

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

Filtro de Kalman dual de estadística robusta para la identificación de un generador eólico en presencia de datos espurios

Robust statistic dual Kalman filter for wind generator identification in presence of outliers

Carlos David Zuluaga Ríos*

Eduardo Giraldo Suárez**

Mauricio Alexánder Álvarez López***

Universidad Tecnológica de Pereira

* Universidad Tecnológica de Pereira, Programa de Ingeniería Eléctrica, joven investigador, magíster en Ingeniería Eléctrica. cardazu@utp.edu.co

** Universidad Tecnológica de Pereira, Programa de Ingeniería Eléctrica, profesor asociado, Ph.D. en Ingeniería. egiraldos@utp.edu.co

*** Universidad Tecnológica de Pereira, Programa de Ingeniería Eléctrica, profesor asociado, Ph.D. en Ciencias de la Computación. malvarez@utp.edu.co

Correspondencia: Carlos David Zuluaga Ríos, Universidad Tecnológica de Pereira, vereda La Julita, Edificio de Ingeniería Eléctrica, Oficina E-230, Pereira, Risaralda, Colombia. Tel: 57 (6) 3137 122 ext. 109.

Resumen

Las turbinas eólicas son una de las tecnologías más avanzadas dentro del conjunto de las energías limpias para obtener energía eléctrica. Debido al comportamiento estocástico del viento, se debe realizar el control de estas turbinas para maximizar la potencia de salida. Con el fin de solucionar el problema de control óptimo se debe conocer la dinámica de la máquina, por ello, los esquemas de identificación de sistemas adquieren importancia en este contexto. Este estudio presenta la identificación de un generador eólico que emplea un filtro de Kalman dual de estadística robusta. Aunque el filtro de Kalman dual (DKF) ha sido utilizado en la identificación de sistemas, su rendimiento está asociado a la ausencia de datos espurios. Para realizar la identificación de este sistema se utilizó una tarjeta de adquisición NI USB-6009. Pese a sus buenas características de muestreo, esta tarjeta incorpora datos espurios que afectan la identificación del sistema. El estudio propone solucionar este problema combinando el DKF y el filtro de Kalman de estadística robusta. La metodología propuesta fue implementada en Matlab, mostrando la identificación del generador eólico en tiempo real ante variaciones en las entradas del sistema, con un comportamiento robusto en presencia de datos espurios, comparado con otra metodología estándar de identificación.

Palabras clave: datos espurios, filtro de Kalman, generador eólico, identificación de sistemas.

Abstract

Among the technologies for clean energy generation, wind turbines are one of the most advanced. Due to stochastic behavior of the wind, a proper control action must be exerted over these turbines in order to get a better use of the output power. A proper control action usually implies the knowledge of the system dynamics, and that is why the identification stage is also crucial.

In this paper, the authors propose a robust statistics dual Kalman filter for identification of a wind generator. Although dual Kalman filter (DKF) has been used in system identification, its performance heavily depends on the absence of outliers in the measured data. However, outliers might be easily produced during the data acquisition stage. In this study, the authors show how by combining the DKF and the robust statistic Kalman filter, the problem of outliers in systems identification can be avoided. The method was implemented in Matlab R2009a®. Results for the identification of a wind generator working online are shown and compared to the performance of dual Kalman filter.

Keywords: outliers, Kalman filter, wind generator, system identification.

Fecha de recepción: 5 de agosto de 2013
Fecha de aceptación: 4 de diciembre de 2013

1. INTRODUCCIÓN

En la actualidad, la energía eléctrica se ha convertido en un componente importante en el desarrollo de cualquier país. La electricidad con base en combustibles ha pasado de ser una ayuda energética a resolver problemas sociales y ambientales, por lo que se sugiere la penetración de las energías limpias, como la energía eólica. Para aprovechar la energía por medio del viento se han desarrollado diferentes tipos de turbinas, que incluyen los generadores de inducción y los generadores síncronos. Una de las complicaciones dentro de estas tecnologías consiste en la amplia variabilidad en la velocidad del viento [1], y es por ello que se deben tener estrategias de control que permitan un mayor aprovechamiento de la energía proveniente del viento. Con el fin de implementar un esquema de control óptimo es necesario conocer la dinámica del sistema, por lo cual se hace necesario recurrir a técnicas de identificación de sistemas.

Para realizar la identificación y control de este tipo de máquinas se deben obtener medidas de velocidad angular, voltajes y corrientes en terminales de la red. Sin embargo, dado que estas medidas se adquieren por medio de sistemas sensoriales, es posible que contengan información errónea o datos anómalos, por lo que se tendría una operación posterior defectuosa. Se debe entonces contar con una identificación que sea capaz de mitigar los efectos asociados a los datos anómalos o espurios.

Por otro lado, una técnica eficiente para realizar la estimación de estados en sistemas dinámicos lineales es el filtro de Kalman [2]. Sus raíces se encuentran en modelos de espacio de estados perturbados por ruidos gaussianos, proporcionando soluciones recursivas al problema de estimación óptima de estas variables; uno de sus inconvenientes, sin embargo, está relacionado con su sensibilidad a datos espurios entregados por los sistemas de medida, haciendo que su rendimiento se vea comprometido.

En la literatura se han presentado diferentes versiones del filtro de Kalman con el fin de volverlo robusto a datos espurios, incluyendo: el filtro de Kalman umbralizado, propuesto en [3], el cual utiliza un análisis comparativo para determinar si hay o no datos espurios. No obstante, el método entrega valores estimados incorrectos cuando la salida estimada se aleja mucho del valor observado. Otra versión es el filtro de Kalman con ponderacio-

nes Bayesianas, propuesto en [4], donde se usa un enfoque similar al de mínimos cuadrados ponderados, pero estas ponderaciones son modeladas probabilísticamente; sin embargo este filtro entrega resultados inadecuados cuando se tienen señales sin tendencia. Otra versión es el filtro de Kalman robusto a datos espurios, propuesto en [5], en el que se supone que el ruido de medida es una variable oculta que se modela probabilísticamente.

Además de realizar la estimación de estados, estas metodologías presentan estructuras recursivas que permiten llevar a cabo la estimación robusta de los parámetros de sistemas dinámicos. No obstante, estos métodos asumen parámetros constantes con el tiempo, introduciendo problemas de estimación cuando las señales presentan cambios fuertes de nivel; igualmente, realizan la estimación de parámetros asumiendo no tener una relación entre las entradas controladas y las salidas observadas. Adicionalmente, los métodos anteriores no han sido utilizados para la identificación de sistemas multivariantes.

