



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Predicción probabilística de corto plazo en generación eólica

Yuber Samir Sánchez Rosas

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica
Bogotá D.C., Colombia
2018

Predicción probabilística de corto plazo en generación eólica

Short term wind power forecasting

Yuber Samir Sánchez Rosas

Tesis presentada a la Universidad Nacional de Colombia como requisito para optar el título de
Maestría en Ingeniería.

Maestría en Ingeniería Eléctrica

Dirigido por:

Ph.D. Fabio Andrés Pavas Martínez

1 de junio de 2018

Línea de Investigación:

Generación eólica, predicción probabilística,
modelamiento matemático

Grupo de Investigación:

Programa de Adquisición y Análisis de Señales PAAS - UN

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica
Bogotá D.C., Colombia

2018

“ Todo tiene su momento oportuno; hay un tiempo para
todo lo que se hace bajo el cielo (...)” Eclesiastés III

Resumen

A través de este proyecto se describe el uso del clasificador del K_p -ésimo vecino más cercano KNN y un estimador del kernel con el fin de predecir la potencia eólica del viento. El modelo es capaz de predecir valores de la potencia eólica generada a la salida de la granja con una antelación de hasta 48h. Los datos usados para el estudio de caso tienen una resolución por cada hora. Estos datos fueron obtenidos de 10 granjas ubicadas en Australia, a dos distintas alturas: 10m y 100m. El modelo desarrollado es un modelo de un único paso. En la primera Fase, el KNN es usado para filtrar los datos. Una vez los datos han sido filtrados y normalizados, el modelo predice escenarios de generación a través del estimador del kernel. Finalmente, algunos indicadores de desempeño son usados para medir los resultados obtenidos respecto del valor real, tales como: EM, EMAN, EPMA y DEE.

Palabras clave: Predicción de la generación, K_p -ésimo vecino más cercano, Predicción, Potencia eólica, Estimador del kernel, EM, EPMA y DEE.

Abstract

This project describes the use of KNN classifier with Kernel Density Estimation (KDE) models to forecasting the wind power. This approach is used to predict hourly values of wind power for horizons of up to 48h. The data used is hourly observation from 10 wind farms at two different heights, 10m and 100m, in Australia. This approached is a one-stage method where first a KNN classifier is applied to the raw data to clean it. Once the dataset has been cleaned and normalized, the forecast of the normalized wind power is calculated using KDE models. MAE, MAPE and SDE performance indices are used to find the performance of this model.

Index Terms: Generation Prediction, KNN, Forecasting, Wind Power, KDS, MAE, MAPE and SDE.

Tabla de Contenido

Resumen	vii
Lista de Tablas	xi
Lista de Figuras	xiii
Lista de Símbolos	xv
1 Introducción	1
1.1 Primeros Modelos	4
1.2 Años 90s	5
1.3 2000 - 2006	8
1.4 2006 - 2016	12
1.4.1 Predicción numérica del clima	12
1.4.2 Métodos estadísticos	13
1.4.2.1 Enfoque estadístico convencional	13
1.4.3 Métodos híbridos	15
1.5 Conclusiones	15
2 Propuesta de Investigación	17
2.1 Conceptos básicos	18
2.1.1 Horizonte de predicción	19
2.1.2 Definición conceptual de modelos en predicción eólica de corto plazo .	19
2.1.3 Formulación del problema de la predicción de la generación eólica . .	20
2.1.4 Métodos de referencia	22
2.1.5 Modelo numérico del clima	23
2.1.6 Enfoque estadístico	24
2.1.6.1 Clasificador no paramétrico	25
2.1.6.2 Ajuste a través de Kernel Density Estimation	25
2.1.7 Índices de desempeño	27
2.2 Objetivo de la Tesis	28
2.2.1 Objetivos Específicos	28
2.3 Solución propuesta	28
2.4 Estructura del documento	29

2.5	Conclusión	32
3	Algoritmo de predicción: Árboles de decisión	33
3.1	Selección del algoritmo	34
3.2	Fase I	34
3.3	Fase II	37
3.4	Fase III	38
3.5	Conclusión	42
4	Caso de estudio	43
4.1	Descripción del conjunto datos	43
4.2	Predicción	47
4.3	Desempeño del modelo	54
4.4	Análisis	58
4.5	Conclusión	62
	CONCLUSIONES	63
	TRABAJO FUTURO	65
	Participación Eventos Nacionales e Internacionales	67
	Apéndice A	69
	Apéndice B	72
	Apéndice C	75
	Apéndice D	78
	Bibliografía	82

Lista de Tablas

2-1	Horizontes de predicción	19
2-2	Funciones típicas de peso.	27
3-1	Tiempos de ejecución del algoritmo en segundos.	34
3-2	Algoritmos de predicción.	35
4-1	Estadísticas descriptivas.	44
4-2	Indices de desempeño.	55
4-3	Estadísticas descriptivas 4-15	59
4-4	Estadísticas descriptivas 4-16	60

Lista de Figuras

2-1	Variación, μ y σ con el tiempo para el grupo datos de entrenamiento.	21
2-2	Funciones de peso	26
2-3	Esquema del modelo implementado.	31
3-1	Descomposición del vector de viento u en componentes zonal u_x y meridional u_y , este-oeste, norte-sur, respectivamente.	36
3-2	Fase II representación esquemática.	37
3-3	Representación esquemática Fase I, II y III del modelo.	41
4-1	Rosas del viento a 10m.	44
4-2	Rosas del viento a 100m.	44
4-3	Serie de tiempo de potencia eólica. En azul, conjunto de datos de <i>entrenamiento</i> del modelo. En rojo, conjunto de datos de <i>validación</i> del modelo. . .	45
4-4	Potencia eólica vs. Velocidad.	46
4-5	Ubicación virtual de las granjas eólicas.	47
4-6	Veinte escenarios para siete días con frecuencia de actualización cada 48h. El área sombreada envolvente en la figura inferior fue construida tomando el valor mínimo y máximo horario de los escenarios	48
4-7	Veinte escenarios para siete días con frecuencia de actualización cada 24h. El área sombreada envolvente en la figura inferior fue construida tomando el valor mínimo y máximo horario de los escenarios.	49
4-8	Veinte escenarios para siete días con frecuencia de actualización cada 12h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.	50
4-9	Veinte escenarios para siete días con frecuencia de actualización cada 6h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.	51
4-10	Veinte escenarios para siete días con frecuencia de actualización cada 4h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.	52
4-11	Veinte escenarios para siete días con frecuencia de actualización cada 2h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.	53
4-12	Distribución diaria de los errores.	56

4-13	Distribución horaria de los errores.	57
4-14	EMA para el modelo de predicción probabilística de la generación eólica propuesto.	58
4-15	Funciones de probabilidad acumuladas	59
4-16	Funciones de probabilidad acumuladas	60
4-17	Distribución horaria de los errores cuando el conjunto de datos de entrenamiento es reducido al 62%.	61
A-1	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 20$	70
A-2	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 20$	71
B-1	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 60$	73
B-2	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 60$	74
C-1	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 60$	76
C-2	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 60$	77
D-1	Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h} y $N_i = 100$, conjunto de datos reducido al 38%.	79
D-2	Escenarios de generación para 7 días con frecuencia de actualización {12h, 6h} y $N_i = 100$, conjunto de datos reducido al 38%.	80
D-3	Escenarios de generación para 7 días con frecuencia de actualización {4h, 2h} y $N_i = 100$, conjunto de datos reducido al 38%.	81

Lista de Símbolos

A continuación se presenta un listado de símbolos usados a lo largo de este documento.

Símbolos con letras alfabéticas

Símbolo	Nombre	Unidad SI
e	Error predicción potencia eólica	
\hat{e}	Error medio	
$\hat{f}(x)$	Función de distribución de probabilidad	
h	Factor de suavidad	
k	Horizonte de la predicción	
k_v	constante Von Karman	0,04
K_p	Número de punto KNN	
$K(x)$	Función de pesos	
N_i	Número de intervalos	
N	Número de intervalos	
\hat{P}_{t+k}	Potencia eólica agregada hasta el tiempo $t + k$	
P_t	Potencia eléctrica	p.u.
P_{nen}	Potencia eléctrica entrenamiento	p.u.
P_{va}	Potencia eléctrica validación	p.u.
P_{max}	Potencia eléctrica máxima	p.u.
t	Tiempo	h
u	Magnitud velocidad viento	m/s
u_*	Velocidad de fricción	m/s
u_{in}	Magnitud Velocidad viento	m/s
\hat{u}	velocidad media	m/s
u_x	Componente zonal	m/s

Símbolo	Nombre	Unidad SI
u_y	Componente meridional	m/s
u_{en}	Velocidad entrenamiento	m/s
u_{va}	Velocidad validación	m/s
z_0	Relación rugosidad	
z_1, z_2	Alturas	m

Símbolos con letras griegas

Símbolo	Nombre	Unidad SI
α	Ángulo velocidad del viento	grados
ρ_k	Coefficiente de correlación	
ϕ_t	Información historia de datos	
μ_t	Media	
σ_t	Desviación Estándar	

Abreviaturas

Abreviación	Definición
KNN	K_p -ésimo vecino más cercano
KDS	Distribución del kernel
EM	Error medio
$EMAN$	Error medio absoluto normalizado
$EPMA$	Error porcentual medio absoluto
DEE	Desviación estándar de los errores

1 Introducción

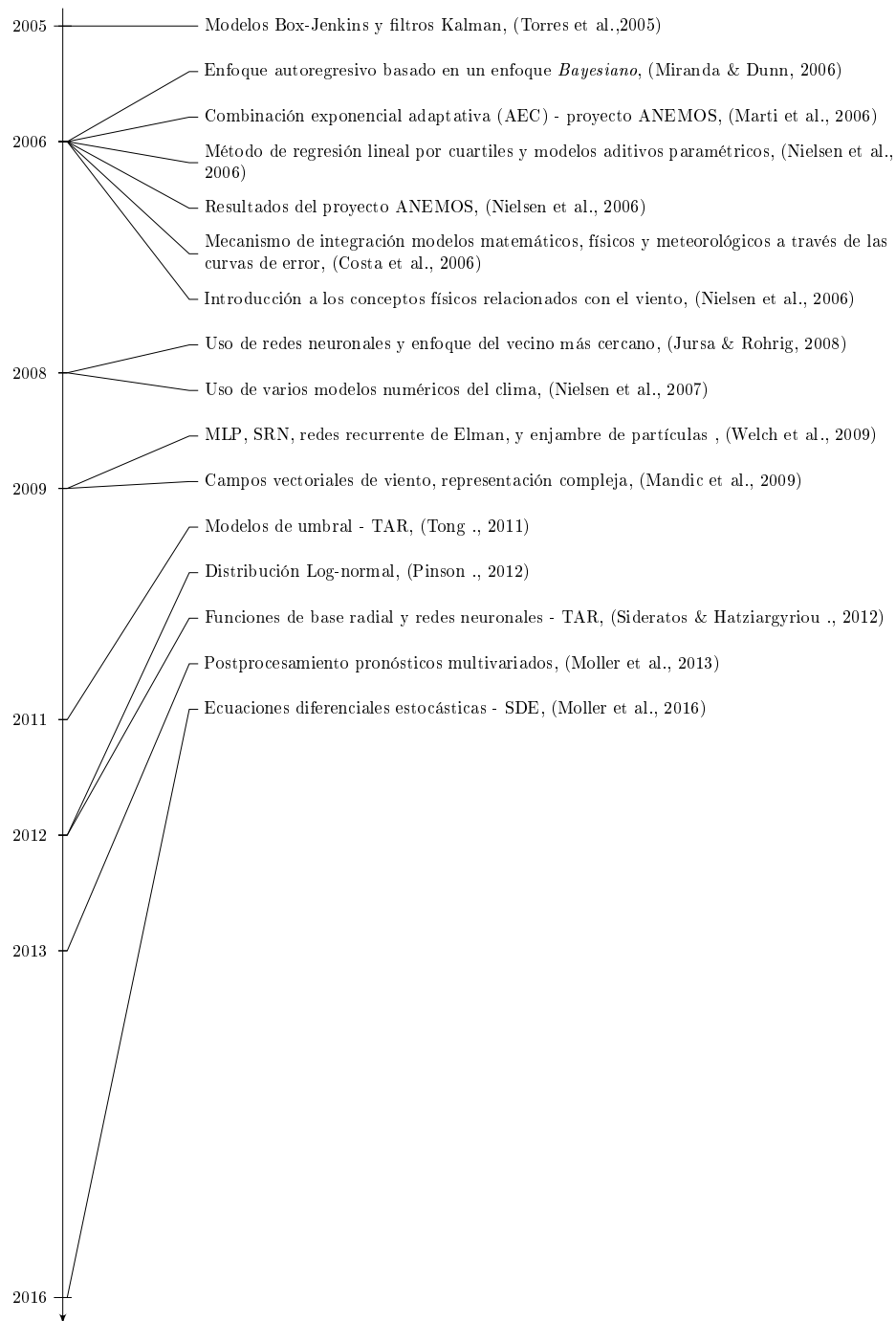
Resumen

En este capítulo se presenta una breve revisión de los métodos y avances en la predicción de la generación eólica a lo largo de sus casi 40 años de historia. Adicionalmente, se presenta la clasificación de los modelos en predicción de generación eólica de corto plazo y se formaliza el problema de su predicción. Además, se presentan los métodos más relevantes y se hace una introducción a los métodos de referencia y tipos de predicción probabilista comúnmente usados.

A continuación se presenta una breve revisión a lo largo de 40 años de historia de la predicción eólica de corto plazo - PECP, desde las primeras ideas y bosquejos, hasta el estado actual de los modelos y herramientas computacionales. La generación de energía eléctrica a partir de la fuerza de viento difiere de otros métodos de generación convencionales debido a la naturaleza *estocástica del viento*. Hasta el año 2006 las principales líneas de investigación que se venían trabajando (enfoque matemático y físico) se han comparado de manera cualitativa. Esta comparación trata de dar una idea clara acerca de la evolución cronológica de la PECP. La comparación entre los modelos de predicción no se hace explícita debido a la carencia de un estándar para la época. Por otra parte, sería necesario que los datos con los que se prueban los modelos sean exactamente los mismos, así, nuestro principal interés es presentar los modelos y herramientas más representativos desde el inicio de la historia de la PECP.

La evolución histórica del estado de arte de la predicción de la generación eólica de corto plazo para un período comprendido entre 1978 y 2016 es descrita a continuación:





1.1. Primeros Modelos

Una de las primeras investigaciones en predicción de corto plazo fue realizada en 1970 en el Laboratorio del Pacífico Noreste (PNL), (Wendell et al., 1978), dicha investigación mostró a las compañías encargadas del suministro eléctrico de la época la importancia de la predicción de corto plazo. Wegley and Formica (1983) estudiaron la predicción de la velocidad del viento a través de tres modelos (persistencia¹, modelo autoregresivo y modelo generalizado de Markov) para tres distintos pasos de tiempo 10, 30 y 60 min. Ellos concluyeron que el comportamiento del modelo de persistencia junto con el modelo generalizado de Markov tenía un mejor desempeño para pasos de tiempo de 10 min y 60 min respectivamente (Wegley et al., 1984). Geerts (1984) desarrolló modelos haciendo uso de filtros Kalman y modelos autoregresivos de media móvil (ARMA). Motivado por los resultados de su experimento, él plantea que adicional a la medida de la velocidad del viento, otras variables deberían usarse de modo que se incremente la precisión del modelo, tales como: *dirección del viento, temperatura, presión*. Dos trabajos seminales publicados en 1984 en la revista de meteorología aplicada (Journal of Applied Meteorology) por Conradsen et al. (1984) y Brown et al. (1984) fueron un símbolo de la transición de modelos para limitar las distribuciones únicamente a modelos dinámicos. En el primero de ellos (Conradsen et al., 1984) se hizo el ajuste de las medidas de la velocidad del viento a una distribución Weibull, mientras que (Brown et al., 1984) fue el primer trabajo enfocado en la búsqueda de un modelos dinámico (series lineales de tiempo) en la predicción de la velocidad del viento y la correspondiente generación de potencia eléctrica. McCarhty uso un programa de calculadora para realizar predicciones sobre algunas granjas de California para el una antelación (1985-1987), dicho programa se ejecutaba sobre una calculadora HP41CX (con una antelación hasta de 24h), el programa se alimentaba de las observaciones meteorológicas locales. Kaminsky et al. (1985) basaron su enfoque sobre la definición de categorías meteorológicas sinópicas. Él trabajó con un paso temporal de 15 min para predecir la velocidad del viento a través de un modelo de regresión sobre los 90 pasos anteriores. Concluyendo que diferentes categorías meteorológicas sinópicas requieren distintos métodos de regresión. Bossanyi (1985) utilizó filtros Kalman para predecir la velocidad del viento con un paso de tiempo de 1 min, comparando el comportamiento de este con el de un modelo de persistencia. Basado en sus observaciones son principales conclusiones fueron:

- El error más bajo en la predicción ocurre para un paso de tiempo de 5 min.
- El comportamiento del modelo respecto al modelo de persistencia mejora cuando el paso de tiempo es de 1 min.

En 1985, Bossany publico los resultados obtenidos al usar un modelo ARMA para predecir

¹La simplicidad del metodo de persistencia consiste en asumir que la velocidad (o la potencia) del viento en un instante determinado será la misma que cuando se hizo la predicción. Si $f(t)$ es el valor de la medida de la velocidad (o potencia) al tiempo t , diremos que la medidad de la velocidad del viento en un instante futuro $(t + \Delta t)$ será $f(t + \Delta t) = f(t)$.

la velocidad del viento con un paso temporal de 2s y 1 min. Durante los años 80s se publicó el primer trabajo significativo. Bailey and Stewart (1987) desarrollaron un método para predecir la velocidad del viento con 24h de antelación, orientado al despacho de carga. El objetivo principal de dicha investigación fue refinar las salidas del servicio meteorológico para un lugar determinado, usando las estadísticas de salida del modelo (MOS). Además, publicaron uno de las primeras revisiones sobre predicción de corto plazo, resumiendo los aspectos más significativos de los modelos desarrollados para la época.

1.2. Años 90s

Durante los años 90s, el incremento de la generación eólica instalada alrededor del mundo (principalmente en Europa y Estados Unidos) atrajo la atención de las compañías del sector eléctrico, así, se desarrollaron investigaciones más profundas en predicción eólica de corto plazo. La gran mayoría de estas investigaciones surgieron debido a la necesidad de integración de la creciente (y desconocida) energía eólica en la red. Troen and Landberg (1990) propusieron un modelo basado en el refinamiento del viento geotrófico² (modelo de predicción del clima de alta resolución de área limitada - HIRLAM) para un lugar en específico, considerando la orografía local, escabrosidad y demás obstáculos del terreno. El modelo se basó en la ley de arrastre geotrófico. Los primeros resultados presentaron mejoras sobre la persistencia para un valor único (velocidad y dirección del viento) con 9h de antelación, considerando únicamente magnitudes de la velocidad del viento mayores a 5m/s. Fellows and Hill (1990) investigaron la predicción de la carga eléctrica (con un paso temporal de 1h y un horizonte con una antelación de hasta 2h) y la predicción de la salida en potencia de las turbinas eólicas (con un paso temporal de 10min y un horizonte con una antelación de hasta 2h). En la predicción de la carga, ellos emplearon un método de reducción³ en tiempo real, con el cual los datos originales fueron reducidos en tendencias anuales y diarias. La tendencia de ambos tipos de datos se predijo a través de un modelo autoregresivo. Sus resultados presentaron mejoras respecto al modelo de persistencia y otros métodos tales como, usar el valor de la carga a la misma hora del día anterior, siempre y cuando los datos sean de tipo estacionario. Para series de tiempo no estacionarias, el método autoregresivo no presentó mejoras significativas sobre los resultados obtenidos a través de la aplicación del método de persistencia. Además, al aplicar el método de reducción en modo fuera de línea, se pudo apreciar una mejora significativa sobre la persistencia. Basados en los resultados obtenidos de la simulación de un sistema eólico/diésel sus principales conclusiones fueron:

- Se logró una reducción en el tiempo de sobrecarga del sistema a diésel.
- El ahorro de combustible fue menos sensible a mejoras de predicción de la generación eólica que a mejoras en la predicción de la carga.

²Aproximación física al viento real.

³Detrended method.

- El modelo de persistencia predice mejor el viento que la carga.

Watson et al. (1992) investigó la reducción del precio de los combustibles fósiles empleando un modelo de predicción numérica del clima (NWP) en conjunto con las estadísticas de salida del modelo (MOS) para predecir la velocidad y dirección del viento con una antelación de hasta 18h y un paso de 1h. Su conclusión fue que el modelo de predicción NWP-MOS puede mejorar de manera significativa el ahorro de combustibles fósiles en comparación con el modelo de persistencia.

Tande and Landberg (1993) probaron redes neuronales para predecir la salida de potencia eléctrica (único valor de potencia). Ellos obtuvieron una pequeña mejora sobre la persistencia. Beyer et al. (1994) usaron redes neuronales (single perceptron, multilayer perceptron, y funciones de base radial) para predecir la velocidad del viento en una única turbina con pasos de tiempo de 1min y 10min. Logrando una mejorar el comportamiento de todos los modelos sobre la persistencia. Para la salida en potencia de la turbina ellos usaron la curva del fabricante. Nuevamente, el comportamiento de todos los modelos estuvo por encima de la persistencia para ambos pasos de tiempo. Comparando sus resultados con los resultados de publicaciones previas, ellos consideraron que mejores resultados para las escalas de tiempo consideradas no eran posibles debido a la carencia de información. Ellos enfatizan que el uso de redes simples genera resultados similares a los generados por enfoques más complejos. Jensen et al. (1994) presentó la herramienta de predicción eólica (WPPT), desarrollada por el departamento de informática y modelamiento matemático (IMM) de la Universidad Técnica de Dinamarca (DTU). Con un paso temporal de 1h y 30min y una antelación de hasta 36h, esta herramienta estuvo orientada al despacho de carga y fue probada en siete granjas eólicas. WPPT fue construida sobre un modelo autoregresivo con la potencia de salida como principal variable y la velocidad del viento como variable exógena. Algunas particularidades de este modelo son:

- La inclusión de un termino armónico que representa el patrón diurno.
- Una transformación de la raíz cuadrada de las variables, de modo que una distribución Gaussiana se aproximaría a la distribución de los errores en la predicción.

Ellos reportaron la estructura del software, algunas gráficas (predicciones reales) y perspectivas para trabajos futuros (predicción meteorológica y variables exógenas). Landberg reportó algunos resultados de la aplicación del modelo RISO (RISO National Laboratory) en 17 granjas de generación eólica en Dinamarca (Landberg, 1994). Madsen publicó un reporte describiendo las experiencias de la DTU con la herramienta WPPT (previamente reportado en Jensen et al. (1994)). En este documento él describió varios enfoques orientados a la predicción de la potencia en granjas de generación eólica (predicción de la potencia directa o indirectamente, considerando o no la dirección del viento, modelos con predicción meteorológica, modelos autoregresivos y redes neuronales). Él detalló un modelo autoregresivo con un termino armónico que representa el patrón diurno y una transformación de la

raíz cuadrada de las variables. Kariniotakis et al. (1996) investigaron la aplicación de redes neuronales y lógica difusa en la predicción de la potencia de salida en las turbinas eólicas, obteniendo un mejor comportamiento sobre la persistencia con el modelo de lógica difusa en horizontes con una antelación de 2h y un paso temporal de 10min. Ellos probaron los modelos con datos de un sistema eólico-diésel, considerando la salida de potencia como variable principal y la velocidad del viento como variable exógena. Para lograr optimizar la arquitectura del modelo, ellos emplearon un algoritmo basado en un modelo no lineal. Lin et al. (1996) realizaron un estudio sobre redes neuronales para predecir la velocidad y dirección del viento con un paso temporal de 1s. Ellos probaron diferentes arquitecturas, diferentes funciones de activación, diferentes métodos de aprendizaje y diferentes longitudes para el conjunto de entrenamiento, logrando una mejora considerable sobre un modelo autoregresivo. Akylas et al. (1997) probaron diferentes enfoques basados en modelos meteorológicos y datos registrados por estaciones de medida en la predicción de la velocidad del viento para luego poder convertir ésta en potencia de salida de las turbinas (a través de las curvas de potencia). Con una antelación de 24h y pasos temporales de 1h. Ellos probaron tres clases de modelos:

- Regresión multivariable.
- Regresión de la predicción meteorológica.
- Corrección de la predicción meteorológica.

Sus conclusiones fueron que el comportamiento del primer modelo escasamente presentaba mejoras sobre la persistencia, mientras que el segundo y tercer modelo mejoraban su comportamiento en gran medida respecto a la persistencia, con una pequeña ventaja del tercer modelo. Kariniotakis et al. (1996) compararon diferentes aproximaciones (método ingenuo⁴, redes neurales, lógica difusa, y modelos autoregresivos, entre otros). Para una antelación de hasta 2h y un paso temporal de 10min, ellos lograron mejorar el comportamiento sobre la persistencia para el modelo basado en lógica difusa. El peor comportamiento lo presentaron el método ingenuo y el modelo autoregresivo (Kariniotakis et al., 1997). Inspirado por modelos desarrollados en economía, Tol (1997) desarrolló un modelo autoregresivo condicional (Generalised Auto Regressive Conditional Heteroscedastic - GARCH) . Bailey et al. (1999) presentó la herramienta Ewind la cuál predice la velocidad y dirección del viento, esta herramienta fue desarrollada por la compañía TRUEWIND. Con una antelación de hasta 48h y pasos de tiempo de 1h. Esta herramienta se basó en un modelo a mesoescala⁵ el cual refinaba las salidas de un modelo de clima regional. La herramienta usaba las estadísticas de salida del modelo para la operación en línea. Dos enfoques han sido ampliamente usados para convertir la predicción de la velocidad del viento en potencia de salida, ellos son: modelos estadísticos

⁴Naive Method.

⁵En Meteorología es el estudio de sistemas del tiempo atmosférico más pequeños que la escala sinóptica meteorológica, pero más grandes que la microescala y la escala de tormenta de los sistemas de nubes cúmulos. Sus dimensiones horizontales generalmente oscilan de cerca de 9 km a varios centenares de km.

y modelos físicos. Haciendo uso de las predicciones de clima alemán, ellos presentaron los resultados de seis lugares al norte de Alemania, con una antelación de hasta 24h. Además, ellos evaluaron la correlación espacial en la desviación de las predicciones, concluyendo que:

- La correlación disminuye a medida que se incrementa la distancia.
- La desviación para horizontes largos de antelación presentó una mayor correlación que para horizontes cortos.

Nielsen et al. (1998) mostró que no es razonable el uso de la persistencia como un modelo de referencia para horizontes de antelación mayores a un par de horas.

