas, debido a su interconexión masiva.
- Uniformidad de análisis y diseño. Son universales, es decir, la misma notación es usada en todos los dominios involucrados en su aplicación.
- Analogía neurobiológica. Diseño basado en el cerebro, que es la prueba viviente del procesamiento paralelo tolerante a fallas. (Haykin, 2005)

La red neuronal artificial es un sistema de procesamiento de información que tiene ciertas aptitudes en común con las redes neuronales biológicas:

- El procesamiento de información ocurre en muchos elementos simples llamados neuronas.
- Las señales son transferidas entre neuronas a través de enlaces de conexión.

- Cada conexión tiene un peso asociado, el cual típicamente multiplica la señal transmitida.
- Cada neurona aplica una función de activación (usualmente no lineal) a su entrada de red (suma de entradas procesadas) para determinar su salida.

La neurona artificial es la unidad principal de procesamiento *Figura 5*. Cada neurona tiene un estado interno, llamado nivel de activación o función de transferencia  $f$ , donde se evalúa la sumatoria de las señales de entrada. Típicamente, una neurona envía su activación como una señal a otras neuronas (Espadas, 2002). Considerando una neurona  $u$ , ilustrada en la *Figura 5*, que recibe las señales de entradas  $x_1, x_2$  hasta  $x_n$ . Los pesos correspondientes son  $w_1, w_2$ , hasta  $w_n$  y un bias  $b$ .

Entonces, la salida de la neurona  $u(x, w)$  se expresa matemáticamente como:

$$u(x, w) = f(x_1w_1 + x_2w_2 + \dots + x_nw_n + b) \quad (1)$$

$$u(x, w) = f\left\{\sum_{i=1}^n (x_iw_i + b)\right\} \quad (2)$$

O en forma matricial como:

$$u(x, w) = f(WX + b) \quad (3)$$

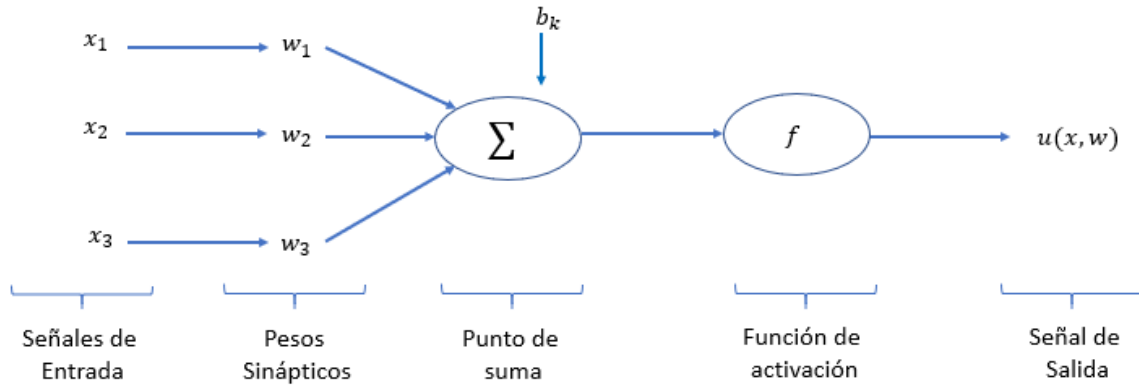


Figura 5. Modelo de una neurona.

Las funciones de activación, denotadas por  $f$  ó comúnmente  $\varphi(v)$ , definen la salida de la neurona en función del potencial de activación. Se incluyen 4 tipos de funciones de activación básicas:

1. Función escalón o umbral. Para este tipo de función de activación, presentada en la Figura 6 se tiene

$$\varphi(v) \begin{cases} 1 & \text{si } v \geq 0 \\ 0 & \text{si } v < 0 \end{cases} \quad (4)$$

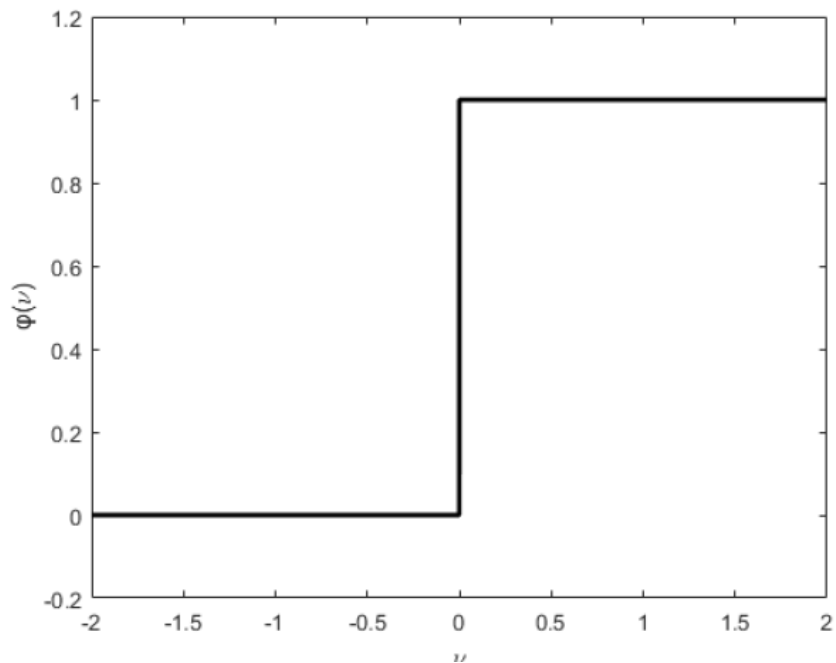


Figura 6. Función de activación escalón.

2. Función sigmoide logística. Es la función más utilizada en la construcción de redes neuronales artificiales. Es estrictamente creciente y exhibe un balance perfecto entre el comportamiento lineal y no lineal. Esta función se define por

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (5)$$

Donde  $a$  es el parámetro que determina la pendiente de la función. Variando el parámetro  $a$  se obtienen las funciones mostradas en la Figura 7.

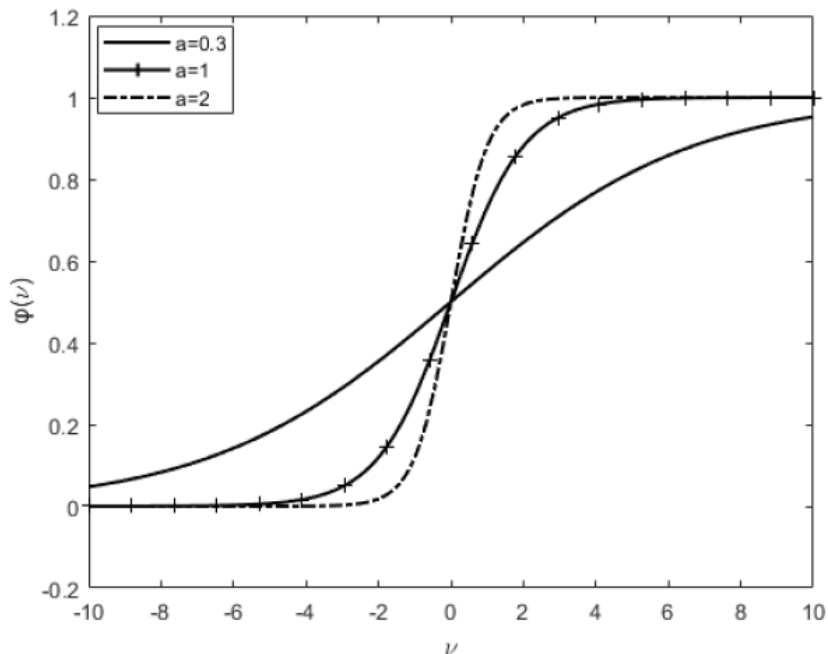


Figura 7. Función de activación sigmoide logística.

3. Función lineal a tramos. Esta función se muestra en la Figura 8. y está definida como

$$\varphi(v) \begin{cases} 1 & \text{si } v \geq 1 \\ v & \text{si } -1 < v < 1 \\ -1 & \text{si } v \leq -1 \end{cases} \quad (6)$$



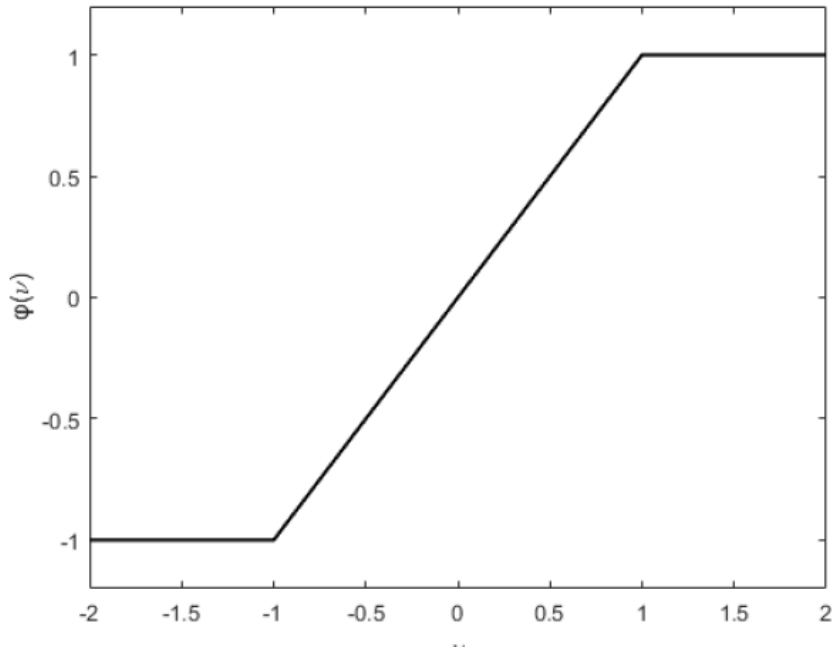


Figura 8. Función de activación lineal a tramos.

4. Función sigmoide tangente hiperbólica. La ecuación que define esta función es

$$\varphi(v) = \frac{1}{1 + e^{-av}} - 1 \quad (7)$$

Esta función se muestra en la Figura 9 para diferentes valores del parámetro  $a$ , el cual también tiene el efecto de variar la pendiente de la función.

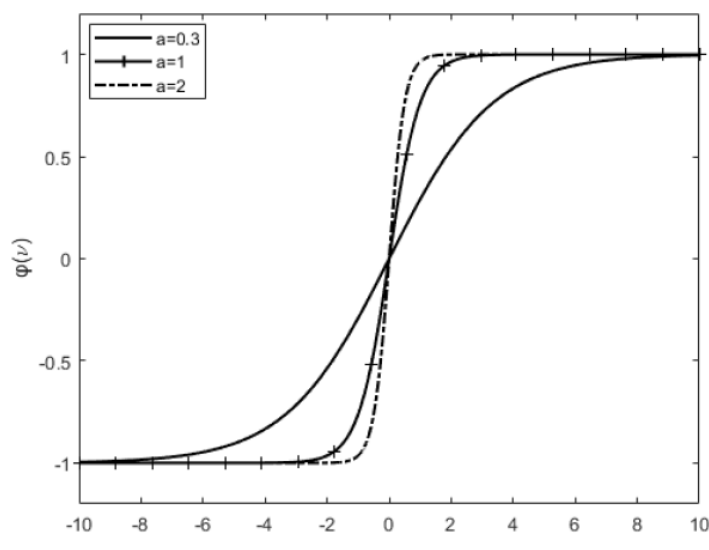


Figura 9. Función de activación sigmoide tangente hiperbólica.

La no linealidad de las funciones de activación constituye la base para que una red neuronal pueda ser el aproximador universal de funciones no lineales.

Una red neuronal artificial es caracterizada por (Fausett, 1994):

1. El patrón de conexión entre las neuronas (Arquitectura). Las redes pueden ser de una sola capa o de múltiples capas.
2. El método para determinar los pesos en las conexiones (entrenamiento, aprendizaje o algoritmo). El aprendizaje se divide en: supervisado o no supervisado. El aprendizaje supervisado puede ser por refuerzo o por corrección.
3. La función de activación.

En general podemos identificar 3 diferentes clases de arquitecturas neuronales:

- Redes estáticas de una capa (Figura 10). No tienen realimentación. Las redes neuronales se organizan por capas. En la forma más simple se cuenta con una capa de entrada de neuronas (nodos de entrada) que proyecta una capa computacional (nodo de salida), pero no viceversa.
- Redes estáticas multicapa (Figura 11). Puede estar completamente conectada (en cada capa cada nodo está conectado a cada nodo de la capa posterior) o parcialmente conectada.
- Redes recurrentes (Figura 12). Presentan al menos un lazo de realimentación, puede tener o no capas ocultas.

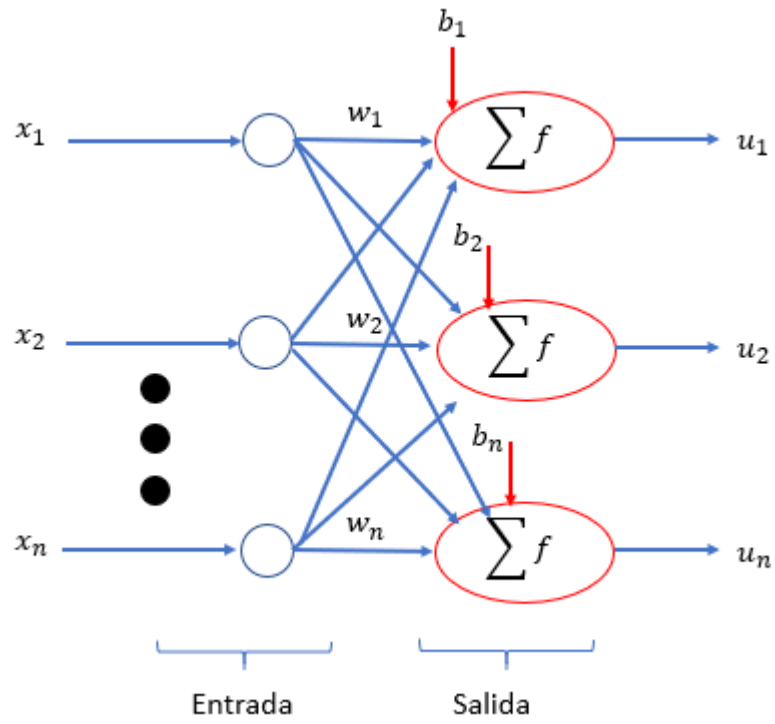


Figura 10. Red neuronal monocapa.

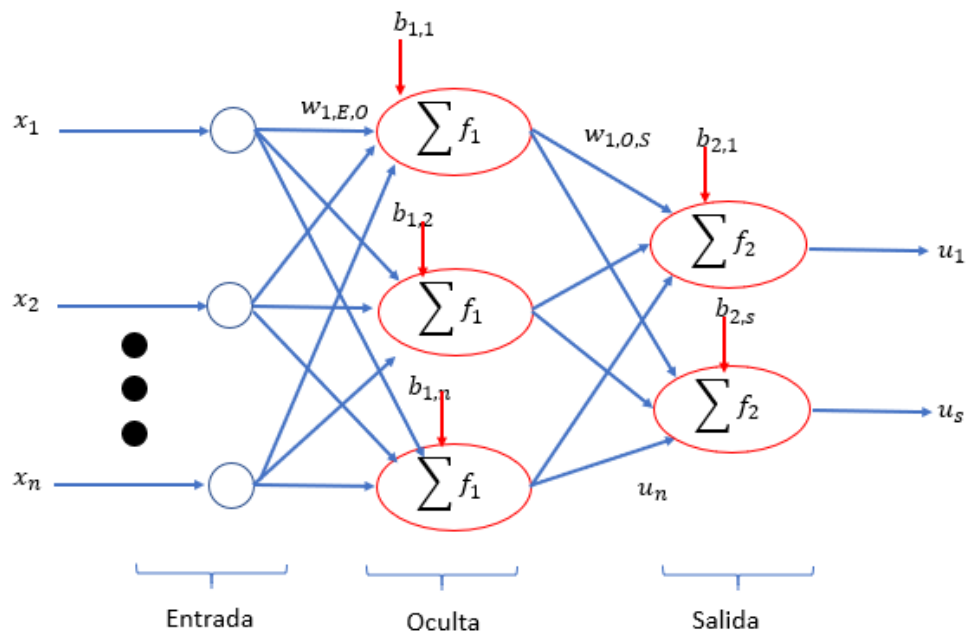


Figura 11. Red neuronal multicapa.