Este estudio propone la combinación de tres filtros secuenciales, uno de ellos corresponde al filtro de Kalman de estadística robusta (rsKF) propuesto en [6], los otros dos filtros componen el filtro de Kalman dual (DKF) [2]. El DKF son dos filtros de Kalman que operan secuencialmente en la estimación de los estados y las matrices de un modelo de espacio de estado de un sistema dinámico lineal variante en el tiempo. Debido a que los parámetros de las máquinas eléctricas cambian con su operación, el DKF utiliza cada dato individual tan pronto como esté disponible para actualizar tanto la estimación del modelo como los estados del sistema. Esta característica hace que estos algoritmos sean atractivos para aplicaciones *online* [2]. Por otra parte, el rsKF es una variación del filtro de Kalman, que desarrolla una ponderación para cada dato de entrada basada en un enfoque de máxima verosimilitud argumentada en la teoría de estadística robusta [7].

Empleando las técnicas anteriormente mencionadas se obtuvieron resultados experimentales sobre un ejemplo de prueba que se realizó a manera de simulación: los datos sintéticos fueron generados a partir de un sistema conocido con el fin de comprobar la eficiencia del rsKF; por otra parte, se aplicó la combinación propuesta (DKF y rsKF) para la identificación en tiempo real de un generador eólico, en donde se utilizó una tarjeta de adquisición NI-USB6009, fabricada por *National Instruments*, para obtener lecturas de la

señal de velocidad y voltaje en terminales, en donde la identificación de sistemas realizada por la metodología propuesta no se ve afectada por los datos espurios entregados por el sistema de adquisición.

El documento se encuentra organizado de la siguiente manera. La sección 2 muestra el modelado del generador síncrono mediante una representación en modelo de espacio de estados. La sección 3 presenta la teoría básica de los filtros de Kalman utilizados. La sección 4 muestra la metodología empleada para realizar la identificación del generador eólico. La sección 5 expone los resultados y la discusión de la utilización de los filtros mencionados para datos sintéticos y la identificación de un generador eólico. Finalmente, se presentan las conclusiones en la sección 6.

2. GENERADOR SÍNCRONO

Dentro de los elementos más importantes de un sistema de potencia está el generador síncrono, ya que con esta máquina se puede controlar la estabilidad del sistema de potencia, conociendo su modelo. Una de las tareas importantes al realizar el control de estas máquinas es tener una estimación apropiada de los parámetros de dicha máquina. Debido a que sus parámetros cambian en el tiempo, se hace necesario tener una etapa de identificación de sistemas. El generador eólico puede ser representado por un generador síncrono descrito como [8],

$$L_s \frac{di_d}{dt} = -R_s I_d + n_p \omega L_s i_q + u_d, \quad (1)$$

$$L_s \frac{di_q}{dt} = -R_s I_d - n_p \omega L_s i_d - K_m \omega + u_d, \quad (2)$$

$$J \frac{d\omega}{dt} = K_m i_q + \tau_L, \quad (3)$$

$$\frac{d\theta}{dt} = \omega, \quad (4)$$

donde u_d es el voltaje en eje directo del generador, u_q es el voltaje en eje en cuadratura, i_d es la corriente en eje directo, i_q es la corriente en eje en cuadratura, ω es la velocidad angular de la máquina, τ_L es el torque mecánico, θ es la posición angular, y K_m es una constante rotórica que depende de la constante del rotor y la inductancia mutua de la máquina.

3. FILTRO DE KALMAN

El filtro de Kalman es un método que nace de la representación de sistemas dinámicos lineales en modelo de espacio de estados, donde interactúan diferentes variables como: salidas, estados, entradas controladas y las entradas no controladas.

La representación de un sistema dinámico en espacio de estados está dada por,

$$\begin{aligned}\mathbf{x}_k &= \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \boldsymbol{\omega}_k, \\ \mathbf{y}_k &= \mathbf{C}_k \mathbf{x}_k + \mathbf{D}_k \mathbf{u}_k + \mathbf{v}_k,\end{aligned}\tag{5}$$

donde $\mathbf{A}_k \in \mathbb{R}^{n \times n}$ es la matriz de transición de estados y en ella está consignada toda la dinámica del sistema, $\mathbf{B}_k \in \mathbb{R}^{n \times p}$ es la matriz de relación de entradas y estados, $\mathbf{C}_k \in \mathbb{R}^{q \times n}$ es la matriz de salida, $\mathbf{D}_k \in \mathbb{R}^{q \times p}$ es la matriz que relaciona las entradas con las salidas. Se asume que ω_k y v_k son ruidos gaussianos de media cero que no están correlacionados, es decir, $\omega_k \sim N(0, Q)$, $v_k \sim N(0, R)$, donde $Q \in \mathbb{R}^{n \times n}$ y $R \in \mathbb{R}^{q \times q}$. Q y R son matrices de covarianza diagonales para el ruido de estado y observación, respectivamente. De aquí en adelante se denominarán los parámetros del modelo como $\theta = \{\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k, \mathbf{D}_k, Q, R\}$.

Dentro de los sistemas dinámicos hay variables a las cuales se puede tener acceso, como las variables de salida y y la señal de control u ; los ruidos siempre van a existir en los sistemas prácticos reales como variables a las que no se tiene acceso. Otras variables de estos sistemas son los estados x y parámetros θ que no pueden ser observados. Pese a que los estados no se observan, son las variables más importantes del sistema dinámico, debido a que pueden ser utilizadas para representar la evolución dinámica del sistema. Uno de los métodos con mayor éxito para estimar los estados en sistemas dinámicos lineales es el filtro de Kalman. El objetivo de esta técnica, desde un punto de vista probabilístico, consiste en obtener el mejor

estado estimado dado los parámetros θ (asumiendo que $B_k = D_k = 0$) y las observaciones $\{y_k\}_{k=1}^N$ [9]. El mejor estimador x_k para en el instante k se puede obtener maximizando la probabilidad conjunta de los estados x_k y las observaciones y_k en el instante k , dado los parámetros θ y las observaciones hasta el instante $k-1$, $p(x_k, y_k | y_1^{k-1}, \theta)$. Del mismo modo, el mejor estimador para x_k se puede obtener minimizando el logaritmo negativo de la probabilidad conjunta de x_k y y_k , lo cual conduce al siguiente problema de minimización,

$$\hat{x}_k = \underset{x_k}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (p_k^{(i)} - h_k^{(i)} x_k)^2 + \sum_{j=1}^q (s_k^{(j)} - m_k^{(j)} x_k)^2 \right\} \quad (6)$$

donde n es el número de estados; q es el número de salidas; $\hat{x}_k^-$ y P_k^- , son la media y la matriz de covarianza de x_k dadas las observaciones $\{y_k\}_{k=1}^N$; los $p_k^{(i)}$ son los elementos del vector $p_k = (P_k^-)^{-\frac{1}{2}} \hat{x}_k^-$; $s_k^{(j)}$ son elementos del vector $s_k = (R)^{-\frac{1}{2}} \hat{x}_k^-$; $H_k = (P_k^-)^{-\frac{1}{2}}$; $M_k = (R)^{-\frac{1}{2}} C_k$ (e.g. $p_k^{(i)}$ es la i -ésima componente del vector p_k ($n \times 1$) y $h_k^{(i)}$ es la i -ésima fila de la matriz H_k ($n \times n$)).