1.3. 2000 - 2006

En este período se prestó especial atención al desarrollo de herramientas en línea que permitieran la operación y evaluación de las predicciones. También, surgieron los primeros intentos de integración de las dos corrientes complementarias de la predicción eólica de corto plazo (modelos matemáticos y modelos físicos). Sfetsos (2000) hizo la comparación entre modelos lineales (modelos autoregresivos) y modelos no lineales (redes neuronales retroalimentadas, redes de funciones de base lineal, redes recurrentes de Elman, modelos ANFIS, redes neuronales lógicas) para predecir el valor medio horario de la velocidad del viento. Él concluyó que:

- El comportamiento de los modelos no lineales superó el comportamiento de los modelos lineales.
- Todos los modelos no-lineales presentaron una desviación estándar similar.
- El modelo de redes neuronales lógicas fue superior a los otros.

Lange and Waldl (2001) presentaron sus conclusiones sobre el modelo de la Universidad de Oldenburg, concluyendo que:

- La incertidumbre en la predicción de la velocidad del viento es independiente de la magnitud de la predicción en velocidad.
- La incertidumbre en la predicción de la potencia está en función de la curva de potencia y el error medio de la predicción en velocidad.
- Para algunos sitios, la incertidumbre en la predicción de la velocidad de viento presenta cierta dependencia de las condiciones climatológicas (por ejemplo, los errores se incrementaron para zonas con bajas presiones).

Watson et al. (2001) presentaron sus resultados preliminares de la operación en línea del modelo del Laboratorio Nacional RISO. Estos se obtuvieron a partir del estudio de 15 granjas eólicas en Irlanda, sin embargo, para la fecha el modelo aún no se había ajustado para las

condiciones locales. Giebel et al. (2002) informaron sobre el lanzamiento de tres proyectos financiados por el Ministerio Danés de Energía. Dichos proyectos tenían como objetivo realizar la integración de los modelos Prediktor y WPPT (el último presentado en 1994) para ser utilizado por todas las empresas Danesas de servicios públicos. Esta nueva herramienta (llamada Zephyr) fue construida teniendo en cuenta las ventajas de ambos modelos. Por un lado, WPPT presentó un buen comportamiento para predicciones con una antelación por debajo de las 8h, mientras que Prediktor lo hacía para un horizonte con una antelación que iba desde las 8h hasta el horizonte máximo del servicio meteorológico, además, Prediktor tenía la posibilidad de realizar predicciones aun sin disponer de datos. La herramienta fue pensada para ser flexible, independientemente del sistema operativo sobre el cual se estuviera trabajando. Giebel et al. (2001) investigaron el refinamiento de la plataforma y las salidas del modelo HIRLAM (con resoluciones espaciales de 0.2 y 0.5) con el fin de predecir la potencia a la salida de las granjas eólicas en España, ubicadas en un terreno moderadamente complejo. Ellos desarrollaron tres modelos:

- Modelo de predicción del viento.
- Modelo de la curva de potencia.
- Modelo de predicción de potencia.

El segundo y tercer modelo fueron construidos basados en modelos de regresión lineal adaptativos, considerando la dirección del viento. Sus principales conclusiones fueron:

- Modelos basados en HIRLAM 0.2 mejoraron su comportamiento en HIRLAM 0.5 para horizonte de tiempo con una antelación hasta 24h.
- Para resoluciones espaciales de 0.2 y 0.5 la reducción basada en las componentes principales mejoró el comportamiento de la interpretación y ajuste del modelo.

Focken et al. (2002) estudió la predicción de la salida agregada de potencia para diferentes granjas con un horizonte de tiempo con una antelación hasta 48h. Ellos observaron una importante reducción del error en la predicción debido al ruido espacial. Considerando 30 granjas eólicas en Alemania. Concluyendo que el error es más sensible al tamaño de la región que a la cantidad de lugares. Costa et al. (2003) informaron acerca del lanzamiento del proyecto UPMPREDICTION, un intento verdadero en la unificación de las dos líneas complementarias en la investigación de la predicción eólica de corto plazo (modelos matemáticos y físicos) y el desarrollo de una herramienta en línea orientada a el mercado local y a la administración de los sistemas eléctricos de potencia. En la primera fase desarrollaron tres tipos de modelos: estadísticos/matemáticos, modelos autoregresivos y modelos de lógica difusa y redes neuronales. De estos modelos, las redes neuronales presentaron un mejor comportamiento sobre la persistencia para tres granjas eólicas en la región noreste de España. Kariniotakis et al. (2003) informaron del lanzamiento del proyecto ANEMOS, un proyecto en conjunto con 7 países (centros de investigación, universidades, industrias,

operadores del sistema, empresas de servicios públicos, servicio meteorológico, agencias de energía de Francia, Irlanda, España, Alemania, Grecia, Dinamarca y Reino Unido) con el propósito de desarrollar un sistema de predicción a gran escala para la integración de granjas eólica terrestres (on-shore) y en el mar (off-shore), considerando un horizonte de predicción con una antelación de hasta 48h. Además, de un horizonte de predicción más largo, con una antelación de hasta 7 días (importantes a la hora de programar mantenimientos). El proyecto fue presentado bajo la siguiente división de tareas:

- Recolección de datos y evaluación de necesidades.
- Evaluación fuera de línea de las técnicas de predicción.
- Desarrollo de modelos estadísticos.
- Desarrollo de modelos físicos.
- Predicción off-shore.
- Desarrollo de la plataforma de predicción ANEMOS.
- Instalación de la plataforma para operación on-line.

Algunos modelos probados en la primera parte del proyecto fueron: Aminies Wind Power Prediction System (AWPPS), Prediktor, WPPT, Zephyr, Previento; LocalPred; Sipreólico; modelo de predicción NTUA's, modelo de predicción RAL's, tecnologías de predicción ASIA. Estos modelos fueron probados en lugares con diferentes tipos de terrenos y bajo distintas condiciones climáticas. Gallardo et al. (2003) informó sobre el lanzamiento del proyecto CASANDRA con un horizonte de tiempo con una antelación de hasta 72h y paso temporal de 1h. La herramienta se basó en un modelo a mesoescala, estadísticas de salida y modelo de la curva de potencia.

Giebel et al. (2003) y Landberg et al. (2003) publicaron dos revisiones del estado del arte de las diferentes líneas de investigación en predicción eólica de corto plazo. Madsen et al. (2004) presentó un protocolo para estandarizar la evaluación de modelos de predicción eólica de corto plazo⁶. Ellos reportaron el uso de este estándar en la base de datos del proyecto ANEMOS. Jiménez et al. (2004) estudió la sensibilidad de las simulaciones del viento bajo el modelo de investigación y predicción meteorológica (WRF) sobre la base de un horizonte de pronóstico de hasta 72h y un paso horario. Ellos consideraron un terreno complejo de 100km×100km en la península Ibérica y dos simulaciones sinópticas diferentes⁷. Ellos concluyeron que al incrementar la resolución del horizonte de antelación se mejoraba los valores de la simulación.

⁶Esto debido a que al comparar el desempeño con un modelo de persistencia, este genera una idea optimista acerca del comportamiento del modelo en cuestión.

⁷Se denomina observación sinóptica al conjunto de medidas de diferentes variables meteorológicas que se realizan a nivel de superficie a determinadas horas, y cuyos fines son contribuir a la elaboración de la predicción meteorológica de la zona y la climatología del lugar donde se realizan.

Bustamante et al. (2004) investigó la predicción horaria del viento a través de modelos autoregresivos y redes neuronales para series de tiempo; aplicando dos métodos de reducción (dinámico y estático) para escalas de tiempo muy pequeñas. Ellos concluyeron que el modelo de reducción dinámico en conjunto con un modelo a mesoescala mejoraba la resolución y el comportamiento de la predicción. Pinson and Kariniotakis (2004) presentaron una metodología para evaluar (en línea) el riesgo relacionado a la predicción de potencia eólica. Ellos emplearon un enfoque de re-muestreo para generar intervalos de confianza. A su vez, ellos recomendaron el uso de *meteo-risk* como indicador de la probabilidad de ocurrencia de grandes errores en la predicción. Kariniotakis et al. (2004a) presentaron los resultados parciales del proyecto ANEMOS. Basados en el estudio de seis granjas de viento distribuidas en la región norte y oeste de Europa (una off-shore, dos in-shore en terreno plano, dos en terreno complejo y una en terreno escabroso), ellos concluyeron entre otras cosas que:

- La predicción numérica del clima juega un papel importante en terrenos complejos.
- El comportamiento de los modelos depende de la complejidad del terreno, siendo menos exactos en terrenos escabrosos.
- El error de predicción es más disperso en terrenos complejos.
- Se debería combinar de varios tipos de predicciones para reducir el error.

Torres et al. (2005) publicaron los resultados para un modelo ARMA (modelo autoregresivo de media móvil) aplicado a la predicción de la velocidad del viento, con un horizonte de antelación de hasta 10 pasos y un paso temporal horario. Considerando un estudio de 9 años de datos en 5 diferentes estaciones de medición. Ellos presentaron un modelo fuera de línea para transformar y estandarizar las series de tiempo, así, aproximar la distribución a una distribución Gaussiana y evitar la estacionalidad. Ajustando el modelo para cada uno de los doce meses del año, ellos lograron mejoras en el comportamiento del modelo respecto a la persistencia, observando que la desviación estandar (RMSE) presentaba cierta dependencia con la velocidad del viento. Madsen et al. (2005) hicieron énfasis en la necesidad de usar una metodología estándar para evaluar el comportamiento de las medidas (uso del error medio absoluto y desviación estándar), ellos sugirieron realizar mediciones del error no solo sobre todo el conjunto de datos, sino que también, sobre divisiones del conjunto de datos (esto permite observar variaciones en el comportamiento del modelo). Además, ellos mencionan que la mejora en el comportamiento de las mediciones depende de la aplicación deseada. Nielsen et al. (2006) emplearon regresión lineal por cuartiles y modelos aditivos paramétricos, ellos investigaron modelos para los cuartiles inferior y superior considerando las siguientes variables de regresión: predicciones de potencia tomadas del modelo WPPT, horizonte de predicción y predicciones del modelo HIRLAM (operado por el Instituto Meteorológico Danés). Además, ellos estudiaron la influencia del índice de riesgo introducido en (Pinson and Kariniotakis, 2004). Sus principales conclusiones fueron:

- La variable mas relevante fue la predicción de potencia.

- El índice de riesgo parece no tener influencia en el estudio por cuartiles.
- El enfoque del modelo aditivo fue adoptado pues permite la inclusión de variables de regresión.

En febrero de 2006 fueron presentados los resultados del proyecto ANEMOS en el marco de la Conferencia Europea de Energía Eólica - EWEC. Algunas contribuciones de dicha reunión fueron:

- Kariniotakis et al. (2006) presentó una revisión de proyecto, haciendo especial énfasis en la aplicación de software para operación en línea en algunas granjas on-shore y off-shore.
- Waldl et al. (2006) mostró algunos aspectos de la operación por más de 1 año de la plataforma ANEMOS.
- Resultados más detallados sobre la operación del proyecto se pueden encontrar en: modelos estadísticos Nielsen et al. (2006); modelos físicos Giebel et al. (2006) y predicciones off-shore Tambke et al. (2006).

Costa et al. (2006) propusieron un mecanismo para la integración de los modelos matemáticos, físicos y meteorológicos. Este mecanismo se basó en el seguimiento del punto de intersección de las curvas de error de los modelos. Después del seguimiento, la idea fue minimizar el error del modelo final.

1.4. 2006 - 2016

A partir del año 2006, se presenta una revisión de los métodos más representativos en la predicción de generación eólica de corto plazo. Entre otros destacan los siguientes: predicción numérica del clima, métodos estadísticos y de aprendizaje y, métodos híbridos

1.4.1. Predicción numérica del clima

El enfoque físico de la predicción de generación eólica, en contraste con el enfoque estadístico, usa un modelo con una descripción detallada de las condiciones locales (geografía del área, orografía, obstáculos, etc) y de la granja eólica (esquema de la granja, curva de potencia, ...). Contrario a los métodos estadísticos, el método de predicción numérica del clima no requiere datos históricos, sin embargo, adquirir los datos físicos es uno de los mayores inconvenientes de este enfoque (Jung and Broadwater, 2014).

Generalmente, un modelo de predicción numérica del clima está dividido en tres componentes: representación adiabática del flujo no viscoso, ecuaciones físicas que describen las variaciones de los procesos meteorológicos (turbulencia, radiación, ...) y las herramientas computacionales de programación (Foley et al., 2012).

En Lange and Focken (2006) se presenta una sólida introducción a los conceptos físicos relacionados con el viento. Lang et al. (2006) presenta un esquema múltiple de predicción en

conjunto (MSEP) aplicado a un estudio comparativo entre los resultados obtenidos en Irlanda y Alemania. Nielsen et al. (2007) demostraron que el uso de varios modelos numéricos del clima, el error tendía a decrecer. Louka et al. (2008) encontraron que el uso de filtros Kalman puede remover el error sistemático asociado con la predicción. Estos modelos presentan un buen comportamiento para horizontes de tiempo con una antelación por encima de 4h. Algunas herramientas de predicción eólica son las presentadas en: (Jensen et al., 1994), (Bailey et al., 1999), (Giebel et al., 2001) y (Giebel et al., 2002). Una de las características más importantes de este modelo es que entrega un único valor de predicción, así, que el uso de la optimización estocástica y evaluación del riesgo es limitado.

1.4.2. Métodos estadísticos

El principal enfoque en la predicción de la velocidad del viento y de la generación eólica se basa en un modelo estadístico. En este se representa la relación entre la velocidad del viento (o la potencia generada) y variables exploratorias (incluyendo modelos de predicción numérica del clima y medidas en línea) Kariniotakis et al. (2004b). En este enfoque una gran cantidad de datos son analizados (sin embargo, el proceso físico no es explícito) para construir un modelo estadístico. Los métodos estadísticos involucran únicamente un paso para convertir las variables de entrada en potencia de salida a través de una relación estadística entre la predicción de la velocidad del viento y la salida de potencia de la granja eólica, contrario a los métodos de predicción numérica del clima, el enfoque estadístico requiere de los datos históricos para entrenar el modelo.

Diferentes enfoques han sido empleados, tales como:

- Redes neuronales artificiales.
- Redes neuronales artificiales y lógica difusa.
- Combinaciones

1.4.2.1. Enfoque estadístico convencional

El enfoque estadístico convencional se utiliza principalmente en horizonte de predicción de muy corto plazo y corto plazo. Generalmente estos son usados como modelos de referencia debido a su exactitud y facilidad de formulación. Este enfoque hace uso de un modelo de series de tiempo para predecir la velocidad (o potencia) del viento. El enfoque estadístico convencional (propuesto por Box-Jenkins) está dividido en cuatro partes: identificación del modelo, estimación, diagnóstico y predicción. Otras técnicas estadísticas usadas son: modelos autoregresivos (AR), media móvil (MA), modelos autoregresivos de media móvil (ARMA), y modelos autoregresivos de media móvil integrada (ARIMA). Los últimos son una generalización de los modelos ARMA. Torres et al. (2005) encontraron que haciendo uso de modelos Box-Jenkins y filtros Kalman era posible lograr una reducción del 20 % en el error comparado

con la persistencia para un horizonte de tiempo con una antelación de 10h, un paso temporal horario y 9 años de datos históricos.

Además del uso de series de tiempo, existen otros enfoques que permiten establecer relaciones estadísticas entre datos como: inteligencia artificial, uso de redes neuronales y lógica difusa. En la literatura reciente, se han referido a estos como métodos de aprendizaje porque ellos aprenden de la relación entre la predicción de la velocidad del viento y la potencia de salida. En algunos estudios se presenta una mejora, dependiendo del horizonte de antelación entre 9.5 % y 28.4 % sobre la persistencia a través del uso de algoritmos genéticos (GE) y sistemas de inferencia difusa (FIS) Damousis and Dokopoulos (2001). Jursa and Rohrig (2008) presentaron un enfoque el cual combinaba redes neuronales y el enfoque del vecino más cercano en un modelo de optimización, logrando una mejora del 10.75 % en la desviación estándar normalizada de la predicción en comparación con la persistencia. Welch et al. (2009) comparó tres tipos de redes neuronales (Multilayer Perceptron - MLP, Redes Recurrentes - SRN y redes recurrentes de Elman) entrenadas usando optimización de enjambre de partículas (PSO) para la predicción de corto plazo de la velocidad del viento. Ramirez-Rosado et al. (2009) encontraron significativas mejoras sobre la persistencia al comparar predicciones para las cuales el modelo numérico del clima era mejorado a través de redes neuronales y otras técnicas de aprendizaje artificial combinadas con modelos de la curva de potencia de la turbina.

Un método novedoso en análisis y modelamiento de campos vectoriales de viento fue presentado por Goh et al. (2004) y desarrollado por Mandic et al. (2009), en este enfoque el vector de viento es representado como una cantidad compleja, dirección y velocidad del viento son modeladas de manera simultánea. En los últimos años, algunos métodos Bayesianos han sido empleados en la predicción de la velocidad del viento. Miranda and Dunn (2006) emplearon un método autoregresivo basado en un enfoque Bayesiano, así, ellos obtuvieron una predicción de la velocidad del viento con un horizonte de antelación de hasta 1h.

Pinson et al. (2009) presentan un método probabilístico basado en la conversión de la serie de errores de la predicción en una variable aleatoria Gaussiana multivariable, para un horizonte de tiempo con una antelación de hasta 2 y 3 días. En Tong (2011) se presenta el desarrollo de modelos de umbral para el análisis de series de tiempo (modelos autoregresivos de umbral - TAR) los cuales son comunes en estudios econométricos. Distribuciones Gaussianas truncadas, restringidas y un modelo generalizado de la distribución Log-Normal fueron algunos de los modelos más relevantes propuestos en el modelamiento de la generación eólica (Pinson, 2012). En Foley et al. (2012) muestra que el error (RMSE) incrementa a medida que el horizonte de antelación aumenta, siendo menor para modelos de predicción de la velocidad del viento que para modelos de predicción de la potencia de salida. Sideratos and Hatziargyriou (2012) presentaron una metodología novedosa para la predicción probabilística de la generación eólica, el método se basa en inteligencia artificial (funciones de base radial y redes neuronales). El comportamiento del modelo fue evaluado en dos granjas eólicas ubicadas en áreas con condiciones climatológicas distintas. Este modelo presenta una mejora del 52 %

sobre la persistencia para el paso 51. En Möller et al. (2013) se propone un método para el postprocesamiento de un conjunto de pronósticos multivariados con el fin de obtener una distribución conjunta de la predicción del clima. Este método se aplicó para predicciones con un horizonte de antelación de hasta 48h y tres variables: temperatura diaria (mínima y máxima), presión a nivel del mar, acumulación de la precipitación y máxima velocidad del viento.

Møller et al. (2016) presentan un nuevo método basado en ecuaciones diferenciales estocásticas (SDE). Ellos consideran una predicción de la generación eólica con una antelación de hasta 48h y un paso temporal de 6h. Una de las mayores ventajas de este modelo es que permite describir: no linealidades, no estacionalidad, variaciones en el tiempo y variaciones dobles.

1.4.3. Métodos híbridos

La idea básica de los modelos híbridos es combinar diferentes enfoques, conservando lo mejor de cada uno de ellos. Sin embargo, no siempre la combinación de varios métodos de predicción resulta en un mejor comportamiento respecto al comportamiento individual. Existen modelos que además de combinar redes neuronales y lógica difusa usan combinaciones lineales adaptativas y diferentes series de tiempo (con distintas variaciones de los parámetros del modelo). En el proyecto ANEMOS (introducido en (Kariniotakis et al., 2003)) consideran un modelo híbrido llamado Combinación Exponencial Adaptativa (AEC) (Marti et al., 2006). En una primera etapa, en conjunto con un modelo AEC se combinaron otros más. En la segunda etapa, el método es usado para realizar combinaciones alternativas de los métodos de la primera etapa. Los resultados obtenidos muestran las ventajas del método propuesto.

1.5. Conclusiones

Esta revisión presenta de manera cronológica la evolución de la predicción de generación eólica de corto plazo a lo largo de sus casi 40 años de historia. Los últimos modelos presentados en la revisión literaria están orientados a la reducción del error en la predicción de la potencia eólica, así, hacer de la generación eólica una fuente de generación atractiva para el operador del sistema.

El principal objetivo de la predicción de la generación eólica es estimar la potencia y velocidad del viento tan rápido y preciso como sea posible. Las diversas herramientas expuestas a lo largo de la revisión literaria evidencian la gran utilidad de la predicción en generación eólica de corto plazo, pues posibilitan hacer un mejor despacho y compromiso de unidades de generación térmica, generación hidráulica y almacenamiento de energía.

El enfoque estadístico utiliza un gran conjunto de datos históricos como entradas al modelo. Dicho enfoque presenta buenos resultados para todos los horizontes de predicción, especialmente para horizontes de corto y muy corto plazo. Además, la comparación entre modelos no

suele ser fácil, pues el comportamiento de cada uno de ellos varía de acuerdo con la situación. Por otra parte, la combinación de modelos permite tomar en consideración lo mejor de cada uno de ellos, este enfoque no solo mejora la precisión de la predicción, sino que también, reduce el riesgo ante eventos extremos. Estos métodos híbridos mejoran el comportamiento tanto de los modelos numéricos del clima (NWP) (en términos de precisión) como el de los modelos estadísticos y técnicas de aprendizaje de máquina. Por último, grandes incrementos en la penetración de la energía eólica al sistema de potencia posibilita el desarrollo de un gran número de herramientas que permitan estimar y manejar lo mejor posible y de acuerdo con la el requerimiento deseado la variabilidad en el flujo de viento.

2 Propuesta de Investigación

Resumen

En este capítulo se presenta la necesidad del desarrollo de herramientas en predicción eólica así como el interés por realizar un trabajo investigativo en predicción eólica de corto plazo. En la primera sección de este capítulo se tratan algunos de los conceptos básicos usados durante el desarrollo de esta tesis. Se deja al lector consultar cada una de las referencias nombradas cuando se quiera tener mayor claridad en los conceptos. Adicionalmente, se presenta la metodología usada y los objetivos propuestos para el tema de investigación abordado.

Varios sistemas eléctricos de potencia hoy en día integran generación eólica en su despacho económico, permitiendo el despacho de granjas de generación eólica en conjunto con plantas de generación convencional. Sin embargo, una alta penetración de potencia eólica implica retos en la planeamiento y operación del sistema de potencia principalmente debidos a la naturaleza fluctuante del viento. Así, la incorporación al sistema eléctrico de potencia de este tipo de generación requiere el desarrollo de nuevas herramientas computacionales capaces de valorar la generación del agregado total de la energía eólica en el tiempo. La creciente tendencia del sistema eléctrico de potencia a incorporar fuentes no convencionales de generación se debe principalmente al aumento de los gases de efecto invernadero en la atmósfera, esto ha generado un ambiente propicio para la incorporación de fuentes no convencionales de energía dentro del portafolio que se maneja actualmente. Aunque el costo de implementación y mantenimiento ha sido un punto en su contra, no va más allá del que se pagará sino se inicia a reducir las emisiones a la atmósfera. En países industrializados la opinión pública demanda a sus gobiernos tomar acciones pertinentes que permitan mitigar los efectos adversos del calentamiento global. Estas acciones requieren un cambio estructural en la economía de largo plazo.

Una parte significativa de dichos cambios debe ser llevada a cabo principalmente por el sector eléctrico. En este sector las medidas adoptadas para cumplir con la reducción de las emisiones consisten en: aumentar el nivel de penetración en general de las energías renovables y en particular de la generación eólica, sustitución de carbón por gas, eficiencia energética

y conservación. Además, el aumento continuo de la demanda energética, los precios al alza de los combustibles fósiles, los desastres naturales que ponen en entredicho la seguridad y conveniencia de la generación nuclear, hacen de la energía eólica una alternativa atractiva, la cual se espera continúe creciendo en los próximos años (Association, 2011).

Alrededor del 6 % de la energía consumida en los Estados Unidos proviene de la generación eólica. Dinamarca es el líder con cerca del 30 % de la demanda nacional suplida desde granjas eólicas, con un incremento esperado del 50 % para 2020 y del 100 % para 2035. Adicional a esto, en el marco de las políticas danesas se planea eliminar el uso de cualquier fuente de combustible fósil para 2050 (DCM, 2012).

En este orden de ideas, el interés por realizar un trabajo investigativo en predicción eólica de corto plazo surge por tres razones fundamentales:

1. La importancia que tiene para el sector eléctrico y en general para toda la sociedad el suministro eléctrico.
2. consecuentemente, la facilidad de generar energía desde la fuerza del viento constituye una nueva fuente de incertidumbre en la operación y planeación de un sistema eléctrico de potencia debido a su dependencia intrínseca con las condiciones climáticas¹ y,
3. la adquisición de conocimiento sobre la conformación de un portafolio de energía con alta penetración de generación de energía proveniente de granjas eólicas.

2.1. Conceptos básicos

Una turbina eólica convierte la energía del viento en energía eléctrica. La cantidad de energía generada depende directamente de la velocidad del viento. La relación *potencia vs. velocidad* es conocida como curva de potencia de la turbina²(ver Figura 4-4). En (Ackermann et al., 2005) se presenta en detalle información sobre el estado actual de los generadores y la electrónica de potencia usada en su implementación.

Debido a la relación potencia vs. velocidad, fluctuaciones en la velocidad del viento se traducen en variaciones de la potencia generada. Dichas fluctuaciones son observadas en diferentes escalas de tiempo. El principal objetivo de esta propuesta es la predicción de la potencia eólica de corto plazo, lo que significa que nuestra escala de tiempo es del orden de horas.

La velocidad del viento esta estrechamente relacionada con las condiciones atmosféricas, periodo del año, cúmulos de nubes y con épocas de lluvia (Vigueras-Rodríguez et al., 2012). Así, son las condiciones atmosféricas quienes la determinan, en consecuencia, dicha velocidad presenta una no-estacionalidad en el tiempo, lo que a su vez resulta en una no-estacionalidad en la potencia generada.

¹Naturaleza estocástica del viento.

²La forma de la curva de potencia es característica de cada generador, de la electrónica de potencia y del sistema de control. Estas curvas se obtienen por parte del fabricante en condiciones ideales, es decir, cuando la turbina se somete a un flujo de viento constante. Dicha curva se divide en cuatro secciones, cada una de las cuales esta en función de la potencia y la velocidad del viento.

2.1.1. Horizonte de predicción

Los horizontes de predicción se suelen agrupar de acuerdo a lo presentado en la Tabla 2-1.

Tabla 2-1: Horizontes de predicción

Horizonte de Predicción	Desde	Hasta
De muy corto plazo	≫ segundos	30min
De corto plazo	30min	6h
De mediano plazo	6h	24
De largo plazo	24h	72h
De muy largo plazo	72h	≫72h

La predicción de la generación eólica en un horizonte de tiempo de muy corto plazo es usada en el mercado eléctrico y en la toma de acciones de regulación. Predicciones de corto plazo, son usadas en el planeamiento del despacho económico de carga, y en la toma de decisiones cuando se produce un incremento/decremento de carga. Decisiones de generación en tiempo real y seguridad operacional en el mercado eléctrico son basadas en predicciones a corto plazo. Finalmente, predicciones de largo plazo y muy largo plazo son usadas en el despacho de unidades de generación, almacenamiento, mantenimiento de las granjas eólica y en la obtención del costo óptimo de operación.

