UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

Construcción y aplicación de herramientas matemáticas para la detección de anomalías en el funcionamiento de aerogeneradores

Agustín López de Lacalle Samaniego

Programa de Posgrado en Ingeniería Matemática
Facultad de Ingeniería
Universidad de la República

Montevideo – Uruguay
Junio de 2020

UNIVERSIDAD DE LA REPUBLICA
URUGUAY

FACULTAD DE INGENIERIA

IMERL

imfia
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE LA REPÚBLICA

Construcción y aplicación de herramientas matemáticas para la detección de anomalías en el funcionamiento de aerogeneradores

Agustín López de Lacalle Samaniego

Tesis de Maestría presentada al Programa de Posgrado en Ingeniería Matemática, Facultad de Ingeniería de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Magister en Ingeniería Matemática.

Director de tesis:
Mag. Ing. Prof. Álvaro Díaz

Director académico:
Dr. Ing. Prof. Gabriel Usera

Montevideo – Uruguay
Junio de 2020

López de Lacalle Samaniego, Agustín

Construcción y aplicación de herramientas matemáticas para la detección de anomalías en el funcionamiento de aerogeneradores / Agustín López de Lacalle Samaniego. - Montevideo: Universidad de la República, Facultad de Ingeniería, 2020.

xi, 144 p. 29, 7cm.

Director de tesis:

Álvaro Díaz

Director académico:

Gabriel Usera

Tesis de Maestría – Universidad de la República, Programa de Ingeniería Matemática, 2020.

Referencias bibliográficas: p. 140 – 144.

1. Predicción de fallas, 2. Aerogenerador, 3. Método predictivo. I. Díaz, Álvaro. II. Universidad de la República, Programa de Posgrado en Ingeniería Matemática. III. Título.

INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

Dr. Prof. Mathias Bourel

Dr. Prof. Martín Draper

Dr. Prof. José León

Montevideo – Uruguay
Junio de 2020

A mi esposa.

Agradecimientos

Quisiera agredecer en primer lugar a Álvaro Díaz. Por su infinita paciencia, por su amplia predisposición, por sus consejos, por compartir conmigo su experiencia y su conocimiento, y por otras tantas cosas más. Sin duda este trabajo no hubiese sido posible sin su dirección.

También agradezco a Gabriel Usera, quien me ha aconsejado en todo momento sobre decisiones importantes tomadas en este transcurso.

A mi familia y amigos, por haberme apoyado y motivado en todo momento. En particular, agradecerle a mi madre y a mi padre, a quienes les debo mis logros. También a mis hermanos, a Daniela y a Jorge. Y a mi abuela que, aunque ya no esté, siempre me acompaña.

Agradezco también a mis compañeros de trabajo que, de una forma u otra, me han acompañado; del IMERL, de CCC y de Copagran.

A la Comisión Académica de Posgrado por haberme dado la posibilidad de realizar mis estudios de posgrado con la mayor dedicación posible, a través del otorgamiento de una beca.

Finalmente, y más importante, quisiera agradecer a mi esposa. Es ella quien estuvo presente en todo momento, desde el comienzo hasta el final de esta etapa, alentándome y acompañándome, tanto en las buenas circunstancias como en las no tan buenas. Su apoyo fue realmente incondicional.

Nunca se puede predecir un acontecimiento físico con una precisión absoluta.

Max Planck

RESUMEN

La predicción de fallas en aerogeneradores es un tema en pleno auge a nivel mundial, para el que diversos autores han propuesto técnicas para reducir los costos de Operación y Mantenimiento asociados. En esta tesis se presentan cuatro herramientas de detección de anomalías en la operación de aerogeneradores, basadas en datos provenientes de sistemas SCADA.

Los métodos presentados buscan responder a distintas naturalezas y enfoques dentro del universo de herramientas existentes para este fin; (1) la construcción de un modelo probabilístico basado en un Proceso Gaussiano, (2) la generación de un modelo no lineal que minimiza el error entre las observaciones y lo modelado, (3) el estudio de la evolución de la curva de potencia a partir de Cópulas y, (4) una técnica de aprendizaje automático basada en Support Vector Machine, incorporando el método de Componentes Principales.

Cada uno de estos métodos son puestos a evaluación en hasta cuatro casos de estudio reales, ya sea con el fin de predecir fallas asociadas al funcionamiento de los aerogeneradores o, identificar cambios en el funcionamiento mediante un monitoreo por condición. Finalmente, los resultados abordados son cuantificados con el fin de comparar el desempeño de estos algoritmos.

Palabras claves:
Predicción de fallas, Aerogenerador, Método predictivo.

ABSTRACT

The prediction of wind turbine failures is a booming topic worldwide, for which various authors have proposed techniques to reduce associated Operation and Maintenance costs. This thesis presents four tools for detecting anomalies in the operation of wind turbines, based on data from SCADA systems.

The methods presented seek to respond to different natures and approaches within the universe of existing tools for this purpose; (1) the construction of a probabilistic model based on Gaussian Processes, (2) the generation of a non-linear model that minimizes the error between observations and modeling values, (3) the study of the evolution of the power curve using Copulas and, (4) a machine learning technique based on Support Vector Machine, incorporating Principal Components Analysis.

Each of these methods are evaluated in up to four real study cases, either in order to predict failures associated to the operation of wind turbines or to identify changes in operation through condition monitoring. Finally, the results addressed are quantified in order to compare the performance of these algorithms.

Keywords:
Faults prediction, Wind turbine generator, Predictive method.

Tabla de contenidos

Capítulo 1

Introducción

La energía eólica es la energía cinética contenida en las partículas de aire en movimiento respecto a la superficie de la Tierra. Una partícula de masa M que se mueva con una velocidad V tendrá una energía cinética como la expresada en la siguiente ecuación.

$$E = \frac{MV^2}{2} \tag{1.1}$$

Si un flujo de aire, de densidad ρ, de velocidad V atraviesa una superficie de área A, suponiendo que el vector velocidad es normal a la superficie, se establece un flujo de energía cinética por unidad de tiempo, es decir, un flujo de potencia, que puede expresarse como sigue.

$$P = \frac{\rho A V^3}{2} \tag{1.2}$$

Esta expresión daría una estimación de la potencia eólica disponible que atraviesa una superficie de área A y que para utilizarse, se debería convertir en una forma útil de energía como puede ser energía mecánica o energía eléctrica, entre otras.

Los equipos utilizados para convertir la energía eólica en energía eléctrica se denominan aerogeneradores y se componen de una turbina eólica, que convierte la energía eólica en energía mecánica disponible en un eje que gira, y un generador de energía eléctrica.

Una clasificación común para los aerogeneradores consiste en ordenarlos de acuerdo a la disposición de su eje principal respecto al suelo o al flujo del viento. Esto desprende dos grandes grupos: los de eje vertical y los de eje horizontal. En la Figura 1.1 ambos aerogeneradores son de eje horizontal. En la Figura

Figura 1.1: Aerogenerador on-shore (izquierda) y aerogenerador off-shore (derecha).

1.2 se presenta un aerogenerador de eje vertical.

Figura 1.2: Aerogenerador de eje vertical.

Los aerogeneradores de eje horizontal están compuestos principalmente por una torre, una góndola (donde se ubican los principales componentes eléctricos) y un rotor. Por otra parte, una de las principales ventajas de los aerogeneradores de eje vertical es la accesibilidad de los componentes eléctricos. Esta diferencia y los principales componentes de cada uno se ilustran en las Figuras 1.3 y 1.4. Los de eje horizontal suelen tener su eje ubicado a cierta altura

Figura 1.3: Principales componentes de un aerogenerador de eje horizontal.

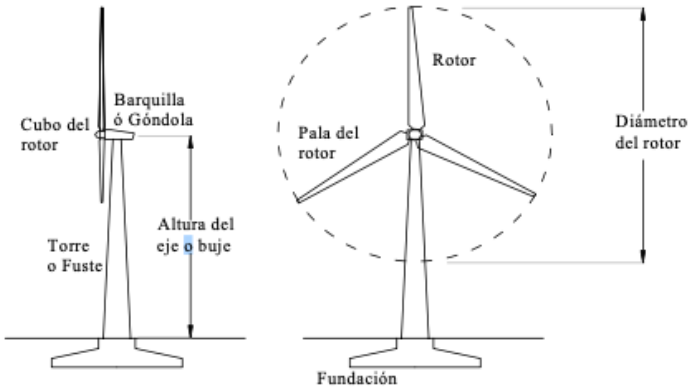
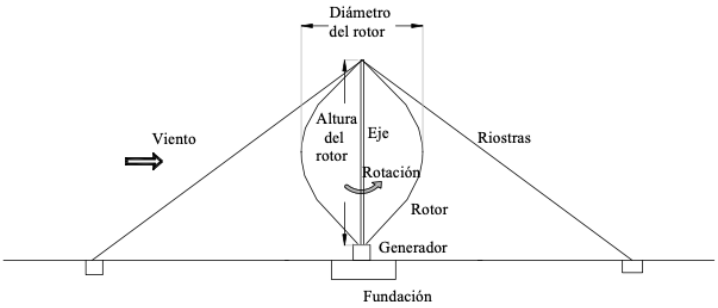


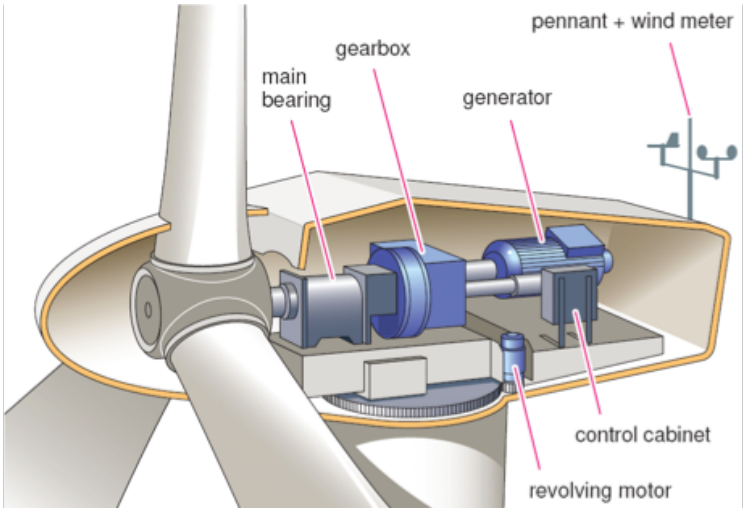
Figura 1.4: Principales componentes de un aerogenerador de eje vertical.



sobre el suelo; para ello es que están montados sobre la torre. El aerogenerador es acoplado a un generador (sincrónico o asincrónico) eléctrico, usualmente a través de una caja multiplicadora. Estos últimos son los dos principales componentes eléctricos del sistema, que se alojan dentro de la góndola, elevada a la altura de la torre. Dentro de la góndola también se encuentran otros componentes como el freno, el transformador, el sistema de orientación, el sistema de detección de vibraciones, la central hidráulica, el sistema de control, el sistema

de desenrrollamiento, entre otros. La Figura 1.5 esquematiza el interior de una góndola típica.

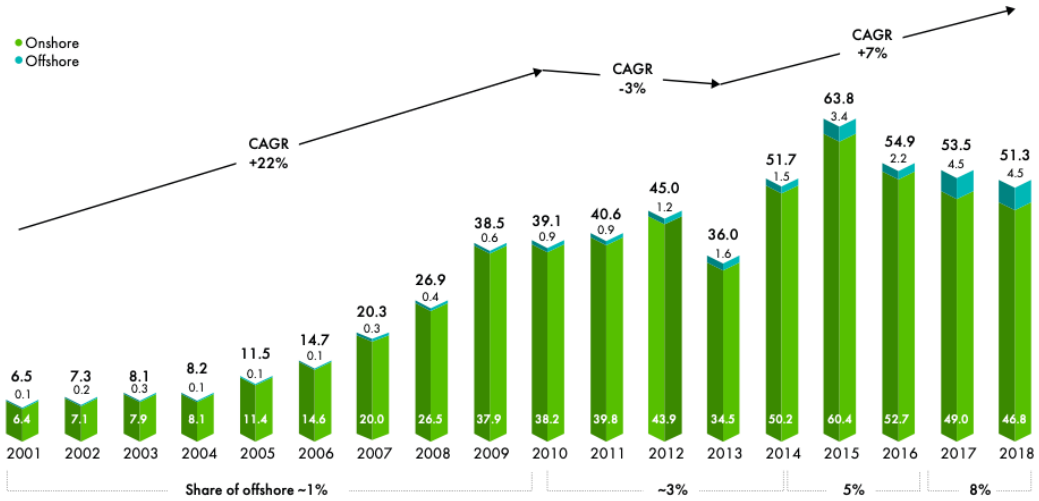
Figura 1.5: Interior de una góndola.



El uso del viento para generar energía tiene sus raíces desde hace años atrás, cuando era empleada con fines domésticos para moler granos. La generación de energía eléctrica a partir del viento comenzó a principios del siglo pasado, pero fue solo luego de los 80s cuando se logró una producción a gran escala. Las siguientes décadas, la energía eólica mostró un continuo crecimiento en lo que refiere a capacidad instalada. La Figura 1.6 muestra la evolución anual de la capacidad global instalada durante los años 2001 y 2018. La energía eólica ha tenido un importante desarrollo internacional, con tasas de crecimiento de la capacidad instalada de 20 % en las últimas décadas (Sawyer et al. (2019)). A fines del 2003 la capacidad instalada estaba por encima de los 40000MW, duplicando la potencia alcanzada en el año 1999 (Amirat et al. (2009)).

En Uruguay se ha dado un importante crecimiento de esta tecnología, constituyendo una de las principales fuentes de transformación de la matriz eléctrica nacional. Este desarrollo fue posible gracias a avances tecnológicos relevantes en las últimas décadas, así como a un marco nacional e internacional favorable. En Uruguay, la instalación del orden de 1000MW de energía eólica en tres años ha implicado importantes desafíos, particularmente generando capacidades antes inexistentes y requiriendo ampliar algunas poco desarrolladas. Así, Uruguay se presenta como un caso de referencia a nivel internacional, con capacidades competitivas a nivel regional en lo que concierne al desarrollo de

Figura 1.6: Evolución anual de la capacidad global instalada de energía eólica.



proyectos de parques eólicos. Al día de hoy, la capacidad instalada de energía eólica asciende los 1500MW, contándose con más de 600 aerogeneradores distribuidos a lo largo de todo el país.

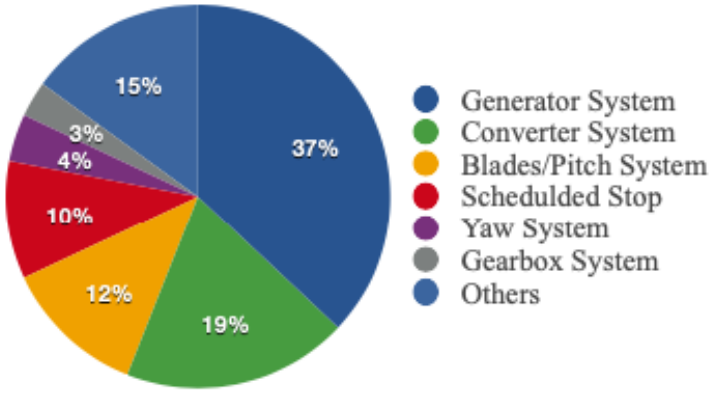
Al ser un aerogenerador un generador eléctrico, con todos los demás componentes que lo constituyen, existen diferentes factores que pueden influenciar en el desempeño de la turbina, como la velocidad de viento o el tamaño de la máquina. Un aerogenerador on-shore promedio de 3MW puede producir más de 6000 MWh en el año. Las turbinas operan con velocidades de viento que van desde los 4m/s a los 25m/s en términos generales. Asimismo, un aerogenerador promedio produce electricidad en el 70-85 % del tiempo (Purarjomandlangrudi (2014)).

Los aerogeneradores usualmente están ubicados en zonas remotas o de difícil acceso. Esto hace que la falla de un componente clave del mecanismo resulte en una parada operacional, produciendo así pérdidas económicas. La detección temprana de anomalías en el funcionamiento permite preservar la capacidad operativa del aerogenerador, facilitar el mantenimiento proactivo y minimizar el tiempo fuera de servicio, maximizando así la productividad. En ese sentido, el análisis predictivo de fallas en aerogeneradores es una rama que está en auge hoy en día, para la que diversos autores han expuesto resultados provenientes de variada naturaleza (Tautz-Weinert and Watson (2016)).

Los costos asociados a Operación y Mantenimiento (OyM) representan

aproximadamente el 10-15 % y 20-25 % de los costos de generación de energía de parques eólicos on-shore y off-shore, respectivamente. En particular, las paradas de un aerogenerador reducen la fiabilidad en la energía eólica e incrementan los costos de OyM. El generador representa aproximadamente el 10 % del costo total de la turbina eólica, cuyas fallas son una de las principales causas de los tiempos de inactividad. La Figura 1.7 ilustra la distribución de paradas, por componentes del aerogenerador, de dos parques eólicos en China (Zhao et al. (2017)).

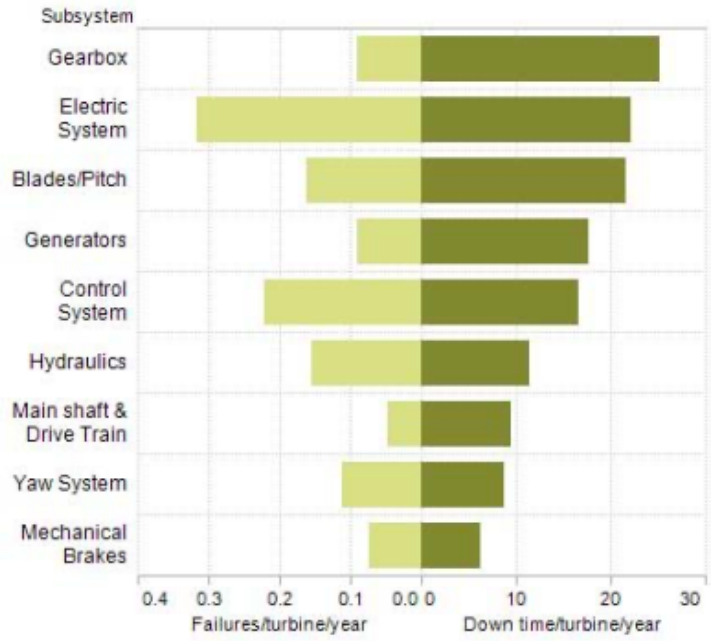
Figura 1.7: Porcentaje de paradas de aerogeneradores de dos parques eólicos en China.



Los costos de mantenimiento se pueden reducir mediante el monitoreo continuo y automatizado de las turbinas eólicas. El monitoreo y análisis de datos permiten ajustes de rendimiento y mantenimiento basados en condiciones más que intervalos de tiempo. La experiencia en otras industrias muestra que el monitoreo de condición detecta fallas antes de que alcancen una etapa catastrófica o de daño secundario, extiende la vida útil de los componentes, permite una mejor planificación logística y de mantenimiento, y puede reducir el mantenimiento de rutina. Un sistema de Supervisory Control and Data Acquisition (SCADA) utiliza datos que ya se recogen en el controlador del aerogenerador y es una forma rentable de monitorear la alerta temprana de fallas y problemas de rendimiento. Un sistema de monitoreo que esté basado en condiciones debe estar diseñado para proporcionar el máximo beneficio dado su elevado costo. Como no todas las fallas pueden detectarse ni prevenirse, tiene sentido enfocarse en las fallas que son más costosas de reparar. La Figura 1.8 muestra las tasas promedio de fallas y los tiempos de inactividad asociados a los dife-

rentes sub-sistemas de los aerogeneradores (información obtenida a partir de un relevamiento bibliográfico basado en artículos que presentan datos reales). En este contexto, el tiempo de inactividad está presentado como un indicador indirecto de costo y de esfuerzo de reparación (Kim et al. (2011)).

Figura 1.8: Tasas promedio de fallas y tiempos de inactividad.

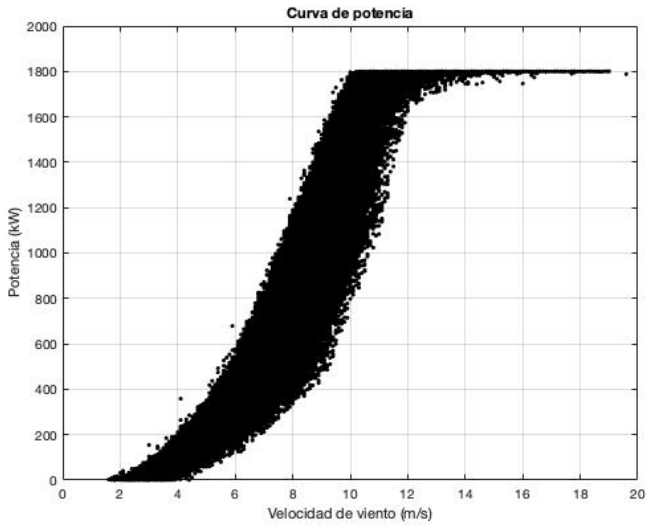


La detección de fallas y monitoreo por condición de aerogeneradores es un campo de investigación totalmente activo. Diferentes técnicas de variada naturaleza han sido desarrolladas con el objetivo principal de reducir los costos de OyM. En particular, la mayoría de estas técnicas están basadas en datos provenientes de sistemas SCADA. En este sentido, Tchakoua et al. (2014) realizan una presentación y clasificación general de métodos de monitoreo por condición. Los autores presentan además una perspectiva a futuro sobre el desarrollo de nuevos algoritmos que puedan surgir de acuerdo al estado del arte presentado en el artículo, a las necesidades industriales del momento y a las capacidades tecnológicas. Tautz-Weinert and Watson (2016) presentan una revisión de las principales técnicas desarrolladas a partir de datos SCADA. Ellos clasifican las técnicas en cinco grupos: métodos de tendencia, de clustering, modelado de comportamiento normal, modelado de daños, y evaluaciones de alarmas y sistemas expertos. Asimismo, Hameed et al. (2007) también presentan una revisión extensiva de los diferentes métodos abordados

hasta el momento, enfocándose en algoritmos desarrollados tanto para el monitoreo del rendimiento de la turbina, como para una detección temprana de anomalías. Finalmente, Wilkinson et al. (2014) realizan una comparación de tres métodos de monitoreo por condición para aerogeneradores: uno basado en tendencia de señales, otro abordado desde los mapas auto-organizados[1], y un modelo físico. Para este último los autores encuentran mejores resultados a la hora de predecir fallas inminentes de componentes.

Diversas técnicas fueron desarrolladas en torno al estudio específico de la curva de potencia del aerogenerador, ya que esta representa una curva de rendimiento del mismo, en el sentido de que es un diagrama de dispersión entre la entrada al sistema (velocidad de viento) y su salida (potencia generada). En la Figura 1.9 se ilustra la curva de potencia de un aerogenerador de 1800 kW. Lydia et al. (2014) realizan un análisis sobre la necesidad de modelar las curvas

Figura 1.9: Curva de potencia de un aerogenerador de 1800 kW de potencia nominal.



de potencia de los aerogeneradores, haciendo una revisión de las diferentes metodologías empleadas para modelar las mismas. Ellos hacen una clasificación en cuanto a técnicas paramétricas y no-paramétricas, complementando con un enfoque crítico de cada una de ellas. Schlechtingen et al. (2009) realizan un

[1]Los mapas auto-organizados son un tipo de red neuronal artificial que es entrenada usando aprendizaje no supervisado para producir una representación discreta del espacio de las muestras de entrada.

estudio comparativo de técnicas basadas en 'data-mining' y abocadas al monitoreo de la curva de potencia del aerogenerador. Ellos construyen tres modelos, evalúan su desempeño, y luego los comparan; un modelo de clustering, uno de redes neuronales y otro de k-vecinos. Gill et al. (2012) proponen un método donde, a partir de datos de los datos de las turbinas eólicas, se estiman funciones de distribución de probabilidad bivariadas que representan la curva de potencia; este método es conocido como Cópulas. Una de las principales virtudes de este trabajo es el reconocimiento temprano de fallas incipientes a partir de los resultados obtenidos de las Cópulas, logrando una caracterización de los mismos asociados a los distintos tipos de fallas como son la degradación de la pala, fallas en el yaw[2] y errores de pitch[3]. En este sentido, Wang et al. (2014) proponen un método probabilístico, desarrollado a partir de la distribución de probabilidad conjunta (Cópulas) de las variables involucradas, para la extracción de outliers de la curva de potencia de un aerogenerador. Por otra parte, Jia et al. (2016) proponen un algoritmo donde modelan la performance del aerogenerador en la zona cuasi-lineal de la curva de potencia a partir de un análisis de componentes principales (PCA), identificando así cambios de la misma a lo largo del tiempo. Pandit and Infield (2018) proponen un algoritmo de monitoreo por condición basado en un modelo de Proceso Gaussiano. Ellos comparan el modelo de la curva de potencia a partir del establecido por la norma International Electro-Technical Comission (2016) con el que ellos proponen, identificando las ventajas de su propuesta. Presentan un caso de estudio asociado a una falla correspondiente al yaw, y hacen énfasis en las fortalezas de su método al momento de la comparación. Skrimpas et al. (2014) realizan un estudio de la curva de potencia observando la evolución temporal de los valores propios correspondientes a la distribución bidimensional de la velocidad de viento y potencia generada. Un aspecto interesante de este abordaje consiste en la clasificación de la potencia en cinco subgrupos: el estudio es llevado a cabo para distintos niveles de potencia asociadas al aerogenerador, logrando de esta forma una mejor resolución de las herramientas y una identificación de la causa del problema raíz más clara. Ellos muestran que un cambio en la evolución temporal de los valores propios asociados a la matriz de covarianza de los datos está asociado a una anomalía. Lapira et al. (2012) proponen tres métodos

[2]El término *yaw* refiere al ángulo formado entre la normal al plano del rotor y la dirección del viento incidente.

[3]El término *pitch* refiere al ángulo que orienta la pala del aerogenerador.

con un enfoque de múltiples regímenes que permiten considerar las condiciones dinámicas del trabajo de la turbina eólica. Estos métodos, desarrollados también a partir de datos provenientes de sistemas SCADA, son sometidos a evaluación con el fin de determinar sus capacidades de medir degradaciones en la turbina antes de que ocurran los eventos de inactividad conocidos a priori. Parte del interés en este trabajo está en la implicación de las redes neuronales para el desarrollo de los métodos. Finalmente, Herp et al. (2016) presentan un modelo orientado a datos para monitorear parques eólicos basado en clasificadores bayesianos y análisis multivariado de la curva de potencia. Los autores introdujeron un nuevo criterio para detectar outliers y varias cotas de control sobre la oblicuidad y curtosis de los datos para la separación en grupos (k-means) y clasificación de turbinas operando con o sin fallas.

Otros abordajes consisten en técnicas que escapan a la curva de potencia, incluyendo el uso de variables adicionales a la velocidad de viento y potencia, presentes en general en el sistema SCADA. En este sentido, Sánchez and Couso (2011) utilizan una generalización del análisis de espectro singular aplicado a datos mal definidos. Esta metodología fue aplicada en 40 aerogeneradores de un parque eólico real, encontrando que oscilaciones en la presión del circuito hidráulico de los frenos estaban correlacionadas con daños a largo plazo en uno de los rodamientos de la turbina. Kim et al. (2011) realizan en primer lugar una exploración de los datos SCADA de aerogeneradores, con el fin de evaluar la capacidad de estos a la hora de desarrollar técnicas de diagnóstico y detección de fallas en el equipo. En base a esta exploración, y a partir de una serie de mediciones de variables, desarrollan algunos algoritmos de detección de anomalías, investigando técnicas de clustering y análisis de componentes principales. Los autores finalmente someten los algoritmos al estudio de casos reales. Catmull (2010) presenta resultados de un método de monitoreo por condición basado en mapas auto-organizados. El autor destaca que esta herramienta, si bien permite determinar cuándo el comportamiento de la turbina comienza a ser anómalo, no tiene el alcance necesario como para definir cuál es la falla en el sistema. La principal virtud de este artículo es la de incorporar el uso de variables de SCADA, distintas a la velocidad de viento y la potencia, a la construcción de un mapa auto-organizado. Yang et al. (2013) proponen un algoritmo multi-variado de monitoreo de condición a partir de datos históricos del SCADA del aerogenerador. Observando las correlaciones entre los registros de diversas variables, construyen una técnica capaz de detectar tanto fallas

incipientes en diversos componentes de la turbina, como de trazar el deterioro continuo del equipo. A partir de la discretización en el tiempo de las curvas bi-variadas de diferentes registros del SCADA, los autores validan sus resultados en dos casos de estudio. Astolfi et al. (2014) presentan un trabajo donde enfocan sus resultados en el análisis de diversas temperaturas registradas en el SCADA del aerogenerador. Todas las temperaturas registradas se asocian a la producción de energía de la turbina, con el fin de desarrollar rutinas automáticas que permitan monitorear la evolución de los datos. La técnica presentada es validada en un parque eólico real, logrando identificar un problema mecánico del equipo y un problema asociado al enfriamiento del mismo.

Algunos autores han implementado métodos aplicados a componentes particulares de los aerogeneradores, sobre todo teniendo en cuenta la información antes presentada en la Figura 1.8. Bertelé et al. (2018) proponen una metodología para identificar si alguna pala se encuentra desalineada, cuantificar dicho desbalance y corregirlo, testeando dicha estrategia mediante simulaciones de rotores desbalanceados con un código aero-servo-elástico utilizando Blade Element Momentum (BEM) para contemplar la aerodinámica del problema. En la misma linea, Mittelmeier and Kuhn (2018) propusieron utilizar datos de SCADA promedio de 1 minuto para estimar la desorientación de un aerogenerador evaluando la relación potencia-yaw. Laouti et al. (2011) emplean una técnica de Support Vector Machine (SVM) con una función de kernel radial para analizar distintos componentes del sistema, tales como actuadores y sensores. Con la duplicación de algunos sensores, fueron capaces de reconocer fallas en el pitch, en el generador y en el rotor. El trabajo presentado por Zhao et al. (2017) describe un algoritmo capaz de reconocer fallas en el generador de la turbina con una precisión del 94 %. La principal virtud que ellos destacan es la falta de necesidad de instalar hardware adicional al SCADA. Purarjomandlangrudi (2014) presenta en su trabajo de tesis dos algoritmos de aprendizaje para la detección de fallas en los rodamientos y para el monitoreo por condición del aerogenerador. A partir de datos de entrenamiento, el autor construye un primer algoritmo que es capaz de reconocer la presencia de datos inusuales y fallas potenciales, el cual aplica a un nuevo conjunto de datos. En segunda instancia, entrenando un algoritmo de SVM, testea un segundo algoritmo encontrando fallas en los rodamientos en un aerogenerador real. Wodecki et al. (2017) presentan una herramienta basada en el Análisis en Componentes Independientes (ICA) para la extracción de información de

medidas de temperatura de grandes cajas multiplicadoras. La aplicación que ellos presentan es para instalaciones en la industria minera, aunque para las mediciones de datos SCADA de aerogeneradores esta herramienta es igual de aplicable. Dada las fluctuaciones y ruido registrado en las señales de adquisición, es difícil en general reconocer patrones a simple vista. Teniendo esto en cuenta, a partir de señales de cuatro cajas multiplicadoras, ellos pueden, con una nueva señal, aplicar reglas de detección automática para reconocer el cambio en el funcionamiento de la caja. Feng et al. (2013) realizan, en primera instancia, una revisión general de modos de fallas de las cajas multiplicadoras de turbinas eólicas. Finalmente, presentan dos casos de estudio de dos aerogeneradores; uno usando datos SCADA y otro usando otro tipo de señales destinadas al control de monitoreo por condición. El primer método está basado en modelos que parten de leyes físicas aplicadas a la caja multiplicadora y las diferentes variables sensadas asociadas; el segundo contempla las amplitudes de las vibraciones de las señales para identificar fallas en la caja. Wang and Infield (2013) exponen una Técnica de Estimación No-Lineal del Estado (NSET) de una caja multiplicadora saludable[4] de un aerogenerador. Los datos SCADA históricos deben capturar la relación entre la caja multiplicadora y distintos componentes de la turbina, tanto en su operación nominal como en sus condiciones extremas para obtener un desempeño satisfactorio del modelo. Comparando el modelo construido con datos de operación de aerogeneradores, y a través de un test estadístico, los autores logran demostrar la efectividad del método en dos situaciones reales.

Existen también métodos destinados al análisis de las vibraciones de determinados componentes del aerogenerador. En ese sentido, Zhang and Kusiak (2012) comparan el desempeño de tres métodos para detectar anomalías en turbinas eólicas, basados en un análisis de vibraciones de las mismas. Estas vibraciones se caracterizan por dos parámetros principales: el sistema de transmisión del equipo y la aceleración de la torre. Para el desarrollo de estos algoritmos, fueron empleados datos SCADA con una frecuencia de muestreo de 10 segundos. Por otra parte, Martínez-Rego et al. (2011) diseñan un sistema de predicción de fallas en aerogeneradores a partir de señales de vibración de turbinas eólicas. El algoritmo que ellos presentan está basado en One-Class SVM: a partir de datos saludables correspondientes al aerogenerador, ellos construy-

[4]Decimos que un aerogenerador, o un componente del mismo, es saludable cuando no conocemos fallas asociadas a su funcionamiento.

yen un modelo que, a posteriori, es capaz de discernir entre datos normales de operación y fallas asociadas. La principal virtud que destacan sobre su trabajo es que su algoritmo tiene la posibilidad de medir la evolución de una falla asociada al equipo. Finalmente, los autores validan su método predictivo en tres escenarios diferentes: un escenario simulado numéricamente, un escenario experimental controlado, y un tercero con datos reales de operación.

Finalmente, otros enfoques tienen asociados otro tipo de datos diferentes a los provenientes de sistemas SCADA, como pueden ser señales de sensores adicionales colocados específicamente para el control de monitoreo por condición. También existen métodos que aplican conjuntos de datos como son los meteorológicos. En este sentido, Beltrán et al. (2012) proponen un método para detectar las desviaciones de la velocidad de viento de los anemómetros de la góndola del aerogenerador, comparándolos con los anemómetros cercanos que haya instalado. Esta comparación se realiza mediante un enfoque para estimar la velocidad de viento de cada góndola, discretizando los datos de velocidad de viento de acuerdo al método de bin[5].

Queda patente entonces que el desarrollo de técnicas que permitan tanto anticipar fallas incipientes en aerogeneradores, como monitorear el estado de condición del mismo, es un tema de interés e investigación activa a nivel mundial en el día de hoy. Asimismo, los métodos existentes son de naturaleza realmente diversa. Esta es la motivación principal de esta tesis. El objetivo de este trabajo es entonces analizar cuatro métodos para el procesamiento de datos SCADA que combinan hallazgos de investigaciones previas, experiencia interna de colegas, y nuevas ideas que fueron surgiendo a lo largo del camino. Cada uno de estos cuatro algoritmos son puestos a evaluación en hasta cuatro casos de estudio reales. Todos los resultados que se pueden obtener mediante estos métodos están condicionados a la disponibilidad de la información asociada a los aerogeneradores. Finalmente, se realiza una evaluación y comparación del potencial de estas técnicas.

Los cuatro métodos elegidos buscan abarcar la mayor parte posible del espectro de técnicas existentes. En primer lugar, el presentado en la sección 3 es un método probabilístico que es capaz de modelar una variable del SCADA a partir de otro subconjunto de variables, basándose en una distribución de probabilidad modelada; representa al grupo de métodos de carácter probabilístico. El método presentado en la sección 4 se basa en una modelación no lineal, que

[5]El método de bin es el propuesto en International Electro-Technical Comission (2016).

surge de la resolución de un problema de mínimos cuadrados, pasando por un algoritmo de selección de datos; este representa al grupo de métodos que también modela una variable del SCADA, pero desde una perspectiva distinta a la probabilística. Luego, el método abordado en la sección 5 se centra fundamentalmente en el estudio de la curva de potencia, que como se mostró, es un enfoque de relevancia dentro del tema de la tesis. Por lo tanto, este método estaría representando al grupo de métodos destinados al estudio de la curva de potencia. Finalmente, en la sección 6, se presenta un método inédito, aunque basado en una composición de métodos de otros autores, que se basa en una técnica de aprendizaje automático (SVM), y que tiene además un enfoque algebraico. En este sentido, los cuatro métodos responden a distintas naturalezas y enfoques, teniendo todos ellos una fuerte componente matemática. Lo anterior ha conducido a la selección de los métodos, generando así una selección de cuatro técnicas que pretenden representar parte del universo existente para el abordaje de esta temática.

Capítulo 2

Descripción de los datos disponibles

En este capítulo se presentan y describen los datos de los aerogeneradores empleados a lo largo de la tesis. En primera instancia, se destaca que, como los datos disponibles corresponden a parques eólicos en funcionamiento, y con el fin de preservar la confidencialidad de las empresas proveedoras de los mismos, se referirá a los parques con nombres que no son reales. Más precisamente, se dispone de datos de cuatro aerogeneradores que son motivo de estudio, pertenecientes a distintos parques, a los que se nombran como A, B, C y D.

Se trabaja en la tesis con datos de SCADA. En general, diversas magnitudes son relevadas en un régimen diezminutal, obteniendo así series temporales de una gran variedad de parámetros influyentes en un aerogenerador, tales como la velocidad de viento, la potencia generada, temperaturas de diversos componentes, presión de diversos fluidos involucrados en la operación, revoluciones del generador, entre otros tantos.

Debe tenerse en cuenta la siguiente consideración a la hora de analizar los datos: es necesario independizar el problema de la variación estacional. Esta independización se realiza de acuerdo a lo expuesto en International Electro-Technical Comission (2016), donde la velocidad de viento es corregida de acuerdo a las mediciones de temperatura ambiente y presión ambiente (involucradas en la densidad del aire de la Ecuación 2.1), según la Ecuación 2.2.

$$\rho = 1.225 \left(\frac{288.15}{T}\right) \left(\frac{P}{1013.3}\right) \tag{2.1}$$

$$V_C = V_M \left(\frac{\rho}{1.225}\right)^{\frac{1}{3}} \tag{2.2}$$

donde V_C y V_M son las velocidades corregidas y medidas, respectivamente; P es la presión ambiente medida en mbar, y T es la temperatura ambiente medida en grados Kelvin. Por no tener registros de presión ambiente en el conjunto de datos de ninguno de los aerogeneradores, la Ecuación 2.1 fue modificada como indica la Ecuación 2.3.

$$\rho = 1.225 \left(\frac{288.15}{T}\right) \tag{2.3}$$

Esta corrección se aplica para todos los aerogeneradores estudiados en este trabajo.

Cabe destacar que los aerogeneradores presentados a continuación no necesariamente disponen de las mismas variables SCADA registradas, lo cual pudo haber limitado algunas aplicaciones que pudieron haber sido implementadas en las secciones subsiguientes. Asimismo, se destaca la incertidumbre asociada a la información asociada a la naturaleza de las fallas ocurridas para los distintos equipos. Esto hace que los resultados presentados en lo que sigue estén fundamentalmente sujetos a esta incertidumbre.

2.1. Aerogenerador A

Los datos provenientes del aerogenerador de A constan de 47 señales relevadas en un régimen diezminutal, entre las cuales se encuentran, además de mediciones de diversos parámetros de la turbina, algunos indicadores referidos a alarmas detectadas por las mismas mediciones, a la limitación de potencia, a la operatividad del aerogenerador durante cada diezminutal, entre otros.

Conjuntamente con los datos obtenidos para el aerogenerador A, también se obtuvieron datos asociados a un aerogenerador, A_0, instalado en el mismo parque eólico y del cual no se registran fallas asociadas. Los datos de A_0 serán empleados en lo que sigue como fuente de datos saludables. Se trata de dos aerogeneradores iguales de 1800kW instalados en un mismo parque eólico de Uruguay.

La información recopilada por la fuente proveedora de los datos de este aerogenerador, narra que a fines de Abril de 2016 y a principios de Setiembre de 2016 se produjeron fallas asociadas a una de las fases del generador.

A raíz de esta información, resulta interesante enfocarse en particular en

determinadas variables del aerogenerador A. A saber: velocidad de viento, potencia generada, las dos temperaturas disponibles de los rodamientos del generador, las tres temperaturas de las fases del generador y las revoluciones del generador. Esta pre-selección está basada fundamentalmente en considerar que estas variables son las que, a priori, pueden estar vinculadas de una forma más directa con la falla ocurrida en A, por lo que serán objeto de estudio en lo que sigue. Otras variables que también interesa incorporar al análisis son las alarmas del SCADA identificadas en los diezminutales[1], la operatividad del aerogenerador durante cada diezminutal[2] y la limitación de producción impuesta externamente por el operario[3].

Los datos del aerogenerador A están comprendidos entre febrero de 2016 y agosto de 2018. Existen en ese período datos faltantes, que bien pueden estar asociados a paradas de mantenimiento preventivo programadas, a interrupciones imprevistas en la producción, o bien a errores en los propios sensores del sistema SCADA. En los 148843 diezminutales que componen cada una de las series, hay 12794 datos faltantes; teniendo entonces un 91 % de diezminutales con registros durante la producción de la turbina. Cabe resaltar que, en las secciones subsiguientes no necesariamente se tendrá en cuenta todo el período disponible a la hora del análisis, sino que el estudio estará enfocado en una parte de este, más precisamente en el que se tiene información externa sobre las fallas.

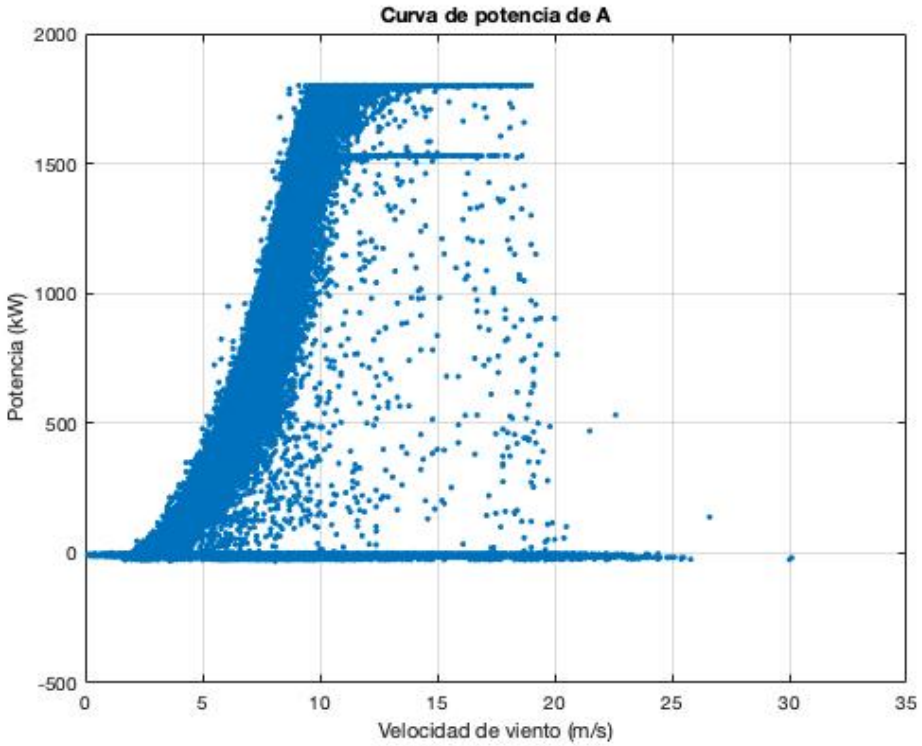
Como se mencionó previamente, los datos precisan ser filtrados en una primera instancia con el fin de remover los puntos correspondientes a una operación anómala a priori. En este sentido, se presenta en la Figura 2.1 la curva de potencia asociada al aerogenerador, que consiste en un gráfico de puntos entre la velocidad de viento y la potencia registrada. Se visualiza que algunas potencias negativas fueron registradas. A su vez, también figuran allí puntos de operación donde el aerogenerador trabajó bajo un límite de potencia impuesto externamente. Finalmente, algunos de los puntos de operación representados

[1]En general los sistemas SCADA en aerogeneradores tienen un registro de alarmas. Este indica la primera alarma registrada por el sistema durante el diezminutal. Las alarmas tienen una codificación numérica que está especificada en el manual del equipo, proporcionado por el fabricante, y pueden deberse a diversos motivos; desde operacionales hasta informativos.

[2]La operatividad del aerogenerador en el diezminutal refiere a la cantidad de segundos en que el equipo estuvo operativo, pudiendo alcanzar hasta un máximo de 600 segundos.

[3]Esta limitación de producción está asociada en general a estrategias de operación impuestas en el parque eólico. Es una limitación de generación de potencia que el equipo recibe externamente, restringiendo la capacidad de generar potencia de la turbina a un valor específico; igual o menor a la potencia nominal del equipo.

Figura 2.1: Curva de potencia asociada al aerogenerador A.



fueron obtenidos con bajo porcentaje de operatividad en el correspondiente diezminutal, así como con alarmas registradas durante ese intervalo. Para los fines propuestos en este trabajo, todos los puntos anteriormente enumerados deben ser filtrados, incluyendo además una extracción manual de outliers de la curva. La curva de potencia pos-filtrado puede verse en la Figura 2.2. La cantidad de datos resultantes con registros, luego de realizado el filtrado, es de 63396; teniendo que la cantidad de datos depurados es de 72653.

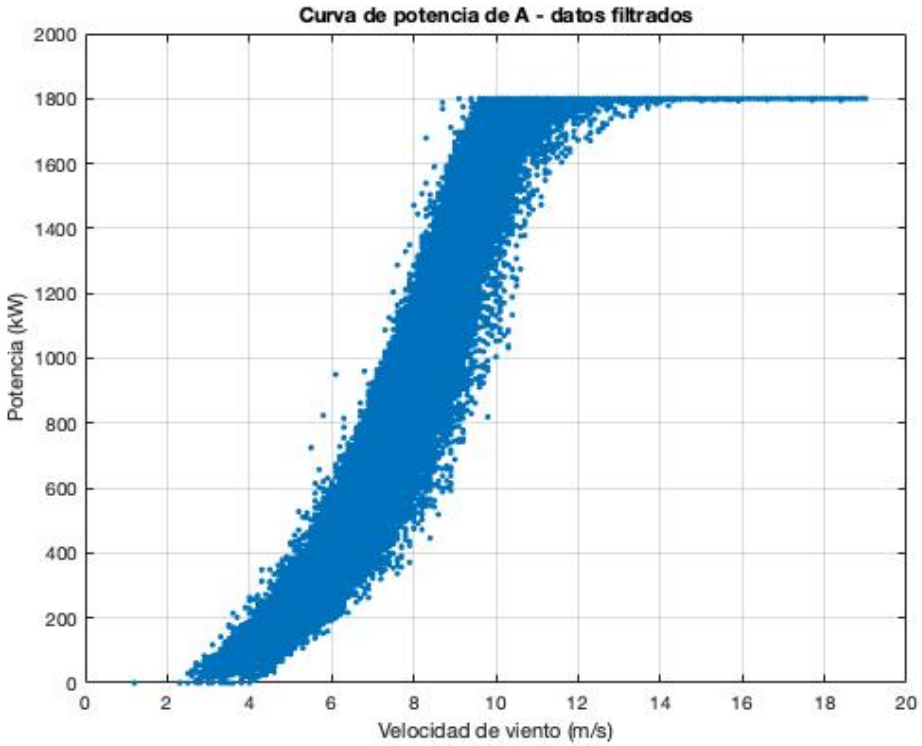
En la Figura 2.3 se presentan algunas medidas estadísticas propias de cada una de las variables. La información allí expresada permite realizar un primer acercamiento hacia los rangos de operación de estas señales. En particular se destaca la presencia de outliers en los valores extremos superiores de la temperatura del rodamiento 1 del generador.

Este mismo procesamiento previo de datos fue aplicado para el aerogenerador A_0, el que resulta en 88939 datos filtrados.

