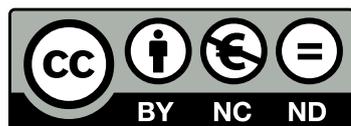


ÁLVARO JARAMILLO DUQUE

GENERACIÓN DE ESCENARIOS DE PRODUCCIÓN
EÓLICA PARA EL ANÁLISIS PROBABILÍSTICO DE
LA OPERACIÓN DE SISTEMAS DE ENERGÍA
ELÉCTRICA.

Esta investigación se ha realizado en el marco del proyecto «Integración de la Energía Eólica en el Mercado de Electricidad», referencia ENE2006-05192 y financiada por el proyecto «Integración de las Energías Renovables en el Mercado de Electricidad», referencia ENE2010-16074 del programa Formación de Personal Investigador (FPI) del Ministerio de Ciencia e Innovación de España.

Esta obra está licenciada bajo Creative Commons.
Reconocimiento - No comercial - Sin obras derivadas
España, License 3.0 2012.



Se permite copiar, distribuir y comunicar públicamente esta obra siempre que:

- Se cite al autor.
- Que el uso que se haga de la obra no sea comercial.
- Que no se pueda crear una obra derivada a partir de la original.

Álvaro Jaramillo Duque: *Generación de escenarios de producción eólica para el análisis probabilístico de la operación de sistemas de energía eléctrica.*, Tesis para la obtención del grado de doctor, © Junio del 2012



UNIVERSIDAD CARLOS III DE MADRID

Tesis para la obtención del grado de doctor.

GENERACIÓN DE ESCENARIOS DE PRODUCCIÓN EÓLICA
PARA EL ANÁLISIS PROBABILÍSTICO DE LA OPERACIÓN DE
SISTEMAS DE ENERGÍA ELÉCTRICA.

AUTOR:

ÁLVARO JARAMILLO DUQUE

DIRECTORES:

EDGARDO DANIEL CASTRONUOVO

JULIO USAOLA GARCÍA

Escuela Politécnica Superior
Departamento de Ingeniería Eléctrica

Leganés, Junio del 2012.

GENERACIÓN DE ESCENARIOS DE PRODUCCIÓN
EÓLICA PARA EL ANÁLISIS PROBABILÍSTICO DE
LA OPERACIÓN DE SISTEMAS DE ENERGÍA
ELÉCTRICA.

Autor: Álvaro Jaramillo Duque
Directores: Edgardo Daniel Castronuovo
Julio Usaola García

Firma del Tribunal Calificador:

Presidente: José Román Wilhelmi Ayza
Universidad Politécnica de Madrid.
Secretaria: María Ángeles Moreno López de Saá
Universidad Carlos III de Madrid.
Vocal: Ismael Sánchez Rodríguez Morcillo
Universidad Carlos III de Madrid.
Vocal: Alberto Berizzi
Politecnico di Milano.
Vocal: José Carpio Ibáñez
Universidad Nacional de Educación a Distancia.
Suplente: Cristian Bovo
Politecnico di Milano.
Suplente: Juan Ignacio Pérez Díaz
Universidad Politécnica de Madrid.

Calificación:

Leganés, Junio del 2012.

Expertos europeos que han valorado positivamente esta tesis:

Experto: Nils Siebert
MINES ParisTech - ARMINES.
Experto: Manuel Matos
Universidade do Porto.

Ohana means family.
Family means nobody gets left behind, or forgotten.
— Lilo & Stitch

a mi familia.

RESUMEN

El objetivo general de la tesis consiste en generar trayectorias de la producción eólica futura para el análisis probabilístico de los sistemas de energía eléctrica. Esta información puede ser utilizada para el comercio en los mercados de energía, para la determinación y la resolución de restricciones en la operación del sistema, para la estimación de las reservas de energía, entre otros usos. De esta forma, se facilita la integración de las fuentes no gestionables, y en particular, de la energía eólica.

Para conseguir este objetivo se desarrolló un modelo estocástico de la producción eólica. Este modelo genera escenarios de producción condicionados por la predicción a corto plazo. El modelo caracteriza las trayectorias simuladas basándose en la información obtenida del comportamiento pasado del parque eólico: producciones reales y sus respectivas previsiones a diferentes horizontes de predicción. Este tipo de información es la que usualmente poseen los productores eólicos y el operador del sistema, de forma que son los únicos datos disponibles para obtener el modelo.

El modelo se desarrolla partiendo de la teoría de las funciones cópula, como una herramienta para el modelado multivariante de variables aleatorias. Para el modelo multivariante es necesario obtener los modelos para las funciones de densidad univariantes de los errores de la predicción. Para obtener estos modelos se parte de los datos de un parque eólico, predicciones y producciones donde se identifican las características de las variables. Para las funciones marginales univariantes, primero se considera un modelo paramétrico basado en la distribución beta; como segunda alternativa, se considera una técnica de estimación de la función de densidad mediante funciones kernel, siendo este un método no paramétrico.

Los dos modelos, paramétrico y no paramétrico, no consiguen modelar de forma adecuada los errores de la predicción eólica cuando la producción real del parque es cero. Para resolver este problema se utiliza la regresión logística, un método que permite estimar la probabilidad de una variable con respuesta cualitativa, en este caso, de las producciones nulas. Una vez que se estima el modelo que obtiene la probabilidad de que ocurra una producción nula, dada la predicción, se adaptan los modelos antes descritos para obtener las funciones de densidad que modelan correctamente los errores en la predicción de producción eólica.

En ambos modelos, paramétrico y no paramétrico, se implementaron estimaciones adaptativas de los parámetros. A medida que pasa el tiempo, el modelo obtiene nueva información del comportamiento

del parque, en este caso, de las producciones reales y sus respectivas predicciones. Los nuevos datos son incluidos y ponderados de acuerdo a su relevancia y antigüedad, dando una característica adaptativa al modelo, de forma que logran replicar los comportamientos estacionales de la producción eólica.

Los modelos estimados para las funciones marginales univariantes son evaluados y comparados durante el procedimiento. Se determina que el modelo no paramétrico, mediante funciones kernel y la regresión logística, tiene un mejor desempeño que la técnica de modelado paramétrica, que utiliza la distribución beta y la regresión logística.

Finalmente, se estima la estructura de dependencia de las variables aleatorias desde el dominio uniforme. Se convierten las variables aleatorias al dominio uniforme, mediante la función de distribución de probabilidad obtenida con las funciones kernel y la regresión logística. De esta forma, se elimina la influencia del dominio original de las variables aleatorias. Con las funciones de densidad univariantes y la estructura de dependencia, se construye un modelo multivariante para la producción eólica, usando las funciones cópula. Con el modelo propuesto se pueden generar trayectorias posibles de la producción y obtener los intervalos de confianza. Dos modelos han sido desarrollados, el primero representa la dependencia temporal (corto plazo) en la producción de un parque y el segundo representa la dependencia espacial-temporal de varios parques en una misma zona.

Con un número considerable de escenarios simulados es posible estimar la incertidumbre de la producción, no solo en términos de potencia, sino también en cantidad de energía desviada. Esta información es utilizada para desarrollar medidas de compensación de los desvíos en la producción eólica, aplicadas a problemas de complejidad real. En el primer problema se propone la utilización de un método de optimización, buscando maximizar los ingresos de la operación conjunta de una central hidroeléctrica reversible y un parque eólico. En el segundo problema se representan diversas medidas de compensación como problemas de optimización, donde se busca cubrir los desvíos de la producción de un grupo de parques eólicos. En los problemas de optimización, el modelo del sistema eléctrico es representado explícitamente.

A partir de las trayectorias simuladas de la producción eólica futura se determinan los niveles de incertidumbre, en términos tanto de potencia como de energía futura. La incertidumbre de la producción es incluida en un problema de optimización que usa una central hidroeléctrica reversible con el objetivo de reducir el costo de los desvíos. En todos los resultados obtenidos para este problema, se consiguió un beneficio al comparar la operación independiente respecto a la operación conjunta. El método propuesto puede ser extendido a otros aspectos de la operación de sistemas de energía eléctrica.

La central hidroeléctrica reversible se usa para minimizar los desvíos debidos a los errores en la predicción de la potencia eólica, cuando el productor eólico participa en el mercado. De esta forma, los agentes pueden reducir el riesgo al que están expuestos debido a la incertidumbre asociada a la producción eólica y a los precios de la energía de reserva.

Finalmente, se comparan diferentes alternativas para compensar los desvíos de la producción eólica en un sistema de energía eléctrica. Las medidas propuestas se prueba en un sistema basado en el caso «IEEE 30 Bus Test Case», incluido en MATPOWER, con el objetivo de simular la complejidad de un pequeño sistema real. Las diferentes alternativas se validan mediante un Flujo de Potencia Probabilístico (FPP), empleando el método de Monte Carlo, donde se considera la incertidumbre de la producción eólica de un grupo de parques eólicos. Mediante este método se puede estimar el valor de cualquiera de las variables del sistema de energía eléctrica por medio de su función de densidad de probabilidad, por ejemplo, el nivel de carga de las líneas de transmisión, las pérdidas del sistema, el uso de los elementos de almacenamiento, etc. La influencia de la producción eólica y los efectos del almacenamiento en el sistema de energía eléctrica pueden apreciarse comparando las variables de salida (generación, intercambio entre áreas, pérdidas del sistema y uso del almacenamiento) del Flujo de Potencia Probabilístico (FPP).

ABSTRACT

The overall objective of the thesis is the generation of trajectories of future wind power production (for a short term horizon, one or two days in advance) for the probabilistic analysis of power systems. This information can be used for trading into markets, for the identification and resolution of constraints on the operation of the system, to estimate energy reserves, among others. This results will facilitate the integration of intermittent energy into the system, in particular the wind energy.

To achieve this objective, a stochastic model of the wind production. This model generates possible production trajectories, conditioned by the short-term wind power prediction has been developed. The model characterizes the simulated trajectories, based on the historical data of the wind farm, namely power production and their forecasts for different horizons. This information is usually held by wind farms owners and system operators.

The model is based on the copula function theory, which is a tool for modeling multivariate random variables. For the multivariate model the univariate density functions of the forecasted error of the wind power is required. These functions have been estimated in this work from the data of a wind farm, predictions and productions, from also the characteristics of the variables were identified. For univariate marginal functions, a parametric model based on the beta distribution was considered in a first trial. As second alternative, the estimation technique called Kernel Density Estimation (KDE), a non-parametric method, was considered.

The two models, parametric and nonparametric, failed to correctly represent the prediction errors of the wind power when the real production is zero. To solve this problem, logistic regression was used, to estimate the probability of a qualitative response variable (in this case the null productions). Once the model for the probability of null production was estimated, for a given value of the prediction, the models described above were adapted for modeling correctly the forecast error of the wind power production.

In both models, parametric and nonparametric, adaptive estimates of the parameters were implemented. This means that a new information obtained from data of the wind farm, in this case of new productions and their predictions, are included and weighted according to their relevance and age, giving to the model an adaptive feature, so the model is able to represent the seasonal behavior of wind production.

The proposed models for the univariate marginal functions were evaluated and compared with real data. It was determined that the non-parametric model using kernel functions and logistic regression performs better than the parametric modeling technique, using the beta distribution and logistic regression.

Finally, the dependence structure of the random variables from the uniform domain was estimated. The forecasted errors were transformed to uniform domain by using the Probability Density Function (PDF), obtained with kernel functions and logistic regression. Through this transformation, the influence of the original domain of the random variables was eliminated. With these elements a multivariate model for the wind power production using the copula function can be built, to generate possible trajectories and to obtain the confidence intervals of the wind power production. The first developed model represents the time dependence (short term) in the wind power production of a wind farm, and the second model represents the spatial-temporal dependence for several wind farms in the same region.

With a considerable number of simulated trajectories, it is possible to estimate the uncertainty of the wind production, not only in terms of power but also in terms of energy. This information is used to develop actions for compensating the wind power deviations in complex problems. As a first problem, the use of an optimization method is proposed, which seeks to maximize the revenues of a wind power producer, reducing the wind power deviations by the joint operation with a hydro-pump plant. In the second problem, different actions of compensation are formulated as optimization problems, in order to reduce the deviations of a group of wind farms. In all the actions, the power system model is explicitly considered in the optimization problem.

The statistical information obtained with the simulated trajectories makes possible to calculate the joint optimal operation of a hydro-pump plant and a wind power producer. The wind power production uncertainty is included in an optimization problem that uses a hydro-pump plant in an optimal way, reducing the imbalance costs. In all the results of this problem, an economic gain was achieved for the operation in cooperation. The procedure can be expanded to independent control centers for renewable energy and some other operational issues.

The hydro-pump plant is used to minimize the imbalances due to the errors in the wind power forecasting when the wind producer trades in energy markets. In this way, the utilities may reduce the risk due to the uncertainty in the wind power prediction and in prices of the reserve market. Eventually, the imbalances in the wind power production could be fully avoided by the action of the hydro-pump plant and others storages devices.

Finally, different alternatives to compensate wind power deviations in power systems are compared. The different approaches were tested on a modified MATPOWER 30-bus test system modeling the complexity of a real system. The different alternatives were validated by using the Probabilistic Load Flow in the electrical power system where wind power uncertainty is considered. With this method the probability density function of any variable in the power system can be estimated. These variables might be the charge level of the transmission line, the losses in the system, the use of storage devices, etc. The wind power and storage effects in the power systems can be observed by comparing the output variables (generation, area interchange flows, system losses and storage use) of the Monte Carlo method in the power systems.

PUBLICACIONES

Algunas ideas y resultados de este trabajo han sido publicadas previamente por el autor:

- Jaramillo Duque, Á., Castronuovo, E. D., Sánchez, I., y Usaola, J. Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer. *Electric Power Systems Research*, 81(9):1767–1777, septiembre 2011.
- Jaramillo Duque, Á., Sánchez, I., Castronuovo, E., y Usaola, J. Simulating scenarios and prediction intervals in wind power forecasting with the beta distribution. En *11th. Spanish-Portuguese Conference on Electrical Engineering (11CHLIE)*, página 5. Zaragoza, Spain, junio 2009.
- Castronuovo, E., Usaola, J., y Jaramillo Duque, Á. Delegated dispatching of wind farms: an optimal approach considering continuous control and interruption capabilities. *Wind Energy*, 12(4):332–347, septiembre 2008.

«La educación puede ser la gran fecundadora o la gran esterilizante de los hombres, y los hombres pueden ser los grandes transformadores de los pueblos o los más formidables obstáculos para su progreso.»

— Héctor Abad Gómez

AGRADECIMIENTOS

Quiero agradecer a Edgardo D. Castronuovo y a Julio Usaola García, por dirigir esta tesis de doctorado.

A Ismael Sánchez y a Miguel García Lobo, por sus ideas y tutorías en los modelos estadísticos.

A todos mis compañeros de la universidad, con los que he compartido más que charlas técnicas y que siempre han estado allí cuando los he necesitado.

A todos mis amigos, algunos en la distancia y a los otros que he descubierto en este viaje que comenzó ya hace algunos años. Gracias por su apoyo y compañía cuando la he necesitado, ya que no todo ha sido fácil.

Y finalmente, un especial agradecimiento a mi familia, que estado en la distancia, siempre me ha dado su apoyo incondicional, y porque este logro también es de ellos.

ÍNDICE GENERAL

I	INTRODUCCIÓN Y OBJETIVOS DE LA TESIS	1
1	INTRODUCCIÓN	3
1.1	Energía eléctrica	3
1.2	Energía eólica del mundo, Europa y España	4
1.2.1	Situación de España	6
1.3	Retos de la energía eólica	7
1.3.1	Estimación de la incertidumbre	8
1.4	Sistemas de almacenamiento de energía	8
1.4.1	Central hidroeléctrica reversible	9
1.5	Algunas aplicaciones del bombeo	11
1.5.1	Para el productor eólico	11
1.5.2	Para un grupo de productores	13
1.6	Objetivos de la tesis	13
1.7	Estructura y contenido de la tesis	14
2	ESTADO DEL ARTE	15
2.1	Introducción	15
2.2	Modelado de la producción eólica	15
2.3	Uso de la incertidumbre en la operación	17
2.4	Almacenamiento en el sistema de energía eléctrica	19
2.5	Definición del problema	20
II	MODELADO DE LA PRODUCCIÓN EÓLICA	23
3	MODELO MULTIVARIANTE Y CÓPULAS	25
3.1	Introducción	25
3.2	Formulación de las funciones cópula	26
3.3	Familias de funciones cópula	28
3.3.1	Cópulas elípticas	28
3.3.2	Cópulas arquimedianas	29
3.4	Estimación de los parámetros para las funciones cópula	30
3.4.1	Método de máxima verosimilitud	30
3.4.2	Método de inferencia de funciones para las marginales	31
3.4.3	Método de la máxima verosimilitud canónica	32
3.5	Modelo de la producción eólica, funciones cópula	32
4	ANÁLISIS DE LOS DATOS	35
4.1	Introducción	35
4.2	Caso de estudio	36
4.2.1	Producción de potencia eólica y su incertidumbre	37
4.2.2	Relación entre la predicción de potencia eólica y el error	38

4.2.3	Relación entre el horizonte de predicción y los errores	41
4.3	Evaluación de las funciones marginales estimadas . . .	43
4.3.1	Transformada integral de probabilidad	44
4.3.2	Continuous Ranked Probability Score (CRPS) . . .	45
5	DISTRIBUCIÓN BETA COMO FUNCIÓN DE DENSIDAD	49
5.1	Introducción	49
5.2	Definición de la distribución beta	50
5.3	Estimación de los parámetros	51
5.4	Varianza de los errores en la predicción	53
5.4.1	Regresión polinómica por intervalos	53
5.4.2	Regresión polinómica directamente de los datos para estimar la varianza en los errores	55
5.4.3	Regresión kernel	58
5.5	Evaluación de la función de distribución	60
6	FUNCIÓN DE DENSIDAD POR MEDIO DE FUNCIONES KERNEL	63
6.1	Introducción	63
6.2	Modelo local de la PDF por medio de funciones kernel	64
6.2.1	Funciones kernel	65
6.3	Funciones de ponderación	68
6.3.1	Ponderación por antigüedad	68
6.3.2	Ponderación según el valor de la predicción	69
6.4	Evaluación de la función de distribución	71
7	REGRESIÓN LOGÍSTICA PARA MODELAR EL ERROR EN LA PREDICCIÓN	73
7.1	Introducción	73
7.2	Formulación	74
7.2.1	Estimación de los coeficientes de la regresión	75
7.3	Integración de los modelos	75
7.4	Evaluación de la función de distribución	77
7.5	Conclusiones	80
8	ESTIMACIÓN DE LA DEPENDENCIA DE LA FUNCIÓN CÓPULA NORMAL	83
8.1	Introducción	83
8.2	Transformación al dominio uniforme	84
8.3	Cálculo estacionario	85
8.4	Estimación recursiva con factor de olvido exponencial	85
8.5	Estimación recursiva con media móvil ponderada exponencialmente	86
8.6	Estructura de dependencia para la cópula Normal	87
8.7	Modelo temporal	87
8.7.1	Simulación de trayectorias	89
8.7.2	Desvíos de energía	91
8.8	Modelo espacial y temporal	92
8.9	Conclusiones	97

III APLICACIONES A SISTEMAS ELÉCTRICOS	99
9 OPERACIÓN CONJUNTA DE UNA CHR Y UN PARQUE EÓLICO	101
9.1 Introducción	101
9.2 Definición del problema	102
9.3 Operación óptima de una Central CHR	103
9.3.1 Resultados	105
9.4 Operación conjunta de una CHR y un productor eólico	106
9.4.1 Desvíos de potencia	107
9.4.2 Reservas de energía	107
9.4.3 Resultados de la operación conjunta	109
9.5 Operación real de la CHR cubriendo los desvíos	111
9.6 Costos de la programación anticipada y real	111
9.7 Conclusiones	120
10 ANÁLISIS DE LOS DESVÍOS EN LA GENERACIÓN EÓLICA	121
10.1 Introducción	121
10.2 Descripción del caso	122
10.2.1 Sistema de potencia	122
10.2.2 MATPOWER como herramienta de simulación	123
10.2.3 Simulaciones	125
10.3 Resultados	127
10.3.1 Análisis	130
10.4 Flujo de Potencia Probabilístico	135
10.4.1 Simulación	136
10.4.2 Análisis	142
IV CONCLUSIONS AND FUTURE WORK	145
11 CONCLUSIONS	147
11.1 Original contributions	150
11.2 Future work	150
11.2.1 Enhancing the integration of stochastic energy sources in large power system, as the Iberian system.	150
11.2.2 Stochastic model.	151
11.2.3 Impact of large-scale storage.	151
11.2.4 Speed up simulations with a GPU	151
11.2.5 Planing evaluation.	151
11.2.6 Unit Commitment and wind farms.	151
V APÉNDICES	153
A MERCADO DE ELECTRICIDAD	155
B IEEE 30 BUS TEST CASE	157
BIBLIOGRAFÍA	161