Resolviendo el problema de minimización se obtiene [9].

$$\hat{x}_k = \hat{x}_k^- + K_k (y_k - C x_k^-) \quad (7)$$

donde,

$$K_k = P_k^- C_k^T (C_k P_k^- C_k^T + R)^{-1} \quad (8)$$

La matriz K_k en la ecuación 8 se conoce como la ganancia de Kalman [2].

El filtro de Kalman es un algoritmo eficiente que utiliza las observaciones ruidosas disponibles para estimar los estados de un sistema dinámico lineal. El filtro cuenta con dos etapas esenciales: una de predicción y una de actualización. En la etapa de predicción se calcula una estimación a priori de los estados ($\hat{x}_k^-$) y de la matriz de covarianza (P_k^-). Luego, con estas dos cantidades se realiza la etapa de actualización, calculando inicialmente la ganancia de Kalman K_k y seguidamente actualizando los estados por medio de la ecuación 7. La matriz de covarianza de x_k también debe actualizarse (para más detalles ver [2]).

Filtro de Kalman dual (DKF)

En el procedimiento de estimación de estados del filtro de Kalman, se asumió que los parámetros A_k , B_k , C_k y D_k y eran conocidos para todos los valores de k . Debido a que los parámetros de las máquinas eléctricas son de difícil acceso y cambian con su operación, es necesario contar con un método que sea capaz de estimar estos parámetros cambiantes. Entre los métodos que han sido empleados para la estimación de los parámetros de máquinas eléctricas se encuentran el método de los mínimos cuadrados [10], el método de los mínimos cuadrados recursivos [11], enjambres de partículas [12], filtro de Kalman [13], redes neuronales artificiales [14], observadores en modos deslizantes [15], entre otros. Algunos de los inconvenientes que presentan los anteriores métodos es que demandan altos costos computacionales o no se pueden usar en aplicaciones en línea. Algunos presentan errores cuadráticos medios considerables en la estimación de parámetros, otros únicamente han sido probados en ambientes controlados, y algunos más no estiman todos los parámetros de la máquina. Por otra parte, el filtro de Kalman dual (DKF) es una herramienta útil y apropiada que, por su enfoque recursivo de estimación simultánea de estados y parámetros de sistemas dinámicos, sobresale cuando se compara contra muchas otras metodologías importantes que hacen parte de la identificación de sistemas. Por consiguiente, a continuación se introduce el DKF para realizar la estimación dual simultánea de los estados del sistema dinámico y las matrices A_k , B_k , C_k y D_k .

Con el fin de modelar la ecuación 5 por medio de una estructura matricial, se definirán los siguientes términos,

$$\begin{aligned} \mathbf{Y}_k &= \begin{bmatrix} \mathbf{x}_{k+1} \\ \mathbf{y}_k \end{bmatrix}, \boldsymbol{\Theta}_k = \begin{bmatrix} \mathbf{A}_k & \mathbf{B}_k \\ \mathbf{C}_k & \mathbf{D}_k \end{bmatrix}, \\ \boldsymbol{\Phi}_k &= \begin{bmatrix} \mathbf{x}_k \\ \mathbf{u}_k \end{bmatrix}, \mathbf{V}_k = \begin{bmatrix} \boldsymbol{\omega}_k \\ \mathbf{v}_k \end{bmatrix}. \end{aligned} \quad (9)$$

Por lo tanto, el sistema mostrado en la ecuación 5 puede ser descrito como

$$\mathbf{Y}_k = \boldsymbol{\Theta}_k \boldsymbol{\Phi}_k + \mathbf{V}_k, \quad (10)$$

Para llevar a cabo la estimación de los parámetros Θ_k se asume que los estados son conocidos. Se plantea para ello un método de estimación dual que trabaja usando el modelo para estimar la señal y luego se emplea la señal para estimar el modelo. Este proceso se puede realizar de forma secuencial tan pronto como esté disponible cada muestra individual, lo cual hace que estos algoritmos sean atractivos para aplicaciones en línea [2]. Por consiguiente, el filtro de Kalman puede ser usado para estimar parámetros y estados en una nueva representación de espacio de estados dada por

$$\Theta_k = \Theta_{k-1} + \mathbf{r}_k, \quad (11)$$

$$\mathbf{Y}_k = \Theta_k \Phi_k + \mathbf{V}_k, \quad (12)$$

donde el modelo descrito por las Ecs. (11) y (12) corresponde a la dinámica de los parámetros y los estados, dentro de un mismo enfoque; $\mathbf{r}_k$ es el ruido presente en la evolución de los parámetros.

De manera similar se puede plantear una función de costo para el modelo de las ecuaciones 11 y 12; y se puede demostrar que la minimización de esta función de costo conduce a un problema de minimización similar al planteado en la ecuación 6, para más detalle ver [2].

Filtro de Kalman de estadística robusta

Con el objetivo de obtener un filtro de Kalman robusto, el filtro de Kalman de estadística robusta (rsKF) introduce una matriz de ponderación W_k para cada observación, tal que la matriz de covarianza de las observaciones y_k sea ponderada mediante W_k . El cálculo de esta matriz de ponderación se basa en la teoría de estimación por máxima verosimilitud para estadística robusta [16]. Un estimador robusto para x_k se obtiene minimizando [6],

$$\hat{\mathbf{x}}_k = \underset{\mathbf{x}_k}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left(p_k^{(i)} - \mathbf{h}_k^{(i)} \mathbf{x}_k \right)^2 + \sum_{j=1}^q \rho \left(s_k^{(j)} - \mathbf{m}_k^{(j)} \mathbf{x}_k \right)^2 \right\} \quad (13)$$

donde los términos $p_k^{(i)}$, $\mathbf{h}_k^{(i)}$, $s_k^{(j)}$ y $\mathbf{m}_k^{(j)}$ fueron definidos en la ecuación 6; ρ es una función de pérdida dada como [16],

$$\rho(y) = \begin{cases} \frac{1}{2}y^2, & |y| \leq c \\ c|y|^2 - \frac{1}{2}c^2, & |y| > c. \end{cases} \quad (14)$$

La función de pérdida en la ecuación 6, es igual a la función identidad. La constante c es seleccionada según el grado de penalización en la función de pérdida ρ .

