

# Pronóstico a corto plazo de velocidad del viento a partir de datos incompletos

Hector Rodriguez Rangel<sup>1</sup>, Noel A. Garcia Carrillo<sup>1</sup>, Juan J. Flores<sup>2</sup>,  
Luis A. Morales Rosales<sup>3</sup>, Giovanni Manjarrez Montelongo<sup>1</sup>

<sup>1</sup> Instituto Tecnológico de Culiacán, División de Estudios de Posgrado e Investigación,  
Culiacan, México

<sup>2</sup> División de Estudios de Postgrado, Facultad de Ingeniería Eléctrica,  
Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

<sup>3</sup> Facultad de Ingeniería Civil,  
CONACYT-Universidad Michoacana de San Nicolás de Hidalgo,  
Morelia, Michoacán, México

hrodriguez@itculiacan.edu.mx, noelandresg@gmail.com, juanf@umich.mx,  
amorales@conacyt.mx, gmanjarrez@itculiacan.edu.mx

**Resumen.** En este artículo se presenta un estudio realizado a la predicción a corto plazo de la velocidad del viento en series de tiempo incompletas. Se ha propuesto realizar este estudio dado el incremento en el interés hacia la transición global de la producción de energías limpias. Siendo el pronóstico de dicha variable de suma importancia para las etapas de planeación, administración y producción. El proceso de pronóstico implica el uso de observaciones pasadas de la variable a pronosticar (velocidad del viento). Para medir la velocidad del viento, las estaciones meteorológicas utilizan dispositivos llamados anemómetros, pero debido a un mal mantenimiento, errores de conexión o desgaste natural, pueden presentar datos falsos o faltantes. En este trabajo, se explora la reconstrucción de la serie de tiempo mediante Redes Neuronales Artificiales, para posteriormente realizar el pronóstico a corto plazo de la serie de tiempo de la velocidad del viento utilizando el método de k Vecinos más Cercanos. Para probar la metodología propuesta, se utilizaron datos recolectados de diferentes locaciones del estado de Michoacán.

**Palabras clave:** pronóstico de la velocidad del viento, reconstrucción de series de tiempo, RNA, kNN.

## Wind Speed Short-term Forecasting from Incomplete Data

**Abstract.** This paper presents a wind speed short-term forecasting from incomplete time series. This study has been proposed given the interest increasing in a global transition of clean energy production. Being

the forecast of interest variable an important process for the planning, administration and production stages. The forecasting process involves the use of historical data of the variable to be forecast (wind speed). Measuring the wind speed, weather stations use devices called anemometers, but due poor maintenance, connection errors or natural wear, they may present false or missing data. In this work, the reconstruction of the time series using Artificial Neural Networks is explored, and then the short-term forecast of the time series of the wind speed using the method of k Nearest Neighbors is presented. To test the proposed methodology, we used data collected from different locations in the state of Michoacan.

**Keywords:** wind speed forecast, time series reconstruction, ANN, kNN.

## 1. Introducción

La energía desempeña un papel de suma importancia en las actividades humanas. Así, la producción de energía se ha convertido en una de las principales cuestiones económicas y medioambientales en todo el mundo [15]. Además, la demanda de energía ha aumentado constantemente con el tiempo, causando daños al medio ambiente al producirla. Según el censo realizado por Observ'ER & Foundation Énergies pour le Monde en 2012, la producción de energía dominante es la producida a través de combustibles fósiles, la cual abarca más de dos tercios de la producción total de energía (68,1 %) [15]. Este tipo de energía, además de ser no renovable, genera daños ambientales irreversibles a nuestro planeta, debido a las grandes cantidades de dióxido de carbono (CO<sub>2</sub>) emitidas a la atmósfera, lo que aumenta los gases de efecto invernadero.

Un mecanismo que busca contrarrestar los daños ambientales y reducir los costos de producción de energía es el uso de energía renovable, también llamada energía limpia. Se encuentra una amplia gama de opciones dentro de las energías renovables como: solar, eólica, biomasa, geotérmica, residuos no renovables e hidroeléctrica, entre otras. De lo anterior, la energía solar y eólica han presentado un crecimiento anual global entre 2002 y 2012 de 50.6 % y 26.1 %, respectivamente [15].

Teniendo en cuenta el gran y continuo crecimiento de estas energías renovables, se han encontrado muchos desafíos (satisfacer la demanda, reducir los costes de producción, mejorar la planificación de las plantas energéticas, etc.). Para llegar a una plena adopción de este tipo de energía es necesario encontrar soluciones eficaces a estos desafíos. Una de las necesidades principales es hacer predicciones de las variables involucradas en la producción de energía. Es decir, tener una estimación de la cantidad de energía que se producirá en los próximos minutos, horas, días o meses [18]. Algunas de las variables involucradas en la producción de energía limpia podrían ser las siguientes: velocidad del viento, temperatura, humedad, presión atmosférica, niveles de presas, causales de rios, etc. En este trabajo, se trata específicamente en el pronóstico de velocidad del viento. La predicción de estas variables puede realizarse a corto, mediano y largo plazo. No hay un marco de tiempo definido para cada uno de los periodos [10].