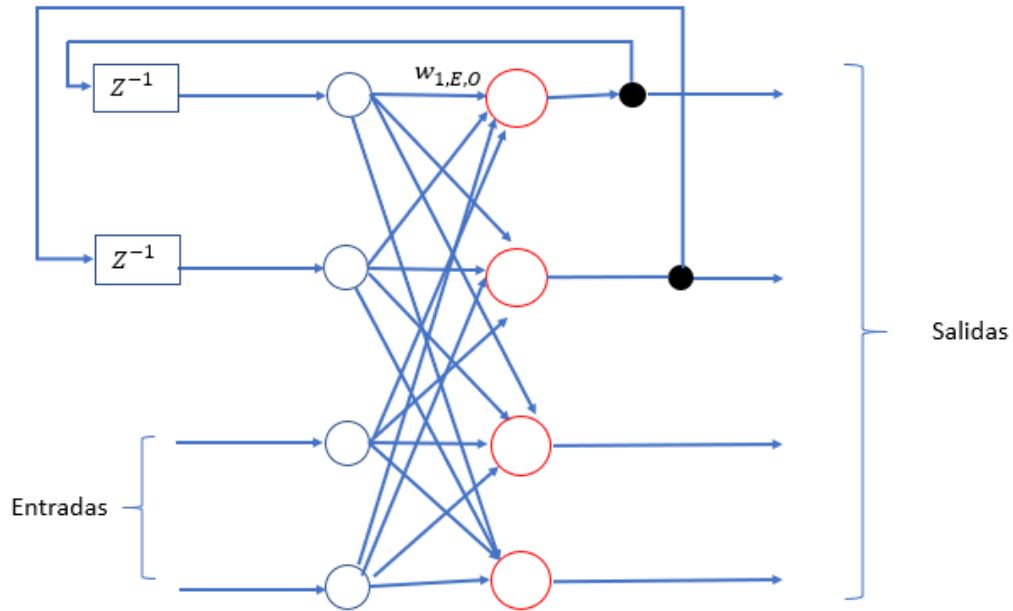


Figura 12. Red neuronal recurrente.

#### 2.4.1 Perceptrón Multicapa

Un perceptrón multicapa (MLP) es un modelo neuronal artificial unidireccional que asigna conjuntos de datos de entrada en un conjunto de salidas apropiadas. Un MLP consta de múltiples capas de nodos de un grafo dirigido, con cada capa totalmente conectada a la siguiente. Excepto para los nodos de entrada, cada nodo es una neurona (o elemento de procesamiento) con una función de activación lineal. MLP utiliza una técnica de aprendizaje supervisado denominado retropropagación para la formación de la red.

## 2.5 Determinación del Orden del Sistema

### 2.5.1 Sistema dinámico determinista

Por sistema dinámico se entiende un conjunto de elementos interconectados entre sí y destinados a cumplir una tarea específica. Este conjunto de elementos, o sistemas, describe un comportamiento cuya evolución depende del tiempo, ya sea de forma explícita o implícita y de forma continua o discreta. Se dice que es determinista porque los valores futuros del sistema pueden ser determinados a partir de los valores actuales. En el caso en que el sistema dependa de una variación temporal discreta también suele denominarse Mapa discreto.

### 2.5.2 Espacio de fase de un sistema dinámico

Se le llama espacio de fase o espacio de estado (de un sistema completamente determinista) al espacio vectorial en el cual un punto representa un estado del sistema y viceversa. En este espacio los ejes coordenados son los estados del sistema dinámico. Por otro lado, el número de estados es igual a la dimensión del sistema y también se denomina grados de libertad del sistema Figura 13.

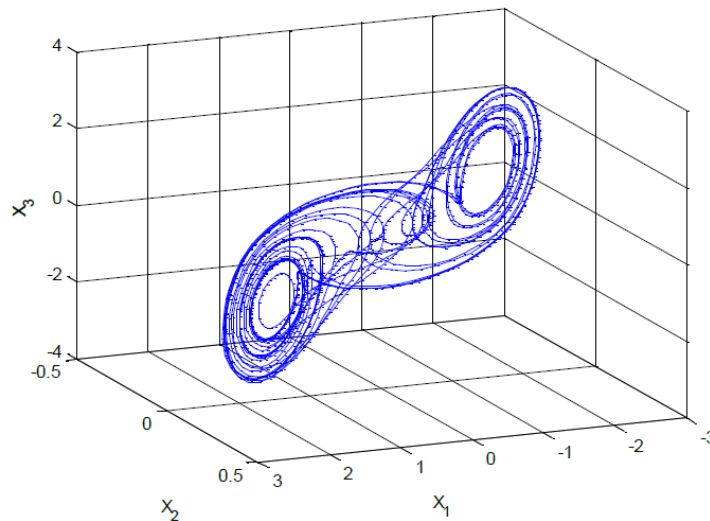


Figura 13. Ejemplo de espacio de fase formado por 3 estados de un sistema dinámico.

### 2.5.3 Atractor extraño

De forma general se puede decir que un atractor es una región o conjunto cerrado en el espacio de fase al cual convergen las trayectorias de un sistema, entendiéndose por trayectoria la evolución temporal del sistema a partir de una condición inicial específica.

Los atractores cumplen las siguientes propiedades:

- El atractor es invariante, cualquier trayectoria que comience dentro del atractor, permanecerá en él en todo momento (siempre y cuando no se perturbe el sistema).
- Al atractor convergen las trayectorias desde un conjunto abierto de condiciones iniciales, al cual se le llama la “cuenca de tracción”.
- El atractor es el mínimo conjunto que cumpla las dos condiciones anteriores, es decir, no contiene subconjuntos propios con sus mismas propiedades.
- Por atractor extraño se entiende una región del espacio de fase donde la evolución del sistema dinámico esta confinada en una región acotada y sin embargo las trayectorias cercanas del atractor se alejan exponencialmente, teniendo en cuenta que la distancia máxima que pueden separarse está limitada por el tamaño del atractor.

### 2.5.4 Teoría de embebimiento o inserción

Al intentar construir un modelo de predicción para una serie temporal, es importante poder caracterizar la dimensión de incrustación de la serie. La dimensión de inserción define el número de valores previos de la serie que van a determinar el siguiente. En otras palabras, es equivalente al orden de una ecuación diferencial que definiría la serie con respecto a los valores previos. No necesariamente nos dice qué valores son importantes, solo cuántos. En el caso simple, suponemos que los valores más recientes determinan el siguiente. Una serie temporal en la cual los dos valores anteriores son suficientes para determinar el actual (usando el ejemplo de una onda sinusoidal: necesitamos el último valor para saber en qué parte de la onda estamos, y

el valor anterior para saber si estamos subiendo o bajando la ola) tiene una dimensión de inserción de dos.

Determinar la dimensión de embedding (inserción) es considerado como uno de los más importantes pasos en el modelado y predicción en series de tiempo no lineales. Un número de métodos han sido desarrollados para la determinar la mínima dimensión de inserción desde el estudio temprano de series de tiempo no lineales.

El método utilizado en el presente trabajo para calcular la dimensión de embedding es el llamado método de Cao o de falsos vecinos cercanos promediados. Supongamos que tenemos una serie de tiempo

$$x_1, x_2, \dots, x_N. \quad (8)$$

Los vectores de tiempo de retardo pueden ser reconstruidos como sigue:

$$y_i(d) = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau}), \quad (9)$$

$$i = 1, 2, \dots, N(d-1)\tau$$

Donde  $d$  es la dimensión de inserción y  $\tau$  es el tiempo de retardo. Nota que  $y_i(d)$  significa el  $i$ -ésimo vector reconstruido con dimensión de inserción  $d$ . Ahora procedemos a definir:

$$a(i, d) = \frac{\|y_i(d+1) - y_{n(i,d)}(d+1)\|}{\|y_i(d) - y_{n(i,d)}(d)\|}, \quad (10)$$

$$i = 1, 2, \dots, N - d\tau,$$

Donde  $\|\cdot\|$  es una medida de distancia Euclidiana y está dado por la norma máxima, i.e.,

$$\|y_k(m) - y_l(m)\| = \max_{0 \leq j \leq m-1} |x_{k+j\tau} - x_{l+j\tau}|; \quad (11)$$

$y_i(d+1)$  es el  $i$ -ésimo vector reconstruido con dimensión de inserción  $d+1$ , i.e.,

$$y_i(d+1) = (x_i, x_{i+\tau}, \dots, x_{i+d\tau}); \quad (12)$$

$n(i, d)$  ( $1 \leq n(i, d) \leq N - d\tau$ ) es un entero tal que  $y_{n(i, d)}(d)$  es el vecino más cercano de  $y_i(d)$  en el espacio de fase  $d$  dimensional en el sentido de distancia  $\|\cdot\|$  que definimos anteriormente.  $n(i, d)$  depende de  $i$  y  $d$ .

Si  $d$  es calificado como una dimensión de inserción por el teorema de embedding o de Takens, entonces 2 puntos cualquiera los cuales permanecen cercanos en el espacio reconstruido  $d$ -dimensional estarán todavía cercanos en el espacio reconstruido  $(d+1)$ -dimensional. Tal par de puntos son llamados verdaderos vecinos, de otra manera, serán llamados falsos vecinos.

Inserción perfecta significa que no existen falsos vecinos. Esta es la idea del método de falsos vecinos, diagnosticar a un falso vecino al ver el  $a(i, d)$ . El problema es cómo escoger este valor límite. De la definición de  $a(i, d)$ , uno puede ver que el valor límite puede ser determinado por la derivada de la señal subyacente, por lo tanto, para diferentes puntos de fase  $i$ ,  $a(i, d)$  puede tener diferentes valores límites, al menos en principio. Además, diferentes datos de series de temporales pueden tener diferentes valores límite. Esto implica que es muy difícil y casi imposible obtener un apropiado y razonable valor límite independiente de la dimensión  $d$  y los puntos de cada trayectoria, así como de los datos considerados de la serie de tiempo.

Para evitar el problema anterior, en su lugar definimos la cantidad siguiente, i.e., el valor promedio de todos los  $a(i, d)$ ,

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i, d) \quad (13)$$

$E(d)$  es dependiente solo de la dimensión  $d$  y del retardo  $\tau$ . Para investigar esta variación de  $d$  a  $d+1$ , definimos

$$E1(d) = E(d + 1)/E(d). \quad (14)$$



Encontramos que  $E1(d)$  para de cambiar cuando  $d$  es mayor que algún valor  $d_0$ . Si la serie de tiempo viene de un atractor. Entonces  $d_0 + 1$  es la mínima dimensión de embedding que podemos encontrar.

### 2.5.5 Selección de tiempo de retardo

El parámetro  $\tau$  es necesario y debe darse antes de que la dimensión mínima de inserción (embedding) se determine numéricamente, independientemente de los métodos que se utilicen. A pesar de que en principio la dimensión de embedding es independiente del tiempo de retardo  $\tau$ , la mínima dimensión de embedding es dependiente de  $\tau$  en práctica. Diferentes valores de  $\tau$  pueden conducir a diferentes dimensiones mínimas de embedding, especialmente para series de tiempo continuas (para series de tiempo de mapas de tiempo discreto, la mejor elección de  $\tau$  es 1).

Se puede entender fácilmente que una buena elección de  $\tau$  puede disminuir la dimensión de inserción mínima que es necesaria para la reconstrucción del espacio de fase. Esta es también una de las razones por las que generalmente elegimos un retardo de tiempo tal que cada componente de los vectores reconstruidos (lineales o no lineales) es independiente en la medida de lo posible. Los métodos para elegir tal  $\tau$  son, por ejemplo, el método de función de autocorrelación y de información mutua.

### 2.5.6 Método basado en la función de autocorrelación.

Una forma de elegir el tiempo de rezago se basa en el cálculo de la función de autocorrelación muestral. Esta se define para una serie de tiempo escalar  $x_n$  de  $N$  muestras como:

$$\rho(\tau) = \frac{\sum_{n=1}^N (x_{n+\tau} - \bar{x})(x_n - \bar{x})}{\sum_{n=1}^N (x_n - \bar{x})^2} \quad (15)$$

Donde  $\bar{x}$  denota la media muestral de la serie  $x_n$ . El criterio para la selección de  $\tau_e$  es elegir el tiempo de rezago donde  $\rho(\tau)$  alcance su primer cero. De esta forma  $x_n$  y  $x_{n+\tau_e}$  estarán linealmente incorrelacionados.

### 2.5.7 Ejemplo de la determinación del orden del sistema.

A continuación, se presenta el ejemplo de la aplicación del procedimiento antes descrito, para el cálculo del orden del sistema de la curva de irradiancia.

Los datos que se utilizaron fueron tomados para 24 horas en intervalos de medición de 5 segundos, para el día primero de febrero de 2017, desde el centro de medición de la facultad de Ingeniería UADY.

En la figura Figura 14 se presenta la gráfica de los datos de irradiancia previamente mencionados.

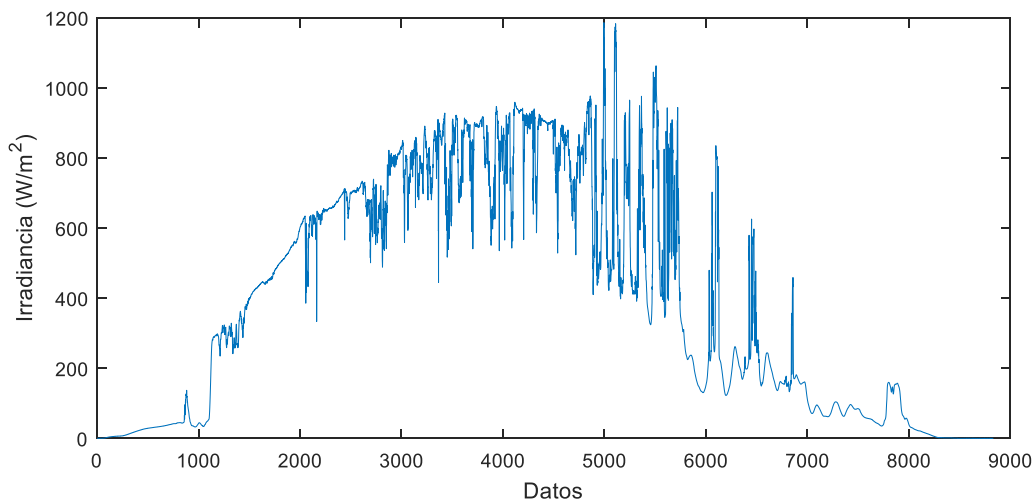


Figura 14. Datos de irradiancia.

Con estos datos se obtiene el parámetro  $\tau$  o tiempo de retardo, con base en el método de la función de autocorrelación.

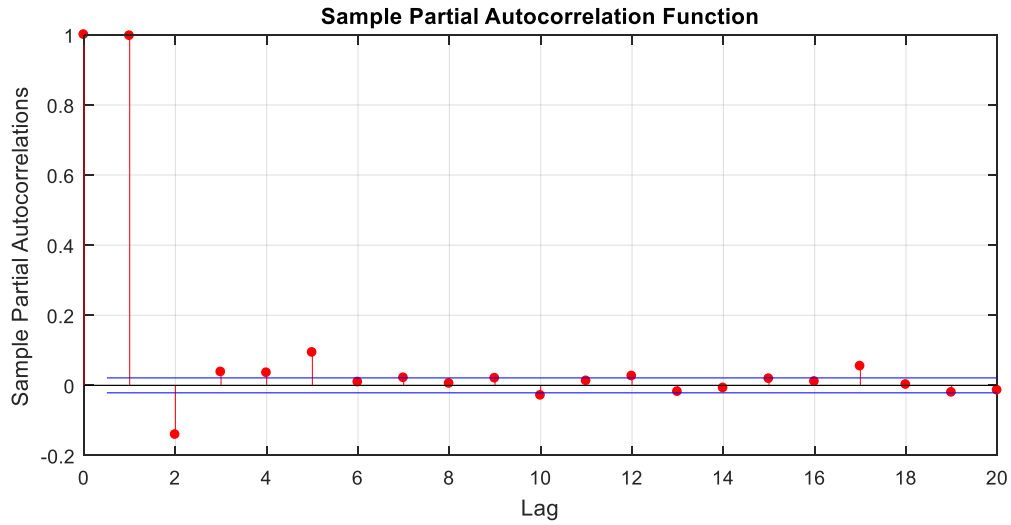


Figura 15. Función de autocorrelación.