Se puede demostrar [6] que la ganancia de Kalman presentada en la ecuación 8 puede ser calculada como

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}_k^T \left(\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R} \frac{1}{2} \mathbf{W}_k \mathbf{R} \right)^{-1} \quad (15)$$

donde $\mathbf{W}_k = \text{diag} \{w_k^{(1)}, \dots, w_k^{(q)}\}$. Las ponderaciones $\{w_k^{(j)}\}_{j=1}^q$ en el instante k se calculan como

$$w_k^{(j)} = \frac{\psi \left(s_k^{(j)} - \mathbf{m}_k^{(j)} \hat{\mathbf{x}}_k^- \right)}{s_k^{(j)} - \mathbf{m}_k^{(j)} \hat{\mathbf{x}}_k^-}, \quad (16)$$

con $\psi(\cdot)$ igual a la función de pérdida de Huber [16],

$$\psi(y) = \begin{cases} y, & |y| \leq c \\ c \operatorname{sgn}(y), & |y| > c, \end{cases} \quad (17)$$

donde $\operatorname{sgn}(\cdot)$ es la función signo. Para los experimentos realizados en este estudio el valor de c fue seleccionado como $c = 1,645$ para una contaminación del 5% de los datos [6].

La aplicación del rsKF se hace de manera similar que el filtro de Kalman estándar (sKF), lo único en que se diferencia es en el cálculo de la ganancia de Kalman.

4. METODOLOGÍA

El desarrollo de este estudio contó con la utilización de dos tipos de datos: sintéticos y datos a partir de un generador eólico real implementado por medio de un generador síncrono. A continuación se describirán de forma

detallada los datos, cómo se realizó la identificación mediante el enfoque propuesto y, finalmente, se explican las métricas de validación para observar la influencia de los datos espurios en las observaciones.

Base de datos sintéticos a partir de un modelo conocido

A partir de un modelo conocido (función coseno) se crearon 1.000 observaciones. A la señal de salida ($q = 1$) se le agregó ruido blanco gaussiano con varianza $\sigma^2 = 0.01$. Los datos espurios se introdujeron en una cantidad igual al 5% de todos los datos. Los datos espurios se ubicaron aleatoriamente como parte de las primeras 500 observaciones. Para el sistema conocido se asumió un único estado ($n = 1$). Además, se asumió $A = C = 1$ y $Q = R = 10^{-4}$ [4]. Igualmente se asumió que las matrices $B = D = 0$.

Base de datos de un generador eólico

En la figura 1 se muestra el montaje que se utilizó para llevar a cabo la identificación del generador eólico. Se empleó un variador de velocidad (VV) (ver apéndice C) que alimenta un motor de inducción para simular el elemento encargado de impulsar el generador eólico a una velocidad específica. En este contexto, el motor de inducción es llamado primo motor. Igualmente, se utilizó un módulo de estado sólido (SSM) (ver apéndice D) para alimentar el campo del generador. Se asumió que el grupo completo es de orden ocho, donde las entradas corresponden al voltaje en el VV y al voltaje en el SSM, donde se hicieron variaciones en sus valores para tener condiciones de velocidad y voltaje variable. Las salidas corresponden a la velocidad del grupo y al voltaje en terminales del generador. Para adquirir los datos a partir del generador eólico se utilizó una tarjeta NI-USB6009 de *National Instruments* (ver apéndice E) con un tiempo de muestreo $t_s = 0,1s$. A cada salida se le agregó un ruido blanco gaussiano, con varianza de $\sigma_{\omega}^2 = 0,1$ para la velocidad; y para la tensión en terminales una varianza de $\sigma_v^2 = 0,01$.

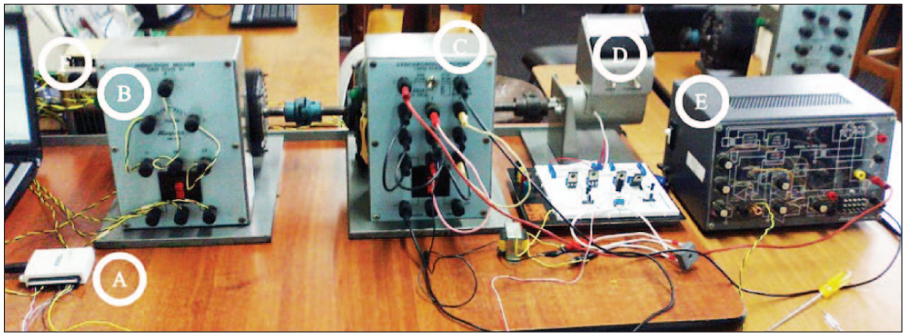


Figura 1. Montaje experimental del generador eólico. A es la tarjeta de adquisición; B es el motor de inducción; C es el generador síncrono; D es el tacogenerador; E es el módulo de estado sólido; y F es el variador de velocidad.

Procedimiento general para la identificación robusta del generador eólico

La sección 2 expuso de forma básica la estructura matemática de las técnicas que se emplearon en este estudio. En esta subsección se hablará de cómo se combinan el DKF y el rsKF. El modelo que combina el DKF y el rsKF se denominará DKF-rsKF. Para explicar cómo se emplearon estos dos métodos en la identificación de sistemas se presenta un diagrama de bloques en la figura 2.