2.1.2. Definición conceptual de modelos en predicción eólica de corto plazo

Típicamente los modelos que proveen datos acerca de la predicción de la generación eólica lo hacen de manera puntual, estos datos corresponden a la producción de potencia más probable para un horizonte dado. Además de este tipo de modelos, existen otros que aportan una mayor cantidad de información acerca de cuál será el valor futuro de la potencia a la salida de la granja eólica. Una revisión completa sobre los distintos métodos en predicción de la generación eólica se puede encontrar en (Jung and Broadwater, 2014)

El uso de la predicción eólica de corto plazo como herramienta de pronóstico de la generación de potencia desde la fuerza del viento, usualmente dentro de un rango que va desde 1 hora hasta 72 horas (Foley et al., 2012), contribuye a una operación segura y económica del sistema eléctrico de potencia. En Okumus and Dinler (2016) se presenta la clasificación de los horizontes de predicción de potencia eólica y algunos de los métodos usados para tal fin. Un punto de interés a la hora de emprender el diseño de un método de predicción de generación

eólica de corto plazo, es que este deberá ser diseñado para ser utilizado durante la operación del sistema, es decir, en tiempo real (online), por otra parte, cuando se trabaja con los datos históricos del sistema se está operando en modo fuera de línea (off-line), generalmente, el último enfoque es usado en investigación.

La predicción de la generación eólica es un problema que involucra variables de diversas fuentes, la gran mayoría de las cuales están estrechamente relacionadas con el comportamiento de clima, razón por la cual el estudio de la predicción de la generación eólica no es un problema trivial. Generalmente, los modelos de predicción contemplan dos etapas, la primera de ellas, *etapa meteorológica*, consiste en determinar la velocidad del viento para un lugar en específico (desde un par de horas hasta días). La segunda etapa, consiste en la *conversión de energía*, esta etapa involucra la transformación de la velocidad del viento a potencia eléctrica.

2.1.3. Formulación del problema de la predicción de la generación eólica

Toda predicción puede ser vista como una forma de extrapolación, un modelo es construido y ajustado a un conjunto de datos, entonces, se obtienen datos totalmente nuevos, basados en la siguiente premisa: “dado un conjunto de datos y asumiendo que la dinámica identificada continuará en el futuro, se predice que...”. Para nuestro caso, nos enfocaremos en la evolución de nuestra variable de interés P_t , potencia eléctrica. Dicha variable es muestreada en el tiempo (usualmente con un intervalo de tiempo constante, por ejemplo: 1h) y su evolución es representada por una *serie de tiempo discreta* $\{P_t, t \in \mathbb{T}\}$ (donde \mathbb{T} es el conjunto de índices de tiempo y está usualmente contenido en \mathbb{Z}^+). Los valores de $\{P_t, t \in \mathbb{T}\}$ pueden ser valores instantáneos o promedios sobre intervalos de tiempo.

Para el caso de la predicción de la generación eólica, la variable de interés P_t es la potencia eléctrica disponible a la salida de la turbina eólica, a la salida de una granja eólica, o para un área con un gran número de granjas eólicas.

Series de tiempo de generación eólica pueden ser caracterizadas como series *no estacionarias* y *no lineales*. Decimos que una serie de tiempo tiene un comportamiento no estacionario, el cual puede ser descrito por sus momentos (media, desviación estándar,...), cuando estos evolucionan con el tiempo. La no linealidad de la serie de tiempo se debe al hecho que estas exhiben un comportamiento que no puede ser explicado a través de modelos lineales. Estas dos propiedades pueden ser observadas por inspección visual de las series de tiempo, Figura 2-1. En nuestro caso, la no estacionalidad está relacionada con la naturaleza propia del viento. La no linealidad se debe al proceso de conversión de energía. Aún cuando la velocidad del viento presentara un comportamiento lineal, al hacer la conversión a potencia, este cambiaría a uno no lineal debido a la forma de las curvas de potencia de la granja (Pinson, 2006).

La predicción del P_t -ésimo valor dado el tiempo t para el tiempo $t + k$, es descrito como

$\hat{P}_{t+k|t}$. Donde $\hat{P}_{t+k|t}$ es el resultado más probable dada la información hasta el tiempo t , siendo k el horizonte de la predicción. La información disponible en el tiempo t consiste en el conjunto de datos históricos ϕ_t .

En la literatura estadística, se hace una distinción entre modelos de una *única variable*, que asumen solamente valores pasados de P_t , este es nuestro caso, y modelos de *múltiples variables* que usan no solo los valores pasados en P_t , sino que además, valores pasados de otras variables. Estas variables se conocen como variables *descriptivas*, pues se espera que su variación ayude a explicar los cambios en P_t . En la predicción de la generación eólica, las variables descriptivas del sistema son principalmente la *velocidad del viento* y su *dirección*, sin embargo, temperatura, humedad y densidad del aire también pueden ser usadas.

Un método probabilístico está formalmente definido como una función de los datos disponibles:

$$\hat{P}_{t+k|t} = f(P_t, P_{(t-1)}, \dots, P_{(t-l)}, x_t, x_{(t-1)}, \dots, x_{t+k|t}) = f(\phi_t) \quad (2-1)$$

Donde ϕ_t es la información en el conjunto histórico de datos. Dependiendo del tipo de método de predicción empleado (de una o varias variables), ϕ_t puede o no contener variables descriptivas.

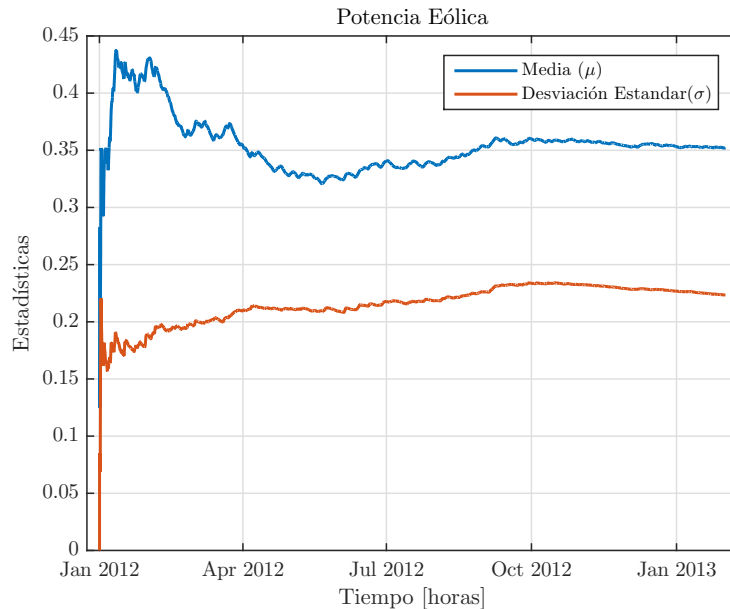


Figura 2-1: Variación, μ y σ con el tiempo para el grupo datos de entrenamiento.

Los métodos de predicción de la potencia eólica suelen ser clasificados en diferentes familias. Métodos simples basados en valores climáticos o valores promedios del histórico de producción son considerados como métodos de referencia debido a su facilidad de implementación e

interpretación. Estos métodos generalmente son usados como insumo de evaluación de otros métodos más elaborados. Los métodos más elaborados pueden ser divididos en dos grupos: *modelos numéricos del clima*, que toman en cuenta condiciones meteorológicas para obtener la potencia de salida de la granja, y *modelos de enfoque estadístico*, que analizan la relación entre el histórico de datos o predicciones meteorológicas, y la potencia de salida, sin tener en cuenta el comportamiento físico.

2.1.4. Métodos de referencia

Usualmente, el método más comúnmente usado en la predicción eólica, o en el campo de la meteorología en general, es el método de la *persistencia*³. Bajo este enfoque se dice que el estado futuro de la generación eólica será el igual al último valor medido, esto es:

$$\hat{P}_{t+k|t} = P_t \quad (2-2)$$

A pesar de la aparente simplicidad, este método es difícil de superar en las primeras horas del pronóstico (para horizontes de tiempo de hasta 4 y 6 horas). Esto es cierto para ubicaciones en las cuales los cambios en la atmósfera son lentos. Una generalización del método de persistencia se obtiene al reemplazar el último valor por el promedio de los n valores medidos.

$$\hat{P}_{t+k|t} = \frac{1}{n} \sum_{i=0}^{n-1} P_{t-i} \quad (2-3)$$

Este tipo de métodos son generalmente conocidos como predictores de media móvil. Cuando n tiende a infinito, estos tienden asintóticamente hacia la media global (también conocida como media climatológica).

$$\hat{P}_{t+k|t} = \hat{P}_t \quad (2-4)$$

El comportamiento de estos métodos ha sido estudiado analíticamente en Madsen et al. (2005). Se demostró que para largos horizontes de tiempo, el método climatológico es dos veces mejor que la persistencia. Consecuentemente, los autores propusieron la unión de estos dos métodos, así, obtener lo mejor del comportamiento de cada uno de ellos. Esto es:

$$\hat{P}_{t+k|t} = \rho_k P_t + (1 - \rho_k) \hat{P}_t, \quad (2-5)$$

³También conocido como “caminata aleatoria.”

Donde ρ_k es el coeficiente de correlación entre P_t y P_{t+k} . Una desventaja de este método es que el valor del coeficiente ρ_k debe ser ajustado usando algunas consideraciones y supuestos, debido a esto, en la práctica, no es ampliamente usado.

2.1.5. Modelo numérico del clima

La generación de energía desde el viento está directamente relacionada con las condiciones climáticas, así, un primer aspecto en la predicción de la generación eólica es la predicción de los valores futuros de las variables climatológicas donde la granja eólica está (o estará) establecida. Dicha predicción es realizable a través del uso de modelos de predicción numérica del clima. Los modelos numéricos del clima están basados en ecuaciones que describen el movimiento de fluidos (el movimiento de la masa de aire) y las fuerzas que lo causan. A partir del conocimiento del estado actual de la atmósfera, el sistema de ecuaciones permite estimar cuál será la evolución del estado de las variables tales como: temperatura, velocidad, humedad y presión en una serie de puntos del sistema.

La evaluación de las condiciones iniciales del sistema de un modelo numérico del clima se lleva a cabo a través de la evaluación de un gran número de medidas desde estaciones meteorológicas sinópticas, globos climatológicos y radiosondas. Los datos son procesados y usados para interpolar valores en áreas donde las medidas son escasas o no están disponibles. La calidad de los datos que sirven de condición inicial para el modelo numérico del clima es de vital importancia. Una estimación deficiente hace que los resultados de modelo sean poco exactos.

La formulación matemática de la evolución de la atmósfera se representa a través de ecuaciones diferenciales no lineales. El sistema de ecuaciones no tiene solución analítica, sin embargo, puede ser solucionada a través de métodos numéricos. Las ecuaciones se solucionan tanto en espacio como en tiempo: las derivadas en función del tiempo son reemplazadas por diferencias finitas, mientras que las derivadas en función del espacio son reemplazadas por diferencias finitas de la red o métodos espectrales (Buizza, 1996). La distancia entre los puntos de la red es llamada *resolución espacial* del modelo numérico del clima. Típicamente, la red tiene una resolución espacial que va desde unos pocos kilómetros hasta 50 km para modelos a mesoescala, mientras que la resolución temporal esta entre 1 hora y 3 horas. Aun con observaciones perfectas y modelos, la naturaleza caótica de atmósfera hace imposible predecir la evolución de las variables meteorológicas más allá de las dos semanas. El objetivo de los modelos numéricos del clima no esta en la descripción de la fluctuación de la potencia en el intervalo de tiempo de la predicción, para ello, el uso de métodos estocásticos es mas aconsejable. Estos último permiten describir el comportamiento de la potencia de salida de acuerdo con las fluctuaciones y turbulencia de la velocidad del viento.

En los modelos numéricos del clima, debido a la resolución espacial, no es posible averiguar sobre los efectos locales del viento a nivel de una granja eólica. Una solución a ello, es extrapolar la predicciones meteorológicas a nivel de la granja de generación, teniendo en

cuenta los accidentes geográficos propias del terreno donde ésta se encuentra localizada. A este proceso se le conoce como *reducción de escala* (downscaling), para realizar una reducción de escala, es utilizada una metodología basada en dinámica computacional de fluidos, una descripción completa se encuentra en (Lange and Focken, 2006).

El proceso de reducción de escala, también involucra un modelamiento del perfil del viento, dicho perfil describe las variaciones de la velocidad media del viento \bar{u} como una función de la altura z sobre el nivel del terreno. El modelo clásico es el modelo logarítmico presentado en la ecuación 2-6

$$\bar{u} = \frac{u_*}{k_v} \ln \left(\frac{z}{z_0} \right), \quad z \geq z_0, \quad (2-6)$$

Donde k es la constante de Von Kármán (típicamente 0.04), u_* es la velocidad de fricción, z_0 esta relacionada con la rugosidad de la superficie, su magnitud varia de acuerdo con el tipo de terreno (0.0002 en la superficie del mar hasta 1 para terrenos con presencia de grandes obstáculos). Reescribiendo la ecuación 2-6 en términos de dos diferentes alturas z_1 y z_2 y después de un simple manipulación, una relación entre la velocidad media del viento y las dos alturas es obtenida:

$$\bar{u}(z_2) = \frac{\ln \left(\frac{z_2}{z_0} \right)}{\ln \left(\frac{z_1}{z_0} \right)} \bar{u}(z_1) \quad (2-7)$$

El uso de la función logarítmica descrita en 2-7, radica en que la velocidad del viento puede ser fácilmente escalable con la altura, multiplicando el valor de la velocidad del viento por un factor que esta dado en función de las dos alturas y el nivel de rugosidad, en Troen and Petersen (1989) se presenta una análisis de los factores a tener en cuenta en la elección de la ubicación y evaluación de granjas eólicas.

Aún cuando este modelo es ampliamente usado y aceptado, su uso debería limitarse a terrenos planos y modelos en tierra. En terrenos complejos, muchas veces los resultados no están de acuerdo con las observaciones y técnicas más avanzadas como la dinámica computacional de fluidos se prefieren (Liu et al., 2016). Además, para granjas fuera de la tierra (off-shore) el logaritmo no es suficiente para describir el perfil de viento, efectos térmicos también deben ser incluidos.

2.1.6. Enfoque estadístico

Los métodos estadísticos de predicción están basados en uno o varios modelos que tratan de establecer la relación entre valores históricos de potencia, así como también, valores históricos y predicciones de las variables descriptivas del sistema. Todos los métodos estadísticos están

basados en el ajuste de un modelo particular al conjunto de datos usados para entrenar el sistema. Este ajuste se reduce a un problema de optimización, donde la función de pérdida a ser minimizada es función del error en la predicción.

Cuando se aplican metodologías de predicción estadística, se tiene dos alternativas: por una parte se pueden diseñar modelos para cada paso de la predicción. Así, para un modelo de n pasos, se deben diseñar n modelos específicos; por otra parte, se podría diseñar un único modelo y usarlo iterativamente para cada paso de la predicción. Esto significa, que la salida del primer paso servirá como entrada a la hora de calcular el segundo paso y así sucesivamente hasta completar el número de pasos de la predicción, como se verá más adelante. La ventaja de la última solución consiste en el entrenamiento y configuración de un único modelo, sin embargo, los errores son acumulativos de un paso a otro dado que el modelo se alimenta con estimaciones y no medidas.

Los modelos estadísticos están usualmente compuestos de una parte autorregresiva, que captura la persistencia del viento, y una parte meteorológica, que consiste en una transformación no lineal de la predicción las variables meteorológicas. Comúnmente, se hace uso de modelos numéricos del clima como entrada al sistema, pues es imposible considerar predicciones de hasta tres días de antelación sin la tener en cuenta la predicción del clima Pinson (2006).

2.1.6.1. Clasificador no paramétrico

Uno de los clasificadores más simples es el K_p -ésimo vecino más cercano (K_p -nearest neighbors) o simplemente KNN. Este clasificador mira K_p puntos en el grupo de datos de entrenamiento del modelo que están más cerca de la entrada x bajo observación. Más formalmente, tenemos que:

$$P(y = c|x, D, K_p) = \frac{1}{K_p} \sum_{i \in N_k(x, D)} \mathbb{I}(y_i = c) \quad (2-8)$$

Donde $N_k(x, D)$ son los índices del K_p -ésimo punto más cercano a x en D e $\mathbb{I}(e)$ es un indicador definido como:

$$\mathbb{I}(e) = \begin{cases} 1 & \text{Si } e \text{ es verdadero} \\ 0 & \text{Si } e \text{ es falso} \end{cases} \quad (2-9)$$

En Robert (2014) se presenta un desarrollo completo y detallado de este tipo de clasificador.

2.1.6.2. Ajuste a través de Kernel Density Estimation

Esta técnica usualmente es conocida como ventana de Parzen-Rosenblatt. La idea básica es estimar la función de densidad para un punto x usando las observaciones en vecindad de x . La distribución del kernel es definida a través de una función de pesos $K(x)$ y un ancho de banda h que controla la suavidad de la curva de densidades resultante, mientras que un

histograma discretiza los valores en barras separadas, el kernel suma las funciones de peso para cada uno de los valores para producir una curva de probabilidad continua y suave. En Zambom and Dias (2012) se presenta un desarrollo completo de este tipo de estimador de densidad y en Xydias et al. (2017) un aplicación del modelo basado en Kernel Density Estimation es usada en el despacho de generadores a gas.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (2-10)$$

Entonces, el kernel estimado es:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right) \quad (2-11)$$

En la Tabla **2-2** se presenta algunas de las funciones mas utilizadas, en la Figura **2-2** se presenta la forma de los kernels.

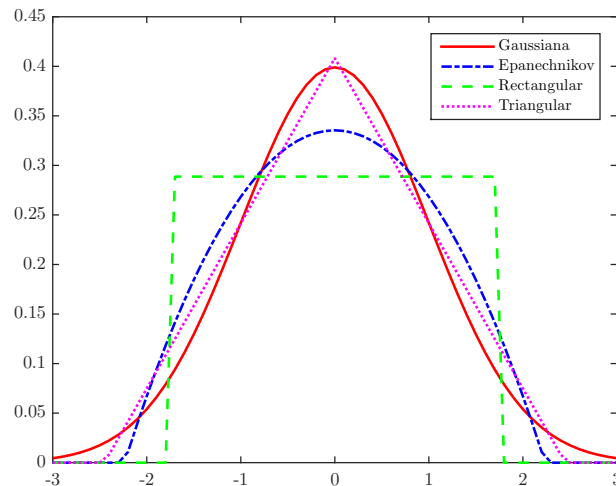


Figura 2-2: Funciones de peso

Tabla 2-2: Funciones típicas de peso.

Kernel	K(x)
Gaussiana	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$
Epanechnikov	$\frac{3}{4}(1-x^2), \{ x \leq 1\}$
Rectangular	$\frac{1}{2}, \{ x \leq 1\}$
Triangular	$\frac{35}{32}(1-x^2)^3, \{ x \leq 1\}$

2.1.7. Índices de desempeño

Debido a la naturaleza variante del viento, la verificación de los modelos usados en la predicción de generación eólica de corto plazo resulta poco trivial. En los modelos estadísticos, los datos juegan un papel clave a la hora de obtener buenos resultados, además, existe una dependencia entre el error asociado a la predicción (RMSE) y el horizonte de tiempo de la predicción. En Madsen et al. (2005) se presenta un conjunto de criterios y protocolos adecuados para la evaluación de la predicción de la generación de potencia eólica. Ellos analizan principalmente tres criterios: consistencia, calidad y valor. Un modelo clásico de predicción del error se define como la diferencia entre el valor medido y el valor predicho. A continuación se presenta un resumen de los modelos estándar de medición del error. Una revisión de los criterios de evaluación en predicción de la generación eólica es descrita en: Zhao et al. (2011), Hou et al. (2001) y Lange (2005).

- Error Medio (EM o MAE)

$$EM_k = \frac{1}{N} \sum_{t=1}^N e_{t+k|t} \quad (2-12)$$

- Error Medio Absoluto Normalizado (EMAN o NMAE)

$$EMAN_k = \frac{1}{P_{max}} \frac{1}{N} \sum_{t=1}^N |e_{t+k|t}| \quad (2-13)$$

- Error Porcentual Medio Absoluto (EPMA o MAPE)

$$EPMA = \frac{1}{N} \sum_{t=1}^N \left| \frac{P_{t+k} - \hat{P}_{t+k}}{P_{t+k}} \right| 100 \quad (2-14)$$

- Desviación Estándar de los Errores (DEE o SDE)

$$DEE_k = \sqrt{\frac{\sum_{t=1}^N [e_{t+k|t} - \bar{e}_k]^2}{N - 1}} \quad (2-15)$$

Donde, $e_{t+k|t} = P_{t+k} - \hat{P}_{t+k}$ es el error correspondiente al tiempo $t + k$ para la predicción hecha al tiempo t .

2.2. Objetivo de la Tesis

El objetivo de esta tesis es contribuir con el avance de conocimiento en predicción eólica a corto plazo a través de la implementación de un modelo predictivo probabilístico.

2.2.1. Objetivos Específicos

- Construir una base de datos con los datos reales de generación eólica, para que estos sean fácilmente manipulables y comprables con los datos del modelo propuesto.
- Desarrollar un modelo computacional que permita analizar el comportamiento de la generación eólica de corto plazo.
- Obtener escenarios que permitan informar sobre incertidumbre de la predicción eólica de corto plazo de acuerdo con la ubicación geográfica y el tiempo.

2.3. Solución propuesta

En enfoque usado en la predicción probabilística de corto plazo en generación eólica ha sido mediante la generación de escenarios de potencia agregada a través del uso de un modelo probabilístico construido a partir de las series de tiempo de la generación eólica. Una ventaja del presente método radica en la independencia de otras variables descriptivas del sistema (dirección del viento, temperatura, humedad, presión, etc).

El modelo está dividido en tres partes: Fase I, Fase II y Fase III, respectivamente.

- **Fase I:** el conjunto de *Datos de entrada* incluye una serie de tiempo de 10200 mediciones horarias de la potencia eólica en [p.u.], normalizadas entre 0 y 1, de acuerdo con la capacidad instalada de la granja de generación. El primer paso en cualquier proceso de predicción consiste en entender con que tipo de datos se esta trabajando. Una forma fácil y efectiva consiste en crear una visualización de los mismos, es decir *graficar los datos* de la serie de tiempo de la potencia eólica. Métodos más avanzados hacen uso de procesamiento de señales o técnicas de agrupamiento para identificar patrones. El clasificador KNN nos permite actuar sobre la series de tiempo de potencia eólica y velocidad horarias e identificar agrupamiento de datos. Este clasificador nos permite

remover algunos valores atípicos⁴ de la serie de tiempo de potencia eólica (*en el Conjunto de datos de entrenamiento*).

Después de la exploración y clasificación inicial de los datos, estos pueden ser cargados en múltiples directorios para su posterior uso en el entrenamiento y validación del modelo de predicción. El conjunto de datos, fue separado en dos: *conjunto de datos de entrenamiento* y *conjunto de datos de evaluación*. El conjunto de datos de validación representa los valores actuales (o valores de la magnitud de la potencia eólica a la salida de la granja de generación a ser estimados), y son usados para evaluar el comportamiento del modelo.

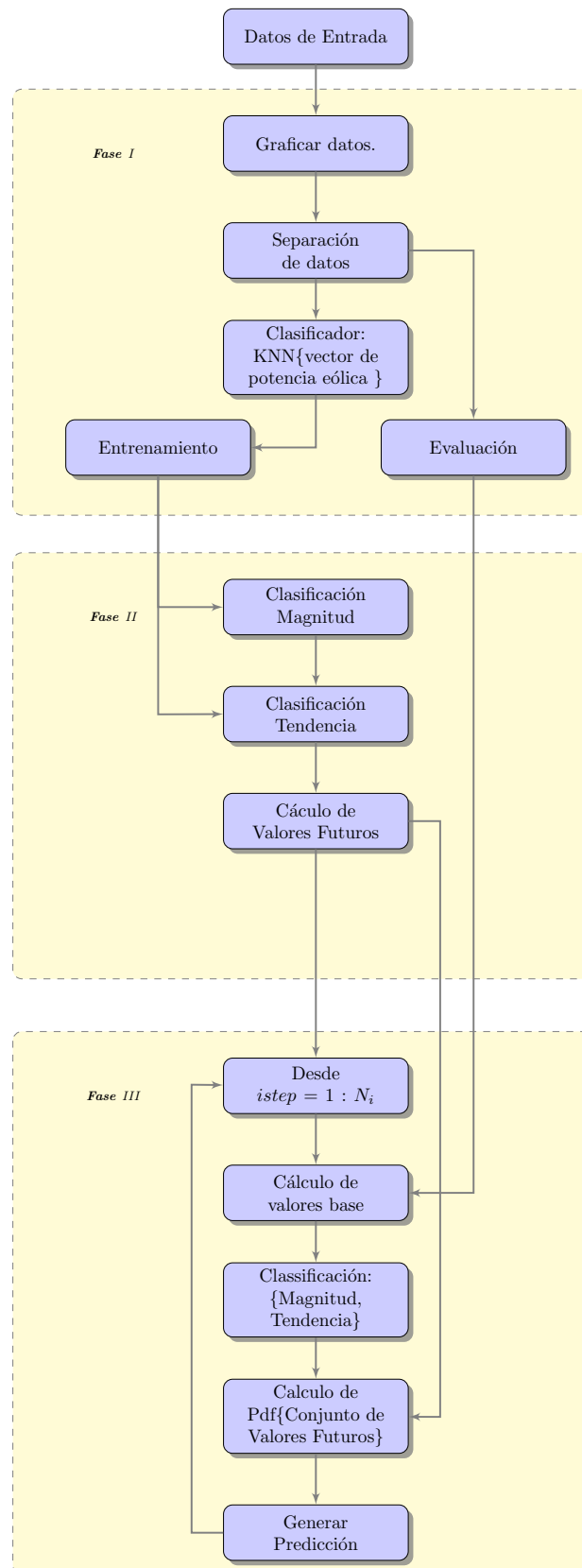
- **Fase II:** durante la etapa de entrenamiento, la relación entre dos valores consecutivos del conjunto de datos de entrenamiento es identificada. Cada valor de la serie de tiempo de generación eólica es clasificado de acuerdo con su magnitud en N intervalos y tendencia (tres posible casos: *Incremento*, *Constante*, *Decremento*), a partir de ahora, nos referiremos al conjunto {magnitud y tendencia} como *clase*. En la etapa final de esta fase de entrenamiento los valores futuros son obtenidos basados en la clasificación de clases de cada uno de los N_i intervalos.
- **Fase III:** en la etapa tres, el modelo, en un paso, genera un valor normalizado de potencia basado en los dos últimos valores de la serie de tiempo. Las condiciones iniciales del modelo en la fase de predicción corresponden a los dos últimos valores de potencia del conjunto de datos de entrenamiento (para nuestro estudio $P_{\{9527\}}$ y $P_{\{9528\}}$). En un primer paso el método reconoce la magnitud de ambos valores, entonces, la tendencia del último valor es identificada. La última parte del modelo consiste en la generación de las funciones de densidad de probabilidad (pdf) del *Conjunto de Valores futuros* calculados en la Fase II. Estas son obtenidas aplicando la metodología de ajuste por Kernel (ver ecuaciones 2-10 y 2-11). Una vez la clase (magnitud y tendencia) es identificada, el modelo genera un valor aleatorio basado en la función de densidad de probabilidad correspondiente. Este proceso es repetido hasta completar N_i pasos. En la Figura 2-3 se presenta el proceso descrito anteriormente de manera esquemática.