Sin embargo, regularmente se utilizan previsiones a mediano y largo plazo para la planificación táctica y estratégica, respectivamente. La predicción a corto plazo se utiliza a nivel operativo y su periodicidad puede ser en escala de tiempo de minutos, horas, o días [10,5].

La capacidad de predecir la velocidad del viento es esencial para la integración correcta de la energía eólica en los sistemas eléctricos. Según Barber et al. [2], la importancia de los pronósticos de viento para la industria de energía eólica se deriva de tres hechos: 1) La potencia agregada producida y consumida en todo el sistema eléctrico debe estar casi en equilibrio en todo momento para garantizar la fiabilidad y la seguridad del suministro. 2) La potencia de un parque eólico es muy variable ya que depende fuertemente de la velocidad y dirección del viento. y 3) No existe un mecanismo eficiente y rentable de almacenamiento de energía eléctrica.

Recientemente, una estimación para el año 2030 menciona que un pronóstico perfecto será valorado en 3 *billones* de dólares anuales [16], para el sistema de energía de Estados Unidos. La estimación fue realizada por el Departamento de Energía [2]. Por lo tanto, la previsión de la velocidad del viento sigue desempeñando un papel importante en la tarea de suministro de energía [5].

En la literatura se encontró que han habido diferentes enfoques utilizados para esta tarea [4,13,9,1]. Chang presenta [5] un estudio que categoriza diferentes enfoques que se ocupan del problema de pronóstico que establece la Persistencia, Física, Estadística, Correlaciones Espaciales, Inteligencia Artificial y Métodos Híbridos. El método Persistence (o Naïve) es un método de pronóstico básico, donde  $Y_{t+\Delta} = Y_t$  [20]. Este método particular se utiliza a menudo como un método basal. Los métodos físicos son predicciones numéricas del tiempo desarrolladas por los meteorólogos para una predicción del tiempo a gran escala [12]. Los métodos estadísticos (ME) utilizan datos históricos para encontrar relaciones dentro de la serie de tiempo de velocidad del viento. Dentro de los métodos estadísticos, tenemos el Auto Regresivo (AR), Auto Regresivo de Media Móvil (ARMA), y el Auto Regresivo Integrado de Media Móvil (ARIMA), entre otros [8,7,6]. Los métodos de Inteligencia Artificial (MIA) utilizan Redes Neuronales Artificiales (RNA) [7,3], Máquinas de Soporte Vectorial (MSV) [21], k-Vecinos Más Cercanos (kNN) [19], entre otros como métodos de predicción. Los MIA mejoran en algunos casos los resultados obtenidos utilizando métodos estadísticos. Los métodos híbridos son una combinación de métodos de predicción [7] (por ejemplo, la combinación de métodos estadísticos y de inteligencia artificial).

Los ME o MIA usualmente realizan el pronóstico usando datos históricos. Pero, si el conjunto de datos presenta huecos en la representación de sus observaciones, el ME puede tener problemas para realizar un pronóstico adecuado. En realidad, las series de tiempo de viento presentan datos faltantes debido a problemas con los sensores (por ejemplo, problemas de comunicación, desgaste natural, falta de mantenimiento, etc.). La velocidad del viento es irregular e intermitente [11], lo que sumado al hecho de que las series temporales se encuentran incompletas, la tarea de pronóstico a corto plazo es aún más difícil.

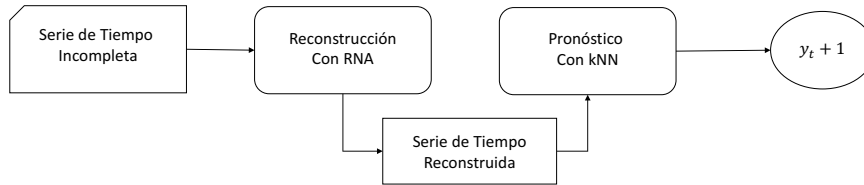
En este artículo, enfrenta del problema de la reconstrucción de la series de

tiempo de la velocidad del viento. Se propone una metodología que se ocupa de este problema, la cual puede trabajar fuera de línea y en tiempo real. La propuesta presentada comienza caracterizando el comportamiento temporal de la serie de tiempo, creando una base de datos de patrones. A continuación, se modela la serie de tiempo utilizando un método híbrido. Este método utiliza una RNA como un enfoque de predicción y Algoritmos Genéticos (AG) para definir la mejor topología de la RNA. Una vez entrenada a la RNA, se procede a reconstruir la serie de tiempo de velocidad del viento. Con estas series de tiempo reconstruidas se procede a realizar pronóstico a corto plazo de la velocidad del viento utilizando el algoritmo de k-Vecinos Más Cercanos (kNN). El resto del trabajo está organizado de la siguiente manera: La Sección 2 describe la metodología propuesta para la realización de la predicción a corto plazo, en donde uno de los pasos previos al pronóstico es el de reconstruir una serie temporal de velocidad del viento, la Sección 3 presenta los experimentos realizados con nuestra propuesta y finalmente las conclusiones de este trabajo se presentan en la Sección 4.