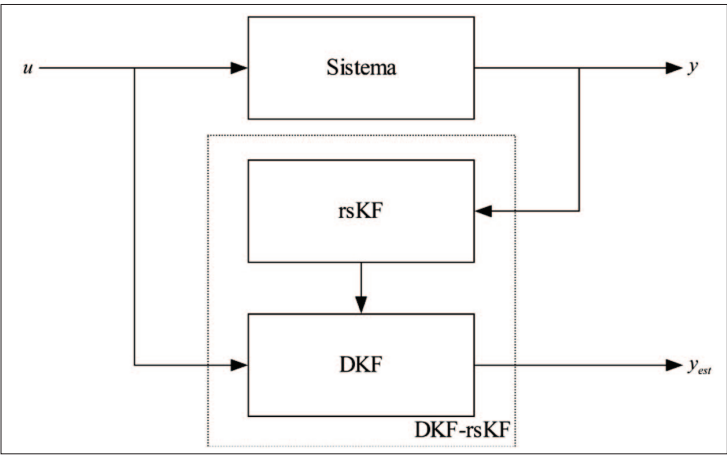


Figura 2. Diagrama de bloques para la representación de la combinación del DKF y el rsKF para la identificación robusta de sistemas multivariables, en particular un generador eólico. La línea punteada corresponde a la combinación secuencial del DKF-rsKF.

Como se ha mencionado, las dos variables que se pueden medir en un sistema dinámico son la señal de entrada o de control u y la señal de salida y . La señal de salida se pasa por el rsKF para obtener una señal que no presente alteraciones frente a datos espurios; esta señal filtrada se convierte en las observaciones disponibles que, junto con la señal de control del sistema, se usa como entrada al DKF que estima las matrices $A_{k'}$, $B_{k'}$, $C_{k'}$ y $D_{k'}$ y los estados $\hat{x}_k$. Luego de obtener los parámetros y estados estimados se calcula la señal de salida estimada.

Validación

Con la finalidad de ilustrar la influencia de los datos espurios presentes sobre la predicción realizada por los métodos (sKF y rsKF) se calcularon dos estadísticas muestrales: la skewness muestral (γ) y la kurtosis muestral (κ), definidas como [17]

$$\kappa = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2} - 3, \quad (18)$$

$$\gamma = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^{3/2}}, \quad (19)$$

donde $\bar{x}$ es la media muestral. El valor de -3 que aparece en la ecuación 18 hace que su valor sea idénticamente cero para el caso en el que los datos sigan una función de distribución gaussiana. Otra métrica usada para observar el rendimiento de los métodos es el error residual E_{res} entre la predicción de la salida, y_{est} y su valor real sin datos espurios y_{clean} (datos limpios) por medio de $E_{res} = \frac{1}{N} \|y_{est} - y_{clean}\|^2$. La kurtosis, la skewness y el error residual solo se calcularon para los datos sintéticos.

La influencia de los datos espurios en las observaciones puede ser demostrada a partir de la skewness y kurtosis, puesto que ambas cantidades son sensibles a datos anormales. Ambos momentos estadísticos han constituido la base para proponer métodos que detectan datos anómalos [17]. Lo que

se pretende es comparar el rendimiento del sKF y del rsKF usando estas métricas.

5. RESULTADOS Y DISCUSIÓN

Primero se presenta un experimento con datos sintéticos, en el que se muestra el desempeño de un filtro de Kalman de estadística robusta comparado con un filtro de Kalman estándar.

En un segundo experimento se empleó el sistema DKF-rsKF, que se mostró en la figura 2, con el propósito de realizar la identificación de un generador eólico.

Datos sintéticos a partir de un sistema conocido

En la figura 3 se muestran los datos sintéticos, a partir de una función coseno ruidosa, que se emplearon en el experimento.

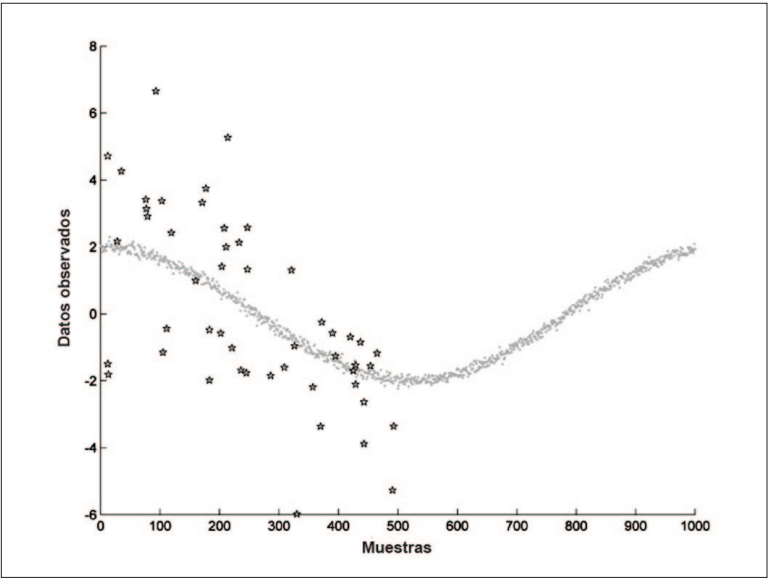
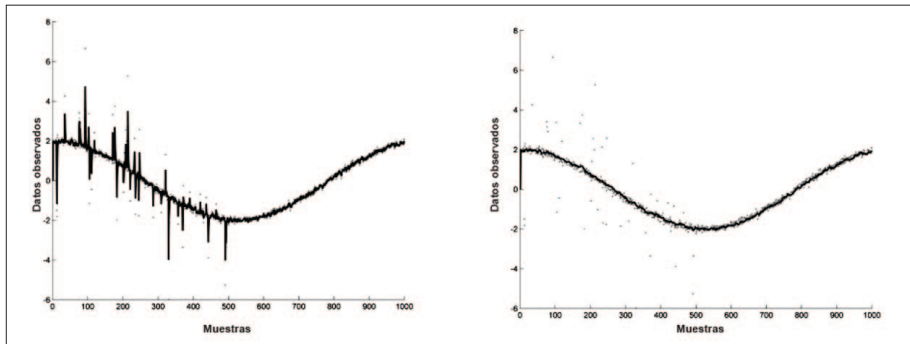


Figura 3. Datos observados y espurios para una función coseno ruidosa. Las estrellas son los datos espurios agregados a la señal coseno, y los puntos son las observaciones de la función coseno.

La figura 4 presenta la respuesta del sKF y rsKF ante el problema de tratamiento de datos espurios en las observaciones.



(a) sKF

(b) rsKF

Figura 4. Predicción de la salida por el sKF (figura de la izquierda) y rsKF (figura de la derecha) para una función coseno ruidosa. Los puntos corresponden a la señal observada y la línea sólida negra es la señal entregada por cada método.