2.4. Estructura del documento

El primer capítulo de este documento consiste en una breve revisión de los métodos y avances en la predicción de la generación eólica a lo largo de sus casi 40 años de historia. El segundo capítulo, ha buscado contextualizar al lector sobre el problema que se aborda y el marco conceptual que se utilizó para resolverlo. El objetivo principal del tercer capítulo es presentar de manera formal los algoritmos usados en el modelamiento del problema a través de pseudocódigos y diagramas de flujo de información. En el capítulo cuatro se presenta una descripción del conjunto de datos de entrada al modelo y los resultados obtenidos bajo el

⁴Para nuestro caso, datos extremos y poco frecuentes.

modelo de preicción propuesto. En el capítulo se aborda el estudio del desempeño del modelo a través del análisis de errores para cada hora (1h a 48h) y cada día (7 días de la semana) de acuerdo con la frecuencia de actualización. Finalmente el capítulo seis está dedicado a conclusiones, recomendaciones y trabajo futuro.

**Figura 2-3:** Esquema del modelo implementado.

2.5. Conclusión

En este capítulo, la propuesta de investigación, objetivos y solución propuesta, junto con los conceptos básicos necesarios fueron tratados.

La importancia de la predicción eólica de corto plazo se ve reflejada en el interés que hay tanto para el sector eléctrico como para toda la sociedad del suministro eléctrico en la adquisición de conocimiento sobre la conformación de un portafolio con alta penetración de generación de energía proveniente de granjas eólicas.

Con el fin de contribuir con el avance de conocimiento en predicción eólica de corto plazo, un enfoque basado en la generación de escenarios de potencia agregada a través de un modelo probabilístico, construido a partir de las series de tiempo de la generación eólica fue introducido. Con el uso del clasificador no paramétrico, ajuste a través de Kernel Density Estimation e índices de desempeño, se introduce las bases para el desarrollo de las tres fases del modelo probabilístico que será implementado para la predicción eólica de corto plazo.

3 Algoritmo de predicción: Árboles de decisión

Resumen

En este capítulo se presentan tres fases del algoritmo usado para realizar la predicción de la potencia eólica. *Fase I* hace referencia al preprocesamiento de los datos de entrada al modelo. En la *Fase II* se describe el ordenamiento y clasificación de los datos en magnitud y tendencia. Finalmente, en la *Fase III* se generan nuevos valores teniendo en cuenta los dos últimos valores del vector de potencia eólica. Para un mejor entendimiento, éste es presentado a través de pseudocódigo y dos esquemáticos.

La energía eólica es de lejos la fuente no convencional de energía que más llama la atención a investigadores e inversores. Sin embargo, en la práctica, numerosos retos deben ser solucionados a través del modelamiento y predicción de la generación eólica en varias escalas temporales (desde un par de horas, hasta días), para luego ser usadas como entradas en la toma de decisiones. En este capítulo se presenta el algoritmo implementado y algunas de sus características más relevantes. El algoritmo es descrito de manera tal que el lector que no esté familiarizado con el tema pueda fácilmente seguir su implementación. En la selección del algoritmo, la revisión histórica del tema fue especialmente útil a la hora de establecer las características deseadas. El algoritmo fue seleccionado basados en cada uno de los tres criterios: rapidez de la predicción, rapidez de entrenamiento y capacidad de memoria. En la Tabla **3-1** se presentan los tiempos medios obtenidos en la ejecución del algoritmo. Estos tiempos fueron registrados para variaciones en el tamaño del conjunto de datos de entrenamiento y el número de escenarios. La complejidad del modelo está estrechamente relacionada con el tiempo requerido para completar la simulación, así, basados en los resultados de la Tabla **3-1**, para 20 repeticiones del modelo, el tiempo medio necesario para completar los 20 escenarios requeridos (3360 puntos en total) fue de 32 segundos, mientras que para 1000 escenarios (para frecuencia de actualización desde 48h y hasta 2h) el tiempo medio de ejecución fue de 1450 segundos. Las simulaciones fueron llevadas a cabo en un equipo con las siguientes especificaciones técnicas: Intel core i7-5500U(2.4GHZ), 8GB de RAM sobre un sistema operativo Windows 10 Home 64-bit.

Tabla 3-1: Tiempos de ejecución del algoritmo en segundos.

N_i	Tamaño Datos	Num. Escenarios	Tiempo medio [s]
100	5928	1000	1450
	9528	20	32

3.1. Selección del algoritmo

No todos los métodos de predicción funcionan para cada problema, e identificar el algoritmo adecuado es un proceso de ensayo y error. Sin embargo, ser consiente de las características de cada algoritmo es conveniente a la hora de realizar la selección. En la Tabla **3-2** se listan algunas de las características de los algoritmos de clasificación más populares, siendo el clasificador de arboles de decisión quién presenta las características más deseadas en términos de rapidez de predicción, rapidez de entrenamiento y capacidad de memoria.

3.2. Fase I

En la primera fase del método, los datos de varias granjas de generación eólica son cargados al programa en múltiples directorios, como se puede apreciar en detalle en el Algoritmo 1. En total, el método trabaja con 10200 datos de potencia eólica (en [p.u.]) y velocidad del viento (en [m/s]) medidos con una resolución horaria. Es recomendable que el conjunto de datos de entrenamiento y validación del método estén guardados en carpetas separadas (MATLAB, 2018).

Algorithm 1 Fase I

1. 1: P_{in} ▷ Potencia eólica
 - 2: u_x ▷ Componente Zonal, diez zonas
 - 3: u_y ▷ Componente Meridional, diez zonas
 - 4: $u_{in} = \sqrt{u_x^2 + u_y^2}$
 - 5: $\hat{P} = P_{in}$
 - 6: $\hat{u} = u_{in}$
 - 7: $L = 9528$
 - 8: $P_{en} = \hat{P}(L)$ ▷ Dividir conjunto de datos. {en, va} = {Entrenamiento, Validación}
 - 9: $P_{va} = \hat{P}(L + 1 : end)$
 - 10: $u_{en} = \hat{u}(L)$ ▷ Dividir conjunto de datos
 - 11: $u_{va} = \hat{u}(L + 1 : end)$
 - 12: **function** KNNCLASSIFIER(u_{nen}, P_{nen}) = (u_{en}, P_{en})
 - 13: ⋮
 - 14: **end function**
-

Tabla 3-2: Algoritmos de predicción.

Algoritmo	Rapidez de predicción	Rapidez de entrenamiento	Uso de memoria	Sintonización	Evaluación General
Log-Regresión (Máquinas de Soporte Vectorial - SVM lineales)	rápido	rápido	pequeño	mínima	Útil para conjuntos de datos pequeños
Arboles de decisión	rápido	rápido	pequeño	moderada	Buen desempeño, en ocasiones presenta un sobreajuste
Log-Regresión (Máquina de Soporte Vectorial - SVM, no Lineales)	lento	lento	medio	moderada	Buena para problemas binarios, y largos conjuntos de datos
Vecino más cercano	moderado	mínimo	medio	mínima	Baja precisión, pero fácil de usar e interpretar
Bayes-Naïve	rápido	rápido	medio	moderada	Ampliamente usado en clasificación de texto y filtro de Spam
Ensembles	moderado	lento	variable	moderada	Alta precisión y bueno desempeño con conjuntos de datos de pequeño y mediano tamaño.
Redes Neuronales	moderado	lento	medio	alto	Popular para clasificación, compresión, reconocimiento y predicción.

Los valores de velocidad del viento vienen dados en una *componente zonal* u_x y una *meridional* u_y son convertidos a rapidez y dirección mediante las relaciones expresadas en 3-1 y 3-2, respectivamente. Las componentes u_x y u_y son proyecciones del vector de velocidad de viento sobre el eje este-oeste y eje norte-sur, ver Figura 3-1.

$$u = \sqrt{u_x^2 + u_y^2} \quad (3-1)$$

$$\alpha = \arctan\left(\frac{u_y}{u_x}\right), \quad 0^\circ \leq \alpha < 360^\circ. \quad (3-2)$$

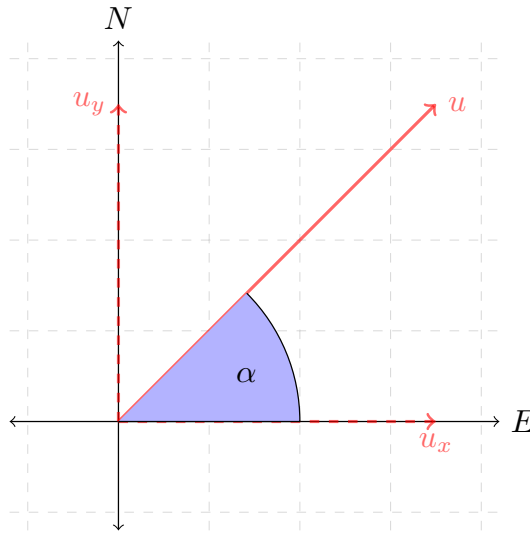


Figura 3-1: Descomposición del vector de viento u en componentes zonal u_x y meridional u_y , este-oeste, norte-sur, respectivamente.

Una vez los valores de velocidad y potencia son cargados, se obtiene el conjunto de datos de entrenamiento del modelo (P_{en}, u_{en}) , y el conjunto de datos de validación (P_{va}, u_{va}) . Con los datos divididos se realiza entonces un filtrado sobre el conjunto de datos de entrenamiento del modelo de predicción. Para ello se hace uso del clasificador KNN. Este clasificador nos permite realizar dos operaciones:

1. identificar agrupamiento de datos.
2. filtrar valores de Potencia-velocidad que no están ubicados en la región de mayor densidad (ver Figura 4-4).

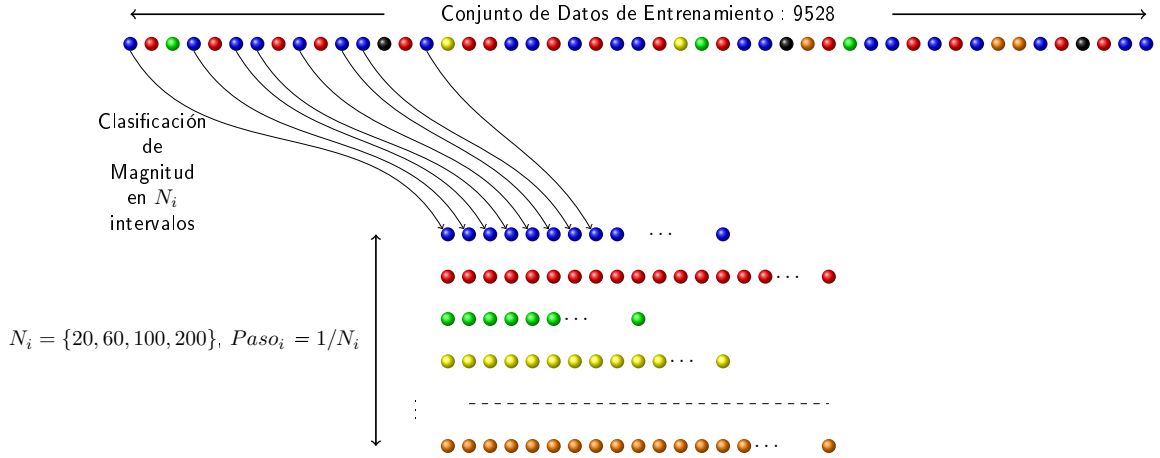


Figura 3-2: Fase II representación esquemática.

Al finalizar la Fase I del método, los datos están listos para ser procesados en la Fase II bajo el esquema establecido en la Figura 2-3.

3.3. Fase II

Durante la Fase II la relación entre dos valores consecutivos del conjunto de datos de entrenamiento es identificada, cada valor de la serie de tiempo es clasificado de acuerdo con su magnitud y tendencia (clase). La clasificación en magnitud se realiza en N_i intervalos (de igual longitud), asumiendo que el valor de P_{nen} ha sido estandarizado, los N_i intervalos estarán distribuidos entre 0 y 1. Para la clasificación de los valores de potencia eólica en tendencia, existen tres posibles clases: incrementos, decremento o constante. Esta clasificación se realiza teniendo en cuenta la clasificación en magnitud del valor en potencia inmediatamente anterior, ver Algoritmo 2, línea 13, 16 y línea 19. Si el valor de $P_{nen}(t_i)$ para el tiempo t_{i-1} es menor o mayor que la magnitud del intervalo, entonces la clase será *incremento* o *decremento*, respectivamente. Si el valor de $P_{nen}(t_i)$ y $P_{nen}(t_{i-1})$ están en el mismo intervalo de magnitud, entonces la clase será definida como *constante*. Basados en los N_i intervalos previamente definidos en esta etapa de entrenamiento, los valores futuros son calculados para cada clase así: si el valor $P_{nen}(t_i) \in N_i$ y su tendencia es "*incremento*", entonces su valor futuro será: $P_{nen}(t_{i+1}) \in N_i$ de este modo, cada valor futuro esta relacionado con una única clase. El la Figura 3-2 se presenta una representación esquemática del clasificador de la Fase II.

3.4. Fase III

Para esta fase, el modelo genera un valor puntual de la predicción basado en los dos últimos valores mas recientes $value_1$ (último valor generado o medido) y $value_2$ (penúltimo valor generado o medido) de la series de tiempo del vector de potencia eólica. Una vez que la clase de $value_1$ es obtenida, el modelo usa el *conjunto de valores futuros* para generar la correspondiente función de densidad de probabilidad, un valor aleatorio es generado. Este proceso es realizado N_i veces usando los dos últimos valores más recientes para generar el siguiente valor. Cada valor obtenido para el tiempo t_i es considerado como $value_1$ (con esto $value_2 = value_1$ para t_{i+1}). El resultado de la predicción es menos preciso a medida que el horizonte de tiempo incrementa. Con el fin de mejorar el desempeño del modelo, en terminos de precisión de la predicción, una herramienta que utiliza éste es la *frecuencia de actualización*. Así, el modelo actualiza los valores $value_1$ y $value_2$ con los valores más recientes de potencia eólica (valores tomados del conjunto de datos de entrenamiento). Los valores de la frecuencia de actualización son definidos de acuerdo con el nivel de precisión deseado.

Algorithm 2 Fase II

1. 1: N_i ▷ Número de intervalos
 - 2: $Paso = 1/N_i$
 - 3: **for** $mag = Paso : Paso : to 1$ **do** ▷ Clasificar magnitud
 - 4: $interval = find(P_{nen} \geq (mag - Paso) \ \& \ P_{nen} < mag)$
 - 5: **end for**
 - 6: **for** $m = 1 : to longitud(N_i)$ **do** ▷ Clasificar Tendencia
 - 7: $index = interval\{1, m\}$
 - 8: **for** $n = 1 : to longitud(index)$ **do**
 - 9: **if** $index(n) == 1 \ || \ index == upperlimit$ **then**
 - 10: flag interval = m
 - 11: flag index = n
 - 12: **else**
 - 13: **if** $P_{nen}(index(n) - 1) < m/N_i - Paso$ **then**
 - 14: Incremento = $P_{nen}(index(n))$
 - 15: Valor futuro incremento = $P_{nen}(index(n) + 1)$
 - 16: **else if** $P_{nen}(index(n) - 1) > m/N_i$ **then**
 - 17: Decremento = $P_{nen}(index(n))$
 - 18: Valor futuro decremento = $P_{nen}(index(n) + 1)$
 - 19: **else**
 - 20: Constante = $P_{nen}(index(n))$
 - 21: Valor futuro constante = $P_{nen}(index(n) + 1)$
 - 22: **end if**
 - 23: **end if**
 - 24: **end for**
 - 25: **end for**
-

El algoritmo 3 describe el conjunto de pasos necesarios para generar un nuevo valor de potencia eólica basado en la función de distribución de probabilidad. Para el primer paso de la predicción, los dos últimos valores del conjunto de datos de predicción son usados como valores iniciales de $value_1 = P_{nen}(end)$ y $value_2 = P_{nen}(end - 1)$. En los restantes pasos del proceso de predicción, el algoritmo verifica si el valor de t corresponde con la frecuencia de actualización, así, $value_1 = P_{va}(tf - 1)$ y $value_2 = P_{va}(tf - 2)$. En el caso específico cuando $tf = 2$, el modelo usa el último valor de $P_{nen}(end)$ como $value_2$ y el valor actual del conjunto de datos de validación como $value_1 = Foreval(tf - 1)$. O $value_1 = Foreval(tf - 1)$ y $value_2 = Foreval(tf - 2)$ en otro caso.

Un esquema completo del clasificador es presentado en la Figura **3-3**, en una primera fase del diagrama, se presentan los datos de entrada del modelo (que han sido previamente procesados en la Fase I), entonces, cada valor de la serie de tiempo de potencia eólica es clasificado en magnitud, de acuerdo al número de intervalos deseados. Una vez esta clasificación es obtenida, cada valor en los los intervalos de magnitud es clasificado en una de tres clases (incremento, decremento o constante). Bajo este procedimiento, al final se obtienen $3N$ conjuntos de datos. En donde cada valor corresponde al siguiente valor en el vector potencia eólica de cada uno de los datos previamente clasificados.

Dependiendo del tamaño del conjunto de datos de entrada al modelo se tiene un compromiso entre precisión en la predicción y requerimientos técnicos. Un conjunto de datos robusto y frecuencias de actualización cortas (2h, 4h y 6h) mejoran el desempeño del modelo, pero requieren un mayor poder computacional.

Algorithm 3 Fase III

```

1. 1: Day
   2:  $nsteps = Day * 48$  ▷ Número de puntos de la predicción
   3: repeticiones
   4: Frecuencia de actualización
   5: for  $r = 1$  to repeticiones do
   6:   for  $tf = 1 : nsteps$  do
   7:     if  $tf == 1$  then
   8:        $value_1 = P_{nen}(end)$ 
   9:        $value_2 = P_{nen}(end - 1)$ 
  10:     else if  $\text{mod}(tf, \textit{Frecuencia de actualización}) == 0 \parallel tf == 2$  then
  11:       if  $tf == 2$  then
  12:          $value_1 = \textit{Foreval}(tf - 1)$ 
  13:          $value_2 = P_{nen}(end - 2)$ 
  14:       else
  15:          $value_1 = P_{va}(tf - 1)$ 
  16:          $value_2 = P_{va}(tf - 2)$ 
  17:       end if
  18:     else
  19:        $value_1 = \textit{Foreval}(tf - 1)$ 
  20:        $value_2 = \textit{Foreval}(tf - 2)$ 
  21:     end if
  22:     for  $magnitud = 1$  to  $\textit{longitud}(N_i)$  do
  23:       if  $value_1 < \textit{magnitud}/N_i$  then
  24:         if  $value_2 < \textit{magnitud}/N_i - nsteps$  then
  25:            $\textit{Intervalo de datos} = \text{Valor futuro incremento}(\textit{magnitud})$ 
  26:         else if  $value_2 : \textit{magnitud}/N_i$  then
  27:            $\textit{Intervalo de datos} = \text{Valor futuro decremento}(\textit{magnitud})$ 
  28:         else
  29:            $\textit{Intervalo de datos} = \text{Valor futuro constante}(\textit{magnitud})$ 
  30:         end if
  31:          $pdfx_i = \textit{fitdist}(\textit{Intervalo de datos})$  ▷ Generar pdf
  32:          $\textit{Foreval}(tf) = \textit{random}(pdfx_i, 1)$  ▷ Generar valor futuro
  33:         break
  34:       end if
  35:     end for
  36:   end for
  37: end for

```

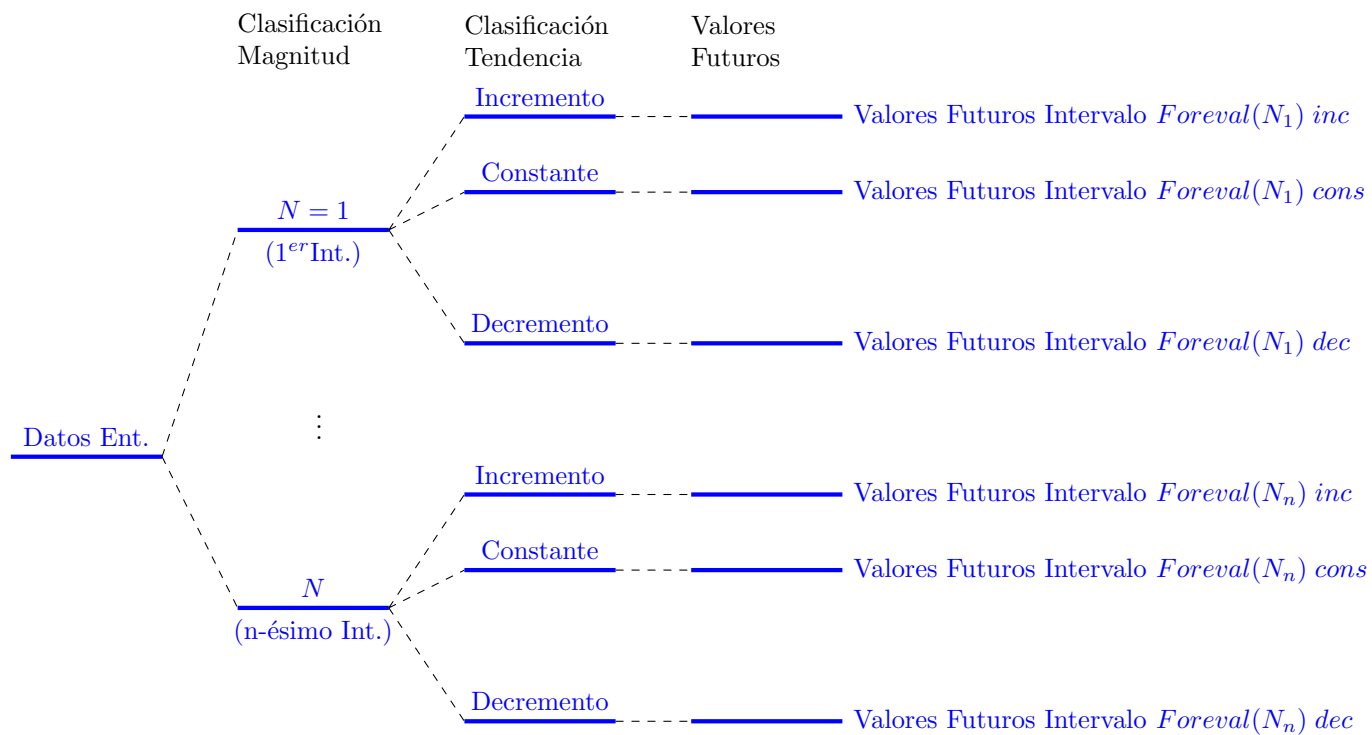


Figura 3-3: Representación esquemática Fase I, II y III del modelo.

3.5. Conclusión

En este capítulo se presentaron los algoritmos usados en la predicción de la potencia eólica. Cada una de las tres fases del algoritmo contribuye a la predicción de escenarios de generación en un horizonte de tiempo de hasta 48h. En la Fase I se realiza una limpieza y división de datos, en la Fase II los datos son clasificados de acuerdo con su magnitud y tendencia. Finalmente, en la Fase III el algoritmo es usado en cascada, basado en los dos últimos valores del agregado de generación eólica (que pueden ser tomados del conjunto de datos de entrenamiento, datos de validación o conjunto de valores estimados), en la estimación de valores futuros de generación eólica. Cabe notar que el tiempo de ejecución del algoritmo no afecta el propósito en tiempo real del modelo, ya que el horizonte de predicción más corto es de 2h. En el siguiente capítulo se presenta un caso de estudio basado en 10200 valores horarios de potencia eólica medida sobre diez granjas de generación.

4 Caso de estudio

Resumen

En la primera sección de este capítulo se presenta una descripción del conjunto de datos usados en el caso de estudio a través de estadísticas y gráficas. Además, se presentan los resultados obtenidos luego de usar el modelo de predicción de generación eólica de corto plazo. Bajo dos condiciones que corresponden a la variación en la frecuencia de actualización del modelo y número de intervalos tenidos en cuenta para la clasificación en magnitud. Adicionalmente, el desempeño del modelo es evaluado a través de tres indicadores de desempeño: error medio absoluto normalizado (EMAN), error porcentual medio absoluto (EPMA), y la desviación estandar de los errores (DEE). Estos índices han sido calculados para mil realizaciones del modelo, variando la frecuencia de actualización. Los resultados son presentados en forma gráfica y tabular horaria y diariamente.

Mediante este caso de estudio se presenta la aplicación del algoritmo desarrollado en la sección anterior a un conjunto en particular. Los datos son divididos en datos de entrenamiento y datos de validación de sistema. El modelo es usado entonces para predecir valores del agregado de potencia eólica en seis horizontes de tiempo (48h, 24h, 12h, 6h, 4h y 2h) y cuatro intervalos de magnitud (20, 60, 100 y 200). Adicionalmente, con el fin de indagar sobre la robustez del modelo, intencionalmente, en uno de los ejercicios, se ha suprimido del conjunto de datos de entrenamiento del sistema el conjunto de datos correspondientes a los meses de enero, febrero y marzo de 2012, así como también los meses de enero y febrero de 2013. En total el conjunto de datos de entrenamiento del sistema fue reducido al 62 %,

4.1. Descripción del conjunto datos

El modelo desarrollado fue aplicado sobre un conjunto de datos reales, datos del agregado de potencia eólica publicados en European Centre for Medium-range Weather Forecasts - ECMWF. El conjunto de datos del agregado de potencia eólica consiste en 10200 datos horarios (un año y tres meses aproximadamente) obtenidos de granjas eólicas ubicadas en

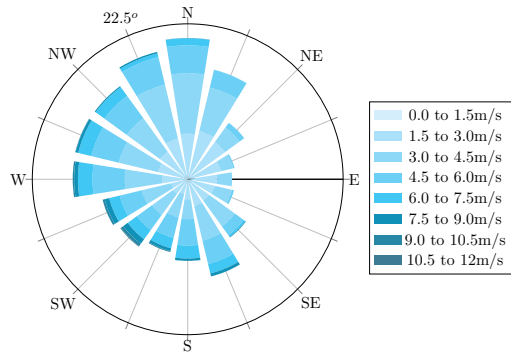


Figura 4-1: Rosas del viento a 10m.

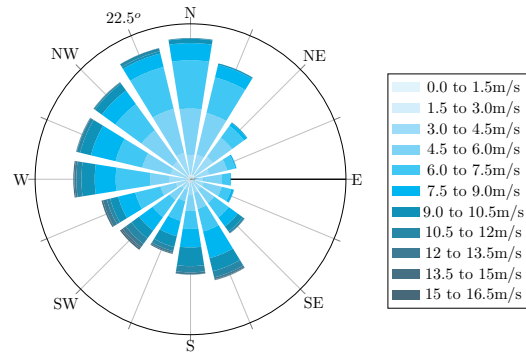


Figura 4-2: Rosas del viento a 100m.