Se visualizan finalmente las series temporales referidas a las variables seleccionadas de interés, luego de aplicado el filtro, en la Figura 2.4.

En la Figura 2.5 se presentan los diagramas de dispersión para todas las

Figura 2.2: Curva de potencia asociada al aerogenerador A para los datos filtrados.



variables pre-seleccionadas. La información contenida en esta figura es una herramienta importante que será empleada en los capítulos siguientes para la selección de variables de interés de los modelos desarrollados.

Figura 2.3: Boxplots de las variables del aerogenerador A pre-seleccionadas.

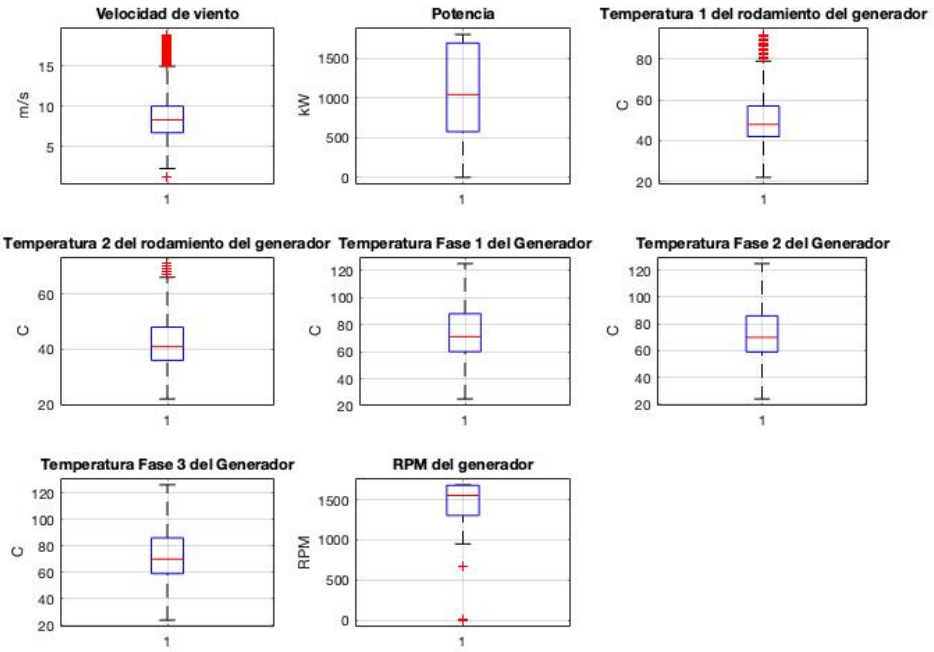


Figura 2.4: Series temporales asociadas a las variables de interés seleccionadas del aerogenerador A luego de aplicado el filtro.

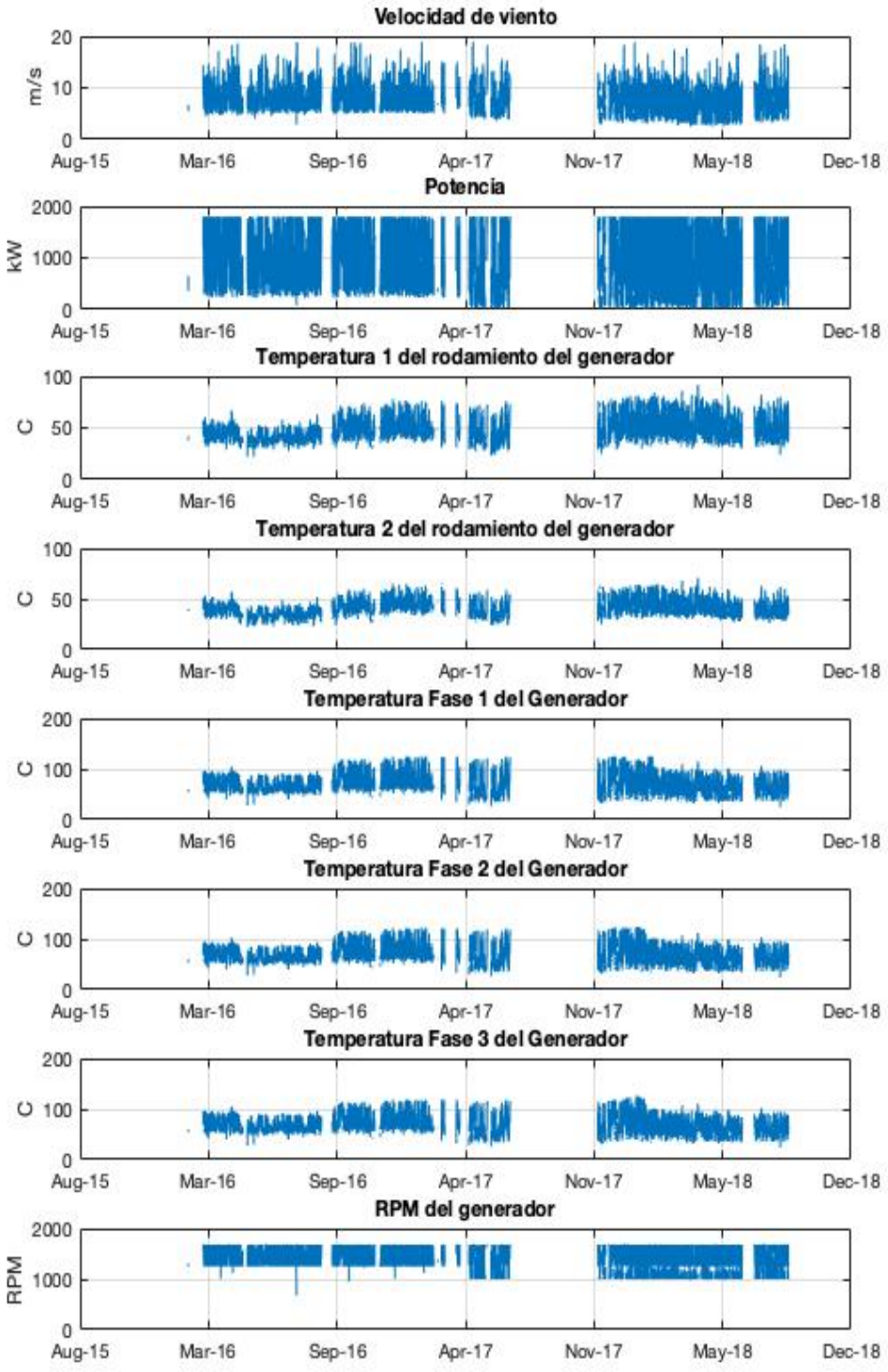
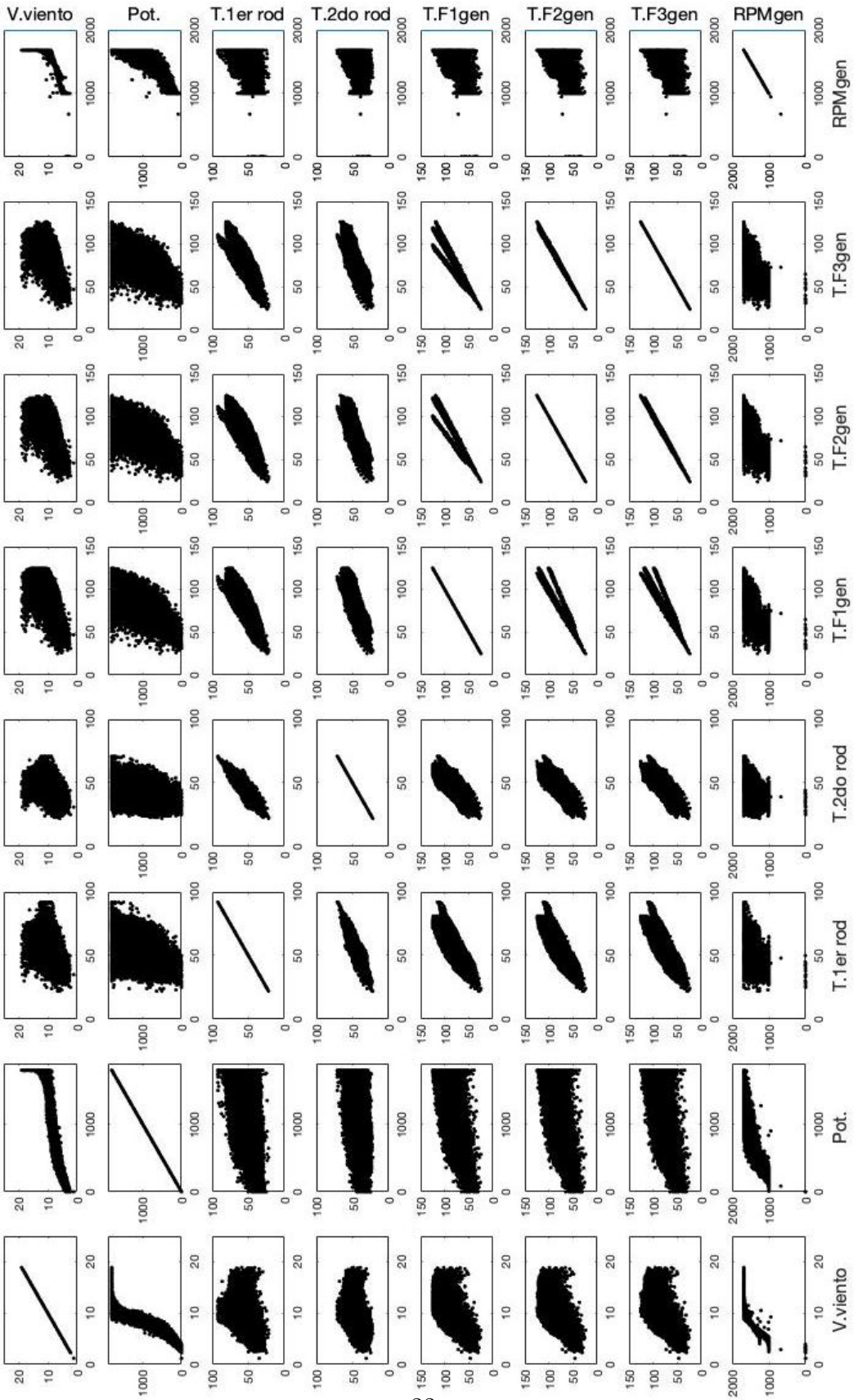


Figura 2.5: Diagrama de dispersión para las variables del aerogenerador A.



2.2. Aerogenerador B

Los datos disponibles del aerogenerador B contienen 19 señales relevadas en un régimen diezminutal. Entre estas se encuentran mediciones de diversos parámetros de la turbina, así como también otros indicadores asociados a alarmas detectadas, limitaciones externas de la potencia, porcentaje de operatividad del equipo en el diezminutal, entre otros. El aerogenerador B es un equipo de 2000kW de potencia nominal, instalado en un parque eólico de Uruguay.

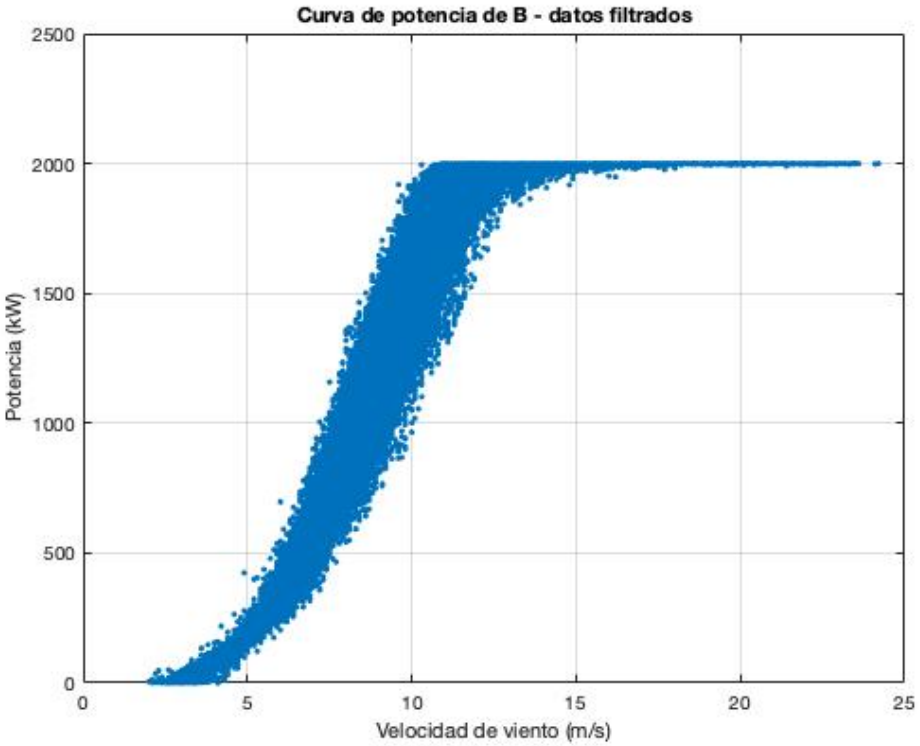
Los datos relevados están comprendidos en el período que va desde abril de 2011 hasta noviembre de 2018. Asimismo, a través del ente proveedor de los datos, se pudo conocer que durante la producción de este equipo se produjo un fallo en abril de 2014 asociado a la caja multiplicadora (CM) del sistema.

A partir de esta información, resulta de interés enfocarse, en particular, en mediciones disponibles que estén relacionadas a este acontecimiento. En este sentido, las variables a analizar son: Temperatura del aceite hidráulico, Temperatura del rodamiento A de la CM, Temperatura del rodamiento B de la CM, Temperatura del rodamiento C de la CM, Temperatura de aceite en la CM, RPM del rotor, Velocidad de viento y Potencia generada. Esta preselección está basada fundamentalmente en considerar que estas variables son las que, a priori, pueden estar vinculadas de una forma más directa con la falla ocurrida en B.

Como en el caso del aerogenerador A, existen en el período antes mencionado datos faltantes. En los 401328 diezminutales que componen el período sensado, hay 11765 datos faltantes; menos del 3 % del total. Asimismo, se destaca nuevamente que, en las secciones subsiguientes, no necesariamente será considerado todo el período disponible para el análisis.

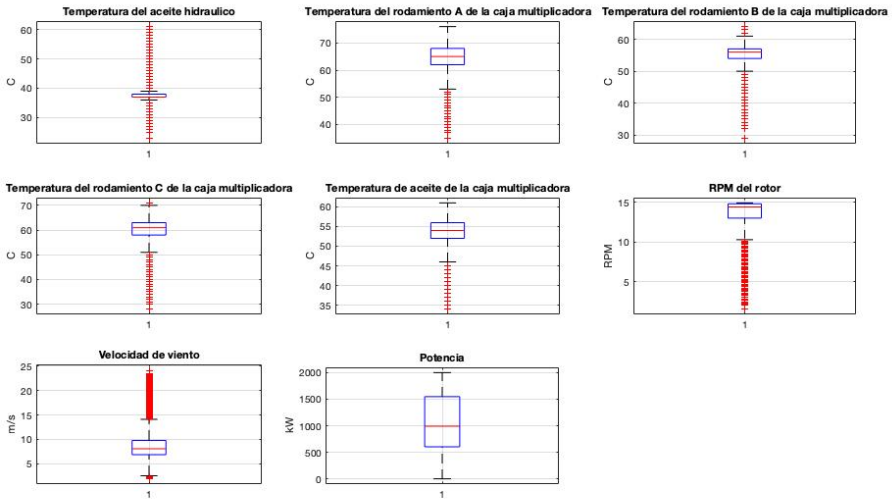
De la misma forma que se expuso en la sección precedente para el aerogenerador A, los datos precisan ser filtrados en una primera instancia con el fin de remover los puntos correspondientes a una operación anómala a priori. La metodología de este filtrado es la misma que la seguida en la sección 2.1, obteniendo una curva de potencia pos-filtrado que se presenta en la Figura 2.6. La cantidad de datos con registros, luego de realizado el filtrado, es de 211784, teniendo que la cantidad de datos depurados en el filtro es de 177779.

Figura 2.6: Curva de potencia asociada al aerogenerador B para los datos filtrados.



En la Figura 2.7 se presentan los boxplots correspondientes a las variables pre-seleccionadas. Esta información permite un primer acercamiento a los parámetros estadísticos más usuales de las señales. Se destaca en particular la

Figura 2.7: Boxplots de las variables del aerogenerador B pre-seleccionadas.



presencia de outliers, en las temperaturas y las RPM del rotor, asociados a los extremos inferiores de estas variables.

En la Figura 2.8 se presentan las series temporales referidas a las variables pre-seleccionadas de interés, luego de ser filtradas.

Finalmente, es de interés conocer cómo se inter-relacionan estas variables. En la Figura 2.9 se presentan los diagramas de dispersión para todas las variables pre-seleccionadas. Estos diagramas serán el punto de partida para la selección de las variables en los modelos desarrollados en los siguientes capítulos, por lo que son de real importancia para lo que sigue.

Figura 2.8: Series temporales asociadas a las variables de interés seleccionadas del aerogenerador B, luego de ser filtradas.

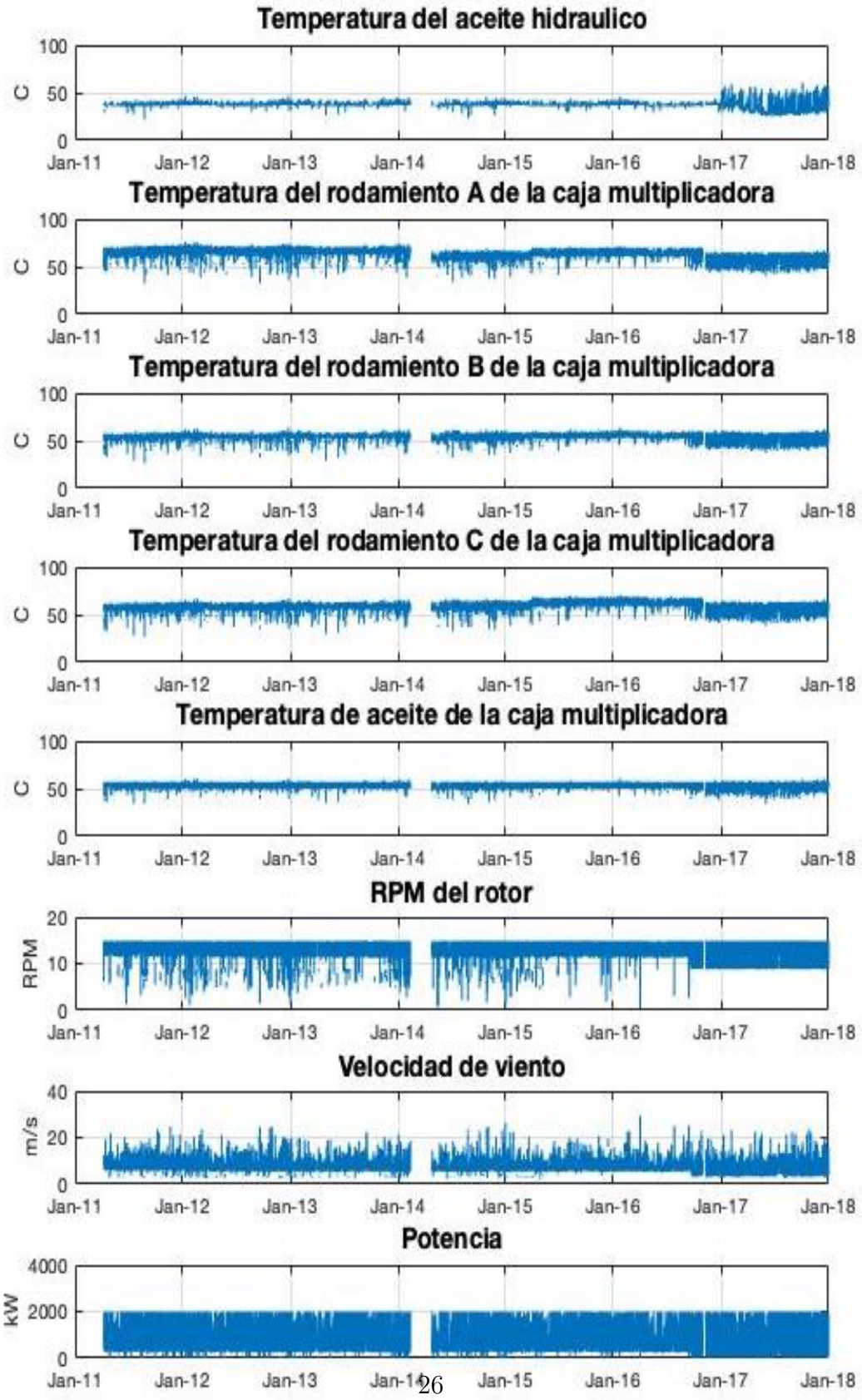
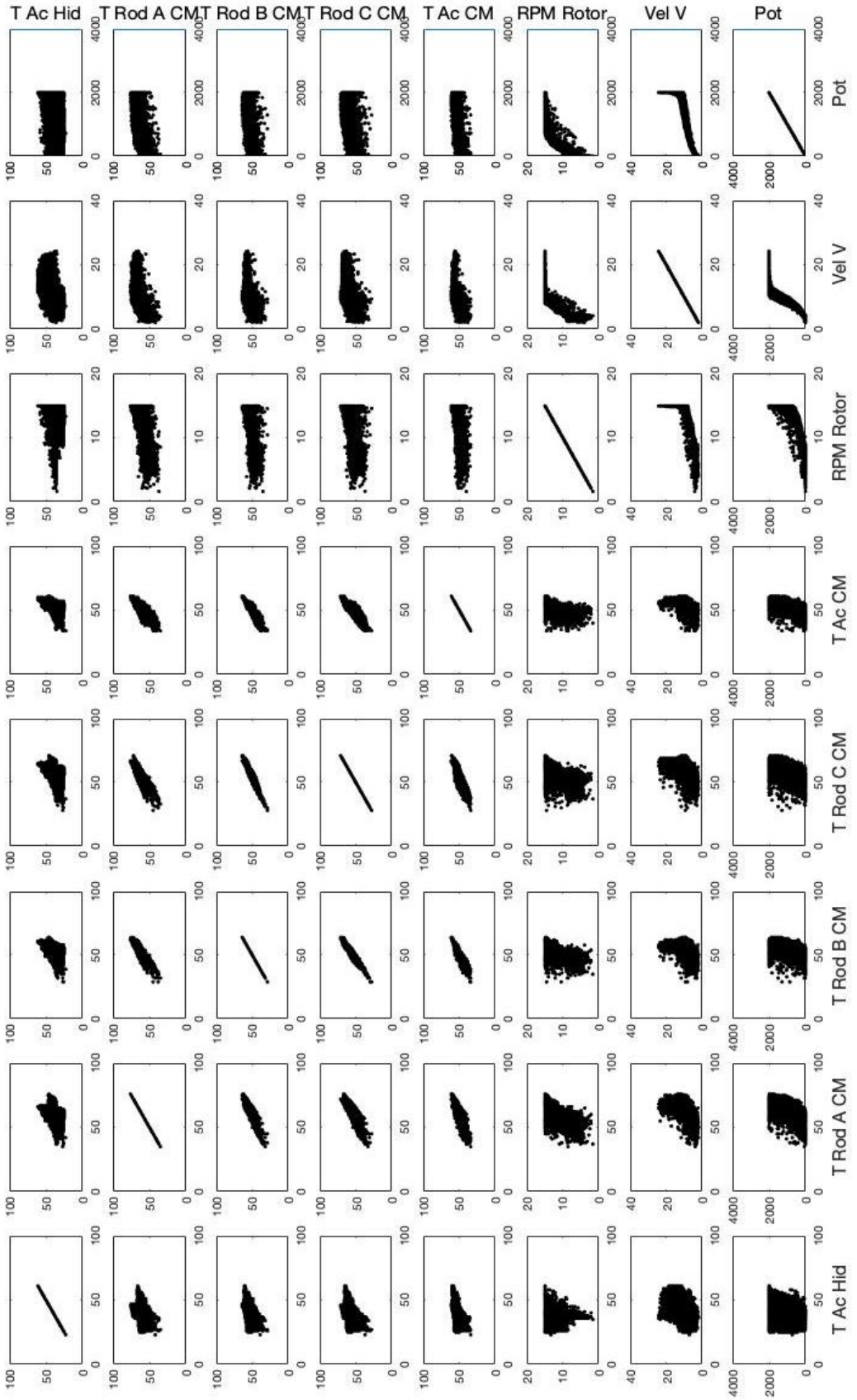


Figura 2.9: Diagrama de dispersión para las variables del aerogenerador B.



2.3. Aerogenerador C

La información recopilada del aerogenerador C consta de 23 señales obtenidas en un régimen diezminutal. Además de diversos parámetros asociados directamente al funcionamiento del equipo, se encuentran los indicadores ya mencionados como lo son las alarmas detectadas, la limitación externa de la potencia y el porcentaje de operatividad del equipo. El aerogenerador C es un equipo de 1800kW de potencia nominal, instalado en un parque eólico ubicado en Uruguay.

Los datos están comprendidos en el período que va desde diciembre de 2012 hasta julio de 2015. Durante este funcionamiento, el aerogenerador registró una falla en marzo de 2015 de la que no se tiene información certera, aunque se sospecha que pudo estar asociada a una de sus palas. En los capítulos que siguen se tomará esta información como hipótesis de trabajo.

Lo anterior hace que resulte de particular interés direccionar el estudio subsiguiente a las variables registradas que están vinculadas directa o indirectamente con la falla. En este sentido, las variables pre-seleccionadas para analizar son: Potencia, Velocidad de viento, RPM del rotor y cargas en las palas A, B y C.

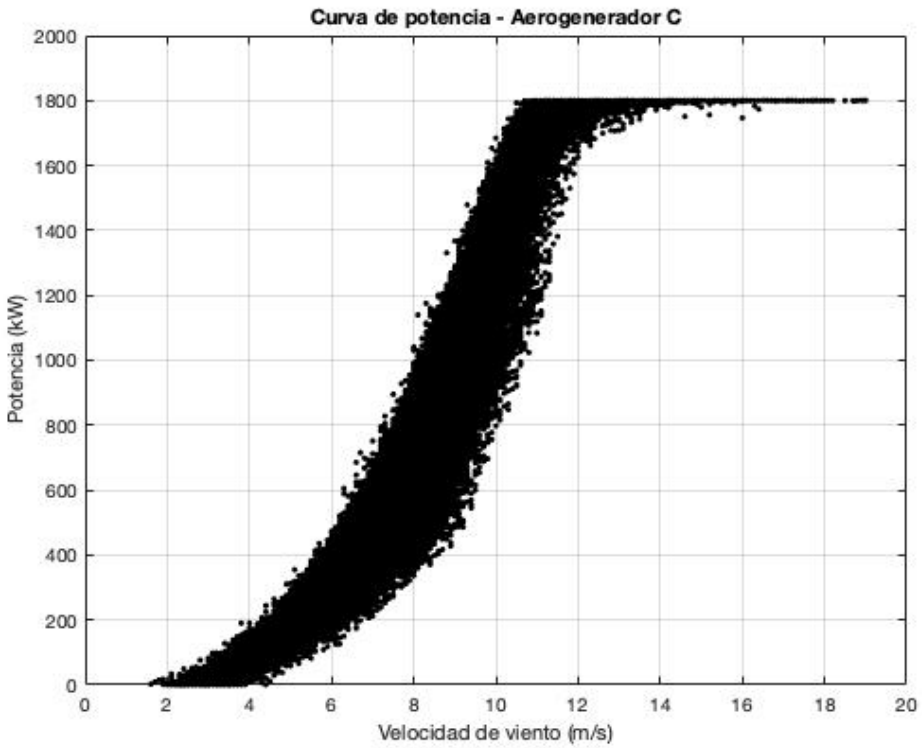
Durante el período de registro de datos, existen algunos diezminutales faltantes, consecuencia de paradas de mantenimiento preventivo programadas, de interrupciones imprevistas en la producción o, de errores en el propio sistema SCADA. En los 133318 diezminutales que componen el período sensado, hay 13799 datos faltantes.

Al igual que como se aplicó en los casos anteriores, los datos fueron filtrados en una primera instancia con el fin de remover los puntos correspondientes a una operación anómala a priori. La metodología de este filtro fue desarrollada en la sección 2.1, obteniendo la curva de potencia pos-filtrado que se presenta en la Figura 2.10.

La cantidad de datos con registros, luego de realizado el filtrado, es de 90692, teniendo que la cantidad de datos depurados por el filtro es de 28827.

En la Figura 2.11 se presentan los boxplots correspondientes a las variables pre-seleccionadas. Allí se presentan algunos de los parámetros estadísticos más usuales para estas señales.

Figura 2.10: Curva de potencia asociada al aerogenerador C para los datos filtrados.



En la Figura 2.12 se presentan las series temporales asociadas a las variables pre-seleccionadas para el análisis, luego de ser filtradas.

Finalmente, en la Figura 2.13 se presentan los diagramas de dispersión para todas las variables involucradas en la pre-selección. Esta información permite conocer cómo se inter-relacionan estas variables, información que será de interés fundamental para la selección de variables en los próximos capítulos.

Figura 2.11: Boxplots de las variables del aerogenerador C pre-seleccionadas.

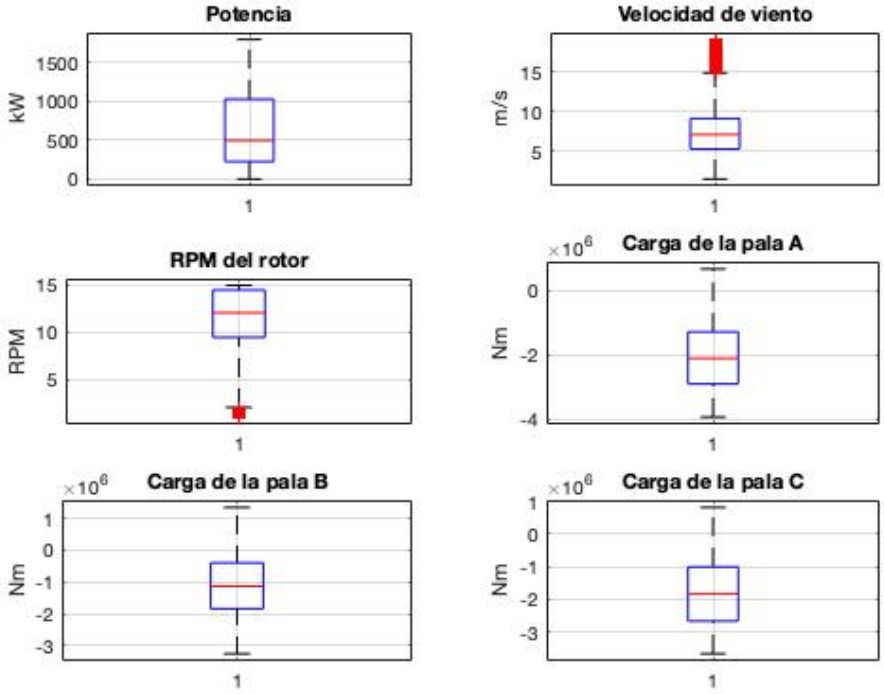


Figura 2.12: Series temporales asociadas a las variables de interés seleccionadas del aerogenerador C, luego de ser filtradas.

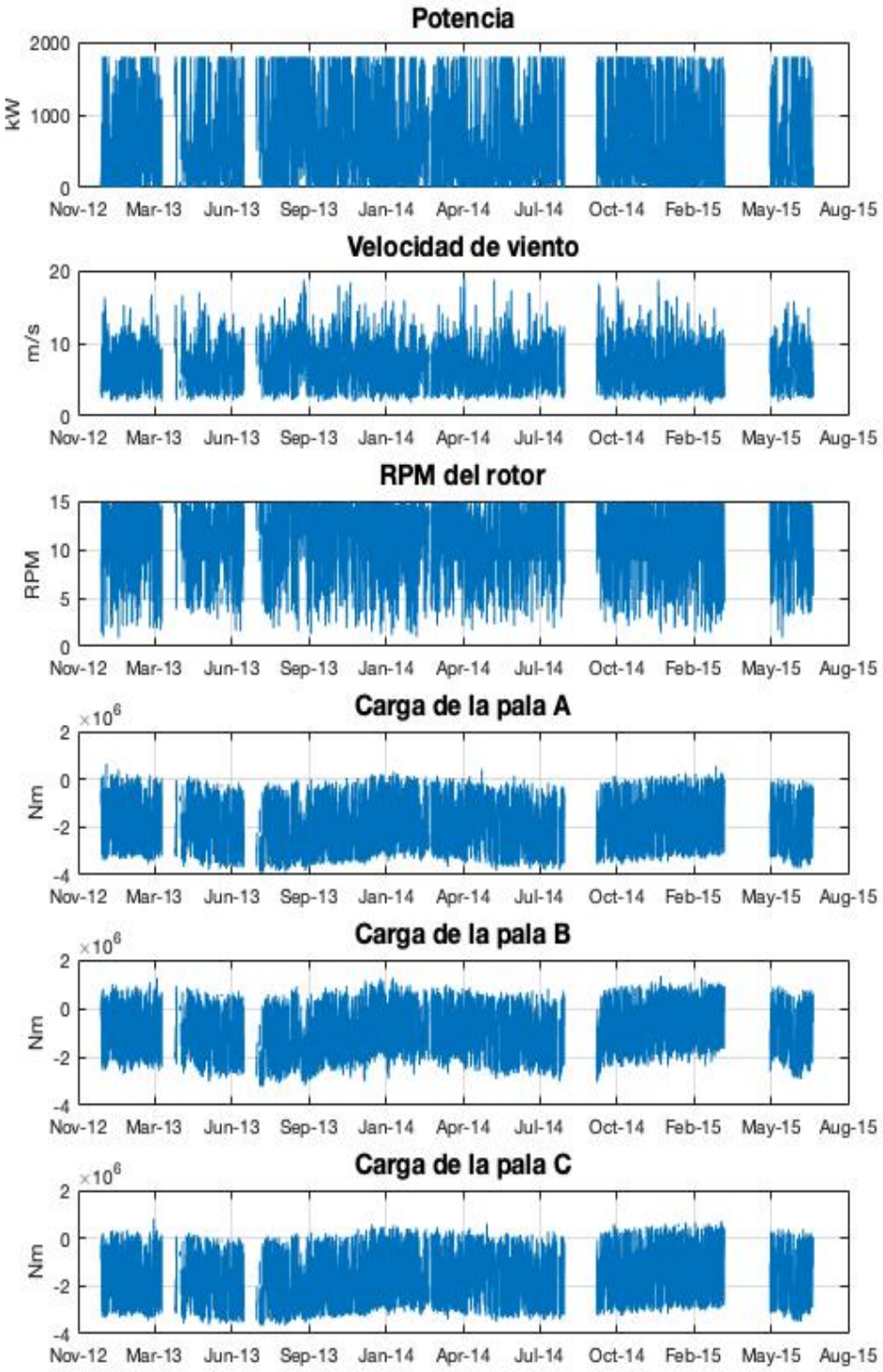
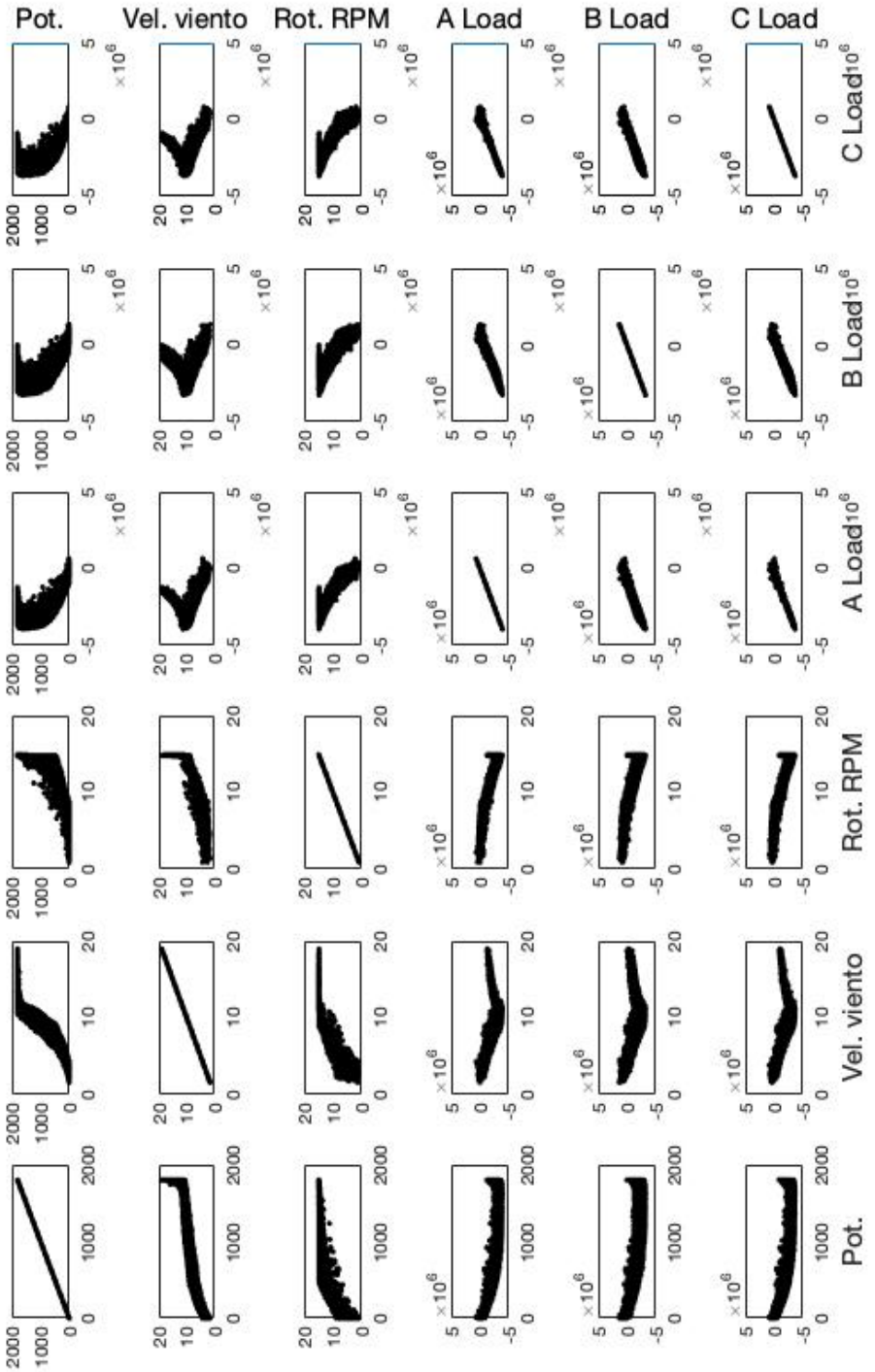


Figura 2.13: Diagrama de dispersión para las variables del aerogenerador C.



2.4. Aerogenerador D

Los datos provenientes del aerogenerador D constan de 7 señales relevadas en régimen diezminutal: potencia, velocidad de viento, dirección de viento, temperatura ambiente, yaw, RPM del rotor y pitch, además de algunos parámetros auxiliares referidos a la condición de operación de la turbina.

El aerogenerador D tiene una potencia nominal de 2000kW y, a diferencia de los aerogeneradores presentados en secciones precedentes, no se tiene información sobre una falla asociada a su funcionamiento. Sin embargo, se conoce que, entre noviembre de 2013 y agosto de 2014, el equipo tuvo una parada de mantenimiento, aunque se desconocen los motivos de la misma. Por lo tanto, el objetivo de los análisis referidos a este equipo será comparar el funcionamiento del aerogenerador entre antes y después de la parada.

Al desconocer si realmente hubo una falla que haya provocado la parada de mantenimiento, no es posible realizar una pre-selección de variables asociadas a la naturaleza del problema. A raíz de esto, la descripción de datos abarcará las 7 señales disponibles.

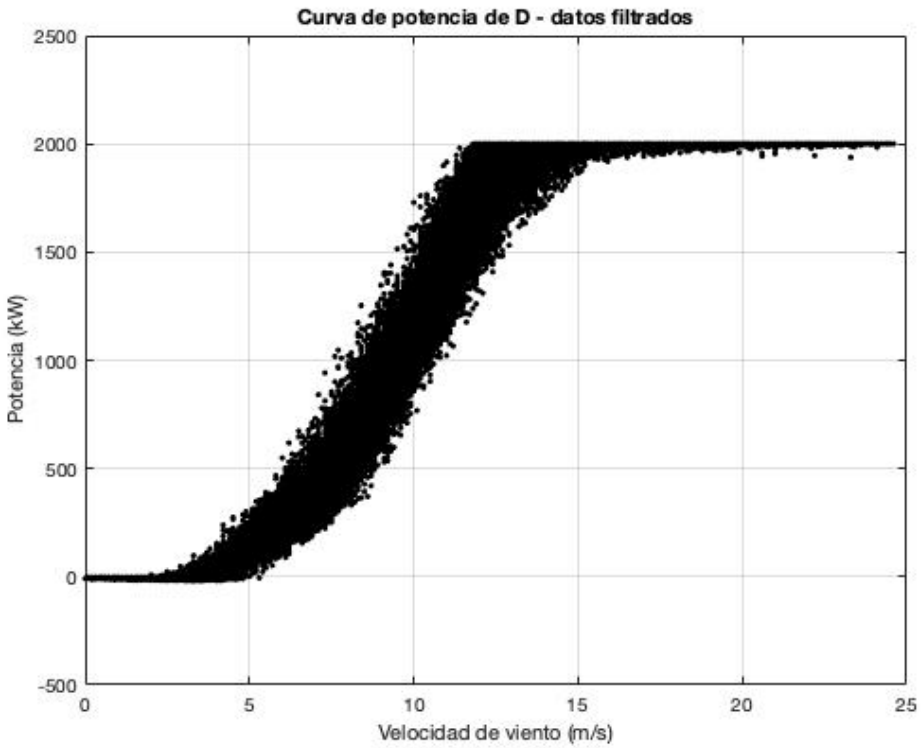
Los datos del aerogenerador D están comprendidos entre enero de 2010 y agosto de 2015. En los 297886 diezminutales que componen el período relevado, hay 3196 datos faltantes; poco más del 1 % del total. Cabe mencionar que, durante la parada de mantenimiento, el sistema SCADA continuó registrando información, aunque valores nulos en todos los casos; por lo que, este período no está considerado como información faltante.

En una primera instancia, los datos precisan ser filtrados con el fin de remover los puntos correspondientes a una operación anómala a priori. En este sentido, en la Figura 2.14 se presenta la curva de potencia pos-filtrado obtenida a partir del filtro aplicado, cuya metodología fue desarrollada en la sección 2.1.

La cantidad de datos con registros, luego de realizado el filtro, es de 244772, teniendo que la cantidad de datos depurados por el filtro es de 49918.

En la Figura 2.15 se presentan los boxplots correspondientes a las variables registradas para el aerogenerador D. Esta información permite hacer un acercamiento primario a las señales relevadas. En particular, se destacan algunos outliers asociados a altos valores del pitch, como también para los bajos valores de las RPM del rotor.

Figura 2.14: Curva de potencia asociada al aerogenerador D para los datos filtrados.



En este sentido, interesa a su vez comparar el comportamiento de las variables entre antes de la parada de noviembre de 2013 y después. En la Figura 2.16 se presentan los boxplots asociados a las variables, comparando el funcionamiento antes y después de la parada. Se destaca que, luego de la parada, la mayor proporción de los valores del pitch se concentran en un rango de operación más acotado; sin embargo, aún así se registran nuevamente numerosos outliers con elevados valores de esta variable. Cabe destacar de todas formas que la información presentada en la Figura 2.16 no evidencia una mejora en el desempeño del equipo luego de la parada.

En la Figura 2.17 se presentan las series temporales asociadas a las variables del aerogenerador D luego de ser filtradas.

Finalmente, en la Figura 2.18 se presentan los diagramas de dispersión para las siete señales involucradas. Esta información permite conocer cómo se inter-relacionan las variables. Se destaca en particular que, algunos pares de variables parecen ser independientes entre sí.

Figura 2.15: Boxplots de las variables del aerogenerador D.

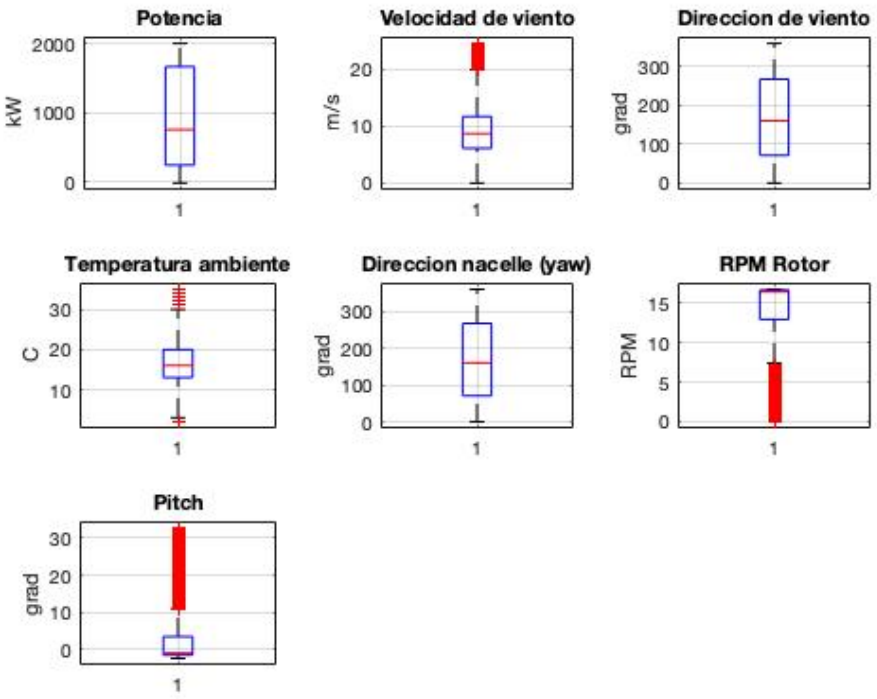


Figura 2.16: Boxplots de las variables del aerogenerador D comparando el comportamiento antes y después de la parada de noviembre de 2013.

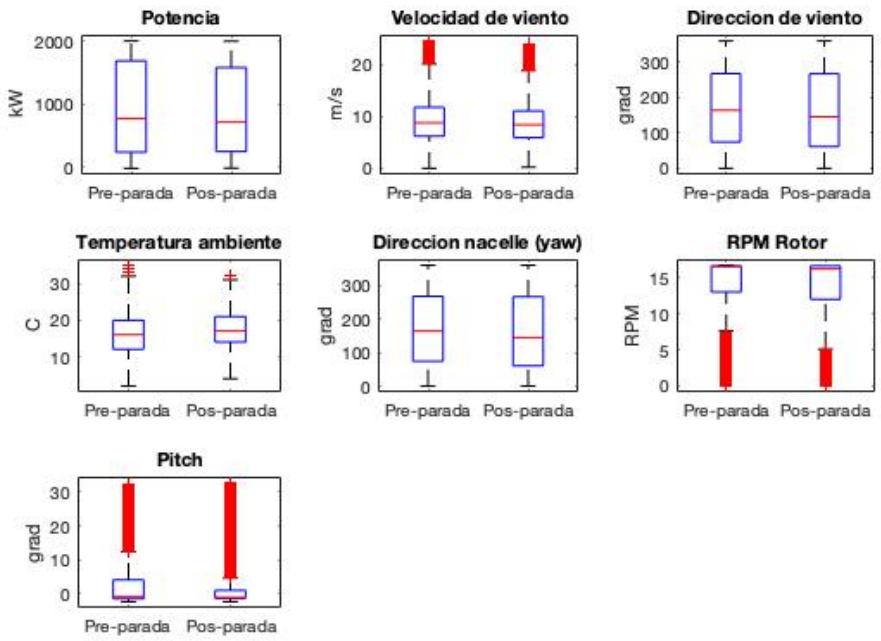


Figura 2.17: Series temporales asociadas a las variables del aerogenerador D, luego de ser filtradas.