La figura 4 muestra cómo el sKF es sensible a estos datos. Por otra parte, el rsKF pondera fuertemente la varianza de las observaciones cuando se tiene un error de estimación grande y, de esta manera, las predicciones del rsKF no se ven perturbadas por la presencia de datos espurios. También en la figura 4 se puede observar que en el caso de no tener datos espurios, es decir en las últimas 500 muestras, ambos filtros entregan estimaciones adecuadas sin importar el comportamiento de las observaciones. El rsKF obtuvo un error residual de $0,0087 \pm 0,0023$, mientras el sKF un error de $0,1728 \pm 0,1945$.

Como se mencionó anteriormente, la kurtosis y la skewness son susceptibles a datos espurios. Por ejemplo, un valor alto de kurtosis o un valor de skewness diferente de cero implican la existencia de datos producidos con distribuciones de colas pesadas en las observaciones. Usualmente, estas distribuciones de colas pesadas son las que mejor describen la incertidumbre en fuentes generadoras de datos espurios [18]. Para este experimento se obtuvo una kurtosis para los datos limpios de $-0,7765 \pm 0,5423$, para el rsKF de $-0,8586 \pm 0,5318$ y finalmente para el sKF de $5,5647 \pm 12,0013$. Se observa entonces que el sKF se ve afectado por los datos espurios presentes

en las observaciones, mientras el rsKF ajusta sus predicciones a los valores similares de la señal sin datos anómalos.

Similarmente, con respecto a la skewness, para los datos limpios se tiene un valor de $-0,007 \pm 0,46$; para el rsKF de $-0,0048 \pm 0,5722$, y para el sKF se obtiene una skewness de $0,0199 \pm 1,5275$.

Resultados en tiempo real para un generador eólico

Los resultados anteriores se basaron en un modelo con datos sintéticos, obteniendo propiedades relevantes para el caso de tratamiento de datos espurios. Posteriormente, con el propósito de observar cómo es el comportamiento del algoritmo de identificación propuesto en un ambiente real, se realizó la identificación de un generador eólico. Su descripción se hizo en la sección 4.

Como primera prueba se analizó la respuesta del DKF ante cambios aleatorios de las señales de entrada. La figura 5 muestra el comportamiento del método de identificación DKF para el generador eólico, en donde se observa que el algoritmo se ve afectado por el ruido presente en la señal y los datos espurios entregados por el módulo de adquisición, como se nota en las muestras 410 y 756 para la velocidad (figura 5 superior) y para las muestras 437, 465 y 862 del voltaje en terminales de la máquina síncrona (figura 5 inferior). En las primeras muestras de esta figura se observa que las señales tienen pequeñas diferencias; a medida que se presentan más muestras, la señal estimada por el DKF se hace muy similar a las observaciones entregadas por la tarjeta de adquisición; y al final se presenta un traslape entre la señal medida y la estimada por el método. Para mostrar lo anterior se realizó un acercamiento de la figura 5 entre las muestras 750 y 760 para la velocidad del generador y entre las muestras 850 y 880 para el voltaje en terminales, con el fin de observar dicha sobreposición (ver figura 6). De la figura 6 es claro que el DKF sigue los datos espurios presentes en las observaciones, lo cual es una propiedad negativa del proceso de identificación.

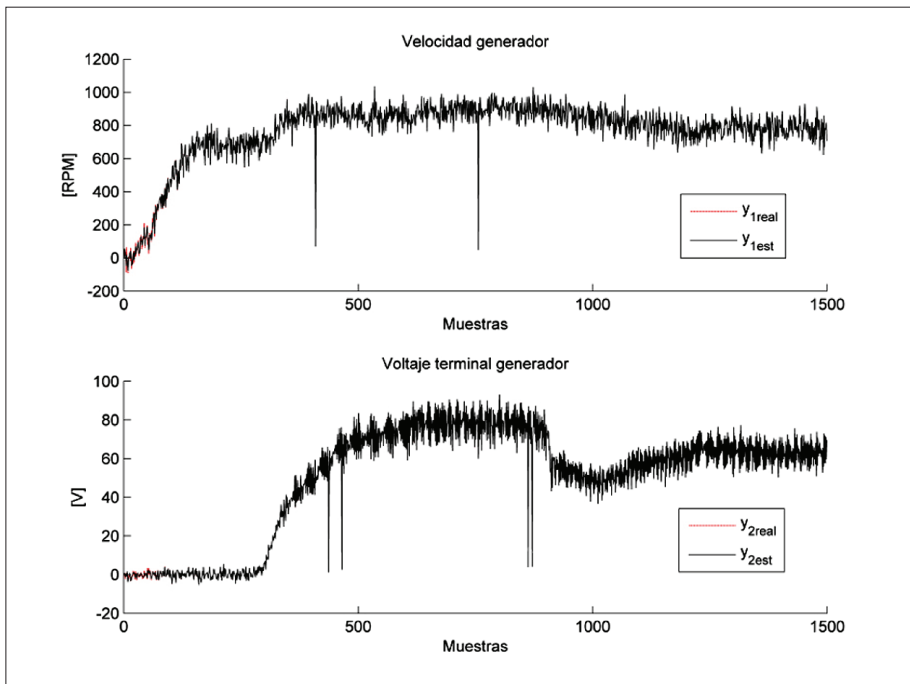
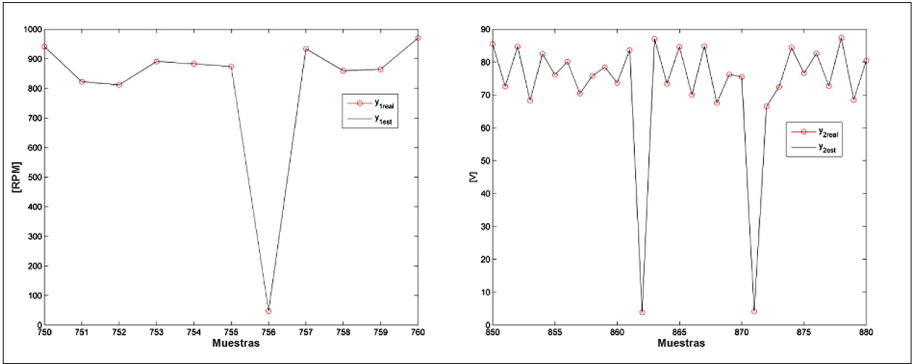


Figura 5. Comportamiento del generador eólico. La línea punteada corresponde a la señal entregada por el sistema de adquisición y la línea sólida pertenece a la señal entregada por el DKF. La figura superior es la respuesta del DKF para la velocidad del generador; la figura inferior muestra la identificación del voltaje en terminales de la máquina.