## **2. Predicción de la velocidad del viento**

El problema del pronóstico de series de tiempo de la velocidad del viento no es algo nuevo. Es una tarea complicada ya que hay una amplia gama de patrones de comportamiento diferentes en los datos a modelar. Cada serie de tiempo cumple con distintas características y al modelar series de tiempo de viento se observa distintos comportamientos, por lo que es necesario modelar cada una de ellas. Con anterioridad, distintos métodos de pronóstico se han implementado con el objetivo de realizar pronósticos [5]. De estos métodos sobresalen los que emplean técnicas de inteligencia artificial como las RNA [7,3], dado que en muchos de los casos este método de pronóstico proporciona mejores resultados que los obtenidos mediante los ME.

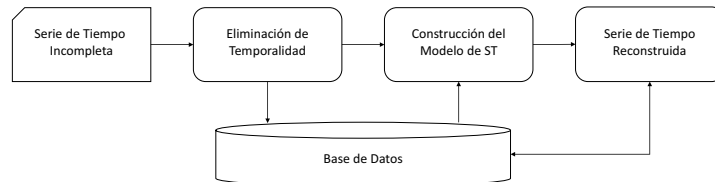
Aunado a que el problema de realizar pronósticos de series de tiempo es naturalmente complicado; si añade el problema de que la serie temporal podría encontrarse incompleta, el problema incrementa su complejidad, aunque, hoy en día hay una serie de algoritmos y técnicas que ofrece una pronóstico eficaz [5,20]. Sin embargo, en la mayor parte de la literatura, el problema de la predicción de la velocidad del viento se aborda suponiendo que la serie temporal está completa (sin ningún hueco). Sin embargo, este no es el caso en realidad. Regularmente, hay problemas con sensores que dan como resultado datos faltantes dentro de la serie de tiempo de velocidad del viento. Para hacer frente a este problema utilizando métodos estadísticos como ARIMA o Holt-Winters se convierte en un trabajo no muy sencillo de realizar, o incluso imposible. Por lo tanto, en este artículo trata el problema de la reconstrucción de las series de tiempo de velocidad del viento como una subtarea previa a la realización del pronóstico a corto plazo y realiza una comparación entre los pronósticos obtenidos con series de tiempo reconstruidas y los obtenidos con las series de tiempo sin reconstruir. El proceso completo de pronóstico se observar en la Figura 1.



**Fig. 1.** Proceso de reconstrucción y pronóstico de la serie de tiempo de la velocidad del viento.

### 2.1. Reconstrucción de series de tiempo

Como ha sido mencionado con anterioridad, nos hemos visto en la necesidad de reconstruir las observaciones faltantes en la serie de tiempo. Para lograr esto es necesario realizar una serie de pasos los cuales se pueden observar en la Figura 2. Una vez completada la reconstrucción de datos se obtiene una nueva serie de tiempo; esta nueva serie de tiempo reconstruida es utilizada para la realización del pronóstico. Un ejemplo de reconstrucción de datos se puede observar en la figura 3.



**Fig. 2.** Diagrama de flujo del proceso de reconstrucción de la serie de tiempo de la velocidad del viento

Para hacerle frente a los huecos dentro de la serie de tiempo, se propone la remoción de la temporalidad. En lugar de tener una ventana deslizante de tamaño  $m$ , el proceso de eliminación de temporalidad crea una base de datos en la que se registran  $m$  características de las series temporales, asociadas con su respectivo valor de pronóstico. Con la base de datos creada, se comienza a construir el modelo de serie de tiempo (CTME). Este proceso utiliza una RNA para modelar la serie de tiempo y Algoritmos Genéticos para definir la arquitectura de la RNA [8,7,6]. Una vez definida la estructura óptima y realizado el proceso de entrenamiento de RNA, se comienza el proceso de reconstrucción

de datos. La Figura 2 muestra el diagrama de flujo del proceso de reconstrucción de la serie temporal de velocidad del viento.

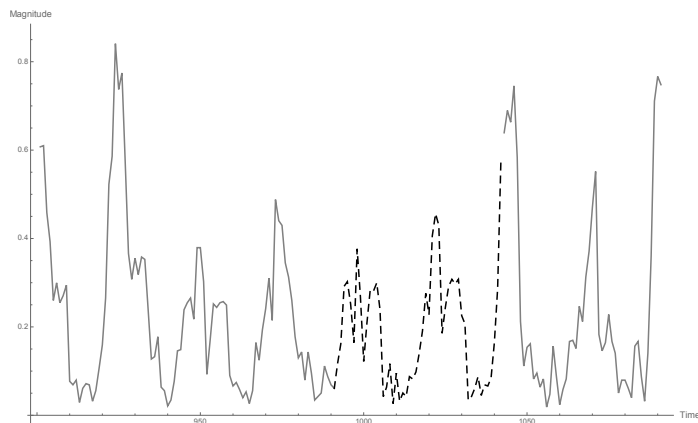


Fig. 3. Ejemplo de reconstrucción de datos en serie de tiempo.