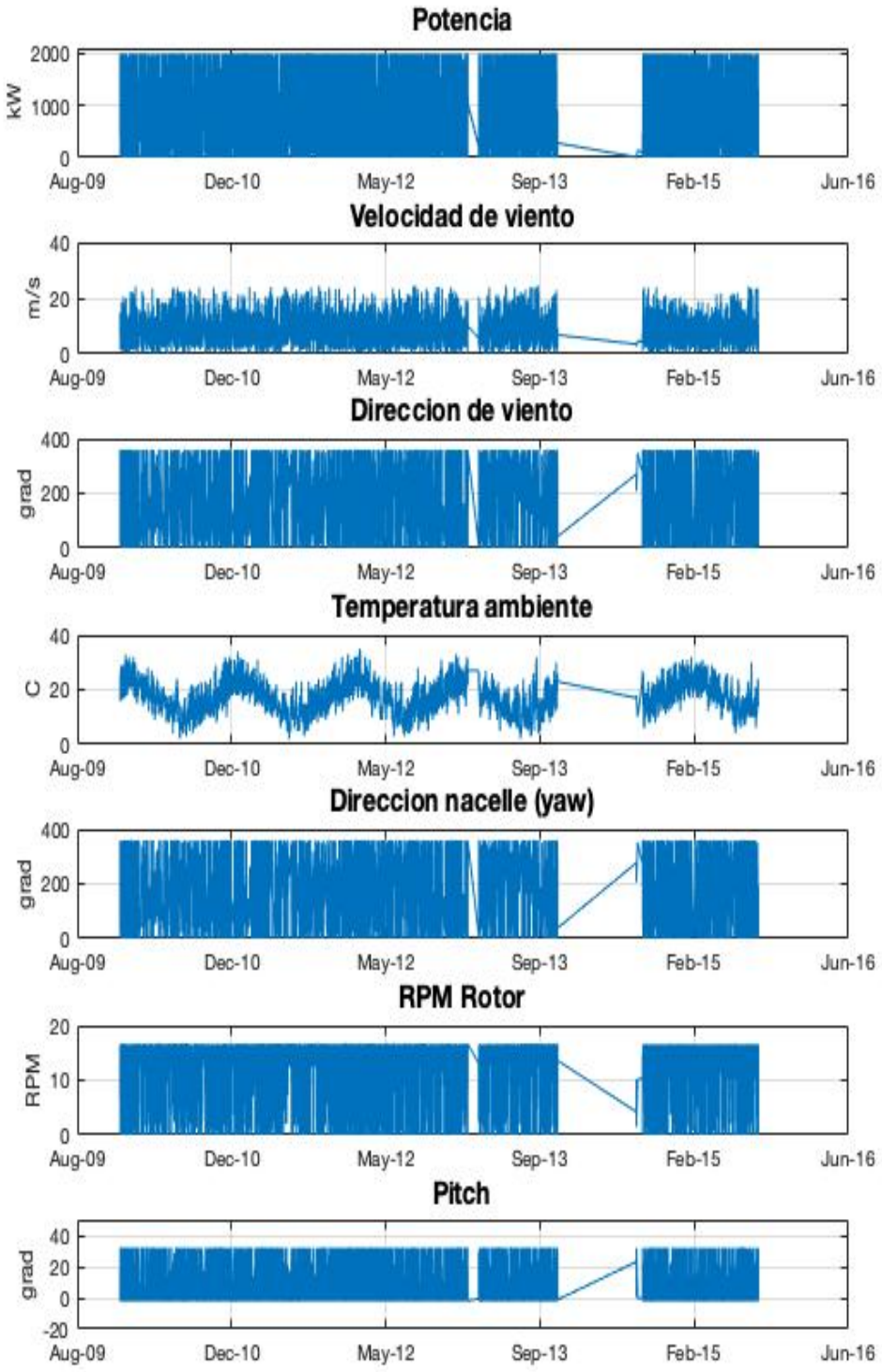
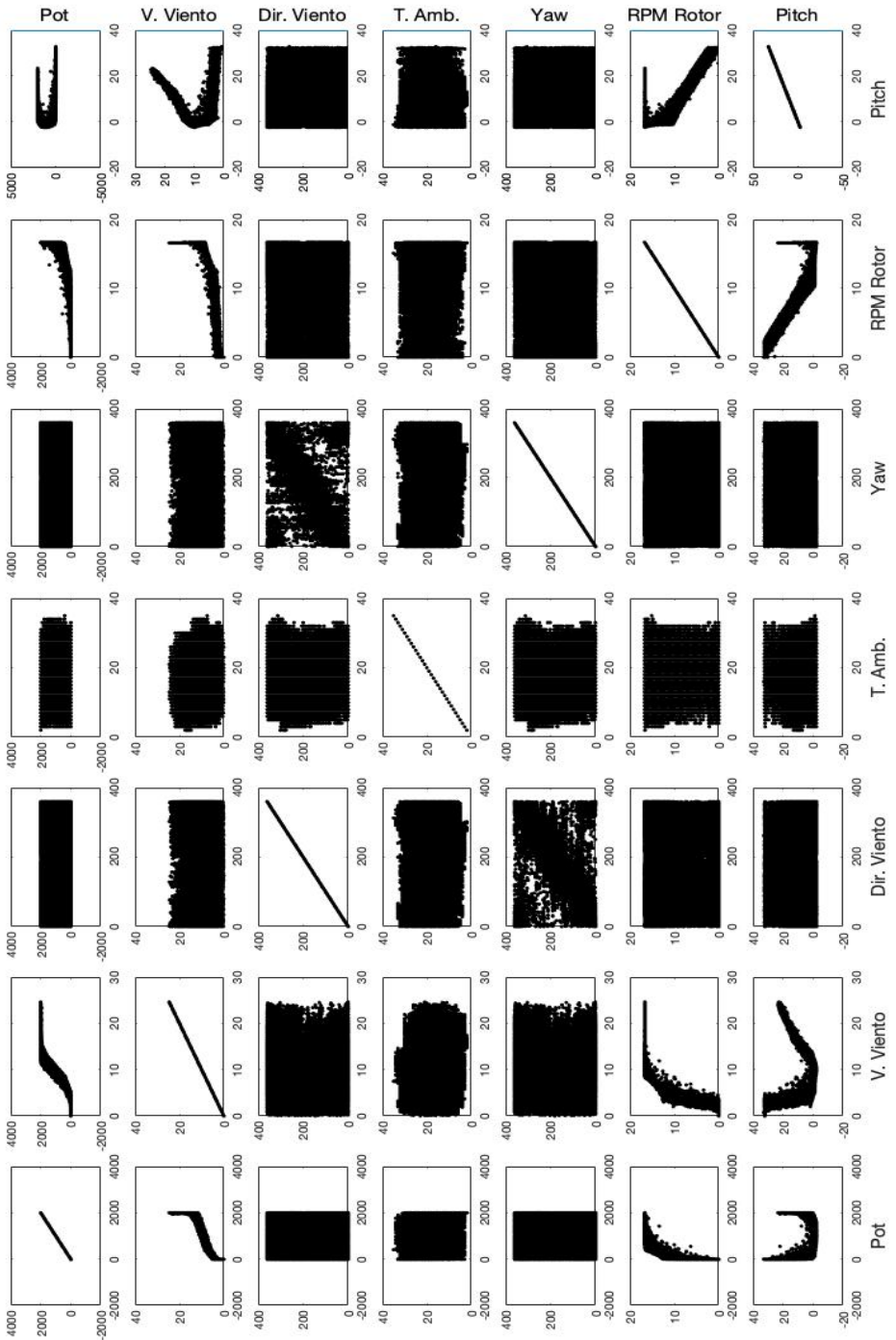


Figura 2.18: Diagrama de dispersión para las variables del aerogenerador D.



Capítulo 3

Proceso Gaussiano (GP): modelo de regresión

En este capítulo se estudia un modelo de regresión basado en un Proceso Gaussiano que tiene como finalidad evaluar la condición de funcionamiento de un aerogenerador, y poder así detectar anomalías que a su vez permitan predecir fallas. Esta herramienta es empleada para modelar, a partir de un conjunto de variables predictoras, una determinada variable del SCADA. Se aborda en este capítulo el desarrollo teórico del mismo y su aplicación a tres de los cuatro aerogeneradores presentados en la sección precedente; por motivos que se expondrán en la sección 7, la aplicación del método al caso del aerogenerador C no está contenida en este capítulo.

3.1. Descripción teórica

En un problema típico de regresión, se tiene que dadas algunas observaciones ruidosas de una variable dependiente y en ciertos valores de la variable independiente x, escalar o vectorial, se busca hallar la mejor estimación de la variable dependiente en un nuevo valor de x. Haciendo alguna suposición sobre la relación de estas variables, $f(x)$, como por ejemplo asumir una forma lineal, se podría aplicar el método de mínimos cuadrados para resolver el problema. La regresión mediante un Proceso Gaussiano es una aproximación más fina, en el sentido de no exigir asociar a $f(x)$ a ningún modelo en particular. Esta herramienta permite representar a $f(x)$ de modo que "los datos hablen"más por sí mismos.

Un Proceso Gaussiano es una generalización no paramétrica de la distribución normal conjunta para un conjunto de variables dado (Rasmussen and Williams (2006)). Matemáticamente se define por sus funciones de media y covarianza, como se expresa en la Ecuación 3.1,

$$Y \sim GP(\mu, \Sigma) \tag{3.1}$$

donde μ es la función media y Σ es la función de covarianza que tiene asociada una función de densidad de probabilidad. Un modelo de Proceso Gaussiano genera datos ubicados en algún dominio de manera que cualquier subconjunto finito del rango siga una distribución gaussiana multivariada. Lo que relaciona una observación a otra es justamente la función de covarianza Σ, que se representa a través de una matriz K, cuyos elementos son $k(x, x')$. En lo que sigue se trabaja con la elección de la función exponencial cuadrática, dada en la Ecuación 3.2.

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{d^2(x, x')}{2l^2}\right] \tag{3.2}$$

Donde $d(x, x')$ representa la distancia entre x y x', σ_f es la varianza de la señal y l es una escala de los datos. Esta función de covarianza tiene parámetros libres σ_f y l, que deben ser ajustados convenientemente para la obtención de un modelo adecuado. Como en general los datos contienen cierto ruido, cada observación está relacionada con la función $f(x)$ de acuerdo a la Ecuación 3.3,

$$y = f(x) + \mathcal{N}(0, \sigma_n^2) \tag{3.3}$$

donde σ_n es una parámetro que también debe ser ajustado. Por lo tanto, para simplificar lo que sigue, se reescribe la función de covarianza de acuerdo a la Ecuación 3.4,

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{d^2(x, x')}{2l^2}\right] + \sigma_n^2 \delta(x, x') \tag{3.4}$$

donde $\delta(x, x')$ es la función Kronecker Delta[1]. Entonces, dado un conjunto de observaciones $Y = \{y_1, \ldots, y_n\}$, el objetivo es predecir el valor de la observación y^*, para un nuevo valor de x^*.

Una vez construido el modelo a partir de la función de covarianza, es decir, a partir de la matriz $(K)_{ij} = k(x_i, x_j)$, el objetivo es determinar el valor de

[1] $\delta(x, x') = 1$ si $x = x'$; $\delta(x, x') = 0$ en otro caso.

y^*. Dado que la suposición clave en este modelo es que los datos se pueden representar como una muestra de una distribución gaussiana multivariada, se tiene lo expresado en la Ecuación 3.5.

$$\begin{bmatrix} Y \\ y^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \tag{3.5}$$

Donde $K_* = [k(x^*, x_1), \ldots, k(x^*, x_n)]$ y $K_{**} = k(x^*, x^*)$. En este contexto se está asumiendo que el conjunto Y tiene media nula.

El interés está entonces en determinar la mejor predicción para y^* dados los datos, es decir, la probabilidad condicional $p(y^*|Y)$. El resultado que se tiene es el presentado en la Ecuación 3.6 (Rasmussen and Williams (2006)).

$$y^*|Y \sim \mathcal{N}\left(K_* K^{-1} Y, K_{**} - K_* K^{-1} K_*^T\right) \tag{3.6}$$

Por lo que la mejor estimación para y^* es la media de la distribución.

$$\bar{y^*} = K_* K^{-1} Y \tag{3.7}$$

A su vez, la incertidumbre en la estimación está dada por su varianza de acuerdo a la Ecuación

$$var(y^*) = K_{**} - K_* K^{-1} K_*^T \tag{3.8}$$

El análisis precedente depende del conjunto de parámetros $\theta = \{l.\sigma_f, \sigma_n\}$. La elección de estos parámetros debe ser realizada adecuadamente para que el resultado del modelo sea satisfactorio. El óptimo a posteriori de θ ocurre cuando $p(\theta|X, Y)$ es máxima. El teorema de Bayes asegura que este problema corresponde a maximizar $\log p(Y|X, \theta)$, dado por la siguiente expresión (Ebden (2015)).

$$\log p(Y|X, \theta) = -\frac{1}{2} Y^T K^{-1} Y - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi \tag{3.9}$$

Mediante un algoritmo de optimización multivariado es posible resolver el problema con el fin de obtener los parámetros del conjunto θ adecuados para la construcción del modelo. Rasmussen and Williams (2006) desarrollan algunas herramientas adicionales para la resolución de este problema de optimización.

3.1.1. Metodología de aplicación

El método estudiado consta de tres etapas globales: *entrenamiento*, *validación* y *testeo*. En la primera, a partir de un conjunto de datos "saludables", se construye el modelo. En la segunda etapa, a partir de otro conjunto de datos de iguales características, se valida el modelo; se verifica que el modelo describe satisfactoriamente los datos. Finalmente, el modelo es empleado para evaluar la condición de funcionamiento del aerogenerador en una última etapa con un tercer conjunto de datos. Este último conjunto eventualmente puede estar asociado a desperfectos de funcionamiento en la turbina. Cabe resaltar que los tres conjuntos de datos son disjuntos entre sí.

En primera instancia, se selecciona un conjunto de datos saludables de *entrenamiento* para construir el modelo. Este está compuesto por n_e datos, d variables predictoras y una determinada variable a modelar, que llamamos y. A su vez, llamamos X_e a la matriz compuesta por los datos de entrenamiento; tenemos entonces que X_e es una matriz de tamaño $n_e \times (d+1)$.

En el caso de estudio, el modelo es construido a partir de X_e mediante la herramienta `fitrgp` de Matlab, que dados la función de covarianza definida por la Ecuación 3.4, el método de resolución del problema de optimización establecido en la Ecuación 3.9, y la inicialización de θ, proporciona la solución para θ junto con el modelo de Proceso Gaussiano.

Este modelo obtenido es validado en segunda instancia con el conjunto de datos de *validación*, X_v. X_v es una matriz de tamaño $n_v \times (d+1)$, donde n_v es la cantidad de observaciones contenidas en estos datos. Para los n_v datos de las d variables predictoras, se obtienen las estimaciones proporcionadas por el modelo, $\bar{y_v}$. $\bar{y_v}$ es una matriz de tamaño $n_v \times 1$. Para la validación se determina la recta de regresión entre las observaciones de la variable modelada y $\bar{y_v}$. Esta validación se considera satisfactoria en el caso de que esa recta de regresión se aproxime a la recta identidad. En caso de que se perciba un sesgo entre el modelo y las observaciones $\bar{y_v}$, se aplica una re-calibración del modelo, cuyos detalles serán expuestos en el caso de estudio concreto que corresponda.

Finalmente, una vez validado el modelo, se avanza a la etapa de *testeo*. Para esta instancia se cuenta con n_t datos de testeo comprendidos en una matriz X_t de tamaño $n_t \times (d+1)$, de las cuales se utilizan las primeras d columnas para obtener las estimaciones mediante el modelo ya construido. A partir del modelo se obtiene la variable estimada $\bar{y_t}$, que se compara con la columna

$(d+1)$-ésima de X_t, y_t, con el fin de evaluar la condición de funcionamiento del aerogenerador. Se define así el residuo de testeo como

$$R_t = y_t - \bar{y}_t \tag{3.10}$$

Con el fin de remover el ruido en forma de picos originados por la presencia de datos faltantes en el registro de X_t, se aplica a R_t un filtro, obteniendo así la señal filtrada R_t^*.

La evolución de R_t^* representa un indicador de la condición de funcionamiento del aerogenerador, en el sentido de que R_t^* representa la desviación de la medición original de la variable modelada respecto al modelo; por esta razón, el residuo del modelo es una herramienta que permite evaluar lo deseado.

En algunos casos, no parece evidente distinguir períodos donde R_t^* tenga un comportamiento inusual. En ese sentido, Wang and Infield (2013) proponen un algoritmo de detección de anomalías aplicado a R_t^*. Un enfoque similar, aunque con algunas modificaciones considerables, se propone en este trabajo, donde el eje temporal en el período de validación es subdividido en ventanas móviles de largo constante LV, y en cada una de ellas se consideran los valores correspondientes de R_v^*, definido de forma análoga a R_t^*. Dos ventanas consecutivas tendrán $LV - 1$ diezminutales en común; el primero de una ventana no estará en presente en la siguiente. En este contexto, LV es considerado una escala temporal dentro del análisis. Si bien todas las ventanas tienen el mismo largo temporal, eventualmente puede ocurrir que la cantidad de datos dentro de una ventana sea inferior a LV, ya sea por ausencia de mediciones o por paradas operativas en ese período. Del mismo modo, R_t^* también es subdividido en ventanas temporales de largo LV, valiendo las mismas observaciones anteriores. El propósito de esto es utilizar un método de Monte Carlo para comparar estadísticamente y en forma reiterada un gran número de veces, de forma cronológica, las ventanas de R_t^* con ventanas aleatorias de R_v^*, y así poder distinguir un cambio en el funcionamiento del aerogenerador.

Se elige este camino (a diferencia de Wang and Infield (2013)) ya que se entiende que, a priori, una ventana particular (o calculada a partir de ventanas particulares) del período de validación no sea preferible a otras, como patrón de comparación. El método de Monte Carlo así implementado tiene la virtud de tomar en cuenta a toda la población de ventanas del período de validación.

Sea $v_v(i)$ la ventana i-ésima de R_v^*, y sea $v_t(j)$ la ventana j-ésima de R_t^*.

Con el objetivo de aplicar el test de diferencia de medias (Sawilowski (2002)), se definen M, s y n como la media, la desviación estándar y la cantidad de datos de la ventana respectiva. Si bien este test es robusto frente a la falta de gaussianidad de los datos (en especial si la cantidad de datos es grande), no lo es en el caso de que la hipótesis de independencia de los propios datos no sea cubierta (von Storch and Zwiers (1999)). En este caso, por tratarse de ventanas temporales asociadas a fenómenos físicos dependientes de variables meteorológicas, en general los datos de las ventanas son correlacionados entre sí. Se propone una corrección para n (ver por ejemplo von Storch and Zwiers (1999)), n_{eq}, que tiene en cuenta la correlación de los datos de la serie de la ventana temporal,

$$n_{eq} = \frac{n}{1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho(k)} \tag{3.11}$$

donde $\rho(k)$ es la función de autocorrelación.

Teniendo en cuenta la consideración anterior, se considera el siguiente estadístico adimensionado, que se puede suponer que sigue una distribución *t de Student*.

$$t = \frac{M_v - M_t}{\sqrt{\frac{s_v^2}{n_{eq_v}} + \frac{s_t^2}{n_{eq_t}}}} \tag{3.12}$$

Este es empleado para evaluar los casos extremos de R_t^*, negativos o positivos, dependiendo de cuál sea el significado físico de la variable modelada. Por otra parte, la cantidad de grados de libertad a emplear está dada por:

$$\nu = \frac{\left(\frac{s_v^2}{n_{eq_v}} + \frac{s_t^2}{n_{eq_t}}\right)^2}{\frac{\left(\frac{s_v^2}{n_{eq_v}}\right)^2}{n_{eq_v}-1} + \frac{\left(\frac{s_t^2}{n_{eq_t}}\right)^2}{n_{eq_t}-1}} \tag{3.13}$$

A partir de lo anterior, y de un nivel de significancia α_W elegido, es posible comparar estadísticamente $v_v(i)$ con $v_t(j)$, y así evaluar si hay una desviación estadísticamente significativa entre una y otra. La hipótesis nula es que las medias en ambas ventanas son iguales.

Luego, se procede a repetir r veces el test, registrando la cantidad de veces que resulta en que la media de $v_t(j)$ es significativamente distinta a la media de $v_v(i)$. Por la elección aleatoria de las ventanas de validación, estos ensayos son independientes entre sí.

Finalmente, es necesario determinar cuántos registros de $v_t(j)$, en las r

repeticiones, hacen que esa ventana sea asignada como anómala; a_j. Para ello, se opta por el percentil 80 de las r repeticiones, es decir, $0.8r$: si $a_j \geq 0.8r$, la ventana $v_t(j)$ será asignada como anómala. Otros percentiles fueron analizados para todos los métodos estudiados en las secciones subsiguientes, como el 70 y 90, por ejemplo; se encontró que el percentil 80 correspondía a un balance entre la cantidad de alarmas y la temprana detección de las fallas conocidas de los aerogeneradores.

De esta forma es posible determinar si $v_t(j)$ es clasificada como anómala, para cada j. Y así, poder identificar funcionamientos anómalos en el aerogenerador.

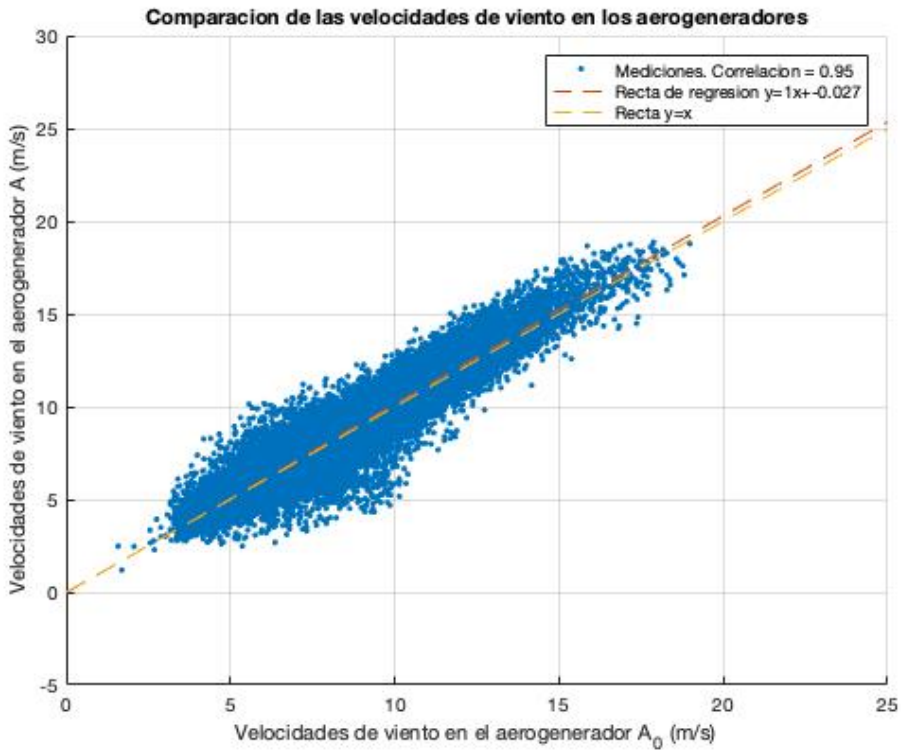
3.2. Resultados

3.2.1. Aplicación para el Aerogenerador A

La metodología desarrollada anteriormente fue aplicada para el estudio de la predicción de fallas en el caso real del aerogenerador A.

Como se mencionó en la sección 2.1, los datos disponibles comprenden a dos aerogeneradores: A y A_0. Por tratarse de dos equipos iguales e instalados en el mismo parque eólico, se optó por tomar el aerogenerador A_0 como base de datos de entrenamiento para generar el modelo de Proceso Gaussiano. Esta decisión se fundamenta principalmente en la hipótesis de que dos aerogeneradores iguales, enfrentados al mismo flujo de viento, en condiciones normales, deberían responder de forma similar; agregando además que el aerogenerador A_0 no tuvo fallas identificadas durante su producción. La selección de los datos provenientes del aerogenerador A_0 está además fundamentada en que los datos considerados saludables del aerogenerador A no son suficientes para construir el modelo, en el sentido de que no se capturarían todos los efectos estacionales ocurridos a lo largo de un año. En la Figura 3.1 se presenta el diagrama de dispersión para las velocidades de viento medidas en los aerogeneradores A y A_0, y en la Figura 3.2 se presenta el diagrama de dispersión para los cosenos de las direcciones de viento medidas en los aerogeneradores. Las mismas muestran que el ajuste es bastante bueno. En algún sentido, esto valida la hipótesis de que ambos generadores observan aproximadamente los mismos flujos de velocidades de viento. Por lo tanto, en lo que sigue se asumirá esta hipótesis de trabajo.

Figura 3.1: Diagrama de dispersión para las velocidades de viento medidas en los aerogeneradores A y A_0.

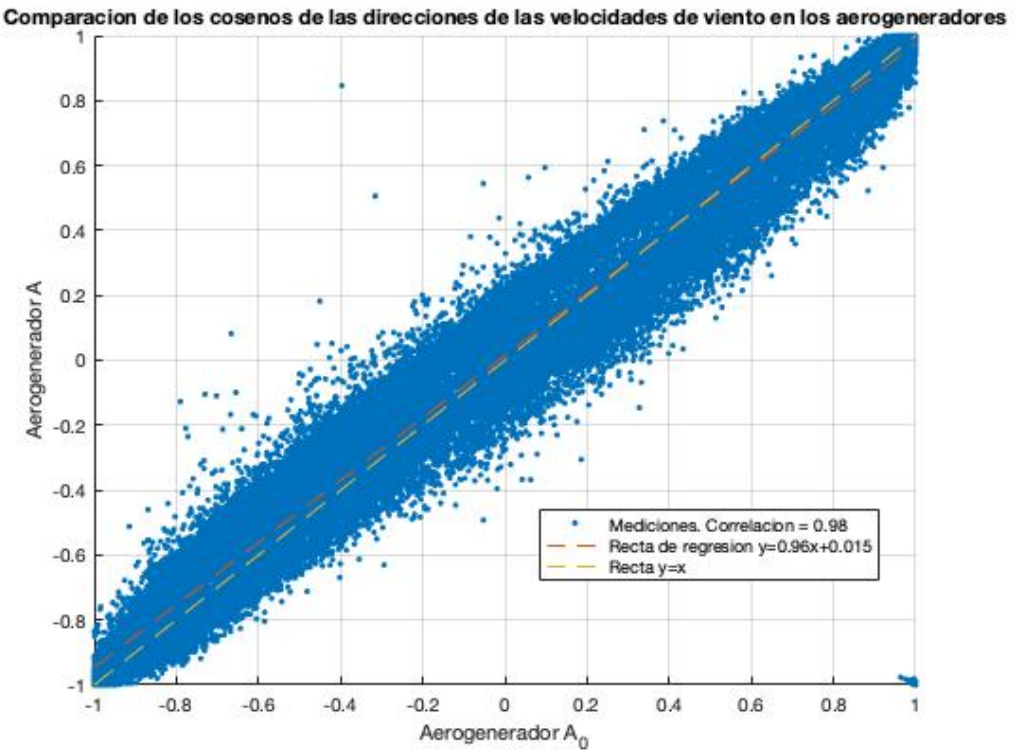


Vista la naturaleza de las fallas asociadas al aerogenerador A, indicadas en la sección 2.1, se decidió tomar la temperatura de la 3ra fase del generador como variable a modelar -variable objetivo-, dentro del conjunto de variables pre-seleccionadas en la sección 2.1. Con el objetivo de definir el conjunto de variables predictoras, se consideran en la Tabla 3.1 las correlaciones entre la variable objetivo y las demás.

Tabla 3.1: Correlaciones entre la temperatura de la fase 3 del generador y las demás variables pre-seleccionadas.

Temperatura de la 3ra fase del generador	1
Velocidad de viento	0.80
Potencia	0.82
Temperatura del rodamiento 1 del generador	0.83
Temperatura del rodamiento 2 del generador	0.82
Temperatura de la 1er fase del generador	0.99
Temperatura de la 2da fase del generador	1.00
Revoluciones del generador	0.77

Figura 3.2: Diagrama de dispersión de los cosenos de las direcciones de viento medidas en los aerogeneradores A y A_0.



A partir de la información que se ilustra en la Figura 2.5 y la que se presenta en la Tabla 3.1, vemos que las variables asociadas a las otras dos fases del generador tienen un comportamiento esencialmente idéntico al de la variable objetivo, generando información redundante. Por lo tanto, se toma la decisión de seleccionar las tres variables con mayor valor de correlación, exceptuando las dos mencionadas anteriormente. Por lo tanto, el conjunto de variables predictoras está compuesto por la potencia, la temperatura del rodamiento 1 del generador y la temperatura del rodamiento 2 del generador.

Con la selección de variables mencionada, se creó el modelo de Proceso Gaussiano con el fin de modelar la temperatura de la 3ra fase del generador, a partir de los datos del aerogenerador A_0, conteniendo un total de 88939 de datos filtrados.

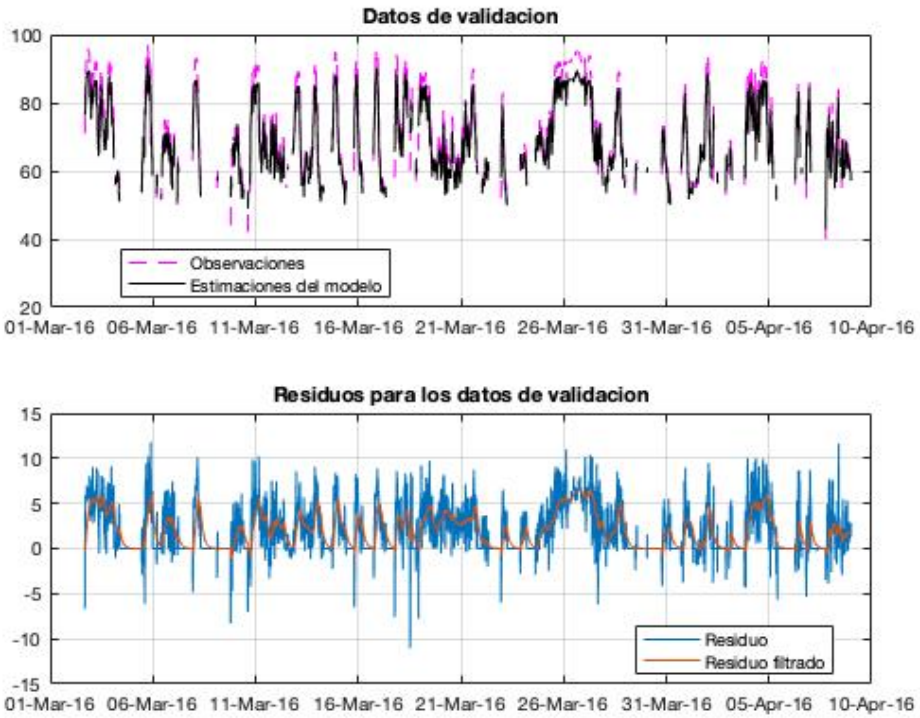
El modelo debe ser validado con el fin de cuantificar qué tan bien responde a las mediciones de la variable modelada. En ese sentido, se selecciona como período de validación los datos del aerogenerador A comprendidos entre febrero de 2016 y abril de 2016, conteniendo un total de 3212 registros, bajo la hipótesis

de que el aerogenerador A funcionaba sin desperfectos en ese entonces. En la Figura 3.3 se presentan las señales y_v y $\bar{y_v}$, como así también la de R_v^*, obtenida a partir del filtro de la Ecuación 3.14, propuesto por Wang and Infield (2013).

$$R_t^*[i] = 0.95R_t^*[i-1] + \frac{1}{39}R_t[i] + \frac{0.95}{39}R_t[i-1] \tag{3.14}$$

Se puede apreciar en la Figura 3.3 que el residuo filtrado R_v^* se mantiene

Figura 3.3: Resultados del modelo para los datos de validación. Comparación entre las mediciones y_v y $\bar{y_v}$ (arriba); evolución de R_v y R_v^* (abajo).



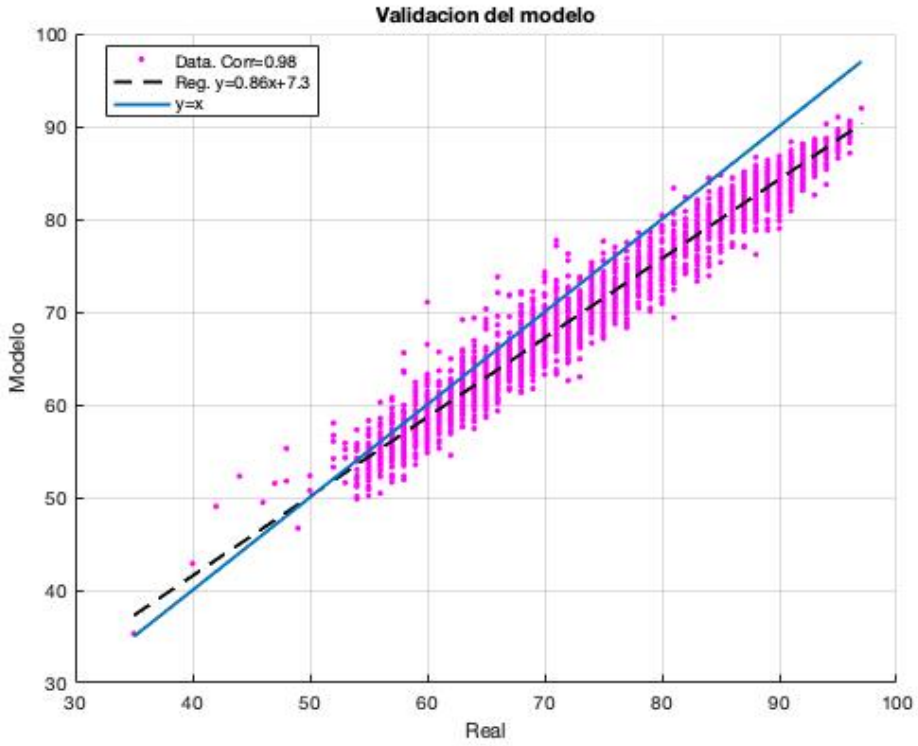
en general en valores bajos de $|R_v^*|$, sin presentar anormalidades durante su evolución. Con el fin de cuantificar esta validación del modelo, se presenta en la Figura 3.4 un diagrama de dispersión entre los valores de y_v y $\bar{y_v}$, junto con la recta de regresión que mejor ajusta a los puntos. Si bien esta recta no difiere apreciablemente de la recta identidad, puede observarse que la recta ajustada a los puntos del modelo subestiman a la realidad; más precisamente para aquellos puntos con temperatura superior a los $52°C$. En este sentido, en este análisis, el ajuste lineal realizado a los puntos del diagrama tiene como finalidad principal la recalibración del modelo construido. Por este motivo, esta

sub-estimación debe ser tenida en cuenta a la hora de aplicar el modelo a X_t: el modelo presenta un sesgo que debe ser corregido en la predicción $\bar{y}_t$. Si la recta de regresión de la Figura 3.4 es de la forma $y = a^R x + b^R$, entonces la corrección del sesgo será implementada de acuerdo a la Ecuación 3.15.

$$\hat{y}_t = \bar{y}_t + (1 - a^R)y_t - b^R \tag{3.15}$$

Donde y_t son las observaciones correspondientes a los datos de testeo, $\bar{y}_t$ es la variable modelada a partir del modelo sin considerar el sesgo obtenido en su construcción, y $\hat{y}_t$ es la variable modelada considerando el sesgo. De esta forma, $\hat{y}_t$ se empleará para lo que sigue en el análisis de los resultados empleados para el conjunto de datos de testeo.

Figura 3.4: Validación del modelo construido.

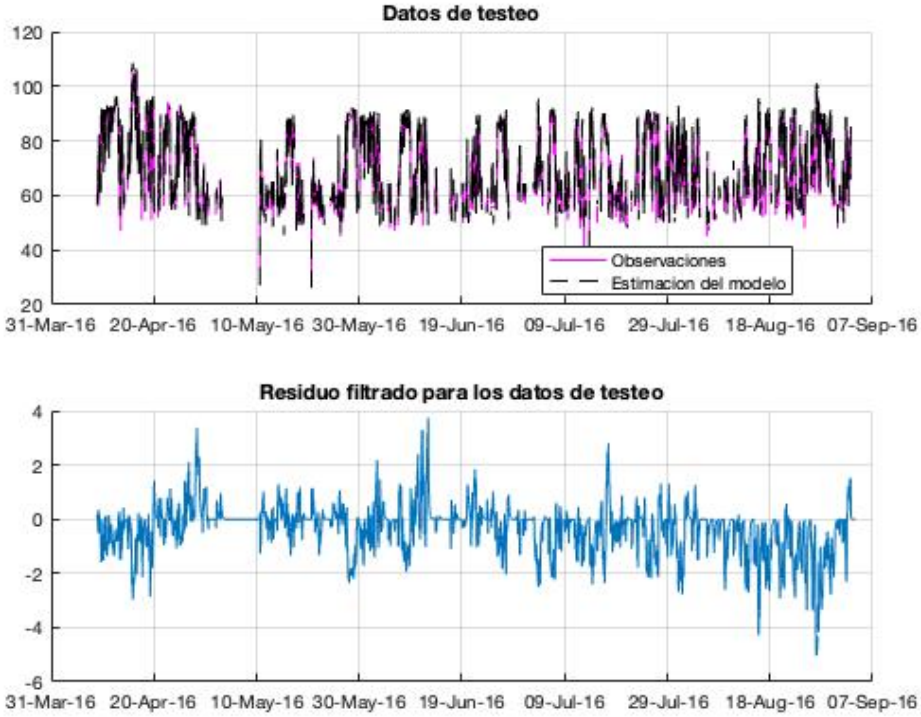


El conjunto de datos de testeo es considerado desde el abril de 2016 hasta el setiembre de 2016, comprendiendo un total de 12162 diezminutales con registros. Dentro de este período de tiempo están comprendidas las dos fallas registradas por el operador del parque eólico, por lo que el objetivo es aplicar el modelo desarrollado con el fin de detectar un funcionamiento anómalo del

aerogenerador de forma anticipada y así poder predecir las fallas.

En la Figura 3.5 se presentan los resultados obtenidos a partir del modelo para los datos de X_t. En la parte superior de la figura aparece la evolución de las señales y_t y $\hat{y}_t$; y en la parte inferior la evolución de R_t^*, obtenido como la salida del filtro de la Ecuación 3.14 de $y_t - \hat{y}_t$ en este caso.

Figura 3.5: Resultados del modelo para los datos de testeo. Comparación entre las mediciones y_t y $\hat{y}_t$ (arriba); evolución de R_t^* (abajo).



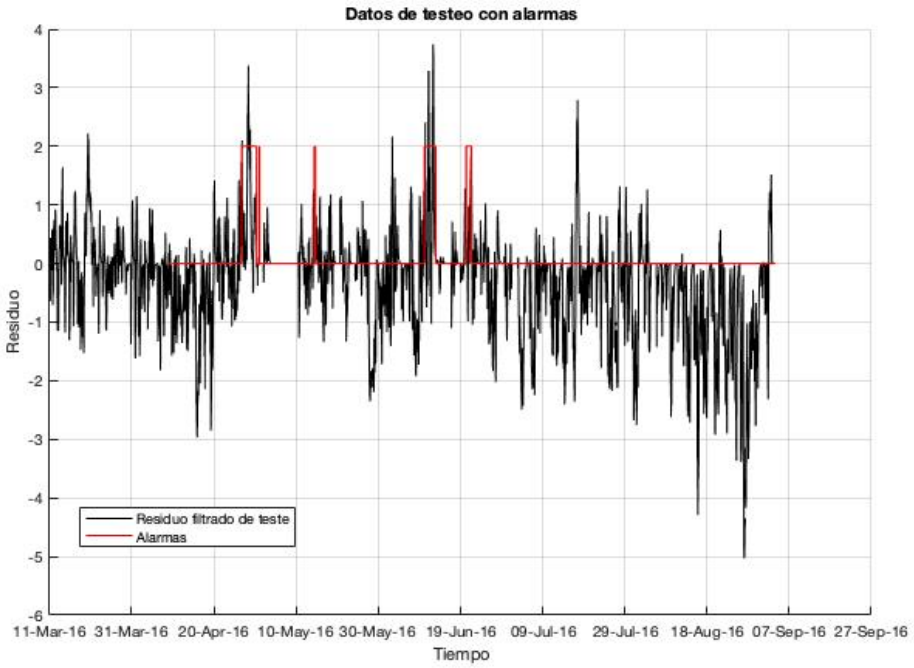
Cabe mencionar que para lo que sigue es de interés exclusivo los valores positivos de R_t^*, ya que los negativos están asociados a puntos donde el modelo sobre-estima las observaciones, no correspondiendo esto a una potencial anomalía. Esto se debe a que la variable modelada, la temperatura de la 3ra fase del generador, presenta riesgo de funcionamiento en valores elevados. Por lo tanto, el test de diferencia de medias que se aplica en lo que sigue es de una cola.

Para realizar el test estadístico a R_t^*, se realizó un estudio de sensibilidad en relación a LV y α_W. Este estudio consistió en calibrar estos parámetros para que la cantidad de alarmas generadas por el método sea reducida. A partir de ello, se optó por trabajar con un valor de LV igual a 432 diezminutales,

es decir, 72 horas, $\alpha_W = 0.01$, y un valor de r igual a 200; en particular, el parámetro LV es una escala de tiempo asociada al alcance predictivo del modelo.

En la Figura 3.6 se presenta, sobre la evolución de R_t^*, las alarmas detectadas por el test estadístico, teniendo en cuenta la consideración anterior. Cabe

Figura 3.6: Evolución de R_t^* junto con las alarmas obtenidas.



mencionar que, a modo de representación gráfica, las alarmas están están asociadas al punto medio de la ventana temporal de largo LV. Las herramientas de detección desarrolladas fueron capaces de detectar algunas alarmas durante el funcionamiento del aerogenerador. En particular, algunas de ellas se concentran inmediatamente antes de la parada de abril 2016. Otro grupo de alarmas son detectadas entre mayo y junio de ese mismo año.

Lo anterior permite una satisfactoria predicción de la falla de abril, teniendo una anticipación de un mes. Las otras alarmas pueden asociarse a una predicción de la falla de setiembre, aunque no con la misma certeza que la predicción de la primera falla. A diferencia de la primera falla, para la parada de setiembre el modelo presenta alarmas con tres meses de anticipación, dejando de reconocer anomalías frecuentes en los dos meses previos.

Luego de la falla ocurrida en abril, el modelo siguió registrando alarmas

durante los siguientes dos meses. Esto puede deberse a que en la primera parada el problema no fue resuelto de forma definitiva, dejando secuelas para lo que luego ocurrió en setiembre.

3.2.2. Aplicación para el Aerogenerador B

El método de Proceso Gaussiano fue aplicado para predecir la falla presentada durante la producción del aerogenerador B.

En esta oportunidad de decidió tomar la potencia del aerogenerador como variable a modelar -variable objetivo-, dentro del conjunto de variables pre-seleccionadas en la sección 2.2. Es necesario entonces definir el conjunto de variables predictoras. Para esto, en la Tabla 3.2 se presentan las correlaciones entre la potencia y las demás variables pre-seleccionadas. A partir de esta

Tabla 3.2: Correlaciones entre la potencia y las demás variables pre-seleccionadas.

Potencia	1
Temperatura del aceite hidráulico	0.46
Temperatura del rodamiento A de la CM	0.80
Temperatura del rodamiento B de la CM	0.79
Temperatura del rodamiento C de la CM	0.73
Temperatura de aceite de la CM	0.66
RPM del rotor	0.81
Velocidad de viento	0.91

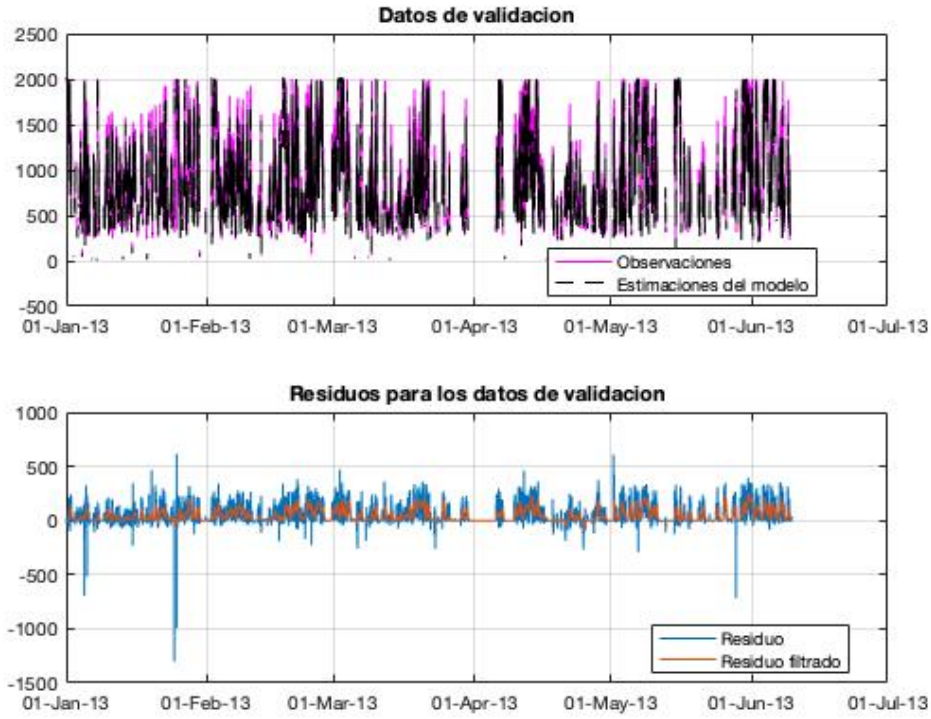
información, se toma la decisión de seleccionar, en primer lugar, la variable con mayor valor de correlación, a saber, la velocidad de viento; y en segunda instancia, la variable con mayor correlación y que este relacionada directamente con la CM, esta es la temperatura del rodamiento A de la CM.

Con la seleccion de variables mencionada, se creó el modelo de Proceso Gaussiano con el fin de modelar la potencia generada por la turbina. El período de entrenamiento seleccionado para este fin está comprendido entre abril de 2011 y diciembre de 2012, conteniendo un total de 52427 datos filtrados.

El modelo construido debe ser validado para cuantificar qué tan bien responde a las mediciones de la variable modelada. En este sentido, se selecciona como período de validación los datos comprendidos entre enero de 2013 y junio de 2013, conteniendo un total de 12086 registros. En la Figura 3.7 se presentan los resultados obtenidos para la etapa de validación. En la parte superior se presentan las estimaciones del modelo $\bar{y}_v$, junto con las observaciones y_v;

mientras que en la parte inferior figuran las evoluciones de R_v y de R_v^*. Con

Figura 3.7: Resultados del modelo para los datos de validación. Comparación entre las mediciones y_v y $\bar{y_v}$ (arriba); evolución de R_v y R_v^* (abajo).