(a) Velocidad del generador (b) Tensión en terminales del generador

Figura 6. Respuesta de la velocidad (figura de la izquierda) y voltaje del generador (figura de la derecha). La línea roja discontinua con círculos corresponde a la señal real medida y la línea negra sólida corresponde a la señal estimada por el método. Se observa que las dos señales se traslapan. Acercamiento de la figura 5 entre las muestras 750 y 760 para la velocidad del generador y entre las muestras 850 y 880 para el voltaje en terminales.

Después de haber mostrado el efecto que tienen los datos espurios en las observaciones usando el DKF para la identificación, se procedió a estudiar la identificación del generador empleando el DKF-rsKF: la señal adquirida por el sistema sensorial es analizada inicialmente por el rsKF para obtener señales no perturbadas por los datos espurios, y a continuación se aplica el DKF para obtener los estados, salidas y matrices estimadas, como se muestra en la figura 2. Igualmente que en el caso anterior, se efectuaron cambios aleatorios en la señal de control y se le agregó ruido blanco gaussiano a cada salida. En la figura 7 se muestra el comportamiento del voltaje en terminales (figura 7 inferior) y la velocidad de la máquina síncrona (figura 7 superior). En estas gráficas se puede observar que los datos espurios y el ruido presente en la señal no afectan la estimación de las señales de salida. Además, se ve que para las muestras 300, 694 y 989 en la velocidad del generador, el algoritmo atenúa el efecto de los datos espurios y entrega una estimación apropiada del valor de la señal en esas muestras. De igual forma, para el voltaje en terminales del generador, en las muestras 612, 1.055 y 1.100, arrojadas por el equipo de adquisición, el algoritmo de identificación multivariable estima de manera adecuada estos valores, brindando robustez en la estimación de la señal de salida del sistema dinámico. Para

observar con claridad la identificación hecha con el DKF-rsKF se realizó un acercamiento de la figura 7 entre las muestras 280 y 370 para la velocidad del generador y entre las muestras 570 y 660 para el voltaje en terminales, como se muestra en la figura 8.

Para obtener una idea de cuánto se tarda el método propuesto en realizar el procedimiento de identificación, se analizó el tiempo de cálculo promedio para cada iteración; para ello se obtuvo el valor promedio de 1.500 iteraciones y su respectiva desviación estándar. El DKF-rsKF presentó un tiempo de cálculo de $0,0120 \pm 0,0039$ segundos, mientras el DKF, un tiempo de cálculo de $0,0109 \pm 7,67e-04$ segundos. El DKF-rsKF tarda 1,1 milisegundos más que el DKF para identificar el generador eólico en presencia de datos espurios, lo cual no se aleja mucho de los tiempos de ejecución aplicando metodologías estándar.

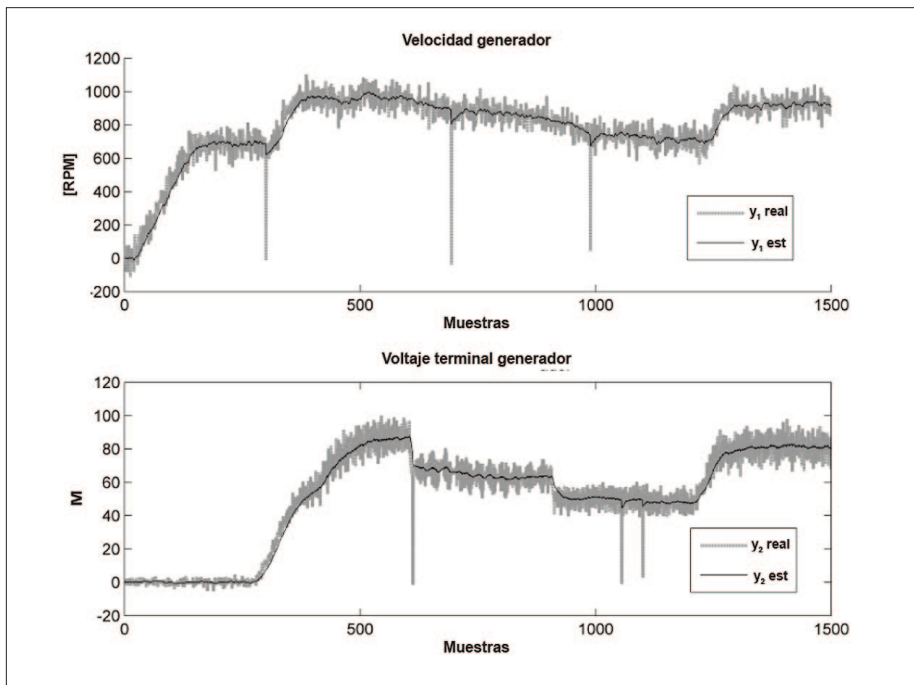
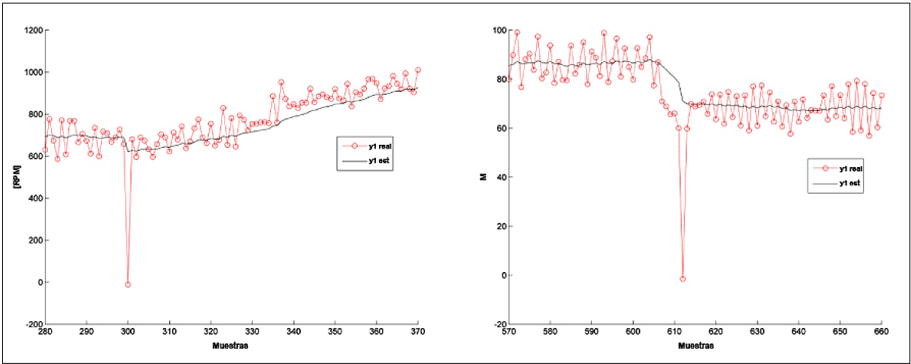


Figura 7. Comportamiento del generador eólico, considerando tratamiento de datos espurios. La línea punteada corresponde a la señal entregada por el sistema de adquisición NI-USB6009 y la línea sólida es la señal entregada por el método robusto. La figura superior es la respuesta del DKF-rsKF para la velocidad del generador; por otro lado, la figura inferior pertenece a la identificación del voltaje en terminales de la máquina.



(a) Velocidad del generador (b) Voltaje en terminales del generador

Figura 8. Respuesta de la velocidad (figura de la izquierda) y voltaje del generador (figura de la derecha). La línea gris discontinua corresponde a la señal real medida y la línea negra sólida corresponde a la señal estimada por el DKF-rsKF. Acercamiento de la figura 7 entre las muestras 280 y 370 para la velocidad del generador y entre las muestras 570 y 660 para el voltaje en terminales.