### 2.1.1 Proceso de eliminación de la temporalidad

Para tratar el problema de los huecos producidos por datos faltantes en las series de tiempo, se propone la creación de una base de datos. Regularmente, para el problema de pronóstico, hay una ventana de tamaño  $m$  que se desliza a través de la serie de tiempo. Pero debido a los agujeros dentro de la serie de tiempo, esta ventana ya no puede ser utilizada de la misma manera. Es por ello que se propone crear una base de datos con los patrones temporales dentro de la serie de tiempo de velocidad del viento. La base de datos registra las características de  $m$  (lecturas anteriores de la serie de tiempo) y las asocia con sus respectivos valores de pronóstico. Después de crear la base de datos, se eliminan todos los registros donde se encuentra ausencia de datos. Al convertir la serie de tiempo a una base de datos, se puede dejar de lado la dimensión de tiempo.

La Tabla 1 muestra el resultado de crear la base de datos y extraer las  $w$  características asociadas con una salida esperada. Cada registro dentro de la base de datos representa las últimas  $m$  observaciones en un tiempo definido ( $t$ ), asociado a la salida esperada  $y_{t+1}$  (para propósitos de entrenamiento). De la Tabla 1 se observa que los registros  $m + 2$ ,  $m + 3$ ,  $m + 4$ ,  $m + 5$ ,  $m + 6$ ,  $m + 7$  presentan un hueco ( $\phi$ ) en su interior. Estos registros se eliminan. Con la base de datos definida (sin los huecos), se puede comenzar a buscar una RNA que modela el comportamiento de la serie del tiempo de la velocidad del viento.

### 2.1.2 Construcción del modelo de series de tiempo

**Tabla 1.** Transformación de la serie temporal en un problema de clasificación.

	entradas ( $W_t$ )					$\hat{y}_{t+1} = P(W_t)$
$m$	$y_1,$	$y_2,$	$a_3,$	$\dots$	$y_m$	$y_{m+1}$
$m + 1$	$y_2,$	$y_3,$	$y_4,$	$\dots$	$y_{m+1}$	$y_{m+2}$
$m + 2$	$y_3,$	$y_4,$	$y_5,$	$\dots$	$y_{m+2}$	$\phi$
$m + 3$	$y_4,$	$y_5,$	$y_6,$	$\dots$	$\phi$	$y_{m+4}$
$m + 4$	$y_5,$	$y_6,$	$y_7,$	$\dots$	$y_{m+4}$	$y_{m+5}$
$m + 5$	$y_6,$	$y_7,$	$\phi,$	$\dots$	$y_{m+5}$	$y_{m+6}$
$m + 6$	$y_7,$	$\phi,$	$y_9,$	$\dots$	$y_{m+6}$	$y_{m+7}$
$m + 7$	$\phi,$	$y_9,$	$y_{10},$	$\dots$	$y_{m+7}$	$y_{m+8}$
$m + 8$	$y_9,$	$y_{10},$	$y_{11},$	$\dots$	$y_{m+8}$	$y_{m+9}$
$m + 9$	$y_{10},$	$y_{11},$	$y_{12},$	$\dots$	$y_{m+9},$	$y_{m+10},$
		$\dots$		$\dots$		
$n - 1$	$y_{n-m},$	$y_{n-m+1},$	$y_{n-m+2},$	$\dots$	$y_{n-1}$	$y_n$

En trabajos recientes, se ha demostrado que las RNA presentan grandes cualidades de clasificación [5]. Una vez definida la base de datos, se aborda el problema de regresión que permite crear un modelo capaz de representar el comportamiento de la serie de tiempo de la velocidad del viento.

En este artículo, la arquitectura propuesta de la RNA es un Feedforward Multilayer Perceptron, entrenado por métodos basados en gradiente. La Figura 4 muestra la arquitectura de la RNA propuesta en este trabajo. Una RNA, como un aproximador universal puede aprender cualquier función dada. Un conjunto de  $m$  observaciones pasadas se consideran como los datos de entrada, la capa oculta consta de  $h$  neuronas ocultas, la capa de salida corresponde al valor de pronóstico  $\hat{y}_{t+1}$ ; se utiliza un sigmoide como función de activación. La RNA que modela la serie de tiempo de velocidad del viento  $\hat{y}_{t+1}$  puede definirse como:

$$\hat{y}_{t+i} = f_1\left(\sum_l^{w+1} w_l x_l\right), \tag{1}$$

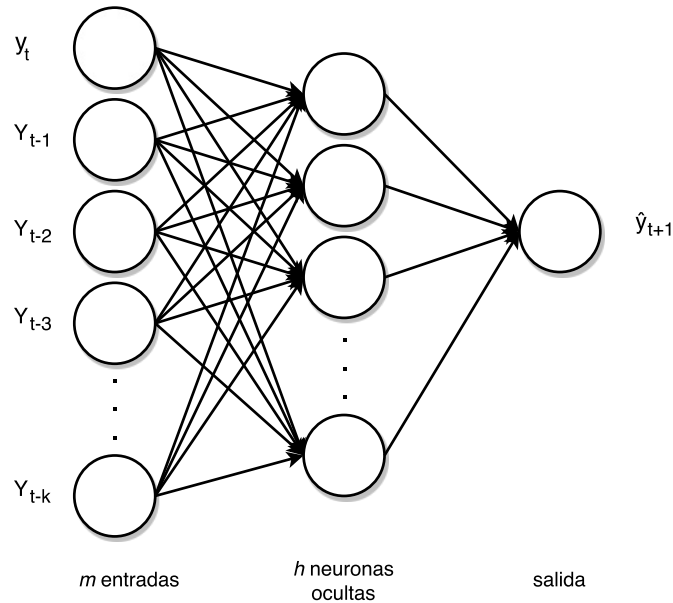
$$x_l = f_2\left(\sum_j^m w_{lj} y_{t-l_j}\right),$$

donde  $f_1$  y  $f_2$  son las funciones de activación, y  $w$  son los coeficientes (también conocidos pesos de sinápticos).