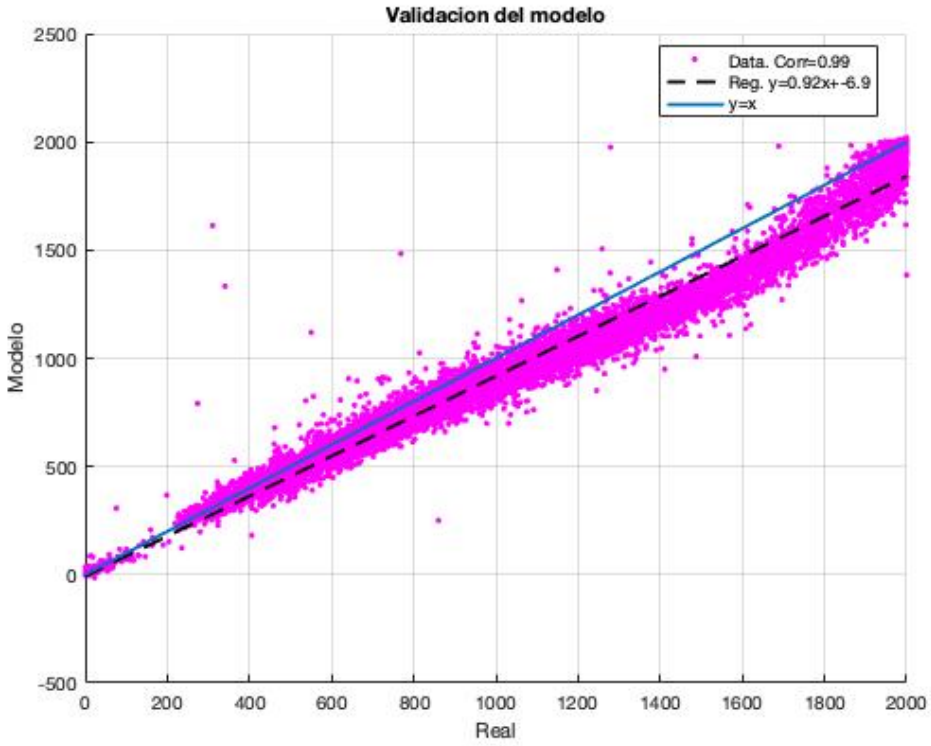


el fin de cuantificar esta validación, en la Figura 3.8 se presenta un diagrama de dispersión entre los valores de y_v y $\bar{y_v}$, junto con la recta de regresión que mejor ajusta a los puntos. Puede observarse que el modelo subestima la recta de ajuste en todo el rango. Por este motivo, esta sub-estimación es tenida en cuenta a la hora de evaluar el modelo en el conjunto de datos X_t. Este sesgo es corregido de acuerdo a la Ecuación 3.15. Por lo tanto, el análisis que sigue será hecho en base a $\hat{y}_t$.

El conjunto de datos de testeo está comprendido en el período que va desde junio de 2013 hasta setiembre de 2014. Cabe recordar que la falla presentada por el aerogenerador B fue en abril de 2014, comprendida en el período de testeo seleccionado. Este período contiene a su vez 32150 diezminutales con registros. El objetivo es entonces aplicar el modelo desarrollado con el fin de detectar un funcionamiento anómalo del aerogenerador de forma anticipada y, así poder predecir la falla.

En la Figura 3.9 se presentan los resultados obtenidos para el conjunto

Figura 3.8: Validación del modelo.



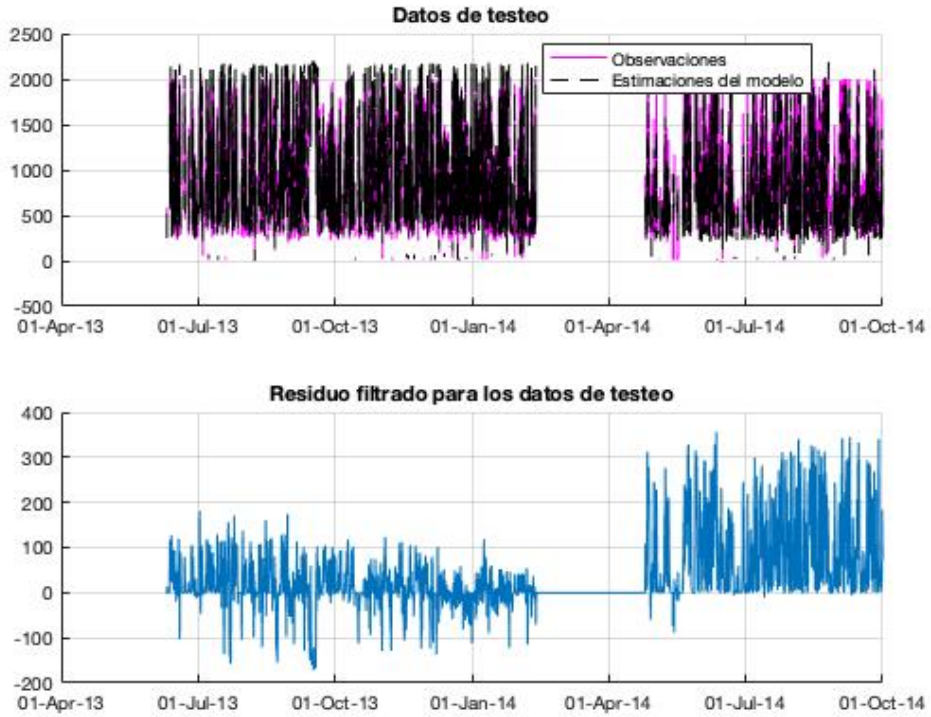
de datos X_t. En la parte superior de la figura se presenta la evolución de las señales y_t y $\hat{y}_t$; mientras que en la parte inferior aparece la evolución de R_t^*, obtenido como la salida del filtro de la Ecuación 3.14, considerando $\hat{y}_t$ en lugar de $\bar{y}_t$.

En lo que sigue, es de interés exclusivo los valores negativos de R_t^*, ya que los positivos están asociados a puntos donde el modelo sub-estima las observaciones, no correspondiendo esto a una potencial anomalía. Esto se debe a que la variable modelada, la potencia, presenta indicios de deterioro en la performance para valores negativos de R_t^*.

Al igual que en la sección 3.2.1, con el fin de determinar los parámetros LV y α_W, se realizó un estudio de sensibilidad en base a estos. A partir de este, se optó por tomar los mismos valores ya empleados en la sección anterior, a saber, $LV = 432$ diezminutales y $\alpha_W = 0.01$. Asimismo, el valor de $r = 200$ también se mantuvo fijo.

En la Figura 3.10 se presenta, sobre la evolución de R_t^*, las alarmas detectadas por el test estadístico. El método de detección desarrollado identificó alarmas durante la producción del aerogenerador, donde su gran mayoría

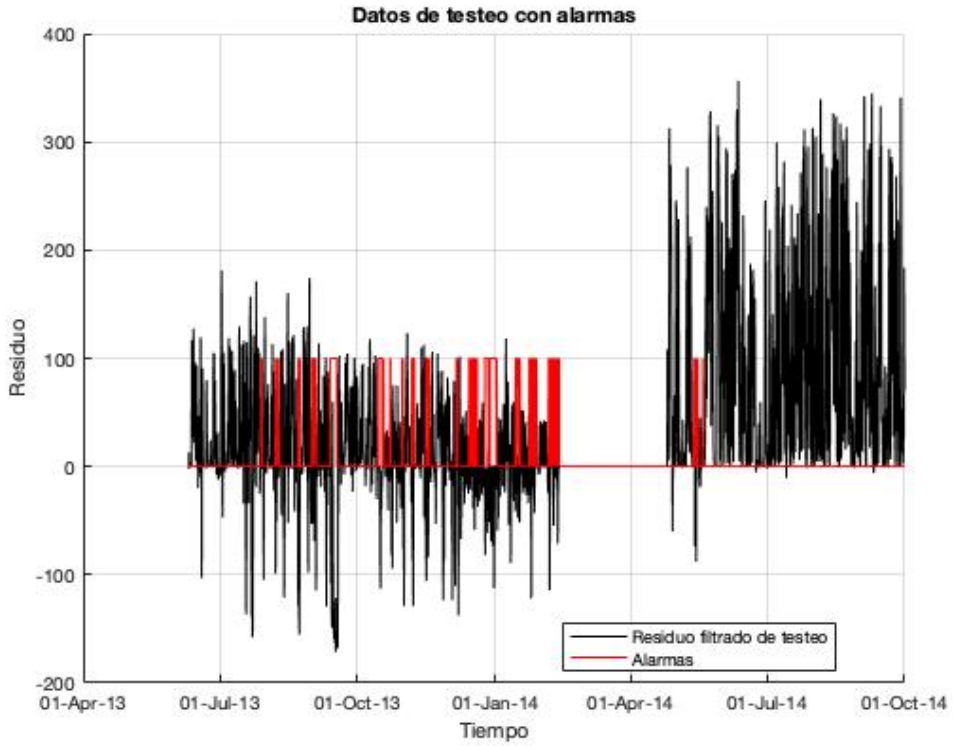
Figura 3.9: Resultados del modelo para los datos de testeo. Comparación entre las mediciones y_t y $\hat{y}_t$; evolución de R_t^*.



se encuentran previo a la falla de abril de 2014. Asimismo, puede observarse en la figura que la herramienta detecta una gran concentración de alarmas justo antes de la parada.

Lo anterior permite observar un deterioro progresivo en la performance del aerogenerador B mediante la detección de las alarmas. Esto a su vez permite predecir la falla de abril de 2014 de forma acertada, teniendo una gran concentración antes de esta ocurrencia.

Figura 3.10: Evolución de R_t^* junto con las alarmas obtenidas.



3.2.3. Aplicación para el Aerogenerador D

El método GP fue aplicado para los datos provenientes del aerogenerador D. Sin embargo, como se adelantó en la sección 2.4, no se conoce una falla asociada a la operación de esta turbina. Por este motivo, la aplicación del método estará centrada en comparar el funcionamiento entre antes y después de la parada, y no en la detección de alarmas asociadas a una falla específica, como fue desarrollado en los casos precedentes.

Se decidió trabajar con la potencia como variable objetivo a modelar. Es necesario entonces definir el conjunto de variables predictoras. Para esto, en la Tabla 3.3 se presentan las correlaciones entre la potencia y las demás variables. A partir de esta, y de la información ilustrada en la Figura 2.18, se toma la decisión de seleccionar como variables predictoras aquellas dos con mayores valores de correlación: la velocidad de viento y las RPM del rotor. Cabe destacar que las demás variables presentan valores de correlación sustancialmente menores.

Con la selección de variables mencionada, se creó el modelo de Proceso

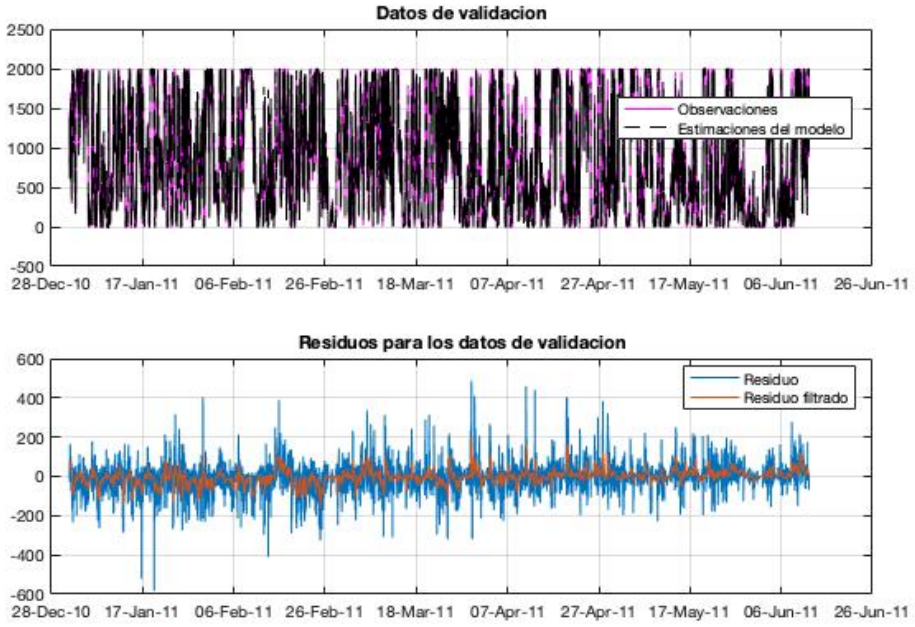
Tabla 3.3: Correlaciones entre la potencia y las demás variables.

Potencia	1
Velocidad de viento	0.95
Dirección de viento	0.10
Temperatura ambiente	-0.15
Yaw	0.10
RPM del rotor	0.66
Pitch	0.10

Gaussiano con el fin de modelar la potencia generada por la turbina D. El período de entrenamiento seleccionado está comprendido entre enero de 2010 y diciembre de 2010, conteniendo un total de 49969 datos filtrados.

Con el fin de validar el modelo obtenido, se seleccionó el periodo de validación comprendido entre enero de 2011 y junio de 2011, conteniendo un total de 23041 datos filtrados. En la Figura 3.11 se presentan, en el panel superior, la comparación entre las observaciones y_v y las estimaciones $\bar{y}_v$, y en el panel inferior, las evoluciones de R_v y R_v^*. Con el fin de cuantificar esta validación,

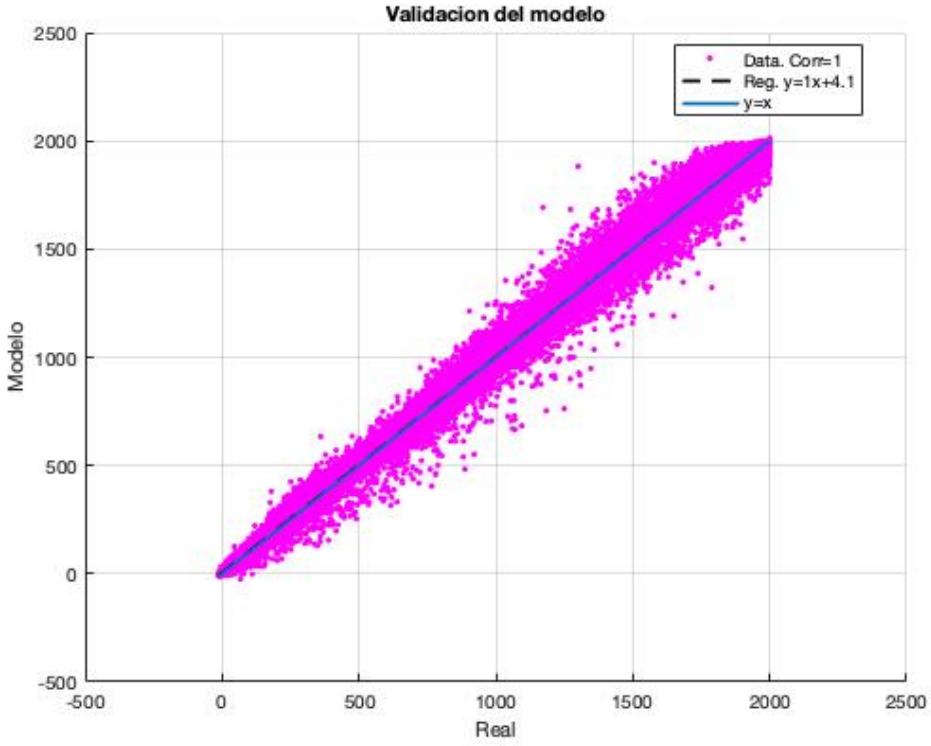
Figura 3.11: Resultados del modelo para los datos de validación. Comparación entre las mediciones y_v y $\bar{y}_v$ (arriba); evolución de R_v y R_v^* (abajo).



en la Figura 3.12 se presenta un diagrama de dispersión entre los valores de las observaciones y_v y de las estimaciones $\bar{y}_v$, junto con la recta de regresión que

mejor ajusta a los puntos. Puede observarse que esta recta aproxima adecuadamente con la recta $y = x$, por lo que puede deducirse que el modelo ajusta satisfactoriamente a las observaciones.

Figura 3.12: Validación del modelo construido.

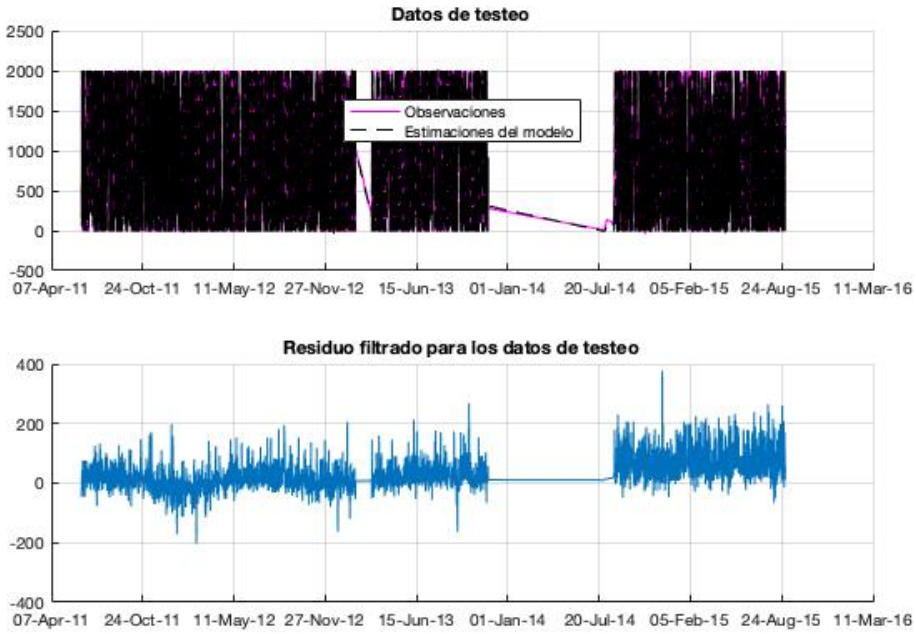


La cuantificación anterior permite avanzar a la etapa de testeo, donde el período destinado para este fin está comprendido entre junio de 2011 y agosto de 2015, comprendiendo el período correspondiente a la parada de mantenimiento antes mencionada. En este sentido, en la Figura 3.13 se presenta, en el panel superior, la comparación entre las mediciones y_t y las estimaciones $\bar{y}_t$, mientras que en el panel inferior figura la señal correspondiente a R_t^*.

Como se introdujo anteriormente, el objetivo en este caso no es identificar alarmas asociadas a un funcionamiento anómalo, ya que no se tiene información sobre alguna falla asociada al equipo. La intención es entonces comparar el funcionamiento del aerogenerador D, a partir de los resultados obtenidos mediante el modelo GP, entre los períodos previos y posteriores a la parada de noviembre de 2013.

En este sentido, un test estadístico de diferencia de medias es aplicado para comparar las señales de R_t^* antes y después de la parada, cuyos valores medios

Figura 3.13: Resultados del modelo para los datos de testeo. Comparación entre las mediciones y_t y $\hat{y}_t$; evolución de R_t^*.



son $14.55kW$ y $70.40kW$, respectivamente. En la sección 3.1.1 fue desarrollado este test, aunque con otra finalidad. De acuerdo a eso, las Ecuaciones 3.11, 3.12 y 3.13 son implementadas para comparar las dos ventanas operacionales de interés.

El nivel de significancia elegido es de 0.05, con 208 grados de libertad (Ecuación 3.13). En este caso en particular, el test debe ser aplicado a dos colas, ya que en principio los dos períodos están en pie de igualdad. Asimismo, el valor de t, calculado a partir de la Ecuación 3.12, es de 13.82. Los resultados de este test estadístico aplicados al caso de estudio muestran que hay suficiente evidencia como para rechazar la hipótesis nula de que los valores medios de ambos períodos son iguales.

Lo anterior confirma que el período posterior a la parada presenta una mejora sustancial en el funcionamiento de la turbina. Las conclusiones abordadas serán exploradas en el capítulo 7. En este sentido, el método GP permitió en este caso realizar un monitoreo por condición, a partir de la señal de R_t^*, del aerogenerador D, pudiendo además detectar que luego de la parada de mantenimiento el equipo presentó una mejoría en su funcionamiento.

Capítulo 4

Técnica de estimación de estado no-lineal (NSET)

En este capítulo se desarrolla una técnica de estimación de estado no-lineal del aerogenerador, que tiene como objetivo estudiar la condición de funcionamiento del equipo, y poder así predecir fallas asociadas al mismo. Guo et al. (2012) y Wang and Infield (2013) presentaron esta técnica, y en este capítulo se desarrolla la misma idea, aunque incluyendo variantes. Finalmente, se aborda la aplicación de esta técnica a los cuatro aerogeneradores presentados en la sección 2.

4.1. Descripción teórica y metodología de aplicación

NSET (por su sigla en inglés) proporciona una técnica de reconocimiento de vectores basada en el estado de condición, que se aplica en este trabajo a la supervisión del estado de condición de turbinas eólicas. Los vectores de estado consisten en la lectura del sensor del SCADA en un diezminutal específico para las variables que están estrechamente vinculadas con la salida del modelo. Este modelo es no-paramétrico, y aprende las relaciones de los datos al calcular la similitud entre las señales de entrada y los vectores de estado históricos almacenados.

Al igual que el método presentado en la sección 3, los datos disponibles se dividen en tres subconjuntos disjuntos de datos: los datos de *entrenamiento*, los de *validación* y los de *testeo*. Los dos primeros corresponden a datos saludables

del aerogenerador, mientras que el último puede eventualmente tener fallas asociadas al funcionamiento del equipo.

Deben seleccionarse un conjunto de variables predictoras, presentes en el registro del SCADA, a partir de las cuales se modela otra variable, también del SCADA, con el fin de realizar el análisis objetivo. Esta selección de variables debe realizarse adecuadamente con la finalidad de estudiar el problema de interés.

Sea M la Matriz de Memoria, en la que se almacenan los vectores de estado asociados a las variables de interés seleccionadas (tanto las predictoras como la variable a ser modelada); estos estados se seleccionarán a partir de una matriz de estados, como se describe a continuación. Cada columna de M (vector de estado) representa un estado de operación del sistema sensado en un determinado diezminutal, y cada fila de M tiene mediciones de un sensor específico del SCADA, es decir, registros de una variable en particular. La Ecuación 4.1 describe la estructura general de la matriz M, donde hay m vectores de observación y n variables involucradas.

$$M = [X(1) \quad X(2) \quad \dots \quad X(m)] = \begin{bmatrix} x_1(1) & x_1(2) & \dots & x_1(m) \\ x_2(1) & x_2(2) & \dots & x_2(m) \\ \vdots & \vdots & \dots & \vdots \\ x_n(1) & x_n(2) & \dots & x_n(m) \end{bmatrix} \tag{4.1}$$

La selección de las variables y la construcción de M son las dos etapas más importantes a la hora de realizar el modelo. Una vez que están determinadas las variables a incluir, se construye M a partir de los datos de *entrenamiento*, los que deben abarcar suficientes estados de operación para cubrir el rango completo de operación normal, incluyendo el comportamiento esperado en condiciones extremas.

Un gran números de estados involucrados en M puede hacer que las operaciones matriciales involucradas en la construcción del modelo consuman excesivos recursos de cómputo. Además, el aumento en el número de estados más allá de cierto límite no contribuye a una mayor precisión del modelo. En consecuencia, se debe utilizar un algoritmo de selección de datos, como el que Guo et al. (2012) presentan, para elegir los vectores de estado de manera uniforme y económica a partir del conjunto de datos de entrenamiento. En este sentido, el número de estados se reduce drásticamente, haciendo que el modelo sea mucho

más efectivo. Esta reducción de estados, es decir, la cantidad de columnas de M, se determina gráficamente, comparando la precisión de la validación y el tamaño de M. De esta forma, se evita que la Matriz de Memoria tenga un tamaño innecesariamente grande.

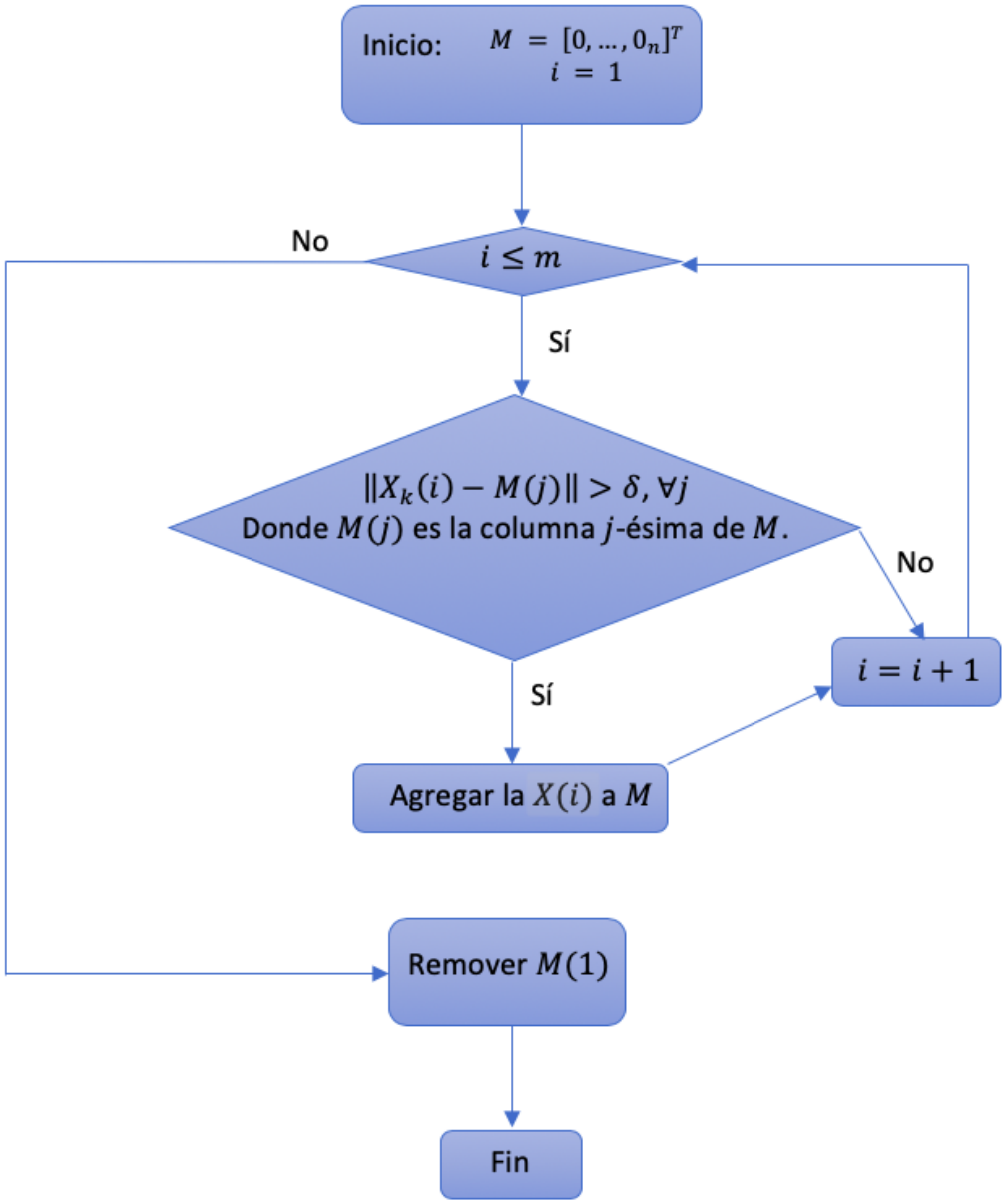
Cabe aclarar que dentro de la construcción de M no se permiten estados repetidos, ya que esto trae como consecuencia una singularidad asociada con la inversión de otras matrices que se introducirán más adelante, operación necesaria para la construcción del algoritmo que se presenta a continuación.

En este sentido, Guo et al. (2012) y Wang and Infield (2013) presentan algoritmos de selección de estados. Sin embargo, en este trabajo se opta por trabajar con otro que resulta más eficiente para los casos de estudio, aunque manteniendo el mismo objetivo conceptual que los presentados por los autores citados. En la Figura 4.1 se presenta un esquema del algoritmo de selección de estados empleado para la construcción de M. El parámetro δ es un valor predefinido positivo cercano a 0, y X_k es la matriz de mediciones normalizadas entre 0 y 1; esta normalización se hace, restando a las señales temporales su valor mínimo, y luego dividiendo entre la diferencia del máximo con el mínimo correspondiente . El fundamento principal del algoritmo es proyectar todos los vectores de estado del conjunto de entrenamiento en el hipercubo $[0, 1]^n$, y que los puntos elegidos de X_k, asociados a las mediciones que luego aparecerán en M, mantengan una distancia mayor a δ entre sí. De esta forma, M estará compuesta por mediciones que cubren todos los rangos operacionales, evitando a su vez repeticiones de estados. La construcción preliminar de M se lleva a cabo aplicando el algoritmo descrito anteriormente. Como se mencionó, δ es un parámetro predefinido, el cual debe ser ajustado para definir el tamaño final de M: cada valor de δ estará asociado a un tamaño de M distinto, que tendrá a su vez asociado una precisión en el modelo.

Dos vectores adicionales deben ser agregados a M: el vector nulo en la columna inicial y el vector de mediciones máximas en la columna final, es decir, las máximas mediciones registradas para las n variables. La razón principal para agregar el vector nulo es mejorar la precisión del modelo, especialmente en el contexto de datos faltantes, mientras que el vector maximal se agrega a M para mejorar la precisión del modelo en el caso de observaciones extremas. Wang and Infield (2013) presentan algunas comparaciones de modelos en los que no se agregan estos dos vectores y en los que sí.

Una vez construida la matriz M a partir de los datos de entrenamiento,

Figura 4.1: Algoritmo de selección de estados para la construcción de M.



esta es utilizada para modelar la performance de la turbina para nuevas observaciones de las n variables seleccionadas. Más precisamente, para una nueva observación X_{obs}, la estimación X_{est} del mismo es obtenida como el producto de M y un vector de pesos $W = [w_1, w_2, \ldots, w_m]^T$ que captura el grado de similaridad entre X_{obs} y cada vector de estado almacenado en M.

$$X_{est} = w_1 X(1) + w_2 X(2) + \cdots + w_m X(m) \tag{4.2}$$

El vector de pesos W es determinado mediante la minimización del cuadrado de la diferencia entre X_{obs} y X_{est}.

$$\Delta^2 = (X_{obs} - X_{est})^2 = (X_{obs} - MW)^2 \tag{4.3}$$

Para minimizar esta diferencia, se determina la primer derivada de Δ^2 y se la iguala a 0.

$$\frac{d\Delta^2}{dW} = \frac{d\,(X_{obs} - MW)^2}{dW} = 0 \Rightarrow M^T\,(X_{obs} - MW) = 0 \tag{4.4}$$

La matriz $M^T M$ no tiene inversa debido a su rango. Esto hace que sea necesario introducir el operador $\otimes$, de forma que el vector de pesos W puede determinarse mediante la Ecuación 4.5 (Wang and Infield (2013), Guo et al. (2012), Black et al. (1998), Herzog et al. (1998)).

$$W = \left(M^T \otimes M\right)^{-1} \left(M^T \otimes X_{obs}\right) \tag{4.5}$$

Se trata de un operador no lineal usado para remplazar la multiplicación estándar de matrices. Si bien existen diversas opciones para el operador $\otimes$, en este caso se opta por trabajar con la distancia Euclidea entre dos matrices, de forma que se considera el siguiente operador.

$$(P \otimes Q)_{ij} = \|P_{(i)} - Q^{(j)}\|_2 \tag{4.6}$$

Donde $P_{(i)}$ es la fila i-ésima de P y $Q^{(j)}$ es la columna j-ésima de Q. De esta forma, el vector de pesos W, determinado mediante la Ecuación 4.5, refleja la similaridad entre el vector X_{obs} y los m estados almacenados en la matriz M. Así, si X_{obs} es similar solamente a un $X(i)$ en particular de la Matriz de Memoria M, su correspondiente peso w_i de W será cercano a la unidad y, cercano a 0 en otro caso (Black et al. (1998), Guo et al. (2012)).

De esta forma, es posible obtener una estimación para un dato observado.

Para el conjunto de datos de validación, se procede a construir M para distintos valores de δ. Para cada una de las matrices obtenidas, con los respectivos valores de δ (cada una con una cantidad m de columnas distinta), se determina $\epsilon^2 = \|Y_{obs} - Y_{est}\|_2^2$, donde Y_{obs}^v y Y_{est}^v son los valores correspondientes a la variable modelada, contenidos en X_{obs}^v y X_{est}^v, respectivamente, de todo el conjunto de datos de validación. Luego, el valor de δ escogido será determinado

a partir de una relación de compromiso entre ϵ^2 y la cantidad de estados en M. Una vez seleccionado δ, la matriz M queda determinada.

La siguiente etapa del método es la *validación*. Esta etapa consiste en la misma metodología desarrollada en la sección 3.1.1 para el modelo analizado en la sección 3. Para cada observación Y^v_{obs} del conjunto de datos de validación, se obtiene una estimación Y^v_{est} a partir de la Ecuación 4.2, procediendo entonces a la validación en sí misma.

Finalmente, una vez validado el modelo se procede a la etapa de *testeo*. Esta etapa consiste en la misma metodología ya explicada en la sección 3.1.1 para la etapa de testeo, valiendo todas las menciones y observaciones realizadas en esa sección. En particular, la entrada i-ésima del residuo R_t es obtenida de acuerdo a la siguiente ecuación, donde $Y^{t,i}_{obs}$ y $Y^{t,i}_{est}$ son la observación i-ésima del conjunto de testeo y la estimación i-ésima asociada, respectivamente.

$$R_t(i) = Y^{t,i}_{obs} - Y^{t,i}_{est} \tag{4.7}$$

4.2. Resultados

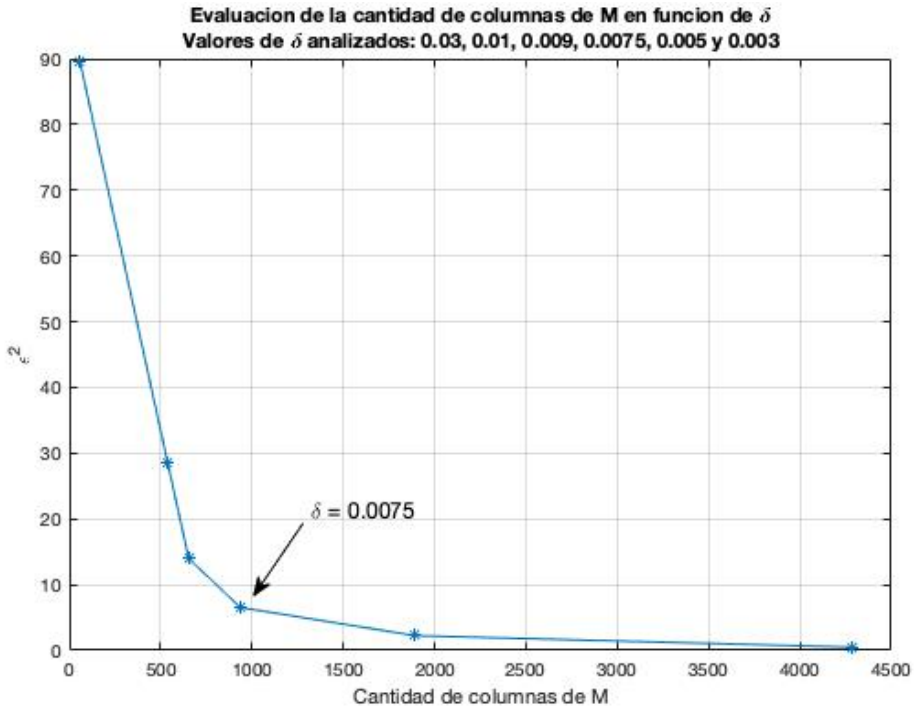
4.2.1. Aplicación para el Aerogenerador A

La metodología desarrollada en la sección 4.1 fue aplicada al conjunto de datos correspondiente al aerogenerador A.

NSET también fue aplicado para modelar la temperatura de la 3ra fase del generador, tal como se desarrolló en la sección 3.2.1, tomando como datos de entrenamiento los asociados al aerogenerador A_0, y los mismos períodos de validación y testeo del aerogenerador A. A su vez, las variables predictoras empleadas fueron también la potencia, la temperatura del rodamiento 1 del generador, y la temperatura del rodamiento 2 del generador; decisión tomada a partir de la información expuesta en la Figura 2.5 y en la Tabla 3.1. De este modo, NSET es desarrollado bajo las mismas condiciones que el Proceso Gaussiano de regresión analizado en la sección 3.

Con el fin de determinar el tamaño de la Matriz de Memoria M asociada al modelo, se procedió a un análisis de sensibilidad en función de δ. En la Figura 4.2 se presentan los resultados de este análisis para valores de δ iguales a 0.03, 0.01, 0.009, 0.0075, 0.005 y 0.003. A partir del resultado mostrado en la Figura 4.2, se optó por construir la matriz M tomando $\delta = 0.0075$, ya que este genera

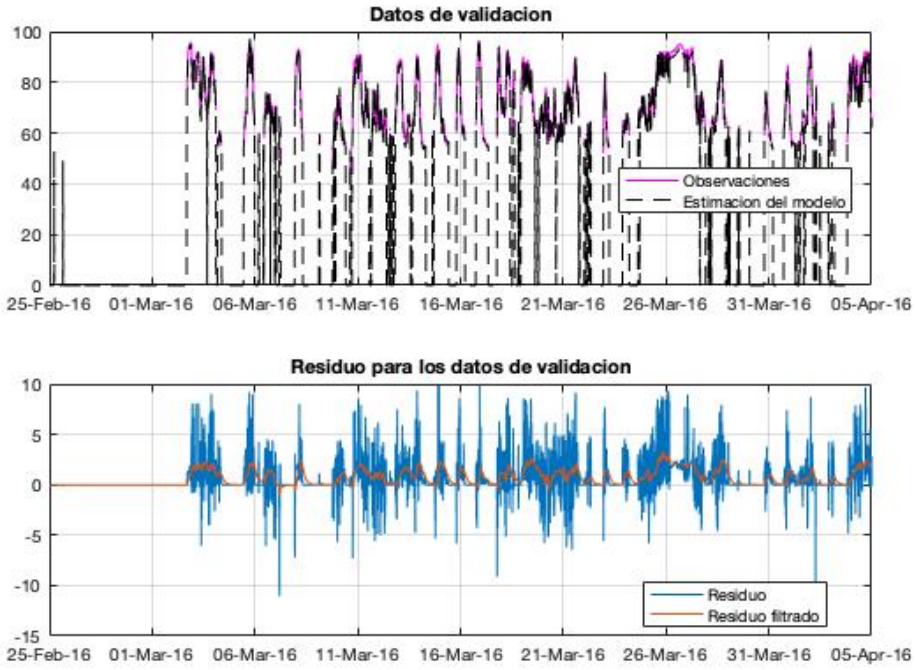
Figura 4.2: Análisis de sensibilidad en función de δ. Variación de ϵ^2 en función de m.



un bajo valor de ϵ^2 y, para valores mayores de m, la precisión del modelo no mejora significativamente. De este modo, M es construida con una cantidad de columnas $m = 937$, es decir, a partir de 937 estados correspondientes al conjunto de datos de entrenamiento, el que comprendía 88939 estados.

Para validar el modelo, se realizaron las estimaciones correspondientes al conjunto de datos seleccionado para tales fines. Se puede apreciar en la Figura 4.3 que el residuo filtrado R_v^* se mantiene en general en valores bajos de $|R_v^*|$, sin presentar anormalidades durante su evolución. Con el fin de cuantificar esta validación del modelo, se presenta en la Figura 4.4 un diagrama de dispersión entre los valores de Y_{obs}^v y Y_{est}^v, junto con la recta de regresión que mejor ajusta a los puntos. Si bien esta recta aproxima razonablemente a la recta identidad, puede observarse que el modelo subestima la recta de ajuste para la mayoría de los puntos de la nube; más precisamente para aquellos puntos con temperatura superior a los $52°C$. Por este motivo, esta sub-estimación debe ser tenida en cuenta a la hora de aplicar el modelo a los datos de testeo: el modelo presenta un sesgo que debe ser corregido en la predicción Y_{est}^t, de acuerdo a la Ecuación 3.15, generando así la nueva predicción $\hat{Y_{est}^t}$, que será empleada en lo que sigue.

Figura 4.3: Resultados del modelo para los datos de validación. Comparación entre las mediciones Y^v_{obs} y Y^v_{est} (arriba); evolución de R_v y R^*_v (abajo).



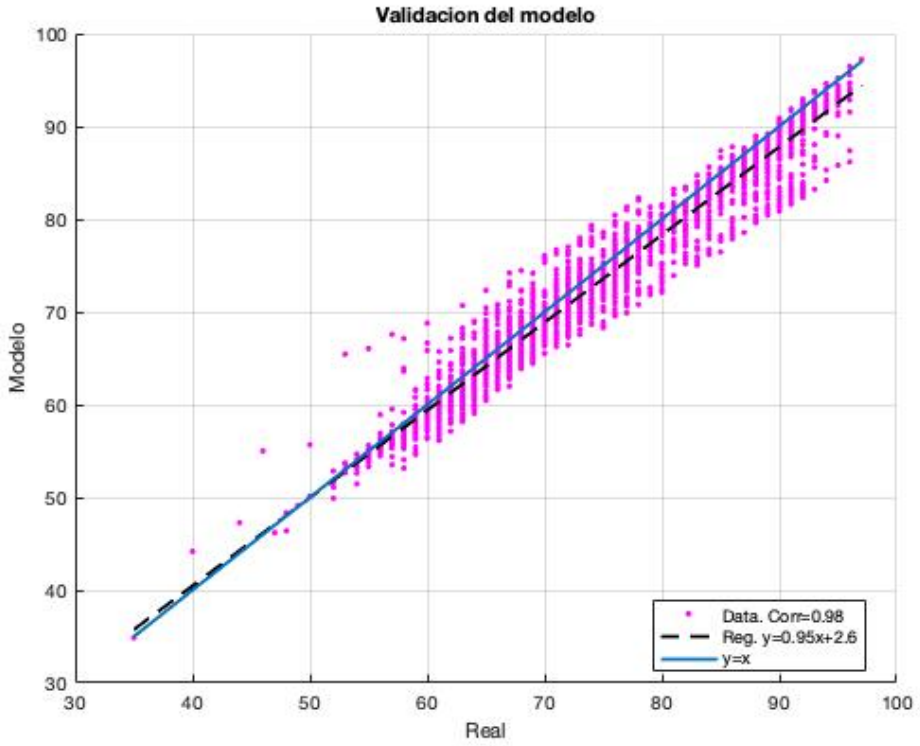
En la Figura 4.5 se presentan los resultados obtenidos a partir del modelo para los datos de testeo. En la parte superior de la figura aparece la evolución de las señales Y^t_{est} y $Y^{\hat{t}}_{obs}$, y en la parte inferior las evoluciones de R_t y R^*_t, obtenido como la salida del filtro de la Ecuación 3.14.

Al igual que en la sección 3.2.1, el análisis es de interés solamente para los valores positivos de R^*_t, ya que la temperatura de la 3ra fase del generador presenta riesgo de funcionamiento en valores elevados. Esto hace que nuevamente el test de diferencia de medias aplicado sea de una cola.

En la Figura 4.6 se presenta, sobre la evolución de R^*_t, las alarmas detectadas por el test estadístico. El método desarrollado en esta sección fue capaz de reconocer algunas alarmas durante el funcionamiento del aerogenerador. En particular, la parada de abril 2016 es anticipada con un mes de antelación. Asimismo, un numeroso conjunto de alarmas está presente inmediatamente a continuación de la parada de abril. Finalmente, la parada de setiembre 2016 es es predicha con dos meses de antelación.

Lo anterior permite una aceptable predicción tanto de la falla de abril 2016

Figura 4.4: Validación cuantitativa del modelo construido.



como de la de setiembre 2016. A su vez, la presencia de una gran cantidad de alarmas luego de la puesta en marcha del aerogenerador en mayo 2016, deja nuevamente la sospecha de que la segunda parada es consecuencia de secuelas que persistieron en el funcionamiento de la turbina luego de la parada de abril.

Figura 4.5: Resultados del modelo para los datos de testeo. Comparación entre Y_{obs}^{t} y $Y_{obs}^{\hat{t}}$ (arriba); evolución de R_t y R_t^* (abajo).

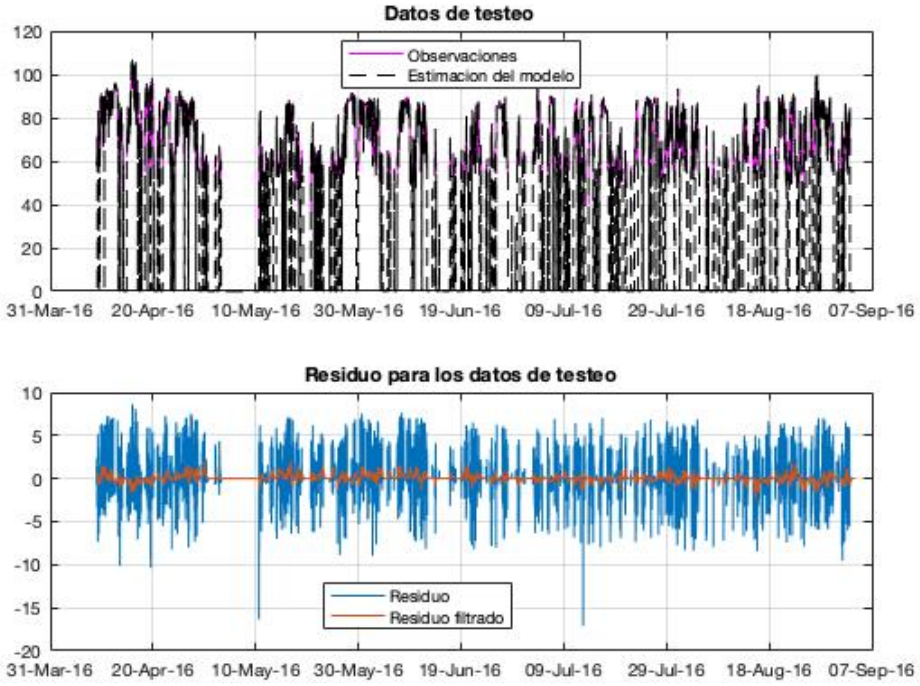
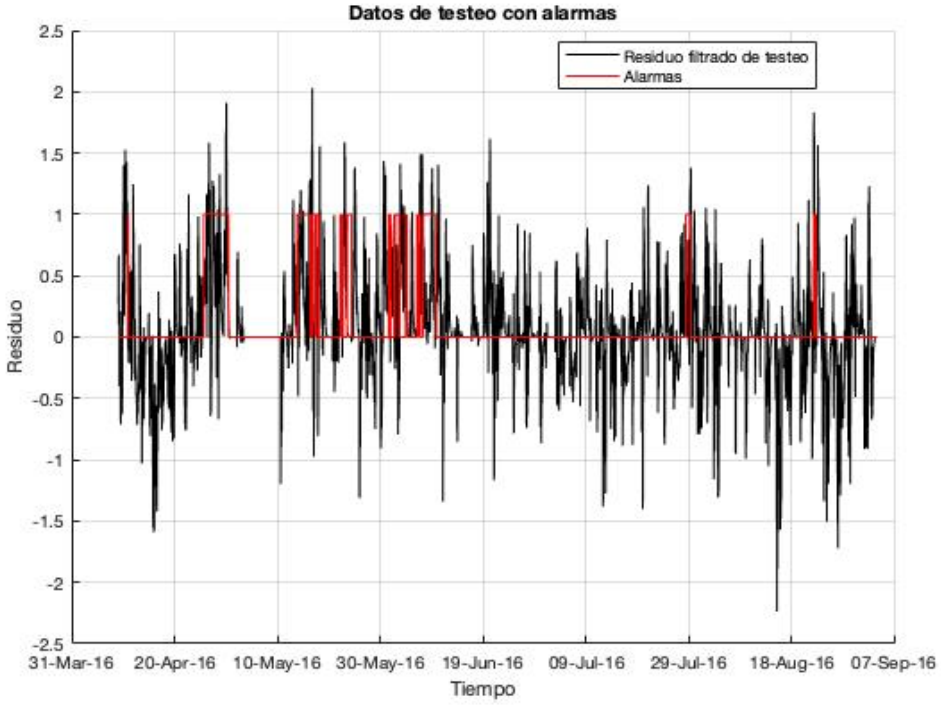


Figura 4.6: Evolución de R_t^* junto con las alarmas obtenidas.