El tiempo de retardo obtenido fue de 6, con el procedemos a calcular la dimensión de embedding del sistema a través del método de Cao, el resultado final es una dimensión de 7 como puede verse en la Figura 16

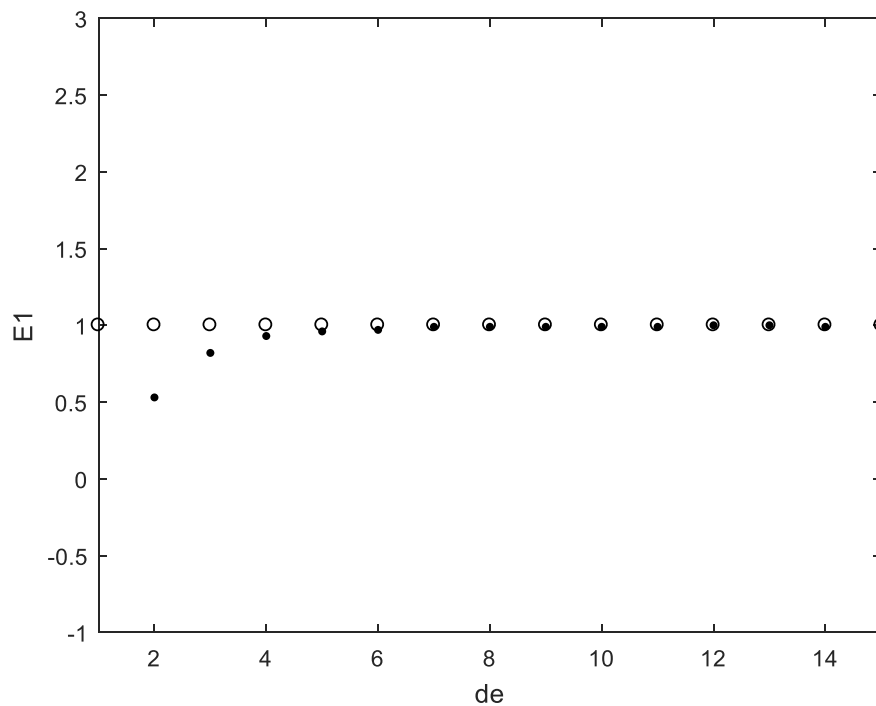


Figura 16. Dimensión de embedding de irradiancia.

## 2.6 Aprendizaje en Redes Neuronales

El conocimiento se refiere a la información almacenada o a modelos y procedimientos usados por una persona o máquina para interpretar, predecir y responder apropiadamente a los estímulos del medio ambiente. En la representación del conocimiento se considera qué información está contenida y cómo está codificada. La red neuronal construye un modelo de la interacción con el medio ambiente, por medio del aprendizaje de relaciones entrada-salida. Este modelo debe ser compatible con el mundo real. La información puede adquirirse por conocimiento a priori o por mediciones obtenidas por sensores. Usualmente la utilización de una red neuronal se divide en:

- a) Aprendizaje por ejemplos (entrenamiento).
- b) Aplicación (generalización).

El aprendizaje por ejemplos es el proceso por el cual los parámetros de la red neuronal (pesos sinápticos) se adaptan. El aprendizaje implica la siguiente secuencia de eventos:

- La red neuronal es estimulada por el medio ambiente.
- La red neuronal ajusta los parámetros.
- La red neuronal genera una nueva respuesta.

Existen dos tipos principales de aprendizaje: aprendizaje supervisado y aprendizaje no supervisado. Se describe únicamente el tipo de aprendizaje supervisado, por ser éste el que se aplicó en el presente trabajo. Los pasos de este tipo de aprendizaje enfocado en la identificación de un sistema dinámico son:

- Disponer de un conjunto de datos de entrenamiento de longitud  $M$ ,  $\{x_i; d_i\}_{i=1}^M$ , donde  $x_i$  es el vector de entrada y  $d_i$  es el vector de la respuesta deseada al vector de entrada.
- Se introducen las entradas  $x_i$  y se espera que la red responda.
- La red responde mediante un vector de salida  $y_i$ .

- Se comparan las señales de la respuesta deseada  $d_i$  y la respuesta de la red  $y_i$ , creando una señal de error  $e_i = d_i - y_i$ .
- Con la señal de error  $e_i$  se ajustan los parámetros de la red neuronal mediante algún algoritmo de aprendizaje. Un algoritmo de aprendizaje es el conjunto de reglas empleadas para determinar el incremento en los pesos sinápticos de las neuronas.

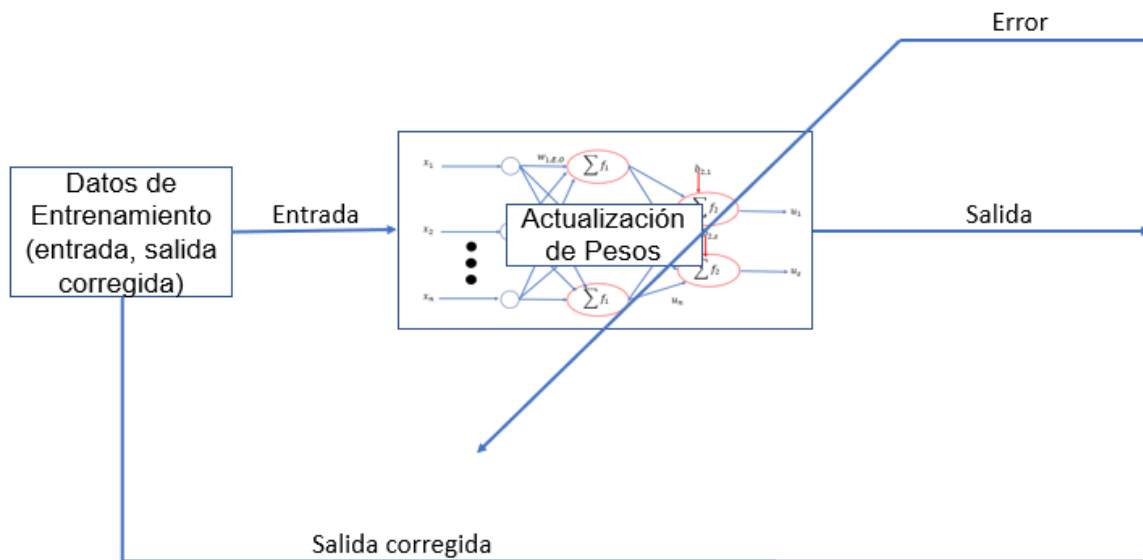


Figura 17. Teoría de aprendizaje.

La secuencia completa de los  $M$  pares entrada-respuesta deseada es conocida como una época. El aprendizaje se detiene cuando la red alcanza algún criterio de error definido antes de iniciar su entrenamiento o cuando se alcanza un número de épocas preestablecido. El algoritmo de aprendizaje empleado en esta tesis está basado en el filtro de Kalman, el cual se describe en la siguiente Sección.

## 2.7 Filtro de Kalman

El filtro de Kalman es un algoritmo que se basa en el modelo de espacio de estados de un sistema para estimar el estado futuro y la salida futura realizando un filtrado óptimo a la señal de salida, y dependiendo del retraso de las muestras que se le ingresan puede cumplir la función de estimador de parámetros o únicamente de filtro. Pero en ambos casos elimina ruido, estas ecuaciones son ampliamente utilizadas ya que incluyen probabilidades estadísticas puesto que toma en cuenta la aleatoriedad tanto de la señal como del ruido. A diferencia de otros tipos de filtros este no requiere de una frecuencia de corte específica debido a que se basa en la característica del ruido permitiendo de esta manera filtrar en todo el espectro de frecuencias. Además, sus ecuaciones solo dependen de una muestra anterior y la muestra presente lo que permite un ahorro considerable de memoria a la hora de ser implementado en un sistema digital y su fácil programación lo hacen muy atractivo ya que se basa en un método recursivo. (Cárdenas, Arias, & Bravo, 2013)

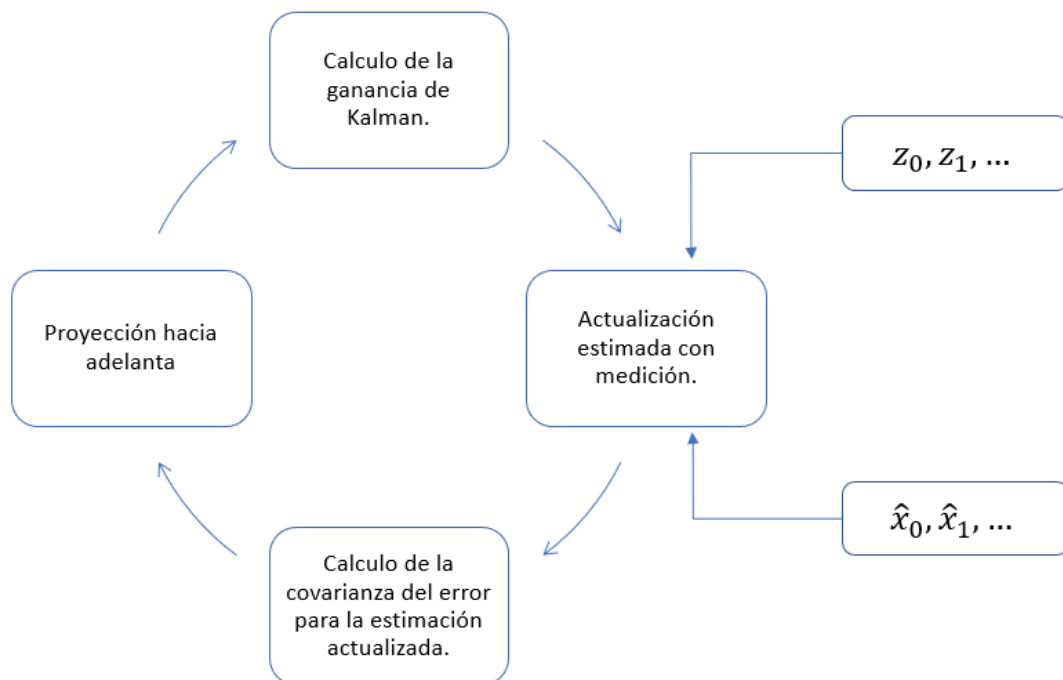


Figura 18. Ciclo del Filtro de Kalman.

Los principales usos del filtro de Kalman han sido en sistemas de control modernos, en el rastreo y navegación de todo tipo de vehículos y en el diseño predictivo de sistemas de estimación y control. Estas actividades técnicas fueron posibles gracias a la introducción del filtro de Kalman (Moreno, 2017). En la aplicación del filtro de Kalman al entrenamiento de redes neuronales, los pesos sinápticos de la red neuronal son los estados que el filtro de Kalman estima, y la salida de la red es la medición usada por el filtro de Kalman. Es por ello que surge la alternativa de visualizar el entrenamiento de redes neuronales como un problema de filtrado óptimo.

Considérese el sistema dinámico lineal, en tiempo discreto, mostrado en la Figura 19. El concepto de estado es fundamental para esta descripción. El vector de estado o simplemente estado, denotado por  $w(k)$ , es definido como el conjunto mínimo de datos suficientes para describir el comportamiento dinámico del sistema; el subíndice  $k$  denota el instante de muestreo. En otras palabras, el estado es la cantidad mínima de datos del comportamiento pasado del sistema que es necesaria para predecir el comportamiento futuro. Normalmente el vector de estados  $w(k)$  es desconocido. Para estimarlo, se usa un conjunto de datos medidos, denotados por  $y(k)$ . EN términos matemáticos, el diagrama a bloques de la figura Figura 19 involucra el siguiente par de ecuaciones:

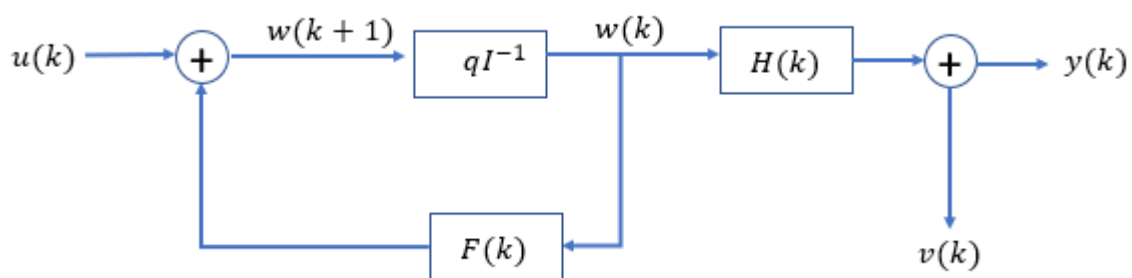


Figura 19. Diagrama de bloques Filtro de Kalman.

Ecuación del proceso

$$w(k + 1) = F_{k+1,k} w(k) + u(k) \quad (16)$$

Donde  $F_{k+1,k} \in \mathcal{R}^{L \times L}$  es la matriz de transición de estados de la iteración  $k$  a la iteración  $k+1$ ;  $w(k) \in \mathcal{R}^L$  es el vector de estados de la iteración  $k$ , y  $u(k) \in \mathcal{R}^L$  representa el vector del ruido del proceso (gaussiano con media cero).

La matriz de covarianza del ruido de proceso se define como:

$$E\{u(n)u^T(k)\} = \begin{cases} Q(k) & \text{para } n = k \\ 0 & \text{para } n \neq k \end{cases} \quad (17)$$

Ecuación de salida

$$y(k) = H(k)w(k) + v(k) \quad (18)$$

Donde  $H(k) \in \mathcal{R}^{m \times L}$  es la matriz de medición;  $y \in \mathcal{R}^m$  es el vector de salida del proceso (medible en la iteración  $k$ ), y  $v \in \mathcal{R}^m$  representa el vector del ruido de medición (gaussiano con media cero).

La matriz de covarianza del ruido de medición se define como:

$$E\{v(n)v^T(k)\} = \begin{cases} R(k) & \text{para } n = k \\ 0 & \text{para } n \neq k \end{cases} \quad (19)$$

Además, el ruido de medición está descorrelacionado del ruido del proceso.

Entonces el vector de estados  $w(k)$  del sistema lineal se puede estimar con el filtro Kalman, el cual queda definido por las siguientes ecuaciones:

Propagación del estado estimado

$$\hat{w}^-(k) = F_{k+1,k} \hat{w}^-(k-1) \quad (20)$$

Propagación de la covarianza del error

$$P^-(k) = F_{k+1,k} P(k-1) F^T(k, k-1) + Q(k-1) \quad (21)$$



Matriz de ganancia de Kalman

$$K(k) = P^-(k)H^T(k)[H(k)P^-(k)H^T(k) + R(k)]^{-1} \quad (22)$$

Actualización del estado estimado

$$\hat{w}(k) = \hat{w}^-(k) + K(k)(y(k) - H(k))P^-(k) \quad (23)$$

Actualización de la covarianza del error

$$P(k) = (I - K(k)H(k))P^-(k) \quad (24)$$

Inicialización para  $k = 0$

$$\hat{w}(0) = E\{w(0)\} \quad (25)$$

$$P(0) = E\{(w(0) - E\{w(0)\})(w(0) - E\{w(0)\})^T\} \quad (26)$$

## 2.8 Filtro Extendido de Kalman

Considere un sistema dinámico no lineal descrito por el siguiente modelo en espacio de estado

$$w(k+1) = f(k, w(k)) + u(k) \quad (27)$$

$$y(k) = h(k, w(k)) + v(k) \quad (28)$$

Como antes,  $u(k)$  y  $v(k)$  son ruidos independientes, blancos, gaussianos, con media cero y con matrices de covarianza  $Q(k)$  y  $R(k)$ , respectivamente.  $f(k, w(k))$  denota la función matricial no lineal de transición, la cual puede ser variante en el tiempo y  $h(k, w(k))$  denota la función matricial de medición no lineal, que también puede ser variante en el tiempo.

La idea básica del filtro extendido de Kalman es linealizar el modelo en espacio de estado de ( 27 y 28 a cada instante de muestreo alrededor del estado estimado más reciente, el cual puede ser tomado como  $\hat{w}(k)$  o  $\hat{w}^-(k)$ . Una vez obtenido el modelo lineal, se aplican las ecuaciones del filtro de Kalman de la Sección anterior. Explícitamente, la aproximación procede en dos etapas:

Etapla 1: se construyen las siguientes matrices

$$F_{k+1,k} = \left. \frac{\partial f(k, w(k))}{\partial w} \right|_{w=\hat{w}(k)} \quad (29)$$

$$H(k) = \left. \frac{\partial h(k, w(k))}{\partial w} \right|_{w=\hat{w}^-(k)} \quad (30)$$

Etapla 2. Una vez que las marices son evaluadas, se emplean en una aproximación en series de Taylor de primer orden para las funciones no lineales  $f(k, w(k))$  y  $h(k, w(k))$  alrededor de  $\hat{w}(k)$  y  $\hat{w}^-(k)$  respectivamente. Específicamente, se aproximan como sigue:

$$f(k, w(k)) \approx F(w, \hat{w}(k))F_{k+1,k}(w, \hat{w}(k)) \quad (31)$$

$$h(k, w(k)) \approx H(w, \hat{w}(k)) + H_{k+1,k}(w, \hat{w}^-(k)) \quad (32)$$

Con esto, las ecuaciones de estado no lineales se aproximan como

$$w(k+1) \approx F_{k+1,k}w(k) + u(k) + d(k) \quad (33)$$

$$\bar{y}(k) \approx H(k)w(k) + v(k) \quad (34)$$

Donde

$$\bar{y}(k) = y(k) - (h(w, \hat{w}^-(k)) - H(k)\hat{w}^-(k)) \quad (35)$$

$$d(k) = f(w, \hat{w}(k)) - F_{k+1,k}\hat{w}(k) \quad (36)$$

(Camperos & García, 2006)

## 2.9 Análisis Wavelet

La composición de una función específica en forma de ondas senoidales a diferentes frecuencias y amplitudes pueden ser derivadas usando la transformada de Fourier. A pesar de que el análisis de Fourier es realmente adecuado para los conjuntos de datos estacionarios, no podrá reconstruir las señales con un alto nivel de diferencia de frecuencias. En el análisis de wavelets, se pueden definir los cambios de función de un período de tiempo al siguiente. Estos cambios pueden ser encontrados al hacer coincidir una función wavelet y variando escalas y posiciones de esta función. En el análisis de wavelet es posible elegir una función de wavelet específica que tenga la mayor similitud con la función deseada. Mientras que, en la transformada de Fourier clásica, solo se pueden usar ondas sinusoidales o coseno. La función  $\varphi(\cdot)$  se llama wavelet madre; Con la traslación y escalamiento de esta wavelet madre se produce los miembros de la familia.

$$\varphi_{\lambda,t}(x) = \frac{1}{\sqrt{\lambda}} \varphi\left(\frac{x-t}{\lambda}\right) \quad (37)$$

Donde  $\lambda > 0$  es escalamiento y este es un número finito,  $t$  es traslación y  $x$  es la señal (Yashar, Palizdan, Huang, & Lee, 2014).