Australia (se desconoce la ubicación), para el período del 1 de enero de 2012 hasta el 31 de marzo de 2013. Aun cuando el modelo únicamente contempla predicciones en dominio de tiempo (no espaciales), se dibuja a modo de ejemplo en la Figura 4-5 una posible distribución espacial de las granja eólicas. En la Figura 4-3 se observa la serie de tiempo de la potencia eólica, mientras que en la Tabla 4-1 se listan algunas estadísticas de los datos. Originalmente los datos vienen contenidos en archivos *.csv*, cada uno de los archivos contiene los siguientes atributos: {Zona, Hora, Tiempo, Potencia, Velocidades (meridional y longitudinal) }, a partir de aquí, el procesamiento de estos se ha realizado en MATLAB.

Tabla 4-1: Estadísticas descriptivas.

Parámetro	Valor
Valor objetivo [p.u.]	Potencia: P_t
Media μ_{10200}	0.3489
Desviación estándar σ_{10200}	0.2208
Valor mínimo P_{min} [p.u.]	0.0
Valor máximo P_{max} [p.u.]	0.9317
Horizonte de antelación	{2h, 4h, 6h, 12h, 24h, 48h}
Número de puntos conjunto de entrenamiento	9528
Número de puntos conjunto de evaluación	672
Tiempo entre medidas	1h
Número de mediciones en potencia (en [p.u.]) por granja	10200

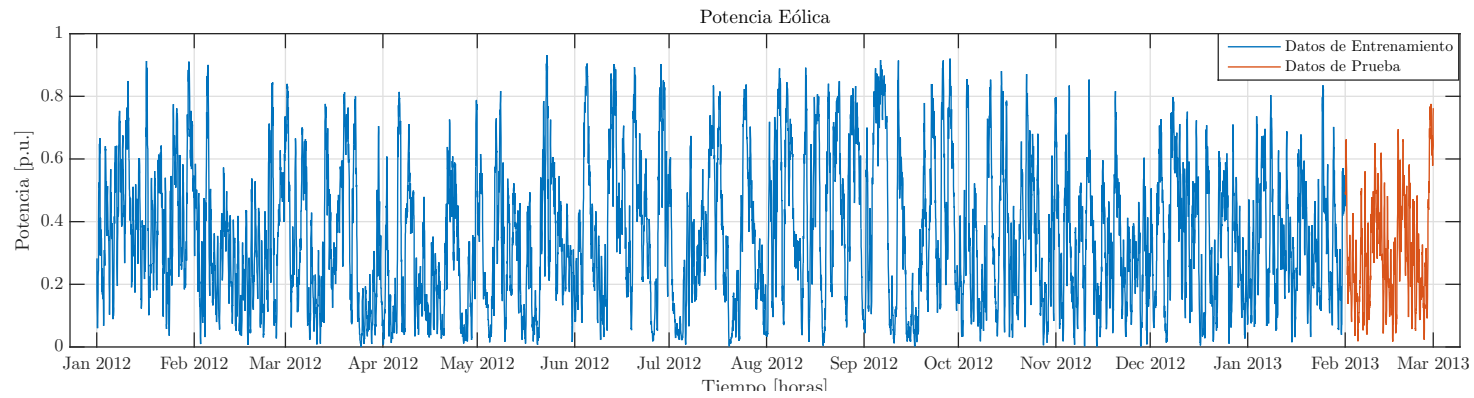


Figura 4-3: Serie de tiempo de potencia eólica. En azul, conjunto de datos de *entrenamiento* del modelo. En rojo, conjunto de datos de *validación* del modelo.

En la Figura 4-1 y 4-2 se presentan dos rosas del viento, una para valores de la velocidad del viento medidas a 10m y otra para 100m medidas sobre el nivel del terreno. La convención de colores hace referencia a la frecuencia de valores en velocidad registrados por cada intervalo. Los valores de velocidad zonal y meridional han sido convertidos en magnitud (rapidez) y dirección de la velocidad del viento haciendo uso de las relaciones presentadas en 3-1 y 3-2. Adicionalmente, los valores en potencia son los mismos para ambas alturas (10m y 100m), en total se tiene medidas del agregado de potencia eólica en diez granjas, el vector final de potencia eólica (ver Figura 4-3) fue obtenido promediando las medidas de potencia eólica sobre las diez granjas de generación eólica.

En un primer paso, los datos fueron procesados de acuerdo con la metodología descrita en la Figura 2-3 y Algoritmo 1. Los datos normalizados son separados en dos grupos: *Conjunto de datos de entrenamiento*, en azul, y *Conjunto de datos de validación*, en rojo Figura 4-3. Los datos correspondientes a enero de 2012 y hasta febrero 2013 fueron utilizados en el entrenamiento del modelo, mientras que los últimos 672 datos (mes de marzo 2013), se utilizaron para la evaluación del mismo. Durante la *Fase II* (etapa de entrenamiento, Algoritmo 2), los datos fueron clasificados respecto a su magnitud y tendencia (incremento, constante o decremento). La clasificación de magnitud se realizó en intervalos de distintos tamaños $N_i = \{20, 60, 100, 200\}$ con pasos $Paso_i = 1/N_i$, en $\{60, 180, 300, 600\}$ clases y $\{48h, 24h, 12h, 6h, 4h, 2h\}$ frecuencias de actualización.

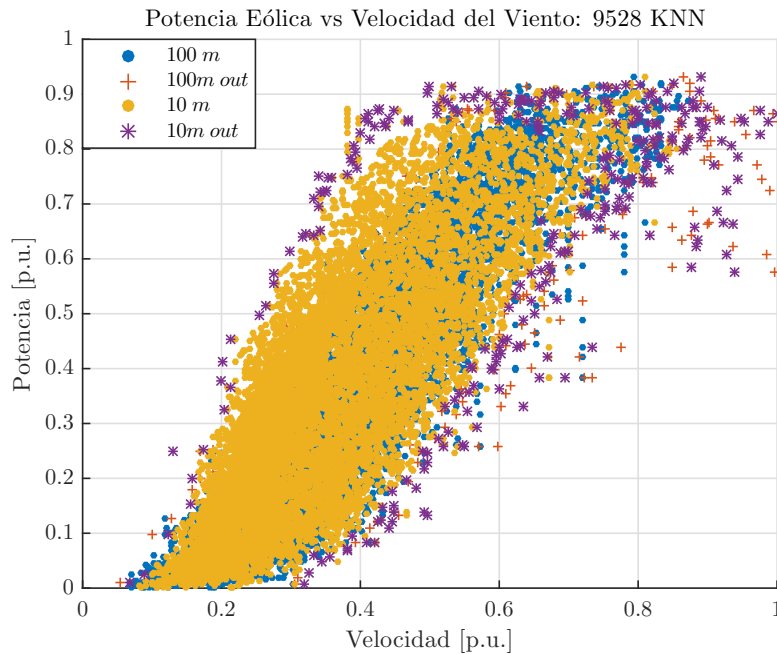


Figura 4-4: Potencia eólica vs. Velocidad.

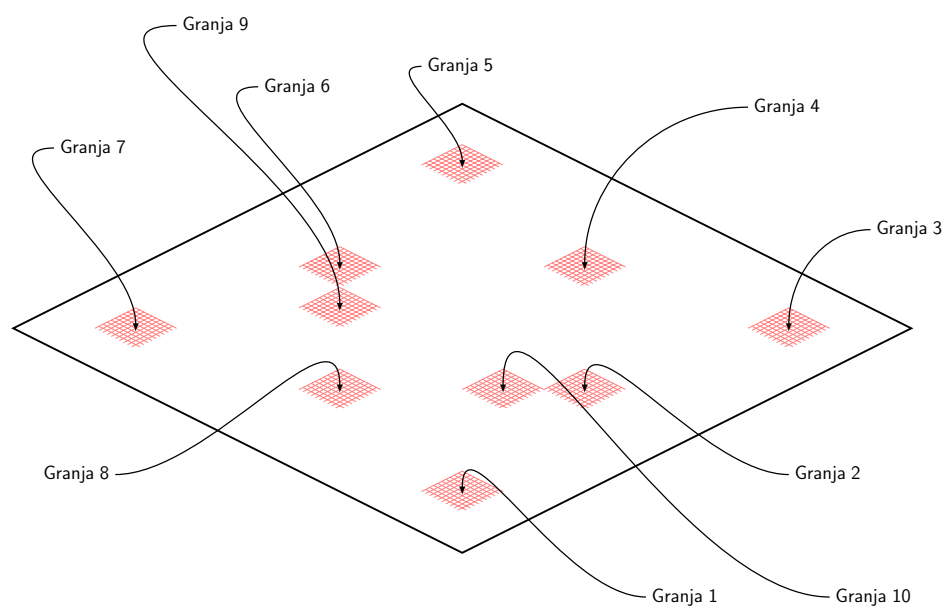


Figura 4-5: Ubicación virtual de las granjas eólicas.

4.2. Predicción

En la *Fase II*, el conjunto de datos de entrenamiento fue clasificado de acuerdo a su magnitud y tendencia. Las Figuras 4-6, 4-7, 4-8, 4-9, 4-10 y 4-11, fueron generadas para intervalos de longitud 0.01 ($N_i = 100$) y tres tendencias (incremento, constante y decremento), entonces, un árbol de clasificación es obtenido con un total de 300 conjuntos de valores futuros (como el presentado en 3-3). Usando 6 distintas frecuencias de actualización {2h, 4h, 6h, 12h, 24h y 48h}, un diferente desempeño en el modelo es obtenido. Para cada una de las frecuencias de actualización el modelo de predicción de la generación eólica entrega un total de 20 escenarios, resultando en 120 predicciones para el mes de marzo de 2013 (336 valores generados en una corrida por escenario). Si el modelo es actualizado con una frecuencia de 2h, entonces el error entre el valor generado mediante el modelo (valor estimado) y el valor real disminuye en comparación con los valores generados por el modelo cuando las frecuencias de actualización son menores.

El número de intervalos de magnitud N es otro factor importante que afecta el desempeño del modelo. El proceso fue repetido para intervalos con magnitud {20, 60 y 200} (ver Apéndice A). Se observa que el efecto de la magnitud del intervalo está directamente relacionado con la frecuencia de actualización. Para valores de la frecuencia de actualización entre 2h y 6h el efecto del tamaño del intervalo se hace despreciable.

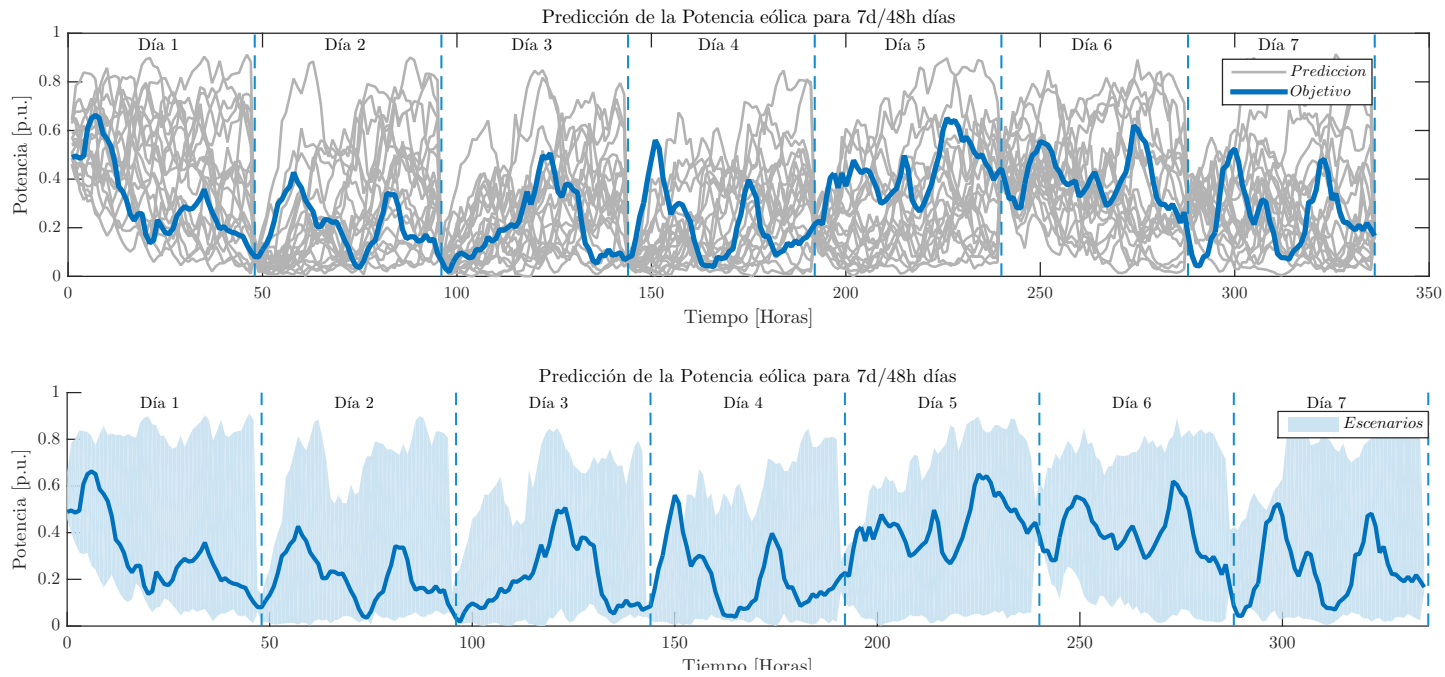


Figura 4-6: Veinte escenarios para siete días con frecuencia de actualización cada 48h. El área sombreada envolvente en la figura inferior fue construida tomando el valor mínimo y máximo horario de los escenarios

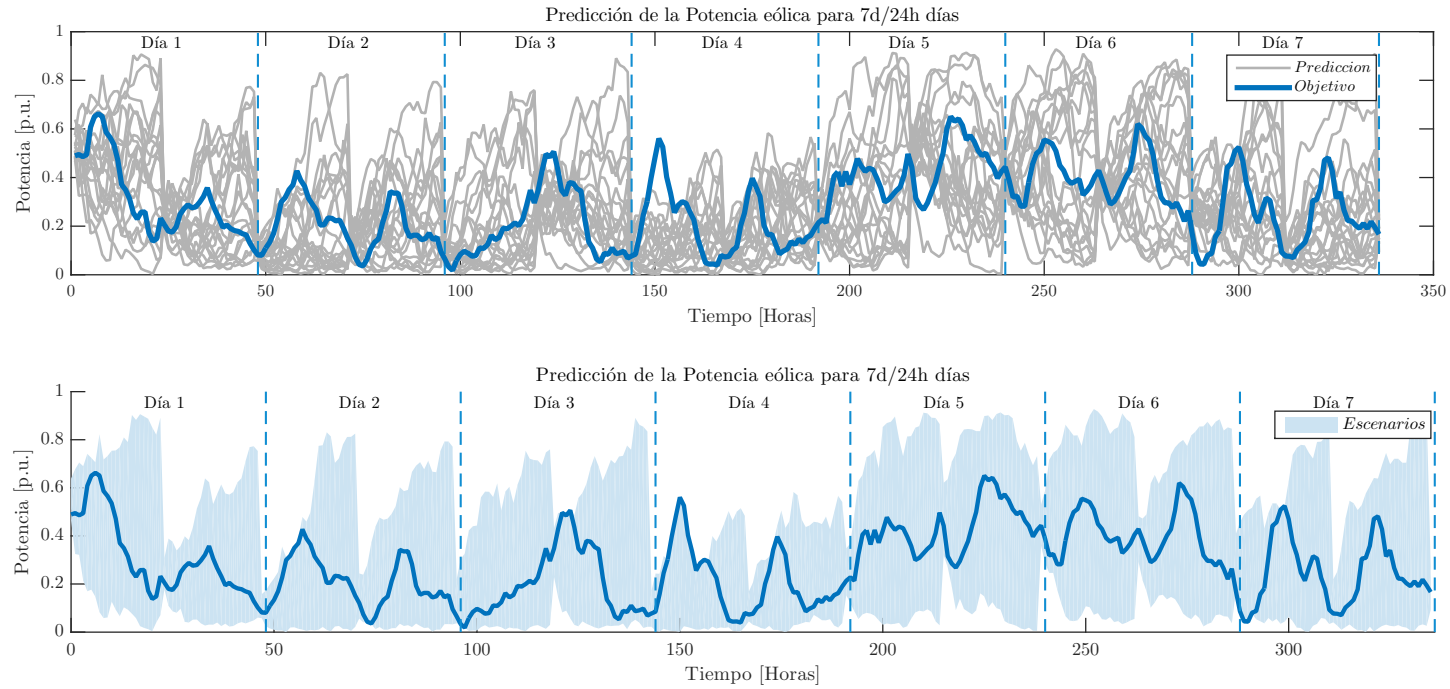


Figura 4-7: Veinte escenarios para siete días con frecuencia de actualización cada 24h. El área sombreada envolvente en la figura inferior fue construida tomando el valor mínimo y máximo horario de los escenarios.

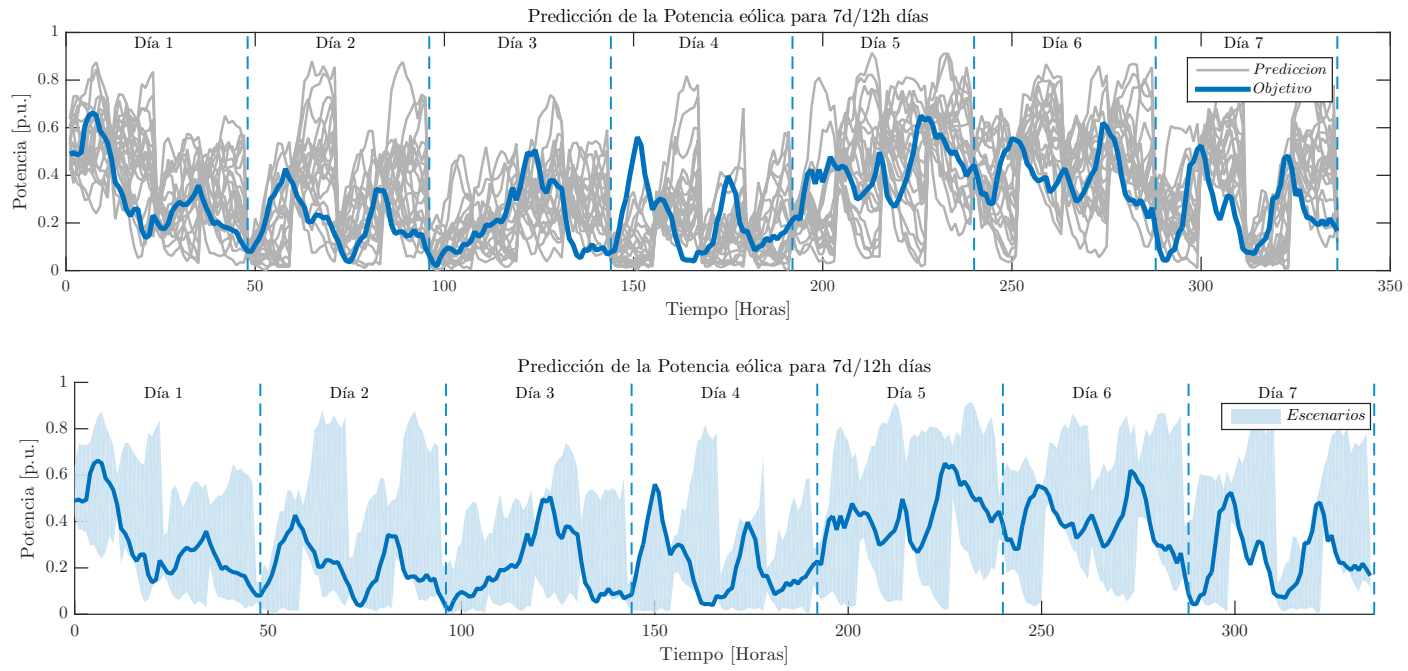


Figura 4-8: Veinte escenarios para siete días con frecuencia de actualización cada 12h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.

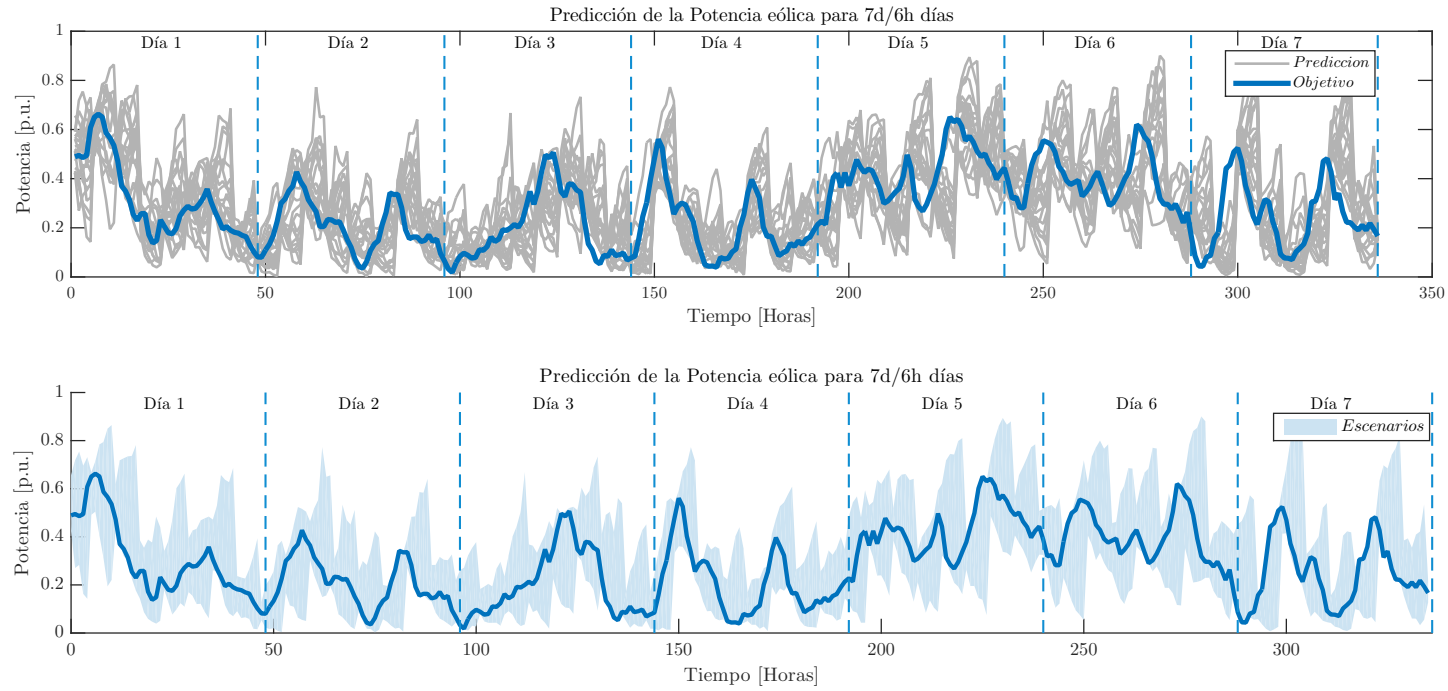


Figura 4-9: Veinte escenarios para siete días con frecuencia de actualización cada 6h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.

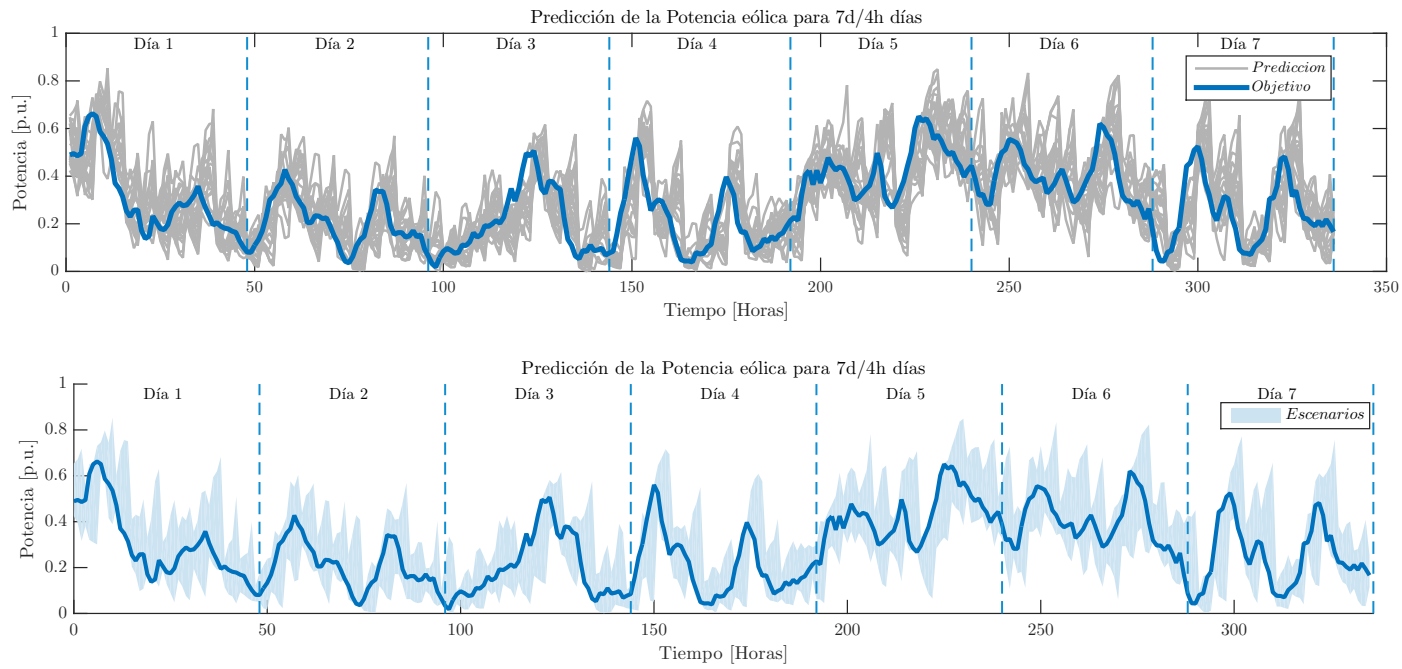


Figura 4-10: Veinte escenarios para siete días con frecuencia de actualización cada 4h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.