Para modelar el comportamiento de la serie de tiempo de velocidad del viento, se necesita proporcionar un modelo preciso. Para definir la exactitud del modelo óptimo, se utilizo la medida estadística de la media del error al cuadrado (MSE), que se define como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2. \tag{2}$$

Esta medida se minimizará durante el proceso de entrenamiento.



**Fig. 4.** Arquitectura de RNA con  $m + 1$  observaciones pasadas,  $h$  neuronas en su capa oculta y una única salida ( $\hat{y}_{t+1}$ ).

Algoritmos genéticos (AG) son una técnica de optimización inspirada en el principio de la evolución de Darwin. Esto es, imita una versión simplista del proceso de evolución biológica, que consiste en crear una población de individuos, donde cada individuo representa una solución prospectiva del problema que se está resolviendo. AG modifica esta población utilizando operadores genéticos: selección, mutación y recombinación [17].

Determinar la mejor arquitectura de RNA y el algoritmo de entrenamiento es un problema de optimización. Esto es, definen la arquitectura de la RNA determinando el número de entradas y el número de neuronas en la capa oculta. Este es un paso crítico en el proceso de entrenamiento que generalmente se realiza por ensayo y error. En trabajos anteriores, el uso de AG ha sido probado dando excelentes resultados [8]. Cada individuo (el cromosoma) en el AG se define como un vector binario que está codificando el número de entradas y el número de neuronas ocultas de la RNA. El algoritmo empleado de AG para encontrar la mejor topología de RNA se describe paso a paso el Algoritmo 1.

El AG comienza generando una población inicial, a partir de esta población se evalúan todos sus individuos. El proceso continua aplicando selección cruza, mutación a los individuos de la población. Este procedimiento se repite hasta que se alcanza un criterio de convergencia. El pseudocódigo de este proceso lo encontramos en el Algoritmo 1. El algoritmo genético proporciona básicamente como resultado el número de entradas, neuronas en la capa oculta y algoritmo de entrenamiento. Después de definir la topología de la RNA, se comienza un



proceso de refinamiento. El refinamiento es el mismo proceso, simplemente se da la oportunidad de un entrenamiento más extenso con la topología de la red ya definida [7,3].

---

**Algorithm 1** Encontrando la mejor topología para la red usando AG

---

```
1: procedure AG
2:   Generar una población inicial
3:   Computar la función de evaluación de cada individuo
4:   while NOT Terminado do
5:     //Producir nueva generación
6:     for Tamaño Población/2 do
7:       BEGIN //ciclo reproductivo
8:         Seleccionar dos individuos de la anterior generación, para el cruce
9:         (probabilidad de selección proporcional a la función de evaluación del
10:        individuo).
11:        Cruzar con cierta probabilidad los dos individuos obteniendo dos
12:        descendientes.
13:        Mutar los dos descendientes con cierta probabilidad.
14:        Computar la función de evaluación de los dos descendientes mutados.
15:        Insertar los dos descendientes mutados en la nueva generación.
16:        Seleccionar los np mejores individuos.
17:       END //ciclo reproductivo
18:   if La población ha convergido then Terminado := True
19:   return mejor individuo
```

---

### 2.1.3 Reconstrucción de datos

Para el proceso de reconstrucción, se utiliza la RNA que modela el comportamiento de la serie de tiempo de velocidad del viento. Usando este modelo, se puede comenzar con el proceso de reconstrucción. Este proceso comienza con un vector de observaciones ordenadas que se evalúan secuencialmente. Cuando se encuentra un hueco o dato faltante en la serie temporal, comienza el proceso de datos de reconstrucción. Este proceso utiliza las  $m$  últimas observaciones del vector utilizado como entrada. La salida de la RNA ya entrenada en los pasos anteriores se utiliza para llenar el vacío. El proceso continúa hasta que se alcanza el final del vector de observación. La Figura 5 muestra el diagrama de flujo del proceso de reconstrucción.