### 2.9.1 Transformada Wavelet Continua (CWT)

En la transformada wavelet continua, la función es transformada por wavelets continuamente a lo largo del tiempo. Como resultado, ambos parámetros  $\lambda$  y  $t$  varían con el tiempo para producir la familia de wavelet. “La idea de la transformación es que para un escalamiento  $\lambda$  y traslación  $t$  de la wavelet madre ( $\varphi$ ), el coeficiente de amplitud el cual hace que  $\varphi_{\lambda,t}$  se ajuste mejor a la señal  $X(\cdot)$  sea calculado “ (Veitch, 2005).

$$\langle x, \varphi_{\lambda,t} \rangle = \int_{-\infty}^{+\infty} \varphi_{\lambda,t}(x) X(x) dx \quad (38)$$

CWT representa toda la información de la señal  $X(\cdot)$ . Así que  $X(\cdot)$  puede ser recuperado usando la transformación inversa:

$$X(t) = \frac{1}{C_\varphi} \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} \langle x, \varphi_{\lambda,t} \rangle \varphi_{\lambda,t}(t) dx \right] \frac{d\lambda}{\lambda^2} \quad (39)$$

$$C_\varphi = \int_{-\infty}^{+\infty} \frac{|\varphi(f)|^2}{f} df \quad (40)$$

Es claro que CWT y  $X(\cdot)$  son diferentes representaciones de la misma información. Pero CWT nos permite conocer más detalles (Yashar, Palizdan, Huang, & Lee, 2014).

### 2.9.2 Transformada Wavelet Discreta (DWT)

CWT produce información valiosa, pero calcular un alto número de dilaciones y traslaciones es un proceso complejo. Si nosotros seleccionamos algunas muestras de la CWT para representar las características clave de la función, entonces tendremos una transformada wavelet discreta (DWT). DWT se realiza en una serie de tiempo  $x(\cdot)$  la cual es definida en un dominio de tiempo finito  $t = 0, 1, \dots, N - 1$ . Las escalas diádicas

se utilizan habitualmente para la dilación  $\lambda(2^{f-1}, j = 1, 2, 3, \dots)$  y valores de traslación correspondientes al intervalo  $2^f$ . En el nivel  $j$  el número de coeficientes wavelet son:  $N_j = N/2^f$  y tiempo asociado  $t = (2n + 1)2^f - 1/2$ ,  $n = 0, 1, \dots, N_{j-1}$ . En DWT, la combinación lineal de una familia de wavelet elegida puede usarse para realizar una función específica (Yashar, Palizdan, Huang, & Lee, 2014).

## 2.10 Redes Neuronales Wavelet (WNNs)

La red neuronal wavelet (WNN) intenta combinar los atributos de las redes neuronales con las características de la transformada wavelet. Por lo general, es una red neuronal feed-forward con una capa oculta y una familia de wavelets como función de activación. Las neuronas en la capa oculta se llaman wavelones. En la bibliografía se suelen sugerir algunas wavelets madre de forma común: la derivada Gaussiana, la segunda derivada Gaussiana, el denominado “Mexican Hat” y la wavelet Morlet.

La selección de la wavelet madre depende de la aplicación y no se limita a las opciones anteriores. La función de activación puede ser un wavenet (wavelets ortogonales) o frame wave (wavelets continuas). En el presente trabajo se utiliza una función wavelet madre Mexican hat, la cual funciona de manera satisfactoria en varias aplicaciones y viene dada por:

$$\varphi(z_{ij}) = 1 - z_{ij}^2 e^{(-\frac{1}{2}z_{ij}^2)} \quad (41)$$

$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}} \quad (42)$$

En la expresión anterior,  $i = 1, \dots, m, j = 1, \dots, k + 1$  y el peso  $w$  corresponde a los factores de traslación  $w_{(\xi)ij}^{[1]}$  y escalamiento  $w_{(\zeta)ij}^{[1]}$ . El vector completo de los parámetros de la red comprende:  $w = (w_i^{[0]}, w_j^{[2]}, w_{k+1}^{[2]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]})$ . Estos parámetros son ajustados durante la fase de entrenamiento.

Las wavelets multidimensionales son calculadas por:

$$\Psi_j(x) = \prod_{i=1}^m \psi(z_{ij}) \tag{43}$$

Donde  $\psi$  es la wavelet madre y con los mismos parámetros de  $i$  y  $m$  descritos con anterioridad (Alexandridis & Zapranis, 2012).

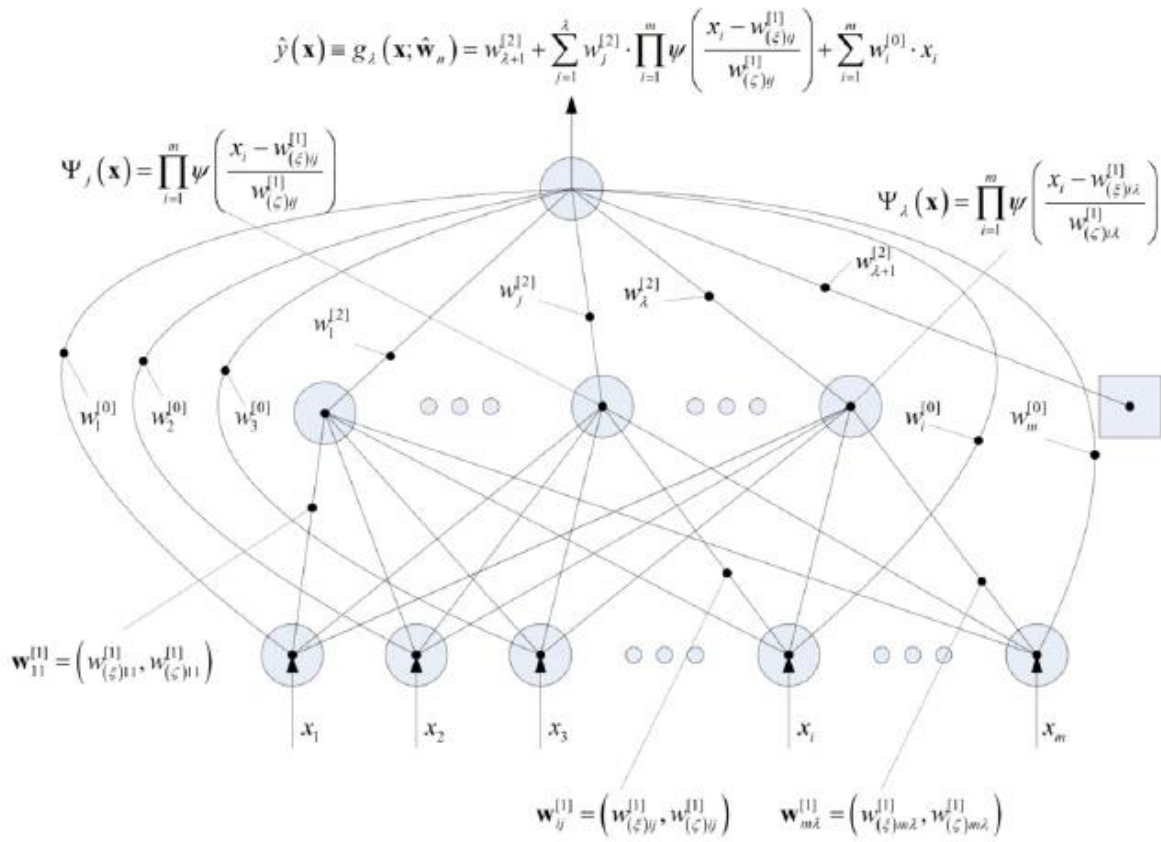


Figura 20. Ejemplo de estructura de red neuronal wavelet.

## 2.11 Modelos Estadísticos

### Algoritmo ARIMA

El promedio móvil autorregresivo integrado (ARIMA) es un procedimiento popular entre los modelos estadísticos para análisis de series de tiempo y aplicaciones de pronóstico (Aznarte, Benítez Sánchez y Lugilde, et al., 2007; Chen, Wang y Huang, 1995; Ture & Kurt, 2006), que fue presentado por Box y Jenkins (1976). El modelo ARIMA se origina del modelo autorregresivo (AR), modelo de promedio móvil (MA), y la combinación de modelos AR y MA (ARMA).

### Modelo AR

En el modelo AR (p), el valor actual de la serie de tiempo se expresa como un agregado lineal de valores p previos y un choque aleatorio como

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (44)$$

Donde

$p$  es el orden del operador de AR(p).

$Z_t$  es la serie temporal observada de carga en el tiempo  $t$ .

$a_t$  es el choque aleatorio en una distribución normal con media cero y varianza en el tiempo  $t$ .

$\phi_1, \dots, \phi_p$  son parámetros de AR(p).

Introduciendo el operador de cambio de retroceso  $B$ , que define  $Z_{t-1} = BZ_t$ , y, consecuentemente,  $Z_{t-p} = B^p Z_t$ , la ecuación (44) puede ser escrita como

$$\phi_p(B)Z_t = a_t \quad (45)$$

Donde  $\phi_p(B)$  es el operador AR(p), el cual es definido como

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (46)$$

### Modelo MA

El modelo MA (q), que expresa el valor actual de una serie temporal como un valor actual y los q valores previos de los choques aleatorios, se escribe como

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (47)$$

O

$$Z_t = \theta_q(B) a_t \quad (48)$$

Donde

q es el orden de MA(q).

$\theta_1, \dots, \theta_q$  son parámetros del modelo MA(q).

$\theta_q(B)$  es el operador MA(q), el cual es definido como

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (49)$$

### Modelo ARMA

Para aumentar la flexibilidad al ajustar series de tiempo reales, las ecuaciones (42 a (47) e combinan para obtener un modelo más general ARMA(p,q), el cual es expresado como

$$\phi_p(B) Z_t = \theta_q(B) a_t \quad (50)$$



## Modelo ARIMA

En general, los modelos de series de tiempo AR, MA y ARMA son estacionarios; sin embargo, si las series no son estacionarias, estas series se transforman en series temporales estacionarias usando el proceso de diferencia. A medida que se completa la estabilización modelada por el modelo ARMA (p, q), el modelo ARIMA (p, d, q) se obtiene de la siguiente manera:

$$\phi_p(B)\Delta^d Z_t = \theta_q(B)a_t \quad (51)$$

Donde  $\Delta^d Z_t = (1 - B)^d Z_t$  Y d es el orden de la diferenciación.

Cuando una serie temporal muestra una estacionalidad potencial indexada por s, es ventajoso utilizar un modelo estacional multiplicado ARIMA (p, d, q)(P, D, Q)<sub>s</sub>. Las series temporales estacionales se transforman en series temporales estacionarias con componentes de tendencias no periódicas. Un modelo ARIMA estacional multiplicado se puede expresar como

$$\phi_p(B)\Phi_p(B^s)\Delta^d\Delta_s^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (52)$$

Donde  $\Delta_s^D Z_t = (1 - B^s)^D Z_t$ , D es el orden de la diferencial estacionaria,  $\Phi_p(B^s)$  y  $\Theta_Q(B^s)$  son los operadores estacionales de AR(p) y MA(p), respectivamente, los cuales son definidos como

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps} \quad (53)$$

$$\Theta_Q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs} \quad (54)$$

Donde  $\Phi_1, \dots, \Phi_p$  son los parámetros de modelo estacional AR(p),  $\theta_1, \dots, \theta_Q$  son los parámetros del estacional MA(q).

## Procedimiento Modelo ARIMA

El procedimiento de modelado ARIMA, desarrollado por Box y Jenkins (1976), consta de tres pasos iterativos: identificación del modelo; estimación de parámetros; y, comprobación de diagnóstico. La Figura 21 representa el proceso de modelado ARIMA. El procedimiento del modelo ARIMA se detalla de la siguiente manera:

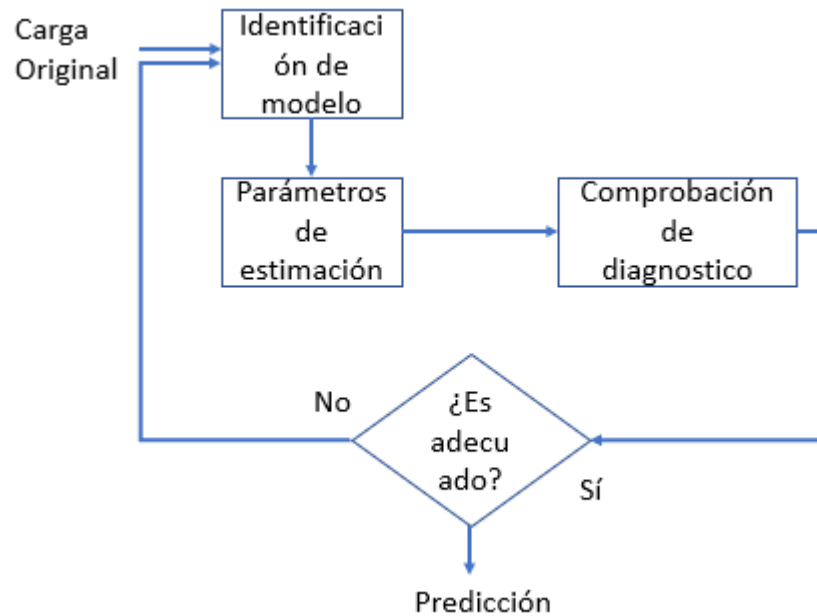


Figura 21. Procedimiento Modelo ARIMA

Paso 1:

Paso 1: identificación del modelo.

Uno determina si la serie de tiempo es estacionaria o no estacionaria. Si la serie temporal no es estacionaria, se transforma en una serie temporal estacionaria utilizando un grado adecuado de diferenciación. El modelo provisional de la serie temporal observada se conjetura a partir de gráficos, estadísticas, una función de autocorrelación (ACF) y una función de autocorrelación parcial (PACF) de series de tiempo. Este proceso requiere habilidad y experiencia. En general, se elige más de un modelo tentativo en este paso.

Paso 2: estimación de parámetros.

Una vez que se formula el modelo tentativo, los parámetros del modelo relacionado se estiman utilizando el esquema de mínimos cuadrados. Se estima que los parámetros tienen un gradiente cero de errores de pronóstico para los datos de carga históricos. El objetivo principal de esta estimación de parámetros es minimizar el error de pronóstico y determinar tanto el orden del modelo como sus parámetros.

Paso 3: comprobación de diagnóstico.

Cuando los parámetros han sido bien estimados, la precisión del modelo tentativo se valida al examinar los residuos ACF y PACF. Los residuos deben simular el proceso de ruido blanco. Además, la prueba Q-statistics se aplica para confirmar el modelo tentativo (O'Donovan, 1983). Si el valor calculado Q excede el valor crítico de  $\chi^2$  obtenido de las tablas chi-cuadradas, el modelo tentativo es inadecuado.

Estos pasos se repiten hasta que se identifica un modelo adecuado.

### 3. Metodología

Nuestra metodología consistió en la recolección de datos de un campo eólico, después se realizó el procesamiento y promediado de los datos obtenidos, calculamos el orden del sistema y se evalúa en diferentes modelos de predicción para finalmente compararlos.