ÍNDICE DE FIGURAS

Figura 1.1	Potencia eólica instalada por país en Europa, datos de 2010.	5
Figura 1.2	Esquema de una central hidroeléctrica reversible.	10
Figura 1.3	Principales centrales hidroeléctricas reversibles en España.	11
Figura 1.4	Central de la Muela, Valencia, España.	12
Figura 3.1	Esquema de la transformación integral de probabilidad.	27
Figura 3.2	Modelado mediante funciones cópula.	28
Figura 3.3	Modelo de la cópula normal	33
Figura 4.1	Relación de la velocidad del viento y la potencia producida por un generador eólico.	38
Figura 4.2	Ejemplo de la predicción y producción de la potencia eólica de un parque	39
Figura 4.3	Ejemplo de la relación entre la predicción de potencia eólica y el error, horizonte de predicción $k = 30$ h.	40
Figura 4.4	Ejemplo de la relación entre la predicción de potencia eólica y el error, horizonte de predicción $k = 5$ h.	42
Figura 4.5	Relación de la predicción eólica y la producción usando un diagrama de cajas.	43
Figura 4.6	Esquema de evaluación de la función de distribución, $U = F(X)$	45
Figura 4.7	Algunas de las posibles formas que puede adquirir el histograma Probability Integral Transform (PIT).	46
Figura 4.8	47
Figura 5.1	Diferentes formas que puede tomar la PDF de la distribución beta.	51
Figura 5.2	Ejemplo de la incertidumbre de la producción eólica usando la distribución beta. En este caso cada una de las distribuciones ha sido condicionada al nivel de potencia prevista y a la varianza de los errores, tabla 5.1.	52
Figura 5.3	Ejemplo de la incertidumbre de los errores en la predicción eólica usando la distribución beta. En este caso cada una de las distribuciones ha sido condicionada al nivel de potencia prevista y a la varianza de los errores, tabla 5.1.	52

Figura 5.4	Ejemplo del polinomio de cuarto orden estimado para la varianza de los errores en función de la potencia prevista y el horizonte de predicción. En este caso $k = 15$ h.	54
Figura 5.5	Ejemplo del polinomio de cuarto orden estimado para modelar la varianza o la desviación estándar de los errores, $k = 22$ h.	56
Figura 5.6	Comparación de diferentes técnicas para estimar la varianza de los errores en la predicción, $k = 29$ h.	60
Figura 5.7	Evaluación de la distribución beta, para diferentes técnicas de estimación de la varianza.	61
Figura 5.8	Evaluación de la distribución beta sin considerar las producciones nulas y sus respectivas predicciones. En este caso la varianza fue obtenida por medio de la técnica de mínimos cuadrados.	62
Figura 6.1	Ejemplo de los valores seleccionados según el ancho de la ventana de datos.	64
Figura 6.2	Funciones kernel.	67
Figura 6.3	Estimación de la PDF usando diferentes funciones kernel. Función teórica: distribución beta, con los parámetros $a = 2$ y $b = 5$	68
Figura 6.4	Ejemplos de la ponderación de una observación según el factor de olvido.	69
Figura 6.5	Ponderación según la distancia relativa.	70
Figura 6.6	Evaluación de la función de distribución estimada mediante funciones kernel.	71
Figura 7.1	Ejemplos de los modelos estimados para la probabilidad de las producciones nulas usando la regresión logística, horizonte de predicción $k = 20$ h.	76
Figura 7.2	Evaluación de los modelos de la probabilidad de que ocurra una predicción nula.	77
Figura 7.3	Escalado de la función de distribución de probabilidad.	78
Figura 7.4	Esquema de evaluación de la función de distribución con la regresión logística.	79
Figura 7.5	Comparación de los modelos propuestos para las funciones marginales, utilizando el criterio de evaluación CRPS.	80
Figura 7.6	Evaluación las funciones de distribución propuestas.	81
Figura 8.1	Matriz de covarianzas, diferentes técnicas de estimación.	88

Figura 8.2	Trayectorias simuladas de producción eólica, modelo temporal.	90
Figura 8.3	Intervalos de predicción basados en la simulación de trayectorias, modelo temporal.	91
Figura 8.4	Desvíos de energía para un período de 24h	92
Figura 8.5	Estructura de dependencia modelo espacial temporal	94
Figura 8.6	Trayectorias simuladas de producción eólica, modelo temporal-espacial.	95
Figura 8.7	Intervalos de predicción basados en la simulación de trayectorias, modelo temporal-espacial.	96
Figura 9.1	Operación óptima para una central hidroeléctrica reversible.	106
Figura 9.2	Operación óptima para una central hidroeléctrica reversible que cubre los errores de un productor eólico.	110
Figura 9.3	Comparación del programa obtenido con antelación y la operación real	112
Figura 9.4	Intervalos de predicción basados en la simulación de trayectorias	115
Figura 9.5	Precio de la energía para días típicos en una semana.	116
Figura 9.6	Precio de los desvíos de la energía para días típicos en una semana.	116
Figura 10.1	«IEEE 30 Bus Test Case», donde se agregó generación eólica en los nudos 19 y 14, además de centrales hidroeléctricas reversibles en los nudos 15 y 21.	123
Figura 10.2	Tensiones en los nudos.	132
Figura 10.3	Potencia activa inyectada por los generadores.	132
Figura 10.4	Potencia reactiva inyectada por los generadores.	133
Figura 10.5	Potencia activa generada por área.	137
Figura 10.6	Potencia reactiva generada por área.	137
Figura 10.7	Potencia aparente generada por área.	138
Figura 10.8	Potencia activa generada por unidad, acciones de compensación.	138
Figura 10.9	Potencia activa generada, método de Monte Carlo.	139
Figura 10.10	Intercambios de potencia activa entre áreas.	140
Figura 10.11	Intercambios de potencia reactiva entre áreas.	140
Figura 10.12	Intercambios de potencia aparente entre áreas.	141
Figura 10.13	Perdidas en el sistema.	141
Figura 10.14	Tensiones en el sistema, caso FPO.	142

ÍNDICE DE TABLAS

Tabla 1.1	Lista de países en el mundo de acuerdo a la potencia eólica instalada, datos de julio de 2011.	5
Tabla 1.2	Potencia eólica instalada en España por comunidad a finales del 2010.	6
Tabla 3.1	Algunos generadores para las funciones cópula arquimedianas.	30
Tabla 5.1	Parámetros de la distribución beta utilizados en las figuras 5.2 y 5.3.	53
Tabla 7.1	Resumen de la evaluación de las diferentes funciones de distribución propuestas.	80
Tabla 9.1	Datos técnicos de la CHR.	105
Tabla 9.2	Costos de operación de la central hidroeléctrica reversible y el productor eólico, invierno. . .	117
Tabla 9.3	Costos de operación de la central hidroeléctrica reversible y el productor eólico, verano. . .	118
Tabla 10.1	Potencia activa, reactiva y aparente generada por área y sistema.	129
Tabla 10.2	Pérdidas en el sistema y costo de la generación.	130
Tabla 10.3	Intercambios de potencia entre áreas.	131

NOMENCLATURA

A continuación se describe la nomenclatura utilizada para las diferentes variables, parámetros y funciones utilizadas en el texto. Los símbolos en **negrita**, denotan vectores y/o matrices. Cuando las variables han sido marcadas con el acento circunflejo, $\hat{\cdot}$, significa que su valor ha sido estimado.

α	Vector que contiene los parámetros que definen la dependencia en una función cópula.
α_A	Parámetro para la familia de funciones cópula arquimedianas.
β	Vector que contiene los parámetros que definen las funciones marginales univariantes de una función cópula.
δ	Ancho de la ventana de datos para el modelo local.
δ_i	Argumento de la tensión en cada nudo. [°]
η_H	Eficiencia de la central hidroeléctrica. [%]
η_P	Eficiencia de la central de bombeo. [%]
κ	Número de clases del histograma PIT .
μ	Representa la esperanza matemática de una variable aleatoria.
ν	Número de grados de libertad de la distribución t Student.
Φ^{-1}	Inversa de la CDF de la distribución Normal.
Φ_Σ	CDF de la distribución Normal multivariante.
ψ	Altura de cada una de las clases del histograma PIT .
ρ	Coefficiente de correlación lineal de Pearson.
ρ_r	Coefficiente de correlación de Spearman.
ρ_w	Densidad del aire. [kg/m ³]
Σ	Matriz de covarianzas.
σ	Representa la desviación estándar.
σ^2	Representa la varianza de una variable aleatoria .
τ	Coefficiente de correlación de Kendall.
ε_i	Perturbación aleatoria con media cero.

φ	Generador para una cópula arquimediana.
A	Área que abarca las palas de un aerogenerador. [m^2]
α	Parámetro de la distribución beta.
\mathcal{B}	Coefficientes de la regresión Logística.
b	Parámetro de la distribución beta.
$C_{P,e}(\cdot)$	Coefficiente de potencia eléctrica.
C	Función cópula.
C_{Σ}	Función cópula normal.
$C_{\Sigma, \nu}$	Función cópula t.
c_i^{buy}	Precio de compra de la energía de reserva, proporción del precio de la energía en el mercado diario en el período i . [%]
c_i^{sell}	Precio de venta de la energía de reserva, proporción del precio de la energía en el mercado diario en el período i . [%]
c_i	Precio horario de la energía en el mercado diario en el período i . [$\text{€}/(\text{pu}\cdot\text{h})$]
c_{Pi}	Precio horario de la energía cuando la central bombea en el período i . [$\text{€}/(\text{pu}\cdot\text{h})$]
$E(\cdot)$	Esperanza matemática.
E_1^{esp}	Nivel especificado de energía almacenada en el embalse. [$\text{pu}\cdot\text{h}$]
E_{n+1}^{esp}	Nivel esperado de energía almacenada en el embalse. [$\text{pu}\cdot\text{h}$]
E_i^R	Capacidad de almacenamiento de energía adicional a reservar en el período i . [$\text{pu}\cdot\text{h}$]
E_i^S	Capacidad de energía almacenada mínima que se debe mantener en el período i . [$\text{pu}\cdot\text{h}$]
E^u	Límite superior de la energía almacenada en el embalse. [$\text{pu}\cdot\text{h}$]
E_i	Energía útil almacenada en el embalse en el período i . [$\text{pu}\cdot\text{h}$]
e	Errores de la predicción eólica. [pu]
$e^{(s)}$	Trayectorias simuladas de los errores en la predicción eólica. [pu]
F	Función de distribución de probabilidad (CDF).
f	Función de densidad de probabilidad (PDF).
f_{Pi}	Función de costo del generador i . [€]

h	Parámetro de suavizado en las funciones kernel.
k	Variable que se refiere al horizonte temporal. [h]
$m(\cdot)$	Función que relaciona la variable dependiente con la explicativa.
n	Número de períodos.
n_g	Conjunto que contiene los índices de los nodos donde está conectado un generador.
n_{TIES}	Conjunto que contiene los índices de los nodos donde hay una línea de transmisión que interconecta dos área del sistema.
npc	Número de parejas concordantes, coeficientes de correlación de Kendall.
npd	Número de parejas discordantes, coeficientes de correlación de Kendall.
\bar{P}_{Hi}	Potencia programada, generación hidroeléctrica en el período i . [pu]
\bar{P}_{Pi}	Potencia programada, consumo bombeo en el período i . [pu]
\mathcal{P}	Probabilidad de ocurrencia del suceso de interés de la regresión Logística.
\hat{P}_{Wi}	Predicción eólica en el período i . [pu]
P_H^u	Límite de operación superior de la central operando como generador. [pu]
P_P^u	Límite de operación superior de la central operando como bomba. [pu]
P_{Wi}^M	Desvío esperado de sobreproducción del parque eólico en el período i . [pu]
P_{Wi}^m	Desvío esperado de subproducción del parque eólico en el período i . [pu]
P_{0i}	Potencia activa en el nudo i del caso FPO ₀ .
P_{Hi}	Potencia producida por la central hidroeléctrica en el período i . [pu]
P_{Pi}	Potencia consumida por la central operando como bomba en el período i . [pu]
P_{Wi}	Potencia producida por el productor eólico en el período i . [pu]

$P_{D i}$	Potencia activa demandada en el nudo i . [MW]
$P_{G i}$	Potencia activa generada por cada central en el nudo i . [MW]
\hat{p}	Predicción eólica. [pu]
$\mathbf{p}^{(s)}$	Trayectorias simuladas de producción eólica. [pu]
p	Producción eólica. [pu]
$Q_{0 i}$	Potencia reactiva en el nudo i del caso FPO ₀ .
$Q_{D i}$	Potencia reactiva demandada en el nudo i . [MVAR]
$Q_{G i}$	Potencia reactiva generada por cada central en el nudo i . [MVAR]
t^{-1}	Inversa de la CDF de la distribución t Student.
$t_{\Sigma, \nu}$	CDF de la distribución t Student multivariante.
U	Variable aleatoria en el dominio uniforme.
u_j	Distancia relativa de cada uno de los datos seleccionados por la ventana de datos.
V_i	Modulo de la tensión en el nudo i . [pu]
$\text{Var}(\cdot)$	Varianza.
v	Velocidad del viento. [m/s]
$W_{hi}(\cdot)$	Función de ponderación para la regresión kernel.
X	Variable aleatoria independiente y continua.
\mathcal{X}	Conjunto de variables explicativas para la regresión Logística.
Y	Variable aleatoria dependiente y continua.
y	Variable respuesta de la regresión Logística.

ACRONIMOS

BOE	Boletín Oficial del Estado
CCG	Centros de Control de Generación
CDF	Cumulative Distribution Function
CECRE	Centro de Control de Régimen Especial
CHR	Central Hidroeléctrica Reversible
CML	Canonical Maximum Likelihood
CRPS	Continuous Ranked Probability Score
CPU	Central Processing Unit
EWMA	Exponentially Weighted Moving Average
FP	Flujo de Potencia
FPO	Flujo de Potencia Óptimo
FPP	Flujo de Potencia Probabilístico
GPU	Graphics Processing Units
IFM	Inference Function for Margins
KDE	Kernel Density Estimation
ML	Maximum Likelihood
MPI	Método de Puntos Interiores
OM	Operador del Mercado
OS	Operador del Sistema
PANER	Plan de Acción Nacional de Energías Renovables
PDF	Probability Density Function
PIT	Probability Integral Transform
RC	Rank Correlation
REE	Red Eléctrica de España
RMSE	Root Mean Square Error
UE	Unión Europea
WWEA	World Wind Energy Association

Parte I

INTRODUCCIÓN Y OBJETIVOS DE LA TESIS

INTRODUCCIÓN

1.1 ENERGÍA ELÉCTRICA

La energía eléctrica se ha convertido en un elemento fundamental en el desarrollo de la sociedad, debido principalmente a su versatilidad frente a otras formas de energía, y a que puede ser transportada y transformada con un alto rendimiento. Para cuantificar el desarrollo de un país o una región, en algunos contextos, es utilizada la cantidad de energía eléctrica demandada como indicador, ya que puede explicar el estado de la industria y economía.

En [82] se estima el agotamiento de los combustibles fósiles: alrededor de 35 años para el petróleo, 107 años para el carbón y 37 años para el gas. Ya que es un hecho el fin de este tipo de fuentes energéticas, además de los problemas ambientales asociados a las mismas, los Estados y empresas se ven en la necesidad de buscar alternativas para su abastecimiento, entre ellas, las energías renovables. No solo para cubrir sus necesidades a futuro, sino para poder garantizar su soberanía y autonomía: depender de otros Estados o empresas proveedoras los hace vulnerables, pierden su autonomía y hace que los recursos energéticos se convierten en elementos estratégicos o de presión.

En la Unión Europea (UE) la dependencia energética registra un aumento constante. En el año 2000, la UE cubría sus necesidades energéticas en un 50 % con productos importados. Según las predicciones, en el peor escenario, si no se toma ninguna medida correctiva, entre el 2020 y el 2030 el porcentaje de dependencia energética podría llegar al 70 %, según el informe de la UE, [87]. Según éste, la dependencia externa implica no solo riesgos económicos, sino también sociales y ecológicos para la UE. Desde el punto de vista geopolítico, el 45 % de las importaciones de petróleo proceden del Oriente Medio, y el 40 % de las importaciones de gas natural viene de Rusia. En el caso particular de la energía eléctrica, las importaciones representan solo el 6 % del total, [29, 87].

En España, a finales del 2010, se contaba con aproximadamente 20 GW de potencia eólica instalada, cubriendo aproximadamente el 16 % de la demanda del sistema de ese mismo año. En lo que respecta a potencia instalada en la península, la eólica representa el 20 % del parque generador, 99,04 GW en el 2010, [79]. Según estas magnitudes y la tendencia al crecimiento de la generación eólica, se hace más importante asegurar una adecuada integración de este tipo de fuente no gestionable dentro del sistema. Para esto, es prioritario desarrollar las tecnologías necesarias para contribuir a este proceso de integración

de la energía eólica, mitigando así las desventajas que tiene frente a otros tipos de fuentes tradicionales.

Para paliar los problemas de previsión del recurso eólico pueden ser utilizados elementos de almacenamiento de energía, como baterías, volantes de inercia, bombeo y condensadores, entre otros, que contribuyen a mitigar los problemas de las fuentes intermitentes.

Adicionalmente tenemos el concepto de redes inteligentes, que entre sus objetivos propone disponer de una capa de comunicación que permita la teleoperación y telemedida a nivel de fuentes de generación y consumo. Este concepto se encuentra en una etapa temprana de desarrollo, por lo que aún no es una alternativa viable para resolver el problema de la integración de las fuentes intermitentes. Sus principales inconvenientes son los desafíos técnicos de instalación y la puesta en marcha que implica.

1.2 ENERGÍA EÓLICA DEL MUNDO, EUROPA Y ESPAÑA

El panorama actual energético en el mundo está cambiando constantemente, en parte debido al cambio climático y al interés por preservar el medio ambiente. También este cambio responde a los conflictos políticos y sociales de los países poseedores de grandes reservas de combustibles fósiles y a eventuales cortes en el suministro, situación que provoca una gran volatilidad en los precios que, usualmente, tienden a subir, [41, 86]. Todos los factores que influyen el panorama actual han impulsado de forma directa o indirecta el desarrollo e instalación de fuentes renovables, por lo que ahora es necesaria su gestión e integración en el sistema.

En junio del 2011, la capacidad instalada de energía eólica en el mundo era de 215 GW, lo que significa un incremento del 18% respecto a los 175 GW de la primera mitad del 2010. En la tabla 1.1 se presenta una lista de los países, de mayor a menor, de acuerdo a la potencia eólica instalada. Según las expectativas y siguiendo las tendencias actuales de crecimiento, se tendrían unos 240 GW de capacidad instalada para finales del 2011, que podrían cubrir alrededor del 3% de la demanda de electricidad mundial, [102].

En la directiva europea de energías renovables (2009/28/CE), donde se ratificaron iniciativas del 2001 y 2003, se marcaron objetivos muy ambiciosos para el 2020. Estos objetivos, llamados comúnmente 20/20/20, se refieren a reducir a un 20% las emisiones de CO₂, aumentar un 20% la eficacia energética y que el 20% de la energía consumida por la UE proceda de fuentes renovables. En lo que se refiere a energía eléctrica, significa que la cuota actual de las energías renovables, 15%, debe aumentar al 34% en el 2020 para la UE, [30].

En la figura 1.1 se presenta un mapa de Europa donde se puede observar la potencia instalada, según datos de finales de 2010.

Tabla 1.1: Lista de países en el mundo de acuerdo a la potencia eólica instalada, datos a julio de 2011, [102].

Puesto	País	GW
1	China	52,8
2	USA	42,4
3	Alemania	28,0
4	España	21,2
5	India	14,6
6	Italia	6,2
7	Francia	6,1
8	Reino Unido	5,7
9	Canadá	4,6
10	Portugal	4,0
	Resto del mundo	29,5
	Total	215,0



Figura 1.1: Potencia eólica instalada por país en Europa, datos de 2010, [30].

1.2.1 *Situación de España*

Según el informe de la World Wind Energy Association (*WWEA*), [102], España es el cuarto país con más potencia eólica instalada en el mundo. A finales del 2010 la potencia eólica instalada era de 20,7 GW y durante ese mismo año se instalaron 1,5 GW, lo que significa un aumento de la potencia instalada del 8% respecto al año anterior, siendo el crecimiento más lento desde el 2003 en términos absolutos [80].