### 2.1.4 Pronóstico a corto plazo mediante Vecinos más Cercanos (kNN)

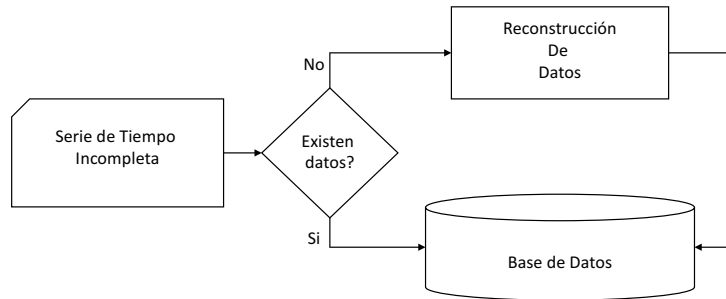


Fig. 5. Proceso de reconstrucción de datos faltantes

Para el pronóstico a corto plazo de la velocidad del viento, se utilizan las series reconstruidas de acuerdo a los pasos definidos previamente. Para el pronóstico se propone utilizar el algoritmo de  $k$  Vecinos más Cercanos (kNN). Este algoritmo es una técnica de clasificación y regresión basada en la similitud de los individuos de una población. La premisa de esta técnica indica que los miembros de una población coexisten rodeados de individuos similares que tienen propiedades similares. Esta idea sencilla es la regla de aprendizaje del los clasificador y regresores kNN. Como se muestra en la Figura 6 donde se busca la similitud de los individuos más cercanos.

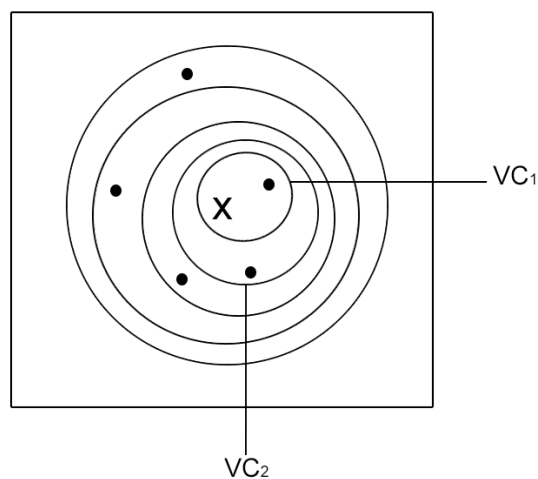


Fig. 6. Ejemplo de clasificación basada en la similitud de los individuos kNN.

Un algoritmo kNN particular, se caracteriza por cuestiones tales como el número de vecinos, el tipo de distancia utilizada, etc. El método usado en este

documento, para realizar el pronóstico de  $y_{t+1}$ , es necesario buscar la similitud del individuo  $x$  en la base de datos definida en la Tabla 1. Donde el individuo  $x$  es definido como las observaciones pasadas de la serie de tiempo en un tiempo  $t$  ( $x = [y_{t-w}, y_{t-w+1}, \dots, y_t]$ ). Después de evaluar las distancias de los individuos de la base de datos, se toman los  $k$  vecinos más cercanos y el pronóstico resultante es el promedio de la columna  $y_{t+1}$  de la Tabla 1. La definición matemática de nuestro modelo de kNN se puede observar a continuación:

$$\hat{y}_{t+1} = \frac{1}{k} \sum_{i=1}^k P(VC(i, y_t)). \quad (3)$$

La distancia euclidiana de dos individuos se define por la Ec. (4):

$$d(r, s) = \sqrt{\sum_{j=1}^m (r_j - s_j)^2}. \quad (4)$$

La ventana que termina en el tiempo  $t$ , denotada por  $w_t$  es la secuencia o vector de las  $m$  mediciones hasta  $t$  como se indica en la Ec. (5):

$$w_t = [y_{t-m+1}, y_t] \in \mathbb{R}^m. \quad (5)$$

La definición recursiva del conjunto de  $k$  vecinos más cercanos es:

$$VC(k, x) = \operatorname{argmin}(d(w, x) | w \in W - V_{k-1}(x)), \quad (6)$$

donde:

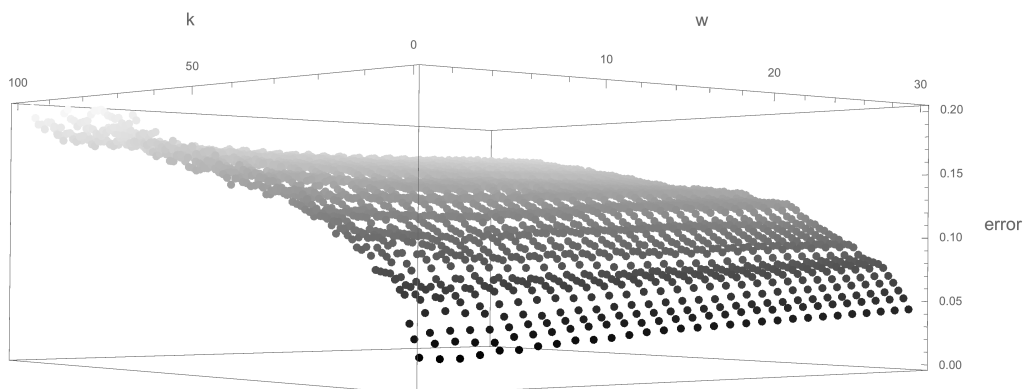
$$V_k(x) = \bigcup_{j=0}^k VC_j(k, x), \quad (7)$$

$$V(0, x) = \phi.$$

Para la determinación de los números adecuados de  $k$  y  $w$  se realizó una exploración en un rango de valores lo suficientemente amplio para establecer las combinaciones que permitan llegar a un valor de error menor. En este caso se realizaron experimentos en donde se iteraron estos valores desde  $k = 1$  hasta  $k = 100$  y desde  $w = 1$  hasta  $w = 100$ . Un ejemplo gráfico de esta exploración se observa en la Figura 7, en donde el eje vertical indica la magnitud de los errores.