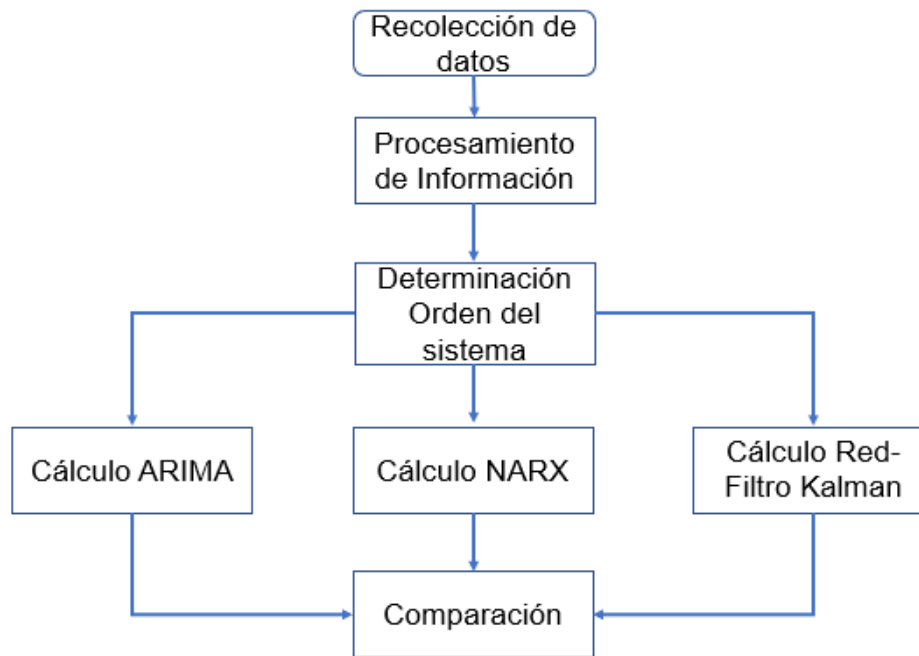


Figura 22. Secuencia de la Metodología

#### 3.1 Procesamiento de Datos

Se obtuvieron datos de velocidad de viento y potencia generada en un campo eólico ubicado en Baja California Sur. A continuación, se listan las características de los datos con los que se realizó el presente trabajo.

Tabla II. Características de los datos empleados

<b>Características</b>	<b>Valor</b>
Altura	60m
Periodo	Febrero – Agosto 2015
Tiempo de muestreo	5 min
Ubicación	Baja California Sur



Figura 23. Ubicación del Campo Eólico



Figura 24. Campo eólico en Baja California Sur

Se obtuvieron datos de velocidad y potencia generada en un campo eólico para los meses de febrero a agosto de 2015, con mediciones cada 5 minutos, esos datos se promediaron y obtuvieron para datos de cada 15 minutos y tener menos tiempo de computo a la hora de predicciones a 1 día. En la Figura 25 se presenta la gráfica con las mediciones de potencia en el periodo ya mencionado.

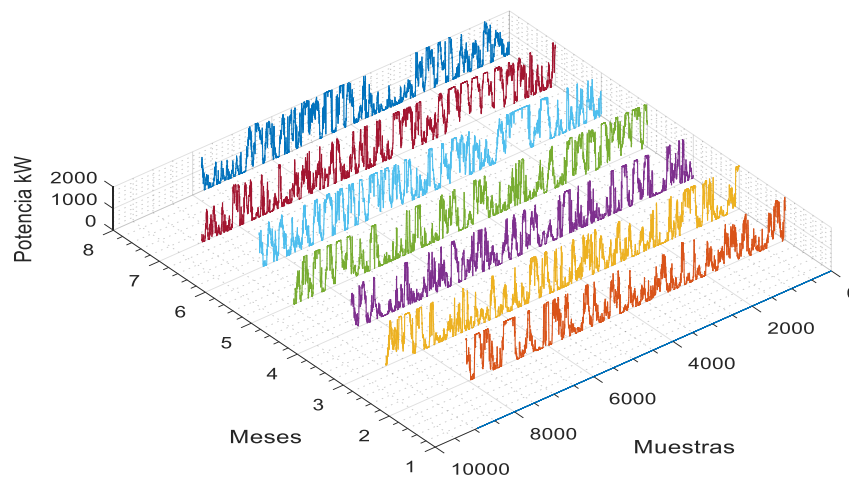


Figura 25. Datos de potencia eólica.

### 3.2 Determinación del Orden del Sistema

A continuación, se procede con la reconstrucción de las dinámicas inmersas en la serie de tiempo. Como se mostró en capítulos anteriores, para este propósito se deben calcular dos parámetros específicos, el primero es el rezago óptimo de embebimiento  $\tau_e$ , y el segundo es la dimensión de embedding  $d_e$ .

Para el cálculo del rezago óptimo de embebimiento  $\tau_e$  nosotros utilizamos la función de autocorrelación, en ella se selecciona el primer valor cercano a cero. En nuestro caso dividimos la serie de tiempo por meses de febrero a agosto y calculamos el rezago para cada mes.

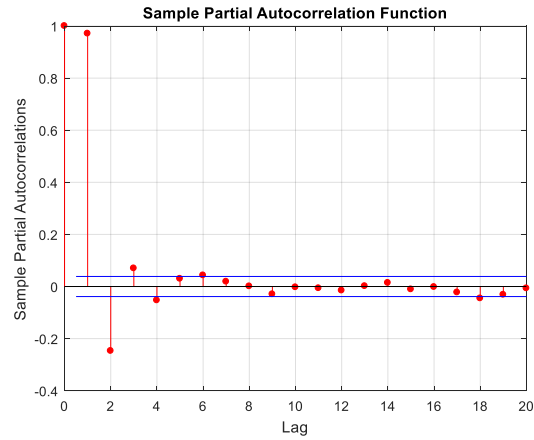


Figura 26. Función de autocorrelación, mes de febrero.

Tabla III. Rezago óptimo de embedimiento.

Mes	Dimensión ( $\tau$ )
Febrero	5
Marzo	4
Abril	5
Mayo	6
Junio	7
Julio	6
Agosto	4

Ahora, para el cálculo de la dimensión de embedding se utiliza el algoritmo propuesto por Cao, el cual usa el valor de  $\tau$ . La figura x muestra la convergencia del método para la serie de tiempo del mes de febrero.

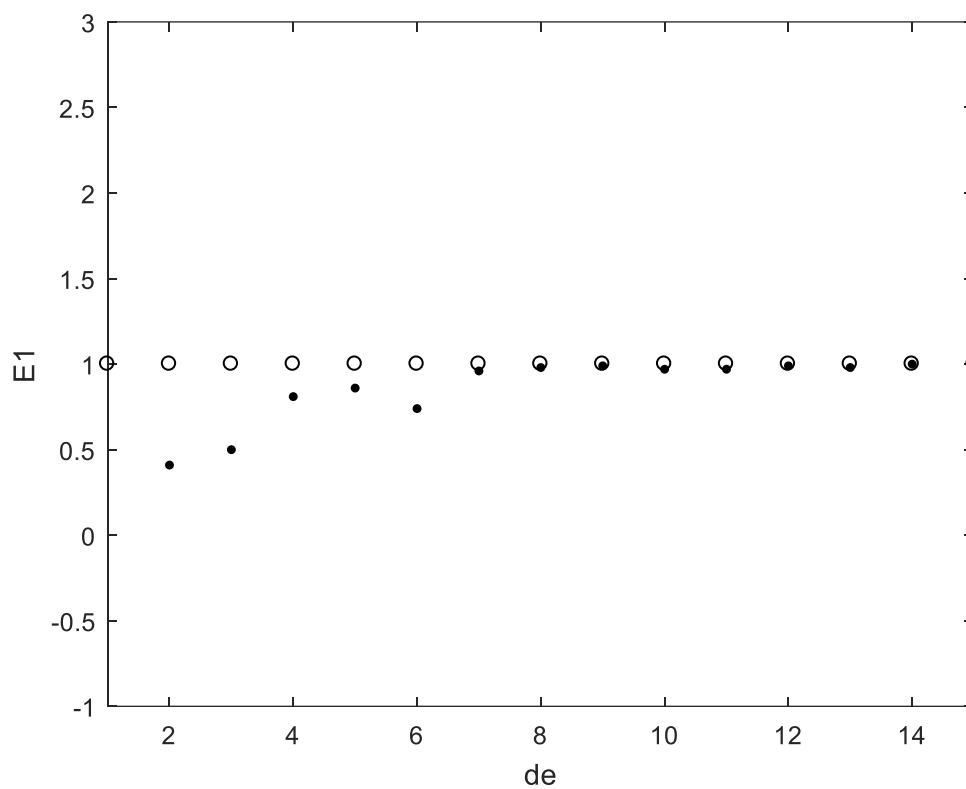


Figura 27. Convergencia de la dimensión de embedding por el método de Cao para el mes de febrero.

Tabla IV. Embedding para los datos de potencia.

Mes	Embedding para serie de Potencia
Febrero	8
Marzo	8
Abril	9
Mayo	8
Junio	9
Julio	8
Agosto	8



### 3.3 Prueba Red Neuronal NARX

Para la demostración de la dimensión de embedding como el rezago optimo en los modelos de predicción, se procedió al cálculo por medio del método de cao y luego se verifico en la red neuronal NARX

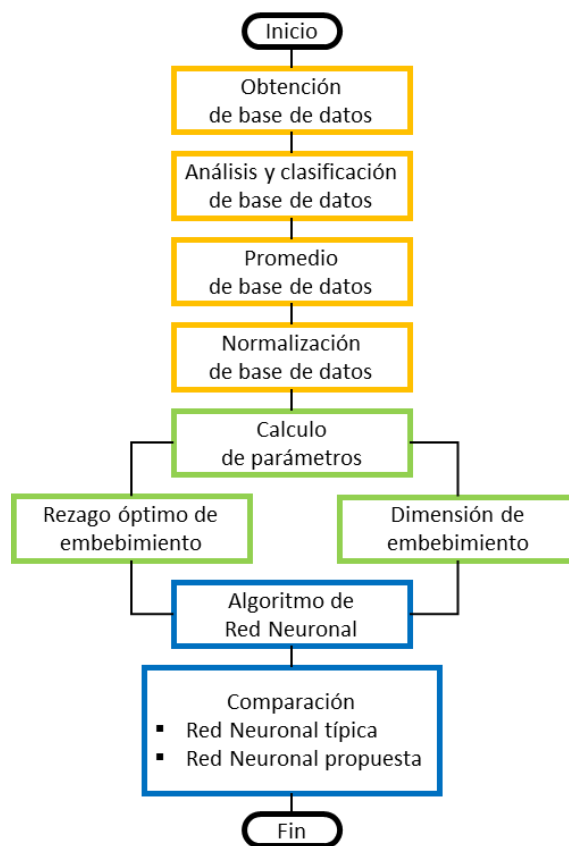


Figura 28. Diagrama de flujo de la prueba en red Neuronal NARX

Se realizó una comparación de un modelo neuronal de 10 neuronas en la capa oculta, variando el número de retardos, la dimensión de embedding de la serie de tiempo se incluye dentro de las pruebas del modelo neuronal, para demostrar lo reportado en la literatura, con esto comprobamos que la dimensión de embedding es el rezago optimo que debemos ingresar en el modelo de predicción.

Tabla V. Respuesta del Modelo Neuronal empleando diferentes retardos

No. de retardos	Regresión Lineal	Épocas
4	$9.73 \times 10^{-1}$	20
5	$9.70 \times 10^{-1}$	11
6	$9.70 \times 10^{-1}$	6
7	$9.70 \times 10^{-1}$	13
8	$9.73 \times 10^{-1}$	8
9	$9.71 \times 10^{-1}$	25
10	$9.73 \times 10^{-1}$	25

Tabla VI. Parámetros del modelo de la Red Neuronal

Parámetro	Valor
No. de Neuronas Ocultas	10
No. de Elementos de retardo	8
No. de Capas Ocultas	1
Función de Transferencia capa Oculta	Tangente Hiperbólica Sigmoide
Función de Transferencia capa de Salida	Lineal
Algoritmo de Entrenamiento	Levenberg-Marquardt
Máximo número de Épocas	8
Periodo de Muestreo	15 min
Ventana de entrenamiento	126

Estan de acuerdo con la revisión del estado del arte previamente descrito, se demostró que emplear la dimensión de embedding, resulta en una optimización adecuada para una convergencia con menor tiempo de computo, y la mejor correlación posible entre los datos de entrada y de predicción.

### 3.4 Cálculo de redes neuronales y método estadístico

Una vez que hemos comprobado y definido el orden del parámetro de incrustación del sistema podemos proceder a realizar las predicciones por medio del perceptrón multicapa y de la red neuronal wavelet con filtro extendido de Kalman y comparar con la red neuronal NARX previamente descrita y utilizada, así como con el método estadístico ARMA.

Para las redes neuronales artificiales, se utilizará una estructura de red neuronal con una capa oculta de 10 neuronas y un número de retardos (embedding) de 8, determinado con anterioridad en la sección 3.2.

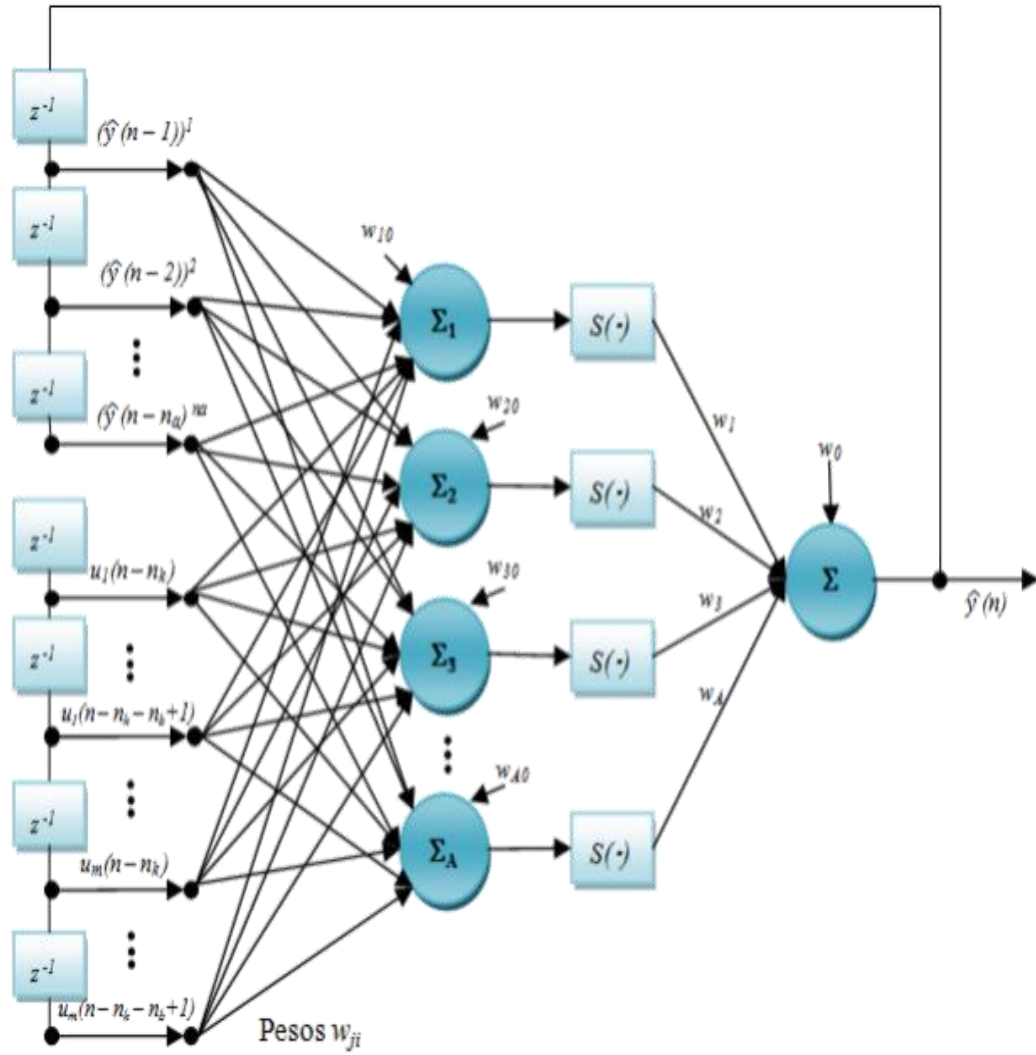


Figura 29. Estructura de red neuronal para el filtro de Kalman

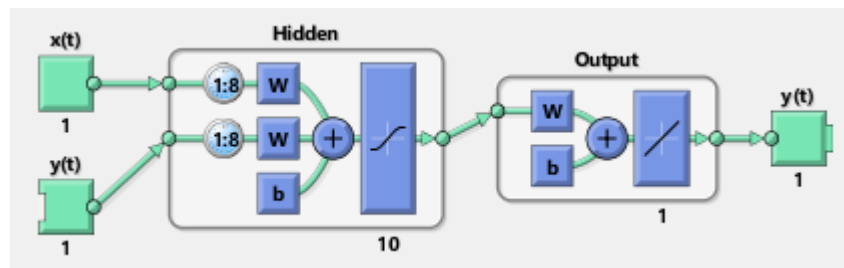


Figura 30. Estructura de red Neuronal NARX.

### 3.5 Criterios Estadísticos

Las ecuaciones estadísticas se utilizan con la finalidad de tener criterios que permitan comparar el desempeño de las arquitecturas de RNAs. A continuación, se describen algunas ecuaciones estadísticas muy utilizadas para comparar grupos de datos generados con redes neuronales.