En la tabla 1.2 se puede apreciar el estado de la potencia instalada por comunidad autónoma. Estas magnitudes reflejan condiciones geográficamente favorables para este tipo de tecnología y la extensión de las mismas.

Tabla 1.2: Potencia eólica instalada en España por comunidad a finales del 2010, [80].

Comunidad	Potencia instalada MW
Castilla y León	4803,8
Castilla-La Mancha	3709,2
Galicia	3289,3
Andalucía	2979,3
Aragón	1764,0
Comunidad Valenciana	987,0
Comunidad Foral de Navarra	968,4
Cataluña	851,4
La Rioja	446,6
Principado de Asturias	356,0
País Vasco	153,3
Región de Murcia	189,9
Canarias	138,9
Cantabria	35,3
Islas Baleares	3,7
Extremadura	0,0
Comunidad de Madrid	0,0
Ceuta y Melilla	0,0

Según el informe de Red Eléctrica de España (*REE*), [79], en 2010 la energía eólica cubrió el 16% de la demanda de electricidad en España. Además, durante este mismo año, la eólica superó los valores máximos históricos de potencia instantánea, a nivel horario y diario. El 9 de noviembre se registró el último récord de energía diaria con 315,3 GWh, lo que significó una cobertura del 43% de la demanda

eléctrica de ese día. En febrero de ese mismo año, la energía eólica cubrió el 21 % del total de la demanda de ese mes.

Debido a la naturaleza no controlable de esta fuente de energía, se pueden presentar situaciones extremas. Un ejemplo es lo que sucedió el 9 de noviembre a las 3:35 horas, donde el 54 % de la demanda era cubierta con energía eólica. Sin embargo, el día 26 de junio a las 10:32 horas, apenas cubría el 1 % de la demanda del sistema, [79].

Según el Plan de Acción Nacional de Energías Renovables (PANER) [63], se espera para el 2020 una potencia instalada de 35 GW de eólica terrestre y 3 GW de eólica marina en España. Estos niveles implican un reto tecnológico, en lo que se refiere a la integración en el sistema.

1.3 RETOS DE LA ENERGÍA EÓLICA

Algunas de las fuentes de energía renovable han sido, tradicionalmente, mucho más costosas que las fuentes convencionales. Por este motivo los gobiernos han subvencionado y adoptado mecanismos de incentivos para fomentar su instalación, [80]. Hoy en día la energía eólica se encuentra en una posición muy competitiva en comparación con otras fuentes convencionales, y se espera que mejore aún más en los próximos años [69].

Sin embargo, todas las fuentes renovables presentan inconvenientes técnicos, entre ellos, la variabilidad de los recursos y, ligado a esto, la dificultad de ajustar la producción a la demanda. Otro aspecto es la dificultad para obtener una predicción fiable sobre la disponibilidad del recurso, ya que cada uno tiene sus características. Por ejemplo, en el caso particular de España, la producción eólica tiene una propiedad estacional, ya que durante los meses de verano hay una producción menor en comparación con el resto del año, [37, 80].

Por otro lado, uno de los obstáculos a los que se enfrenta la energía eólica es la capacidad de evacuación de las redes de transmisión, ya que las zonas con mayor potencial eólico se encuentran concentradas geográficamente. Por este motivo, las redes de transmisión limitan la cantidad de potencia que se puede instalar en una región. Con la instalación de nuevas líneas y la adecuación de las subestaciones es posible aumentar la capacidad de evacuación, pero los costos se elevan. Además de esto, están las restricciones por operación, gestión y estabilidad del sistema eléctrico, que están relacionadas con la topología y las características de la red.

Según las tendencias globales, las energías renovables y, sobre todo, la eólica, seguirán creciendo en las próximas décadas, [102]. Así, se hace necesario desarrollar estas tecnologías, y otras que permitan mitigar los problemas y desventajas que éstas tienen. Esto implica que, en el futuro, los parques eólicos tendrán que cumplir con los requisitos y consignas de regulación primaria y de inercia [80].

Tal como se dijo antes, uno de los aspectos débiles de la energía eólica es la dificultad de estimar con precisión la disponibilidad del recurso en un determinado momento y lugar. Por tanto, cuando se trabaja con estimaciones, siempre se está expuesto a un riesgo que es necesario estudiar y estimar.

1.3.1 *Estimación de la incertidumbre*

Debido a la naturaleza del viento, no es posible prever con exactitud la energía eléctrica que será producida a partir de éste. Para solucionar este problema se combina la producción eólica con otras fuentes de energía, con el objetivo de asegurar la cobertura de la demanda. Adicionalmente, la operación conjunta haría posible una mejor integración dentro de los mercados de energía, sujetos a regulaciones basadas en las tecnologías tradicionales como la térmica, hidroeléctrica o nuclear, en los que es posible asegurar una capacidad energética de forma anticipada.

Para garantizar una adecuada integración de la generación eólica en los sistemas eléctricos se debe considerar la incertidumbre asociada a este tipo de fuente no gestionable. Ninguna predicción de potencia eólica es perfecta, por lo que siempre se producen desvíos respecto a la potencia eólica esperada, afectando así el equilibrio del sistema eléctrico, y más aún cuando el nivel de penetración de potencia eólica instalada es alto. Por esta razón es necesario desarrollar herramientas que describan mejor el comportamiento de la producción, de modo que sea posible obtener estimaciones de potencia y energía. Este tipo de información es necesaria para gestionar los posibles desvíos y garantizar el funcionamiento del sistema eléctrico.

Como medida para gestionar los desvíos, en algunos casos [3, 8, 24, 74, 94], son utilizados dispositivos de almacenamiento de energía que permiten absorber la sobreproducción para luego aprovecharla de forma controlada, cuando el sistema lo requiere; o cuando hay un desvío negativo (subproducción), para mantener el equilibrio en el sistema. Lo que se busca es utilizar o almacenar la energía eólica disponible en todo momento.

1.4 SISTEMAS DE ALMACENAMIENTO DE ENERGÍA

Una de las ventajas de instalar sistemas de almacenamiento es que, además de permitir una mejor integración de las energías renovables, darían más flexibilidad al sistema eléctrico. Operando de forma conjunta, un productor eólico y un sistema de almacenamiento podrían comportarse como una fuente convencional gestionable, lo cual reduciría el riesgo al que están expuestos los agentes cuando van a los mercados. También, se podría aprovechar el recurso eólico, incluso

con las restricciones de la red o cuando la demanda es baja, de forma que se almacene la energía y pueda ser utilizada en otro momento.

Existen muchos tipos de tecnologías de almacenamiento, como baterías, volantes de inercia, sistemas electromagnéticos, aire comprimido, bombeo hidráulico e hidrógeno, entre otras; múltiples áreas de la ciencia están involucradas en el desarrollo de estas tecnologías, haciendo que se mantengan en un constante desarrollo.

Cada tecnología de almacenamiento se caracteriza por: cantidad de energía que pueden acumular, tiempos de carga y descarga, rendimientos, número de ciclos de carga o descarga, costos de puesta en marcha y mantenimiento, y disponibilidad de ubicaciones, entre otras, lo que la hace apta para diferentes aplicaciones.

Hoy en día, las centrales hidroeléctricas reversibles son la opción más usada para almacenar energía, [58, 104], gracias a su eficiencia, su capacidad de control y sus dimensiones, características las hace aptas para muchos tipos de aplicaciones en los sistema de energía eléctrica. Es posible modificar embalses ya existentes dedicados al riego, al abastecimiento de agua potable o al uso industrial anexando equipos de bombeo, lo cual potencia aún más esta alternativa [98]. España cuenta con un número considerable de centrales hidroeléctricas reversibles y, según datos obtenidos del Boletín Oficial del Estado (BOE), se espera contar con 3,2 GW más de centrales hidroeléctricas reversibles en el 2020.

1.4.1 Central hidroeléctrica reversible

Las centrales hidroeléctricas reversibles utilizan una de las tecnologías más maduras, ya que han sido utilizada desde los años veinte [80]. Para almacenar energía con una central hidroeléctrica reversible se bombea agua desde una presa de nivel bajo, a una presa con un nivel superior. Un esquema de esto se presenta en la figura 1.2.

Existen dos tipos básicos de centrales de bombeo, de acuerdo a su funcionamiento:

CENTRALES DE BOMBEO PURO: en este tipo de centrales, la presa del nivel superior solo tiene aportaciones de agua debido al bombeo. Por esta razón hace falta tener una cantidad adecuada de agua almacenada en la presa superior para poder generar electricidad.

CENTRALES DE BOMBEO MIXTO: en este tipo de centrales hay aportaciones de agua provenientes de ríos u otros embalses, por lo que es posible su operación como central de generación en cualquier momento, y solo dependiendo de su cota de agua almacenada.

Las centrales hidroeléctricas reversibles pueden conseguir un rendimiento que está alrededor del 70 %, requieren bajo mantenimiento y son alternativas para el almacenamiento de energía, tanto en pe-

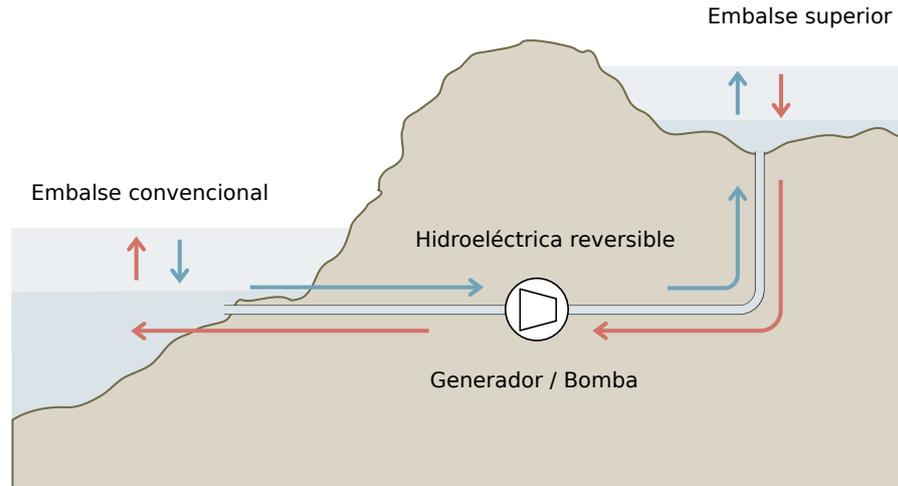


Figura 1.2: Esquema de una central hidroeléctrica reversible.

riodos diarios como semanales. Presentan algunos inconvenientes, ya que necesitan de lugares geográficamente aptos, el costo de puesta en marcha es elevado, producen un impacto ambiental y son proyectos que requieren de varios años hasta poder estar operativos.

1.4.1.1 Situación en España

En la figura 1.3 se pueden ver localizadas algunas de las centrales hidroeléctricas reversibles más importantes en España. Además se muestran algunas en proyecto o construcción. El gobierno de España ha agilizado la puesta en marcha de algunos proyectos de bombeo puro, al considerarlos fuera del ámbito de aplicación del Real Decreto Legislativo 1302/1986, por lo que no se deben someter al procedimiento de evaluación de impacto ambiental. Con estas medidas se busca aumentar la capacidad de almacenamiento del sistema y, a la vez, facilitar la gestión de toda la energía que es inyectada en éste. En España, según datos obtenidos de [62], se espera contar con 3 GW más de centrales de bombeo puro para el 2016, lo que significaría 5,7 GW; esto sin considerar las centrales hidroeléctricas mixtas. Otro proyecto importante, que también influirá en España, es el que aprovecha el río Tâmega, al norte de Portugal, que tendrá una potencia de 1135 MW. Estos proyectos buscan dar garantías al sistema, haciendo posible el aumento de las energías renovables.

En la figura 1.4 se observan dos fotos de la central de la Muela, en el río Júcar, con una potencia de 546 MW, donde se instalarán 848 MW más en una segunda fase del proyecto. Como este proyecto hay muchos más en España [63], lo que ratifica la tendencia hacia un sistema con elementos de almacenamiento.

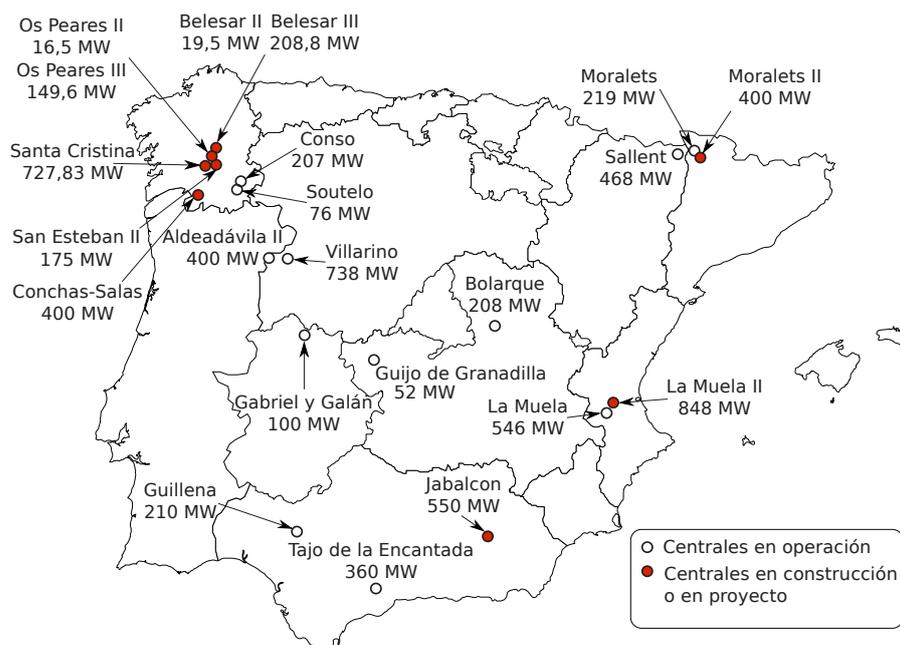


Figura 1.3: Principales centrales hidroeléctricas reversibles en España.

1.5 ALGUNAS APLICACIONES DEL BOMBEO

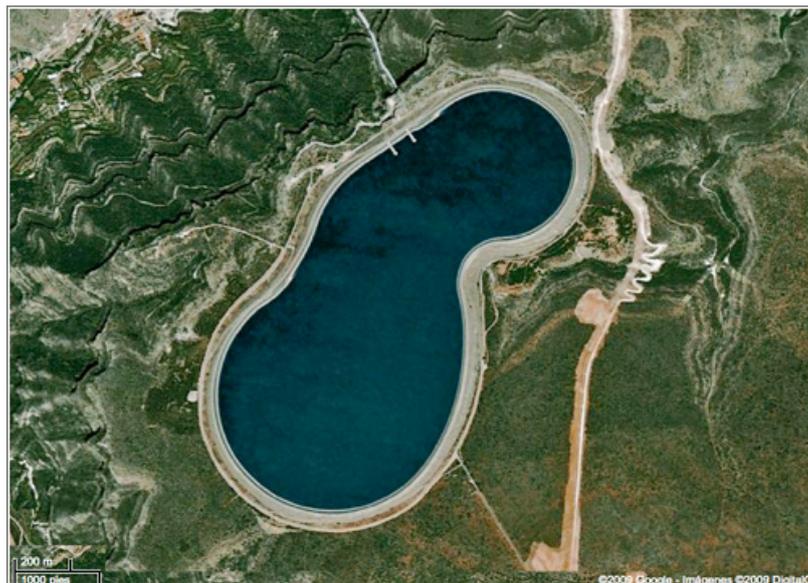
1.5.1 Para el productor eólico

Cualquier productor eólico busca maximizar sus beneficios. Para lograrlo puede optar por una estrategia punta-valle, donde utiliza la energía eólica en *periodos valle* para bombear agua, y luego utiliza el agua almacenada en *periodos punta* para obtener el mayor beneficio. Esta estrategia depende de la diferencia del precio de la energía entre los periodos punta y valle; cuanto mayor sea la diferencia, mayor será el beneficio.

Otra aplicación es la compensación de los desvíos de la producción. Los productores eólicos se comprometen con una cierta cantidad de energía a futuro, basándose en las predicciones de potencia. Cuando no son cumplidos los compromisos, el productor debe pagar por los desvíos. En este caso, se podría utilizar una central hidroeléctrica reversible que asumiera los desvíos. Si el productor eólico produce menos energía que la esperada, la central produciría la energía faltante; en el caso contrario, si el productor eólico produce más energía de la esperada, la central operaría como bomba, almacenando el exceso de energía. Por tanto, si se elige esta medida, el beneficio máximo se obtiene al minimizar los desvíos entre la producción real y la previsión del productor eólico.



(a) Foto aérea, [45].



(b) Foto tomada desde un satélite, [38].

Figura 1.4: Central de la Muela, Valencia, España.

1.5.2 *Para un grupo de productores*

En España, un productor eólico puede ir al mercado de electricidad por medio de un agregador que agrupa la oferta de muchos productores. De esta forma, el representante puede estar interesado en compensar los desvíos, pero de forma global. Para este fin es necesario considerar la ubicación de los parques y la dependencia que existe en la producción de todos ellos. Las centrales hidroeléctricas reversibles reducirían los desvíos, ya sea de forma global, o incluso de forma zonal, donde se apliquen restricciones del sistema de energía eléctrica. Debido a la complejidad del sistema, se delegan en los Centros de Control de Generación (CCG) la tarea de gestionar la energía eólica generada, de forma que ésta sea aprovechada. Además, se consideran las restricciones técnicas del sistema, con el objetivo de mantener la fiabilidad y estabilidad de éste, o para no sobrepasar la capacidad de los elementos de la red [17].

El Operador del Sistema (OS) es el encargado de asegurar el suministro de energía eléctrica. Por esta razón, para el operador también es necesario prever la producción de los parques eólicos, ya que son una de las principales fuentes de incertidumbre en la estimación de la reserva. El OS no solo busca aprovechar la producción proveniente de la generación eólica durante las horas valle, sino también la que proviene de la energía nuclear en la cual, por restricciones técnicas, es complejo modificar el punto de operación. Esta estrategia también está ligada a la intención de alisar la curva de la demanda, que es otro objetivo del OS, [33, 64, 65].

Gracias a las características técnicas de las centrales hidroeléctricas reversibles, éstas también pueden contribuir en la tarea de mantener la frecuencia del sistema de energía eléctrica cuando se presenta una perturbación debida a la pérdida repentina de un generador o de una carga. En el futuro, las centrales hidroeléctricas reversibles y todos los elementos de almacenamiento serán fundamentales en la implementación de las redes inteligentes que buscan mejorar la gestión de los recursos energéticos [33, 64, 65].

1.6 OBJETIVOS DE LA TESIS

El objetivo general de esta tesis es generar trayectorias de producción eólica futura para el análisis probabilístico de los sistemas de energía eléctrica. Esta información puede ser utilizada para el comercio en los mercados de energía, para la determinación y la resolución de restricciones en la operación del sistema y para la estimación de las reservas de energía, entre otros usos. De esta forma se facilita la integración de la energía intermitente y, en particular, de la energía eólica.

Además se busca mejorar la integración de la energía eólica en el sistema eléctrico, determinando la incertidumbre y las posibles tra-

yectorias de la producción eólica. Con esta información es posible operar la reserva del sistema para maximizar el uso de la energía eólica y garantizar el funcionamiento del sistema.

Como primer objetivo específico se obtendrá un método para anticipar posibles trayectorias de la producción eólica futura. Se analizará la dependencia temporal a corto plazo y la dependencia espacial que existe entre parques eólicos localizados en una región geográfica y, de esta forma, se obtendrán trayectorias o escenarios de producción eólica a futuro, ya sea para un parque o para un grupo de parques.

Como segundo objetivo específico se analizarán algunas medidas de compensación de los desvíos en la producción aplicados a casos realistas. Primero se estudiará el caso de un productor eólico que busca maximizar sus beneficios reduciendo los desvíos respecto a la producción con que se ha comprometido. En segundo lugar se modelará un sistema eléctrico donde se analizarán diversas medidas para compensar los desvíos en la producción de un grupo de parques eólicos.

1.7 ESTRUCTURA Y CONTENIDO DE LA TESIS

La tesis está dividida en cuatro partes. En la primera parte se presenta la introducción, objetivos, estado del arte y definición del problema. En la segunda parte se explica el método utilizado para obtener escenarios de la producción eólica a futuro utilizando las funciones cópula. La tercera parte presenta algunas aplicaciones donde se busca mejorar la integración de la generación eólica por medio de la operación conjunta de centrales hidroeléctricas reversibles. Finalmente, en la cuarta parte se presentan las conclusiones generales de la tesis y las expectativas de trabajo en el futuro.