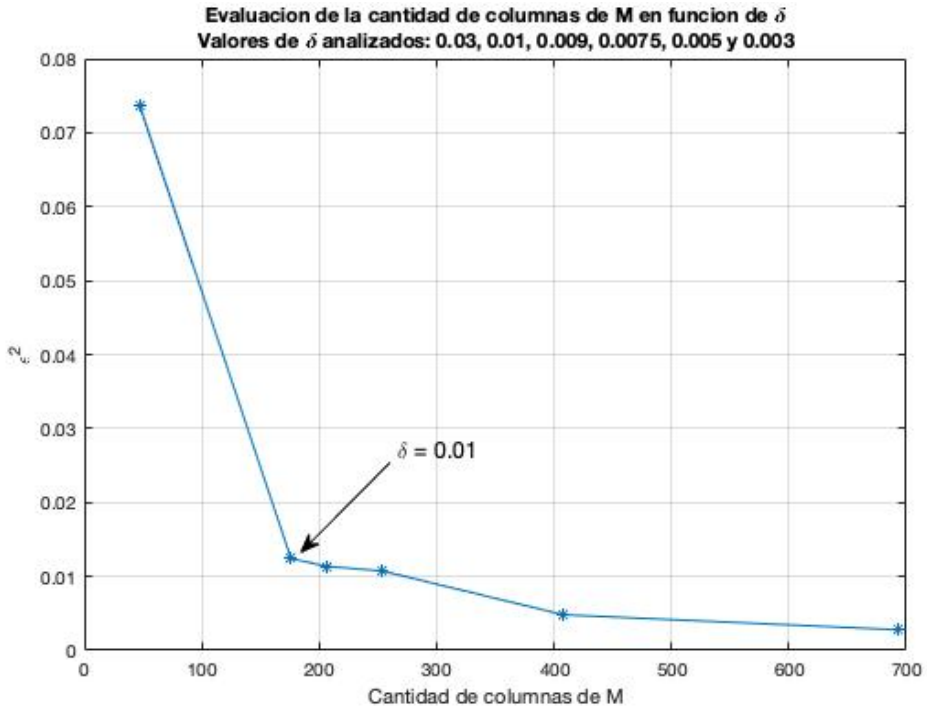


4.2.2. Aplicación para el Aerogenerador B

NSET fue aplicado, en este caso, para modelar la potencia generada por el aerogenerador B. Tal como se desarrolló en la sección 3.2.2, tomando como período de entrenamiento los datos comprendidos entre abril de 2011 y diciembre de 2012 (compuesto por 52427 datos filtrados en total), como período de validación los comprendidos entre enero de 2013 y junio de 2013 y, como período de testeo los sensados entre junio de 2013 y setiembre de 2014, este método fue aplicado tomando la temperatura del rodamiento A de la CM y la velocidad de viento como variables predictoras. Esta selección está fundada en la información mostrada en la Figura 2.9 y en la Tabla 3.2.

Para determinar la Matriz de Memoria M asociada al modelo, se condujo a un análisis de sensibilidad en función de δ. En la Figura 4.7 se presentan los resultados obtenidos para el error del modelo en función de la cantidad de estados m para cada valor de δ. Este análisis llevó a la decisión de tomar $\delta = 0.01$, parámetro que genera una matriz M de 176 columnas. Como se verá a

Figura 4.7: Análisis de sensibilidad en función de δ. Variación de ϵ^2 en función de m.

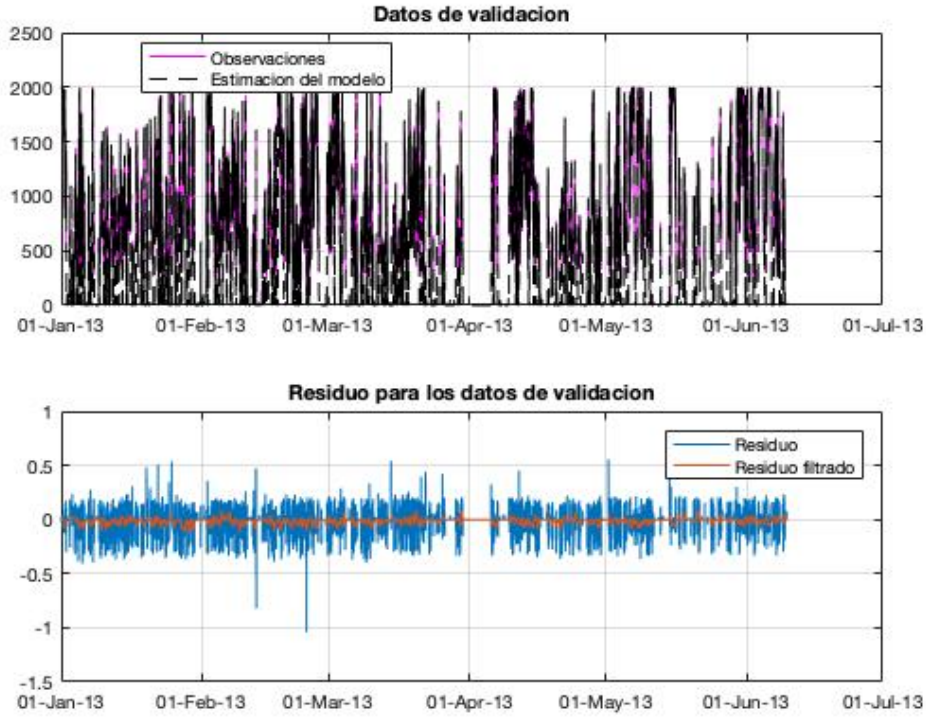


continuación, el valor de δ seleccionado conduce a un ajuste satisfactorio en la estimación. Una vez determinada la matriz M, queda entonces definido el

modelo NSET.

Con el fin de validar este modelo, se realizaron las estimaciones correspondientes al conjunto de datos de validación. En la Figura 4.8 se muestran las evoluciones temporales de Y^v_{obs} y Y^v_{est} en el panel superior y, las evoluciones temporales de R_v y R^*_v en el panel inferior. Con el fin de cuantificar esta validación, en la Figura 4.9 se presenta un diagrama de dispersión entre los valores de Y^v_{obs} y Y^v_{est}, junto con la recta de regresión que mejor ajusta a los puntos. Puede observarse que esta recta coincide prácticamente con la recta identidad durante todo el rango de operación. Por este motivo, se considera que el modelo no presenta sesgo apreciable en este caso y, por lo que el ajuste a las observaciones es adecuado.

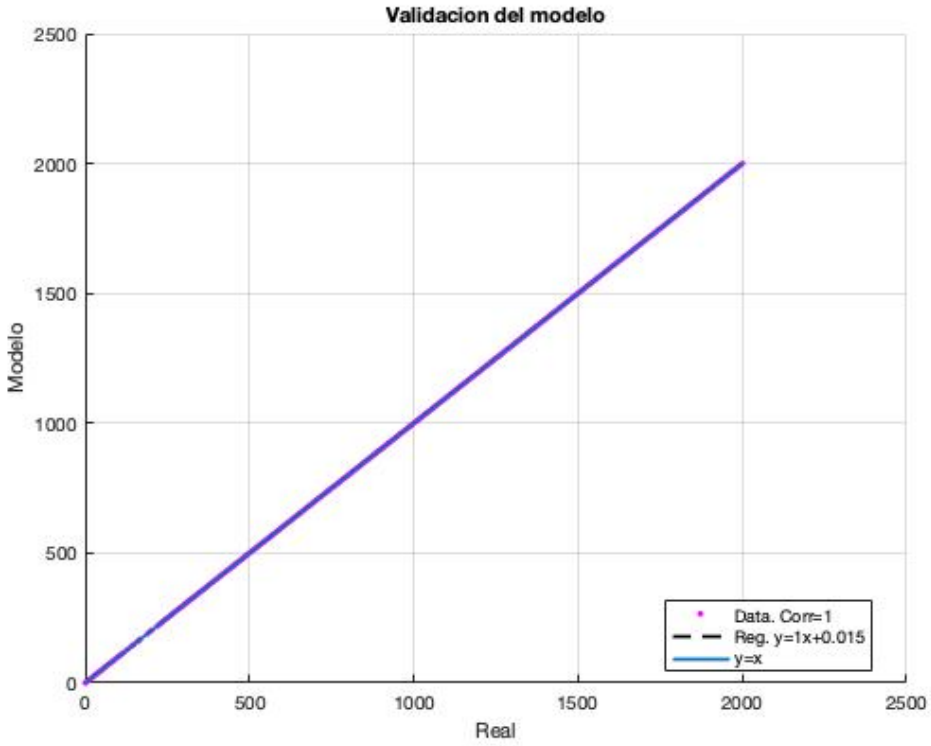
Figura 4.8: Resultados del modelo para los datos de validación. Comparación entre las mediciones Y^v_{obs} y Y^v_{est} (arriba); evolución de R_v y R^*_v (abajo).



En la Figura 4.10 se presentan los resultados obtenidos, a partir del modelo generado, para el conjunto de datos de testeo. En la parte superior de la figura se presenta la evolución temporal de las señales Y^t_{obs} y Y^t_{est}, mientras que en la parte inferior se presentan las señales de R_t y R^*_t.

Como se mencionó en la sección 3.2.2, son de interés únicamente los valo-

Figura 4.9: Validación cuantitativa del modelo construido.



res negativos de R_t^*, ya que la potencia generada de la turbina presenta una potencial anomalía en valores negativos de este residuo. Esto hace que el test de diferencia de medias sea aplicado solo a una cola.

En la Figura 4.11 se presentan las alarmas generadas por el test estadístico, junto con el residuo R_t^*. El método presentado en esta sección fue capaz de reconocer alarmas durante la producción del aerogenerador B. Las alarmas detectadas se encuentran todas antes de la falla producida en abril de 2014. En particular, se aprecia una gran concentración de alarmas durante la última etapa productiva del equipo previo a la falla. Lo anterior permite concluir que el método genera una respuesta satisfactoria en cuanto a la predicción de la falla ocurrida en abril de 2014. Asimismo, puede interpretarse, a partir de los resultados, que la falla está asociada a una degradación de la performance, dada la concentración de alarmas durante el semestre previo a la parada.

Figura 4.10: Resultados del modelo para los datos de testeo. Comparación entre Y^t_{obs} y $Y^{\hat{t}}_{obs}$ (arriba); evolución de R_t y R^*_t (abajo).

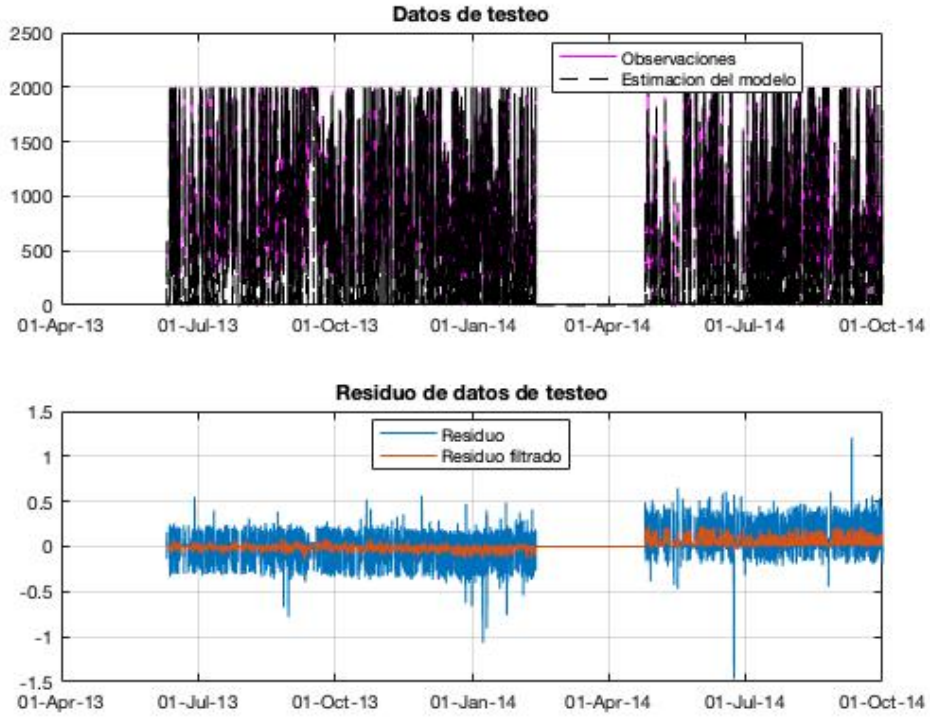
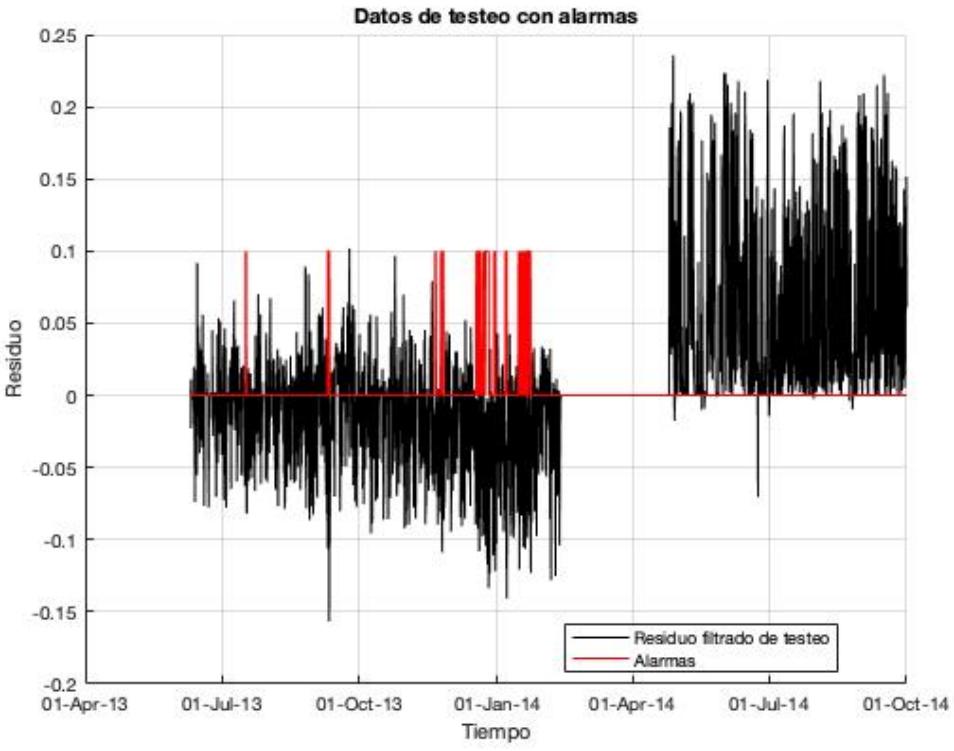


Figura 4.11: Evolución de R_t^* junto con las alarmas obtenidas.



4.2.3. Aplicación para el Aerogenerador C

NSET fue aplicado también para predecir la parada ocurrida durante la operación del aerogenerador C.

Debida a la naturaleza de la falla, se optó por tomar como variable objetivo una de las cargas de la pala dentro del conjunto de variables pre-seleccionadas en la sección 2.3; más precisamente, la carga de la pala C. Con el fin de determinar a su vez el conjunto de variables predictoras, en la Tabla 4.1 se presentan las correlaciones entre la variable objetivo a modelar y las demás variables pre-seleccionadas. A partir de esta información, junto con la que proporciona la

Tabla 4.1: Correlaciones entre la carga de la pala C y las demás variables pre-seleccionadas.

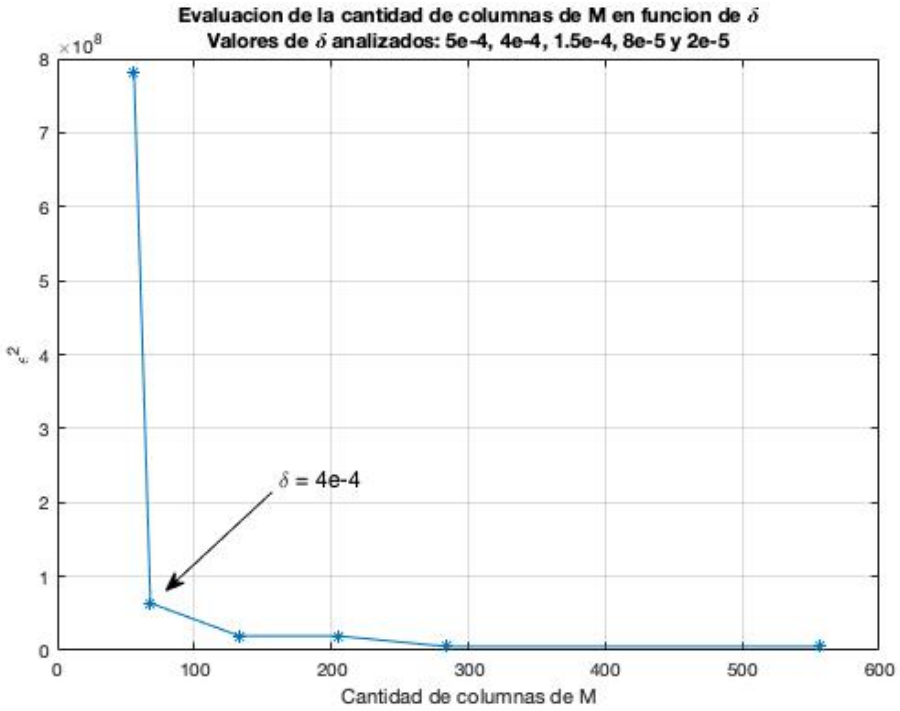
Carga de la pala C	1
Potencia	-0.83
Velocidad de viento	-0.86
RPM del rotor	-0.95
Carga de la pala A	0.99
Carga de la pala B	0.99

Figura 2.13, se observa en primer lugar que las variables asociadas a las otras dos cargas de las palas tienen un comportamiento esencialmente idéntico al de la variable objetivo, generando información redundante. Por lo tanto, se opta por seleccionar las demás variables como predictoras: potencia, velocidad de viento y RPM del rotor.

El conjunto de datos de entrenamiento para generar este modelo es el comprendido entre diciembre de 2012 y diciembre de 2013. Con la finalidad de determinar la Matriz de Memoria M, se procedió con un análisis de sensibilidad en función de δ. En la Figura 4.12 se presentan los resultados obtenidos para el error del modelo en función de la cantidad de estados m, correspondientes a cada valor de δ. Este análisis condujo a la decisión de optar por un valor de $\delta = 4 \times 10^{-4}$, correspondiendo este valor con una matriz M de 68 estados (frente a los 38577 presentes en el conjunto de datos de entrenamiento). Si bien puede observarse en la Figura 4.12 que para mayor valores de m, el error ϵ^2 disminuye, se verá en la etapa de validación que el valor seleccionado de δ conduce a un ajuste satisfactorio en la estimación del modelo.

El período correspondiente al conjunto de datos de validación está comprendido entre enero de 2014 y junio de 2014. En la figura 4.13 se presenta,

Figura 4.12: Análisis de sensibilidad en función de δ. Variación de ϵ^2 en función de m.



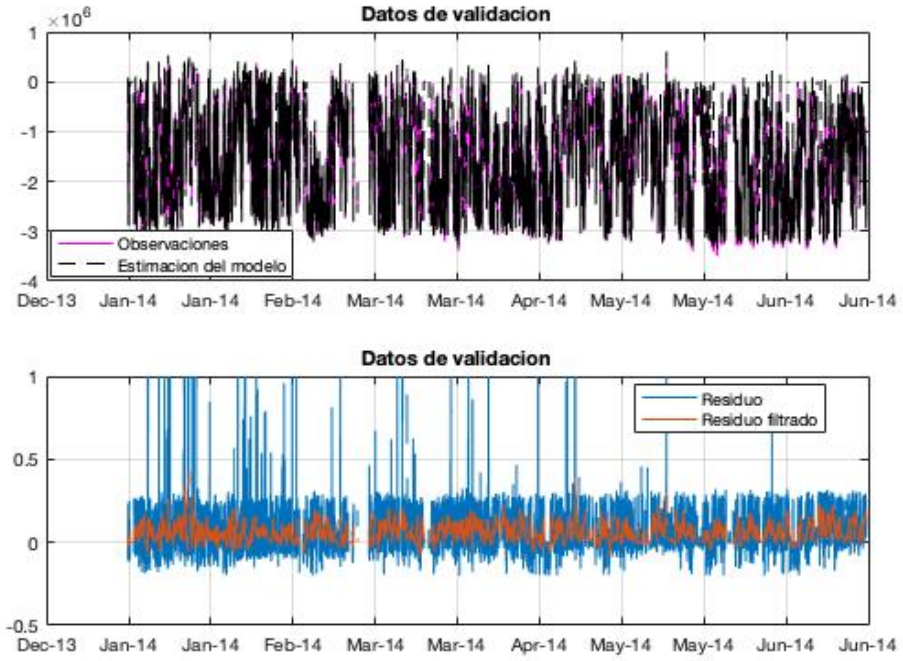
en el panel superior, la comparación entre las evoluciones temporales de Y_{obs} y Y_{est}, y en el panel inferior, las evoluciones temporales de R_v y R_v^*. Estas señales permiten visualizar los resultados obtenidos para el conjunto de datos de validación. Con el fin de cuantificar esta validación, en la Figura 4.14 se presenta un diagrama de dispersión entre Y_{obs}^v y Y_{est}^v, junto con la recta de regresión que mejor ajusta a los puntos. Puede apreciarse allí que esta recta de regresión ajusta satisfactoriamente a la recta identidad. Es por esto que se considera que el modelo no presenta sesgo apreciable en este caso, y por lo tanto, el ajuste de las observaciones es aceptable.

En la Figura 4.15 se presentan los resultados del modelo correspondientes al conjunto de datos de testeo, comprendidos entre julio de 2014 y julio de 2015. En el panel superior de la figura, se presenta la comparación entre las evoluciones temporales de Y_{obs}^t y Y_{est}^t; y en el panel inferior, la evolución temporal de las señales R_t y R_t^*.

Tanto los valores negativos como positivos de R_t^* son de interés para el análisis, por lo que el test de diferencia de medias es aplicado a dos colas.

En la Figura 4.16 se presentan los resultados finales para la aplicación de

Figura 4.13: Resultados del modelo para los datos de validación. Comparación entre las mediciones Y^v_{obs} y Y^v_{est} (arriba); evolución de R_v y R^*_v (abajo).



NSET al aerogenerador C. Allí figura la evolución temporal de R^*_t junto con las alarmas obtenidas a partir del test estadístico. El método fue capaz de reconocer alarmas en un período de funcionamiento previo a la fecha de la falla. Estas alarmas están localizadas en el mes previo a la falla. Asimismo, puede observarse una concentración de alarmas en octubre de 2014. El grupo de alarmas identificado previo a la parada, permite interpretar los resultados como una predicción de la falla. Finalmente, las alarmas de 2014 pueden interpretarse como falsas alarmas.

Los resultados anteriores permiten afirmar que NSET es capaz de predecir la falla ocurrida en el aerogenerador C.

Figura 4.14: Validación cuantitativa del modelo construido.

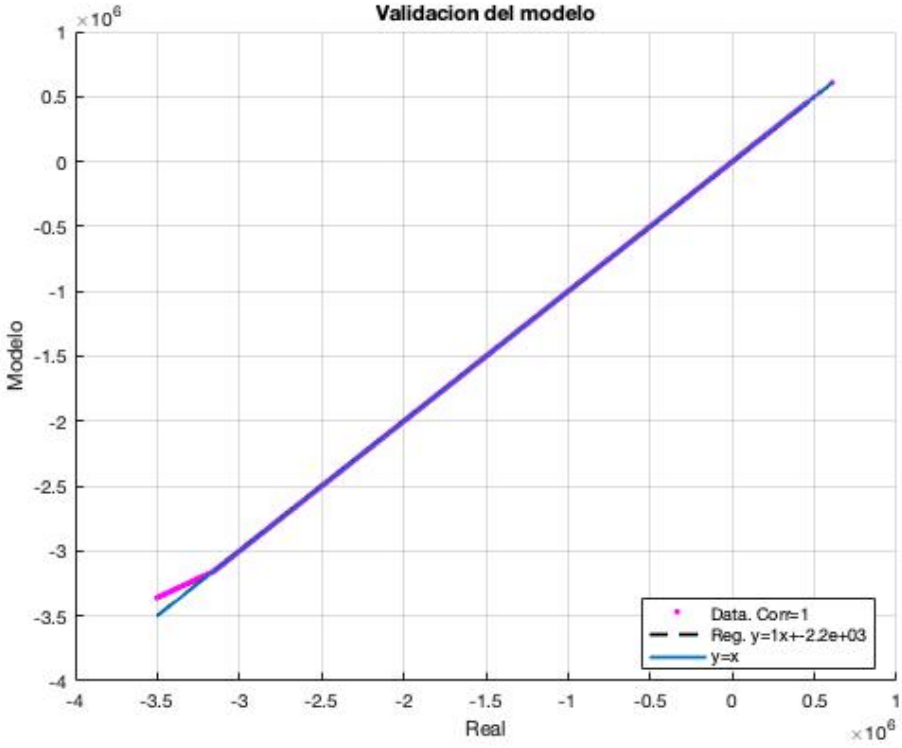


Figura 4.15: Resultados del modelo para los datos de testeo. Comparación entre Y^t_{obs} y $Y^{\hat{t}}_{obs}$ (arriba); evolución de R_t y R^*_t (abajo).

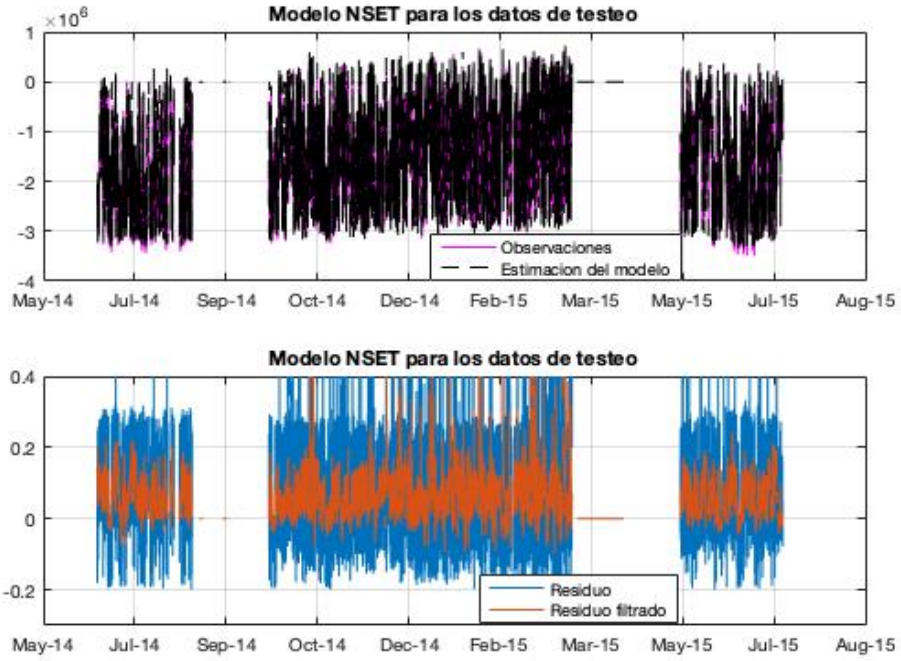
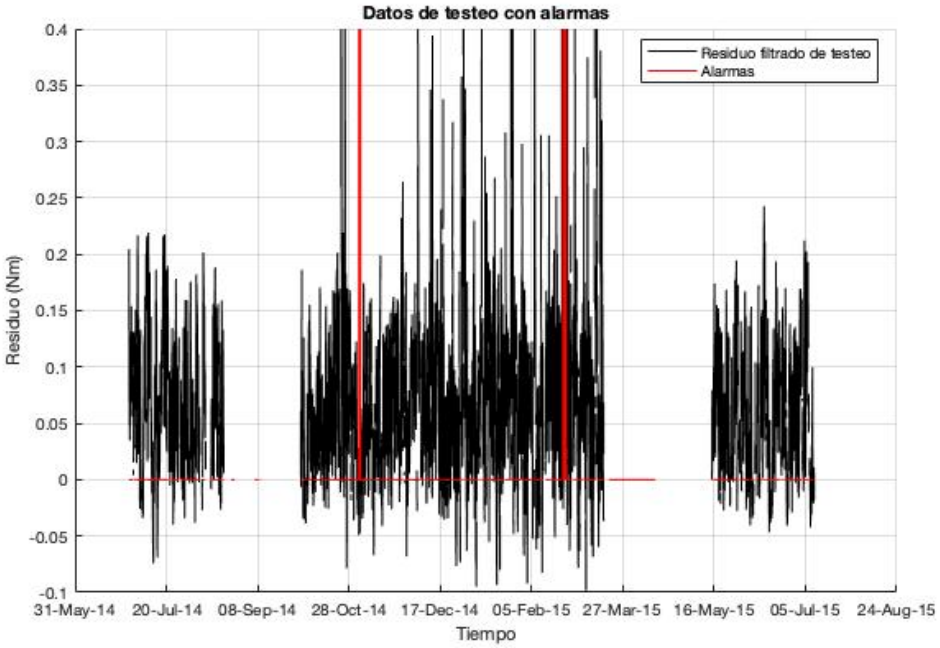


Figura 4.16: Evolución de R_t^* junto con las alarmas obtenidas.



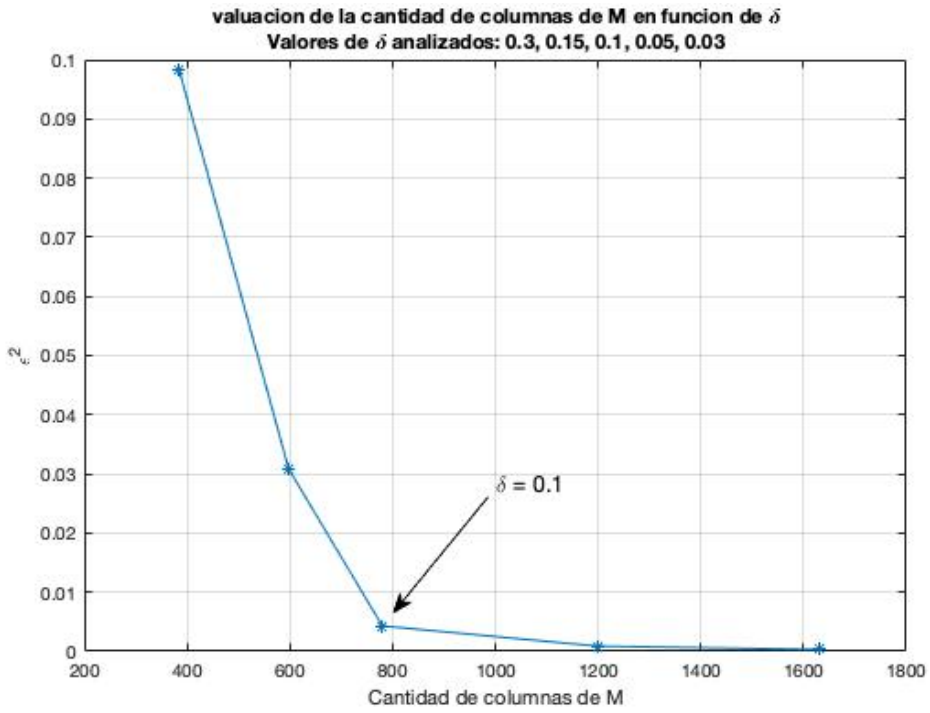
4.2.4. Aplicación para el Aerogenerador D

El método NSET fue aplicado para el conjunto de datos provenientes del aerogenerador D.

Al igual que en la sección 3.2.3, la variable a modelar fue la potencia, tomando como variables predictoras la velocidad de viento y las RPM del rotor. Asimismo, el período de entrenamiento está también comprendido entre enero de 2010 y diciembre de 2010.

Con el fin de determinar el tamaño de la Matriz de Memoria M asociada al modelo, se condujo a un análisis de sensibilidad entre el error del propio modelo y la cantidad de estados de M, realizado a partir del parámetro δ. En la Figura 4.17 se presentan los resultados obtenidos de este análisis para valores de δ iguales a 0.3, 0.15, 0.1, 0.05 y 0.03. La decisión tomada a partir de

Figura 4.17: Análisis de sensibilidad en función de δ. Variación de ϵ^2 en función de m.

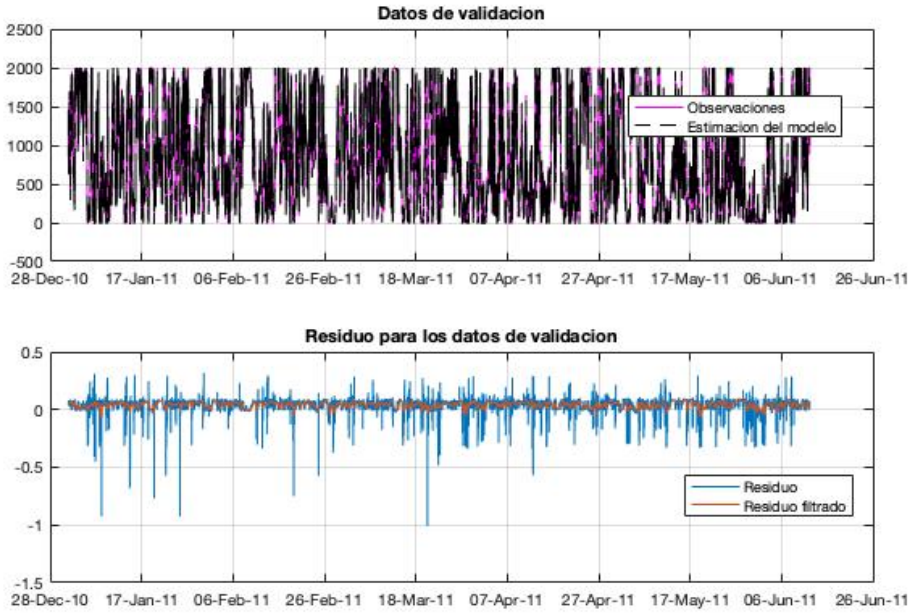


estos resultados fue de trabajar con un valor de δ igual a 0.1, ya que este genera un valor bajo de ϵ^2. Si bien para mayores valores de estados m, los valores del error del modelo disminuyen, se considera que la precisión del modelo no mejora en forma significativa. Esto debe ser verificado en la etapa de validación del modelo. Así, con el valor de δ seleccionado, la matriz M consta de 780 estados

correspondientes al conjunto de datos de entrenamiento, el que consta de 49969 estados en total.

Para validar el modelo, se realizaron las estimaciones correspondientes al conjunto de datos destinados para este fin. Este período es el mismo que el empleado en la sección 3.2.3 para la validación. En la Figura 4.18 se presenta, en el panel superior, la comparación entre las observaciones Y^v_{obs} y las estimaciones Y^v_{est}, y en el panel inferior, la evolución temporal de los residuos R_v y R^*_v. Esta validación debe ser cuantificada. En ese sentido, en la Figura 4.19 se

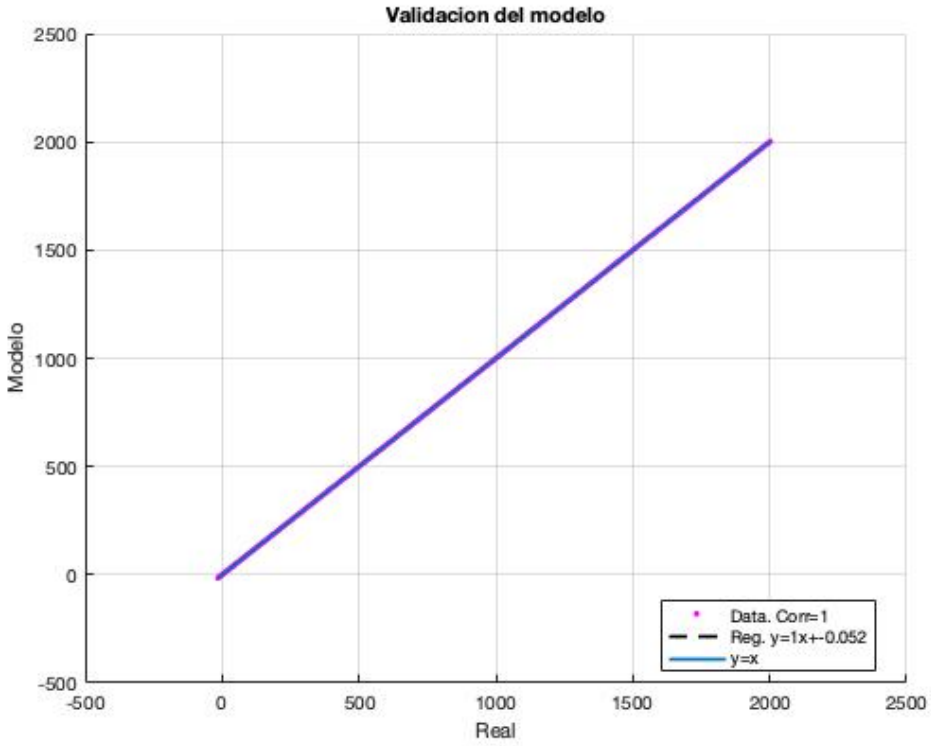
Figura 4.18: Resultados del modelo para los datos de validación. Comparación entre las mediciones Y^v_{obs} y Y^v_{est} (arriba); evolución de R_v y R^*_v (abajo).



presenta un diagrama de dispersión entre las observaciones Y^v_{obs} y las estimaciones del modelo generado Y^v_{est}, junto con la recta de regresión que ajusta a estos puntos. Puede apreciarse que la recta de ajuste coincide en todo el rango de operación con la recta identidad, traduciéndose esto en un buen ajuste del modelo para las observaciones.

La cuantificación de la validación del modelo permite avanzar a la etapa de testeo, donde el período de datos correspondiente a esta finalidad es el descrito en la sección 3.2.3. En la Figura 4.20 se presenta, en el panel superior, la comparación entre las observaciones Y^t_{obs} y las estimaciones del modelo Y^t_{est}, y en el panel inferior, las evoluciones temporales de R_t y R^*_t.

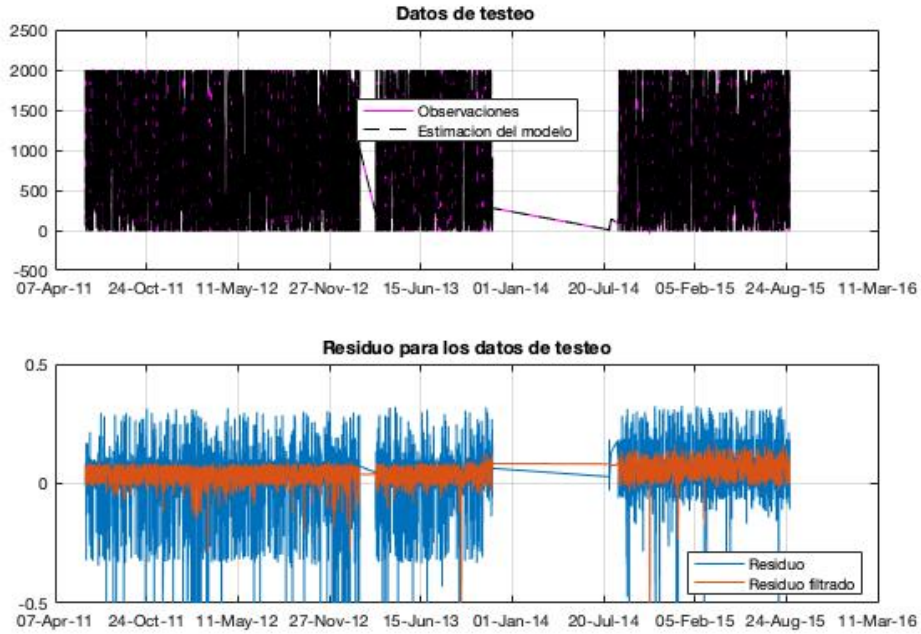
Figura 4.19: Validación cuantitativa del modelo construido.



Como se mencionó en secciones precedentes, no se tiene información sobre alguna falla ocurrida durante el funcionamiento del aerogenerador D. Esto hace que, análogamente a lo analizado en la sección 3.2.3, los resultados de NSET no sean empleados para detectar alarmas asociadas a anomalías en el funcionamiento. Sin embargo, la señal de R_t^* presentada en la Figura 4.20 es empleada con la finalidad de comparar el funcionamiento del equipo entre los períodos pre-parada y pos-parada. En este sentido, el test estadístico de comparación de medias desarrollado en la sección 3.1.1 es aplicado para comparar los datos de R_t^* antes y después de la parada de noviembre de 2013; tal como se implementó en la sección 3.2.3. Se destaca que los valores de las medias correspondientes a R_t^* antes y después de la parada valen $0.037kW$ y $0.063kW$, respectivamente.

Los resultados de este test estadístico, empleando un nivel de significancia del 5 % y una cantidad de grados de libertad igual a 963, muestran que hay suficiente evidencia como para rechazar la hipótesis nula de que los valores medios de R_t^* antes y después de la parada son iguales. El valor de t, calculado de acuerdo a la Ecuación 3.12, es de 14.87.Lo anterior se traduce en

Figura 4.20: Resultados del modelo para los datos de testeo. Comparación entre Y_{obs}^{t} y $\hat{Y_{obs}^{t}}$ (arriba); evolución de R_t y R_t^* (abajo).



que el desempeño del aerogenerador D vio una mejora luego de la parada de mantenimiento. Esta conclusión será abordada nuevamente en el capítulo 7.

Se concluye entonces que NSET fue aplicado para generar un monitoreo por condición, a partir de la señal de R_t^*, del aerogenerador D, pudiendo inferir además que el período correspondiente a la pos-parada de mantenimiento presentó una mejora en la performance del equipo.

Capítulo 5

Cópulas

En este capítulo se presenta el desarrollo de una técnica para evaluar el estado de condición de un aerogenerador a través de su curva de potencia. Esta herramienta ha sido explorada con estas finalidades en distintos trabajos. En este sentido, Gill et al. (2012) y Stephen et al. (2011) desarrollaron en sus respectivas publicaciones la técnica de Cópulas para describir la performance del aerogenerador. A su vez, Wang et al. (2014) presentaron un modelo basado en Cópulas para eliminar outliers de la curva de potencia.

Se presenta a continuación el desarrollo de esta herramienta, junto con la aplicación de la misma para los datos de tres de los cuatro aerogeneradores presentados en la sección 2; por motivos que se expondrán en la sección 7, la aplicación del método al caso del aerogenerador B no está contenida en este capítulo.

5.1. Descripción teórica

Es necesario enmarcar teóricamente las Cópulas para luego poder realizar la aplicación correspondiente. Esta contextualización está basada fundamentalmente en los contenidos de Nelsen (2006). En este sentido, deben realizarse algunas consideraciones preliminares.

El objetivo de esta sección es presentar las herramientas que permitan conducir al Teorema de Sklar, el que dará el principal resultado empleado en las secciones subsiguientes.

Sean S_1 y S_2 dos subconjuntos no vacíos de $\mathbb{R}$, y sea H una función de dos variables, tal que $dom(H) = S_1 \times S_2$. Sea $B = [x_1, x_2] \times [y_1, y_2]$ un rectángulo

cuyos vértices están contenidos en $dom(H)$. Luego, el H-volumen de B queda definido por:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \tag{5.1}$$

Si se definen las diferencias de primer orden de H sobre el rectángulo B como:

$$\Delta_{x_1}^{x_2} H(x, y) = H(x_2, y) - H(x_1, y) \tag{5.2}$$

$$\Delta_{y_1}^{y_2} H(x, y) = H(x, y_2) - H(x, y_1) \tag{5.3}$$

Entonces el H-volumen del rectángulo B es la diferencia de segundo orden de H sobre B:

$$V_H(B) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y) \tag{5.4}$$

Una función real de dos variables H se dice *2-creciente* si $V_H(B) \geq 0$ para todos los rectángulos B cuyos vértices pertenecen a $dom(H)$.

Para ejemplificar el concepto anterior, sea H una función definida sobre $I^2 = [0, 1] \times [0, 1]$ tal que $H(x, y) = (2x - 1)(2y - 1)$. Luego, H es 2-creciente. Cabe notar que también se trata de una función decreciente en x para cada $y \in (0, 1/2)$, y de una función decreciente en y para cada $x \in (0, 1/2)$.

Lema: Sean S_1 y S_2 dos sunconjuntos no vacíos de $\mathbb{R}$, y sea H una función 2-creciente con dominio $S_1 \times S_2$. Sean $x_1, x_2 \in S_1$, con $x_1 \leq x_2$, y sean $y_1, y_2 \in S_2$, con $y_1 \leq y_2$. Luego, la función $t \mapsto H(t, y_2) - H(t, y_1)$ es no-decreciente en S_1, y la función $t \mapsto H(x_2, t) - H(x_1, t)$ es no-decreciente en S_2.

Como una aplicación inmediata del lema anterior, se puede mostrar que con una hipótesis adicional, una función H 2-creciente, es no-decreciente en cada uno de sus argumentos. Supongamos que S_1 tiene al menos un elemento a_1, y que S_2 tiene al menos un elemento a_2. Decimos que la función H, con $dom(H) = S_1 \times S_2$ y $rec(H) = \mathbb{R}$ es *asentada* si $H(x, a_2) = 0 = H(a_1, y), \forall (x, y) \in S_1 \times S_2$. Tenemos entonces el siguiente lema.

Lema: Sean S_1 y S_2 dos subconjuntos no vacíos de $\mathbb{R}$, y sea H una función asentada 2-creciente con $dom(H) = S_1 \times S_2$. Luego, H es no-decreciente en cada uno de sus argumentos.

Supongamos ahora que S_1 y S_2 tienen, cada uno, un elemento más grande, b_1 y b_2 (que eventualmente podrán ser $+\infty$), respectivamente. Podemos decir que $H : S_1 \times S_2 \to \mathbb{R}$ *tiene marginales*, y que los marginales de H son las

funciones F y G, definidas por:

$$dom(F) = S_1 \quad F(x) = H(x, b_2), \forall x \in S_1 \tag{5.5}$$

$$dom(G) = S_2 \quad G(y) = H(b_1, y), \forall y \in S_2 \tag{5.6}$$

Para ejemplificar lo anterior, sea $H : [-1, 1] \times [0, +\infty) \to \mathbb{R}$ tal que

$$H(x, y) = \frac{(x + 1)(e^y - 1)}{x + 2e^y - 1}$$

Luego, H es asentada porque $H(x, 0) = 0$ y $H(-1, y) = 0$, y H tiene marginales $F(x)$ y $G(y)$ dados por $F(x) = H(x, \infty) = \frac{x+1}{2}$ y $G(y) = H(1, y) = 1 - e^{-y}$.

Tenemos entonces el siguiente lema para funciones asentada, 2-creciente, y con marginales.

Lema: Sean S_1 y S_2 dos subconjuntos no vacíos de $\mathbb{R}$, y sea H una función asentada, 2-creciente, y con marginales, con $dom(H) = S_2 \times S_2$. Sean (x_1, y_2) y (x_2, y_2) dos puntos cualquiera en $S_1 \times S_2$. Tenemos entonces que:

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_2)| + |G(y_2) - G(y_1)| \tag{5.7}$$

Se tiene que una sub-Cópula bidimensional es una función C' con las siguientes propiedades.

1. $dom(C') = S_1 \times S_2$, donde S_1 y S_2 son subconjuntos de $I = [0, 1]$ conteniendo el 0 y el 1.
2. C' es asentada y 2-creciente.
3. Para todo $u \in S_1$ y para todo $v \in S_2$, $C'(u, 1) = u$ y $C'(1, v) = v$.

Cabe notar que para todo $(u, v) \in dom(C')$, $0 \leq C'(u, v) \leq 1$, por lo que $rec(C') = I$.