#### Coeficiente de Correlación

El coeficiente de correlación de Pearson R es una medida de la relación lineal entre dos variables cuantitativas. Este coeficiente es un índice que no depende de las magnitudes de las variables, siempre y estas sean cuantitativas. Matemáticamente se puede relacionar los datos experimentales y simulados con la fórmula:

$$R = \frac{\sum_{i=1}^n (Y_{Ri} - \bar{Y}_R)(Y_{Pi} - \bar{Y}_P)}{\sqrt{\sum_{i=1}^n (Y_{Ri} - \bar{Y}_R)^2 (Y_{Pi} - \bar{Y}_P)^2}} \quad (55)$$

Donde

$$Y_R = \text{Potencia eólica real}$$

$$Y_P = \text{Potencia pronosticada}$$

Estas variables se aplicarán de a las siguientes ecuaciones de la misma manera.

#### Error Cuadrático Medio

El desempeño de estos modelos ha sido evaluado en base al error cuadrático medio (MSE). Cada error o residual se eleva al cuadrado; luego estos valores se suman y se dividen entre el mismo número de observaciones. Este enfoque penaliza los errores mayores de pronóstico ya que eleva cada uno al cuadrado. Este criterio es de los más importantes, pues es preferible un método de pronóstico que produzca errores moderados que otro que por lo regular tenga errores pequeños y ocasionalmente

arroje algunos en extremo grandes. Idealmente se desea que el resultado de EMC sea 0, indicando un pronóstico perfecto. Se define como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_P - Y_R)^2 \quad (56)$$

Esta es una cantidad conocida, calculado dada una muestra particular (y por lo tanto es dependiente de la muestra). Dado que el MSE es la métrica más utilizada para presentar los resultados en los estudios observados, con el fin de permitir algunas comparaciones entre modelos, hemos tratado de describir resultados con respecto a esta métrica.

### **Error porcentual absoluto medio**

El MAPE (Mean Absolute Percent Error, por sus siglas en inglés) mide la magnitud del error en términos porcentuales. Este es calculado como el promedio del error porcentual sin signo. Matemáticamente:

$$MAPE = \frac{1}{n} \sum \frac{|Y_R - Y_P|}{|Y_P|} \quad (57)$$

### **Raíz del error cuadrático medio**

El RMSE (Root Mean Squared Error, por sus siglas en inglés) es una medida del grado de dispersión de los datos con respecto al valor promedio. Es conocida como la desviación estándar para una distribución de probabilidad discreta:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_R - Y_P)^2} \quad (58)$$

### Error Relativo Absoluto

ARE (Absolute Relative Error, por sus siglas en inglés) es una herramienta que comúnmente se utiliza para evaluar los métodos de predicción, la ecuación que permite calcular el ARE es:

$$ARE = \left| \frac{Y_R - Y_P}{Y_R} \right| \quad (59)$$

## 4 Resultados

Como ya se había comentado antes se probaron diferentes escenarios con 4 principales algoritmos de predicción:

Caso 1: redes NARX

Caso 2: Perceptrón multicapa

Caso 3: red neuronal wavelet con filtro de Kalman

Caso 3: método ARMA.

Para el caso de las redes neuronales artificiales, se trabajó con un número estándar de 10 neuronas en la capa oculta.

## 4.1 Red Neuronal NARX

En esta red se trabajó un esquema algoritmo de aprendizaje en el que se manejó el 70% de los datos para el entrenamiento, 15% para la validación y un 15% para el testeo. En la Figura 31 se presenta el resultado de la red neuronal NARX para el mes de febrero, en la Figura 32 se muestran las correspondientes leyendas para cada arreglo de datos.

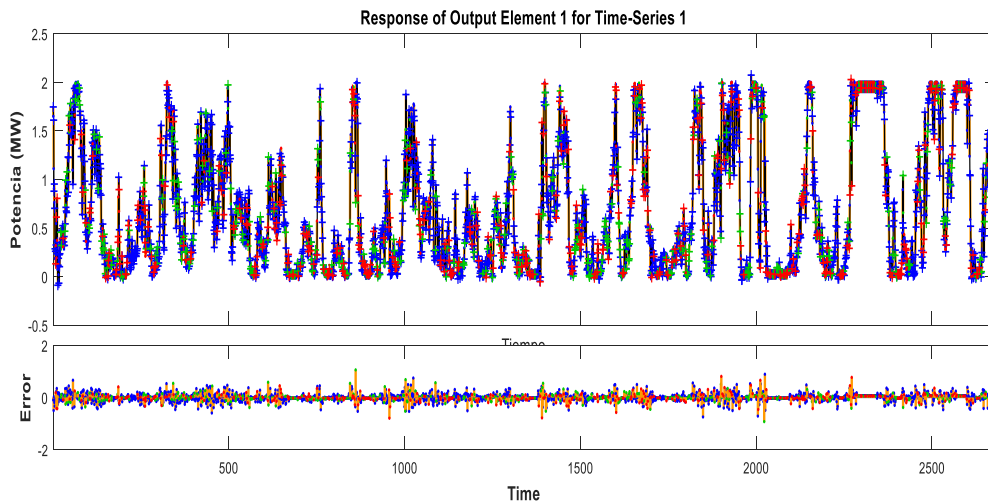


Figura 31. Resultado red NARX (mes de febrero)

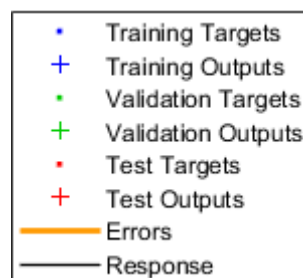


Figura 32. Legendas de gráfica red NARX

En la Figura 33 se presenta la relación para la serie de datos real y de predicción usada en la verificación, se puede apreciar un mejor contraste que en la Figura 31 y las discrepancias en la tendencia. Cabe aclarar que los datos fueron tomados de manera



aleatoria en la serie, por eso se puede ver que abarcan un mayor rango de cobertura en diferentes días del mes y no solamente en una sección de la serie de tiempo.

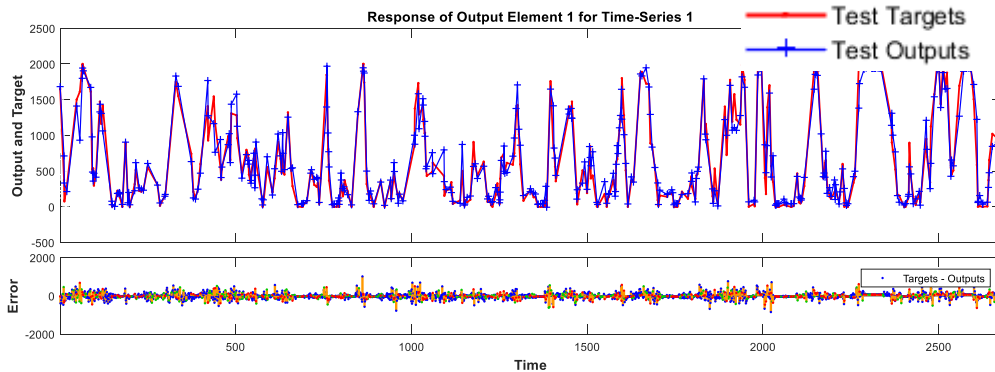


Figura 33. Testing con red NARX

La Figura 34 nos ofrece el resultado del número de épocas que es necesario recorrer en el proceso de aprendizaje para obtener el menor error cuadrático medio posible. De los resultados de las evaluaciones, con los parámetros especificados del número de retardos y de neuronas, obtuvimos el menor número de épocas en los entrenamientos fue de 8 con tiempo de 5.84 segundos.

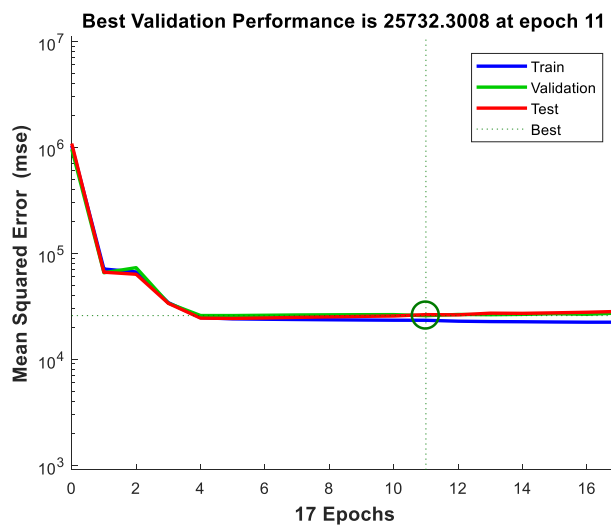


Figura 34. Descenso del error vs épocas

## 4.2 Predicción Perceptrón

Se utilizó un perceptrón multicapa para la predicción de la potencia eléctrica, se emplearon en el entrenamiento 1 capa oculta con 10 neuronas y función de activación tangente sigmoide, en la figura de la( 35 a la( 37 se puede observar un proceso de evolución del entrenamiento, mostrando el proceso de aprendizaje a 5 épocas, 10 y llegando hasta las mil.

En la figura( 38 podemos observar de igual forma como se obtiene el mejor performance de entrenamiento, con un error mínimo desde la época 389.

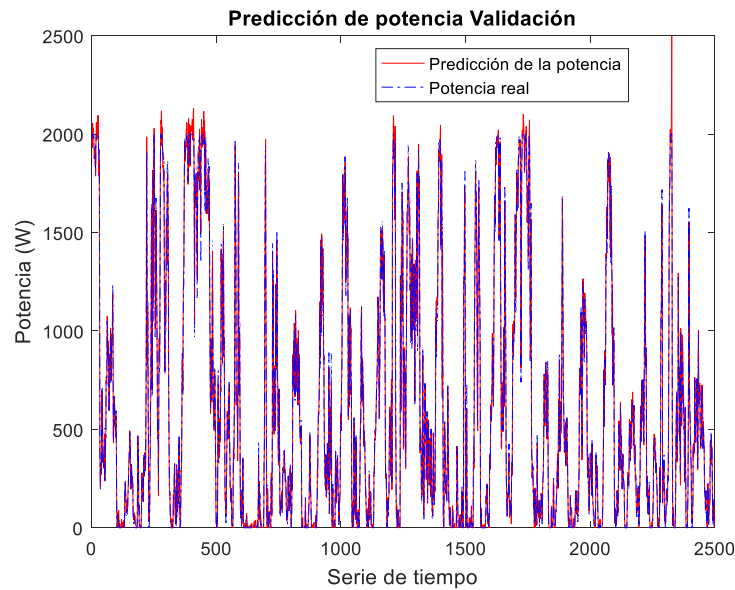


Figura 35. Perceptrón, 5 épocas.

Nkjk

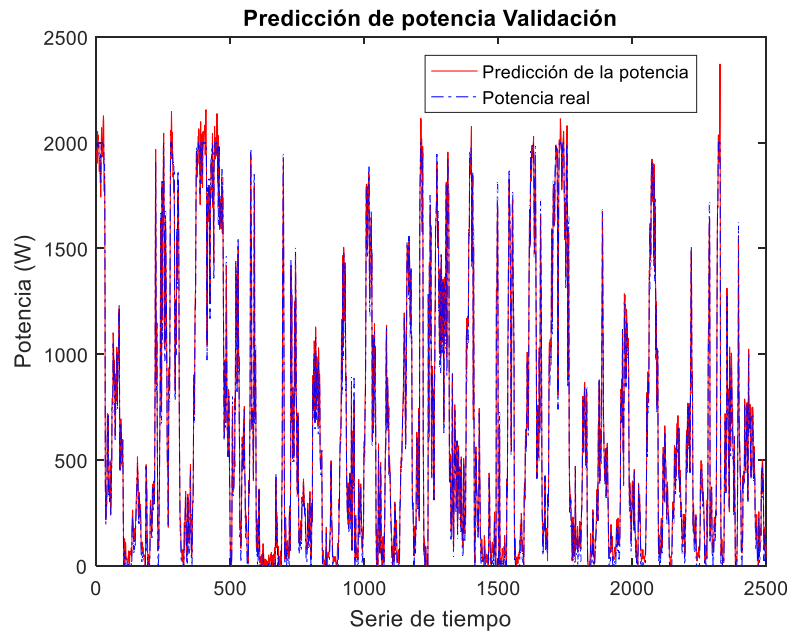


Figura 36. Perceptrón, 10 épocas.

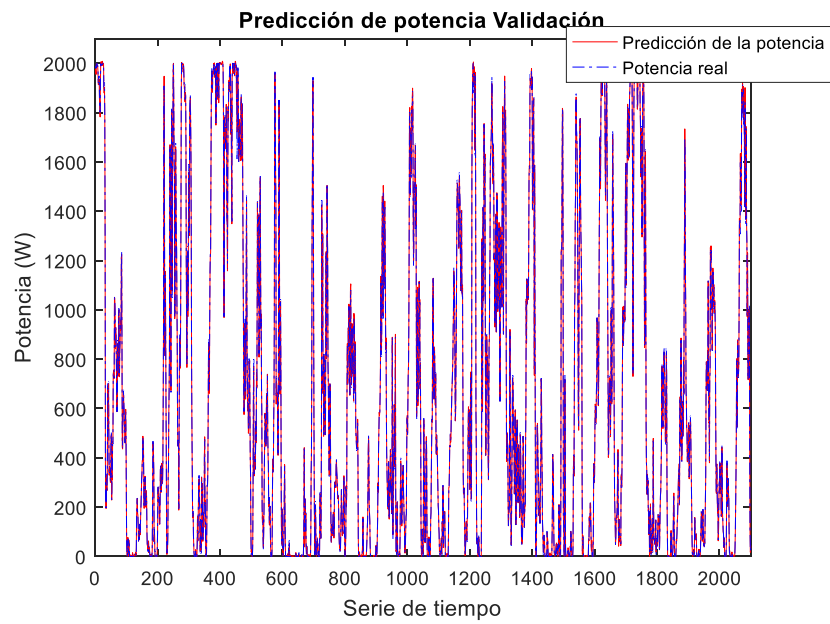


Figura 37. Perceptrón, 1000 épocas.

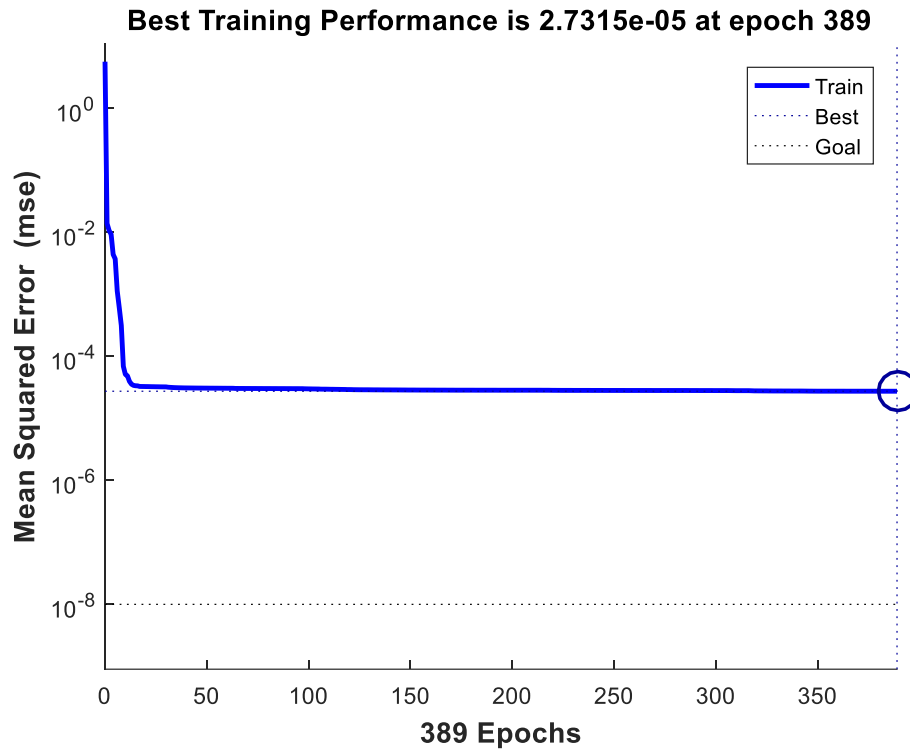


Figura 38. Evolución de entrenamiento.

#### 4. Red Neuronal Wavelet con Filtro Extendido de Kalman (NWFEK)

Para la realización de la predicción por medio de la red neuronal wavelet con filtro extendido de Kalman se procedió a hacer un entrenamiento con los datos obtenidos para el mes de febrero, de aquí se obtuvieron los pesos óptimos para la red neuronal, a través del algoritmo de filtro extendido de Kalman, y se tomaron para utilizarlos en el testing o prueba de la predicción con los datos siguientes.

##### Entrenamiento

En la Figura 39 se muestra la gráfica con el entrenamiento de los datos de potencia para el mes de febrero. En la configuración del algoritmo se definió un número máximo de 10 épocas para la reducción del error, se llegó a un error óptimo en

menos del número máximo de épocas. En las figuras de la ( 40 a la ( 42 se muestran los errores obtenidos del análisis de las iteraciones y el entrenamiento de los datos.