2.1 INTRODUCCIÓN

Las principales variables que introducen incertidumbre en la operación del sistema de energía eléctrica son la demanda, el precio de la energía y la producción de la potencia eólica. Dependiendo del rol de cada agente, es necesario tener herramientas de estimación para cada una de estas variables.

La estimación de la potencia demandada y el precio de la energía son problemas que surgieron con los sistemas de energía eléctrica. Estos temas han sido estudiados ampliamente y hoy en día se cuenta con herramientas que logran dar buenas estimaciones para ambos problemas [22, 25, 40]. Sin embargo, la predicción del viento, que hasta hace poco era un problema exclusivo de la meteorología, se ha tornado de gran interés por su impacto económico y técnico en países donde la potencia eólica instalada tiene una cuota significativa. Un dato fiable sobre la velocidad del viento implica una buena estimación de la producción de potencia eólica. Adicionalmente, hay que considerar otros factores funcionales de la producción eólica que son ajenos al viento y que afectan directamente la potencia producida. Por esta razón son necesarias herramientas que proporcionen una predicción precisa de la producción eólica [54].

2.2 MODELADO DE LA PRODUCCIÓN EÓLICA

Para la coordinación de la producción eólica con cualquier otro tipo de fuente de energía y/o almacenamiento, es necesario estimar de una forma fiable los posibles desvíos entre la predicción a corto plazo y la producción real. En [52] se utiliza una función que estima los posibles errores, usando como única variable el horizonte de predicción. Esta función está basada en el promedio de diferentes métodos para predecir los errores.

La predicción del viento se modela como una serie ARMA en [59], convirtiendo este dato posteriormente en potencia eléctrica, mediante el uso de las curvas de los generadores. Este modelo es utilizado en un problema de optimización que busca aprovechar de forma eficiente la energía generada por un grupo de parques eólicos y un conjunto de centrales hidroeléctricas en el mismo cauce de un río, cuando existe la posibilidad de bombear agua de un embalse a otro. En el problema de optimización se tiene también como restricción la capacidad de las líneas de transmisión que conectan las centrales hidroeléctricas

y los parques eólicos, que limitan la potencia inyectada en el sistema eléctrico.

En [71], la producción eólica es modelada como una variable aleatoria, y la función marginal es estimada a partir de los datos de viento, que son convertidos en potencia eléctrica por medio de las curvas de los generadores eólicos. En este caso se buscaba determinar la dependencia entre diferentes zonas de producción eólica y la demanda del sistema.

Basado en las predicciones y sus respectivas observaciones de producción, en [76] y [72] son estimadas las funciones marginales de la producción como funciones no paramétricas. Luego estas funciones son utilizadas para modelar la dependencia temporal y espacial respectivamente.

En [37], se propone un modelo regional para la producción eólica, donde se modela la producción por medio de una técnica no paramétrica.

El trabajo desarrollado por [26] analiza una zona con gran penetración eólica en Irlanda, con la intención de cuantificar la incertidumbre del sistema y así poder estimar la reserva necesaria para asegurar el funcionamiento del sistema eléctrico. Para este caso, la incertidumbre de la potencia eólica y la demanda del sistema son modelados conjuntamente por una distribución Normal. El modelo está condicionado por el horizonte de predicción y la desviación estándar de los errores de la predicción eólica y de la demanda. Los modelos están basados en datos reales de un parque eólico y los resultados son obtenidos con una herramienta de cálculo difuso. Para el cálculo de la incertidumbre de los errores de la predicción, en una zona donde existen múltiples parques eólicos, se utiliza una función de correlación que se basa en la distancia que existe entre los diferentes parques eólicos.

En [43] se presenta un análisis de los errores en la predicción eólica, donde se muestra una dependencia no sólo del horizonte de predicción, sino también del nivel de potencia prevista. Este concepto es presentado y aplicado en [74, 76], donde se cuantifica la incertidumbre de la predicción eólica a partir de un modelo construido con los datos históricos de un parque eólico, predicciones y producciones reales, donde se considera tanto el horizonte de predicción como los niveles de potencia prevista.

En [93], la producción eólica y la energía demandada son analizadas como dos variables aleatorias para determinar la incertidumbre y cuantificarla. El caso de estudio es una isla, donde la proporción de energía eólica es relevante en comparación con la demanda total. En ese trabajo, el error en la predicción no es relacionado con el horizonte de predicción y los desvíos son calculados de forma absoluta. Finalmente, el método se aplica utilizando perfiles de producción de días típicos y cuantificando los desvíos de una forma económica a partir de escenarios con diferentes niveles de incertidumbre.

La producción de la potencia eólica es representada como una distribución beta en [7], y es usada para calcular la función de probabilidad de los errores en la predicción. Sin embargo, es introducido un factor de ponderación que se ajusta de forma empírica. Para el cálculo de los errores, el horizonte de predicción es usado para determinar los parámetros que condicionan la función de probabilidad.

En [51], se utiliza la técnica de estimación mediante funciones kernel para obtener función de densidad de la producción eólica, ya que el objetivo era obtener la forma de la función de densidad. Para este caso, las producciones nulas son consideradas y modeladas como una discontinuidad y son introducidas como una parte discreta en la distribución que modela la producción eólica. La probabilidad se calcula como la proporción de producciones nulas, dada la potencia prevista.

En [71], se utiliza una función cópula para modelar la dependencia espacial en la producción de diferentes parques eólicos en Holanda. En este caso particular, la variable aleatoria es la velocidad del viento y el comportamiento de la demanda del sistema. La dimensión del problema es de 16 variables aleatorias, donde se utiliza la cópula normal para determinar la dependencia estocástica de estas variables.

En [39], son utilizadas las funciones cópula para determinar la dependencia estocástica del viento en diferentes regiones de Alemania. Con esta información se estima la producción eólica para nuevos emplazamientos con el objetivo de encontrar los lugares con mejores características, a la vez que se reduce el riesgo de inversión. En este caso la dimensión del problema es de 40 variables y se utiliza la cópula normal.

En [97], se presenta un modelo que utiliza una función cópula para estudiar la integración de fuentes de energía intermitentes, fotovoltaica y eólica. Los centros de generación se localizan en una red de distribución en el área de Davarzan en Irán. Las variables del problema son: la velocidad del viento y la radiación solar. En este caso la dimensión del problema es de 11 variables y se utiliza la cópula normal.

Es necesario desarrollar modelos que se adaptan a los datos reales, incorporando las características de las variables. Este trabajo se fundamenta en la necesidad de tener información fiable para la toma de decisiones.

2.3 USO DE LA INCERTIDUMBRE EN LA OPERACIÓN DE LOS SISTEMAS DE ENERGÍA ELÉCTRICA

Un productor eólico oferta la energía en el mercado basándose en la información obtenida por medio de las herramientas de predicción de las que dispone. Sin embargo, si pudiera operar de forma conjunta con una central hidroeléctrica reversible, tendrían la posibilidad de

almacenar energía, llevando agua a un embalse para luego producir energía en períodos donde se obtenga el máximo beneficio, basándose en la diferencia de los ingresos en el mercado y los costos variables. Los costos variables para una central térmica dependen de los precios de combustible, operación y mantenimiento. Para una central hidroeléctrica reversible y un parque eólico no hay gasto en combustible, ya que sus insumos son agua y viento, y para este tipo de casos los costos variables son insignificantes. Finalmente, el objetivo es maximizar los beneficios de forma conjunta en el mercado.

En [107], se presenta un análisis probabilístico que considera el impacto de algunas fuentes de incertidumbre sobre el Flujo de Potencia Óptimo (FPO). El impacto de las variables es abordado por medio de un análisis probabilístico, empleando el método de Monte Carlo, y un proceso de optimización determinista. El objetivo es proponer un método para la evaluación del riesgo en la operación de sistemas de energía eléctrica, enfocado a la toma de decisiones para el OS. El método se desarrolla y aplica a un sistema de 5 nudos, y luego es probado en el sistema IEEE de 30 nudos para ilustrar los resultados.

En [99] se presenta un procedimiento para calcular el flujo de carga probabilístico en un sistema de energía eléctrica, considerando la incertidumbre de la generación eólica. La función de densidad de la producción eólica es obtenida a través de una aproximación cuadrática de la curva típica de viento-potencia de un aerogenerador. El sistema de energía eléctrica es modelado linealmente (flujo de carga en DC). Finalmente los modelos son integrados y reducidos usando técnicas de convolución para reducir el tiempo de cálculo. El método es puesto a prueba en un sistema de 14 nudos, sobre el cual son desarrolladas diferentes pruebas.

En [96] se presenta un flujo de potencia probabilístico donde la generación eólica es modelada con la función beta. El método se basa en las propiedades de la función de densidad de probabilidad y la expansión de Cornish-Fisher, adecuada para funciones no gaussianas. El método presentado se pone a prueba en el sistema tipo llamado «IEEE-RTS96», donde se comparan los resultados con el método de Monte Carlo para comprobar la precisión del procedimiento.

En [57, 107], se utiliza una versión modificada de MATPOWER [108] para desarrollar un análisis probabilístico que considera el impacto de algunas fuentes de incertidumbre sobre el Flujo de Potencia Óptimo (FPO).

Se pueden distinguir dos esquemas de operación conjunta entre una central hidroeléctrica reversible y un productor eólico. En el primero, la programación de la operación de bombeo depende fundamentalmente de la producción eólica, aunque ocasionalmente hay aportes externos de energía que se utilizan para almacenar agua en el embalse. Este tipo de esquema se utiliza frecuentemente en regiones aisladas eléctrica o físicamente, como las islas [4, 12, 13, 98]. En

algunas islas, el uso de bombeo, puede llegar a reemplazar la necesidad de instalar una central térmica [13]. Adicionalmente, el uso de bombeo, como una forma de almacenamiento de energía eólica, resuelve no sólo los problemas ya antes mencionados, sino que además plantea soluciones y alternativas para las reservas de agua potable, la irrigación, la protección contra incendios e incluso la desalinización de agua marina [13, 98].

En el segundo esquema de operación, la central hidroeléctrica reversible opera de forma independiente y además cubre los desvíos en la predicción de un productor eólico, de forma que maximiza los beneficios económicos al mantener la producción eólica lo más cercana posible al programa presentado con antelación. A su vez, el productor eólico puede aportar los excedentes de energía para alimentar la central hidroeléctrica reversible, cuando ésta opera como bomba. Este último esquema está sujeto a restricciones técnicas del sistema, del mercado y, a su vez, a intereses económicos de los agentes que intervienen [5, 18, 36]. En regiones interconectadas y con buenos perfiles de viento es posible que, por restricciones de carácter técnico en el sistema eléctrico, sea limitada la cantidad de energía aportada por los productores eólicos [59]. Esta labor es desarrollada en España por Centro de Control de Régimen Especial (CECRE), lo que obliga en algunos casos a desaprovechar parte de la energía eólica [17]. La operación conjunta proporciona no solo una alternativa para utilizar este excedente, sino también para cubrir los errores en la programación de los parques eólicos.

2.4 ALMACENAMIENTO EN EL SISTEMA DE ENERGÍA ELÉCTRICA

El almacenamiento de energía se realiza mediante dispositivos o medios físicos que la almacenan para realizar operaciones útiles en un momento posterior. El almacenamiento en los sistemas de energía eléctrica permite guardar el exceso de energía, de modo que ésta pueda ser utilizada cuando la demanda de electricidad es mayor.

En [8] es utilizada la incertidumbre de la predicción eólica obtenida de acuerdo a la método descrito en [7] para determinar la dimensión de un dispositivo de almacenamiento que cubrirá los desvíos del productor eólico.

En [94], es presentado un estudio basado en el sistema holandés, donde son utilizados sistemas de almacenamiento alimentados por centrales térmicas de carbón, que son más baratas que las centrales térmicas de gas, y se utilizan solo en los periodos punta. En este caso, se realiza un balance neto de las emisiones de CO₂ en el sistema. El resultando es un aumento de éstas, ya que el sistema de almacenamiento es alimentado por energías no renovables.

En [24] es presentado un modelo con el concepto de central virtual. Este modelo consta de un elemento de almacenamiento y un parque

eólico. El objetivo es maximizar los beneficios de forma conjunta en un entorno de mercado, minimizando los desvíos debidos a los errores en las predicciones de potencia eólica.

En [74], es utilizada una técnica que simula posibles trayectorias de producción eólica a corto plazo. Con esta información se estima la dimensión de una central de almacenamiento de energía, con el objetivo de cubrir los desvíos de la predicción eólica. Este problema es desarrollado en un contexto de mercado de energía, donde los desvíos en la producción eólica tienen un valor económico.

En [3] se presenta un modelo conjunto de una central hidroeléctrica reversible y un productor eólico como una solución para redes aisladas, con el fin de aumentar la penetración de la energía eólica y los beneficios económicos. Este trabajo presenta un método para el dimensionado óptimo de los distintos componentes de una central hidroeléctrica reversible, diseñada para almacenar los excedentes de energía eléctrica proveniente de un parques eólico, los cuales son rechazados debido a las limitaciones impuestas por el OS. El método es probado utilizando datos reales de la energía rechazada proveniente de un número de parques eólicos instalados en la isla de Creta, en Grecia.

2.5 DEFINICIÓN DEL PROBLEMA

Tal y como se dijo en los objetivos de la tesis, sección 1.6, para generar trayectorias de la producción eólica futura es necesario determinar la incertidumbre asociada a este tipo de fuente no gestionable. El modelo para generar trayectorias de producción tendrá que obtenerse a partir de la información que poseen los productores eólicos y el operador del sistema. Usualmente éstos solo disponen de un registro histórico de las producciones eólicas obtenidas a partir de los dispositivos de medida y las respectivas predicciones, que en la mayoría de los casos son deterministas. Esto significa que en los registros solo se cuenta con un valor para la potencia prevista para cada horizonte de predicción y el valor respectivo de la potencia observada. El modelo de la incertidumbre de la producción eólica deberá considerar los casos de producción nula, los cuales son usuales a nivel de parques y puede llevar a errores.

Para un parque eólico, un modelo temporal a corto plazo considera la producción promedio por hora durante dos días; esto quiere decir que el modelo tienen una dimensión de 48 variables. Adicionalmente, si se modela la dependencia espacial que existe con otro parque, la dimensión del problema se duplica, en este caso, 48 variables por cada parque. Para simular trayectorias de producción eólica futura se debe buscar un método que se adapte a las características y dimensiones de este problema. Por estas razones se exploran en este trabajo algunos de los métodos más utilizados para obtener funciones conjuntas,

sobre todo aquellos métodos que han sido utilizados en problemas de dimensión y complejidad semejante. De esta forma es posible realizar una simulación utilizando el método de Monte Carlo, donde se modelará una realización completa de un posible escenario de producción eólica temporal y/o espacial. Con un conjunto de realizaciones simuladas se podrá estimar la incertidumbre futura de las variables que dependen de la producción eólica. Esta información es útil para estimar la reserva del sistema, realizar análisis de la operación, participar en mercados y gestionar el riesgo, entre otras aplicaciones.

Para abordar el problema de la operación conjunta de un productor eólico y una central hidroeléctrica reversible, primero se propone la solución de un problema de optimización que busca obtener el mayor beneficio de una central hidroeléctrica reversible operando conjuntamente con un parque eólico. Como entrada en el problema de optimización se utilizarán los resultados del modelo de producción eólica a corto plazo (modelo temporal) para estimar la incertidumbre de esta fuente no gestionable. Como resultado se espera obtener el programa óptimo de la central hidroeléctrica reversible que maximice el beneficio de la operación conjunta.

Finalmente, se comparan diferentes alternativas para compensar los desvíos de la producción eólica por parte del OS como problemas de optimización. En estos casos, el modelo del sistema eléctrico es representado explícitamente y se considera la generación de un grupo de parques eólicos. Para cuantificar el impacto de la incertidumbre sobre el sistema de energía eléctrica se aplicará el método de Monte Carlo combinado con un proceso de optimización determinista. De esta forma se espera evaluar la fiabilidad en sistemas de energía eléctrica. Los resultados serán una herramienta para evaluar los efectos de la incertidumbre sobre el rendimiento y operación del sistema de energía eléctrica.

Para obtener la solución de los Flujo de Potencia (FP) y los Flujo de Potencia Óptimo (FPO), será utilizado MATPOWER [108, 112]. Este programa es software libre, por lo que es posible modificarlo y extenderlo y así programar funciones objetivo alternativas a la opción por defecto. De esta forma será posible modelar el sistema eléctrico y evaluar diferentes alternativas para la gestión de la reserva.

Parte II

MODELADO DE LA PRODUCCIÓN EÓLICA

En esta parte de la tesis se pretende desarrollar un modelo estocástico de la producción eólica. Con este modelo se busca obtener escenarios de producción dada la predicción de potencia a futuro. Además se buscará estimar, en base a la predicción, la incertidumbre de la potencia y la energía a futuro (corto plazo) de un parque o de un grupo de parques eólicos.

Primero se presenta un resumen de las funciones cópula como herramienta para el modelo multivariante. Luego se analizan los datos de un parque eólico, predicciones y producciones. Una vez identificadas las características de las variables, se continúa con la estimación de las funciones de densidad univariantes para la producción eólica, con diferentes horizontes de predicción. Para las funciones marginales univariantes se considerará un modelo paramétrico basado en la distribución beta y una alternativa basada en la técnica de estimación de la función de densidad mediante funciones kernel, siendo éste un método no paramétrico.

Para modelar los errores en la predicción, cuando la producción es cero, se utiliza la regresión logística para estimar la probabilidad de una variable con respuesta cualitativa.

Finalmente son presentadas y analizadas diferentes formas de obtener la estructura de dependencia de las variables aleatorias desde el dominio uniforme.

Una vez obtenidos todos los elementos necesarios para construir un modelo multivariante que simule la producción eólica usando funciones cópula, se presenta el modelo temporal de la producción eólica de un parque. Luego éste se extiende para modelar la producción de varios parques en una misma zona (modelo espacial-temporal).

Resumen

En este capítulo se presenta un resumen de la teoría de las funciones cópula para realizar modelos multivariantes. Además son presentadas algunas técnicas para estimar los parámetros de estas funciones y, finalmente, se expone el procedimiento que se desarrollará en los capítulos siguientes para conseguir el modelo multivariante con el que se obtendrán las trayectorias de la producción eólica futura.

3.1 INTRODUCCIÓN

Cuando se trabaja con energía eólica, una de las tareas más difíciles es considerar la incertidumbre asociada a la producción futura. Desde el punto de vista técnico, todos los errores en la previsión de energía eólica deben ser compensados por otro generador para suministrar la potencia demandada por la carga. Desde el punto de vista del mercado de electricidad, la incertidumbre en la predicción aumenta el riesgo al que los productores eólicos están expuestos, ya que los desvíos en la producción usualmente provocan una reducción en los ingresos. Estos factores aumentan el precio de la energía, siendo éstas las razones principales que llevan a tratar de cuantificar la incertidumbre en la previsión de la energía eólica y, de esta forma, conseguir una operación más eficiente.

Los productores eólicos se basan en herramientas de previsión físicas, estadísticas o una combinación de ambas, para estimar la producción de potencia eólica en un período a corto plazo. Esta información generalmente es requerida por el OS para realizar los programas de operación y los agentes deben presentarlo con anterioridad a la operación. No obstante, esta información es utilizada también para el desarrollo de estrategias de mercado por las que estos datos son también necesarios y valiosos.

Usualmente la información entregada por las herramientas de predicción consiste en la estimación de la producción en un horizonte de tiempo, así como los niveles de fiabilidad asociados a ésta. Sin embargo, dicha información, generalmente, no es suficiente si se requiere conocer el orden de los posibles desvíos de potencia y energía a futuro.

La cuantificación de los posibles desvíos de la energía y la definición de los intervalos de predicción son necesarias para conseguir una operación apropiada del sistema eléctrico, a través del dimen-

sionamiento de los elementos de almacenamiento, del desarrollo de programas de reservas y de estrategias de mercado, además de otros problemas de la operación del sistema.

La motivación para desarrollar el presente trabajo nace de la necesidad de cuantificar los posibles desvíos de potencia y energía, considerando la dependencia entre los errores y la producción eólica en diferentes horizontes. La Distribución de Probabilidad Conjunta podría ser una solución para cuantificar los posibles desvíos de energía durante un período futuro, pero las características de no-linealidad de la producción eólica hacen esta tarea demasiado compleja y fuera del alcance de esta tesis. Por esta razón se realiza un análisis de los datos de un parque y se propone un método alternativo para simular posibles producciones de potencia eólica para un horizonte de corto plazo.