Una Cópula bidimensional es una sub-Cópula bidimensional C cuyo dominio es I^2. Dicho de otro modo, una Cópula es una función $C : I^2 \to I$ con las siguientes propiedades.

1. Para todo $u, v \in I$, $C(u, 0) = 0 = C(0, v)$, $C(u, 1) = u$ y $C(1, v) = v$.
2. Para todo $u_1, u_2, v_1, v_2 \in I$ tales que $u_1 \leq u_2$ y $v_1 \leq v_2$, $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$

Como $C(u, v) = V_C([0, u] \times [0, v])$, puede interpretarse a $C(u, v)$ como una asignación de un número en I al rectángulo $[0, u] \times [0, v]$.

Una función de distribución es una función F con $dom(F) = \mathbb{R}$ tal que:

- F es no-decreciente.
- $F(-\infty) = 0$ y $F(\infty) = 1$.

Una función de distribución conjunta es una función H con $dom(H) = \mathbb{R}^2$ tal que:

- H es 2-creciente.
- $H(x, -\infty) = H(-\infty, y) = 0$, y $H(\infty, \infty) = 1$.

Por lo tanto, H es asentada, y como $dom(H) = \mathbb{R}^2$, H tiene marginales F y G dados por $F(x) = H(x, \infty)$ y $G(y) = G(\infty, y)$.

Interesa finalmente formular el Teorema de Sklar (Sklar (1959)).

Teorema: Sea H una función de distribución conjunta con marginales F y G. Entonces existe una Cópula C tal que, para todo $x, y \in \mathbb{R}$,

$$H(x, y) = C(F(x), G(y)) \tag{5.8}$$

Si F y G son continuas, entonces C es única. Recíprocamente, si C es una Cópula, y F y G son funciones de distribución, entonces $H(x, y) = C(F(x), G(y))$ es una función de distribución conjunta con marginales F y G.

El Teorema de Sklar proporciona entonces las condiciones necesarias para asegurar la existencia y unicidad de la Cópula C. Este último resultado es de relevancia para fundamentar la metodología descrita en la sección 5.1.1.

5.1.1. Metodología de aplicación

El resultado obtenido del Teorema de Sklar hace que las Cópulas sean una herramienta que facilitan la descripción de estructuras de dependencia compleja, así como también permite obtener una relación con las distribuciones marginales dentro de una sola función.

Dadas dos variables aleatorias y continuas, X e Y, la función de distribución conjunta $H(x, y)$ queda definida por $H(x, y) = p(X \leq x, Y \leq y)$. Las distribuciones marginales acumuladas de H están dadas por $F(x) = p(X \leq x, Y \leq \infty)$ y $G(y) = p(X \leq \infty, Y \leq y)$.

Estas distribuciones marginales acumuladas pueden ser usadas para transformar las variables aleatorias originales, X e Y, en nuevas variables U y V

con densidades marginales uniformes en I.

$$u = F(x) \tag{5.9}$$

$$v = G(y) \tag{5.10}$$

En estas condiciones, para cualquier distribución continua bivariada, el Teorema de Sklar asegura la existencia de una Cópula C que verifica la Ecuación 5.8, que puede reescribirse de acuerdo a la siguiente ecuación.

$$C(u, v) = H(x, y) \tag{5.11}$$

Por lo tanto, se tiene que $H(x, y) = C(F(x), G(y)) = p(U \leq u, V \leq v)$. Invirtiendo las Ecuaciones 5.9 y 5.10, tenemos que

$$C(u, v) = H\left(F^{-1}(u), G^{-1}(v)\right) \tag{5.12}$$

La Ecuación 5.12 explicita la utilidad de las Cópulas en el sentido de que permite que las distribuciones marginales y la estructura de dependencia se especifiquen por separado. La estimación de las Cópulas se puede realizar ajustando familias de Cópulas parametrizadas a los datos o, alternativamente, mediante una serie de técnicas no-paramétricas como la estimación de las funciones de densidad.

La metodología a aplicar consiste en un método no-paramétrico para la estimación de la Cópula. Consiste en estimar las distribuciones marginales acumuladas y la distribución bivariada a partir de un gran conjunto de datos de referencia.

En este sentido, el método presentado en este trabajo consiste en estimar la Cópula para representar la relación existente entre los datos de velocidad de viento y de potencia de un aerogenerador; la curva de potencia. Para esto, se considera la curva de potencia como una distribución conjunta bivariada. Es necesario realizar una correcta estimación de las distribuciones marginales acumuladas de la velocidad de viento y la potencia, para garantizar que las variables transformadas tengan una distribución uniforme.

Este método requiere dividir el total del conjunto de datos en dos subconjuntos: uno de *entrenamiento* y otro de *testeo*.

La estimación de la Cópula se realiza a partir del conjunto de datos de

entrenamiento, que se elige bajo la hipótesis de que son datos saludables de funcionamiento del aerogenerador. Estos se transforman en el espacio de la Cópula a partir de las distribuciones marginales acumuladas de la velocidad de viento y la potencia. I^2 es sub-dividido en una grilla de 100×100 elementos. En cada una de los 10000 elementos de la grilla, se realiza un conteo de la cantidad de puntos de los datos de entrenamiento que caen allí, para estimar la densidad de la Cópula en esa región. Cabe mencionar que existen otros estimadores más sofisticados para estimar densidades que permitirían también determinar la densidad de la Cópula en esas regiones.

Entonces, sean w_e y p_e las series de velocidad de viento y potencia, respectivamente, asociadas al conjunto de datos de entrenamiento. A partir de estas, se estiman las distribuciones marginales acumuladas de cada una, F y G. Así, los datos de entrenamiento son transformados a las variables de las Cópulas de acuerdo a las siguientes ecuaciones.

$$u_e = F(w_e) \tag{5.13}$$

$$v_e = G(p_e) \tag{5.14}$$

De esta forma queda determinada la Cópula para puntos de operación saludables. La discretización antes mencionada permite determinar la densidad de esta Cópula estimada, para que luego sea comparada con la del conjunto de datos de testeo.

La etapa de testeo consiste en tomar las nuevas series de velocidad de viento y potencia asociadas a este conjunto de datos, w_t y p_t. A partir de las funciones F y G estimadas, se transforman estas nuevas señales al espacio de la Cópula. En la discretización de I^2, se busca finalmente comparar la densidad del conjunto de datos de entrenamiento con la generada por el conjunto de datos de testeo.

Finalmente, se desarrollan algunos indicadores para comparar las densidades generadas por ambos conjuntos de datos en el espacio de la Cópula, de forma de cuantificar el estado de condición de la turbina. Como los datos del conjunto de testeo deberían estar, con cierta dispersión, sobre la línea $v = u$ del espacio de la Cópula, dos indicadores directos son el promedio de los residuos y del cuadrado de los residuos, en relación con $v = u$.

$$R = \frac{\sum_{i=1}^{n} (v_i - \bar{v}_i)}{n} = \frac{\sum_{i=1}^{n} (v_i - u_i)}{n} \tag{5.15}$$

$$Q^2 = \frac{\sum_{i=1}^{n} (v_i - \bar{v_i})^2}{n^2} = \frac{\sum_{i=1}^{n} (v_i - u_i)^2}{n^2} \tag{5.16}$$

Donde u_i y v_i son los puntos de testeo en el espacio de la Cópula, $\bar{v_i}$ es el valor esperado de v_i en un funcionamiento saludable, que es igual a u_i, ya que es la proyección vertical de v_i en la recta $v = u$, y n es la cantidad de puntos de testeo.

Luego, se define un tercer indicador.

$$\chi^2 = \frac{\sum_{u_i=1}^{100} \sum_{v_j=1}^{100} \left(N_{u_i,v_j} - \bar{N}_{u_i,v_j}\right)^2}{n^2} \tag{5.17}$$

Donde N_{u_i,v_j} y $\bar{N}_{u_i,v_j}$ son la cantidad de puntos observados y la cantidad de puntos esperados, para el conjunto de datos de testeo, en la sub-división (u_i, v_j) de la grilla. $\bar{N}_{u_i,v_j}$ se obtiene como la estimación de la densidad de la Cópula obtenida en el entrenamiento para esa región de la grilla. Cabe destacar que el indicador definido en la Ecuación 5.17 está inspirado en la distribución Chi-cuadrado, aunque no necesariamente tiene esa distribución estadística.

Estos tres estimadores son calculados a lo largo del tiempo durante el funcionamiento del aerogenerador, de forma de obtener una cuantificación de la condición de estado del mismo.

Cabe destacar que los valores R, Q^2 y χ^2 tienen interpretaciones diferentes. R permanecerá cerca de 0 si la distribución es simétrica con respecto a $v = u$ y no detectará cambios en la variabilidad de los datos; por otra parte, valores negativos de R darán un indicio de decaimiento en la performance de la turbina, detectando conjuntos de datos que se desvían de esa recta. Por otra parte, Q^2 permite la detección de cambios en la variabilidad de los datos mientras se mantiene la simetría alrededor de $v = u$. Así, valores altos de Q^2 corresponderán a períodos potencialmente anómalos. Finalmente, χ^2 indicará qué tan apropiado es el modelo obtenido en la fase de entrenamiento para el nuevo conjunto de datos; valores altos de χ^2 corresponderán a un desajuste del modelo, pudiendo ser esto una eventual anomalía.

5.2. Resultados

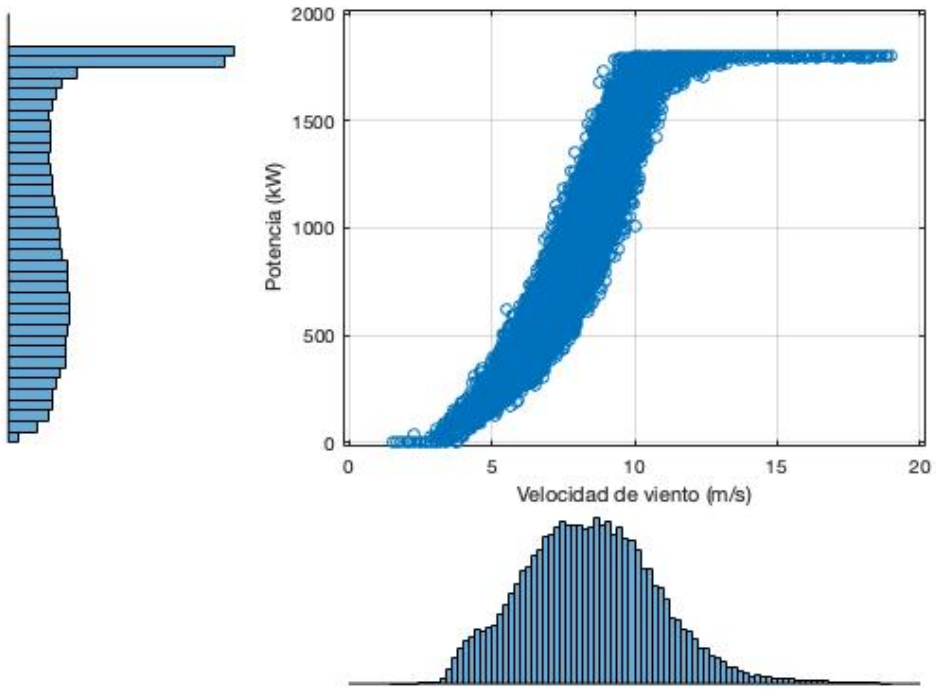
5.2.1. Aplicación para el aerogenerador A

La metodología descrita en la sección 5.1.1 fue aplicada al conjunto de datos asociado al aerogenerador A con el fin de aplicar el método de monitoreo por condición, para reconocer anticipadamente las fallas asociadas al funcionamiento de este equipo.

En este sentido, los datos del aerogenerador A_0 son considerados en esta sección, por los motivos ya fundamentados en la sección 3.2.1, como datos saludables de operación del aerogenerador A. Con este conjunto de datos se realiza la estimación de la Cópula asociada a A.

Con el fin de estimar las funciones F y G, en la Figura 5.1 se presenta el diagrama de dispersión de la curva de potencia del aerogenerador A_0, junto con las distribuciones marginales de w_e y p_e. A partir de estas distribuciones,

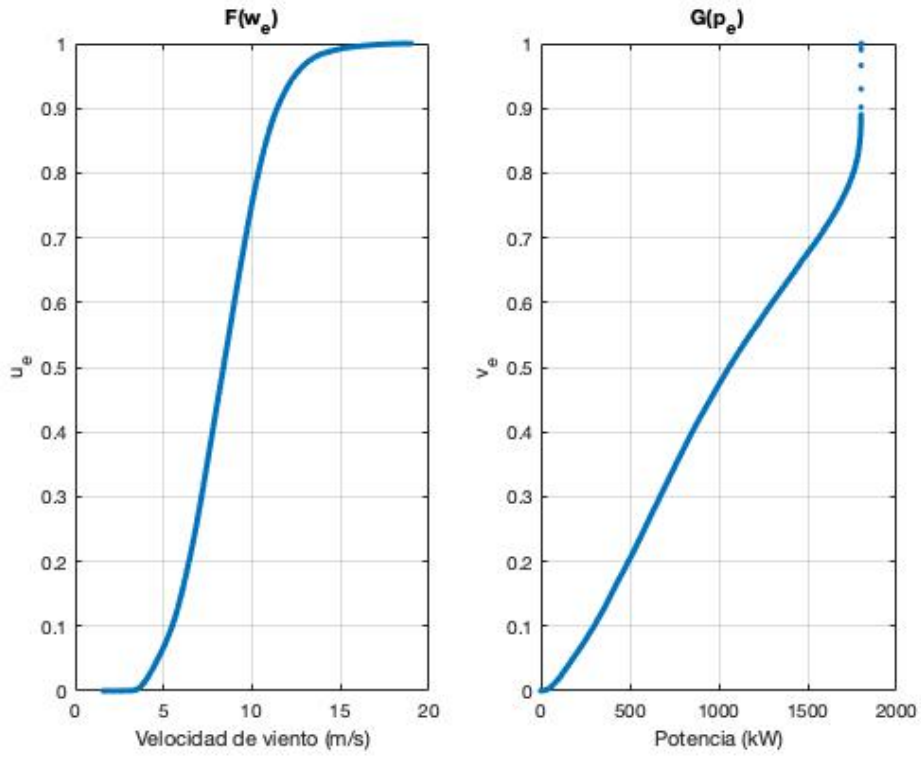
Figura 5.1: Diagrama de dispersión de la curva de potencia, junto con las distribuciones marginales de w_e y p_e, de los datos del aerogenerador A_0.



es posible determinar las distribuciones marginales acumuladas de w_e y p_e, F y G, respectivamente, que se presentan en la Figura 5.2. Estas funciones son

las que permiten definir las nuevas variables u_e y v_e para pasar al espacio de la Cópula.

Figura 5.2: Distribución marginal acumulada de w_e, F (izquierda); distribución marginal acumulada de p_e, G (derecha).

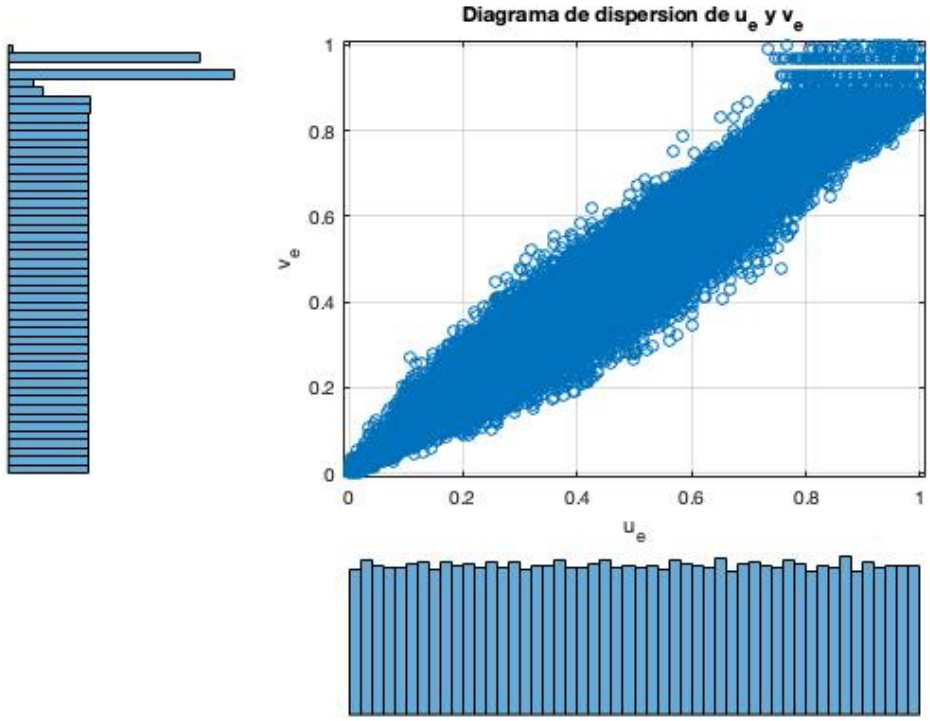


En este sentido, en la Figura 5.3 se presenta el diagrama de dispersión del las variables u_e y v_e, generando así la Cópula para el conjunto de datos de entrenamiento. A su vez, en esta figura se generaron también las distribuciones de estas nuevas variables. Como era de esperar, estas distribuciones son prácticamente uniformes en I. Para el caso de v_e, más precisamente para valores cercanos a 1, hay una cierta dispersión, debida a la menor correlación entre la potencia y la velocidad de viento para potencias próximas a la nominal.

La información representada en la Figura 5.3 es empleada, a través de la sub-división descrita en la sección 5.1.1, con el fin de determinar la densidad de la Cópula para los datos de entrenamiento. En este sentido, en la Figura 5.4 se presenta la Cópula obtenida para los datos del aerogenerador A_0.

En esta figura puede observarse cómo los datos se concentran, con cierta dispersión, alrededor de la recta $v = u$, tal como estaba previsto. A su vez, la dispersión observada en la Figura 5.3, se refleja aquí en el entorno del punto

Figura 5.3: Diagrama de dispersión para u_e y v_e, junto con las distribuciones de probabilidad correspondientes a estas variables.



$(1, 1)$. Estos resultados expuestos, obtenidos de un funcionamiento saludable, son los que se utilizan para comparar con el funcionamiento del aerogenerador A.

De esta forma, en la Figura 5.5 se presentan los indicadores resultantes de comparar la Cópula presentada con las respectivas Cópulas del aerogenerador A, obtenidas a lo largo de su funcionamiento. Estos indicadores fueron calculados de acuerdo a las Ecuaciones 5.15, 5.16 y 5.17. Su evolución temporal permite realizar un monitoreo por condición del aerogenerador A.

Los valores de R mostrados en la Figura 5.5 presentan una caída previo a la parada de abril de 2016. El decrecimiento en este indicador también es notorio en los primeros meses posteriores a esta parada, dando un indicio de que la turbina aún presentaba algún desperfecto en su funcionamiento en ese entonces. De esta forma, la parada de abril es anticipada a través de este indicador. Asimismo, los valores más bajos de R durante su evolución se dan en el mes de agosto de 2016, previo a la parada de setiembre de ese año. Consecuentemente, esta parada también es anticipada por este método. A modo ilustrativo, en la

Figura 5.4: Cópula para los datos del aerogenerador A_0.

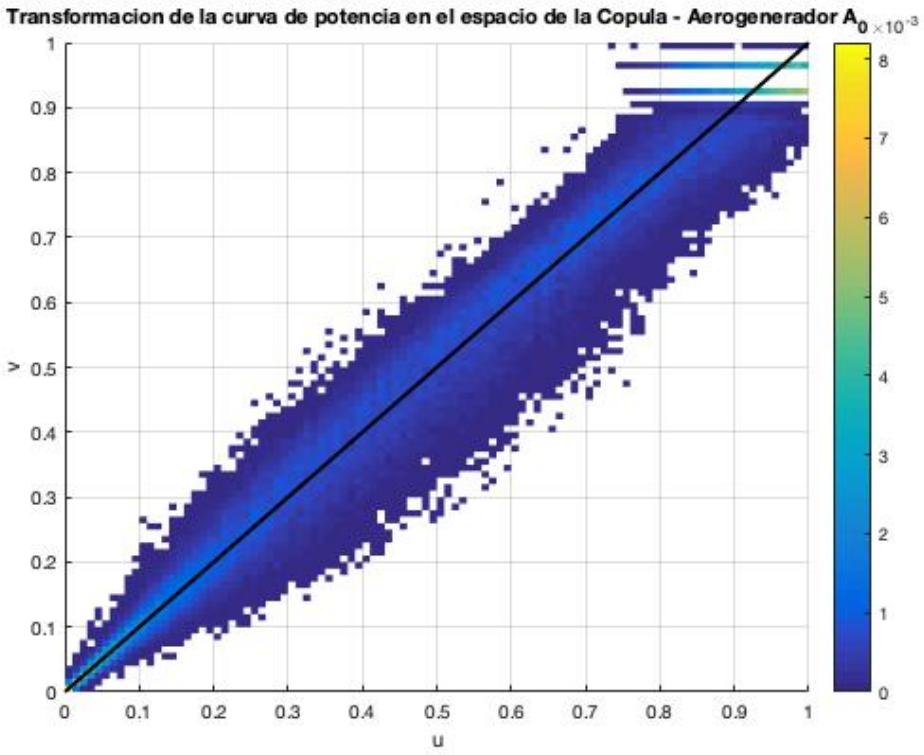


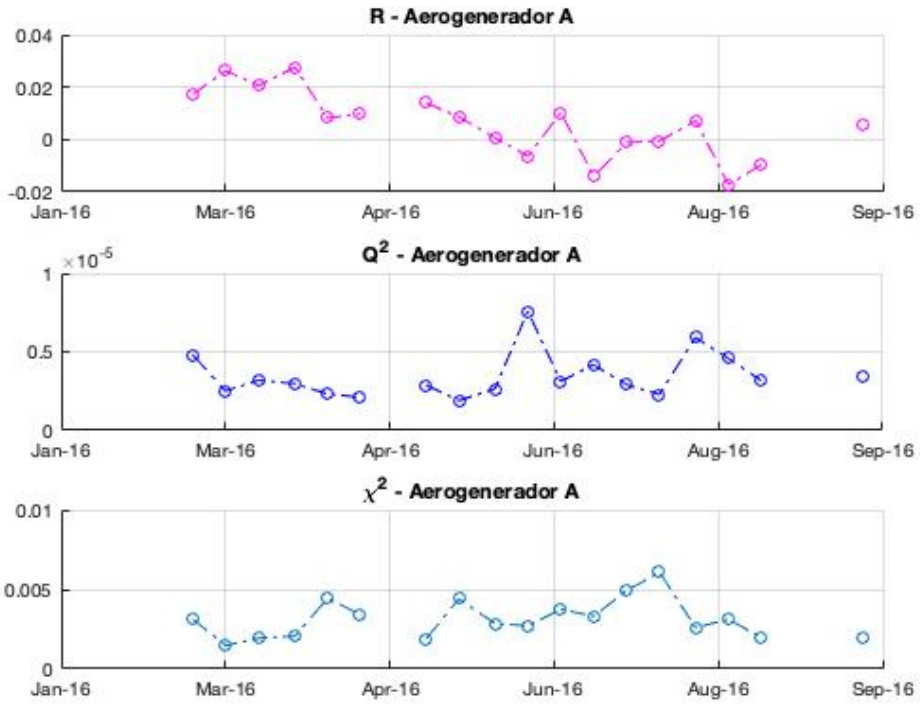
Figura 5.6 se presenta la Cópula generada por los datos de la segunda decena de agosto de 2016. Allí se observa cómo la densidad pierde la simetría respecto a la recta $v = u$, detectando una dispersión mayor sobre la región inferior a esta recta; traduciéndose esto en una caída de la performance del equipo.

Por otra parte, los valores de Q^2 presentan algunos picos durante su evolución, más precisamente en los meses de junio y de agosto, previo a la parada de setiembre de 2016. Estos altos valores de este indicador se traducen en un aumento en la dispersión de los datos alrededor de la recta $v = u$, significando esto una inestabilidad en la respuesta de la turbina. Q^2 no presenta anomalías en su evolución previo a la parada de abril.

Finalmente, los valores de χ^2 son estables durante toda la evolución, aunque se registran dos picos en los meses de julio y agosto de 2016, pudiendo esto ser un indicio de que el modelo de la Cópula generada no se adapta adecuadamente durante esos períodos. Este es un indicio de que la turbina presenta un comportamiento inusual en estos períodos, que eventualmente es síntoma de una falla inminente.

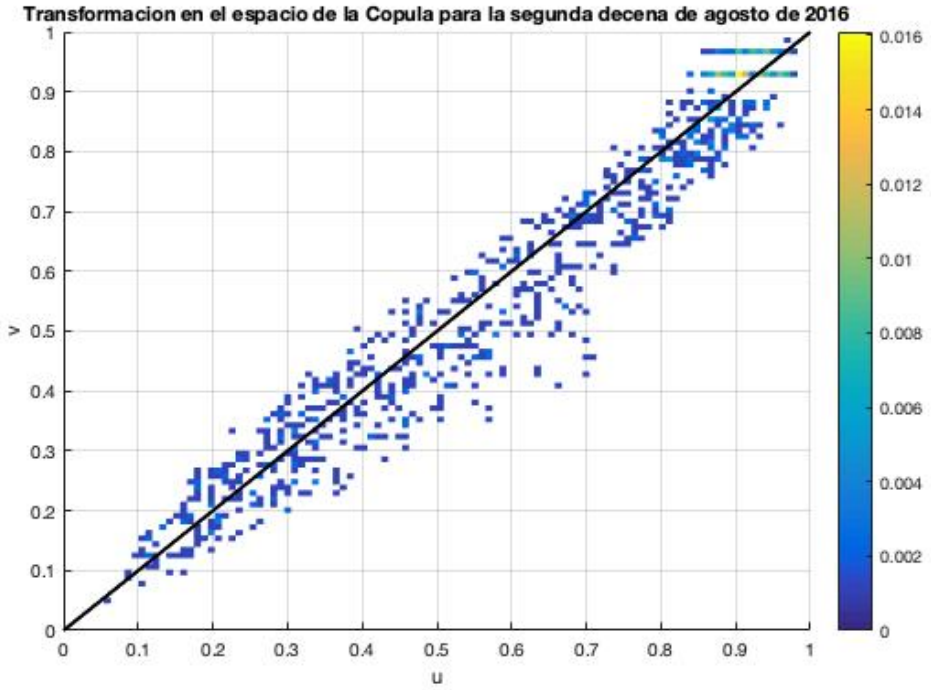
De esta forma, los indicadores R, Q^2 y χ^2 permiten anticipar, a través

Figura 5.5: Indicadores comparativos: R (arriba), Q^2 (medio) y χ^2 (abajo).



de un monitoreo por condición del aerogenerador A, las dos fallas conocidas durante su producción. La parada de abril es anticipada principalmente por R, mientras que la parada de setiembre es predicha por anomalías de R y Q^2.

Figura 5.6: Cópula para el aerogenerador A durante la segunda decena de agosto de 2016.



5.2.2. Aplicación para el Aerogenerador C

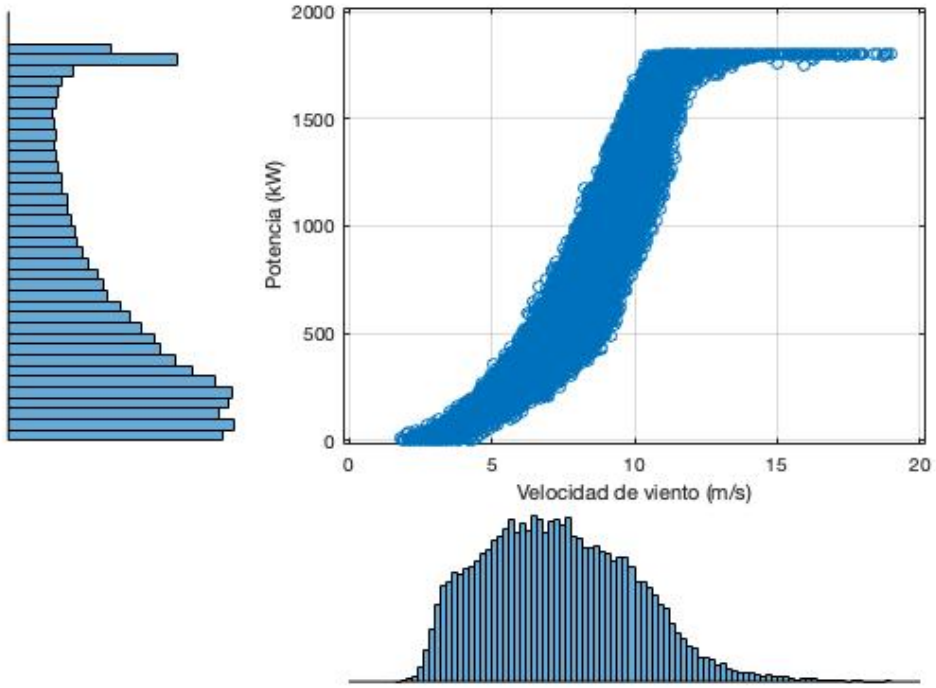
El método desarrollado fue aplicado también para los datos disponibles correspondientes al aerogenerador C.

El período de entrenamiento empleado para la construcción del modelo está comprendido entre diciembre de 2012 y junio de 2014. A partir de este conjunto, se realiza la estimación de la Cópula asociada a C.

En la Figura 5.7 se presenta un diagrama de dispersión de la curva de potencia asociada al conjunto de datos de entrenamiento, junto con las respectivas distribuciones marginales de w_e y p_e. A partir de esta información, es posible estimar las funciones F y G correspondientes; las distribuciones marginales acumuladas de w_e y p_e, respectivamente. Estas se presentan en la Figura 5.8. Las estimaciones de F y G permiten definir las nuevas variables u_e y v_e para pasar al espacio de la Cópula.

En este sentido, en la Figura 5.9 se presenta el diagrama de dispersión de las variables u_e y v_e, junto con las respectivas distribuciones de probabilidad. Las distribuciones asociadas a las variables u_e y v_e se asemejan, como era de esperar, a la distribución uniforme en I.

Figura 5.7: Diagrama de dispersión de la curva de potencia, junto con las distribuciones marginales de w_e y p_e, de los datos de entrenamiento del aerogenerador C.

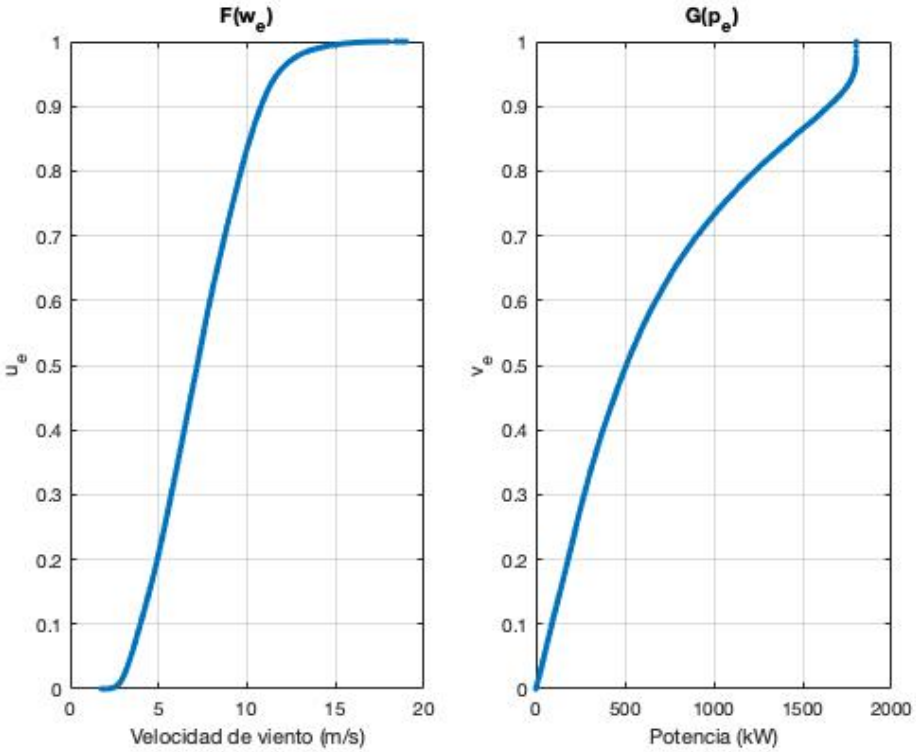


La información presentada permite entonces obtener, a partir de la subdivisión introducida en la sección 5.1.1, la densidad de la Cópula para el conjunto de datos de entrenamiento. En este sentido, en la Figura 5.10 se presenta la densidad de la Cópula para estos datos. Allí puede verse que los puntos de la Cópula se concentran alrededor de la recta $v = u$. Esta densidad será empleada en lo que sigue para comparar puntos de funcionamiento saludable con los puntos de operación asociados al conjunto de datos de testeo, con el fin de identificar con antelación la falla asociada al aerogenerador C.

El período de testeo está comprendido entre julio de 2014 y julio de 2015. En este sentido, en la Figura 5.11 se presenta la evolución temporal de cada uno de los indicadores R, Q^2 y χ^2 introducidos en la sección 5.1.1.

Los valores de R mostrados en el primer panel de la Figura 5.11 presentan un decaimiento en la segunda decena de días de marzo de 2015, justo antes de la parada asociada a la falla registrada en el equipo. Esto permite asociar esta disminución de R con una predicción, de hasta 10 días, del problema asociado

Figura 5.8: Distribución marginal acumulada de w_e, F (izquierda); distribución marginal acumulada de p_e, G (derecha).



a la parada.

Luego, durante la evolución de Q^2 se presentan algunos valores elevados. En particular, uno de ellos está ubicado justamente antes de la falla ocurrida, lo que permite asociar ese valor con una alarma.

Algo similar ocurre con el indicador χ^2. Durante su evolución, se presentan dos valores elevados, donde justamente uno de ellos está ubicado inmediatamente antes de la parada de marzo de 2015, siendo este valor además el más elevado de toda la serie. Esto permite asociarse de forma directa con una alarma en el funcionamiento del aerogenerador C.

De esta forma, el decaimiento de R justo antes de la parada correspondiente a la falla es interpretado como una predicción de la misma. Asimismo, los valores elevados de Q^2 y χ^2 justo antes de la parada también son un indicio de una anomalía en el funcionamiento del equipo durante ese período. Finalmente, el hecho de que los tres indicadores señalan simultáneamente una posible anomalía potencial solamente en marzo de 2015 dan fuerza a la presunción de la ocurrencia de la misma en esa fecha.

Figura 5.9: Diagrama de dispersión para u_e y v_e, junto con las distribuciones de probabilidad correspondientes a estas variables.

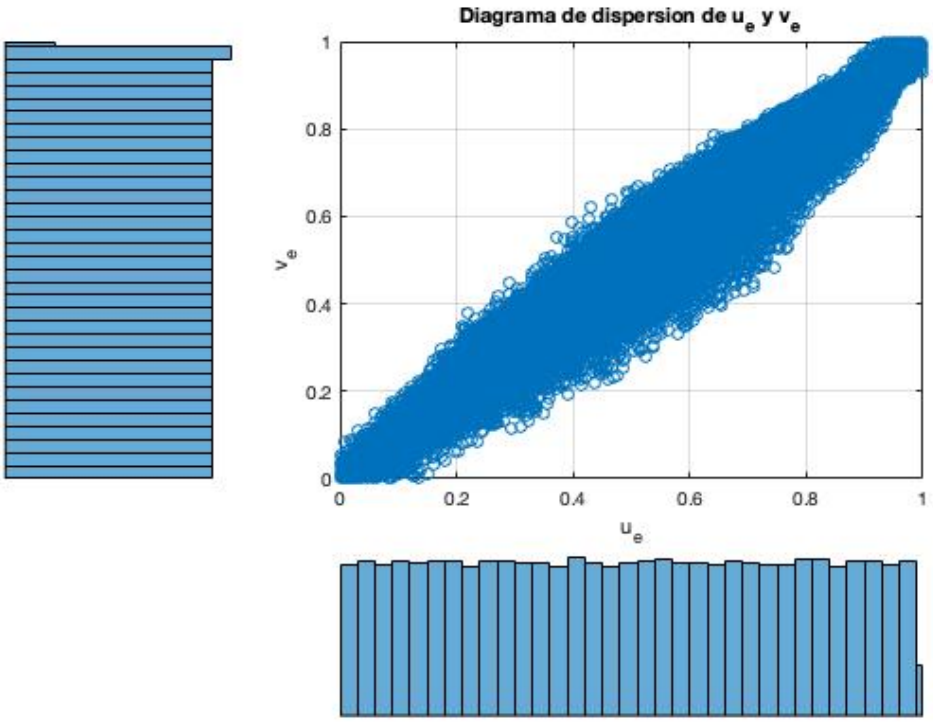


Figura 5.10: Cópula para los datos del aerogenerador C.

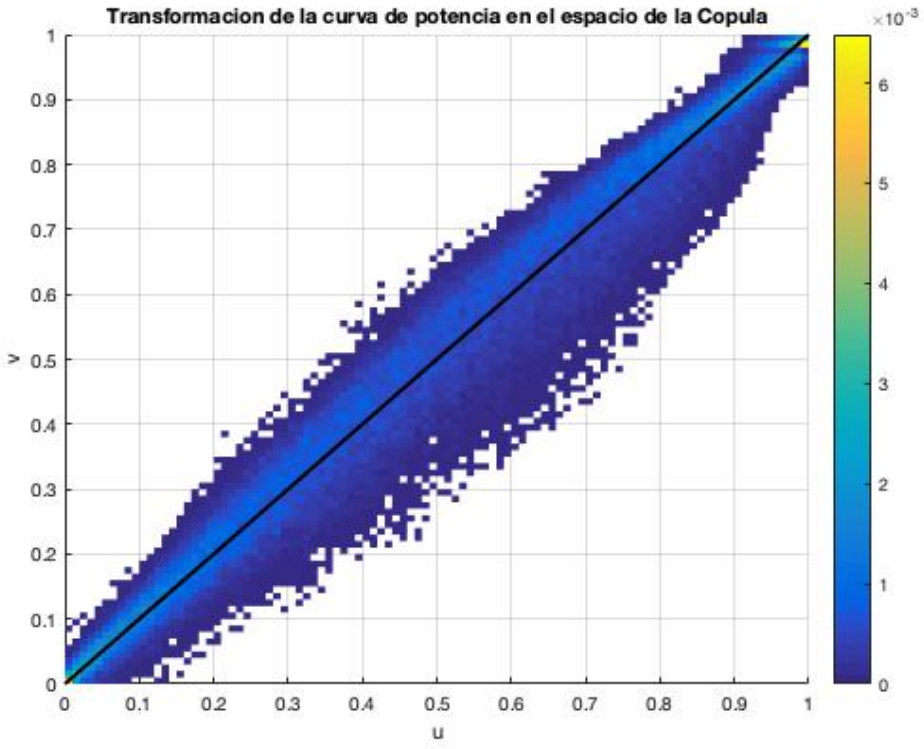
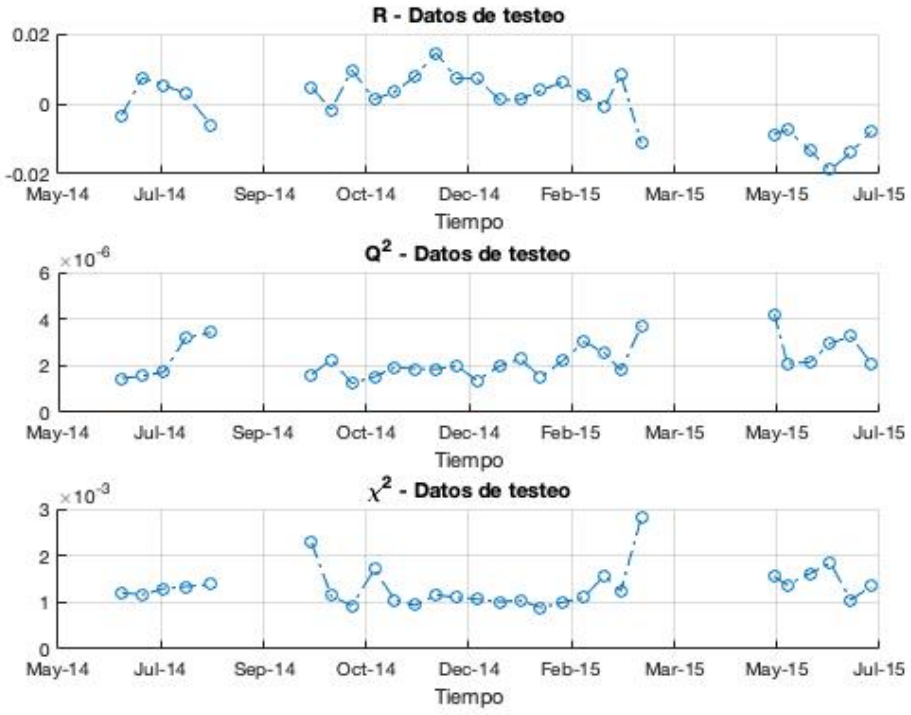


Figura 5.11: Indicadores comparativos: R (arriba), Q^2 (medio) y χ^2 (abajo).



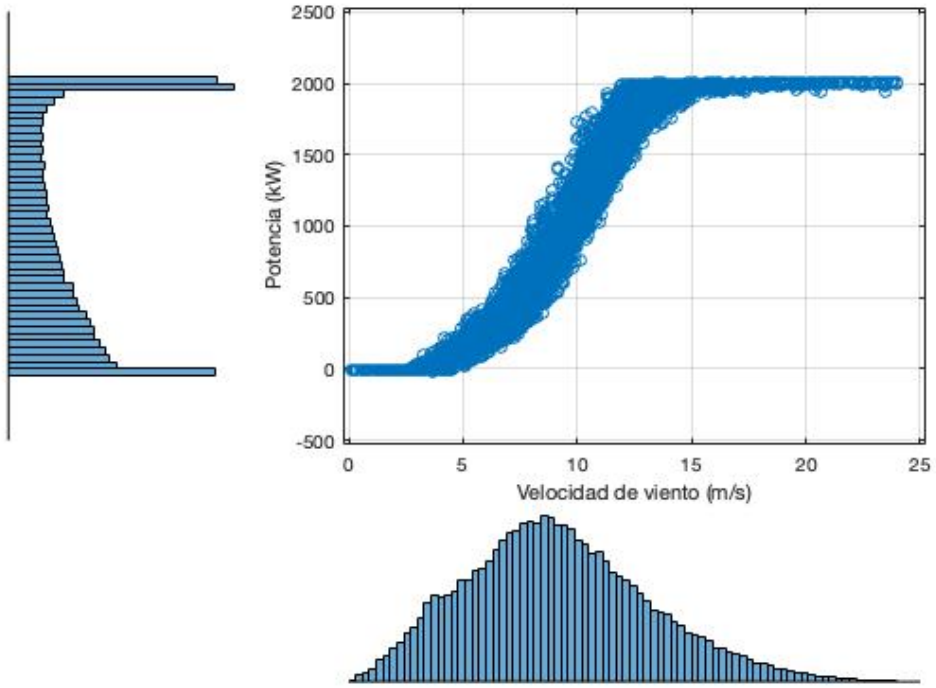
5.2.3. Aplicación para el Aerogenerador D

El método de Cópulas fue empleado para realizar un monitoreo por condición del funcionamiento del aerogenerador D.

El período de entrenamiento empleado para tales fines está comprendido entre enero de 2010 y junio de 2011. A partir de este conjunto de datos, es posible realizar la estimación de la Cópula asociada a D.

En la Figura 5.12 se presenta un diagrama de dispersión de la curva de potencia del aerogenerador D para el período de datos de entrenamiento, así como las distribuciones marginales de w_e y p_e. Esta información permite la estima-

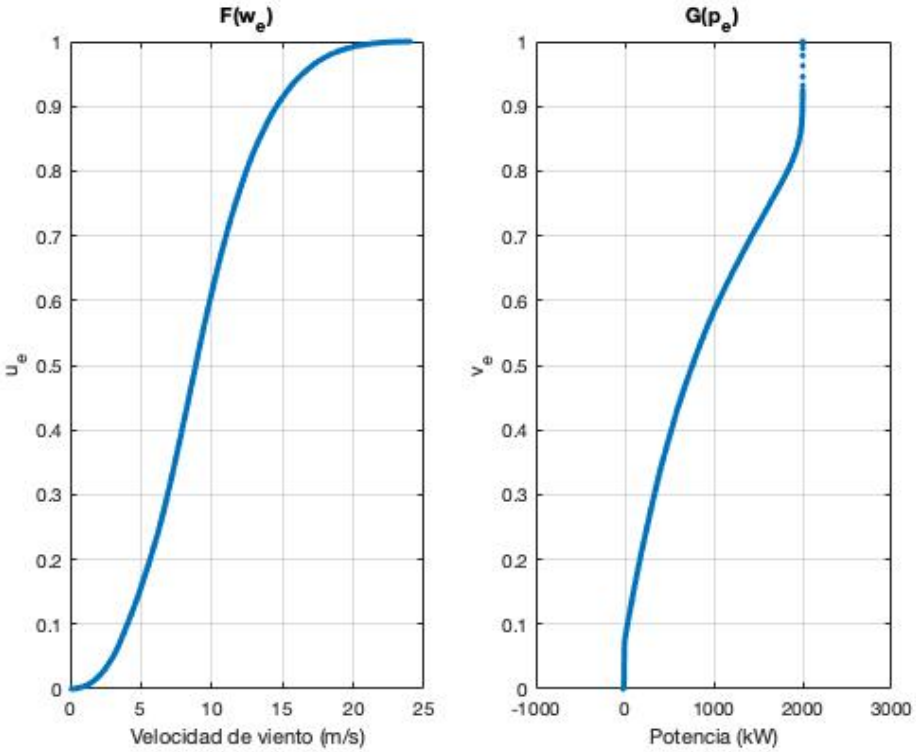
Figura 5.12: Diagrama de dispersión de la curva de potencia, junto con las distribuciones marginales de w_e y p_e, de los datos de entrenamiento del aerogenerador D.



ción de las distribuciones marginales acumuladas de w_e y p_e, F y G respectivamente. Las estimaciones de estas funciones se presentan en la Figura 5.13. Asimismo, esta información permite obtener las nuevas variables u_e y v_e para pasar al espacio de la Cópula.

En este sentido, en la Figura 5.14 se presenta un diagrama de dispersión entre los valores obtenidos de u_e y v_e, junto con las distribuciones de proba-

Figura 5.13: Distribución marginal acumulada de w_e, F (izquierda); distribución marginal acumulada de p_e, G (derecha).



bilidad asociadas a estas variables. Estas últimas se asemejan, tal como era previsible, a la distribución uniforme en I.