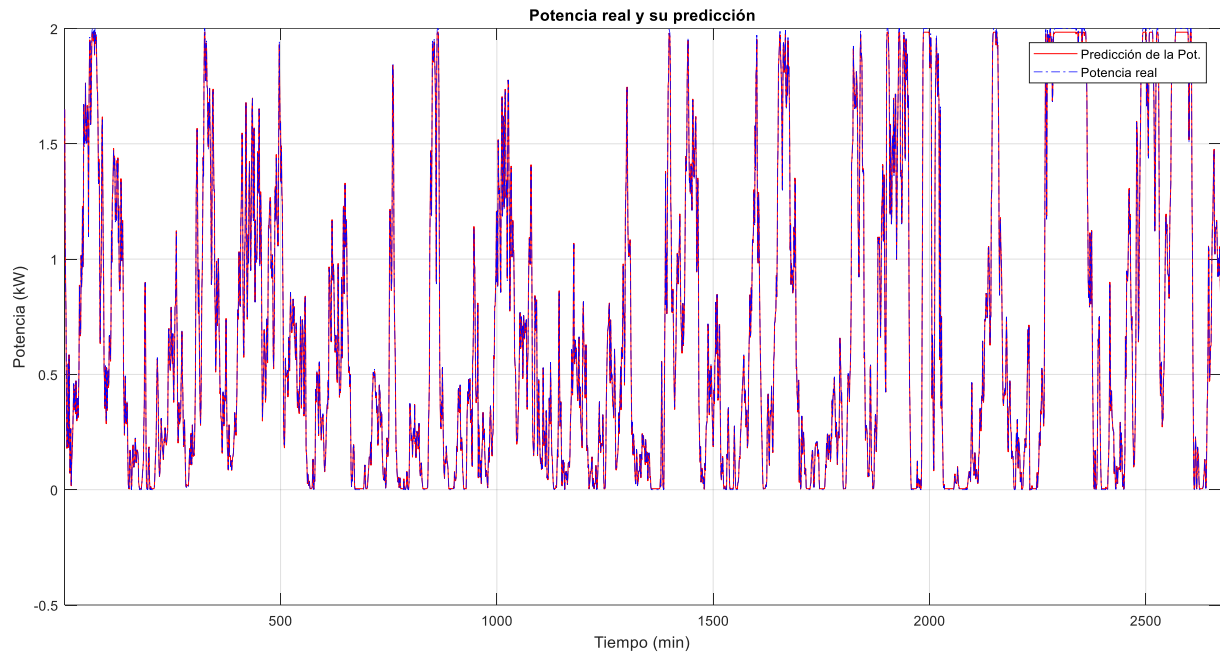


Figura 39. Entrenamiento de red neuronal wavelet con Filtro extendido de Kalman

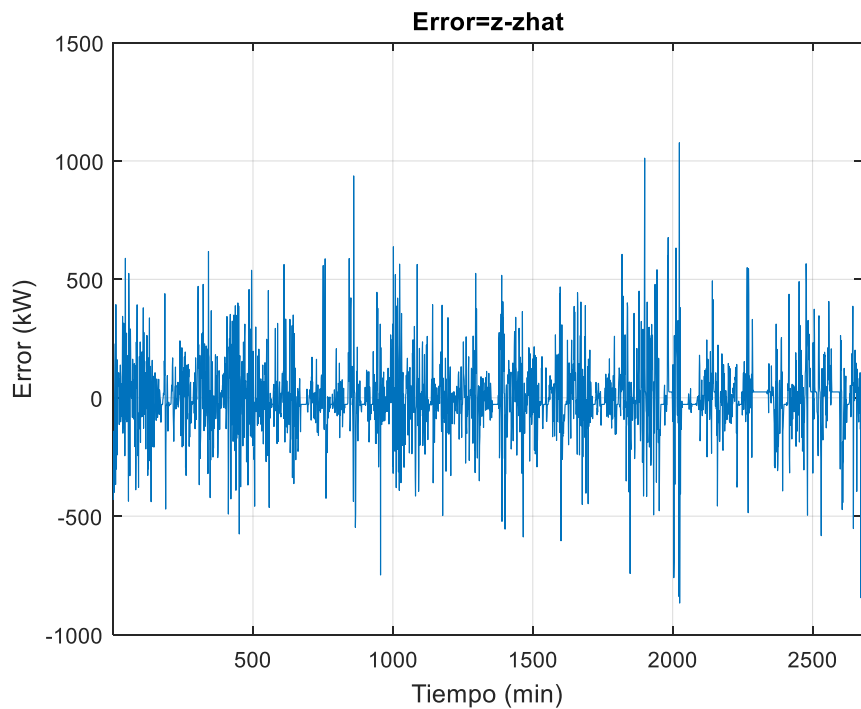


Figura 40. Error de estimación para el entrenamiento NWFEK

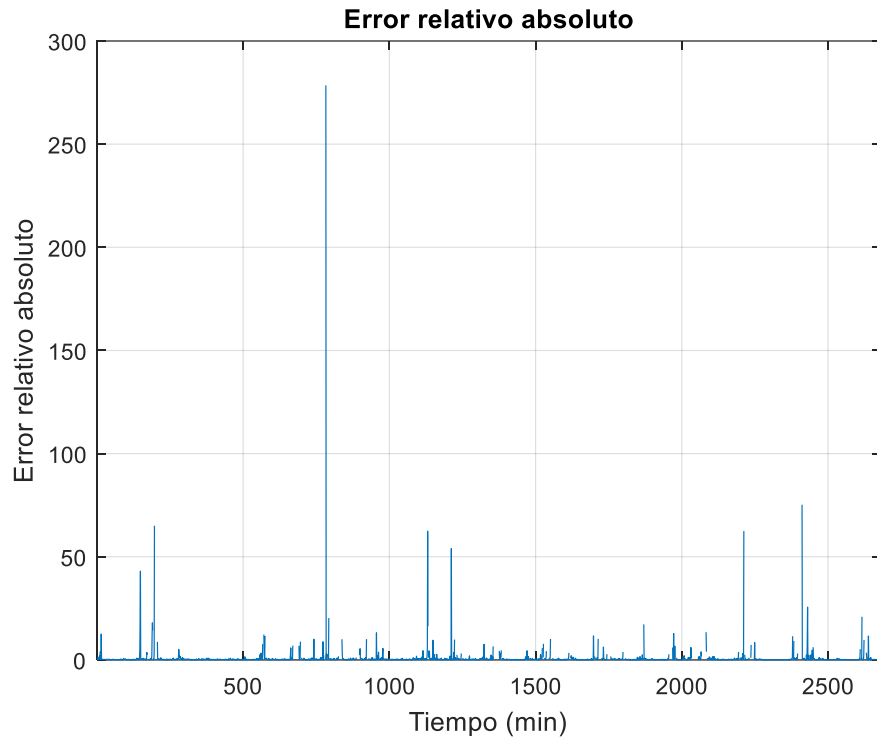


Figura 41. Error relativo absoluto para NWFEK

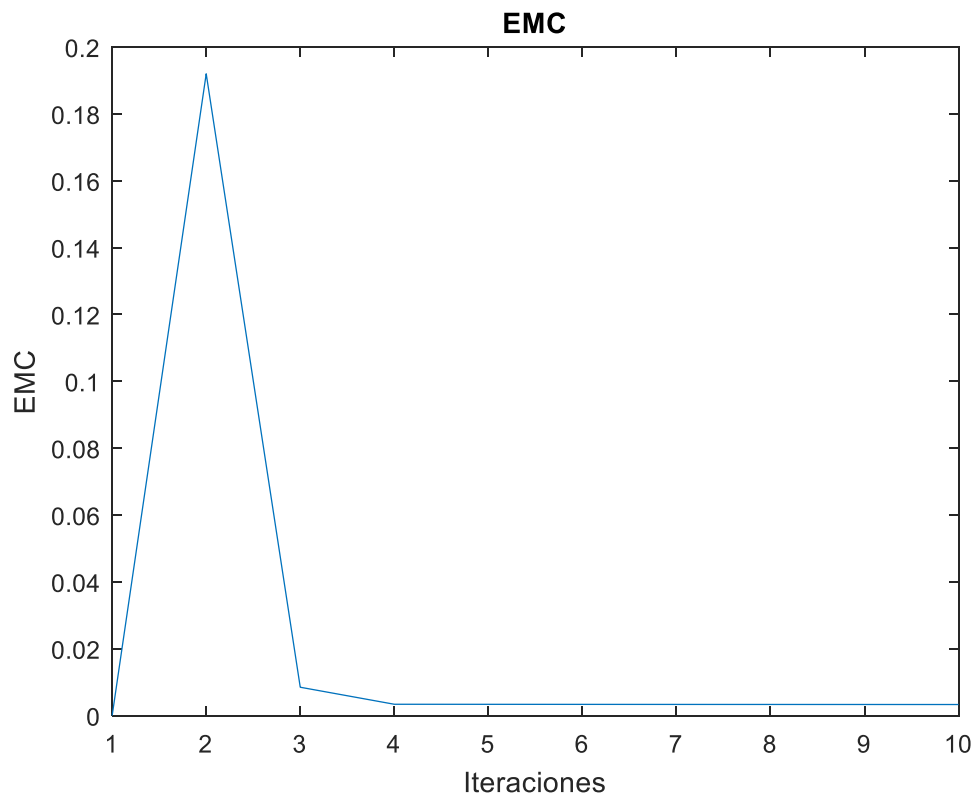


Figura 42 Evolución error medio cuadrático.

En la Tabla VII se muestran los resultados de los entrenamientos de la red wavelet con el filtro extendido de Kalman, representados mediante los errores medidos.

Tabla VII. Resultados en etapa de entrenamiento NWFEK

<b>Parámetro</b>	<b>Training</b>
RMSE	162.1080
MAPE	3.6034e+03
MSE	2.6279e+04
R	0.9683

## Testing

Del resultado del entrenamiento se obtienen los pesos óptimos, calculados mediante el NWFEK, para evaluar la red neuronal y obtener los resultados de la nueva predicción. En la Figura 43 se muestra la comparación de la predicción y los datos reales para el mes de marzo, en la Figura 44 se muestra solo una sección de los datos para apreciar de mejor manera las variaciones entre las líneas de tendencia, las cuales no son parecen ser tan significativas, y podemos comprobarlo observando los errores y el coeficiente de correlación que aparecen en la Tabla VIII

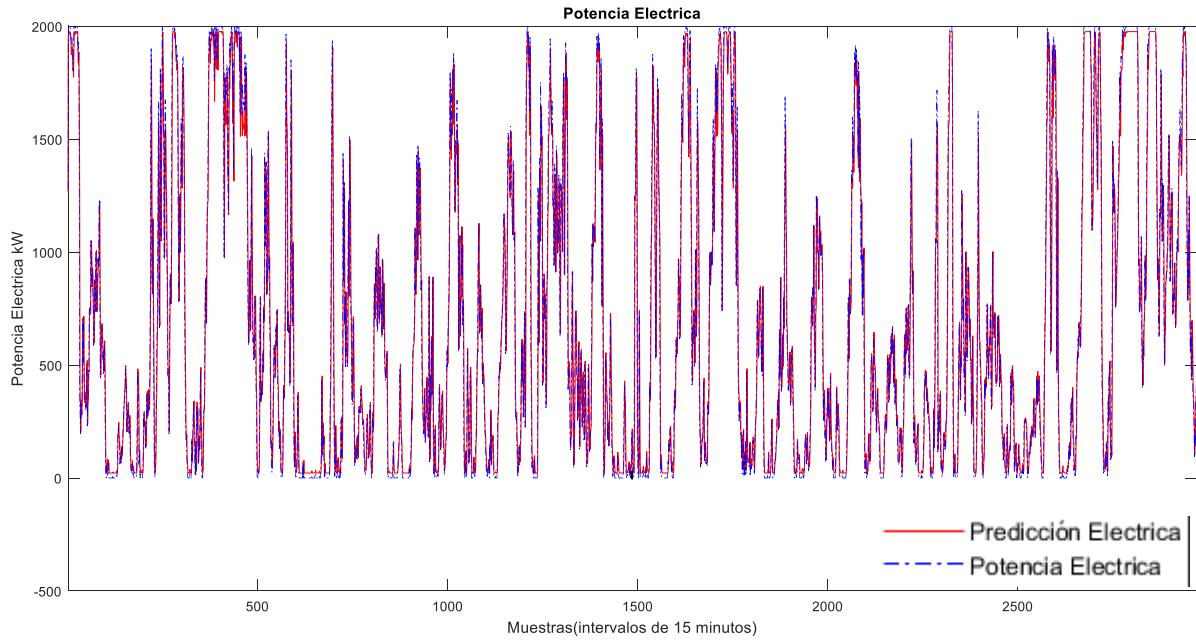


Figura 43. Testing (marzo) con NWFEK

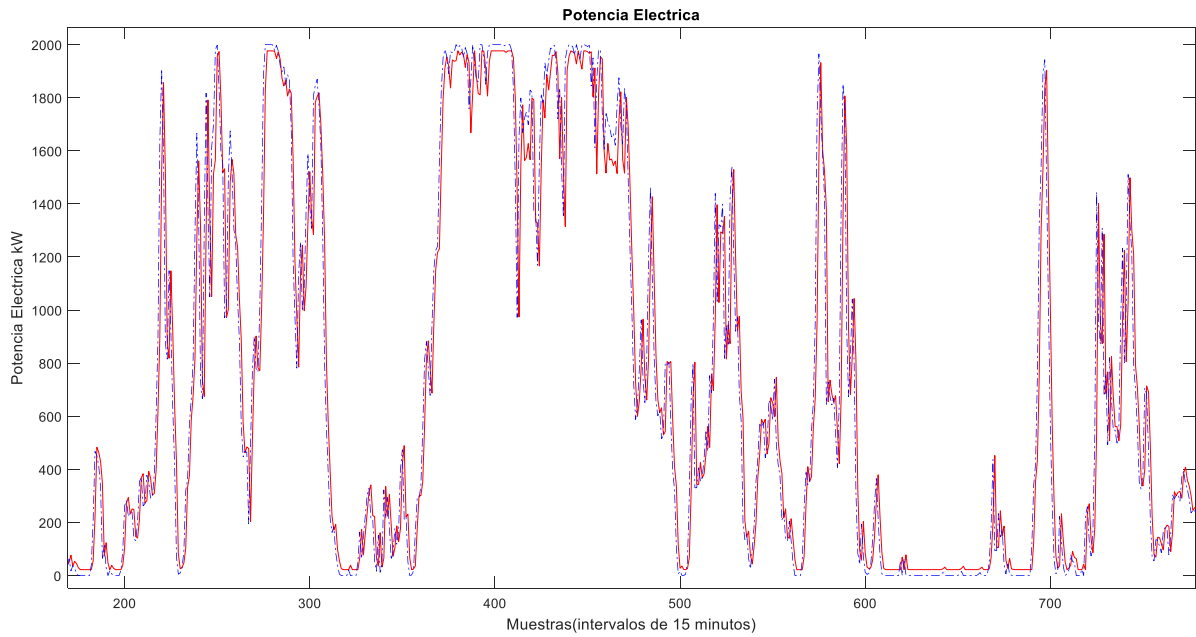


Figura 44. Zoom Testing



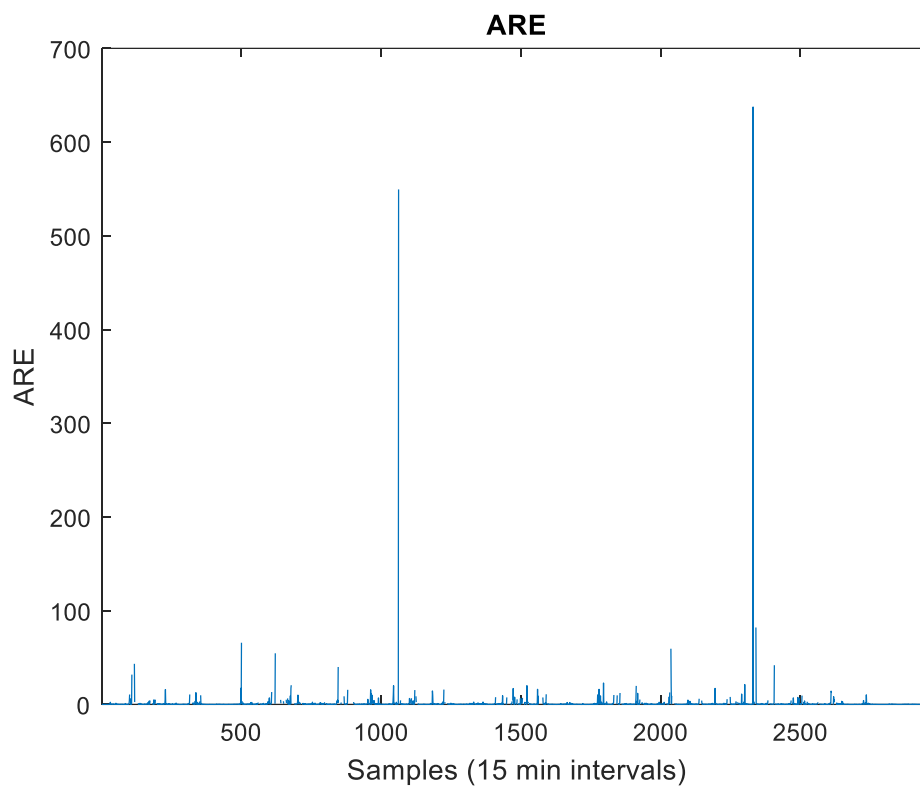


Figura 45 Error relativo absoluto para etapa de testing NWFEK

Tabla VIII. Resultados en etapa de entrenamiento NWFEK

Parámetro	Testing
RMSE	175.3629
MAPE	2.9313e+03
MSE	3.0752e+04
R	0.9645

### 4.3 Métodos Estadísticos ARMA

Los resultados de la predicción con el modelo ARMA se presentan en la Figura 46, y en la Figura 47 se puede observar una parte de los datos donde se puede apreciar mejor la tendencia de los datos. Se observa a simple vista que la predicción tiene dificultad en apegarse a la curva original, sobre todo en los picos, sin embargo, logra llevar de buena manera la tendencia.