Para realizar un modelo multivariante de la producción eólica será utilizada la teoría de funciones cópula, por ser un método robusto y utilizado en problemas de gran dimensión y complejidad semejante [39, 71, 72, 97, 103]. Para obtener un modelo con las funciones cópula se presenta un procedimiento que consta de dos pasos: primero, se determinan las distribuciones de probabilidad univariantes de cada variable, como si se tratase de variables independientes; segundo, se determina la estructura de dependencia de la función cópula, en otras palabras, la dependencia que existe entre las variables aleatorias univariantes. Esta estructura se puede obtener desde el dominio uniforme de las variables aleatorias; de esta forma se evita la influencia de las funciones marginales en la estimación [71]. Desde un punto de vista práctico, estimar el modelo en varios pasos reduce la complejidad en la estimación de cada uno de los parámetros, tanto de las funciones marginales como de la estructura de dependencia.

A continuación se presenta una breve descripción de las funciones cópula, con el objetivo de tener los elementos necesarios para realizar y entender un modelo multivariante.

3.2 FORMULACIÓN DE LAS FUNCIONES CÓPULA

En la teoría de probabilidad y estadística, la cópula es una función que pueden describir la dependencia de un conjunto de variables aleatorias. El concepto fue introducido por Sklar en 1959, [85], y es reconocido como un poderoso instrumento para modelizar la dependencia entre variables aleatorias.

En muchos campos de la ciencia el uso de las funciones cópula es relativamente nuevo, pero se encuentra en rápido crecimiento. Ejemplos de su uso en demografía, finanzas, medicina e ingeniería se pueden encontrar en [71, 103], siendo una garantía de su madurez, simple implementación y gran potencial.

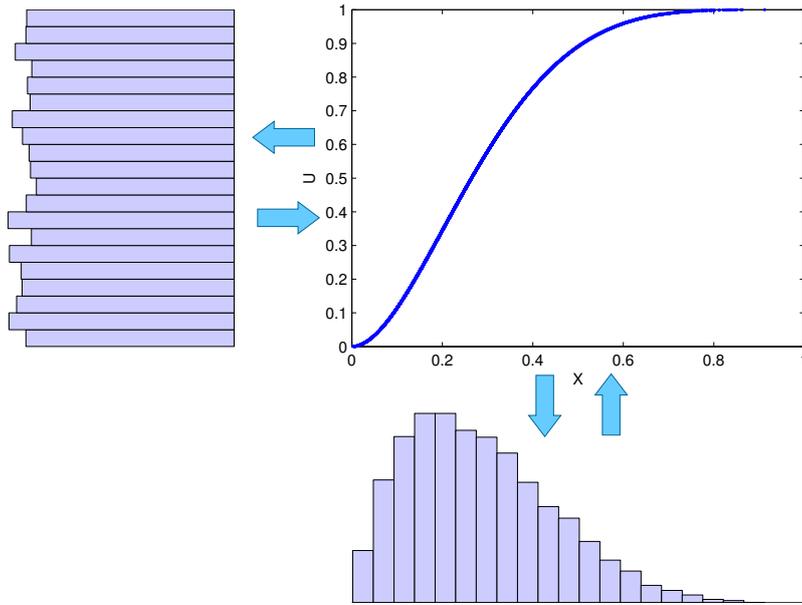


Figura 3.1: Esquema de la transformación integral de probabilidad.

Una función cópula es una distribución multivariante cuyas distribuciones marginales son uniformes, $[0, 1]$. Sea (X_1, X_2, \dots, X_d) , un vector de variables aleatorias continuas, con sus respectivas funciones de distribución de probabilidad, (Probability Density Function (PDF)), continuas, es decir $F_i(x) = P(X_i \leq x)$. Entonces, por definición, para toda variable X_i se cumple lo siguiente:

$$X_i = F_i^{-1}(U_i) \Leftrightarrow U_i = F_i(X_i) \tag{3.1}$$

Donde U_i es una variable aleatoria uniforme, con valores entre 0 y 1. La Transformación Integral de Probabilidad es más conocida en inglés como Probability Integral Transform (PIT). Un esquema de la transformación es mostrado en la figura 3.1, [71]. Por tanto, para el vector (X_1, X_2, \dots, X_d) se obtiene la siguiente expresión:

$$(U_1, U_2, \dots, U_d) = (F_1(X_1), F_2(X_2), \dots, F_d(X_d)) \tag{3.2}$$

Para el vector de variables aleatorias (X_1, X_2, \dots, X_d) , la función cópula se define como:

$$C(u_1, u_2, \dots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d) \tag{3.3}$$

Donde F es una distribución multivariante con sus respectivas funciones de distribución de probabilidad (en inglés Cumulative Distribution Function (CDF)), $F_1(X_1), F_2(X_2), \dots, F_d(X_d)$. Según el teorema de Sklar, [85], existe una función cópula C , para toda variable aleatoria x en el dominio de F , que contiene la información necesaria para

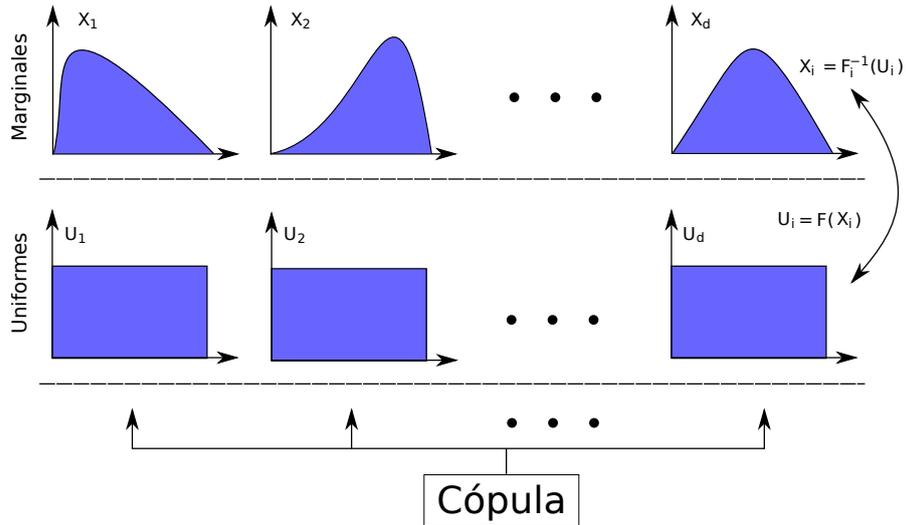


Figura 3.2: Modelado mediante funciones cópula.

modelar la dependencia entre las variables aleatorias (X_1, X_2, \dots, X_d) . Las funciones $F_i(x_i)$ contienen los respectivos modelos de las funciones de distribución de cada variable aleatoria. Por tanto, se puede reescribir la ecuación (3.3) de la siguiente forma:

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \quad (3.4)$$

De las ecuaciones (3.1) y (3.4), se tiene que:

$$C(u_1, u_2, \dots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_d^{-1}(u_d)) \quad (3.5)$$

Por tanto, una función cópula se define a partir de las funciones marginales de cada una de las variables F_i , y por la estructura de dependencia presente en la función cópula C . Un esquema de la formulación puede ser visto en la figura 3.2.

3.3 FAMILIAS DE FUNCIONES CÓPULA

Las dos familias de funciones cópula más usadas y aplicadas son las elípticas y las arquimedianas, las cuales son explicadas a continuación.

3.3.1 Cópulas elípticas

Las funciones cópula elípticas son las funciones cópula de distribuciones de perfil elíptico, como la normal y la t-student multivariantes. Una descripción más detallada se puede encontrar en [32, 68].

Este tipo de cópula se emplean frecuentemente en finanzas y gestión del riesgo por su simple implementación. Otra de sus ventajas es

la posibilidad de obtener funciones de probabilidad condicionada a otra variable, lo que es muy útil para elaborar predicciones [103].

3.3.1.1 Cópula normal

Esta cópula se deduce de la distribución Normal multivariante en \mathbb{R}^d . La formulación para la cópula Normal, está dada por:

$$C_{\Sigma}(u_1, u_2, \dots, u_d) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_d)) \quad (3.6)$$

Donde Φ^{-1} es la inversa de la CDF de la distribución normal, y Φ_{Σ} es la CDF de la distribución Normal multivariante, con media cero y matriz de covarianza Σ .

Un esquema de este modelo se presenta la figura 3.3.

3.3.1.2 Cópula t

Así como la cópula normal está definida por la distribución normal multivariante, la distribución t de Student multivariada da origen a la cópula t. La formulación de la cópula t viene dada por:

$$C_{\Sigma, \nu}(u_1, u_2, \dots, u_d) = t_{\Sigma, \nu}(t^{-1}(u_1), t^{-1}(u_2), \dots, t^{-1}(u_d)) \quad (3.7)$$

Donde el parámetro ν representa el número de grados de libertad, t^{-1} es la inversa de la CDF, de la distribución t Student y $t_{\Sigma, \nu}$ es la CDF de la distribución t Student multivariante parametrizada. Como en el caso anterior, Σ es la matriz de covarianza

3.3.2 Cópulas arquimedianas

Una cópula arquimediana es construida a través de una función denominada como *generador* y representada por φ . Se describe de la siguiente forma:

$$C(u_1, \dots, u_d) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)) \quad (3.8)$$

Donde φ^{-1} es la inversa del generador φ . Para que la ecuación (3.8) cumpla la definición de una función cópula, el generador φ tiene que ser una función completamente monótona. El generador determina de forma única la función cópula, [103]. En la práctica, las funciones cópula arquimedianas están extendidas porque permiten representar la dependencia, en el modelo multivariante, con un solo parámetro denominado α_{Λ} . En la tabla 3.1, son mostrados algunos de los generadores más usados para las funciones cópula arquimedianas, [103].

Una descripción más detallada de las funciones cópula arquimedianas y sus parámetros se puede encontrar en [68].

Tabla 3.1: Algunos generadores para las funciones cópula arquimedianas.

Familia	Parámetro	Generador $\varphi(l)$	Inverso del generador $\varphi^{-1}(s)$
Clayton	$\alpha_A \geq 0$	$l^{-\alpha_A} - 1$	$(1 + s)^{-1/\alpha_A}$
Frank	$\alpha_A \geq 0$	$\ln\left(\frac{e^{\alpha_A l} - 1}{e^{\alpha_A} - 1}\right)$	$\alpha_A^{-1} \ln(1 + e^s(e^{\alpha_A} - 1))$
Gumbell	$\alpha_A \geq 1$	$(-\ln l)^{\alpha_A}$	$\exp(-s^{-1/\alpha_A})$

3.4 ESTIMACIÓN DE LOS PARÁMETROS PARA LAS FUNCIONES CÓPULA

Para toda función cópula, independientemente de su familia, es necesario estimar los parámetros que la definen. Estos parámetros pueden ser estimados principalmente a partir de dos métodos: inferencia paramétrica y no paramétrica. El método totalmente paramétrico, a su vez, puede ser dividido en dos enfoques. En el primero se maximiza una función de probabilidad, donde las variables representan los parámetros de las funciones de densidad y la estructura de dependencia de la función cópula. Este enfoque se conoce como el método de máxima verosimilitud, en inglés Maximum Likelihood (ML). Cuando la dimensión del problema es grande, este método precisa de mucho tiempo de cálculo.

Como alternativa a este método hay uno más simple, pero menos exacto, denominado Método de Inferencia de Funciones para las Marginales, conocido en inglés como Inference Function for Margins (IFM), que consta de dos pasos. En el primero se determinan los parámetros de las funciones de densidad de forma individual, reduciendo así la complejidad del problema. Una vez que se han obtenido los parámetros óptimos de cada una de las marginales, se introducen en la función cópula, para así encontrar el parámetro óptimo que determina la dependencia de las variables.

A continuación se formulan algunos de los métodos que se emplean frecuentemente para estimar los parámetros que definen una función cópula. Los parámetros a estimar dependen de la función cópula escogida, y pueden ser estimados de formas alternativas. Una descripción más detallada de las técnicas que se emplean para estimar los parámetros de las funciones cópula, se puede encontrar en [68].

3.4.1 Método de máxima verosimilitud

Este método de estimación de parámetros, conocido en inglés como Maximum Likelihood (ML), puede ser aplicado a cualquier familia de distribución de probabilidad que tenga un cierto número de paráme-

tros. El conjunto de parámetros de la distribución de probabilidad es introducido en una función de optimización, donde se maximiza la probabilidad de todos los datos observados.

Sea un grupo de n observaciones independientes de una distribución multivariante, $\{(X_{i,1}, \dots, X_{i,d})^T : i = 1, \dots, n\}$. Supóngase que la distribución multivariante está definida por un número de marginales, cada una con su CDF F_j y su PDF f_j donde $j = 1, \dots, d$, y una cópula definida con densidad c , donde β es el vector que contiene los parámetros de las marginales y α el vector que contiene los parámetros de la cópula. El vector de parámetros a estimar se define como $\theta = (\beta^T, \alpha^T)^T$, y la función de verosimilitud como, [103]:

$$L(\theta) = \sum_{i=1}^n \ln c\{F_1(X_{i,1}; \beta), \dots, F_d(X_{i,d}; \beta); \alpha\} + \sum_{i=1}^n \sum_{j=1}^d \ln f_i(X_{i,j}; \beta) \quad (3.9)$$

Mediante el método ML se obtiene el valor de $\hat{\theta}_{ML}$ que maximiza el valor de la ecuación (3.9).

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} L(\theta) \quad (3.10)$$

3.4.2 Método de inferencia de funciones para las marginales

Este método, conocido en inglés como Inference Function for Margins (IFM), consta de dos pasos, [103]:

Primero se determinan los parámetros de las distribuciones marginales por separado de los parámetros de la cópula.

$$\hat{\beta}_{IFM} = \arg \max_{\beta} \sum_{i=1}^n \sum_{j=1}^d \ln f_j(X_{i,j}; \beta_j) \quad (3.11)$$

Cuando cada marginal F_i tiene su propio conjunto de parámetros, se define el vector $\beta_i = (\beta_1^T, \dots, \beta_d^T)^T$ que contiene los parámetros de la función marginal. Para este caso, el primer paso consiste en estimar por medio del método ML, los parámetros de cada marginal al maximizar la expresión siguiente:

$$\hat{\beta}_j_{IFM} = \arg \max_{\beta_j} \sum_{i=1}^n \ln f(X_{i,j}; \beta) \quad (3.12)$$

A continuación, a partir de los parámetros estimados de las funciones marginales, se calculan los parámetros de la función cópula, maximizando la siguiente expresión:

$$\hat{\alpha}_{\text{IFM}} = \arg \max_{\alpha} \sum_{i=1}^n \ln c (F_1(X_{i,1}; \hat{\beta}_{\text{IFM}}), \dots, F_d(X_{i,d}; \hat{\beta}_{\text{IFM}}); \alpha) \quad (3.13)$$

3.4.3 Método de la máxima verosimilitud canónica

Este método, conocido en inglés como Canonical Maximum Likelihood (CML), se diferencia del método IFM porque no hace ninguna presunción sobre las distribuciones marginales. El proceso se realiza en dos pasos, [103]:

Primero, partiendo del conjunto de observaciones $\{(X_{i,1}, \dots, X_{i,d})^T : i = 1, \dots, n\}$, se transforman las variables aleatorias utilizando la PIT, a partir de la CDF empírica de las mismas, en variables aleatorias uniformes, $(U_{i,1}, \dots, U_{i,d})^T$. En segundo lugar se calculan los parámetros de la cópula, maximizando la siguiente expresión:

$$\hat{\alpha}_{\text{CML}} = \arg \max_{\alpha} \sum_{i=1}^n \ln c (U_{i,1}, \dots, U_{i,d}; \alpha) \quad (3.14)$$

3.5 MODELO DE LA PRODUCCIÓN EÓLICA, FUNCIONES CÓPULA

De la literatura científica se puede deducir que una de las funciones cópula más usadas para modelos multivariantes de gran dimensión es la normal, [39, 71, 72, 97]. Esta función cópula permite modelar de forma simple problemas multivariantes de gran dimensión, a diferencia de la familia de funciones cópula arquimedianas, donde se limita usualmente a modelos bivariantes. Por estas razones se utilizará la función cópula normal para desarrollar el modelo estocástico para obtener las trayectorias de la producción eólica futura. Un esquema de esta modelo puede ser visto en la figura 3.3. Otra de las ventajas de la cópula Normal es la posibilidad de obtener funciones condicionadas a partir de variables de entrada. Esta característica particular será utilizada para modelar la producción eólica.

Para realizar un modelo multivariante utilizando la teoría de funciones cópula, en el caso de un modelo temporal de un parque (donde las variables están dadas por la producción eólica a diferentes horizontes de tiempo), se parte del análisis de los datos de un parque eólico, capítulo 4. Una vez definidas las características de las variables aleatorias se procederá a estimar una función marginal que represente el comportamiento de la producción eólica. En los capítulos 5, 6 y 7 se analizarán y evaluarán diferentes opciones para las funcio-

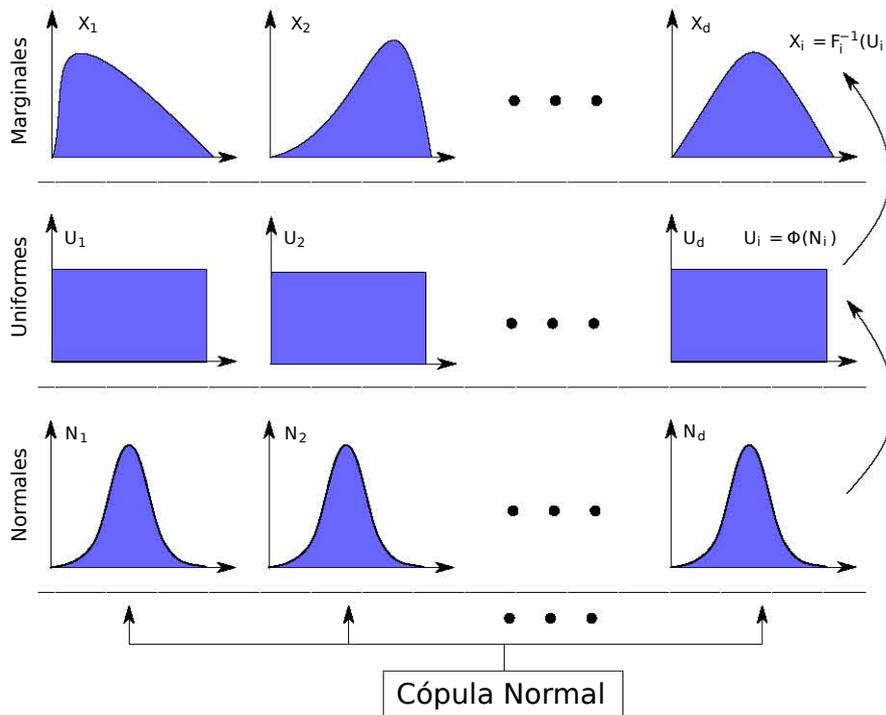


Figura 3.3: Modelo de la cópula normal

nes marginales que modelarán el comportamiento de la producción eólica.

En el capítulo 8 se analizarán diferentes formas para determinar la estructura de dependencia entre las variables aleatorias: para un modelo temporal de un parque y para un modelo espacial-temporal de varios parques. Las variables serán la predicción eólica a diferentes horizontes por cada parque. Puesto que se utilizará la cópula normal, la estructura de dependencia se puede obtener desde el dominio uniforme de las variables aleatorias aplicando la PIT a las mismas; con esto se elimina la influencia de las funciones marginales en la estimación, [71]. Este método se basa en la formulación del Canonical Maximum Likelihood (CML), pero se estimarán los parámetros de forma adaptativa con el objetivo de modelar el comportamiento estacional y las características de la producción eólica. Por esta razón no es posible obtener la estructura de dependencia por medio de un proceso de optimización, tal como es planteado en los métodos descritos en la sección 3.4. Estos conceptos serán explicados con más detalle en el capítulo 8.

Una vez obtenidas las distribuciones marginales y la estructura de la dependencia, se completará el modelo con las funciones cópula, que será utilizado para obtener las trayectorias de producción eólica dada la predicción de potencia a futuro.

Resumen

En este capítulo son presentados los datos de un parque eólico, producciones y sus respectivas previsiones a diferentes horizontes de predicción. Con estos datos se pretende definir de forma general las relaciones y características que existen entre las predicciones y los errores en la predicción a corto plazo. Una vez definidas estas características se continuará en los capítulos próximos con la estimación de una función marginal genérica para el modelo multivariante usando las funciones cópula, para obtener las trayectorias de producción eólica dada la predicción de potencia a futuro.