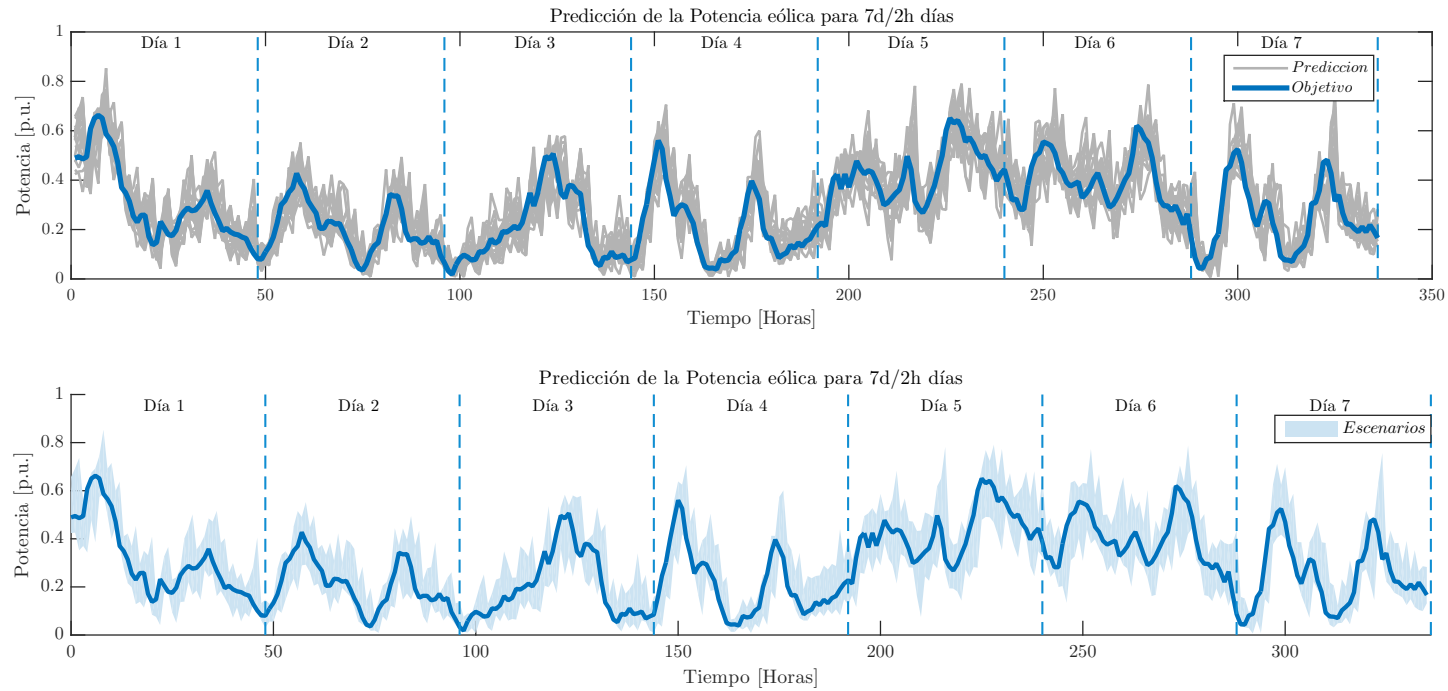


Figura 4-11: Veinte escenarios para siete días con frecuencia de actualización cada 2h. El área sombreada envolvente fue construida tomando el valor mínimo y máximo horario de los escenarios.

4.3. Desempeño del modelo

En la Tabla 4-2 se presentan los índices de desempeño usados en la evaluación del modelo de predicción de generación eólica. Como se aprecia en las Figuras 4-6, 4-7, 4-8, 4-9, 4-10 y 4-11, el desempeño del modelo mejora a medida que la frecuencia de actualización tiende a 1h (paso de la muestra), lo cual es consecuente con la presentado en la Tabla 4-2 donde los índices de desempeño del modelo se hacen más pequeños cuando la frecuencia de actualización varía de 48h a 2h. Debido a la estructura del modelo, es posible la propagación del error. El error asociado a la predicción P para el tiempo t contribuye al valor estimado de P para el tiempo $t + k$ creando un error en cascada. La propagación del error se ve interrumpida cuando el modelo es actualizado con los valores $value_1$ y $value_2$ del agregado de potencia eólica P_{va} , entonces, el modelo genera un nuevo valor de P sin tener en cuenta los errores previos.

La fluctuación de la potencia eólica se hace mayor en el día 3, día 4 y día 7, razón por la cual el error es mayor. Por otra parte, la potencia generada en el día 5 y día 6 presenta menores variaciones, razón por la cual los índices de desempeño se ven reducidos. El valor del EPMA varía entre 49.29% hasta 135.20% para una frecuencia de actualización de 48h. Cuando la frecuencia de actualización varía de 48h a 2h, el EPMA varía entre 17.16% y 41.00%, con un mejoramiento medio de 63.17% en el desempeño.

En la Figura 4-12 se presenta la distribución diaria de los errores. Para cada una de las cajas, la línea roja representa la mediana, y los ejes de las cajas azules representan el 25_{avo} y 75_{avo} percentil. Los errores horarios fueron calculados y los resultados se pueden apreciar en la Figura 4-13, el modelo se ejecutó 1000 veces para cada una de las frecuencias de actualización. En todos los casos la mediana es cercana a cero para cada hora, lo cual indica que el modelo está libre de errores sistemáticos. El incremento en la frecuencia de actualización se traduce en un menor error en la predicción, mejorando la precisión del modelo de potencia eólica P para el tiempo P_{t+k} .

Adicionalmente, para mejorar el desempeño del modelo, el conjunto de datos de entrenamiento debe ser actualizado con los datos disponibles más recientes. En este sentido, las predicciones del modelo no son precisas cuando éste ha sido entrenado con datos del invierno, pero se quieren predecir datos del verano. En la Figura 4-14 se presenta el comportamiento de los errores absolutos medios para cada una de las frecuencias de actualización. Es visible la mejora que imponen las frecuencias de actualización mayores en el error medio absoluto, éste tiende a decrecer para valores de la frecuencia de actualización que van desde las 48h, 24h, 12h, 6h, 4h y 2h.

Tabla 4-2: Índices de desempeño.

Frecuencia de Actualización	Index	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7
Cada 48 horas	EPMA(%)	94.23	115.04	127.56	137.29	49.29	50.61	135.20
	SDE	0.12	0.13	0.13	0.18	0.09	0.10	0.13
	NMAE	0.32	0.26	0.25	0.29	0.32	0.28	0.31
Cada 24 horas	EPMA(%)	69.07	83.82	124.98	111.94	44.87	46.14	97.48
	SDE	0.11	0.11	0.13	0.16	0.12	0.10	0.14
	NMAE	0.25	0.21	0.23	0.25	0.28	0.28	0.25
Cada 12 horas	EPMA(%)	72.95	90.82	123.74	111.70	37.81	46.05	114.12
	SDE	0.14	0.13	0.14	0.17	0.12	0.12	0.18
	NMAE	0.25	0.22	0.22	0.23	0.23	0.24	0.30
Cada 6 horas	EPMA(%)	41.55	66.62	85.68	65.73	31.79	35.09	90.37
	SDE	0.08	0.10	0.11	0.09	0.13	0.13	0.18
	NMAE	0.16	0.16	0.16	0.15	0.19	0.20	0.22
Cada 4 horas	EPMA(%)	35.03	51.69	68.73	58.73	25.66	25.65	70.62
	SDE	0.08	0.07	0.10	0.12	0.09	0.08	0.12
	NMAE	0.14	0.12	0.13	0.15	0.15	0.14	0.17
Cada 2 horas	EPMA(%)	25.79	33.58	41.00	36.30	17.16	19.42	37.85
	SDE	0.06	0.04	0.06	0.07	0.06	0.06	0.08
	NMAE	0.10	0.08	0.09	0.10	0.10	0.11	0.11

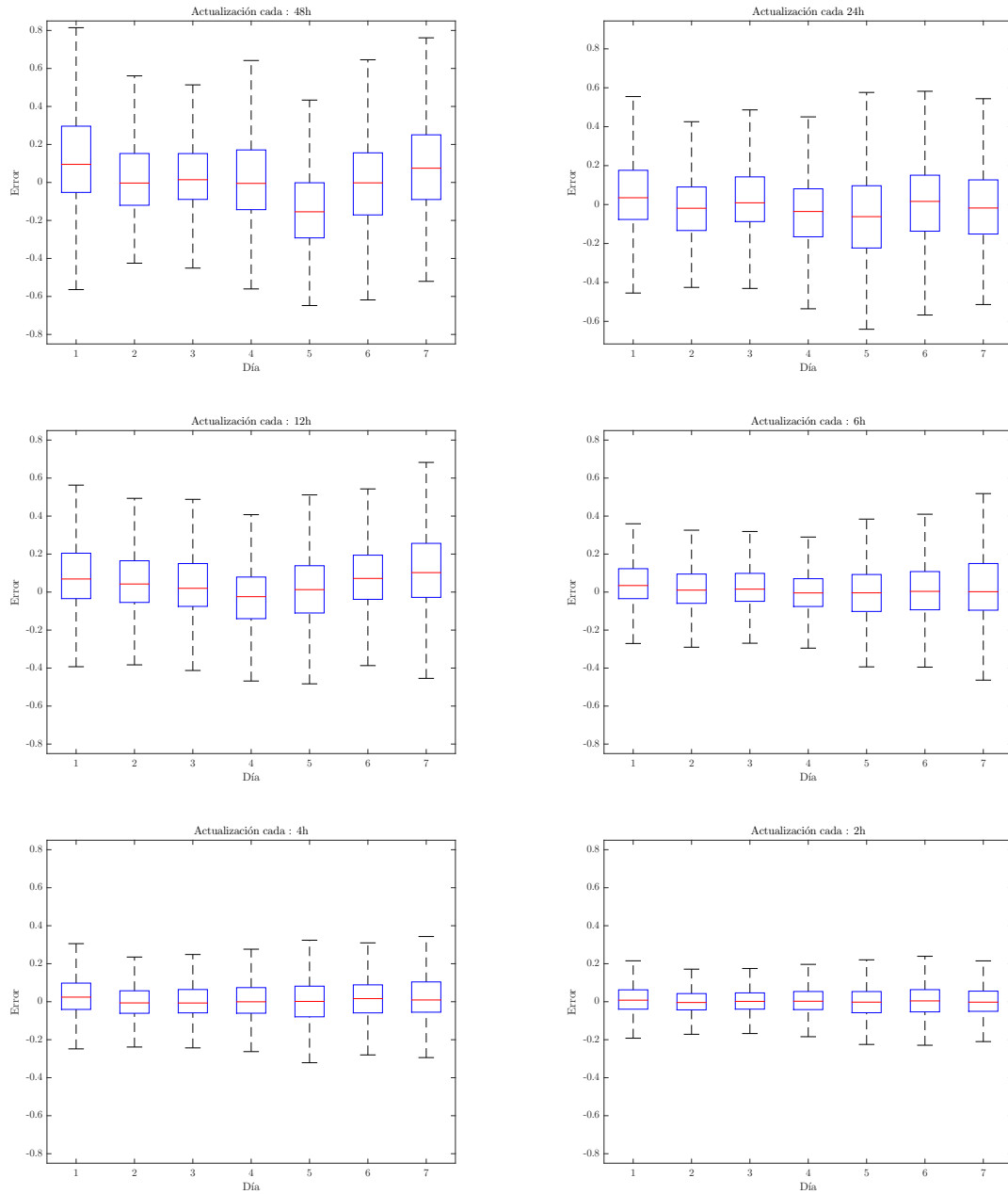


Figura 4-12: Distribución diaria de los errores.

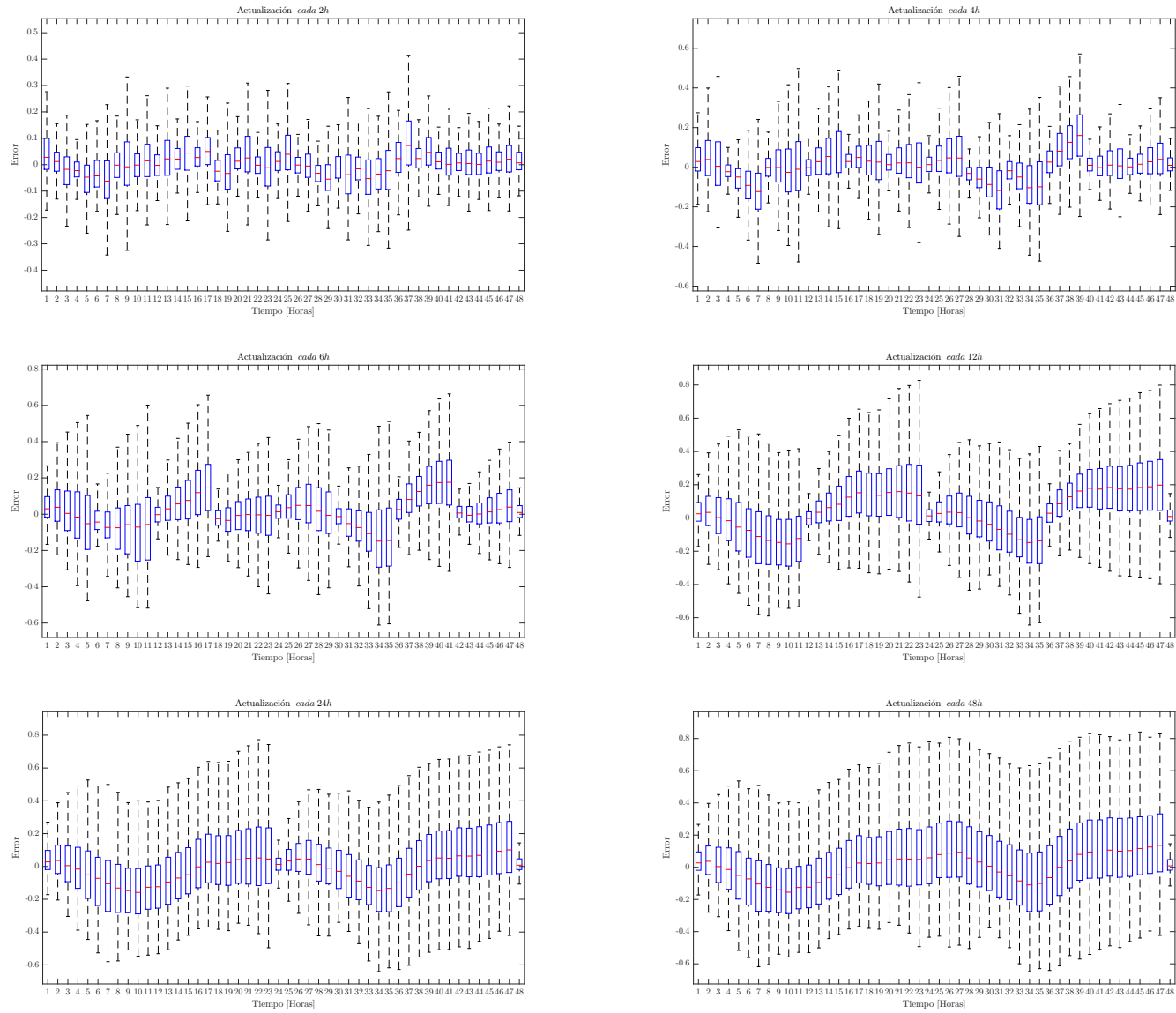


Figura 4-13: Distribución horaria de los errores.

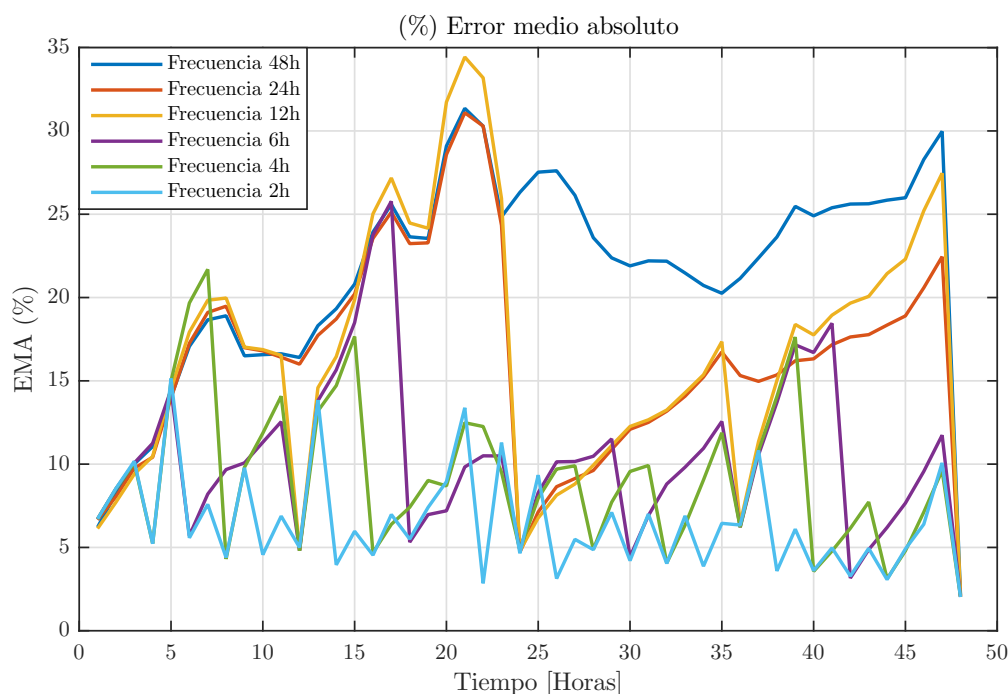


Figura 4-14: EMA para el modelo de predicción probabilística de la generación eólica propuesto.

4.4. Análisis

En la Figura 4-15 se presentan seis funciones de probabilidad acumulada para cada una de las frecuencias de actualización del modelo y la función de distribución acumulada del conjunto de datos de validación del modelo. Principalmente, el modelo presenta un mejor ajuste para frecuencias de actualización mayores. Lo cual nos confirma una vez más la mejora sustancial que tiene el uso de datos recientes en el modelo de predicción probabilística de corto plazo. Similarmente, la función de probabilidad acumulada ha sido obtenida para los errores medios absolutos (EMA), siendo la función de probabilidad acumulada con frecuencia de actualización 2h quién presenta un ΔP_t menor. Las anteriores observaciones son verificables de forma gráfica y tabular. Las Tablas 4-3 y 4-4 contienen algunos de las estadísticas descriptivas acerca del comportamiento del efecto de que tiene la actualización de frecuencia sobre el modelo. Otro aspecto a resaltar es el desempeño del modelo en términos de errores absolutos. En la Figura 4-17 se presenta el comportamiento de los errores para diferentes frecuencias de actualización, cuando se ha reducido el tamaño del conjunto de datos de entrenamiento del modelo al 62% (esto es 5928 valores de potencia eólica), aquí, es importante resaltar que gráficamente no se percibe cambio alguno entre el error presentado en la Figura 4-13 y el obtenido en la Figura 4-17, sin embargo, es aconsejable recurrir a

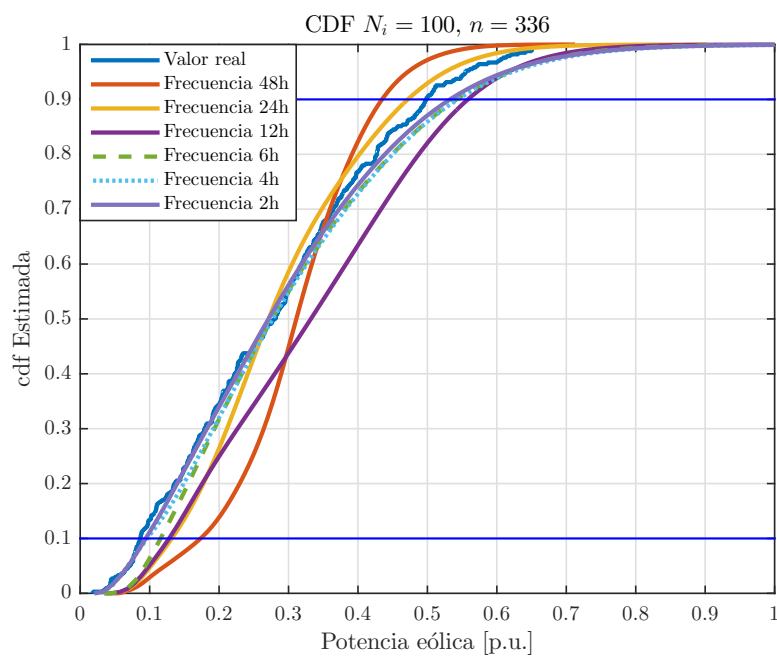


Figura 4-15: Funciones de probabilidad acumuladas

un análisis estadístico de los errores para observar cuál es la menor cantidad de datos con el cual el modelo será capaz de trabajar y entregar un error “aceptable”.

Tabla 4-3: Estadísticas descriptivas 4-15

Frec. Actualización	μ	σ	Mediana	Minímo	Máximo
48h	0.31	0.10	0.31	0.07	0.54
24h	0.29	0.12	0.27	0.06	0.54
12h	0.33	0.15	0.34	0.06	0.61
6h	0.3	0.15	0.28	0.06	0.62
4h	0.30	0.15	0.28	0.04	0.68
2h	0.30	0.15	0.28	0.04	0.68

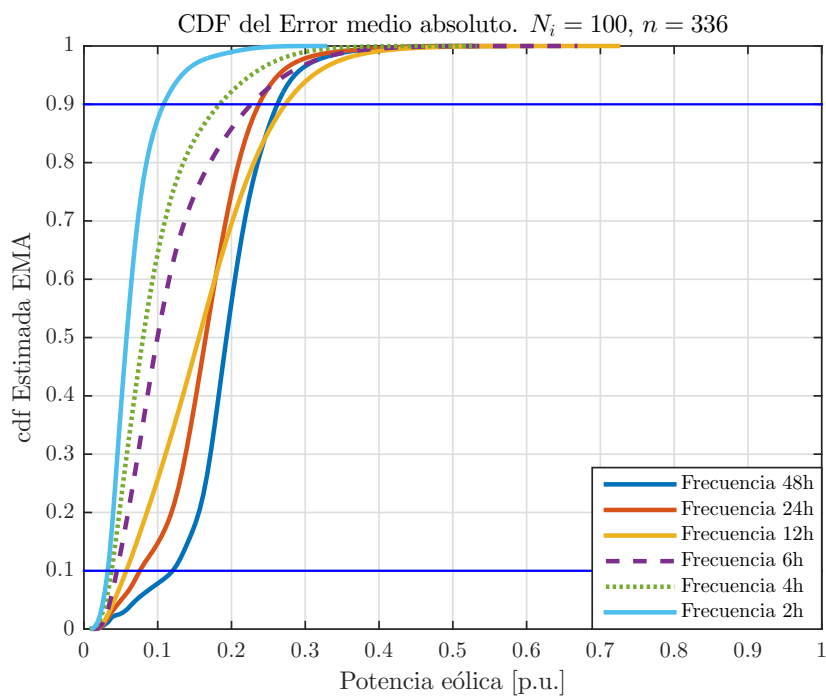


Figura 4-16: Funciones de probabilidad acumuladas

Tabla 4-4: Estadísticas descriptivas 4-16

Frec. Actualización	μ	σ	Mediana	Minímo	Máximo
48h	0.19	0.06	0.19	0.02	0.43
24h	0.16	0.06	0.17	0.02	0.43
12h	0.16	0.08	0.16	0.02	0.43
6h	0.12	0.07	0.10	0.02	0.39
4h	0.10	0.06	0.08	0.01	0.30
2h	0.07	0.03	0.06	0.01	0.22

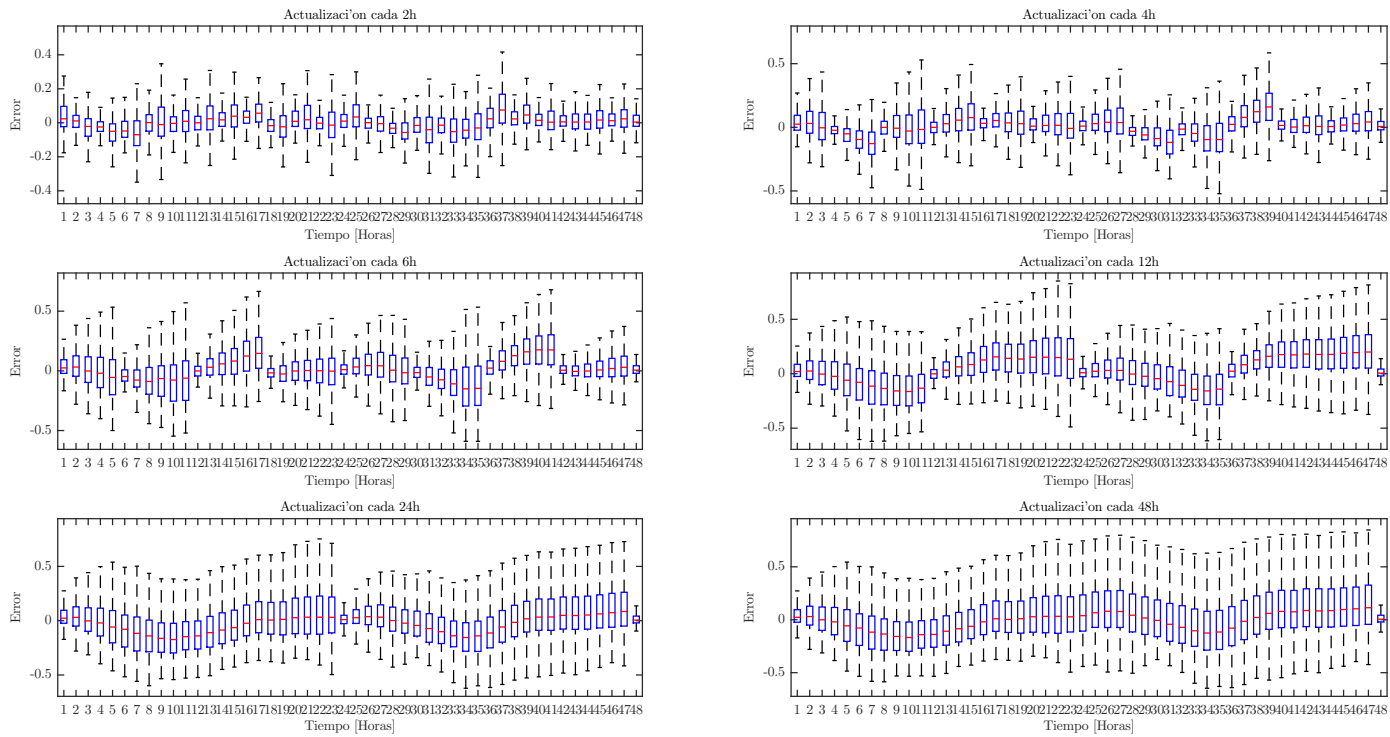


Figura 4-17: Distribución horaria de los errores cuando el conjunto de datos de entrenamiento es reducido al 62%.

4.5. Conclusión

En este capítulo un caso de estudio fue presentado, el modelo fue alimentado con 10200 mediciones horarias del agregado de potencia eólica. El conjunto de datos fue dividido en conjunto de datos de entrenamiento y conjunto de datos validación. Adicionalmente, el modelo de predicción de la generación eólica está en la capacidad de ser actualizado en horizontes de tiempo de entre 2h y 48h dependiendo de la precisión que se busque en la estimación.

Admeás, un análisis del desempeño del modelo de predicción de la generación eólica de corto plazo es estudiado, a través d ela variación de la frecuencia de actualización y el tamaño del conjunto de datos de entrenamiento. Índices de desempeño tales como: error medio(EM), error medio absoluto normalizado(EMAN), error porcentual medio absoluto (EPMA) y desviación estándar de los errores(DEE). Haciendo uso de la frecuencia de actualización, el modelo presenta un mejor desempeño cuando ésta es mayor (tiende al valor de la resolución del conjunto de datos de entrenamiento del modelo P_{nen}).