En la Tabla IX se puede apreciar los resultados para los errores y la correlación obtenida de la comparación de la predicción y los datos reales.

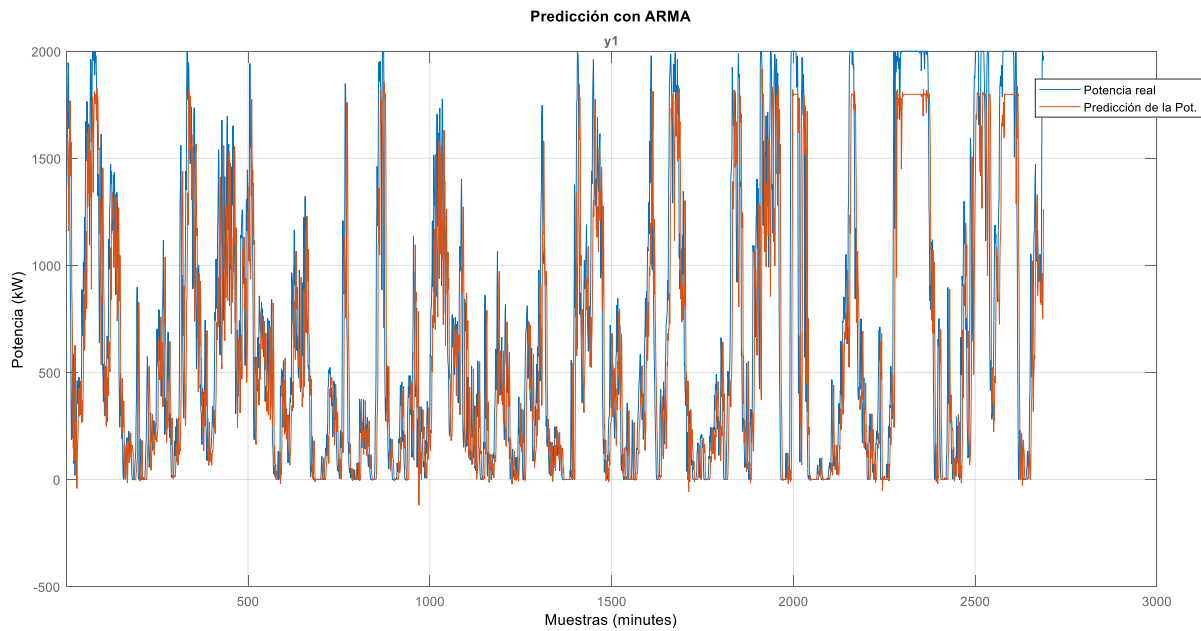


Figura 46. Resultado modelo ARMA

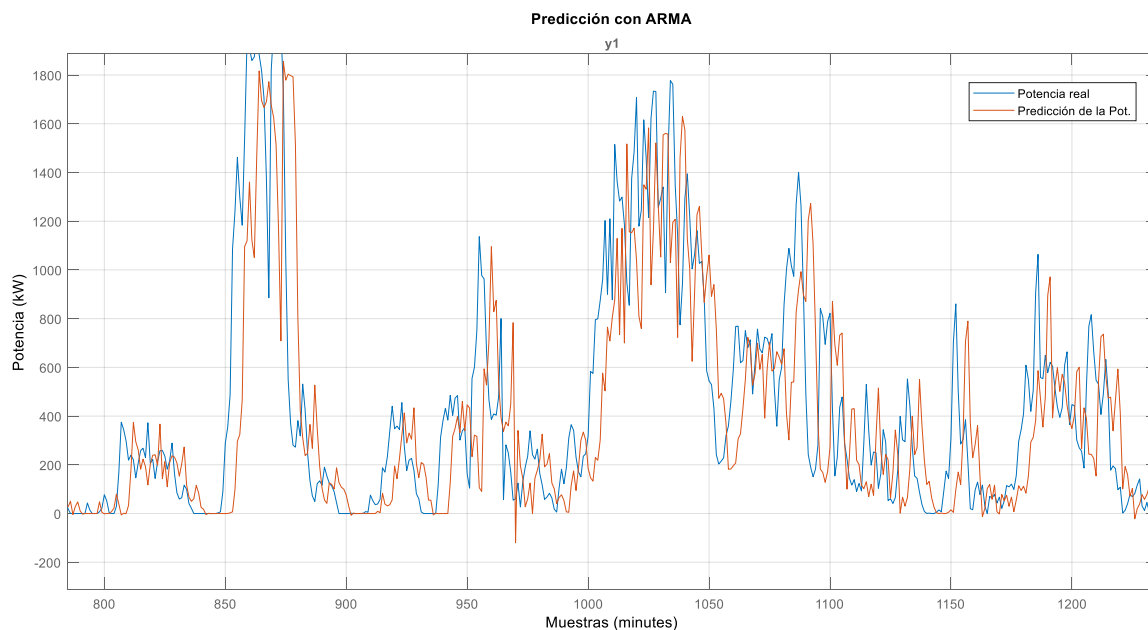


Figura 47. Zoom resultados modelo ARMA

Tabla IX. Resultados método ARMA

Parámetro	ARMA
MAPE	2.552e+04
MSE	3.547e+04
R	0.7552

#### 4.1 Análisis Comparativo de Resultados

En la Tabla X podemos observar la comparación de los resultados de los diferentes algoritmos utilizados en la predicción, los primeros 3 basados en redes neuronales. De estos datos se puede apreciar que la red neuronal NARX podría tener los mejores resultados con respecto a correlación, MSE y tiempo de computo, incluso sobre la

red neuronal con NWF EK, pero la desventaja es que lleva a cabo un proceso aleatorio y se necesita de ejecutarlo varias veces o realizar varias corridas en las cuales se van obteniendo resultados diferentes, a diferencia de los métodos con Filtro de Kalman y ARMA, en los cuales solamente es necesaria una corrida que va minimizando los errores y las correlaciones con una menor incidencia de variación.

Tabla X. Comparación de Resultados

Parámetro	NARX	Perceptrón	NWF EK	ARMA
No. de Neuronas Ocultas	10	10	10	—
No. de Elementos de retardo	8	8	8	—
No. de Capas Ocultas	1	1	1	—
Función de Transferencia capa Oculta	Sigmoide	Tangente	Tangente	—
Función de Transferencia capa de Salida	Lineal	Lineal	Lineal	—
Algoritmo de Entrenamiento	Levenberg-Marquardt	Levenberg-Marquardt	FEK	—
Máximo número de Épocas	8	389	4	—
Periodo de Muestreo	15 min	15 min	15 min	15 min
MSE	2.4159e+04	2.7315e-05	3.0752e+04	3.547e+04
R	0.973	0.999	0.9645	0.7552
Tiempo de computo (segundos)	15.84	31	90.97	39.90

## 5 Conclusiones

En el presente trabajo se diseñó e implementó un algoritmo para predecir la serie de tiempo de la potencia producida mediante un sistema de energía eólica para aerogeneradores a 60 metros de altura en una planta de Baja California Sur, usando tres arquitecturas de redes neuronales con diferentes tipos de entrenamiento y un método estadístico de predicción ARMA, para obtener un error de estimación mínimo en la predicción de la variable de interés.

Este trabajo fue motivado dado que la energía renovable ha tenido un aumento significativo en la estructura energética mundial donde los sistemas eólicos han demostrado un rápido desarrollo y grandes beneficios. Se observan y prevén grandes sistemas de producción eólicos, incluso en la península de Yucatán, así como un gran impacto en los sistemas de energía ya existentes.

Se obtuvieron datos de potencia producida por las turbinas eólicas y mediciones de velocidad de viento cada 5 minutos, las cuales se promediaron para presentarlas cada 15 minutos. Después de la captura y procesamiento de los archivos de datos, se obtuvo la serie de tiempo de la producción de potencia y la serie de tiempo de velocidad de viento. Mediante métodos descriptivos se analizaron tablas y gráficos para los datos de las series de tiempo.

Para aprovechar la información de la serie de tiempo de la mejor forma posible, se encontró una dimensión del sistema mediante la aplicación de teoría de embebimiento a la serie de tiempo de potencia. El resultado del método de Cao para la serie de tiempo de potencia fue de 8, lo que indica cuantos retrasos deben ser empleados como dimensión mínima para aproximar la función objetivo.

Se desarrolló un modelo de predicción de generación de potencia mediante un perceptrón multicapa, con algoritmo de entrenamiento Levenberg-Marquardt, en el cual se observó la evolución de la adaptación de la predicción según el número de épocas transcurridas en el aprendizaje.

Un segundo modelo de predicción, el cual es una red neuronal wavelet entrenada con el filtro extendido de Kalman. Este demostró ser el de mejor resultado en cuanto a la predicción con respecto a la eficiencia en el aprendizaje y reducción de número de épocas de entrenamiento con respecto a los demás métodos, pero también requiere un mayor tiempo y fuerza de cómputo.

Con base a todo lo anterior, los resultados de este trabajo exponen que es posible modelar el comportamiento de generación de potencia mediante redes neuronales artificiales. Ante lo ya mencionado, las redes neuronales son una potente herramienta útil para encontrar relaciones fuertemente no lineales de grandes volúmenes de datos y pueden ser usadas en múltiples áreas de la ciencia.

En trabajos futuros se plantea utilizar este método para la predicción de potencia en campos eólicos a pasos de una hora mediante series de tiempo de rangos de una hora, aplicados a la potencia eléctrica o velocidad de viento.

## Referencias

- Aguilar, M. S., Rodriguez, G. M., & Ríos, M. C. (2006). Caracterización Estadística y Optimización de Redes Neuronales Artificiales para Pronóstico de Series de Tiempo: Pronóstico de un Solo Período. *Computación y Sistemas Vol. 10*, 69-81.
- Alanís, A. (2009). High Order Neural Networks for Wind Speed Time Series Prediction. *Proceedings of the IEEE International Joint Conference on Neural Networks*, 76-80.
- Alexandridis, A. K., & Zaprani, A. D. (2012). Wavelet Neural Networks: A practical guide. *Neural networks*, 1-27.
- Arana, C. B. (2015). Pronóstico de la demanda eléctrica para sistemas de generación interconectados a la red. *Tesis maestría en Ingeniería*. Universidad Autónoma de Yucatán, Mérida, Yucatán.
- Balluff, S., Bendfeld, J., & Krauter, S. (2015). Short term wind and energy prediction for offshore wind farms using neural networks. *IEEE*, 379-382.
- Bingheng Wu, M. S. (2014). Wind power prediction system for wind farm based on auto regressive statistical model and physical model. *Journal of Renewable and Sustainable Energy*, 1-14.
- Box, G., & Jenkins, G. (1976). Time Series Analysis: Forecasting and Control. *Time Series Analysis: Forecasting and Control*.
- Cámara, E. H. (2016). Administración de Flujo de potencia entre sistemas híbridos fotovoltaico-celda de combustible y el almacenamiento por medio de redes neuronales de alto orden. *Tesis posgrado en Energía Renovable*. Centro de Investigación Científica de Yucatán A.C., Mérida, Yucatán, México.
- Camperos, E. N., & García, A. Y. (2006). *Redes Neuronales conceptos fundamentales y aplicación a control automático*. Guadalajara: PEARSON.
- Cárdenas, J. A., Arias, M. A., & Bravo, V. A. (2013). ANÁLISIS Y APLICACIÓN DEL FILTRO DE KALMAN A UNA SEÑAL CON RUIDO ALEATORIO. *Scientia et Technica Vol. 18* , 267-274.
- Chatfield, C. (2005). Time-series Forecasting. *Significance*, 131-133.
- El-Fouly, T. H., El-Saadany, E. F., & Salama, M. M. (2008). One Day Ahead Prediction of Wind Speed and Direction. *IEEE Transaction on Energy Conversion, Vol. 23 No. 1*, 191-201.
- Espadas. (2002). *Reconocimiento de Patrones usando redes neuronales artificiales*. mérida, Yucatán: Monografía de Licenciatura.
- Espadas, G. (2002). *Reconocimiento de patrones usando redes neuronales artificiales*. Meriada, Yucatán: Universidad Autónoma de Yucatán.
- Fausett, L. (1994). *Fundamentals of Neural Networks*. Prentice Hall.
- Feng, Y., & Lin, H. (2013). Simulation of Wind Power System Involving Flywheel Energy Storage Unit Based on Wind Speed Forecasting by RBF Neural Network. *IEEE International Conference on Renewable Energy Research and Applications*, 763-768.

- Gamez, M. E., Sanchez, E. N., & Ricalde, L. J. (2011). Optimal Operation via a Recurrent Neural Network of a Wind-Solar Energy System. *IEEE*, 2222-2228.
- García, G. A., & Salcedo, O. (2010). Predicción de Fallos en Redes IP empleando Redes Neuronales Artificiales. *Polibits*, 67-75.
- Ghareeb, A. T., & Mohammed, O. A. (2013). Wavelet-Adaptive ANN Forecaster for Renewable Energy Sources for Continuous Supply in Microgrid Applications. *Power and Energy Society General Meeting (PES), IEEE*.
- Haykin, S. (1999). *Neural Networks*.
- Haykin, S. (2005). *Neural Networks. A Comprehensive Foundation*. India: Rentice Hall.
- Holtschneider, T. (2013). Optimization of Electricity Pricing Considering Neural Network based Model of Consumers Demand Response. *IEEE Computational Intelligence Applications in Smart Grid (CIASG)*, 154-160.
- Jie, W., Gao, L., Sizhe, C., & Jun, Z. (2009). Control Technologies in Distributed Generation System Based on Renewable Energy. *Asian Power Electronics Journal*. Vol. 3 No. 1, 39-52.
- Jiménes, L. F. (s.f.). Modelos avanzados para la predicción a corto plazo de la producción eléctrica en parques eólicos. *Tesis Doctoral*. Universidad de la Rioja, Logroño.
- Kimura, K., & Kimura, T. (2013). Neural Networks Approach for Wind-Solar Energy System with Complex Networks. *IEEE*, 1-5.
- Medeiros, M. C., & Pedreira, C. E. (2011). "What are the effects of forecasting linear time series with neural networks?". *Logistic and Transportation Review*, Vol. 31, 239-251.
- Moreno, R. A. (2017). *PREDICCIÓN DE GENERACIÓN DE POTENCIA DE ARREGLOS FOTOVOLTAICOS MEDIANTE REDES NEURONALES DE ALTO ORDEN*. MÉRIDA, YUCATÁN: UADY.
- Novakovic, B., Pashaie, R., & Nasiri, A. (2014). Neural Network Based Energy Storage Control for Wind Farms. *IEEE*.
- Shahat, A. E., Haddad, R. J., & Kalaani, Y. (2015). An Artificial Neural Network Model for Wind Energy Estimation. *IEEE*.
- Shing-Yeu, L. (2013). Distributed optimal power flow for smart grid transmission system with renewable energy sources. *Energy* 56, 184-192.
- Veitch, D. (2005). Wavelet Neural Networks and Their Application in the Study of Dynamical Systems. *University of York, York*.
- Wen-hui, Z., Jin, M., & Zheng-zhong, Z. (2014). The Day-Ahead Neural Network Wind Power Prediction Method in Wind Farms. *Mechatronics and Automatic Control Systems, Lecture Notes*, 283-291.
- Yashar, F., Palizdan, N., Huang, Y. F., & Lee, T. S. (2014). Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (WNNs). *Agricultural Water Management*, 26-36.



Zhang, G. P. (2001). A simulation study of artificial neural networks for nonlinear time series forecasting. *Computers & Operations Research*. Vol. 28, 381-396.