4.1 INTRODUCCIÓN

Tanto el productor eólico como el OS necesitan tener estimaciones de la posible producción. La importancia de las predicciones radica en la ayuda que aportan para planificar y anticiparse a los eventos futuros que afectarán el sistema. Esta información ayuda a gestionar las reservas y a maximizar el uso de los recursos. Si estos objetivos se consiguen, también se maximizan los beneficios en la operación del sistema.

En las últimas décadas se ha dado un gran avance en las herramientas de predicción, en parte debido al incremento de la capacidad de los centros de cálculo. Actualmente es posible utilizar, analizar y operar grandes cantidades de datos y realizar estimaciones complejas en tiempos relativamente cortos, sin la necesidad de grandes inversiones. Por las razones antes mencionadas, las herramientas de predicción han podido introducir modelos cada vez más complejos con el objetivo de mejorar los resultados o para entregar información más completa, [37].

La información que entregan las herramientas de predicción eólica suele ser determinista, lo que quiere decir que se proporciona un número que representa la estimación de la producción a una hora determinada. Esta información no es suficiente si se requiere conocer el orden de los posibles desvíos de potencia y energía a futuro.

En la sección 3.5, se describió de forma general el procedimiento que se empleará para conseguir el modelo multivariante de la producción eólica. En este capítulo se hará un análisis de las producciones y sus respectivas previsiones a diferentes horizontes de predicción, para un parque eólico determinado. Puesto que es necesario compren-

der y determinar las características de las variables, una vez definidas estas características, se continuará con la estimación de las funciones marginales a partir de este análisis.

Tal y como es presentado en [37], la incertidumbre de la producción eólica, es desarrollada a partir del modelo del error en la producción. La predicción, $\hat{p}_{t+k|t}$, es un valor determinista puntual y los errores, un valor probabilista, $\hat{f}_{e,k}$. Entonces, la incertidumbre de la producción eólica, $\hat{f}_{p,k}$, se consigue sumando a la PDF de los errores el valor de la predicción:

$$\hat{f}_{p,k} = \hat{p}_{t+k|t} + \hat{f}_{e,k} \quad (4.1)$$

Obteniendo el modelo de la función marginal de los errores, se podrá simular el comportamiento de la producción eólica. Esta afirmación es abordada nuevamente en los siguientes capítulos, donde se estima dicha marginal. Sin embargo es importante mencionarlo ahora, ya que se hace un mayor énfasis en las características de los errores en la predicción.

4.2 CASO DE ESTUDIO

Se analizarán datos pertenecientes a un parque eólico, donde se dispone de la producción de potencia eólica, hora a hora, y la previsión de esta, hecha k horas antes, durante dos años consecutivos. Por motivos de confidencialidad, solo es posible afirmar que el parque se encuentra localizado en España y que sus características corresponden a las de un parque típico.

Los datos de la predicción de potencia del parque eólico son obtenidos con un sistema de predicción a corto plazo. Esta herramienta obtiene la predicción por medio de modelos estadísticos que utilizan como datos de entrada predicciones meteorológicas de la velocidad y dirección del viento, además de las telemidas de potencia generadas del parque. Una descripción más detallada de los algoritmos de esta herramienta de predicción se encuentra en [88, 90]. Estos algoritmos también son utilizados en la herramienta de predicción SIPREOLICO, utilizada por el Operador del Sistema (OS) español.

Del conjunto de datos, producciones y sus respectivas predicciones, se estimará una distribución de probabilidad que representa la producción eólica.

Debido a interrupciones en los datos de entrada o en el funcionamiento del parque, no todos los datos son válidos, lo cual provoca salidas erróneas en la herramienta de predicción, por lo que son descartados algunos registros de los datos originales.

Una vez eliminados los registros erróneos, los datos se dividen en dos conjuntos. El primer conjunto se denomina datos de entrenamiento y corresponde a la primera mitad de todos los datos, producciones

y sus respectivas predicciones. Este conjunto está orientado a ser utilizado en el proceso de estimación de las funciones marginales. En el resto del documento se hará referencia a este conjunto como *datos de entrenamiento*.

El segundo conjunto, se denomina en el resto del documento como *datos de evaluación*. Este conjunto corresponde a la segunda mitad de los datos del parque. Una vez estimados los modelos de las funciones marginales, estos datos son utilizados para verificar si el modelo simula el comportamiento real de las variables. El proceso de evaluación se explica con más detalle en la sección 4.3.

La denominación *datos históricos* es utilizada para referirse al conjunto de datos disponibles en el presente. Un productor eólico o el OS, mantienen un registro del comportamiento de los parques, que consta de las medidas y las respectivas predicciones. Este conjunto es parecido a los datos de entrenamiento, con la diferencia de que los datos históricos crecen a medida que pasa el tiempo.

A continuación se presentan algunos conceptos básicos sobre el funcionamiento y producción de la energía eólica, antes de comenzar con el análisis de los datos.

4.2.1 Producción de potencia eólica y su incertidumbre

La relación entre la velocidad del viento y potencia producida en un generador eólico está dada por la siguiente ecuación:

$$P(v) = \frac{1}{2} \rho_w A v^3 C_{P,e}(v) \quad (4.2)$$

Donde A es el área que abarca las palas al girar en $[m^2]$, v es la velocidad del viento en $[m/s]$, ρ_w es la densidad del viento en $[Kg/m^3]$ y $C_{P,e}(\cdot)$ es el coeficiente de potencia eléctrica que relaciona el rendimiento mecánico y eléctrico del aerogenerador, [81]. Los constructores de generadores eólicos utilizan túneles de viento para mantener las condiciones constantes y de esta forma poder caracterizar sus productos. De esta forma se obtiene la curva característica del aerogenerador 4.1a [88]. Esta curva se aproxima al modelo teórico expresado en la ecuación (4.2). Sin embargo, en la figura 4.1b [88], se puede observar una familia de puntos que corresponde a la velocidad del viento medida frente a la potencia de salida de un generador eólico ubicado en un parque. En esta gráfica se puede ver que para una misma velocidad de viento, existen varios valores de potencia producida. Con este ejemplo se podría asegurar que, aún contando con una predicción perfecta de la velocidad del viento, y utilizando la ecuación (4.2) para obtener la estimación de la potencia, siempre se tendrían errores o desvíos, debido a que hay factores como la energía del viento y la dirección del mismo que pueden cambiar continuamente, afec-

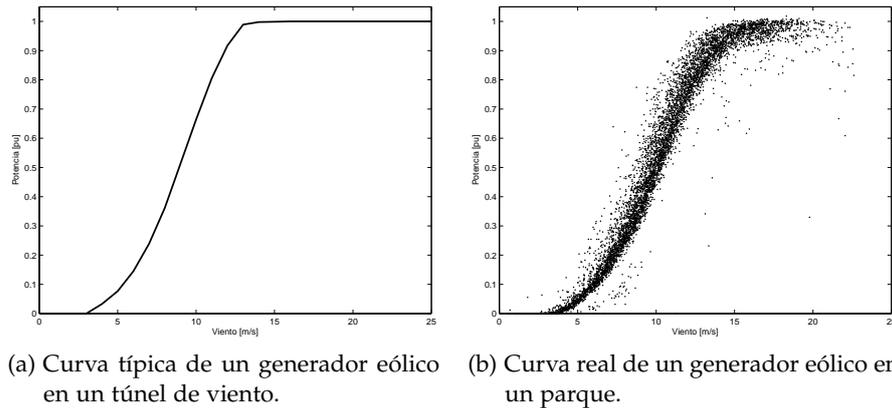


Figura 4.1: Relación de la velocidad del viento y la potencia producida por un generador eólico.

tando la potencia de salida de los aerogeneradores. Adicionalmente, considerando los fallos en los aerogeneradores, las salidas por mantenimiento o los tiempos en orientarse hacia el viento se hace aún más compleja la tarea de realizar estimaciones para parques eólicos y zonas donde se concentren un grupo de ellos.

En [59, 71, 97, 99] y otros, es utilizada la ecuación (4.2) y predicciones de viento para estimar la producción eólica. Esta aproximación es bastante imprecisa si se compara con el caso real de un aerogenerador, tal como se puede ver en la figura 4.1b. Es mucho más conveniente estimar los modelos considerando las producciones y predicciones del conjunto de datos históricos.

En trabajos como [26], se hace un análisis de la producción eólica y se modela el error en la predicción en función del horizonte de predicción y la potencia instalada. En [52] es usada una aproximación empírica para calcular el error en la predicción, como una proporción de la producción eólica y dependiendo solamente del horizonte de predicción. En [76], los errores en la predicción son modelados como una función no paramétrica, considerando el horizonte de predicción el nivel de la potencia prevista.

En el presente análisis se buscará determinar la relación que existe entre el horizonte de predicción, el nivel de potencia prevista y los errores de la predicción. De este modo se puede formular un modelo para la función marginal de los errores que sea fiable y represente sus características.

4.2.2 Relación entre la predicción de potencia eólica y el error

En la figura 4.2 se observa un ejemplo del problema de la incertidumbre en la predicción eólica. La línea vertical situada en el cero del eje de abscisas, representa el tiempo presente t , hacia la izquierda se encuentran los datos históricos del parque, producciones y sus

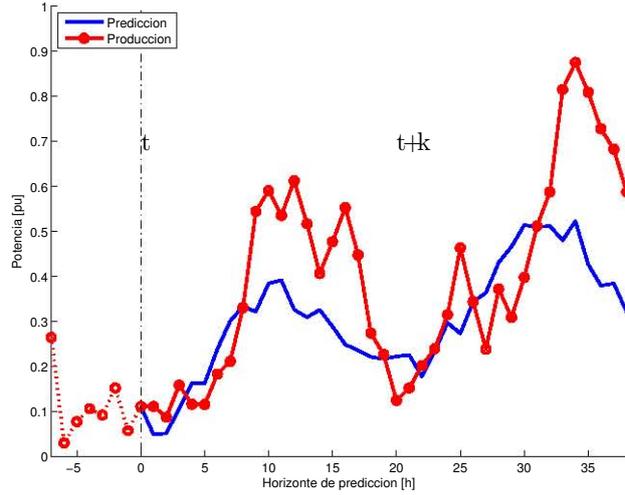


Figura 4.2: Ejemplo de la predicción ($\hat{p}_{t+k|t}$), y la producción (p_{t+k}), de la potencia eólica de un parque, donde $t+k|t$ se refiere a la predicción hecha para la hora $t+k$ en el tiempo t , donde $k = [1, 2, \dots, h]$

respectivas predicciones (las cuales no son dibujadas); a la derecha de la línea vertical, se encuentra un ejemplo de la previsión a futuro, representada con una línea continua; y la producción real del parque, representada con la línea continua con círculos. En esta figura se pueden observar los desvíos entre la predicción y la producción real, en términos de potencia para cada hora y, adicionalmente, el área entre las dos curvas, representa los desvíos en términos de energía. Los datos utilizados para el ejemplo de la figura 4.2 corresponden a datos reales, tomados del caso de estudio descrito en la sección 4.2.

Los errores en la predicción son calculados como la diferencia entre la potencia producida y las predicciones hechas con anterioridad para cada período de producción.

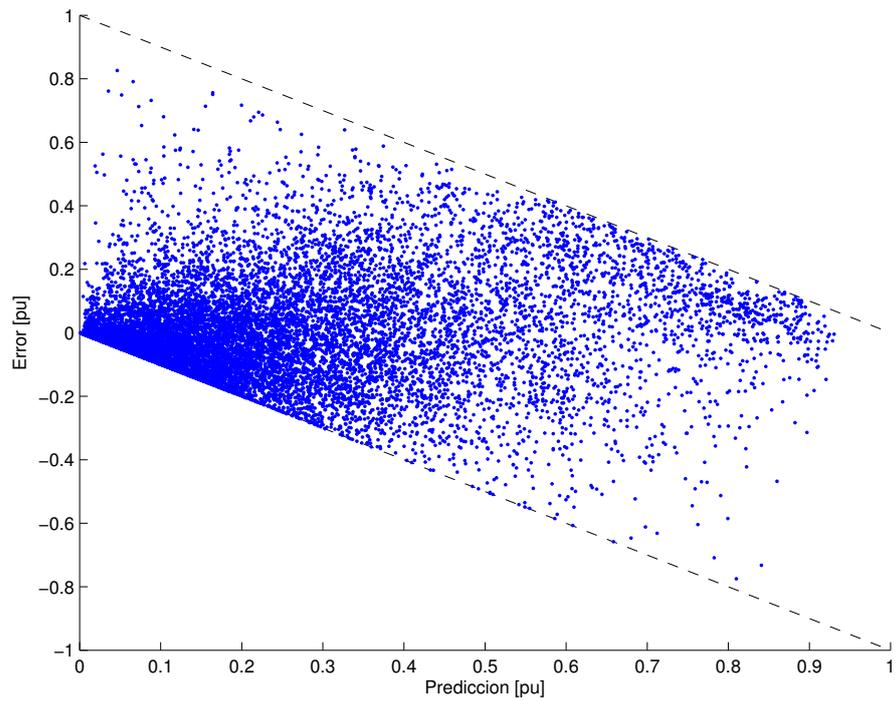
$$e_{t+k|t} = p_{t+k} - \hat{p}_{t+k|t} \quad \forall t, k \quad (4.3)$$

$$\mathbf{e}_t = [e_{t+1|t}, e_{t+2|t}, \dots, e_{t+h|t}]^T \quad \forall t, k \quad (4.4)$$

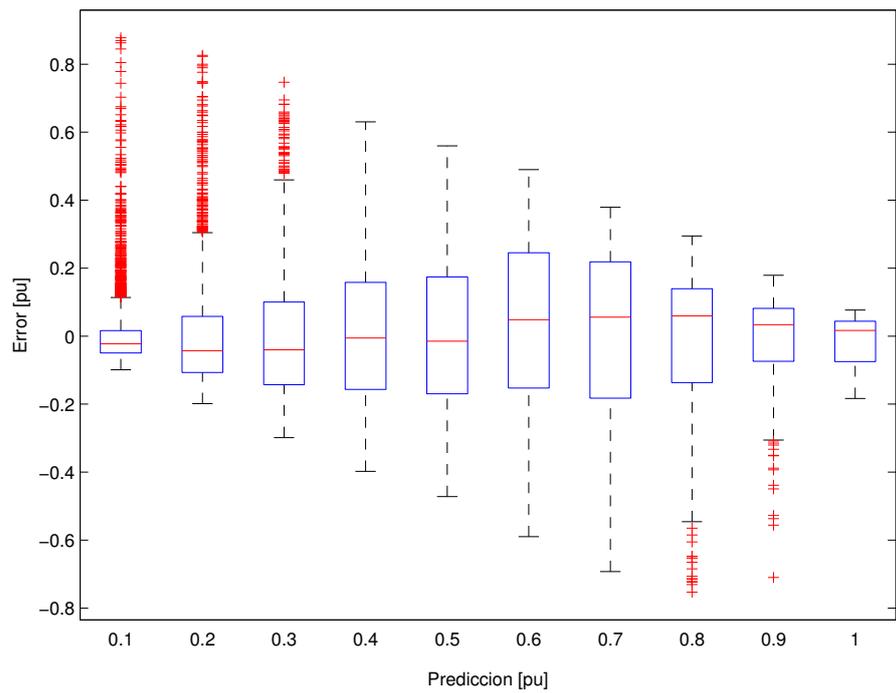
Donde la notación $t+k|t$, se refiere a las predicciones hechas para la hora $t+k$ en la hora t ; y $k = [1, 2, \dots, h]$. Los errores se calcularon para todos los datos del caso de estudio.

En la figura 4.3, se utiliza una nube de puntos, figura 4.3a y un diagrama de caja, figura 4.3b, para mostrar la relación que existe entre la predicción y el error. Para este caso en particular son tomados todos los datos del caso de estudio, con horizonte de predicción $k = 30$ h.

En la figura 4.3b, se puede identificar la simetría y la forma de la distribución de los datos. Para valores de predicción bajos y altos se



(a) Nube de puntos.



(b) Diagrama de caja.

Figura 4.3: Ejemplo de la relación entre la predicción de potencia eólica y el error, horizonte de predicción $k = 30$ h.

puede observar que los errores están concentrados y que tienen una forma asimétrica. Para predicciones cercanas al 50 % de la predicción, los errores tienen una distribución casi simétrica y están más dispersos. Estas características se pueden asumir como generales para todos los horizontes de predicción, tal y como se ve en la figura 4.4b, donde, de igual forma a la gráfica anterior, son tomados todos los datos del caso de estudio, pero en este caso el horizonte de predicción es $k = 5$ h. En esta figura se puede identificar que la forma en que se distribuyen los errores, respecto al nivel de potencia prevista, se mantiene casi constante. La diferencia radica en que la varianza de los errores del horizonte $k = 30$ h es mayor que la del horizonte $k = 5$ h. Esta relación se tratará con más detalle en la siguiente sección.

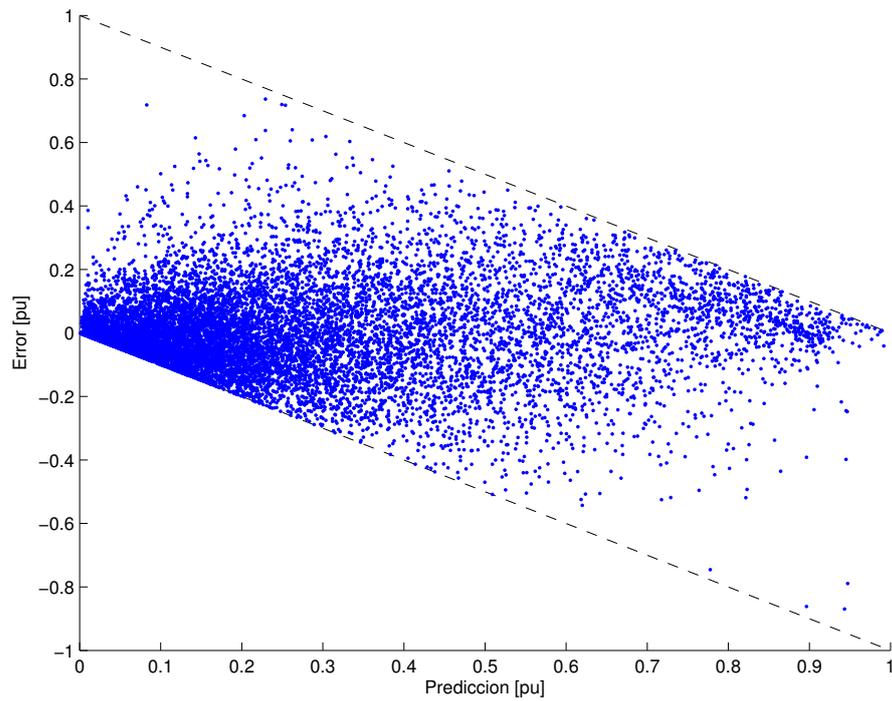
4.2.3 Relación entre el horizonte de predicción y los errores

En la figura 4.5, se dibuja la relación entre el horizonte de predicción y la magnitud de los errores en la predicción. Para este diagrama de caja son utilizados todos los datos del caso de estudio descrito al comienzo del capítulo. Para esta figura no se realizó ninguna discriminación por el nivel de potencia prevista. En la figura 4.5 se pueden observar, como características generales, que la varianza y la magnitud de los errores aumenta con el horizonte de predicción. Adicionalmente se puede asumir que la distribución de los errores, para cada horizonte de predicción, es aproximadamente simétrica.