### 3. Resultados

Para probar la aptitud de nuestra propuesta, se realizaron experimentos con series de tiempo de la velocidad del viento procedentes de anemómetros ubicados en seis diferentes lugares dentro del estado de Michoacán, México, con registros a intervalos horarios. Dado que los sensores son susceptibles a fallas de funciones y errores de lectura, se tiene el problema de la obtención de series de tiempo con datos faltantes.



**Fig. 7.** Exploración de valores óptimos de  $k$  y  $w$  para el pronóstico mediante kNN.

Se realizaron varios experimentos aplicando el enfoque del método de reconstrucción de la serie de tiempo descrito en la Figura 2. Cada experimento comienza eliminando la temporalidad de la serie temporal creando una base de datos siguiendo el procedimiento descrito en la sección anterior. Una vez obtenida la base de datos, definimos una RNA que modela el comportamiento de la serie de tiempo de velocidad del viento.

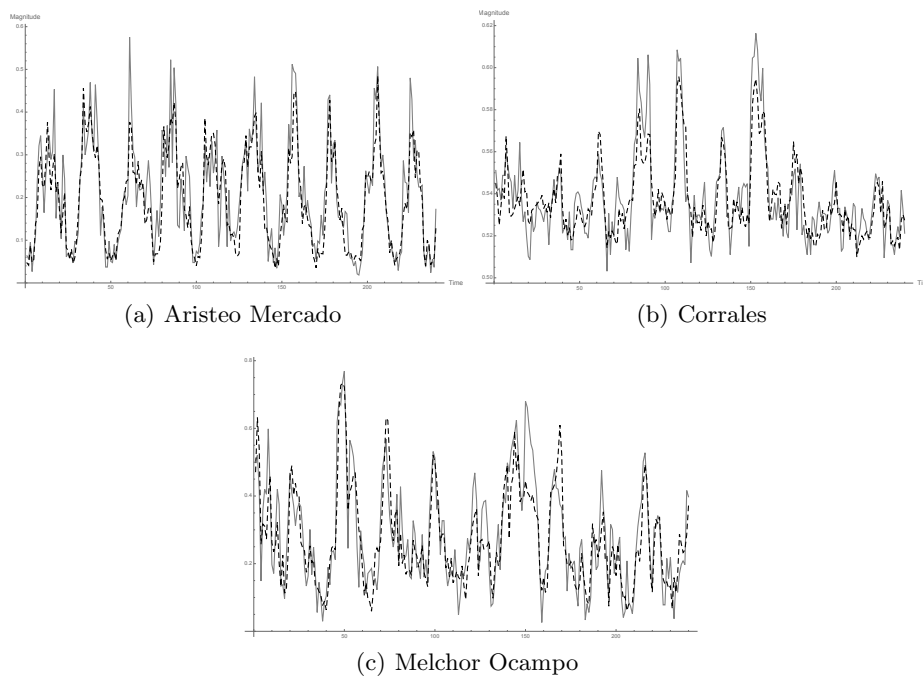
La topología obtenida en esta etapa se utiliza en un proceso de refinamiento, que sólo aumenta el número de épocas, en este caso, se usó hasta 3,000 épocas.

Una vez realizada la reconstrucción y obtenida la nueva serie de tiempo reconstruida; ésta es utilizada como entrada para el proceso de pronóstico mediante el algoritmo de kNN. Realizamos experimentos con seis series de tiempo de distintas estaciones de viento. Paralelamente se realizó el mismo proceso de pronóstico con las series de tiempo sin reconstruir. La aptitud de los pronósticos realizados se muestra en la Tabla 2.

**Tabla 2.** Comparativa de errores definidos por la medida estadística MSE obtenidos en los pronósticos utilizando RNA y kNN en series sin reconstruir y reconstruidas.

	MSE en Pronóstico RNA vs kNN					
	Sin reconstruir			Reconstruida		
	RNA	kNN	Naïve	RNA	KNN	Naïve
1 La Palma	0.0078	<b>0.0028</b>	0.0067	0.0079	0.0034	0.0086
2 La Piedad	0.0076	<b>0.0047</b>	0.0073	0.0077	0.0052	0.0081
3 Markazuza	0.0010	0.0008	0.0011	0.0010	<b>0.0001</b>	0.0005
4 Aristeo Mercado	0.0071	0.0037	0.0084	0.0073	<b>0.0037</b>	0.0087
5 Corrales	0.0022	0.0008	0.0012	0.0037	<b>0.0001</b>	0.0006
6 Melchor Ocampo	0.0144	<b>0.0073</b>	0.0112	0.0137	0.0082	0.0115

Un ejemplo gráfico del proceso de reconstrucción se muestra en la Figura 3. En la Figura 3, la línea continua representa los datos reales, y los punteados representa la reconstrucción. Una vez realizado la reconstrucción se procedió a realizar el pronóstico mediante el uso de kNN. En la Figura 8 muestra un ejemplo de pronóstico de tres series de tiempo correspondientes a Aristeo Mercado, Corrales, y Melchor Ocampo. Para esta Figura 8, la línea continua representa los datos reales y la punteada representa al pronóstico.