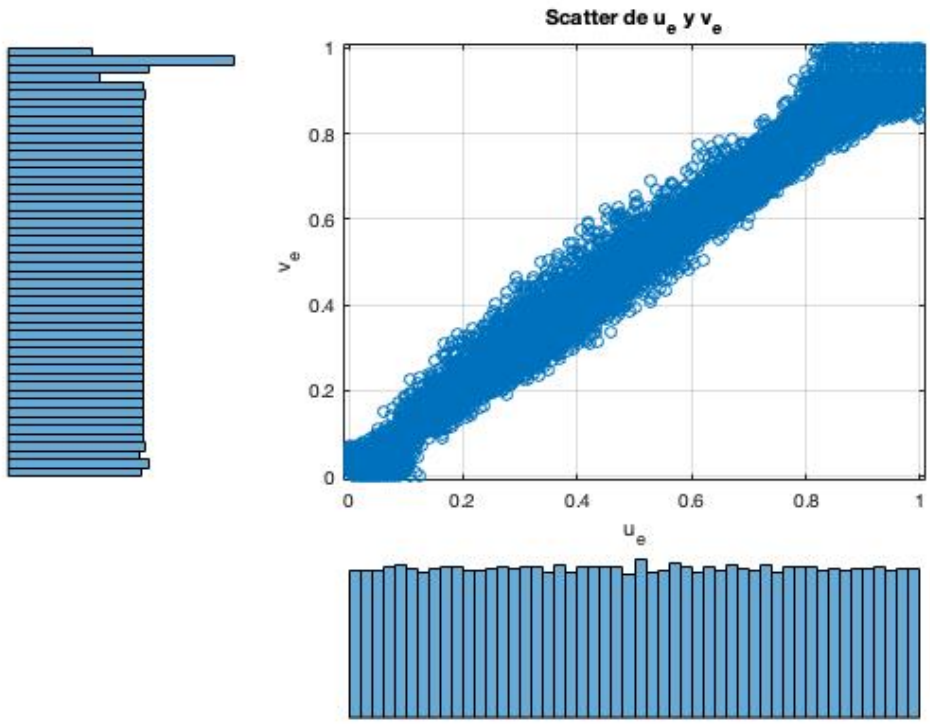
Esto permite obtener, a partir de la sub-división presentada en la sección 5.1.1, la densidad de la Cópula para el conjunto de datos de entrenamiento. En la Figura 5.15 se presenta la densidad de la Cópula del aerogenerador D. Puede observarse que los puntos de la Cópula se concentran alrededor de la recta $v = u$. Aunque se aprecia también que hay una cierta dispersión en las regiones cercanas a los puntos $[0, 0]$ y $[1, 1]$. Esta es consecuencia de la baja dependencia que hay entre la velocidad de viento y la potencia en los puntos asociados a bajas velocidades de viento o potencias cercanas a la nominal.

La densidad de la Cópula presentada anteriormente será utilizada para comparar con las densidades asociadas al conjunto de datos de testeo. Este período está comprendido entre junio de 2011 y agosto de 2015.

En la Figura 5.16 se presentan las evoluciones temporales de los indicadores R, Q^2 y χ^2, asociadas al conjunto de datos de testeo.

Como se mencionó en capítulos precedentes, no se dispone de información

Figura 5.14: Diagrama de dispersión para u_e y v_e, junto con las distribuciones de probabilidad correspondientes a estas variables.

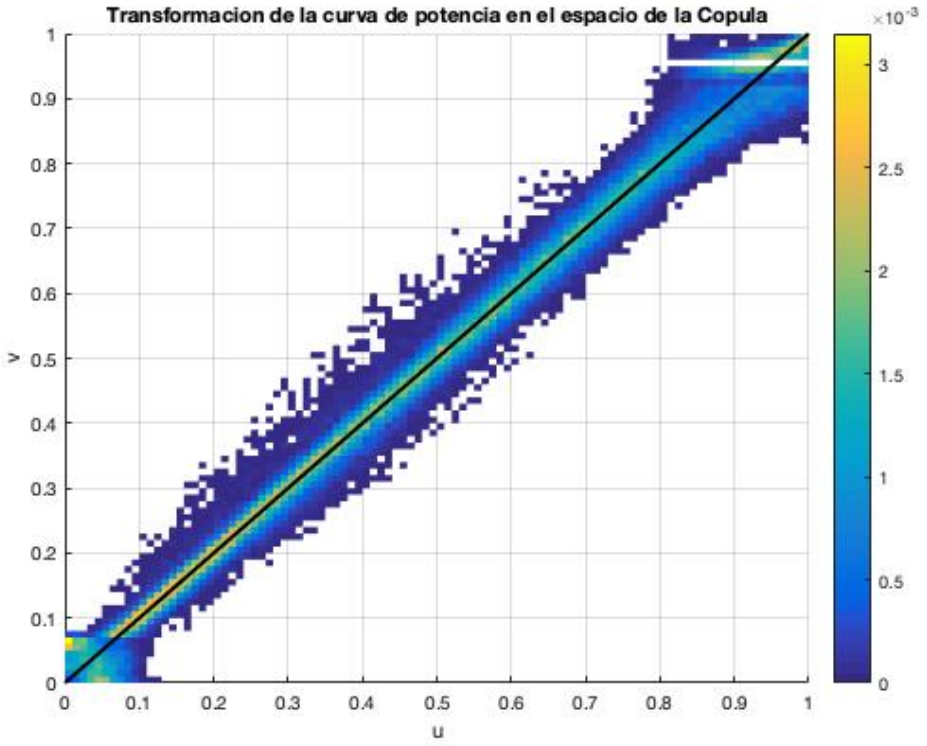


acerca de alguna falla asociada al aerogenerador D durante el período de funcionamiento. Esto hace que el análisis de la Figura 5.16 no se centre en la predicción de fallas. Sin embargo, el objetivo es comparar la performance del equipo entre antes y después de la parada de mantenimiento de noviembre de 2013.

Puede apreciarse que el indicador R de la Figura 5.16 presenta un notorio aumento luego de la parada de mantenimiento, manteniéndose esta tendencia durante el resto del período relevado. Esto se traduce directamente como una mejora en la performance del aerogenerador, expresada a través de la curva de potencia. A modo ilustrativo, en la Figura 5.17 se presenta la densidad de la Cópula obtenida, a partir de las funciones F y G ya determinadas, para el mes de mayo de 2015. Puede observarse que, en términos generales, los puntos se concentran por encima de la recta $v = u$, correspondiéndose esto con una mejora en comparación con la operación durante el período previo a la parada.

El indicador Q^2 también presenta un notorio aumento en sus valores luego de la parada de noviembre de 2013. Esto se traduce en una pérdida de si-

Figura 5.15: Cópula para los datos del aerogenerador D.



metría de la densidad de la Cópula respecto a la recta $v = u$. Esto también se ejemplifica en la Figura 5.17.

Finalmente, el indicador χ^2 también presenta valores elevados para el período de la pos-parada, si se los compara con los obtenidos durante período previo. Esto corresponde a que la Cópula del último período no está representada adecuadamente por la obtenida durante el período de entrenamiento. Esta discrepancia puede interpretarse también como una consecuencia de que el aerogenerador D tuvo un cambio en su funcionamiento luego de la parada.

A partir del análisis anterior, es posible concluir que el método de la Cópula permite inferir que la condición de funcionamiento del aerogenerador D tuvo un cambio luego de la parada de noviembre de 2013.

Figura 5.16: Indicadores comparativos: R (arriba), Q^2 (medio) y χ^2 (abajo).

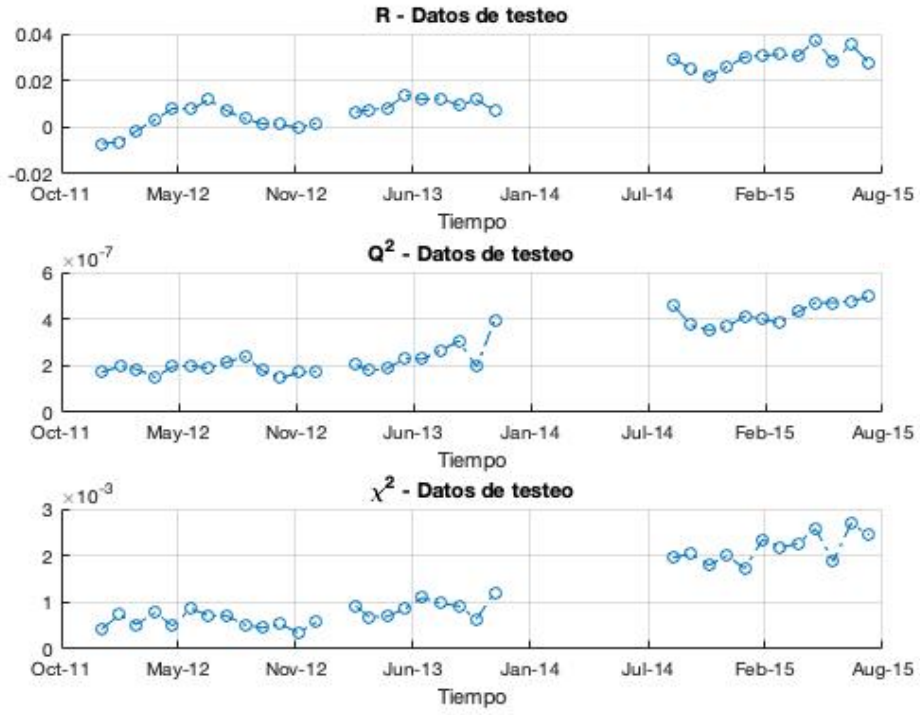
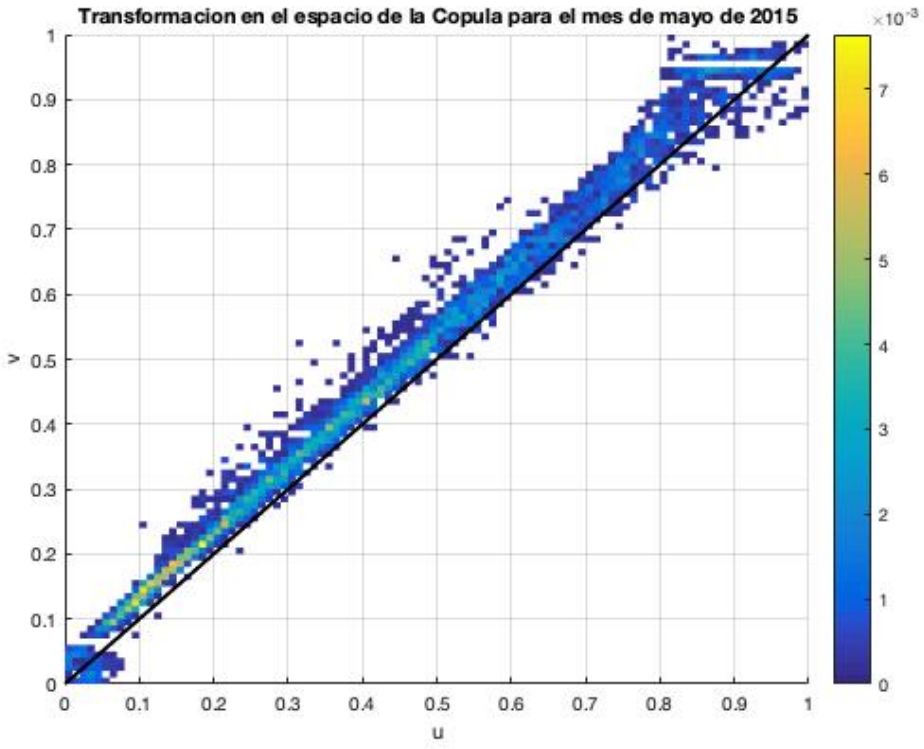


Figura 5.17: Cópula para el aerogenerador D durante el mes de mayo de 2015.



Capítulo 6

Análisis de Componentes Principales interpretado a través de One Class - Support Vector Machine (PC-1cSVM)

En este capítulo se desarrolla un método para predecir fallas en el funcionamiento de un aerogenerador. Se trata de una técnica inédita, aunque basada en una composición de métodos presentados por otros autores, que se compone de dos etapas fundamentales: un análisis de componentes principales y la aplicación de una técnica de SVM.

Skrimpas et al. (2014) exploraron la técnica de detección de anomalías en la curva de potencia usando valores propios de la matriz de covarianza de los datos, mostrando resultados interesantes y dejando abierto el problema de introducir variables de otra naturaleza además de la velocidad de viento y la potencia.

Por otra parte, la aplicación de SVM fue explorada por diversos autores. A modo de ejemplo, Ma and Perkins (2003) emplearon la técnica de One Class - SVM para detectar anomalías en series temporales, aunque no aplicado directamente a aerogeneradores. Asimismo, Martínez-Rego et al. (2011) generaron un modelo de One Class - SVM para predecir fallas en aerogeneradores, a partir de señales de vibración de los mismos. Laouti et al. (2011) emplearon un enfoque de SVM para clasificar datos de operación, y de esta forma identificar fallas vinculadas al pitch, al generador y al rotor.

El método propuesto en este capítulo es una combinación de los dos enfoques, que tiene por objetivo predecir fallas en el funcionamiento de la turbina. El desarrollo de esta herramienta es presentada a continuación, junto con la aplicación de la misma para los aerogeneradores de estudio.

6.1. Descripción teórica

PC-1cSVM es una técnica de predicción de fallas en el funcionamiento del aerogenerador que está constituida por dos partes esenciales. En primer lugar, la información seleccionada del SCADA es procesada mediante un análisis de componentes principales (PC). En una segunda instancia, los datos obtenidos en la primera etapa son clasificados mediante la técnica de One Class - SVM (1cSVM). En esta sección se describen ambas herramientas, haciendo hincapié en la ventaja de relacionar una con otra.

6.1.1. Componentes principales (PC)

El método de componentes principales es una técnica multivariada que se utiliza en muchas áreas de investigación, con diversos objetivos (Jackson (1991), Preisendorfer (1988), entre otros). En este trabajo, se utilizará específicamente la estructura de los valores propios, como se explica a continuación.

Sea un conjunto de n observaciones de K variables, organizadas en un matriz W de tamaño $K \times n$, de forma que cada observación i, con $i = 1, 2, \ldots, n$, está contenida en el vector w_m, con $m = 1, 2, \ldots, K$, de componentes w_{m_i}. Definimos los vectores X_m como sigue.

$$x_m = w_m - \bar{w_m} \tag{6.1}$$

Donde $\bar{w_m}$ es el valor medio de w_m en las n observaciones. Se obtiene así la matriz X de tamaño $K \times n$.

El método de componentes principales busca obtener nuevas variables u_m, con $m = 1, 2, \ldots, K$, no correlacionadas entre sí, y que sean combinaciones lineales de las variables x_m, tales que contengan una fracción decreciente de la varianza total de las variables originales. Esas nuevas variables u_m (que son vectores de largo n) se llaman componentes principales, y tienen asociada una matriz U.

Sea S_X la matriz de covarianza de X, de forma que:

$$S_X = \frac{XX^T}{n-1} \tag{6.2}$$

El problema se resuelve hallando los vectores y valores propios de S_X (e_m y λ_m, con $m = 1, 2, \ldots, K$), que verifican la siguiente relación.

$$S_X e_m = \lambda_m e_m \tag{6.3}$$

Los vectores e_m forman una base ortonormal de $\mathbb{R}^K$ y se organizan en una matriz ortogonal E, donde la columna m-ésima de E es e_m.

Geométricamente, el vector e_1 apunta en la dirección (en $\mathbb{R}^K$) en la que los datos exhiben la mayor variabilidad. Así, e_1 está asociado con el mayor valor propio de S_X, que por convención se asigna a λ_1 ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$). De forma general, e_m apunta en la dirección en la que los datos tienen mayor variabilidad posible, sujeto a que $e_m \perp \{e_1, e_2, \ldots, e_{m-1}\}$.

La ecuación 6.3 puede escribirse de forma matricial de acuerdo a la siguiente expresión.

$$S_X E = E\Lambda \tag{6.4}$$

Donde Λ es una matriz diagonal con los valores propios λ_m.

Este conjunto de versores propios define un nuevo sistema de coordenadas en el que ver los datos.

$$U = E^T X \tag{6.5}$$

Sea S_U la matriz de covarianza de las componentes principales, o sea la matriz de covarianza asociada a U. Luego, S_U se obtiene como la diagonalización de S_X de acuerdo a la siguiente relación.

$$S_U = var\left(E^T X\right) = E^T S_X E = E^{-1} S_X E = \Lambda \tag{6.6}$$

Esto implica que la varianza de la componente principal u_m es igual a λ_m. En este sentido, la variación total exhibida por los datos originales está representada por las K componentes principales (Wilks (2011)).

El proceso anterior puede ser conducido también a través de la matriz de correlaciones R, obteniendo así la matriz Λ correspondiente. La matriz de

correlaciones es la matriz de covarianza de los datos z estandarizados.

$$z_m = \sigma_m^{-1} x_m \tag{6.7}$$

Donde σ_m es la desviación estándar de la serie temporal de w_m. La principal virtud de emplear el análisis anterior a través de R está en independizar el problema de las unidades de las K variables que, en el caso de interés, suelen en general ser distintas. Cabe destacar que en este caso se verifica la siguiente relación.

$$\sum_{m=1}^{K} \lambda_m = K \tag{6.8}$$

De esta forma, los valores propios λ_m provenientes de un conjunto de observaciones de K variables, con $m = 1, \ldots, K$, representan la varianza de los datos estandarizados, en las direcciones especificadas por los versores e_m correspondientes. Cabe destacar que los versores propios e_m, obtenidos a partir de la matriz de correlaciones, no necesariamente coinciden con los obtenidos a partir de la matriz de covarianza. Asimismo, asumiendo por convención que $\lambda_1 \geq \cdots \geq \lambda_K$ y, asumiendo que las K variables involucradas tienen un cierto grado de correlación entre ellas, una estrategia para identificar cambios en el comportamiento de los datos es analizar la evolución temporal de los valores propios de menor magnitud; los de mayor magnitud estarán asociados a las direcciones preferenciales en que los datos presentan mayor varianza (Skrimpas et al. (2014)). Así, aumentos considerables en algún λ_m, con $m = K - k + 1, \ldots, K$, con $k < K$, podrá traducirse como una anomalía en las observaciones.

Las series temporales de estos k valores propios serán el input para la siguiente instancia del método.

6.1.2. One Class Support Vector Machine (1cSVM)

Sean N muestras temporales de $L = \{\lambda_{K-k+1}, \ldots, \lambda_K\} \in \mathbb{R}^k$, con $k < K$, de forma que se obtienen k series temporales. Es necesario entonces desarrollar una herramienta que permita, a partir de estas k series, identificar anomalías durante la evolución temporal para nuevas observaciones de L. Para ello empleamos la técnica de 1c-SVM. El método desarrollado en este trabajo busca encontrar la ecuación de un hiper-plano, en un espacio de dimensión mayor a

k, que permita separar los vectores L de funcionamiento saludable del aerogenerador, de los que representan anomalías en los datos.

Sea L_e la matriz de tamaño $N \times k$, cuyas filas son los vectores L asociados a un funcionamiento saludable del aerogenerador, con el que se obtendrá la ecuación del hiper-plano; datos de entrenamiento. Los parámetros del hiper-plano buscado son determinados a partir de la resolución del siguiente problema de optimización primal cuadrático (Scholkopf et al. (2001)).

$$\min_{\mathbf{w},\xi_j,\rho} \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\mu N}\sum_{j=1}^{N}\xi_j - \rho \tag{6.9}$$

$$\text{sujeto a: } \mathbf{w}.\phi(y_j) \geq \rho - \xi_j, \quad \xi_j \geq 0, \quad \forall j = 1, \ldots, N$$

Donde $\phi : \mathbb{R}^k \to \mathcal{F}$ es una función que mapea los vectores y en un espacio $\mathcal{F}$ de mayor dimensión; $\mathbf{w}$ y ρ determinan el hiper-plano; ξ_j son variables de holgura; y_j los vectores de entrenamiento, obtenidos a partir de los valores propios introducidos anteriormente como se explica a continuación; y $\mu \in (0, 1]$ es un parámetro propio del algoritmo.

En la Figura 6.1 se presenta un esquema de la separación de los puntos a través del hiper-plano: lo que se busca es el mejor hiper-plano, en algún sentido, que separe los puntos del origen, permitiendo una cantidad igual a μN de outliers que caigan entre el origen y este hiper-plano. Se espera así, que los puntos de entrenamiento, asociados a un funcionamiento saludable, se encuentren del lado $+1$ del hiper-plano. Por este motivo, debe aplicarse una transformación a los vectores L de forma de cumplir con este requerimiento.

$$\tau : \mathbb{R} \to [0, K] \text{ tal que } \tau(L_m) = K - L_m \tag{6.10}$$

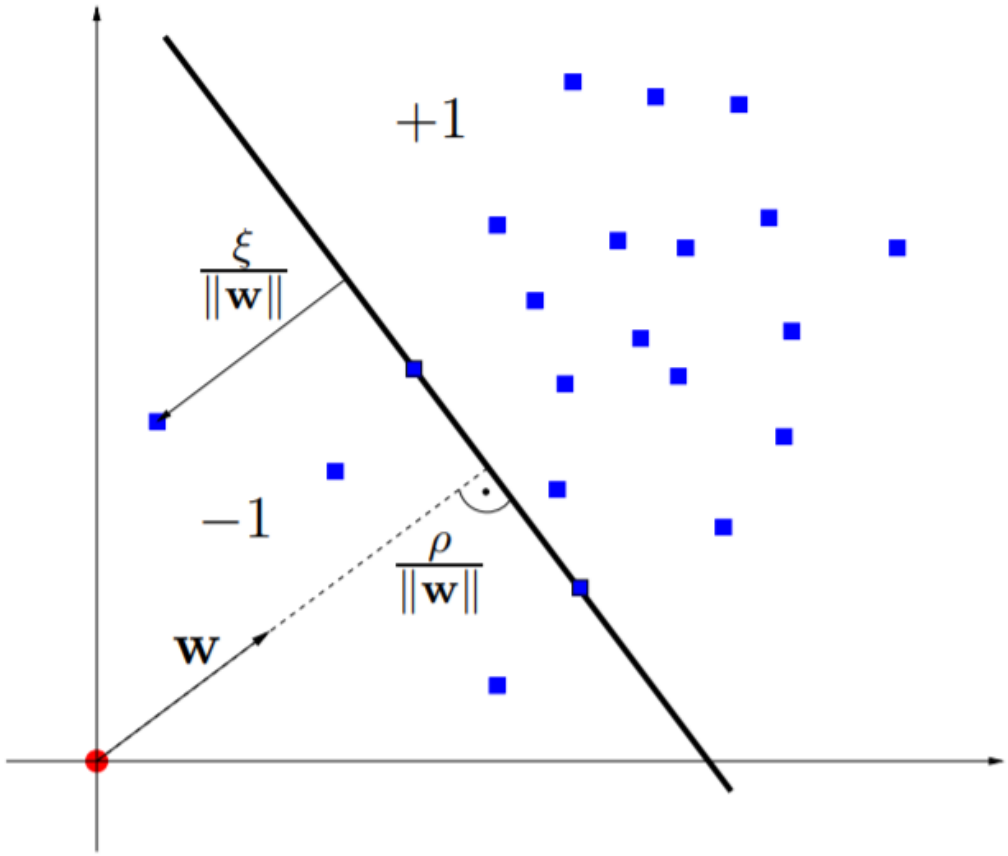
Así, los vectores de entrenamiento y_j del problema 6.9 quedan determinados por la siguiente ecuación.

$$y_j = \left(\tau(L_{K-k+1}), \ldots, \tau(L_K)\right), j = 1, \ldots, N \tag{6.11}$$

Una vez que los parámetros $\mathbf{w}$ y ρ son determinados, una nueva observación, con un vector $\tilde{y}$ asociado, será clasificada a partir de la siguiente función de decisión.

$$f(\tilde{y}) = sgn\left(\mathbf{w}.\phi(\tilde{y}) - \rho\right) \tag{6.12}$$

Figura 6.1: Esquema de separación de de los puntos a través del hiper-plano.



Resta entonces encontrar la solución óptima del problema expresado en la Ecuación 6.9. Para ello, introducimos el Lagrangiano $\mathcal{L}$.

$$\mathcal{L}(\mathbf{w}, \Xi, \rho, \Omega, \Gamma) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\mu N}\sum_{j=1}^{N}\xi_j - \rho - \sum_{j=1}^{N}\omega_j\left(\mathbf{w}.\phi(y_j) - \rho + \xi_j\right) - \sum_{j=1}^{N}\gamma_j\xi_j \tag{6.13}$$

Donde $\Xi = (\xi_1, \ldots, \xi_N)$ y, $\Omega = (\omega_1, \ldots, \omega_N)$ y $\Gamma = (\gamma_1, \ldots, \gamma_N)$ son los multiplicadores de Lagrange, que verifican $\omega_j \geq 0$ y $\gamma_j \geq 0$, $\forall j = 1, \ldots, N$. Igualando las derivadas de $\mathcal{L}$ (respecto a las variables primales $\mathbf{w}$, Ξ y ρ) a 0, tenemos las siguientes tres ecuaciones.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{j=1}^{N}\omega_j\phi(y_j) \tag{6.14}$$

$$\frac{\partial \mathcal{L}}{\partial \Xi} = 0 \Rightarrow \omega_j = \frac{1}{\mu N} - \gamma_j, \forall j = 1, \ldots, N \tag{6.15}$$

$$\frac{\partial \mathcal{L}}{\partial \rho} = 0 \Rightarrow \sum_{j=1}^{N} \omega_j = 1 \tag{6.16}$$

Como $\omega_j \geq 0$, de la Ecuación 6.15 se desprende que $\gamma_j \leq \frac{1}{\mu N}, \forall j = 1, \ldots, N$. Asimismo, como $\gamma_j \geq 0$, la Ecuación 6.15 puede reescribirse como sigue.

$$0 \leq \omega_j = \frac{1}{\mu N} - \gamma_j \leq \frac{1}{\mu N}, \forall j = 1, \ldots, N \tag{6.17}$$

Definimos entonces el conjunto de Vectores de Soporte como $SV = \{y_j : \omega_j > 0\}$. De esta forma, la Ecuación 6.14 puede reescribirse de acuerdo a la siguiente expresión.

$$\mathbf{w} = \sum_{y_j \in SV} \omega_j \phi(y_j) \tag{6.18}$$

Esto hace que la función de decisión de la Ecuación 6.12 pueda reescribirse como sigue.

$$f(\tilde{y}) = sgn\left(\sum_{y_j \in SV} \omega_j \phi(y_j)\phi(\tilde{y}) - \rho\right) = sgn\left(\sum_{y_j \in SV} \omega_j k(y_j, \tilde{y}) - \rho\right) \tag{6.19}$$

Donde $k(y_i, y_j) = \phi(y_i).\phi(y_j)$, ya que las imágenes de ϕ pueden ser evaluadas con una función kernel. Esto hace innecesario el conocimiento de la forma explícita de la función ϕ.

Sustituyendo las Ecuaciones 6.18 y 6.19 en la Ecuación 6.13, se obtiene la formulación del problema dual, expresado en la siguiente ecuación.

$$\min_{\Omega} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j k(y_i, y_j) \tag{6.20}$$

$$\text{sujeto a: } 0 \leq \omega_j \leq \frac{1}{\mu N}, \forall j = 1, \ldots, N, \quad \sum_{j=1}^{N} \omega_j = 1$$

Finalmente, se puede probar que, en la solución óptima, las restricciones del problema expresado en la Ecuación 6.9, se vuelven igualdades si ω_j y γ_j son no nulos, lo que es equivalente a que $0 < \omega_j < \frac{1}{\mu N}$. Luego, el valor de ρ puede ser determinado, ya que para cualquiera de esos ω_j, el y_j correspondiente

satisface la siguiente ecuación.

$$\rho = \mathbf{w}.\phi(y_j) = \sum_{i=1}^{N} \omega_i k(y_i, y_j) \tag{6.21}$$

El problema dual expresado en la Ecuación 6.20, por ser convexo, independientemente de si el problema primal lo es, es más beneficioso de resolver. Además tiene otra virtud frente al problema primal, que es que es un problema en un sólo parámetro (de N componentes), mientras que el problema primal es en tres parámetros (de al menos $k + N + 1$ componentes en total). Esto también hace que la resolución del problema dual sea más conveniente.

Así, a través de la resolución del problema dual de la Ecuación 6.20, es posible determinar ρ mediante la Ecuación 6.21, y de esta forma clasificar nuevas observaciones mediante la función de decisión de la Ecuación 6.19.

De esta forma, para un conjunto de N observaciones de K variables, es posible clasificar los datos obtenidos a partir del análisis de componentes principales, con el fin de identificar variabilidades significativas en su conjunto, que se traduzcan en anomalías de los datos SCADA, y conduzcan así a predecir potenciales fallas del aerogenerador.

6.1.3. Metodología de aplicación

Las herramientas descritas en la sección precedente hacen posible obtener una metodología para predecir fallas en el funcionamiento del aerogenerador.

Dada una matriz X_e de entrenamiento de datos SCADA, compuesta por n_e observaciones de K variables, el primer paso consiste en la obtención de los datos primarios para procesar. Cabe destacar que el análisis se centrará únicamente en los puntos asociados a la parte cuasi-lineal de la curva de potencia, dado que en esta región se encuentra la mayor información asociada a la degradación del funcionamiento del aerogenerador (Jia et al. (2016)). Para ello, el primer paso consiste en, a partir de la curva de potencia, seleccionar el rango de interés para llevar a cabo el análisis.

Luego, el período de entrenamiento seleccionado es subdividido en ventanas móviles de largo constante LV y paso r. De esta forma, dos ventanas consecutivas tendrán $LV - r$ diezminutales en común; los primeros r diezminutales de una ventana no estarán presentes en la siguiente. Nuevamente, este parámetro LV, junto con r, son considerados escalas temporales dentro del

análisis. Cabe mencionar que, si bien todas las ventanas tienen el mismo largo temporal, eventualmente puede ocurrir que la cantidad de datos dentro de una ventana sea inferior a LV, consecuencia de falta de mediciones. Se considera la matriz de correlaciones R_e de los datos, de tamaño $K \times K$, para las ventanas temporales i. Para cada una de estas matrices, se toman los k valores propios asociados de menor magnitud, donde k será tomado igual a 2 en este trabajo.

Los vectores de entrenamiento, que luego son empleados para la obtención del hiper-plano, son obtenidos mediante la Ecuación 6.11. De esta forma, para el conjunto de datos de entrenamiento, se obtiene una serie de vectores $y_i \in [0, K]^k$.

El problema dual presentado en la Ecuación 6.20 es resuelto, a partir de los vectores de entrenamiento correspondientes, en el entorno `cvx` de Matlab. La solución permite determinar el valor de ρ a partir de la Ecuación 6.21, junto con el conjunto SV y así, poder determinar la función de decisión expresada en la Ecuación 6.19. Cabe destacar que la función kernel empleada en este trabajo es una función lineal, dada por la siguiente ecuación.

$$k(y_i, y_j) = y_i^T y_j \tag{6.22}$$

Una vez determinada la función de decisión, es posible determinar, para nuevos vectores, de qué lado del hiper-plano se encuentran. La etapa de testeo consiste entonces en tomar un conjunto de datos a evaluar, compuestos por n_t observaciones de K variables. La obtención de los datos primarios para procesar es igual que en el caso anterior. Esto es, seleccionar el rango de interés correspondiente a la parte cuasi-lineal de la curva de potencia; subdividir los datos en ventanas temporales de largo constante LV y paso r; considerar, en cada una de las sub-ventanas, la matriz de correlación R_{t_i}; tomar los k valores propios asociados de menor magnitud, y obtener los puntos en $[0, K]^k$ mediante la transformación de la Ecuación 6.11. Así, con la función de decisión, es posible determinar qué puntos de operación corresponden a una potencial anomalía en el funcionamiento.

6.2. Resultados

6.2.1. Aplicación para el Aerogenerador A

La metodología desarrollada en la sección precedente fue aplicada para estudiar el conjunto de datos del aerogenerador A. En este sentido, y por los motivos ya expuestos en la sección 3.2.1, los datos asociados al aerogenerador A_0 fueron considerados como datos saludables de entrenamiento. PC-1cSVM fue aplicado para analizar el conjunto de variables que comprende la temperatura de la 3ra fase del generador, la potencia, la temperatura del rodamiento 1 del generador, y la temperatura del rodamiento 2 del generador; de esta forma $K = 4$.

La fracción μ de outliers en el problema de optimización fue fijada en 0.05; este valor fue seleccionado con el fin de que los análisis correspondientes a los aerogeneradores estudiados en las secciones próximas sean llevados a cabo bajo el mismo valor del parámetro. Asimismo, el parámetro LV fue fijado en 432 diezminutales (correspondientes a 72 horas de operación), al igual que en secciones precedentes, mientras que el paso r fue tomado en 43 diezminutales (aproximadamente 10 % de LV). Finalmente, como se mencionó anteriormente, la cantidad de valores propios empleados para el análisis es de $k = 2$.

En la Figura 6.2 se presenta la evolución temporal de los 2 valores propios de menor magnitud obtenidos a partir de la matriz de correlación de los datos subdivididos en ventanas temporales. Allí puede vislumbrarse en una primera instancia algunos aumentos puntuales, aunque de forma individual, para cada valor propio, durante la evolución del funcionamiento del aerogenerador. Estos datos son procesados de acuerdo a la metodología descrita con el fin de obtener las alarmas asociadas al método.

En la Figura 6.3 se presenta la evolución temporal de la temperatura de la 3ra fase del generador, junto con las alarmas detectadas por el método PC-1cSVM para analizar los datos de testeo. Allí puede verse que tanto la parada de abril como la de setiembre pueden ser anticipadas. Algunas alarmas aparecen algunos días antes previo a la primera parada. A su vez, la segunda parada es anticipada por algunas alarmas que aparecen durante los meses previos a setiembre de 2016. Se tiene entonces que este método da indicios de predecir las dos fallas ocurridas en el aerogenerador A.

Figura 6.2: Evolución temporal de los 2 valores propios de menor magnitud obtenidos a partir de la matriz de correlación de los datos de las ventanas temporales.

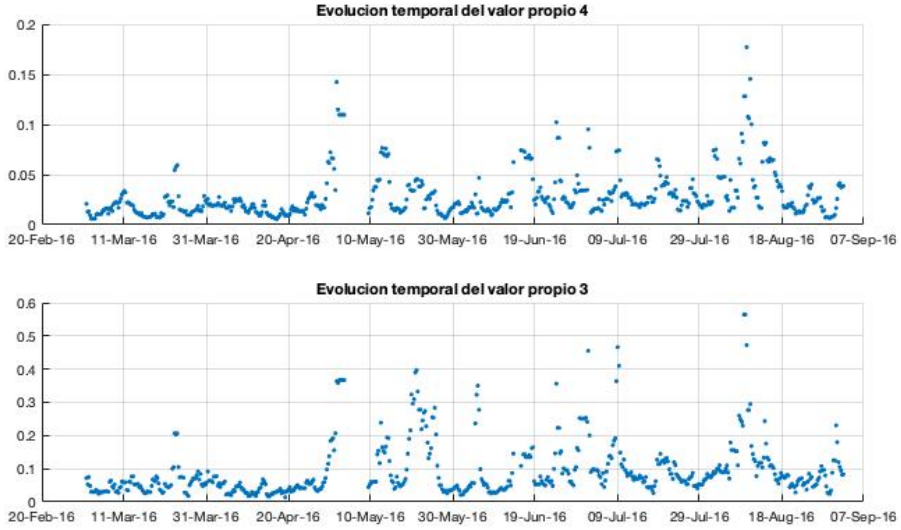
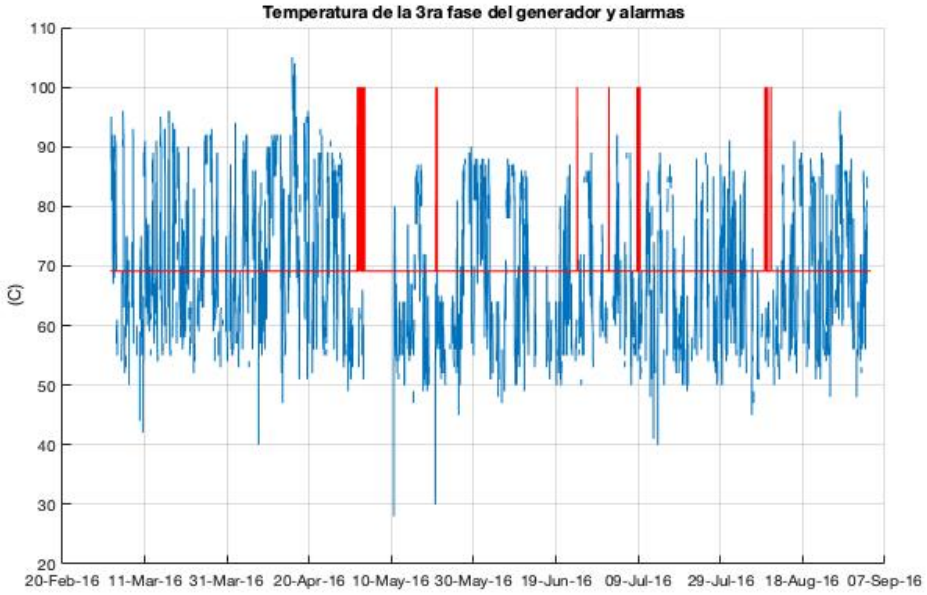


Figura 6.3: Evolución temporal de la temperatura de la 3ra fase del generador, junto con las alarmas registradas por el método PC-1cSVM.



6.2.2. Aplicación para el Aerogenerador B

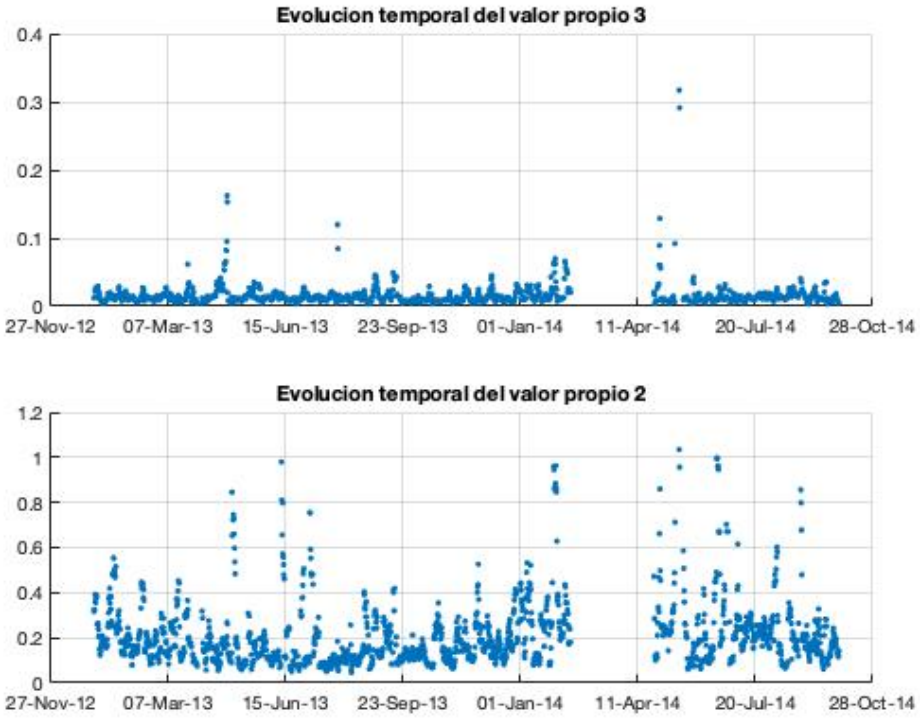
PC-1cSVM también fue aplicado para el caso del aerogenerador B. Por los motivos ya expuestos en la sección 3.2.2, las variables a analizar son la potencia,

la velocidad de viento y la temperatura del rodamiento A de la CM, de forma que $K = 3$. Asimismo, el período de entrenamiento empleado es el mismo que fue utilizado en las secciones precedentes asociadas al aerogenerador B: abril de 2011 a diciembre de 2012.

A su vez, al igual que en la sección 6.2.1, la fracción μ de outliers en el problema de optimización fue fijada en 0.05. El parámetro LV fue fijado en 432 diezminutales y el paso r en 43 diezminutales. La cantidad de valores propios empleados para el análisis también es $k = 2$.

En la Figura 6.4 se presenta la evolución temporal de los 2 valores propios de menor magnitud obtenidos a partir de la matriz de correlación de los datos de las ventanas temporales, correspondientes al conjunto de datos de testeo; comprendidos entre enero de 2013 y setiembre de 2014. Pueden observarse allí el comportamiento de cada valor propio a lo largo de tiempo, de forma independiente.

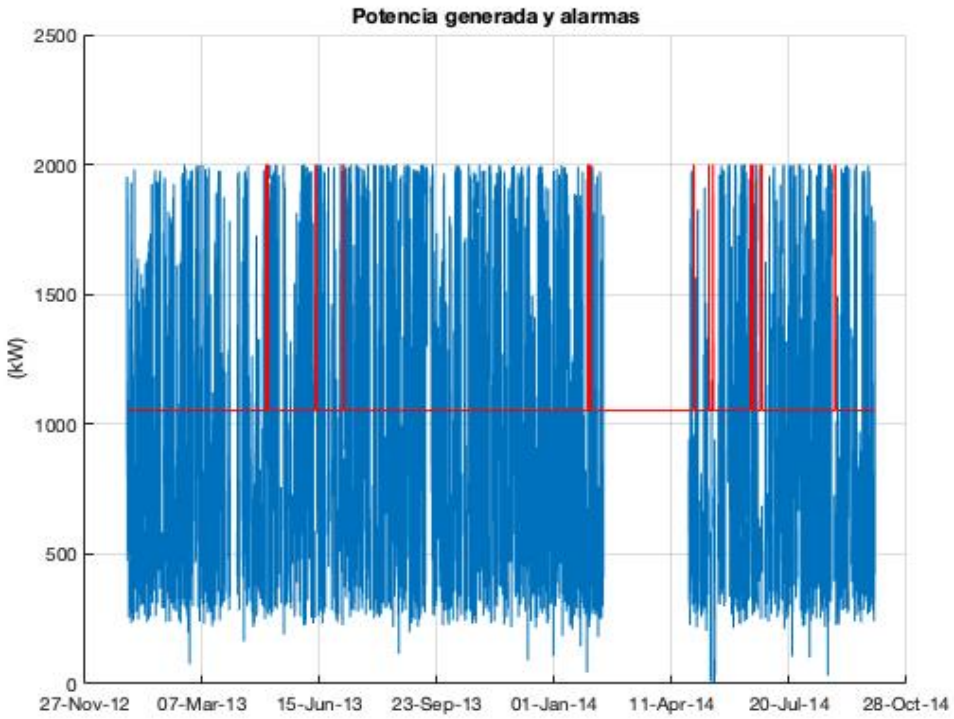
Figura 6.4: Evolución temporal de los 2 valores propios de menor magnitud obtenidos a partir de la matriz de correlación de los datos de las ventanas temporales.



En este sentido, en la Figura 6.5 se presenta la evolución temporal de la potencia generada por el aerogenerador B durante su funcionamiento en el

período de testeo, junto con las alarmas identificadas por el método. Puede

Figura 6.5: Evolución temporal de la potencia generada, junto con las alarmas registradas por el método PC-1cSVM.



observarse en la figura que el método fue capaz de reconocer algunas alarmas justo antes de la parada de enero de 2014. Asimismo, se identificaron algunas alarmas en meses anteriores a la parada, así como también luego de la misma; se considera que estas son falsas alarmas generadas por el método.

Lo anterior permite predecir la falla asociada a la CM del equipo con una anticipación de algunos días. Aunque cabe destacar la presencia de falsas alarmas.

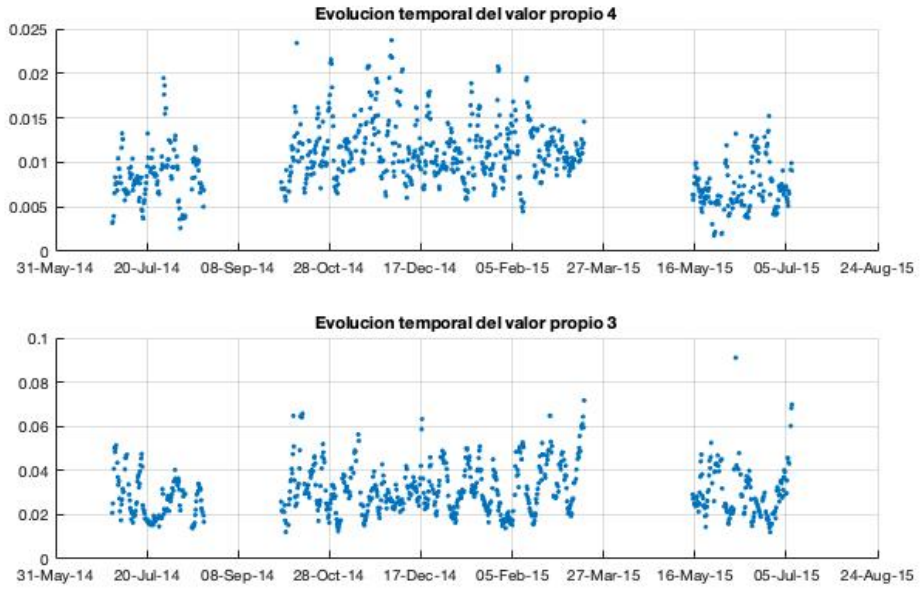
6.2.3. Aplicación para el Aerogenerador C

La aplicación de este método fue llevada a cabo para los datos correspondientes al aerogenerador C. Al igual que en lo expuesto en la sección 4.2.3, las variables a analizar son la carga en la pala C, la potencia, la velocidad de viento y las RPM del rotor, de forma que $K = 4$. El período de entrenamiento del modelo está comprendido entre diciembre de 2012 y junio de 2014.

La fracción μ de outliers en el período de entrenamiento se mantuvo en 0.05, mientras que el parámetro LV también se mantuvo igual a 432 diezminutales, al igual que el paso r en 43 diezminutales. Finalmente, la cantidad de valores propios empleada para el análisis también es $k = 2$. A partir de estos parámetros y de los datos de entrenamiento, se obtuvo el hiper-plano correspondiente que permite clasificar los puntos de testeo.

En la Figura 6.6 se presenta la evolución temporal de los correspondientes 2 valores propios de la matriz de correlación de las ventanas temporales, asociadas a los datos de testeo. Este período de testeo está comprendido entre julio de 2014 y julio de 2015. En esta figura puede observarse cómo evoluciona

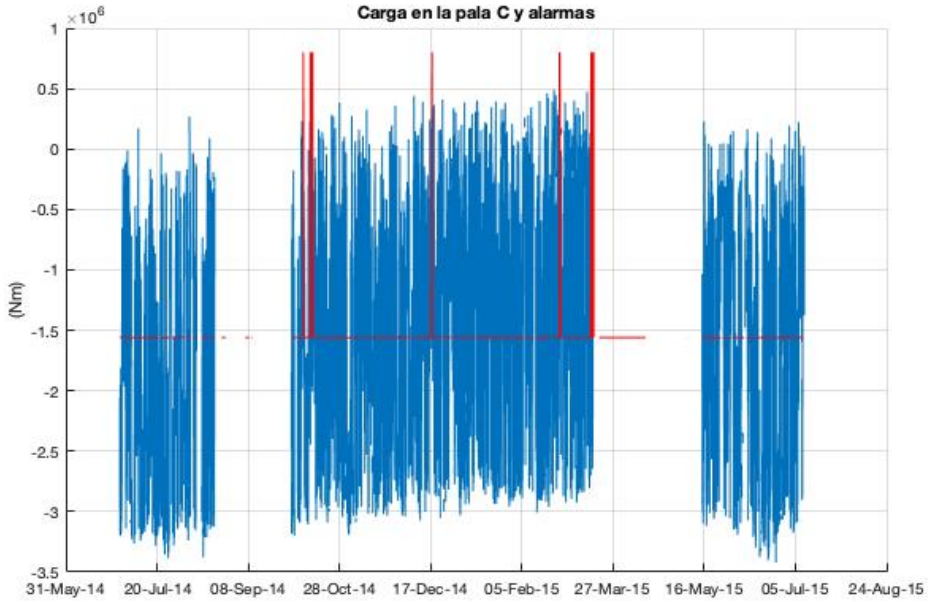
Figura 6.6: Evolución temporal de los 2 valores propios de menor magnitud obtenidos a partir de la matriz de correlación de los datos de las ventanas temporales.



cada valor propio de forma independiente.

A partir de la función de decisión determinada, los puntos de operación fueron clasificados con el fin de identificar alguna alarma en el funcionamiento del equipo. En este sentido, en la Figura 6.7 se presenta la evolución temporal de la carga de la pala C correspondiente a los datos de testeo, junto con las alarmas identificadas por PC-1cSVM. En primer lugar, puede observarse que algunas alarmas son identificadas justo antes de la parada de marzo de 2015. Esto permite una predicción de la falla. Se visualizan a su vez algunas alarmas algunos meses antes de la parada, pudiendo ser estas falsas alarmas generadas

Figura 6.7: Evolución temporal de la carga en la pala C, junto con las alarmas registradas por el método PC-1cSVM.



por el método.