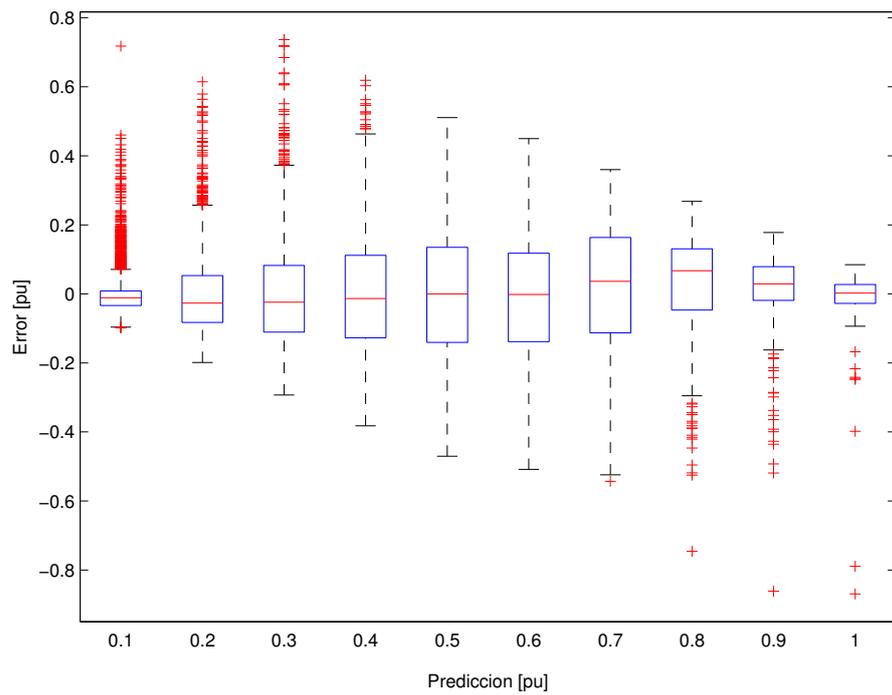
Como características generales para los errores en la predicción de potencia eólica, se ha observado que:

- La variabilidad aumenta con el horizonte de predicción k .
- La variabilidad depende del horizonte de predicción, k , y del valor de la predicción, $\hat{p}_{t+k|t}$.
- La magnitud de los errores aumenta con el horizonte de predicción k .
- Para predicciones bajas y altas, los errores se agrupan de forma asimétrica y están concentrados.
- Para predicciones cercanas al 0,5 pu, los errores son casi simétricos y están dispersos.

Con estas características generales sobre los errores de la predicción, se buscará un modelo univariante que represente el comportamiento de la variable. Antes de continuar, se definirán algunos conceptos necesarios para evaluar estos modelos.



(a) Nube de puntos.



(b) Diagrama de caja.

Figura 4.4: Ejemplo de la relación entre la predicción de potencia eólica y el error, horizonte de predicción $k = 5$ h.

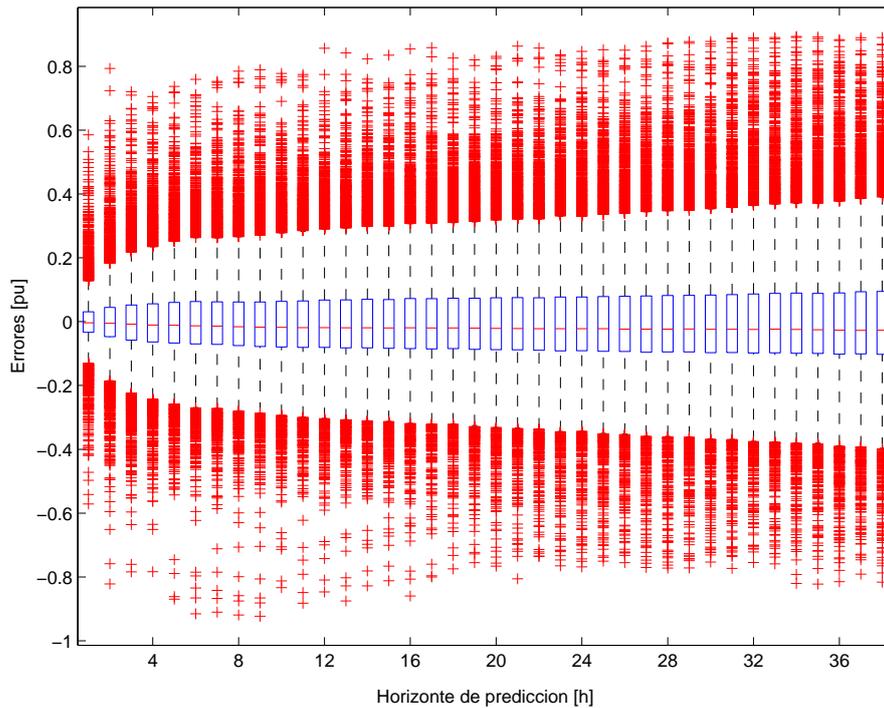


Figura 4.5: Relación de la predicción eólica y la producción usando un diagrama de cajas.

4.3 EVALUACIÓN DE LAS FUNCIONES MARGINALES ESTIMADAS

Tal y como se dijo en la sección 4.2, los datos del caso de estudio se separaron en dos conjuntos, el primer conjunto se denomina datos de entrenamiento y corresponde a la primera mitad de los datos disponibles. Este conjunto está orientado a la obtención de las funciones marginales de los errores. El segundo conjunto, que se denomina datos de evaluación, corresponde a la segunda mitad de los datos disponibles. Estos datos son utilizados para verificar si el modelo simula el comportamiento de los errores.

Por tanto, se busca un modelo que simule el comportamiento de los errores en la predicción, a partir del conjunto de datos de entrenamiento. Luego, el modelo obtenido se evaluará utilizando la segunda mitad de los datos. Sí el modelo obtenido simula el comportamiento del conjunto de datos de evaluación, quiere decir que el modelo simula de forma adecuada las características de los errores.

Debido a que las funciones marginales para el modelo de la producción eólica son condicionadas, no es posible utilizar los métodos convencionales para determinar los parámetros de las funciones cópula, descritos en la sección 3.4. Estos métodos son concebidos para modelos estacionarios, donde la función marginal que describe la variable es única.

Para determinar si el modelo simula de forma adecuada las variables, serán utilizados los siguientes métodos, conocidos en inglés

como Probability Integral Transform (PIT) y Continuous Ranked Probability Score (CRPS). Estos métodos son usados en la literatura especializada para evaluar y comparar este tipo de modelos [15, 37, 51]. Con la evaluación se podrá cuantificar y comparar el grado de ajuste de los modelos. De esta forma se podrá determinar de forma simple si el modelo replica el comportamiento de la variable real. A continuación serán explicados brevemente los métodos que se utilizan para evaluar y comparar los modelos de las marginales.

4.3.1 Transformada integral de probabilidad

Partiendo del conjunto de datos de entrenamiento, se estiman modelos para las funciones de densidad. Posteriormente cada modelo estimado, es evaluado con el conjunto de datos de evaluación. Para la evaluación, se aplica la inversa de la transformada integral de probabilidad, PIT, al conjunto de datos de evaluación, en este caso los errores de la predicción. Por tanto, cada uno de los errores será transformado al dominio uniforme utilizando la inversa de la función de distribución de probabilidad, de la siguiente forma:

$$F_e^{-1}(e_{t+k|t}) = U_{t+k|t} \quad \forall t, k \quad (4.5)$$

La variable uniforme $U_{t+k|t}$ esta comprendida entre los valores $[0, 1]$. Si el modelo es correcto, los valores se distribuirán uniformemente en todo el intervalo, de modo que si se obtiene un histograma, todas las alturas de las clases tendrán aproximadamente la misma altura. Un ejemplo de este proceso se ilustra en la figura 4.6. En el resto del documento, cuando se hable del *histograma PIT*, se hará referencia a este proceso de evaluación de la función marginal.

El grado de ajuste del histograma PIT, se puede evaluar comparando éste con una distribución uniforme ideal. Para cuantificar el grado de ajuste, se calcula la raíz del error cuadrático medio, más conocido en inglés como Root Mean Square Error (RMSE), entre la altura relativa de cada una de las clases del histograma PIT, y la altura relativa de una función uniforme ideal. Habitualmente se suele dividir el histograma en veinte clases ($\kappa = 20$), para su evaluación, [37].

$$\text{RMSE}_{\text{PIT}} = \sqrt{\frac{1}{\kappa} \sum_{i=1}^{\kappa} \left(\psi_i - \frac{1}{\kappa} \right)^2} \quad (4.6)$$

Donde ψ se refiere a la altura relativa de cada una de las clases del histograma PIT.

Este método alternativo de evaluación de las funciones marginales permite evaluar el ajuste de las funciones estimadas, [15].

En la figura 4.7, se muestran varios ejemplos para interpretar las formas que puede adquirir el histograma PIT. En el primer cuadro

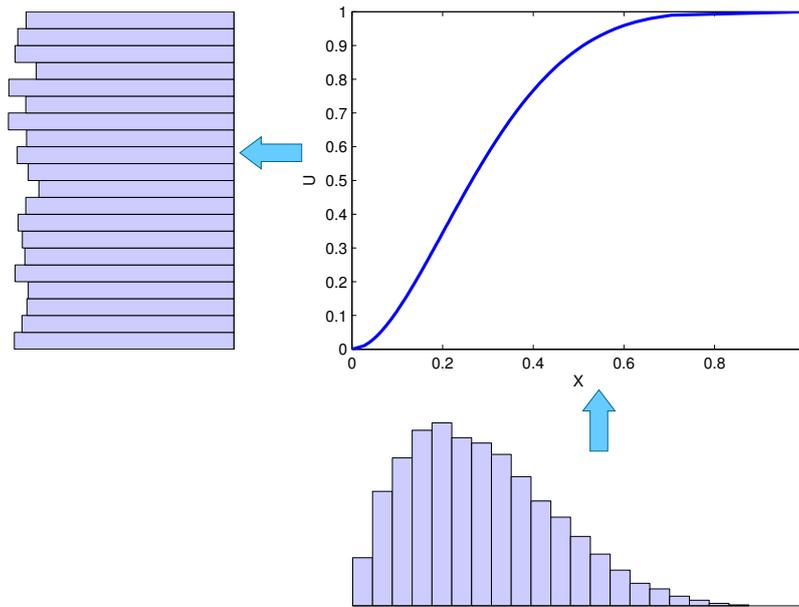


Figura 4.6: Esquema de evaluación de la función de distribución, $U = F(X)$.

de la columna izquierda se muestran la distribución real, y que se espera estimar (en este caso es una función normal con media cero y desviación estándar uno $N(0; 1)$) y la distribución prevista. De arriba a abajo las distribuciones previstas son, $N(0; 1)$, $N(0; 0,5)$, $N(0; 2)$ y $N(1; 1)$. En la columna de la derecha se representan los histogramas PIT obtenidos para cada uno de las funciones estimadas. Cada histograma PIT fue obtenido con 10000 puntos generados a partir de la distribución de destino, [15].

4.3.2 Continuous Ranked Probability Score (CRPS)

El CRPS, [71], es un método para evaluar la función probabilista de una variable continua. Este método es equivalente al Brier Score¹, [11], integra todos los valores posibles y es una generalización del Ranked Probability Score, [28], que se utiliza para evaluar las predicciones probabilistas sobre las variables ordinales. El CRPS es sensible a la distancia, es decir, es capaz de penalizar las predicciones que están muy lejos de la observación real. En esencia, el CRPS mide la diferencia entre la CDF de la predicción y la observación, figura 4.8. Para que el resultado sea sensible a la distancia de la observación, son calculados los errores cuadráticos entre la CDF de la predicción y la observación, [15].

Del conjunto de datos de evaluación $\{(e_{i,1}, \dots, e_{i,d})^T : i = 1, \dots, n\}$, y siendo $\hat{f}_{e(i,k)}$ la PDF estimada de los errores y $\hat{F}_{e(i,k)}$ su respec-

¹ El Brier Score mide la desviación cuadrática media entre las probabilidades predichas de un conjunto de eventos y sus observaciones. Una puntuación más baja representa una mayor precisión del modelo.

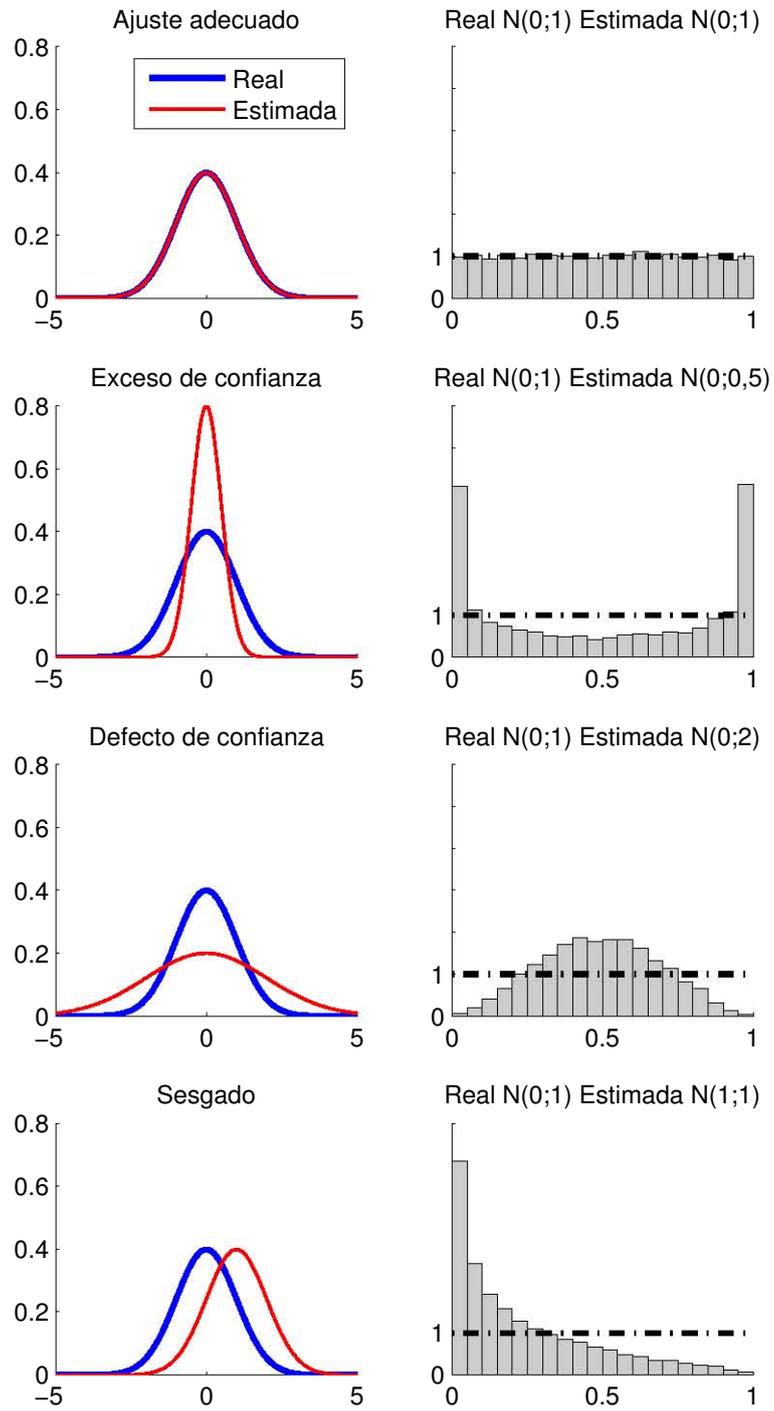
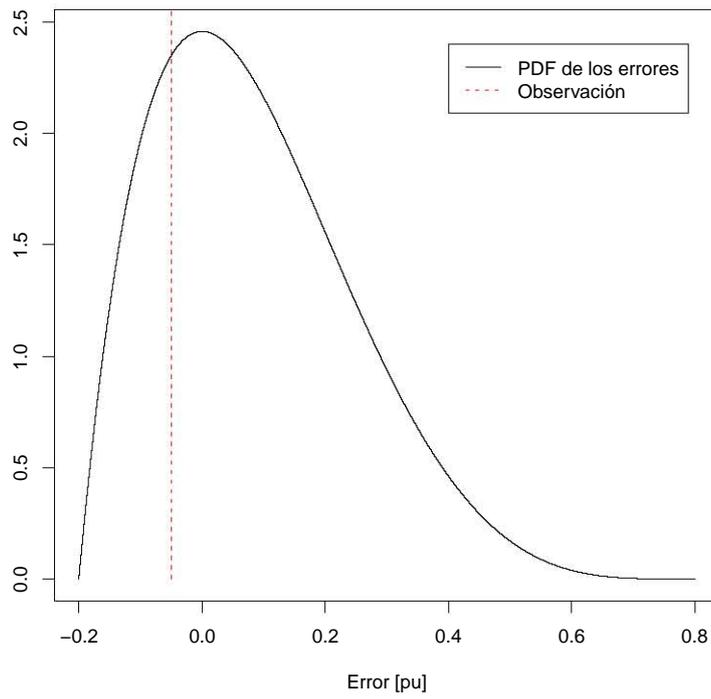
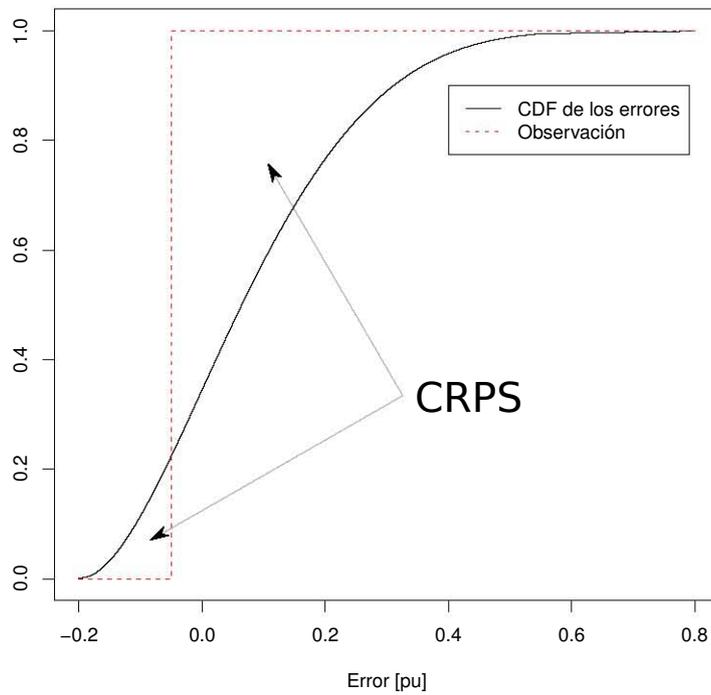


Figura 4.7: Algunas de las posibles formas que puede adquirir el histograma PIT.



(a) Función de densidad de probabilidad.



(b) Función de distribución de probabilidad estimada.

Figura 4.8: .

tiva CDF (para cada horizonte de predicción $k = 1, \dots, d$), el valor promedio del CRPS se calcula de la siguiente forma:

$$\overline{\text{CRPS}}_k = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} (\hat{F}_{e_{(i,k)}}(x) - H(x, e_{i,k}))^2 dx \quad (4.7)$$

Donde H es la función escalón de Heaviside:

$$H(x, e_{i,k}) = \begin{cases} 0, & x < e_{i,k} \\ 1, & x \geq e_{i,k} \end{cases} \quad (4.8)$$

En lo sucesivo, cuando se utilice el CRPS para evaluar las marginales, se presentará el valor promedio de las observaciones por cada horizonte de predicción.

En la figura 4.8a se muestra un ejemplo de una PDF estimada para el error en la predicción, $\hat{f}_{e,k}$, donde el valor del error observado es de $e = -0,05$ pu. En la figura 4.8b se muestra la respectiva CDF estimada del mismo ejemplo, $\hat{F}_{e,k}$. El área entre el error observado y la CDF estimada (mostrada por las dos flechas), es el valor que entrega la expresión (4.7) para una sola observación. Sí el modelo estimado representara perfectamente la variable se obtendría un valor cero. Para el caso particular de estimar la distribución de los errores de la predicción, no se podría tener un CRPS igual a cero, ya que se depende de la herramienta de previsión. El CRPS se utilizará para comprar diferentes modelos de la función marginal del error.

DISTRIBUCIÓN BETA COMO FUNCIÓN DE DENSIDAD

Resumen

La distribución beta, según la literatura, se ha utilizado para modelar la incertidumbre de la producción eólica. Con el objetivo de validar el modelo, en este capítulo se utilizarán los datos de un parque eólico para estimar los parámetros de la distribución beta, y así modelar la incertidumbre de la producción eólica. Para hacer el modelo más robusto, se exploran alternativas diferentes a las empleadas en la literatura especializada para estimar los parámetros de la distribución beta. Finalmente el modelo es evaluado utilizando datos reales de un parque eólico.

5.1 INTRODUCCIÓN

En este capítulo se presenta un método paramétrico para modelar la incertidumbre producción eólica utilizando la distribución beta. La idea principal del modelo es presentada en [9] y adaptada en [7, 50], donde son introducidos cambios a la propuesta inicial. Otros autores han utilizado aproximaciones semejantes para modelar la producción eólica, entre ellos [10, 31, 43, 96].