**Fig. 8.** Resultado del pronóstico kNN de las estaciones Aristeo Mercado, Corrales, Melchor Ocampo, con serie reconstruida por RNA.

#### 4. Conclusiones

Este artículo ha propuesto un enfoque de reconstrucción de series de tiempo basado en una metodología neural-evolutiva híbrida. Esta metodología permite determinar las arquitecturas óptimas de RNA para modelar series temporales incompletas y posteriormente para poder reconstruirlas. Los experimentos se realizaron considerando múltiples series temporales con datos faltantes. Se generaron modelos y reconstrucción de datos. Se realizó una comparación entre

los pronósticos obtenidos a partir de series de tiempo reconstruidas y series de tiempo con datos faltantes. Se determinó que en algunos casos la reconstrucción puede mejorar la eficiencia del pronóstico pero en otros caso no ocurría lo mismo. Se requiere de realizar un análisis estadístico para determinar la superioridad de un método sobre otro. Los experimentos se realizaron utilizando la plataforma Python utilizando la biblioteca Neurolab [14] para las implementaciones de RNA.

## Referencias

1. Alexiadis, M., Dokopoulos, P., Sahsamanoglou, H., Manousaridis, I.: Short-term forecasting of wind speed and related electrical power. *Solar Energy* 63(1), 61–68 (1998)
2. Barber, C., Bockhorst, J., Roebber, P.: Auto-regressive hmm inference with incomplete data for short-horizon wind forecasting. In: *Advances in Neural Information Processing Systems*. pp. 136–144 (2010)
3. Barbounis, T.G., Theocharis, J.B., Alexiadis, M.C., Dokopoulos, P.S.: Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on Energy Conversion* 21(1), 273–284 (2006)
4. Cassola, F., Burlando, M.: Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. *Applied energy* 99, 154–166 (2012)
5. Chang, W.Y.: A literature review of wind forecasting methods. *Journal of Power and Energy Engineering* 2(04), 161 (2014)
6. Erdem, E., Shi, J.: Arma based approaches for forecasting the tuple of wind speed and direction. *Applied Energy* 88(4), 1405–1414 (2011)
7. Flores, J.J., Graff, M., Rodriguez, H.: Evolutive design of arma and ann models for time series forecasting. *Renewable Energy* 44, 225–230 (2012)
8. Flores, J.J., Loaeza, R., Rodríguez, H., Cadenas, E.: Wind speed forecasting using a hybrid neural-evolutive approach. In: *MICAI 2009: Advances in Artificial Intelligence*, pp. 600–609. Springer (2009)
9. Foley, A.M., Leahy, P.G., Marvuglia, A., McKeogh, E.J.: Current methods and advances in forecasting of wind power generation. *Renewable Energy* 37(1), 1–8 (2012)
10. Giacomoni, M., Kanta, L., Zechman, E.: Complex adaptive systems approach to simulate the sustainability of water resources and urbanization. *Journal of Water Resources Planning and Management* 139(5), 554–564 (2013)
11. Hayashi, M., Kermanshahi, B.: Application of artificial neural network for wind speed prediction and determination of wind power generation output. In: *Proceedings of ICEE*. pp. 12–15 (2001)
12. Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., Yan, Z.: A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews* 13(4), 915–920 (2009)
13. Liu, H., Tian, H., Pan, D., Li, Y.: Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. *Applied Energy* 107, 191–208 (2013)
14. Neurolab: Neurolab a python plugin. Available: <https://pypi.python.org/pypi/neurolab> [Online]
15. Observ, E.: Worldwide electricity production from renewable energy sources. Ninth inventory (2012)

16. Piwko, D., Jordan, G.: The economic value of day-ahead wind forecasts for power grid operations. In: 2010 UWIG Workshop on Wind Forecasting (2010)
17. Rangel, H.R., Puig, V., Farias, R.L., Flores, J.J.: Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks. *Journal of Hydroinformatics* (2016), <http://jh.iwaponline.com/content/early/2016/08/31/hydro.2016.199>
18. Rodriguez Rangel, H., Puig, V., Farias, R.L., Flores, J.J.: Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks. *Journal of Hydroinformatics* (2016)
19. Yesilbudak, M., Sagiroglu, S., Colak, I.: A new approach to very short term wind speed prediction using k-nearest neighbor classification. *Energy Conversion and Management* 69, 77–86 (2013)
20. Zhao, X., Wang, S., Li, T.: Review of evaluation criteria and main methods of wind power forecasting. *Energy Procedia* 12, 761–769 (2011)
21. Zhou, J., Shi, J., Li, G.: Fine tuning support vector machines for short-term wind speed forecasting. *Energy Conversion and Management* 52(4), 1990–1998 (2011)