CONCLUSIONES

En este estudio se presentó una metodología para la identificación de un generador eólico implementado a través de un generador síncrono. El sistema anterior corresponde a un sistema multivariable acoplado, el método propuesto permitió realizar la estimación del modelo sin verse afectado por los datos espurios, lo que sí pasa con el filtro de Kalman dual estándar. También se pudo demostrar que la identificación del generador eólico que exhibe comportamientos no lineales se realiza satisfactoriamente empleando una combinación de métodos lineales que incluyen el DKF y el rsKF.

En particular, se mostró que el filtro de Kalman estándar se ve afectado por datos espurios, obteniendo métricas de validación erróneas, como el caso de la skewness, la kurtosis y el error residual.

Igualmente, se demostró que el rsKF es un método capaz de atenuar el efecto de los datos espurios en las observaciones, alcanzando medidas para la kurtosis y skewness similares en comparación con las señales sin datos espurios. También logró un error residual de 0,87%, el cual es menor que el error obtenido por el sKF (17,82%).

Agradecimientos

Este trabajo se desarrolló en el marco del proyecto de investigación: “Identificación de parámetros de sistemas multivariantes basado en el filtro de Kalman y filtro de partículas dando robustez a outliers”, en el programa de jóvenes investigadores “Virginia Gutiérrez de Pineda” y fue financiado por medio de Colciencias y la Universidad Tecnológica de Pereira.

Apéndice A, motor de inducción

Modelo IM-100 de *Hampden Engineering Corporation*, motor jaula de ardilla con estator bobinado, trifásico, cuatro polos, 240V, 1.5A, 60Hz y 1725 RPM.

Apéndice B, generador síncrono

Modelo IM-100-3 de *Hampden Engineering Corporation*, generador trifásico, cuatro polos, 240V, 1.5A, 60Hz, 0.33HP y 1800 RPM.

Apéndice C, variador de velocidad-Varispeed-606PC3

Equipo que controla la velocidad de máquinas, especialmente motores; el *Varispeed-606PC3*, elaborado por *Yaskawa*, proporciona un modo remoto de operación. Sirve para suministrar un voltaje de salida con una frecuencia proporcional a un voltaje DC de referencia.

Apéndice D, módulo de estado sólido H-IEC-DC2

Es un sistema que realiza el control por medio de rectificadores controlados de silicio (SCR). Este módulo, fabricado por *Hampden Engineering Corporation*, es un equipo para uso académico. El módulo también permite entregar un voltaje variable entre 0-125V DC proporcional a un voltaje de referencia de 0-10V.

Apéndice E, módulo de adquisición

La tarjeta NI-USB6009 de *National Instruments* brinda funcionalidad de adquisición de datos básica para aplicaciones como registro de datos simple y medidas portátiles. Algunas de sus características incluyen: 8 entradas

analógicas, 2 salidas analógicas, rango de voltaje máximo de entrada [-10 10]V y rango de voltaje máximo de salida [0 5]V.

REFERENCIAS

- [1] S. Thomsen, K. Rothenhagen, and F. Fuchs, "Online parameter identification methods for doubly fed induction generators", in *Power Electronics Specialists Conference, 2008. PESC 2008. IEEE*, junio, pp. 2735-2741.
- [2] S. Haykin, *Kalman Filtering and Neural Networks*. John Wiley & Sons, 2001.
- [3] I. C. Schick and S. K. Mitter, "Robust recursive estimation in the presence of heavy-tailed observation noise", pp. 1045-1080, 1994.
- [4] J.-A. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection", *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, octubre-noviembre 2007, pp. 1514-1519.
- [5] G. Agamennoni, J. Nieto, and E. Nebot, "An outlier-robust Kalman filter", in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, mayo 2011, pp. 1551-1558.
- [6] T. Cipra and R. Romera, "Kalman filter with outliers and missing observations", *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, vol. 6, n.º 2, pp. 379-395, 1997.
- [7] P. J. Huber, "Robust estimation of a location parameter", *The Annals of Mathematical Statistics*, vol. 35, n.º 1, pp. 73-101, 1964.
- [8] J. Chiasson, *Modeling and High-Performance Control of Electric Machines*. IEEE Press/Wiley-Interscience, 2005.
- [9] A. Nelson, "Nonlinear estimation and modeling of noisy time-series by dual Kalman filtering methods", Ph.D. dissertation, septiembre 2000.
- [10] A. Mansouri and F. Krim, "Genetic algorithms based parameters identification of induction machine armature model", in *Power Engineering and Optimization Conference (PEOCO), 2011 5th International*, junio 2011, pp. 182-189.
- [11] Y. He, Y. Wang, Y. Feng, and Z. Wang, "Parameter identification of an induction machine at standstill using the vector constructing method", *Power Electronics, IEEE Transactions on*, vol. 27, n.º 2, pp. 905-915, febrero 2012. DOI: 10.1109/TPEL.2010.2089699
- [12] P. Kou, J. Zhou, C. Wang, H. Xiao, H. Zhang, and C. Li, "Parameters identification of nonlinear state space model of synchronous generator", *Engineering Applications of Artificial Intelligence*, vol. 24, n.º 7, pp. 1227-1237, 2011. DOI: 10.1016/j.engappai.2011.05.012

- [13] M. Huang, W. Li, and W. Yan, "Estimating parameters of synchronous generators using square-root unscented Kalman filter", *Electric Power System Research*, vol. 80, n.º 9, pp. 1137-1144, 2010.
- [14] H. Sediki, A. Bechouche, D. O. Abdeslam, and S. Haddad, "Adaline approach for induction motor mechanical parameters identification", *Mathematics and Computers in Simulation*, n.º 0, 2012. DOI: 10.1016/j.matcom.2012.05.003
- [15] A. Meghnous, M. Pham, and X. Lin-Shi, "Dynamic identification of a synchronous machine using an extended sliding mode observer", *Mathematics and Computers in Simulation*, n.º 0, 2012. DOI: 10.1016/j.matcom.2012.05.020
- [16] P. Huber, *Robust Statistics*, 2nd ed. New York: Wiley Series in Probability and Statistics, 2009.
- [17] R. Pearson, "Exploring process data", *Journal of Process Control*, vol. 11, n.º 2, pp. 179-194, 2001. DOI: 10.1016/S0959-1524(00)00046-9
- [18] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer Verlag New York, Inc., 2006.