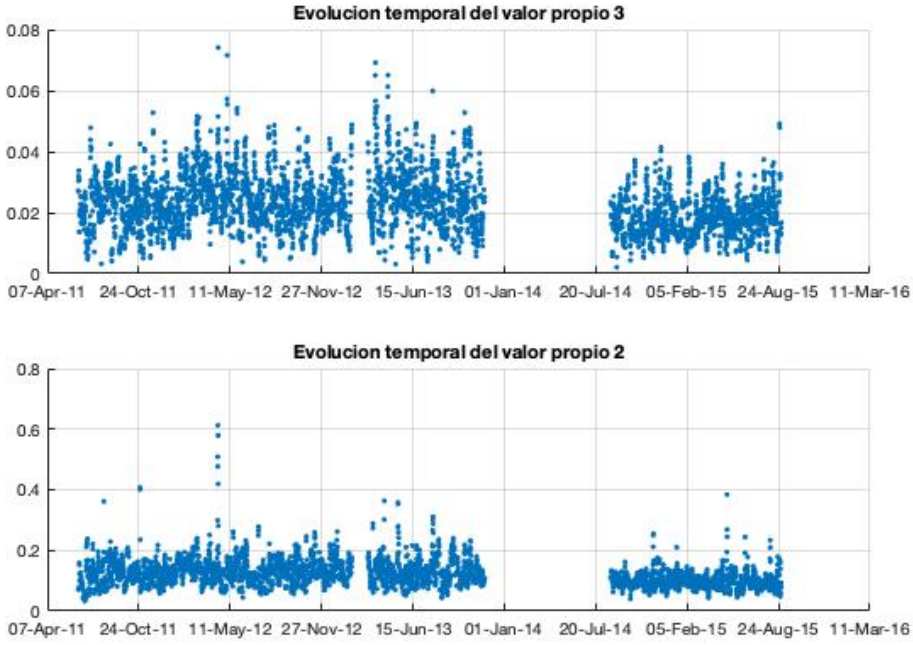
6.2.4. Aplicación para el Aerogenerador D

La aplicación de este método fue realizada para los datos provenientes del aerogenerador D, aunque con ciertas salvedades. Esto último se debe a que, al no tener información sobre la existencia de una falla asociada al funcionamiento del equipo, el objetivo del análisis es otro. Se buscará, a partir de PC, comparar el funcionamiento entre el período asociado a las pre-parada y el asociado a la pos-parada de mantenimiento de noviembre de 2013.

Al igual que en secciones 3.2.3 y 4.2.4, las variables involucradas en este análisis son la potencia, la velocidad de viento y las RPM del rotor.

El parámetro LV en este caso se mantuvo en 432 diezminutales, así como también el paso r se mantuvo en 43 diezminutales. La cantidad de valores propios k empleadas para el análisis también fue igual a 2. En la Figura 6.8 se presenta la evolución temporal de los 2 valores propios asociados a la matriz de correlación de los datos de las ventanas temporales correspondientes al período de testeo; período comprendido entre junio de 2011 y agosto de 2015. Puede

Figura 6.8: Evolución temporal de los 2 valores propios de menor magnitud obtenidos a partir de la matriz de correlación de los datos de las ventanas temporales.

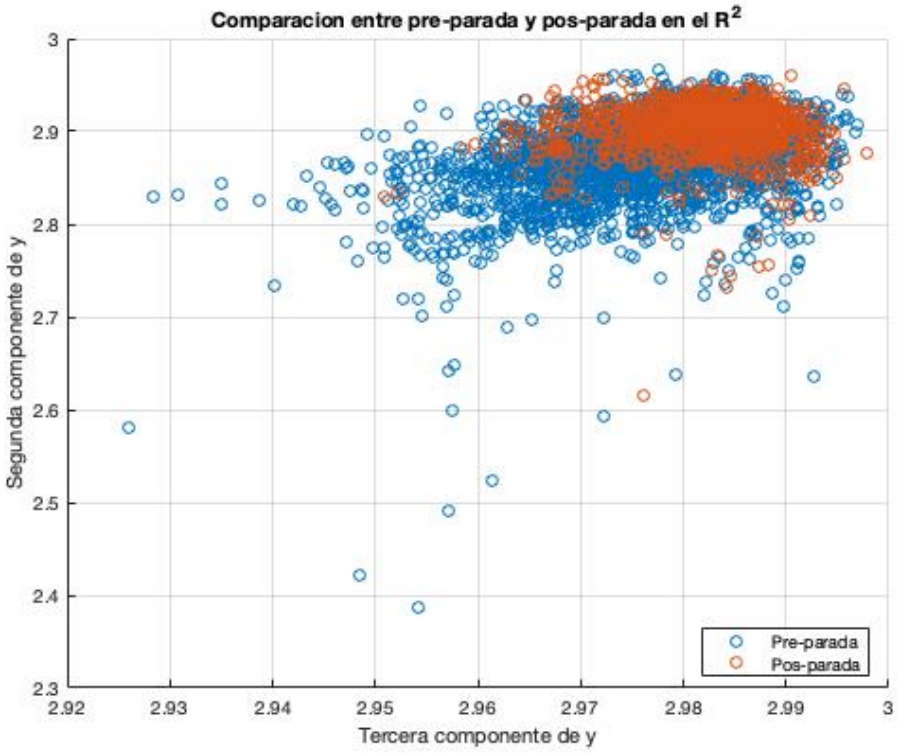


observarse allí una disminución de la variabilidad de los datos luego de la parada de noviembre de 2013, que se traduce a partir de la disminución en la

magnitud de los 2 valores propios.

Lo anterior puede visualizarse de forma alternativa si se proyectan los puntos y, calculados de acuerdo a la Ecuación 6.11, en el plano $\mathbb{R}^2$. En ese sentido, en la Figura 6.9 puede compararse esta representación entre el período asociado a la pre-parada y al de la pos-parada. Se observa que los puntos correspondien-

Figura 6.9: Representación en $\mathbb{R}^2$ de los puntos y de testeo.



tes al funcionamiento posteriores a la parada, se concentran sobre la esquina derecha-superior del panel.

Lo anterior se traduce como una disminución de la variabilidad de los datos luego de la parada de mantenimiento, que se interpreta como una mejora en la performance del equipo.

Las observaciones realizadas sobre las Figuras 6.8 y 6.9 pueden cuantificarse con el test estadístico empleado en la sección 3.2.3. El test de diferencia de medias busca comparar la evolución temporal de cada uno de los k valores propios entre antes y después de la parada. Las Ecuaciones 3.11, 3.12 y 3.13 son empleadas para comparar estas dos ventanas operacionales con un nivel de significancia de 0.05 para ambos test. La cantidad de grados de libertad para el test asociado al valor propio de menor magnitud es de $\nu = 727$, mientras que

el valor de t obtenido es igual a 10. Para el test asociado al otro valor propio, los valores de ν y t obtenidos son de 988 y 14, respectivamente.

Los resultados de los dos test (uno para cada valor propio) muestran que hay suficiente evidencia como para rechazar la hipótesis nula de que los valores medios de ambos períodos son iguales. De esto se puede concluir que parte del método desarrollado en este capítulo permite inferir que el aerogenerador D vio una mejora en su funcionamiento luego de los trabajos de mantenimiento realizados durante la parada de noviembre de 2013.

Capítulo 7

Discusión de resultados

El objetivo de este capítulo es establecer una discusión sobre los resultados abordados en las secciones precedentes. Se buscará comparar los resultados obtenidos para los diferentes métodos y aerogeneradores, así como también elaborar alguna herramienta que permita cuantificar su calidad.

Cabe destacar en primer lugar que hubieron dos casos para los cuales no fueron presentados los resultados; la aplicación de GP para el aerogenerador C y, la aplicación de Cópulas para el aerogenerador B. En el primer caso, el modelo construido no permitió avanzar con la metodología descrita en la sección 3.1.1. Asimismo, el método de Cópulas aplicado al aerogenerador B no permitió predecir la falla asociada al equipo, en el sentido de que los indicadores (R, Q^2 y χ^2) no evidenciaron ninguna anomalía en su evolución.

Con el fin de realizar un primer acercamiento a los resultados obtenidos, aunque de forma cualitativa, en la Tabla 7.1 se presenta un resumen de los aciertos y desaciertos de las aplicaciones de los métodos en los diferentes casos de estudio. Los resultados de la Tabla 7.1 muestran que los métodos fueron capaces de anticipar las fallas en un 90 % de los casos, y con gran claridad en el 50 % del total de los casos. Se destaca que el aerogenerador C no presenta ningún "**Sí**". Esto podría llegar a ser consecuencia de la incertidumbre asociada a la información disponible sobre la falla de este equipo, tal como se mencionó en la sección 2.3.

A partir de las metodologías descritas para los diferentes métodos, se elaboran algunos cuantificadores que buscan determinar la capacidad de detección de los métodos y determinar las falsas alarmas que los mismos proporcionan. Esto último tiene importancia fundamentalmente por las consecuencias

Tabla 7.1: Resumen cualitativo del desempeño de los métodos para los diferentes casos de estudio.
"Sí": El método anticipó la falla.
"No": El método no anticipó la falla.
"**Sí**": El método anticipó la falla con notoria claridad (criterio subjetivo).
Para el caso del aerogenerador D: los códigos son los mismos, pero en relación a la detección de un cambio en el funcionamiento.

Aerogenerador	GP	NSET	Cópulas	PC-1cSVM
A (primera falla)	**Sí**	**Sí**	Sí	**Sí**
A (segunda falla)	Sí	Sí	**Sí**	Sí
B	**Sí**	**Sí**	No	Sí
C	No	Sí	Sí	Sí
D	**Sí**	**Sí**	**Sí**	**Sí**

económicas que tienen las paradas innecesarias. Asimismo, estos indicadores son comparables entre los distintos métodos.

Los resultados obtenidos para los métodos desarrollados en las secciones 3 y 4 están basados en ventanas temporales móviles de longitud LV. Las alarmas obtenidas en los diferentes casos están asociadas a estas ventanas, por lo tanto, es posible definir, para los métodos GP y NSET, el indicador w_1 como el cociente entre la cantidad de ventanas temporales clasificadas como alarmas y la cantidad total de ventanas del período de testeo que fueron evaluadas por el test de comparación de medias descrito en la sección 3.1.1. Así, w_1 indica la tasa de alarmas detectada por estos métodos. Para el caso de Cópulas (sección 5), se define el indicador w_2 como el cociente entre la cantidad de puntos temporales que fueron considerados como anómalos a partir de los indicadores R, Q^2 o χ^2 (definidos en la sección 5.1.1) y, la cantidad total de puntos temporales en los que fueron calculados los indicadores. De esta forma, w_2 también indica la tasa de alarmas en el método de Cópulas. Finalmente, el indicador w_3 corresponde a la tasa de alarmas para del método PC-1cSVM (sección 6). Análogamente a w_1, w_3 se define como el cociente entre la cantidad de ventanas móviles de largo LV clasificadas como alarmas y la cantidad total de ventanas que fueron evaluadas por el método en el período de testeo. De esta forma, los indicadores w_1, w_2 y w_3 permiten cuantificar en qué proporción del período de testeo fueron determinadas las alarmas.

Asimismo, interesa conocer cuáles de las alarmas identificadas por los diferentes métodos están asociadas a una predicción cercana a las respectivas fallas de los aerogeneradores. Por este motivo, es necesario establecer un período de

tiempo previo a la falla en el cual se considera que las alarmas detectadas en este están asociadas directamente a la falla. En este sentido, se opta por tomar los tres meses previos a las respectivas fallas; las alarmas identificadas en estos tres meses son entonces consideradas como una predicción directa de las fallas. Así, se define el indicador x_1 como el cociente entre la cantidad de ventanas temporales clasificadas como alarmas en los tres meses previos a la falla y, la cantidad total de ventanas temporales clasificadas como alarmas; este indicador abarca los métodos GP y NSET. En esta misma línea, se define el indicador x_2, para Cópulas, como el cociente entre la cantidad de puntos temporales que fueron considerados como anómalos en los tres meses previos a las fallas y, la cantidad total de puntos temporales que fueron considerados como anómalos. Finalmente, análogamente a x_1, se define el indicador x_3 para el método PC-1cSVM. De esta forma, los indicadores x_1, x_2 y x_3 son una medida del poder de predicción de los métodos desarrollados para los casos particulares de estudio.

Interesa cuantificar las falsas alarmas generadas, entendidas como las que aparecen fuera de los tres meses previos a la falla. En este sentido, se define y_1 como $1 - x_1$. Cabe destacar que esto tiene sentido siempre y cuando el caso de estudio corresponda a una única falla; en el caso particular del aerogenerador A, el análisis involucra dos fallas distintas, por lo que el indicador y_1 se define como $1 - x_1^{(1)} - x_1^{(2)}$, donde $x_1^{(1)}$ corresponde al valor del indicador x_1 de la primera falla y, $x_1^{(2)}$ al de la segunda. Análogamente, se definen y_2 y y_3 considerando la misma observación anterior. De esta forma, los indicadores y_1, y_2 y y_3 permiten cuantificar las falsas alarmas generadas por los diferentes métodos en los casos particulares de estudio.

Finalmente, con el fin de cuantificar el alcance de cada una de las predicción, se define un cuarto tipo de indicadores. z_1 es la cantidad de días que hay entre la falla y la detección de la primera alarma, dentro de los tres meses establecidos, para los métodos GP y NSET. Análogamente, se definen z_2 y z_3, para los métodos Cópulas y PC-1cSVM, respectivamente. De esta forma, los indicadores z_1, z_2 y z_3 permiten cuantificar el alcance de las predicciones.

Por cómo fueron desarrollados los cuatro métodos, resulta inmediato observar que los indicadores asociados a GP, NSET y PC-1cSVM son totalmente comparables. Por otra parte, los indicadores asociados a Cópulas no son comparables con los anteriores; teniendo en cuenta que el paso temporal utilizado en este método es de un orden significativamente mayor al empleado para los

otros, por lo que los valores de los indicadores estarán condicionados a esta diferencia.

Antes de presentar los resultados de los cuantificadores introducidos, cabe recordar que es posible realizar la comparación para los aerogeneradores A, B y C; el abordaje de los resultados obtenidos para el aerogenerador D debe realizarse de forma particular.

En primera instancia, en la Tabla 7.2 se presentan los cuantificadores w_1, w_2 y w_3, en términos porcentuales, para los casos particulares de estudio de los aerogeneradores A, B, y C. Los valores reflejados en la tabla representan la

Tabla 7.2: Comparación de la proporción de alarmas generadas en los diferentes casos de estudio.

Aerogenerador	GP (w_1 %)	NSET (w_1 %)	Cópulas (w_2 %)	PC-1cSVM (w_3 %)
A	5.78 %	9.11 %	55.56 %	3.04 %
B	8.99 %	3.10 %	Sin resultados	2.52 %
C	Sin resultados	0.24 %	18.52 %	1.98 %

tasa de alarmas detectadas por los diferentes métodos en los distintos casos de estudio. Se destaca en primer lugar la constancia en la tasa del cuantificador w_3, cuyo valor se mantiene entre aproximadamente 2 % y apenas por encima de 3 % para los tres casos; caracterizando así al método PC-1cSVM por su baja generación de alarmas. Asimismo, el indicador w_1 para el caso de NSET varía entre una tasa prácticamente nula y casi 10 %. Para el caso de GP, los dos resultados que se obtuvieron están comprendidos dentro de un rango de 4 % (5 % a 9 %, aproximadamente). Finalmente, es notorio que el indicador w_2 toma valores muy distintos en los tres casos de estudio; esto se debe a que el paso temporal de los resultados de Cópulas es grande en comparación con el de los demás métodos.

En la Tabla 7.3 se presentan los indicadores x_1, x_2 y x_3, en términos porcentuales, para los casos particulares de estudio de los aerogeneradores A, B y C. Los datos de la tabla muestran qué proporción de las alarmas detectadas en cada caso corresponden a una predicción de la falla que sufrieron los aerogeneradores. Se destaca en particular el alto poder de predicción de los métodos GP y PC-1cSVM para los aerogeneradores A y B. Por otra parte, el valor más bajo observado (sin considerar el caso de x_2 para el aerogenerador B) corresponde al cuantificador x_1 asociado a NSET para el caso de la segunda falla del aerogenerador A. Asimismo, el valor más alto también corresponde a

Tabla 7.3: Comparación de la proporción de alarmas generadas, asociadas a la predicción de las fallas, en los diferentes casos de estudio.

Aerogenerador	GP (x_1 %)	NSET (x_1 %)	Cópulas (x_2 %)	PC-1cSVM (x_3 %)
A (primera falla)	45.55 %	30.75 %	20.00 %	38.89 %
A (segunda falla)	51.22 %	8.25 %	60.00 %	50 %
B	18.29 %	78.30 %	Sin resultados	21.74 %
C	Sin resultados	48.91 %	20.00 %	31.58 %

NSET, asociado al caso del aerogenerador B. Finalmente, cabe mencionar la particularidad de los valores del indicador x_2, ya que la identificación de las alarmas fue realizada de forma cualitativa, a diferencia de los demás casos.

Cabe resaltar que los indicadores anteriores fueron obtenidos mediante la selección subjetiva de un período de tres meses previo a la falla. Esto no necesariamente puede ser así en todos los casos, ya que el aerogenerador puede presentar un desgaste paulatino en su performance, comenzado previamente a estos tres meses, que puede traducirse en anomalías. Sin embargo, es una herramienta que permite realizar una comparación cruzada para los métodos propuestos.

Con el fin de cuantificar las falsas alarmas generadas en los casos de estudio, se presentan en la Tabla 7.4 los cuantificadores y_1, y_2 y y_3, en términos porcentuales, para los casos particulares de estudio de los aerogeneradores A, B y C. Estos valores presentan una tasa de las falsas alarmas generadas por

Tabla 7.4: Comparación de la proporción de falsas alarmas generadas en los diferentes casos de estudio.

Aerogenerador	GP (y_1 %)	NSET (y_1 %)	Cópulas (y_2 %)	PC-1cSVM (y_3 %)
A	3.23 %	61.00 %	20.00 %	11.11 %
B	81.71 %	21.70 %	Sin resultados	78.26 %
C	Sin resultados	51.09 %	80.00 %	68.42 %

los distintos métodos para los diferentes casos de estudio. Se destaca que los mejores resultados obtenidos, en el sentido de escasez de falsas alarmas, están asociados al método GP para el caso del aerogenerador A.

Finalmente, en la Tabla 7.5 se presentan los cuantificadores z_1, z_2 y z_3, medidos en días, para los casos particulares de estudio de los aerogeneradores A, B, y C. Los valores presentados muestran que el menor alcance de predicción se da para el caso de la primera falla del aerogenerador A con el método PC-1cSVM. Asimismo, la misma falla tiene solo 10 días de anticipación en el

Tabla 7.5: Comparación de los alcances de las predicciones (en días) asociadas a las fallas en los diferentes casos de estudio.

Aerogenerador	GP (z_1)	NSET (z_1)	Cópulas (z_2)	PC-1cSVM (z_3)
A (primera falla)	10	21	20	5
A (segunda falla)	90	37	60	71
B	90	88	Sin resultados	81
C	Sin resultados	56	30	25

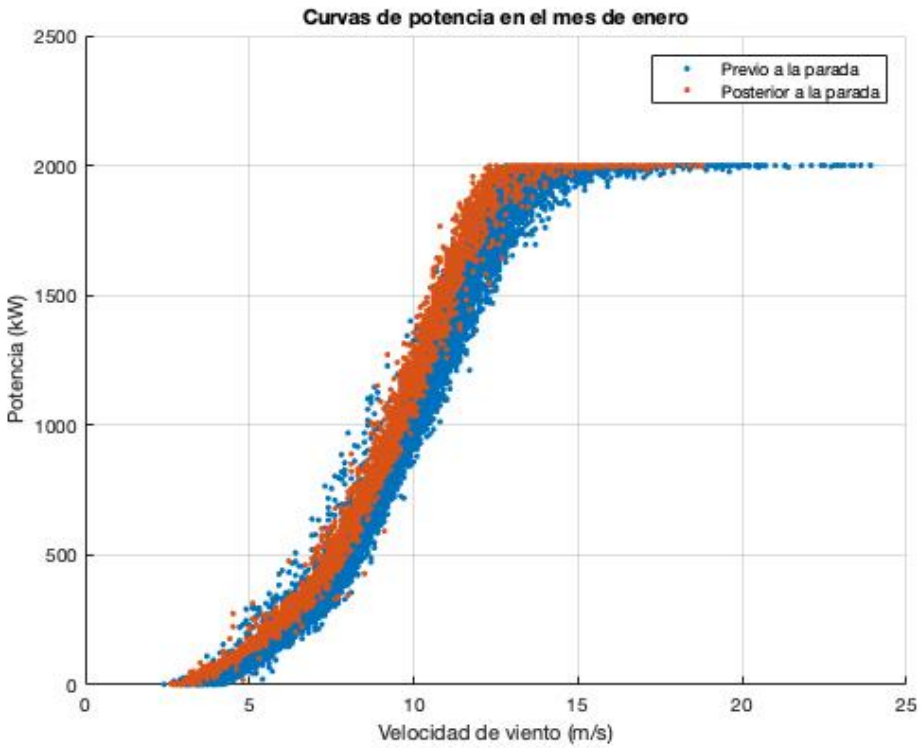
caso de GP. En los demás casos, se tiene al menos 3 semanas de antelación de las fallas, considerándose este un buen alcance de predicción desde el punto de vista operacional. Para tener una idea más clara sobre la calidad de la predicción, puede ser conveniente considerar valores de más de un cuantificador, por ejemplo, los de x y z.

Los valores de los cuantificadores presentados permiten realizar una comparación entre los resultados obtenidos, aunque deben tenerse ciertos cuidados en las conclusiones a extraerse, por las observaciones realizadas anteriormente. Podemos concluir que los métodos GP, Cópulas y PC-1cSVM lograron una buena predicción de las fallas del aerogenerador A, con un bajo nivel de falsas alarmas; por otro lado, NSET genera una buena predicción de la primera falla y no de la segunda, generando una tasa considerable de falsas alarmas y un alcance de predicción de más de tres semanas. Para el caso del aerogenerador B, NSET logra una muy buena predicción de la falla, con una baja cantidad de falsas alarmas y un alto alcance; mientras que los otros métodos proporcionan una baja proporción de los indicadores x, generando esto un incremento en las falsas alarmas; con excepción de Cópulas, para el cual no se obtuvieron resultados. Finalmente, no se logró un modelo viable mediante la técnica de GP en el caso del aerogenerador C. Por otra parte, los otros tres métodos generaron una predicción de calidad media para las fallas de este equipo, teniendo una tasa de falsas alarmas superior al 50 % en los tres casos. Las conclusiones anteriores revelan que ninguno de los cuatro métodos destaca frente a los demás en cuanto a sus buenos resultados, aunque NSET y PC-1cSVM fueron los dos que generaron una predicción al menos aceptable en todos los casos de estudio.

Por otra parte, es necesario interpretar las conclusiones parciales extraídas de los diferentes métodos para el caso del aerogenerador D. En los cuatro casos pudo observarse un cambio en el funcionamiento del equipo luego de la parada ocurrida. Con el fin de validar estas conclusiones, se procede con la implemen-

tación de una técnica que permita evidenciar esto. Más precisamente, se evalúa la evolución de la curva de potencia distinguiendo los períodos pre-parada y pos-parada del aerogenerador D, separando el análisis para los distintos meses del año. En este sentido, y a modo ilustrativo, en las Figuras 7.1 y 7.2 se presentan las curvas de potencia del aerogenerador D correspondientes a los datos de enero y junio, respectivamente, distinguiendo los puntos de operación asociados al período pre-parada y pos-parada. En ambas figuras puede

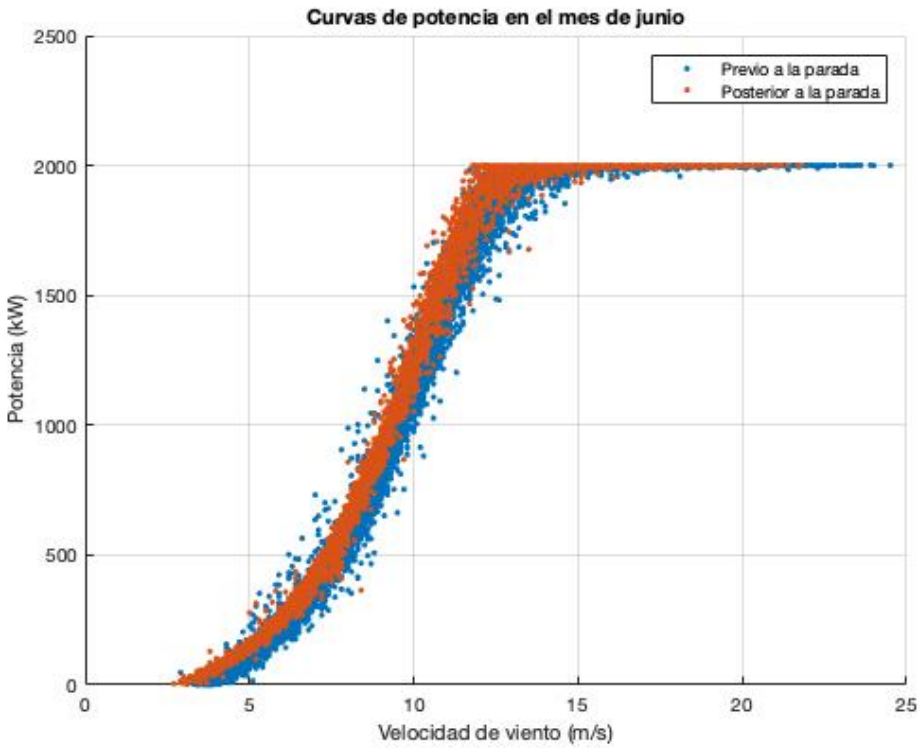
Figura 7.1: Curva de potencia del aerogenerador D, para los períodos pre-parada y pos-parada, de los datos correspondientes al mes de enero.



observarse que los puntos correspondientes al segundo período se ubican en la parte superior de la curva, traduciéndose esto en un mejor funcionamiento del equipo. Los restantes diez meses presentan un comportamiento similar en lo que respecta a la ubicación de los puntos en la curva de potencia, por lo que podría generalizarse la conclusión anterior para todos los meses de operación.

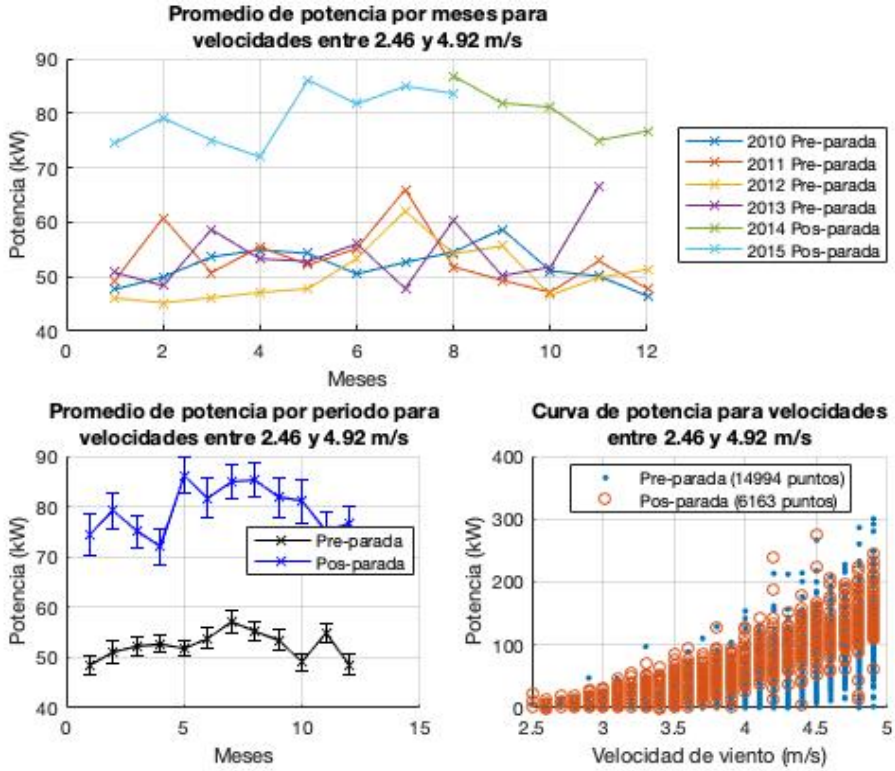
Con el fin de analizar más detenidamente este comportamiento de la curva de potencia, se seleccionaron bins de velocidad de viento, de forma de ver localmente la evolución de la curva de potencia para estos dos períodos. En este sentido, en las Figuras 7.3 y 7.4 se presentan las evoluciones temporales de los

Figura 7.2: Curva de potencia del aerogenerador D, para los períodos pre-parada y pos-parada, de los datos correspondientes al mes de junio.



valores medios de potencia para dos bins de velocidad de la curva de potencia (con distintos niveles de producción), la evolución temporal de los valores medios de potencia para esos bins de velocidad, distinguiendo por períodos pre-parada y pos-parada (incluyendo sus respectivos intervalos de confianza al 95 %) y, el tramo de curva de potencia asociado a esos bins de velocidad. En la Figura 7.3 puede observarse cómo la producción aumenta significativamente en los años 2014 y 2015, correspondientes al período pos-parada. Esto se ratifica en lo presentado en el panel inferior-izquierdo, donde los valores medios de la potencia generada por el equipo son superiores para el período pos-parada. De la Figura 7.4, donde el bin de velocidades de viento consideradas corresponde a la parte de alta producción de la parte cuasi-lineal de la curva de potencia, pueden extraerse las mismas conclusiones anteriores. Además, tal como era de esperarse, se observa un aumento en la producción para los meses de invierno; producto de la relación entre la densidad del aire y la temperatura del mismo. Las figuras correspondientes a los demás bins de velocidades reflejan características similares a las mencionadas anteriormente, por lo que las con-

Figura 7.3: Aerogenerador D: Evolución temporal de los valores medios de potencia comprendidos en el bin de velocidad (2.46, 4.92) m/s (arriba); evolución temporal de los valores medios de potencia, para los períodos pre-parada y pos-parada, en el bin de velocidad (2.46, 4.92) m/s (abajo-izquierda); curva de potencia para los períodos pre-parada y pos-parada en el bin de velocidad (2.46, 4.92) m/s (abajo-derecha).

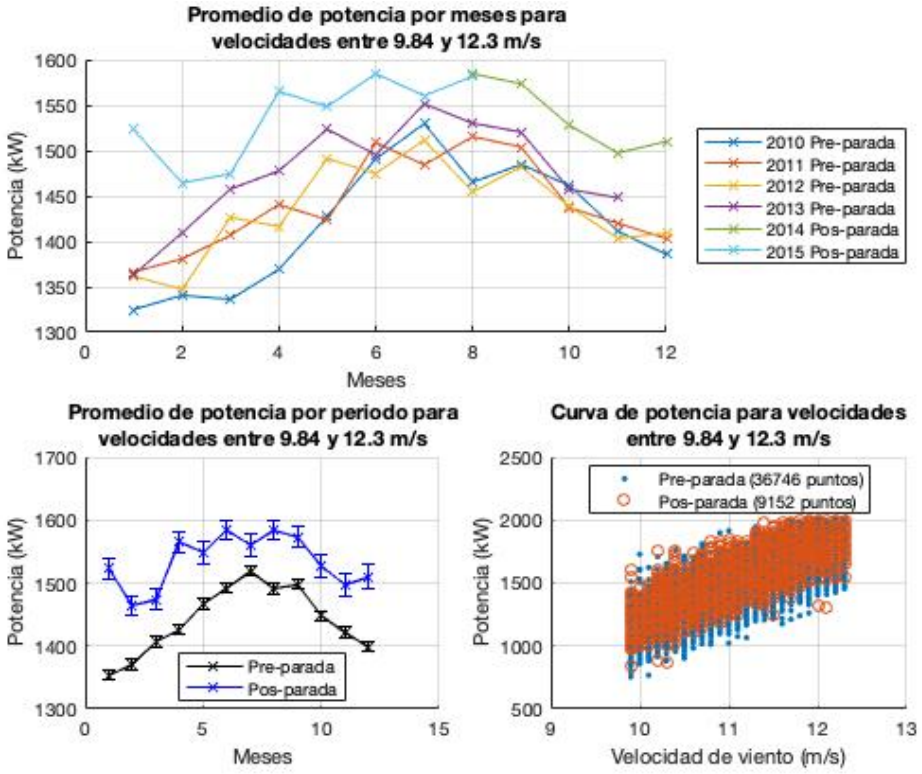


clusiones previas pueden extenderse para toda la parte cuasi-lineal de la curva de potencia.

De esta forma, es posible validar las conclusiones obtenidas para los métodos aplicados al caso del aerogenerador D, concluyendo finalmente que el cambio en el funcionamiento que presentó el equipo luego de la parada corresponde a una mejora en su performance.

Así, los resultados abordados por los diferentes métodos en los distintos casos de estudios fueron cuantificados y comparados, sin distinguir de forma evidente una capacidad predictiva superior de uno frente a otro en términos generales. Asimismo, las conclusiones determinadas para los distintos métodos en el caso del aerogenerador D fueron reafirmadas mediante una técnica elemental de comparación de períodos de funcionamiento.

Figura 7.4: Aerogenerador D: Evolución temporal de los valores medios de potencia comprendidos en el bin de velocidad (9.84, 12.3) m/s (arriba); evolución temporal de los valores medios de potencia, para los períodos pre-parada y pos-parada, en el bin de velocidad (9.84, 12.3) m/s (abajo-izquierda); curva de potencia para los períodos pre-parada y pos-parada en el bin de velocidad (9.84, 12.3) m/s (abajo-derecha).



Capítulo 8

Conclusiones

Se presentó en esta tesis la construcción y aplicación de cuatro herramientas matemáticas destinadas a la detección de anomalías en el funcionamiento de aerogeneradores, basadas en el registro de datos en sistema SCADA. Las mismas fueron aplicadas en hasta cuatro casos diferentes de estudio, teniendo cada uno sus particularidades a la hora del análisis. Los métodos mostraron ser capaces tanto de predecir fallas específicas asociadas a la operación de los equipos, como de identificar períodos de cambios en el funcionamiento de los mismos. Finalmente, los cuatro métodos fueron sometidos a comparación en la sección 7, evidenciando que ninguno de ellos generó un mayor poder predictivo, en términos generales, sobre otro.

En algunos casos puntuales, algunos métodos no permitieron avanzar con la metodología propuesta para los fines establecidos. En los demás casos, condujeron a resultados aceptables y satisfactorios, dependiendo del caso de estudio, en cuanto a la predicción de las respectivas fallas. Se pudo detectar la posibilidad de una futura falla en el 90 % de los casos, y en la mitad del total de los casos esta detección fue muy clara. Todos los resultados obtenidos estuvieron sujetos a determinadas hipótesis de trabajo establecidas a lo largo del desarrollo y, por lo tanto, también las conclusiones extraídas de estos resultados; desde la información sobre los orígenes de las fallas de los aerogeneradores estudiados no siempre disponible, hasta la selección de determinados parámetros asociados a la metodología de implementación de los distintos métodos. Si bien estas hipótesis de trabajo generan cierto nivel de incertidumbre en los resultados obtenidos, fueron asumidas con la mayor rigurosidad posible.

En particular, en la sección 7 se seleccionó un período de tres meses pre-

vio a cada falla, a partir del cual fueron cuantificados los resultados obtenidos en todos los métodos. Este parámetro debe ser escogido a partir de algunos factores que escapan al alcance de este trabajo, como por ejemplo, los costos involucrados en hacer una parada de mantenimiento innecesaria, asociada a una falsa alarma; cuál es el tiempo entre paradas programadas para determinadas tareas de mantenimiento; cuáles son los costos que implican determinadas fallas, entre otros. Si bien esos tópicos no se trataron en esta tesis, la información presentada en las Figuras 1.7 y 1.8 dan un indicio para profundizar sobre este terreno.

Los métodos GP, NSET y PC-1cSVM pueden ser empleados con una cantidad de variables SCADA indeterminada. Si bien en los casos de estudio las aplicaciones fueron llevadas a cabo con una dimensionalidad relativamente baja, esta podría aumentarse con el fin de involucrar nueva información. Esto podría llegar a potenciar el poder descriptivo de los métodos y, en consecuencia, su poder predictivo.

Resulta de interés explicar cuál es la virtud, desde el punto de vista operacional, que proporcionan estos métodos. Las cuatro herramientas presentadas están basadas en un conjunto histórico de datos de entrenamiento, el cual debe ser saludable para que los resultados conduzcan a interpretaciones acordes a lo esperado. En ese sentido, el operador debe disponer de datos saludables que sirvan como antecedentes para ser utilizados como datos de entrenamiento y validación. A partir de esto, las condiciones están dadas para la construcción de los modelos presentados y, así, poder evaluar en tiempo real la condición del aerogenerador, identificando eventualmente la presencia de anomalías.

A su vez, desde el punto de vista operacional, el operador podría tener que tomar una decisión ante la presencia de una alarma generada por un determinado método. En este caso, bien podría estar ante una inminente falla, o bien estar presenciando una falsa alarma. Esta decisión está estrechamente vinculada a los costos mencionados anteriormente. Sin embargo, una posible alternativa podría ser la implementación de los cuatro métodos en forma simultánea; ante una primera alarma, el operador debería estar alerta y analizar la evolución en tiempo real de los resultados de los otros métodos, basándose fundamentalmente en los resultados obtenidos de los indicadores expuestos en la sección 7. Esta alternativa no es más que la combinación de los cuatro métodos, generando un nuevo método más robusto que busca mejorar el poder predictivo que presentan individualmente.

Resulta evidente que en los casos de estudio presentados el análisis estuvo dirigido a fallas donde la información sobre su origen fue considerada a priori. En la operación cotidiana de un aerogenerador, esta información claramente no está disponible. Por lo tanto, a partir de un conjunto de datos saludables y de un método en particular, podrían generarse un conjunto de modelos donde cada uno de estos comprenda distintas variables predictoras, enfocadas al análisis de distintos componentes del aerogenerador.

Asimismo, resulta de interés destacar que si bien los estudios llevados a cabo en este trabajo comprendieron aerogeneradores considerados de forma individual, en general estos se encuentran dispuestos en parques eólicos, donde la interacción entre ellos afecta su funcionamiento. Esto debe tenerse en cuenta tanto a la hora de comparar resultados de métodos aplicados a dos aerogeneradores cercanos, como al momento de utilizar datos de un aerogenerador para modelar el funcionamiento de otro; tal como se desarrolló en el caso del aerogenerador A.

En relación a lo anterior, es natural preguntarse si las conclusiones extraídas para el caso del aerogenerador D pudieran haber sido anticipadas antes de la parada. Como se vio en las secciones precedentes, los resultados presentados no evidenciaron un cambio en el funcionamiento en el período previo a la parada. Por lo tanto, una posible respuesta es que un operador podría haber comparado el funcionamiento de este aerogenerador con el de equipos de iguales características cercanos a este -teniendo en cuenta las consideraciones expuestas en el párrafo anterior-, comparando la evolución del estado de condición de ambos equipos.

Como fue expuesto en el Capítulo 1, existe una amplia metodología para tratar el tema de predicción de fallas en aerogeneradores. Los métodos propuestos en esta tesis son tan solo un subconjunto de las herramientas existentes, aunque, sin ser necesariamente mejores en algún sentido, tienen un importante grado de versatilidad entre ellos. Esto deja abierta la línea de estudio de nuevas metodologías de aplicaciones puntuales para aerogeneradores instalados en Uruguay.

Se concluye entonces que la construcción y la aplicación de los cuatro métodos presentados, basados fundamentalmente en trabajos previos de diversos autores, permitieron el abordaje a la detección de anomalías en el funcionamiento de aerogeneradores, mostrando aceptables resultados y, dejando en evidencia la posibilidad de profundizar en futuras líneas de trabajo relacionadas a la

temática.

Referencias bibliográficas

Amirat, Y., Benbouzid, M. E. H., Al-Ahmar, E., Bensaker, B., and Turri, S. (2009). A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. *Renewable and Sustainable Energy Reviews, Elsevier*, 3(9):2629–2636.

Astolfi, D., Castellani, F., and Terzi, L. (2014). Fault prevention and diagnosis through scada temperature data analysis of an on-shore wind farm. *Diagnostyka*, 15(2):71–78.

Beltrán, J., Guerrero, J. J., Melero, J. J., and Lombart, A. (2012). Detection of nacelle anemometer faults in wind farm minimizing the uncertainty. *Wind Energy*, 16(1):939–952.

Bertelé, M., Bottasso, C. L., and Cacciola, S. (2018). Automatic detection and correction of pitch misalignmen in wind turbine rotors. *European Academy of Wind Energy*, 3(1):791–803.

Black, C. L., Uhrig, R. E., and Hines, J. W. (1998). System modeling and instrument calibration verification with a nonlinear state estimation technique. *Proceedings of the Maintenance and Reliability Conference*, 1(1):1–15.

Catmull, S. (2010). Self-organising map based condition monitoring of wind turbines. *European Wind Energy Association*, 1(1):1–6.

Ebden, M. (2015). Gaussian processes: A quick introduction. *arXiv*, 2(1505.02965v2):1–13.

Feng, Y., Qiu, Y., Crabtree, C. J., Long, H., and Tavner, P. J. (2013). Monitoring wind turbine gearboxes. *Wind Energy*, 1(1):1–13.

Gill, S., Stephen, B., and Galloway, S. (2012). Wind turbine condition assessment through power curve copula modelling. *IEEE Transactions on sustainable energy*, 3(1):94–101.

Guo, P., Infield, D., and Yang, X. (2012). Wind turbine generator condition-monitoring using temperature trend analysis. *IEEE Transactions on Sustainable Energy*, 3(1):124–133.

Hameed, Z., Hong, Y., Cho, Y., Ahn, S., and Song, C. (2007). Condition monitoring and fault detection of wind turbines and related algorithm: A review. *Renewable and Sustainable Energy Reviews*, 13(1):1–39.

Herp, J., Pedersen, N. L., and Nadimi, E. S. (2016). Wind turbine performance analysis based on multivariate higher order moments and bayesian classifiers. *Cotrol Engineering Practice*, 49(1):204–211.

Herzog, J. P., Wegerich, S. W., Gross, K. C., and Bockhorst, F. K. (1998). Mset modeling of crystal river-3 venturi flow meters. *6th International Conference on Nuclear Engineering*, 6(1):1–13.

International Electro-Technical Comission, W. T. (2016). Part 12-1: Power performance measurements of electricity producing wind turbines. Technical Report 61400-12-1, IEC, Washington, D.C.

Jackson, J. E. (1991). *A User's Guide to Principal Components*. Wiley Series in Probability and Statistics, New York.

Jia, X., Jin, C., Buzza, M., Wang, W., and Lee, J. (2016). Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves. *Renewable Energy*, 99(1):1191–1201.

Kim, K., Parthasarathy, G., Uluyol, O., Foslien, W., Sheng, S., and Fleming, P. (2011). Use of scada data for failure detection in wind turbines. *2011 Energy Sustainability Conference and Fuell Cell Conference*, 1(1):1–9.

Laouti, N., Sheibat-Othman, N., and Othman, S. (2011). Support vector machines for fault detection in wind turbines. *Proceedings of the 18th World Congress The International Federation of Automatic Control*, 18(1):7067–7072.

Lapira, E., Brisset, D., Ardakani, H. D., Siegel, D., and Lee, J. (2012). Wind turbine performance assessment using multi-regime modeling approach. *Renewable Energy*, 45(1):86–95.

Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30(1):452–460.

Ma, J. and Perkins, S. (2003). Time-series novelty detection using one-class support vector machines. *Proceedings of the International Joint Conference on Neural Networks*, 3(1):1741–1745.

Martínez-Rego, D., Fontela-Romero, O., and Alonso-Betanzos, A. (2011). Power wind mill fault detection via one-class ν-svm vibration signal analysis. *Proceedings of International Joint Conference on Neural Networks*, 1(1):511–518.

Mittelmeier, N. and Kuhn, M. (2018). Determination of optimal wind turbine alignment into the wind and detection of alignment chenges with scada data. *European Academy of Wind Energy*, 3(1):395–408.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, Portland, USA.

Pandit, R. K. and Infield, D. (2018). Scada based wind turbine anomaly detection using gaussian process (gp) models for wind turbine condition monitoring purposes. *IET Renewable Power Generation*, 12(2):1249–1255.

Preisendorfer, R. W. (1988). *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, New York.

Purarjomandlangrudi, A. (2014). Application of machine learning technique in wind turbine fault diagnosis. M.Sc. dissertation, Queensland University of Technology, Brisbane, Australia.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT, London, England, 1 edition.

Sawilowski, S. S. (2002). Fermat, schubert, einstein, and behrens-fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$. *Journal of Modern Applied Statistical Methods*, 1(2):461–472.

Sawyer, S., Fried, I., Shukla, S., and Liming, Q. (2019). Global wind report 2018. *Global Wind Energy Council*, 1(1):1.

Schlechtingen, M., Santos, I. F., and Achiche, S. (2009). Using data-mining approaches for wind turbine power curve monitoring: A comparative study. *IEEE Transactions on Sustainable Energy*, 4(3):671–679.

Scholkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471.

Sklar, M. J. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8(1):229–231.

Skrimpas, G. A., Sweeney, C. W., Marhadi, K. S., Jensen, B. B., Mijatovic, N., and Hoboll, J. (2014). Detection of wind turbine power performance abnormalities using eigenvalue analysis. *Annual Conference of the Prognostics and Health Managment Society*, 11(1):1–7.

Stephen, B., Galloway, S., McMillan, D., Hill, D. C., and Infield, D. (2011). A copula model of wind turbine performance. *IEEE Transactions on Power Systems*, 26(2):965–966.

Sánchez, L. and Couso, I. (2011). Singular spectral analysis of ill-known signals and its application to predictive maintenance of windmills with scada records. *Soft Comput*, 16(1):755–768.

Tautz-Weinert, J. and Watson, S. J. (2016). Using scada data for wind turbine condition monitoring - a review. *IET Renewable Power Generation*, 11(4):382–394.

Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T. A., and Ekemb, G. (2014). Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, 7(1):2595–2630.

von Storch, H. and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, New York.

Wang, Y. and Infield, D. (2013). Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring. *IET Renewable Power Generation*, 7(4):350–358.

Wang, Y., Infield, D. G., Stephen, B., and Galloway, S. J. (2014). Copula based model for wind turbine power curve outlier rejection. *Wind Energy*, 17(11):1677–1688.

Wilkinson, M., Darnell, B., van Delft, T., and Harman, K. (2014). Comparision of methods for wind turbine condition monitoring with scada data. *IET Renewable Power Generation*, 8(4):390–397.

Wilks, D. S. (2011). *Statistical methods in the armospheric sciences*. Elsevier, UK.

Wodecki, J., Stefaniak, P., Sawicki, M., and Zimroz, R. (2017). Application of independent component analysis in temperatur data analysis for gearbox fault detection. *Cyclostationarity: Theory and Methods III. Applied Condition Monitoring*, 6(1):187–198.

Yang, W., Court, R., and Jiang, J. (2013). Wind turbine condition monitoring by the approach of scada data analysis. *Renewable Energy*, 53(1):365–376.

Zhang, Z. and Kusiak, A. (2012). Monitoring wind turbine vibration based on scada data. *Journal of Solar Energy Engineering*, 134(1):1–12.

Zhao, Y., Li, D., Dong, A., Kang, D., Lv, Q., and Shang, L. (2017). Fault prediction and diagnosis of wind turbine generators using scada data. *MDPI - Energies*, 10(1210):1–17.