CONCLUSIONES

- A través del estudio de los métodos de predicción eólica de corto plazo, en una primera etapa se propuso trabajar con datos del agregado de potencia eólica en Colombia, más específicamente con datos del parque eólico Jepírachi EPM. Sin embargo, fue imposible que obtener respuesta alguna por parte de la parte del área técnica encargada, razón por la cual se usaron datos del histórico de potencia eólica de 10 granjas ubicadas en el continente australiano.
- El modelamiento de cualquier tipo de predicción basado en datos reales de generación esta estrechamente ligado con la calidad de la información histórica que se pose sobre el fenómeno bajo estudio. Gran parte del tiempo invertido en la implementación del modelo de predicción fue dedicado al ordenamiento y selección del conjunto de datos requeridos para el entrenamiento de este y su posterior validación. Adicionalmente, una correcta selección y representación de estos es de ayuda en el entendimiento del comportamiento de las masas de aire en las granjas de generación.
- La predicción de la generación eólica de corto plazo es un tema ampliamente investigado desde múltiples puntos de vista. Muchos de los modelos vigentes a la fecha están basados en estudios realizados a mediados de los 80's. Sin embargo, debido a limitaciones propias de la época, el procesamiento de grandes cantidades de información era limitado. Hoy en día los ordenadores están en la capacidad de realizar cálculos complejos usando la mayor cantidad de información disponible. Esto ha permitido el desarrollo de modelos más complejos y con mayor grado de detalle en la predicción de la generación de potencia eólica de corto plazo.
- Al inicio de esta tarea, fue abrumador encontrar tan diversos y variados enfoques para un mismo propósito. Al final, uno de los mayores aprendizajes consistió en que primero se debe responder a la pregunta ¿Cuál es el propósito al cual mi predicción desea atender?, será usado en el planeamiento del sistema o despacho, tal vez regulación gestión de costos. Sea cual sea la razón, un horizonte de tiempo fijo debe ser establecido. De esta forma, es más sencillo atender a cada una de las múltiples soluciones existentes y quedarse con la más conveniente según sea el caso.
- Sin importar el método seleccionado y el número de variables de entrada al modelo de predicción, siempre el supuesto será el mismo: Basados en el conjunto de observaciones

anteriores, y suponiendo que el comportamiento futuro de la generación eólica se semejante al histórico, se espera que ...

- Un modelo de predicción de la generación eólica fue implementado. El modelo consta en total de tres fases: *preprocesamiento*, *entrenamiento* y *predicción*. Gracias a un clasificador es posible preprocesar los datos, entonces, aplicando KDS es posible encontrar diversos escenarios de generación. Una vez el modelo fue implementado, diversas preguntas son resueltas ¿Cuál es el comportamiento de los datos de acuerdo con el horizonte de predicción? ¿Cuál es la correlación entre número de intervalos y la frecuencia de actualización?.
- El modelo de predicción probabilística de corto plazo genera escenarios compuestos de valores puntuales del agregado de potencia eólica, mediante un proceso en cascada, donde los dos últimos valores más actuales del agregado de potencia eólica son usados para generar un nuevo valor. La *frecuencia de actualización* permite obtener valores de la potencia eólica con un mayor grado de precisión, esto, para valores de frecuencia en la vecindad del paso de la muestra.
- El modelo de predicción probabilística implementado puede ser usado en la generación de escenarios en el despacho óptimo de granjas eólicas. Una de las mayores ventajas del modelo es su independencia de otras variables climatológicas, como pueden ser: temperatura, presión, humedad, entre otras.

TRABAJO FUTURO

- El modelo descrito en este documento genera un valor futuro de potencia eólica basado en los dos valores más recientes, así, como también las funciones de distribución de probabilidad construidas al final de la clasificación en magnitud y tendencia. Aquí, dos opciones elementales del modelo se deberían investigar: la mínima cantidad de datos necesarios para generar resultados "aceptables" (dependiendo del propósito de la predicción) y ¿cuál debe ser el número de valores que se deben tener en cuenta para realizar la clasificación en tendencia?
- El modelo está basado en un estimador del kernel de una única variable. En trabajos futuros, se deberá plantear el estudio de estimadores del kernel de múltiples variables.
- En este trabajo únicamente se usó el modelo para datos de potencia eólica. Sin embargo, el mismo puede ser usado para predecir valores futuros de generación solar y algunas convencionales (hidroeléctrica).
- No todos los valores del conjunto de datos de entrenamiento del modelo aportan información útil al modelo. Un enfoque práctico, consiste en ver cuáles datos se deben tener en cuenta a la hora de realizar una predicción de tipo probabilística.
- Se debe procurar por el desarrollo de un modelo multivariable, donde, los datos de entrada estén en función del histórico de la velocidad, dirección y potencia generada en las granjas, además, sería de utilidad contar con un modelo que este en la capacidad de entregar datos no solamente en el dominio del tiempo, sino que también en el espacio (mapa de colores de las granjas de generación).
- El modelo desarrollado trabaja únicamente con 10200 datos con un paso de 1h. Sería un bonito ejercicio aumentar el tamaño del conjunto de datos de entrenamiento del modelo a 200000, así, lograr capturar más información, y en conjunto con un método de optimización, lograr el mayor provecho de la información.

Participación eventos nacionales e internacionales

- Dentro del desarrollo de esta tesis de maestría se tuvo la oportunidad de participar en los siguientes eventos:
 1. Pasantia de Investigación en la University of Wisconsin-Milwaukee, (USA)
 2. Poster competition: University of Wisconsin-Milwaukee, (USA)
 3. Poster competition: IEEE Milwaukee Section Larry Hause Poster Competition

- Producción bibliográfica
 1. Martinez, F. A. P., Vivas, O. A. G., & Rosas, Y. S. S. (2017). Cuantificación del ahorro de energía eléctrica en clientes residenciales mediante acciones de gestión de demanda. Revista UIS Ingenierías, 16(2), 217-226.. **Ver adjunto.**

Apéndice A: 20 Intervalos ($N_i = 20$)

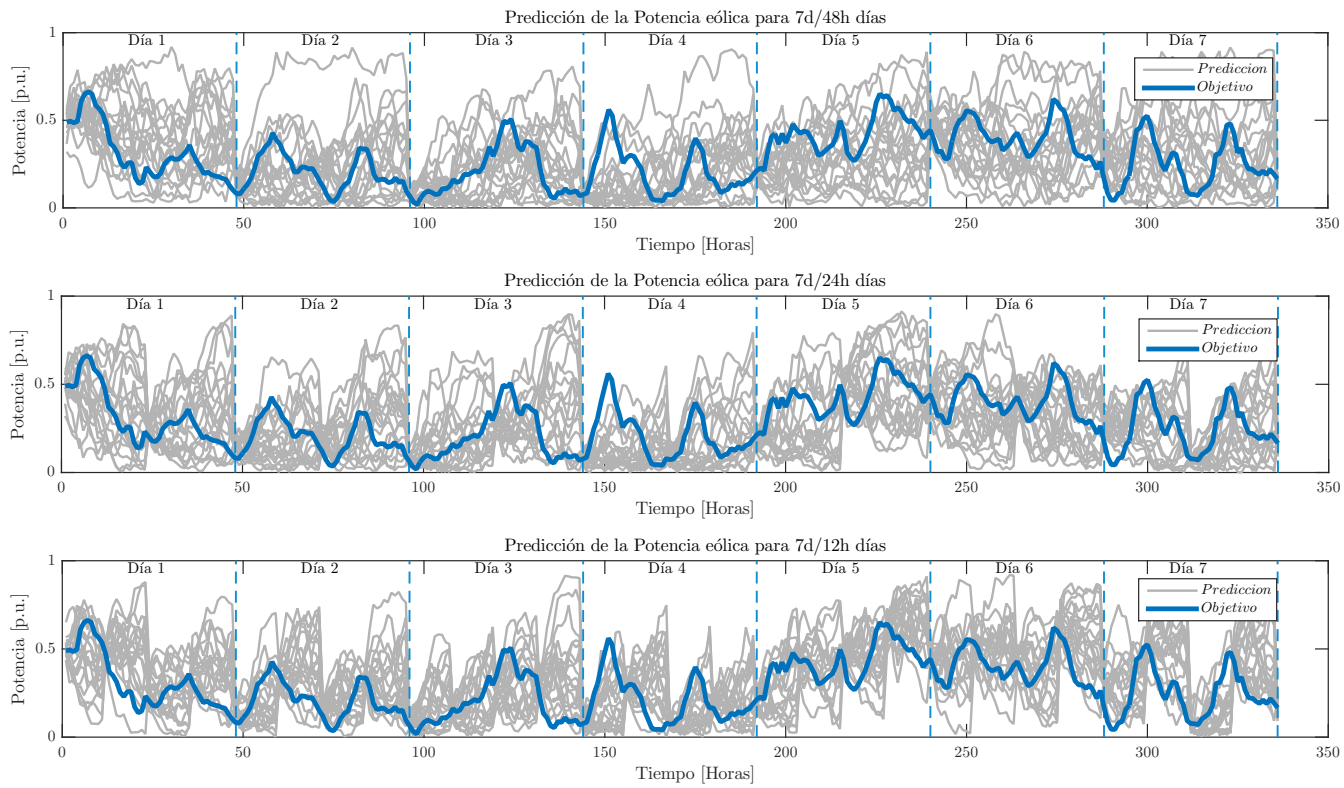


Figura A-1: Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 20$.

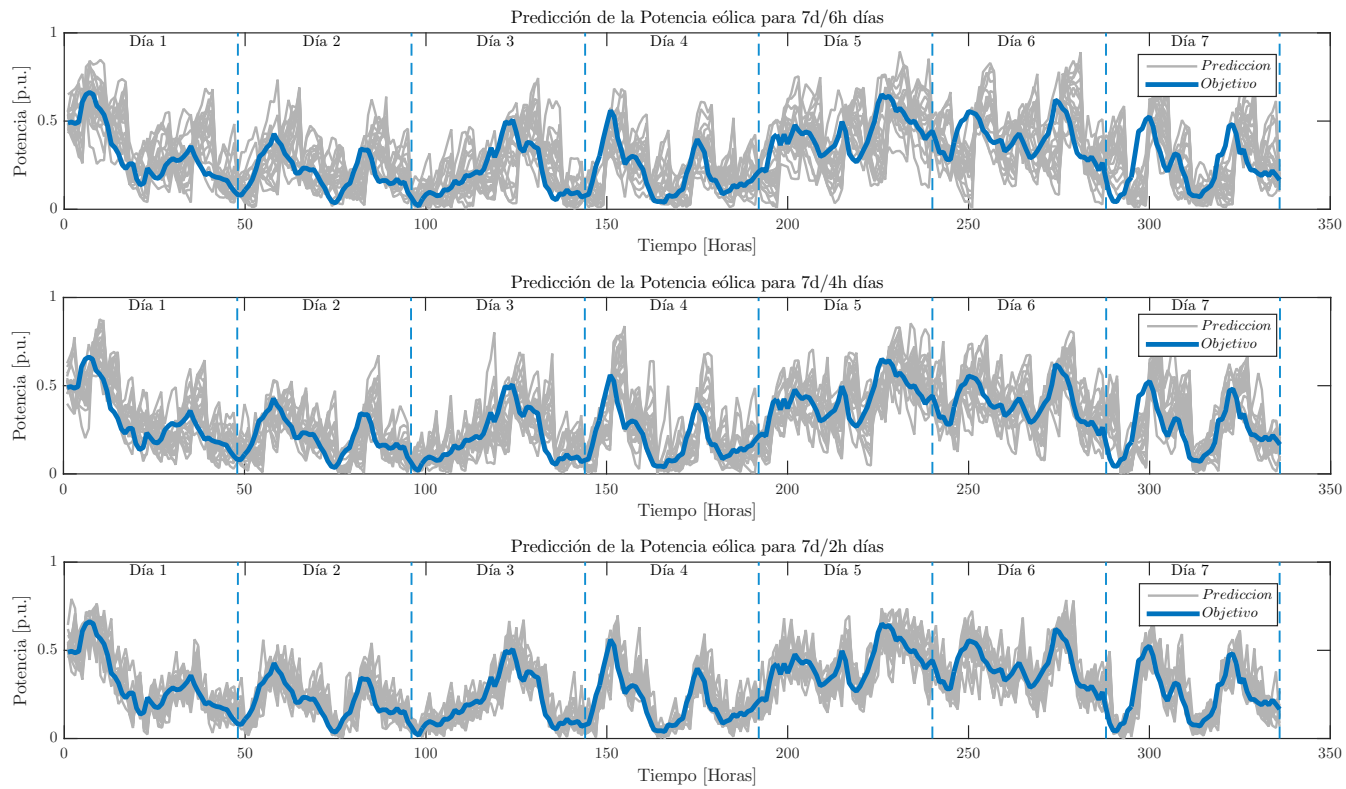


Figura A-2: Escenarios de generación para 7 días con frecuencia de actualización $\{48h, 24h, 12h, 6h, 4h, 2h\}$ y $N_i = 20$.

Apéndice B: 60 Intervalos ($N_i = 60$)

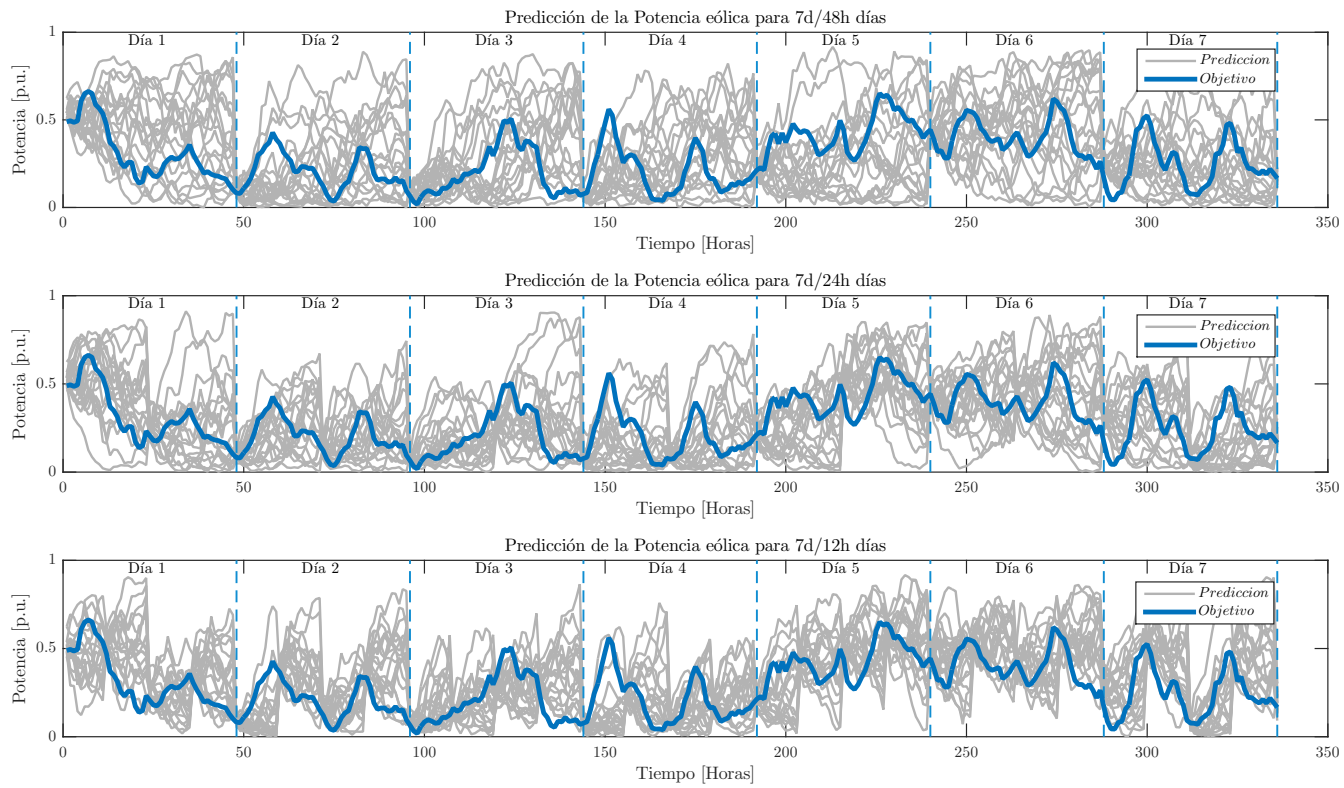


Figura B-1: Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h, 12h, 6h, 4h, 2h} y $N_i = 60$.

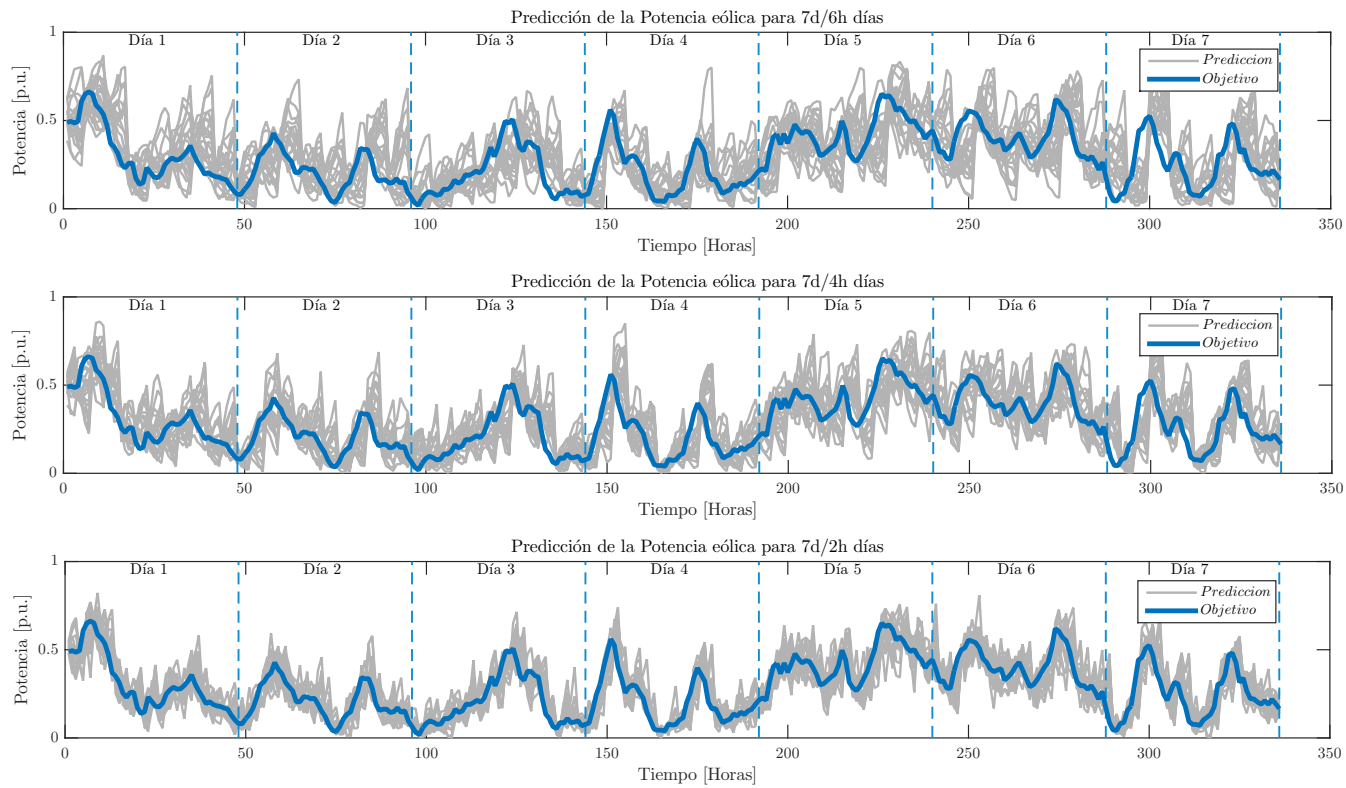


Figura B-2: Escenarios de generación para 7 días con frecuencia de actualización $\{48h, 24h, 12h, 6h, 4h, 2h\}$ y $N_i = 60$.

Apéndice C: 200 Intervalos

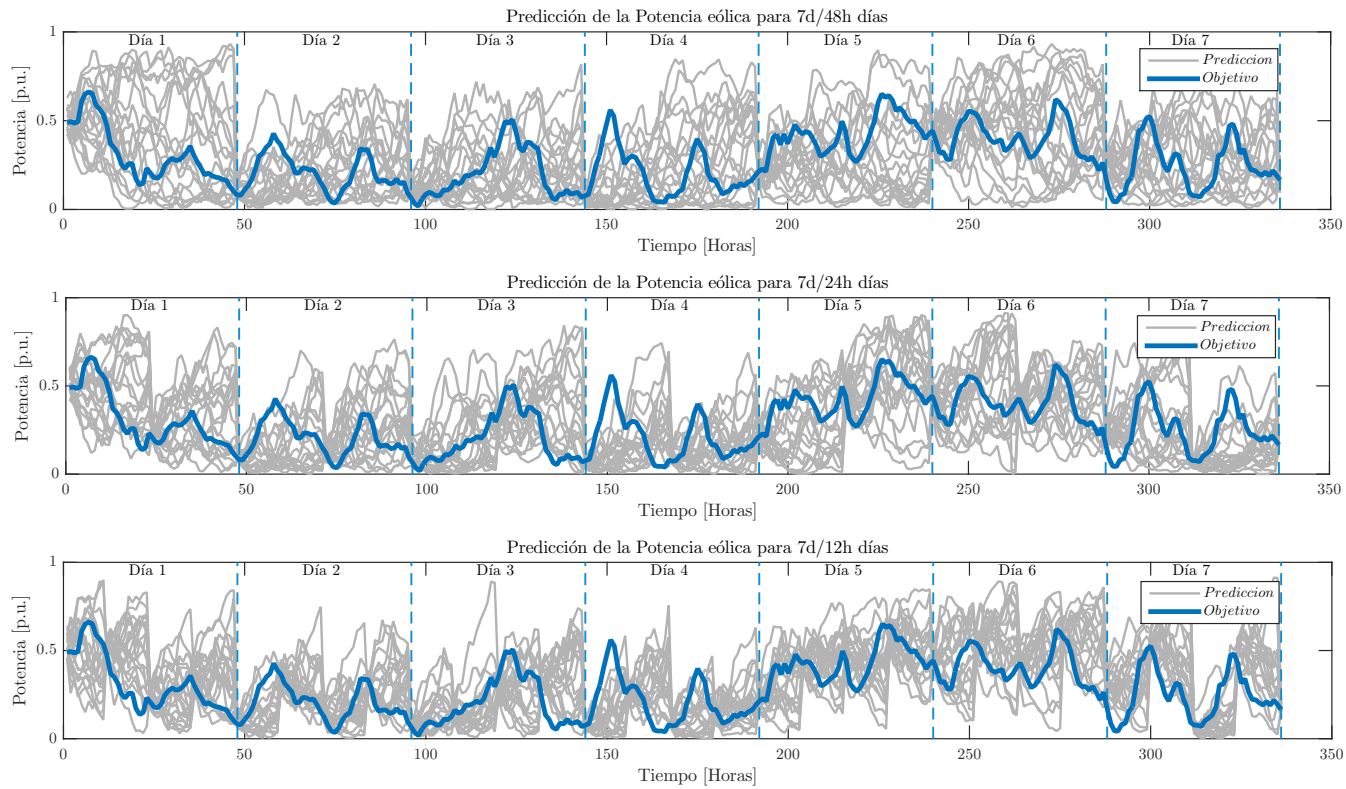


Figura C-1: Escenarios de generación para 7 días con frecuencia de actualización $\{48h, 24h, 12h, 6h, 4h, 2h\}$ y $N_i = 60..$

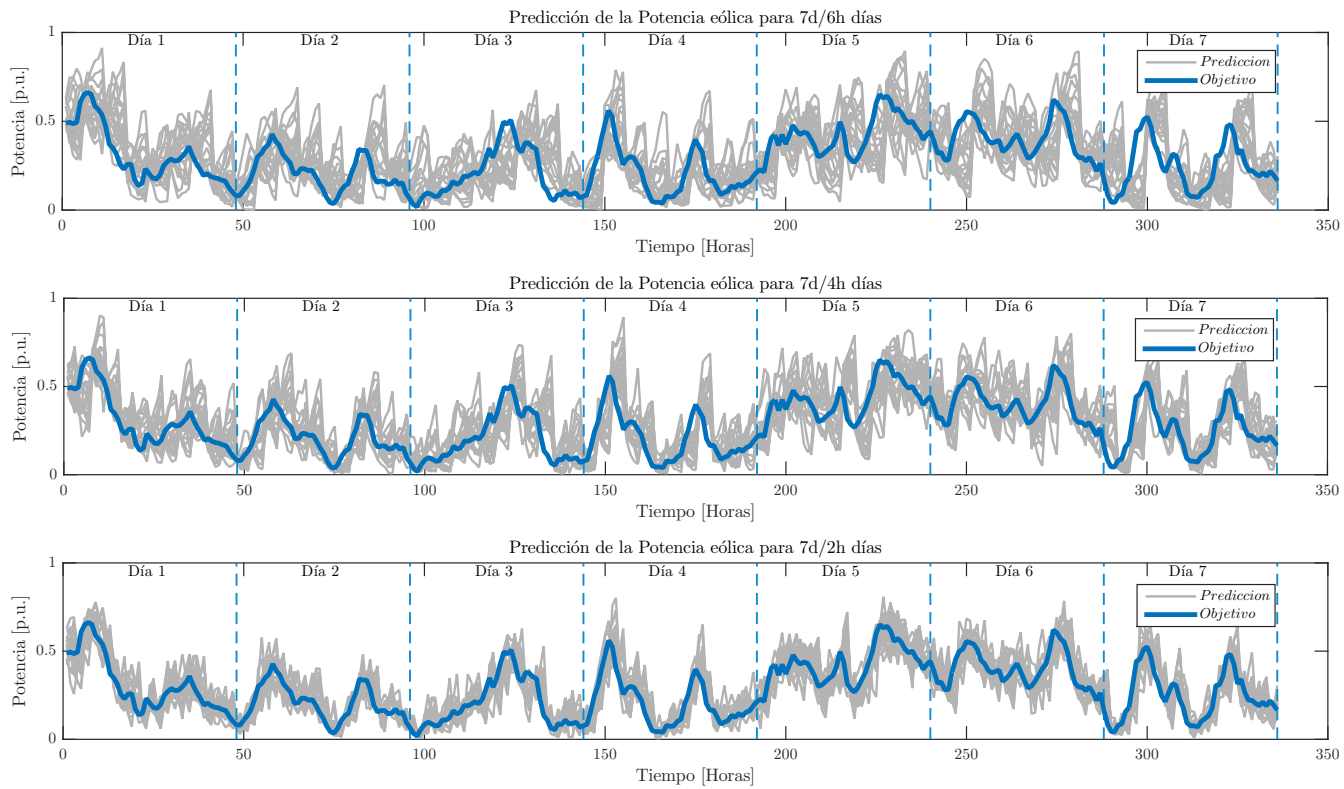


Figura C-2: Escenarios de generación para 7 días con frecuencia de actualización $\{48h, 24h, 12h, 6h, 4h, 2h\}$ y $N_i = 60$.

Apéndice D: Reducción conjunto de datos entrenamiento al 38 %

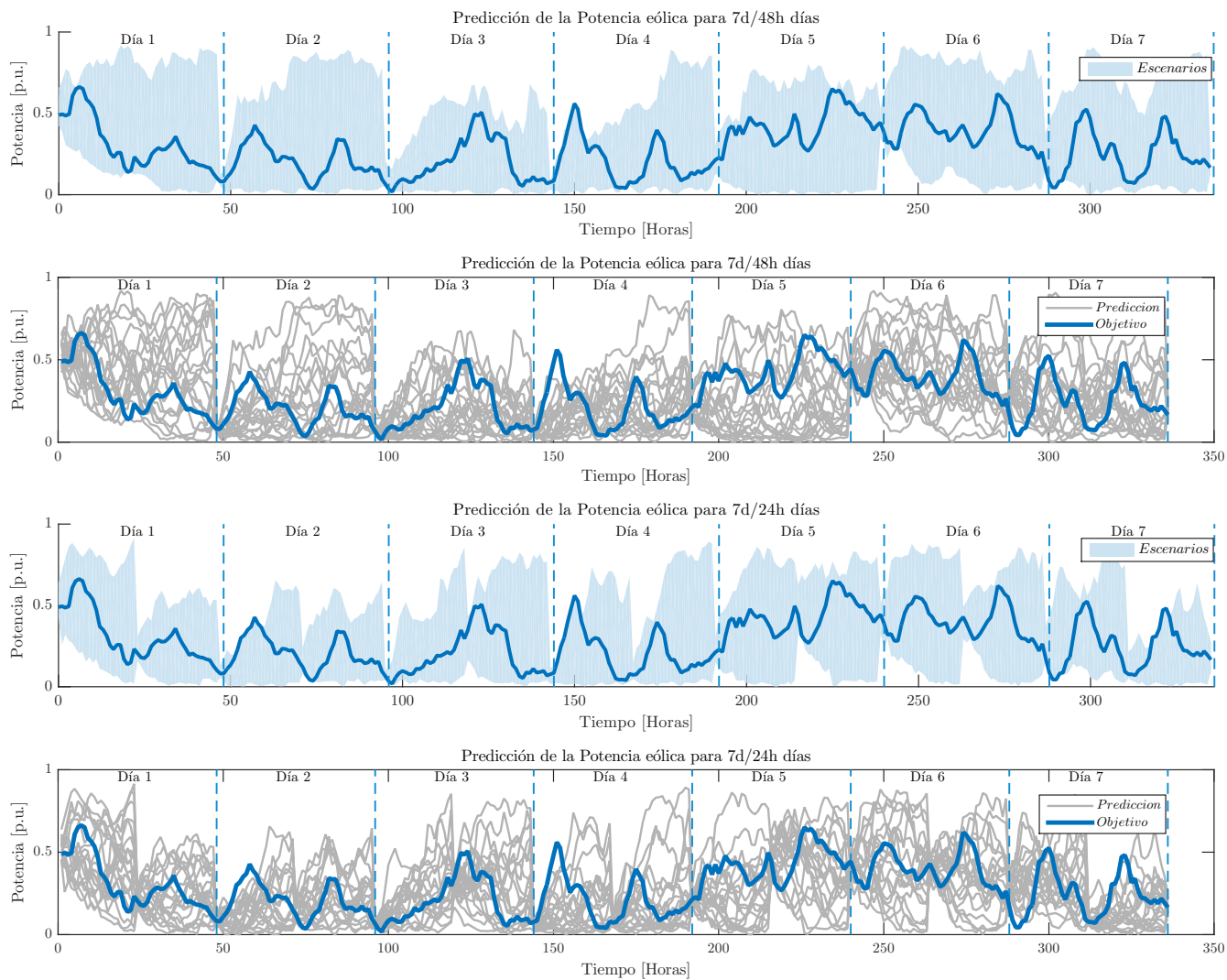


Figura D-1: Escenarios de generación para 7 días con frecuencia de actualización {48h, 24h} y $N_i = 100$, conjunto de datos reducido al 38 %.

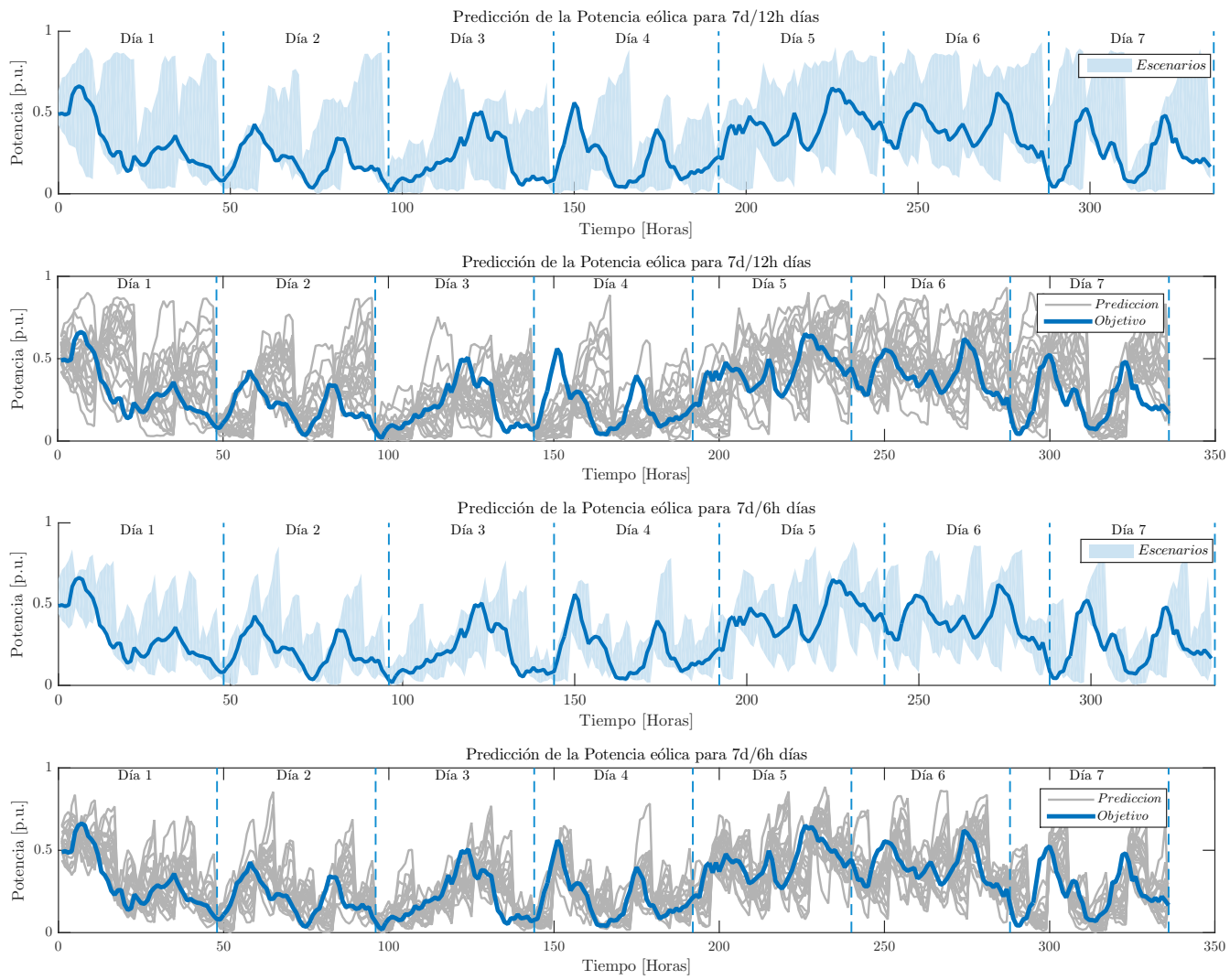


Figura D-2: Escenarios de generación para 7 días con frecuencia de actualización {12h, 6h} y $N_i = 100$, conjunto de datos reducido al 38 %.

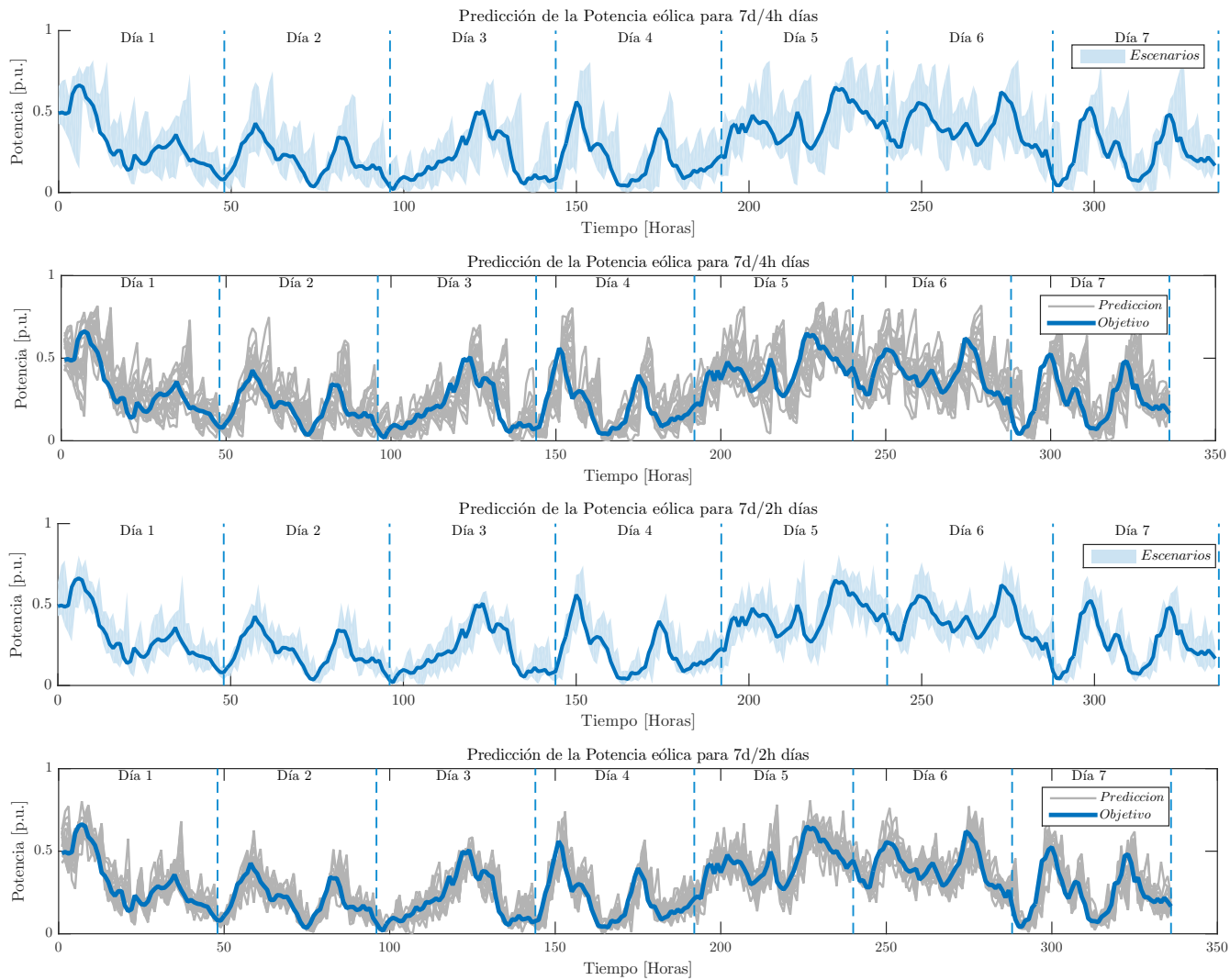


Figura D-3: Escenarios de generación para 7 días con frecuencia de actualización $\{4h, 2h\}$ y $N_i = 100$, conjunto de datos reducido al 38 %.

Bibliografía

- Ackermann, T. et al. (2005). *Wind power in power systems*, volume 140. Wiley Online Library.
- Akylas, E., Tombrou, M., Panourgias, J., and Lalas, D. (1997). The use of common meteorological predictions in estimating short term wind energy production in complex terrain. In *EWEC-CONFERENCE-*, pages 329–332. BOOKSHOP FOR SCIENTIFIC PUBLICATIONS.
- Association, E. W. E. (2011). *Pure power-wind energy targets for 2020 and 2030*. Ewea.
- Bailey, B., Brower, M. C., and Zack, J. (1999). Short-term wind forecasting: development and application of a mesoscale model. In *EWEC-CONFERENCE-*, pages 1062–1065.
- Bailey, B. and Stewart, R. (1987). Wind forecasting for wind power stations. *Proceedings of the ninth British wind*.
- Beyer, H., Degner, T., Hausmann, J., Hoffmann, M., and Rujan, P. (1994). Short term prediction of wind speed and power output of a wind turbine with neural networks. In *Proceedings of the EWEC*, volume 94, pages 349–352.
- Bossanyi, E. (1985). Short-term wind prediction using kalman filters. *Wind Engineering*, 9(1):1–8.
- Brown, B. G., Katz, R. W., and Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meteorology*, 23(8):1184–1195.
- Buizza, R. (1996). Chaos and weather prediction january 2000. *Analysis*, 12:1–7.
- Bustamante, E., Cofino, A., Navarro, J., Gutierrez, J., and Roldan, A. (2004). Wind speed downscaling combining mesoscale and neural autoregressive models. In *Proceedings of European wind energy conference, London*.
- Conradsen, K., Nielsen, L., and Prahm, L. (1984). Review of weibull statistics for estimation of wind speed distributions. *Journal of Climate and Applied Meteorology*, 23(8):1173–1183.
- Costa, A., Crespo, A., and Migoya, E. (2003). First results from a prediction project. In *Proceedings of the European Wind Energy Conference, EWEC*, volume 3.

- Costa, A., Crespo, A., Navarro, J., Palomares, A., and Madsen, H. (2006). Modelling the integration of mathematical and physical models for short-term wind power forecasting. In *Proceedings of European wind energy conference, Athens*, volume 18.
- Damousis, I. G. and Dokopoulos, P. (2001). A fuzzy expert system for the forecasting of wind speed and power generation in wind farms. In *Power Industry Computer Applications, 2001. PICA 2001. Innovative Computing for Power-Electric Energy Meets the Market. 22nd IEEE Power Engineering Society International Conference on*, pages 63–69. IEEE.
- DCM (2012). Danish ministry of climate, energy and building. energy policy report. Technical report.
- Fellows, A. and Hill, D. (1990). Wind and load forecasting for integration of wind power into a meso-scale electrical grid. In *Proceedings of the European community Wind Energy conference*, pages 10–14.
- Focken, U., Lange, M., Monnich, K., Waldl, H.-P., Beyer, H. G., and Luig, A. (2002). Short-term prediction of the aggregated power output of wind farm—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(3):231–246.
- Foley, A. M., Leahy, P. G., Marvuglia, A., and McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1):1–8.
- Gallardo, C., Gaertner, M., Tejada, C., Martínez, N., Calabria, S., Martínez, N., et al. (2003). The casandra project: first results of a new wind power operational forecasting system in spain. In *Proceedings of European wind energy conference, Madrid*.
- Geerts, H. (1984). Short range prediction of wind speeds: a system-theoretic approach. In *Proceedings of the European Wind Energy Conference, Hamburg (DE)*, pages 594–599.
- Giebel, G., Badger, J., Perez, I. M., Louka, P., Kallos, G., Palomares, A., Lac, C., and Descombes, G. (2006). Short-term forecasting using advanced physical modelling: The results of the anemos project. In *Proceedings of the European Wind Energy Conference*.
- Giebel, G., Landberg, L., Kariniotakis, G., and Brownsword, R. (2001). Prediction models in complex terrain. In *Proceedings of European wind energy conference*.
- Giebel, G., Landberg, L., Kariniotakis, G., and Brownsword, R. (2003). State-of-the-art methods and software tools for short-term prediction of wind energy production. In *EWEC 2003 (European Wind Energy Conference and exhibition)*, pages CD–ROM.
- Giebel, G., Landberg, L., Nielsen, T. S., and Madsen, H. (2002). The zephyr project—the next generation prediction system. In *Proceedings of the 2001 European Wind Energy Conference, EWEC*, volume 1, pages 777–780.

- Goh, S. L., Popovic, D., and Mandic, D. (2004). Complex-valued estimation of wind profile and wind power. In *Electrotechnical Conference, 2004. MELECON 2004. Proceedings of the 12th IEEE Mediterranean*, volume 3, pages 1037–1040. IEEE.
- Hou, D., Kalnay, E., and Droegemeier, K. K. (2001). Objective verification of the samex'98 ensemble forecasts. *Monthly Weather Review*, 129(1):73–91.
- Jensen, U., Pelgrum, E., and Madsen, H. (1994). Short term prediction of wind speed and power output of a wind turbine with neural networks. In *Proceedings of the EWEC*, pages 353–6.
- Jiménez, P. et al. (2004). Sensitivity of mm5 wind prediction to domain size and pbl schemes. In *Proceedings of*.
- Jung, J. and Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews*, 31:762–777.
- Jursa, R. and Rohrig, K. (2008). Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. *International Journal of Forecasting*, 24(4):694–709.
- Kaminsky, F., Kirchhoff, R., and Sheu, L. (1985). Time series models of average wind speed within synoptic weather categories. In *Proceedings of the fourth ASME wind energy symposium*, pages 215–9.
- Kariniotakis, G., Halliday, J., Brownsword, R., Marti, I., Palomares, A. M., Cruz, I., Madsen, H., Nielsen, T., Nielsen, H. A., Focken, U., et al. (2006). Next generation short-term forecasting of wind power—overview of the anemos project. In *European Wind Energy Conference, EWEC 2006*, pages 10–pages.
- Kariniotakis, G., Marti, I., Casas, D., Pinson, P., Nielsen, T. S., Madsen, H., Giebel, G., Usaola, J., Sanchez, I., Palomares, A., et al. (2004a). What performance can be expected by short-term wind power prediction models depending on site characteristics. In *Proc. of EWEC*, volume 4.
- Kariniotakis, G., Mayer, D., Moussafir, J., Chevallaz-Perrier, R., Usaola, J., Sanchez, I., Marti, I., Madsen, H., Nielsen, T. S., Lac, C., et al. (2003). Anemos: development of a next generation wind power forecasting system for the large-scale integration of onshore & offshore wind farms. In *Ewec 2003*, pages CD–ROM.
- Kariniotakis, G., Nogaret, E., and Stavrakakis, G. (1997). Advanced short-term forecasting of wind power production. In *European wind energy conference. EWEC'97*, pages 751–754. Irish Wind Energy Association.

- Kariniotakis, G., Pinson, P., Siebert, N., Giebel, G., and Barthelmie, R. (2004b). The state of the art in short-term prediction of wind power - from an offshore perspective.
- Kariniotakis, G., Stavrakakis, G., and Nogaret, E. (1996). A fuzzy logic and neural network based wind power model. In *Proceedings of the European Wind Energy Conference*, pages 596–599.
- Landberg, L. (1994). Implementing wind forecasting at a utility. In *Proceedings of European wind energy conference0*, pages 357–60.
- Landberg, L., Giebel, G., Nielsen, H. A., Nielsen, T., and Madsen, H. (2003). Short-term prediction-an overview. *Wind Energy*, 6(3):273–280.
- Lang, S., Möhrle, C., Jørgensen, J., Gallachóir, B., and McKeogh, E. (2006). Application of a multi-scheme ensemble prediction system for wind power forecasting in ireland and comparison with validation results from denmark and germany. In *Proceedings of European wind energy conference, Athens*.
- Lange, M. (2005). On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors. *Journal of solar energy engineering*, 127(2):177–184.
- Lange, M. and Focken, U. (2006). *Physical approach to short-term wind power prediction*. Springer Science & Business Media.
- Lange, M. and Waldl, H.-P. (2001). Assessing the uncertainty of wind power predictions with regard to specific weather situations. In *Proceedings of the European Wind Energy Conference, Copenhagen, Denmark*, pages 695–698.
- Lin, L., Eriksson, J., Vihriala, H., and Soderlund, L. (1996). Predicting wind behavior with neural networks. In *Proceedings the European Wind Energy Conference*, pages 655–658.
- Liu, Y., Wang, Y., Li, L., Han, S., and Infield, D. (2016). Numerical weather prediction wind correction methods and its impact on computational fluid dynamics based wind power forecasting. *Journal of Renewable and Sustainable Energy*, 8(3):033302.
- Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., and Kallos, G. (2008). Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(12):2348–2362.
- Madsen, H., Kariniotakis, G., Nielsen, H. A., Nielsen, T. S., and Pinson, P. (2004). A protocol for standardizing the performance evaluation of short-term wind power prediction models. *Technical University of Denmark, IMM, Lyngby, Denmark, Deliverable ENK5-CT-2002-00665*.

- Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A., and Nielsen, T. S. (2005). Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, 29(6):475–489.
- Mandic, D., Javidi, S., Goh, S., Kuh, A., and Aihara, K. (2009). Complex-valued prediction of wind profile using augmented complex statistics. *Renewable Energy*, 34(1):196–201.
- Marti, I., Kariniotakis, G., Pinson, P., Sanchez, I., Nielsen, T., Madsen, H., Giebel, G., Usaola, J., Palomares, A. M., Brownsword, R., et al. (2006). Evaluation of advanced wind power forecasting models—results of the anemos project. In *European Wind Energy Conference, EWEC 2006*, pages 9–pages.
- MATLAB (2018). *Mastering Machine Learning A Step-by-Step Guide with MATLAB*. MATLAB.
- Miranda, M. S. and Dunn, R. W. (2006). One-hour-ahead wind speed prediction using a bayesian methodology. In *Power Engineering Society General Meeting, 2006. IEEE*, pages 6–pp. IEEE.
- Möller, A., Lenkoski, A., and Thorarinsdottir, T. L. (2013). Multivariate probabilistic forecasting using ensemble bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139(673):982–991.
- Møller, J. K., Zugno, M., and Madsen, H. (2016). Probabilistic forecasts of wind power generation by stochastic differential equation models. *Journal of Forecasting*.
- Nielsen, H. A., Madsen, H., and Nielsen, T. S. (2006). Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, 9(1-2):95–108.
- Nielsen, H. A., Nielsen, T. S., Madsen, H., Pindado, M. J., and Marti, I. (2007). Optimal combination of wind power forecasts. *Wind Energy*, 10(5):471–482.
- Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L., and Giebel, G. (1998). A new reference for wind power forecasting. *Wind energy*, 1(1):29–34.
- Okumus, I. and Dinler, A. (2016). Current status of wind energy forecasting and a hybrid method for hourly predictions. *Energy Conversion and Management*, 123:362–371.
- Pinson, P. (2006). *Estimation of the uncertainty in wind power forecasting*. PhD thesis, École Nationale Supérieure des Mines de Paris.
- Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):555–576.

- Pinson, P. and Kariniotakis, G. (2004). On-line assessment of prediction risk for wind power production forecasts. *Wind Energy*, 7(2):119–132.
- Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., and Klöckl, B. (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind energy*, 12(1):51–62.
- Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., Monteiro, C., Sousa, J., and Bessa, R. (2009). Comparison of two new short-term wind-power forecasting systems. *Renewable Energy*, 34(7):1848–1854.
- Robert, C. (2014). Machine learning, a probabilistic perspective.
- Sfetsos, A. (2000). A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewable energy*, 21(1):23–35.
- Sideratos, G. and Hatziaargyriou, N. D. (2012). Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Transactions on Power Systems*, 27(4):1788–1796.
- Tambke, J., Von Bremen, L., Barthelmie, R., Palomares, A. M., Ranchin, T., Juban, J., Kariniotakis, G., Brownsword, R., and Waldl, I. (2006). Short-term forecasting of offshore wind farm production—developments of the anemos project. In *2006 European Wind Energy Conference, EWEC*, pages 13–pages.
- Tande, J. and Landberg, L. (1993). A 10 sec. forecast of wind turbine output with neural networks. In *1993 European Community Wind Energy Conference. Proceedings*. HS Stephens and Associates.
- Tol, R. S. (1997). Autoregressive conditional heteroscedasticity in daily wind speed measurements. *Theoretical and applied climatology*, 56(1):113–122.
- Tong, H. (2011). Threshold models in time series analysis—30 years on. *Statistics and its Interface*, 4(2):107–118.
- Torres, J. L., Garcia, A., De Blas, M., and De Francisco, A. (2005). Forecast of hourly average wind speed with arma models in navarre (spain). *Solar Energy*, 79(1):65–77.
- Troen, I. and Landberg, L. (1990). Short term prediction of local wind conditions.
- Troen, I. and Petersen, E. L. (1989). *European wind atlas*. Risø National Laboratory.
- Vigueras-Rodríguez, A., Sørensen, P., Viedma, A., Donovan, M., and Lázaro, E. G. (2012). Spectral coherence model for power fluctuations in a wind farm. *Journal of Wind Engineering and Industrial Aerodynamics*, 102:14–21.

- Waldl, I., Dierich, F., Kariniotakis, G., Bocquet, A., and Virlot, S. (2006). The anemos wind power forecasting platform technology-techniques and experiences. In *EWEC 2006*, pages 4–pages.
- Watson, R., Landberg, L., Costello, R., McCoy, D., and O’Donnell, P. (2001). Implementation and evaluation of the predikator wind power forecasting tool in ireland.
- Watson, S., Halliday, J., and Landberg, L. (1992). Assessing the economic benefits of numerical weather prediction model wind forecasts to electricity generating utilities. *Wind Energy Conversion 1992*.
- Wegley, H. and Formica, W. (1983). Test application of a semi-objective approach to wind forecasting for wind energy applications. Technical report, Pacific Northwest Lab., Richland, WA (USA).
- Wegley, H., Kosorok, M., and Formica, W. (1984). Subhourly wind forecasting techniques for wind turbine operations. Technical report, Pacific Northwest Lab., Richland, WA (USA).
- Welch, R. L., Ruffing, S. M., and Venayagamoorthy, G. K. (2009). Comparison of feedforward and feedback neural network architectures for short term wind speed prediction. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 3335–3340. IEEE.
- Wendell, L. L., Wegley, H. L., and Verholec, M. G. (1978). Report from a working group meeting on wind forecasts for wecs operation. Technical report, Battelle Pacific Northwest Labs., Richland, WA (USA).
- Xydas, E., Qadrdan, M., Marmaras, C., Cipcigan, L., Jenkins, N., and Ameli, H. (2017). Probabilistic wind power forecasting and its application in the scheduling of gas-fired generators. *Applied energy*, 192:382–394.
- Zambom, A. Z. and Dias, R. (2012). A review of kernel density estimation with applications to econometrics. *arXiv preprint arXiv:1212.2812*.
- Zhao, X., Wang, S., and Li, T. (2011). Review of evaluation criteria and main methods of wind power forecasting. *Energy Procedia*, 12:761–769.