Comparando la distribución beta, por ejemplo, con la distribución normal utilizada por algunos autores [18, 75], ésta resulta mucho más adecuada para representar la incertidumbre de la producción eólica, ya que la distribución beta está acotada entre $[0, 1]$, tal como la producción eólica, por lo que puede ser utilizada directamente, sin necesidad de ser adaptada o escalada. Su comportamiento está condicionado por dos parámetros que pueden ser estimados del conjunto de datos de entrenamiento.

En este capítulo se pretende modelar la incertidumbre de la producción eólica mediante la distribución beta. Una vez obtenidas las funciones marginales univariantes, se podrá simular, a través de éstas y de las funciones cópula, el comportamiento de la producción eólica. Adicionalmente, para hacer el modelo más robusto, se exploran alternativas a las encontradas en la literatura especializada para estimar los parámetros.

Como se dijo antes, se utilizarán los datos del caso de estudio descritos en la sección 4.2. Con el conjunto de datos de entrenamiento se obtendrán los parámetros del modelo, y con el conjunto de datos de evaluación se verificará el modelo.

Primero se definirán algunos conceptos sobre la distribución beta necesarios para comprender el modelo de la producción eólica.

5.2 DEFINICIÓN DE LA DISTRIBUCIÓN BETA

En la teoría de la probabilidad y la estadística, la distribución beta, que pertenece a la familia de las distribuciones continuas, es definida en el intervalo $(0; 1)$ y parametrizada por dos números positivos que generalmente son denotados por a y b , y cuya PDF es la siguiente:

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \quad (5.1)$$

Donde B es la función beta:

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad (5.2)$$

El valor esperado y la varianza de una variable aleatoria X con distribución beta es:

$$E(X) = \frac{a}{a+b} \quad (5.3)$$

$$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)} \quad (5.4)$$

La distribución beta está condicionada por a y b . Estos parámetros condicionan la forma de la PDF, un ejemplo de esto se puede ver en la figura 5.1. Para modelar la producción eólica, son útiles todas las distribuciones beta, menos aquellas que son cóncavas hacia arriba. Por esta razón se deberá evitar estimar parámetros que generen PDF inadecuadas para modelar la incertidumbre de la producción eólica.

La figura 5.2 muestra ejemplos de la incertidumbre de la producción eólica, siendo modelada por la distribución beta. Para estos ejemplos han sido dibujadas diferentes curvas, y cada una de ellas fue condicionada al nivel de potencia prevista y a la varianza de los errores. Los valores y parámetros son listados en la tabla 5.1. En la sección 5.3 se explicará cómo estimar los parámetros a partir de los datos del conjunto de entrenamiento.

Sea $\hat{f}_{p,k}$ la PDF estimada para modelar la incertidumbre de la producción eólica, en este caso, utilizando la distribución beta. Entonces, según la ecuación (4.1), se puede definir el modelo de los errores de la predicción de la siguiente forma:

$$\hat{f}_{e,k} = \hat{f}_{p,k} - \hat{p}_{t+k|t} \quad (5.5)$$

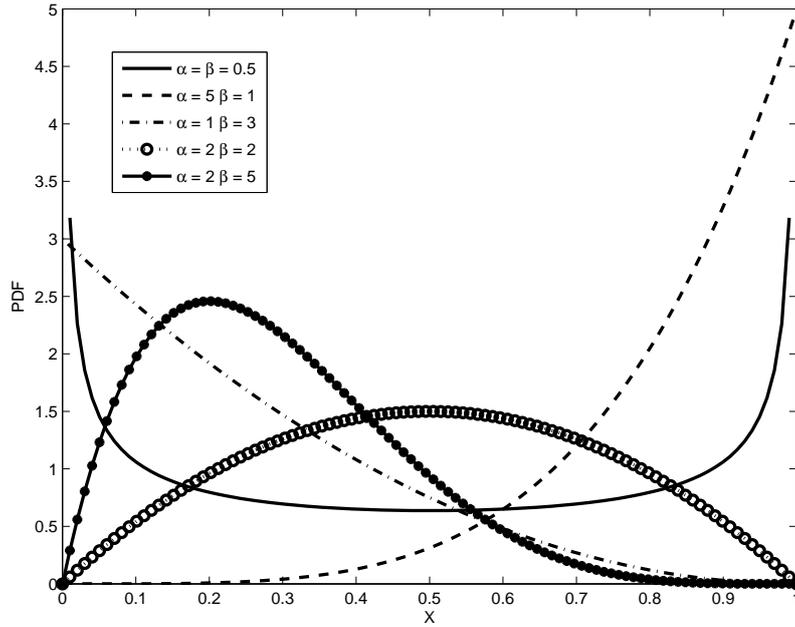


Figura 5.1: Diferentes formas que puede tomar la PDF de la distribución beta.

En la figura 5.3 son mostradas sólo tres curvas, para evitar que se solaparan las líneas, de la PDF de los errores utilizando la distribución beta. Igual que para el caso anterior, cada una de las curvas está condicionada al nivel de potencia prevista y a la varianza de los errores. Los valores y parámetros son presentados en la tabla 5.1.

A continuación se presenta un método para estimar los parámetros de la distribución beta utilizando los errores en la predicción. Tal y como fue descrito en el capítulo 4, el modelo considera el nivel de potencia prevista y el horizonte de predicción.

Los parámetros estimados de la distribución beta se denominarán $\hat{\alpha}$ y $\hat{\beta}$.

5.3 ESTIMACIÓN DE LOS PARÁMETROS

Los parámetros que condicionan la distribución beta, $\hat{\alpha}_{t,k}$ y $\hat{\beta}_{t,k}$, pueden ser definidos en función de la media y la varianza estimada, $\hat{\mu}_{t,k}$ y $\hat{\sigma}_{t,k}^2$ respectivamente, como es mostrado en las ecuaciones (5.6) y (5.7).

$$\hat{\alpha}_{t,k} = \hat{\mu}_{t,k} \left(\frac{\hat{\mu}_{t,k} (1 - \hat{\mu}_{t,k})}{\hat{\sigma}_{t,k}^2} - 1 \right) \quad \forall k \quad (5.6)$$

$$\hat{\beta}_{t,k} = (1 - \hat{\mu}_{t,k}) \left(\frac{\hat{\mu}_{t,k} (1 - \hat{\mu}_{t,k})}{\hat{\sigma}_{t,k}^2} - 1 \right) \quad \forall k \quad (5.7)$$

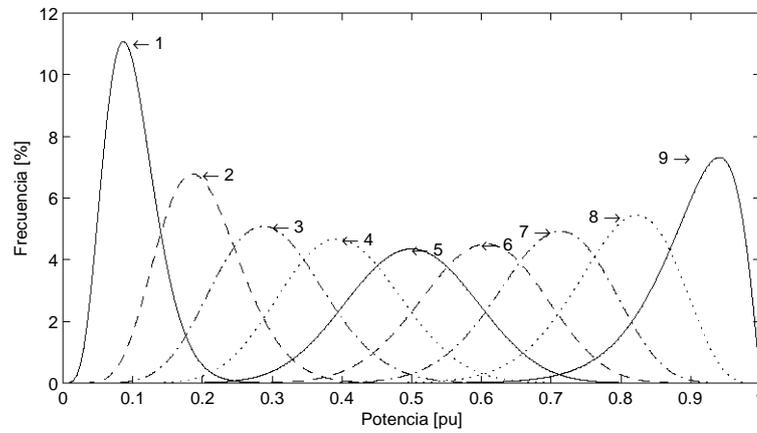


Figura 5.2: Ejemplo de la incertidumbre de la producción eólica usando la distribución beta. En este caso cada una de las distribuciones ha sido condicionada al nivel de potencia prevista y a la varianza de los errores, tabla 5.1.

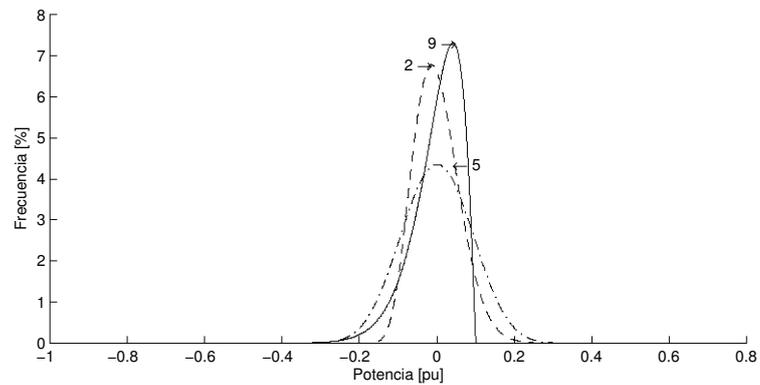


Figura 5.3: Ejemplo de la incertidumbre de los errores en la predicción eólica usando la distribución beta. En este caso cada una de las distribuciones ha sido condicionada al nivel de potencia prevista y a la varianza de los errores, tabla 5.1.

Tabla 5.1: Parámetros de la distribución beta utilizados en las figuras 5.2 y 5.3.

Número PDF	\hat{p}_{pu}	σ^2	a	b
1	0,1	$1,4 \times 10^{-3}$	6,3	57,0
2	0,2	$3,5 \times 10^{-3}$	8,9	35,8
3	0,3	$6,0 \times 10^{-3}$	10,2	23,8
4	0,4	$7,0 \times 10^{-3}$	13,3	20,0
5	0,5	$8,0 \times 10^{-3}$	15,1	15,1
6	0,6	$7,5 \times 10^{-3}$	18,6	12,4
7	0,7	$6,4 \times 10^{-3}$	22,3	9,5
8	0,8	$5,5 \times 10^{-3}$	22,5	5,6
9	0,9	$4,0 \times 10^{-3}$	19,4	2,2

El valor medio se considera como la esperanza de la producción futura, $\hat{\mu}_{t,k} = \hat{p}_{t+k|t}$, [9, 31]. De esta forma se introduce en el modelo el valor de la potencia prevista y el horizonte de predicción.

Para estimar la varianza se presentarán diferentes opciones, con el objetivo de comparar varios modelos para la función marginal de los errores de predicción. La evaluación será presentada al final del capítulo.

5.4 VARIANZA DE LOS ERRORES EN LA PREDICCIÓN

La estimación de la varianza es determinante para el modelo de la distribución beta [7] y, según el análisis del capítulo 4, está condicionada al nivel de potencia prevista y al horizonte, por tanto $\hat{\sigma}_{t,k}^2 = \sigma^2(k, \hat{p}_{t+k|t})$.

A continuación se presentan diferentes formas de estimar el valor de la varianza de los errores, ya que es un parámetro determinante para el modelo de incertidumbre de los errores en la predicción utilizando la distribución beta.

5.4.1 Regresión polinómica por intervalos

Como primera aproximación se presenta el método utilizado por [7, 43], donde la varianza de la distribución es estimada, para cada horizonte k , como un polinomio de cuarto orden en función de la potencia prevista, $\hat{\sigma}_{t,k}^2 = \sigma^2(k, \hat{p}_{t+k|t})$. Este polinomio es estimado por medio de la técnica de los mínimos cuadrados, usando los datos de entrenamiento descritos en la sección 4.2.

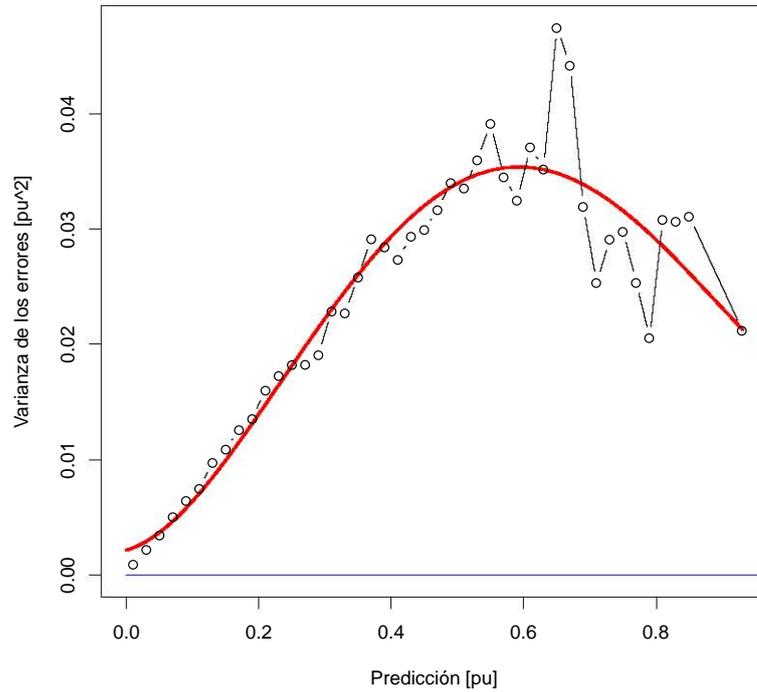


Figura 5.4: Ejemplo del polinomio de cuarto orden estimado para la varianza de los errores en función de la potencia prevista y el horizonte de predicción. En este caso $k = 15$ h.

El conjunto de datos de entrenamiento es subdividido en intervalos de 0,02 pu de acuerdo a la potencia prevista (el último intervalo es más grande para asegurar que haya datos suficientes para el cálculo, ya que el rango de operación del parque eólico usualmente no supera el 80% de su capacidad nominal, [7, 43]). De este modo se tienen subconjuntos de previsiones de potencia eólica con sus respectivos errores de la predicción. Para cada uno de los intervalos se calcula la varianza de los errores de la predicción y, finalmente, se estima el polinomio de cuarto orden que se ajusta a los datos por medio de la técnica de mínimos cuadrados. Este procedimiento es desarrollado y aplicado en [7, 43]. Un ejemplo de los resultados obtenidos se puede ver en la figura 5.4, para un horizonte de predicción $k = 15$ h. En esta figura se muestran los valores calculados de la varianza de los errores en cada intervalo, representados como un círculo. El polinomio estimado por medio de la técnica de mínimos cuadrados es dibujado como la línea continua sobre los círculos. La línea continua horizontal, marca el cero en el eje de las ordenadas.

Para los otros horizontes el modelo es obtenido de igual forma: estimando el polinomio de cuarto orden que se acerca más a la varianza de los errores de todos los intervalos.

En los resultados se encontró que la función estimada tomaba valores negativos para predicciones bajas, lo cual no es correcto según la definición de la varianza. Un ejemplo de esto se puede observar en la figura 5.5a, donde el horizonte de predicción es $k = 22$ h. Para solucionar este problema se estimó de igual forma la desviación estándar. Por tanto, se calculó la desviación estándar de cada intervalo y se estimó el polinomio de cuarto orden que se ajusta mejor a todos los puntos. En el caso de los errores de la predicción, la desviación estándar tiene una magnitud mayor comparada con la varianza, el nuevo polinomio estimado no toma valores negativos en los extremos. Un ejemplo se puede ver en la figura 5.5b, donde se solucionó el problema específico mostrado en la figura 5.5a.

El método utilizado por [7, 43], donde se estima la varianza o la desviación por medio del método de mínimos cuadrados, no garantiza, para todos los horizontes, que el polinomio de cuarto orden tome valores positivos.

Otro inconveniente de este tipo de modelo es que la estimación agrupa las observaciones por intervalos y no considera el número de datos dentro de cada intervalo. Por ejemplo, para predicciones bajas hay una gran concentración de observaciones, pero para predicciones altas hay pocas, tal como se ve en las figuras 4.3a y 4.4a. El problema radica en que los intervalos no tienen el mismo número de observaciones, pero la varianza o desviación estándar calculada por cada intervalo influye de igual forma en el polinomio estimado por medio del método de mínimos cuadrados.

Para que cada observación influya de igual manera en el modelo resultante, a continuación se estimará la varianza de los datos directamente, sin realizar cálculos parciales en intervalos, como se hizo antes. De esta forma se busca garantizar que cada dato influya de igual manera en el resultado final.

5.4.2 Regresión polinómica directamente de los datos para estimar la varianza en los errores

Siendo una variable aleatoria X con media $\mu = E(X)$, se define la varianza, $\text{Var}(X)$, como:

$$\text{Var}(X) = E[(X - \mu)^2] \quad (5.8)$$

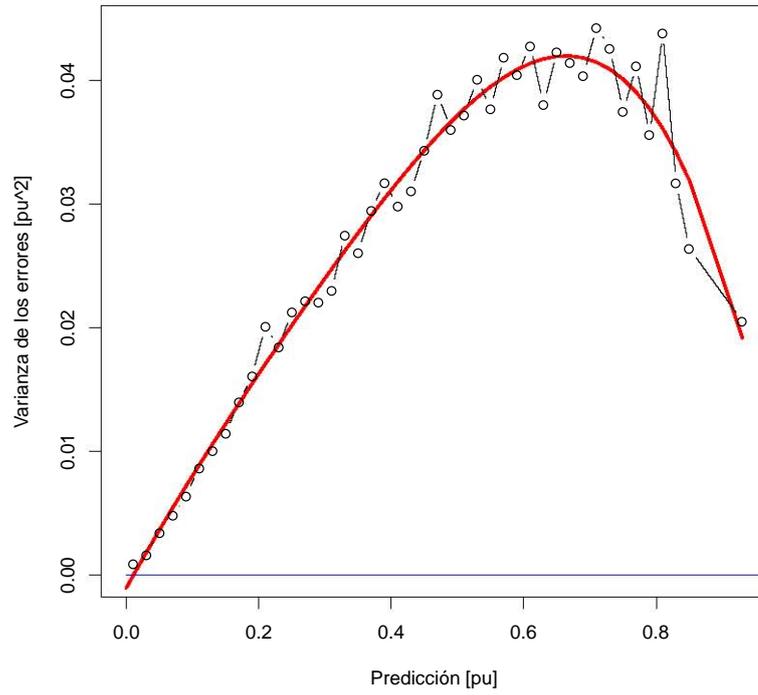
$$= E[(X^2 - 2X\mu + \mu^2)] \quad (5.9)$$

$$= E(X^2) - 2\mu E(X) + \mu^2 \quad (5.10)$$

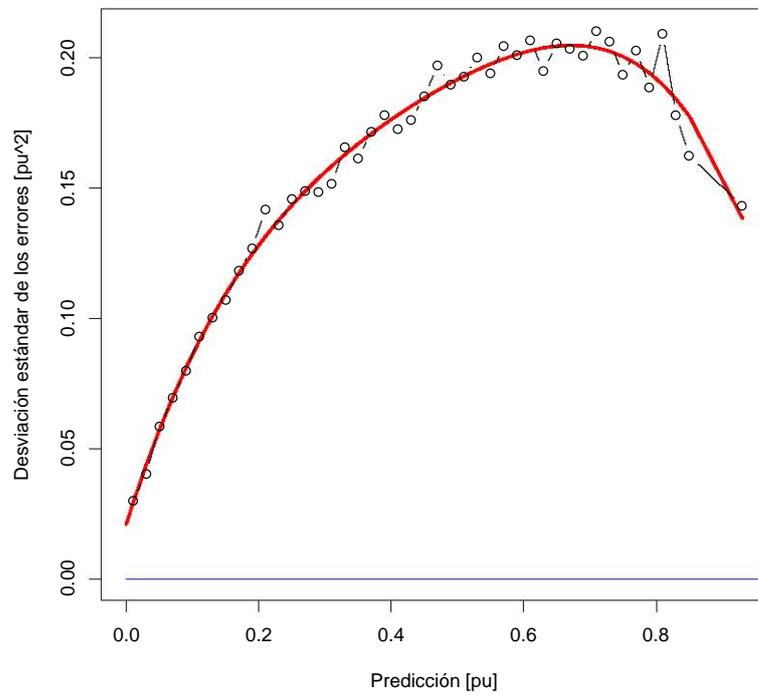
$$= E(X^2) - 2\mu^2 + \mu^2 \quad (5.11)$$

$$= E(X^2) - \mu^2 \quad (5.12)$$

Para obtener el modelo de la varianza en los errores, se asume que los errores de la predicción tienen media cero, $\mu_e = 0$, [7]. Esta



(a) Estimación de la varianza.



(b) Estimación de la desviación estándar.

Figura 5.5: Ejemplo del polinomio de cuarto orden estimado para modelar la varianza o la desviación estándar de los errores, $k = 22$ h.

aproximación se acepta de forma común en este tipo de análisis, [23]. Por tanto, la varianza de los errores se puede expresar a partir de la ecuación (5.12), como:

$$\text{Var}(e_{t+k|t}) = E(e_{t+k|t}^2) \quad (5.13)$$

Siendo (x_i, y_i) un par de observaciones e $i = 1, \dots, n$ el número de éstas. Se desea encontrar una función $\hat{f}(x_i) = y_i + \varepsilon_i$, tal que ε_i es una perturbación aleatoria con media cero.

La función estimada \hat{f} , es aquella que obtiene el valor mínimo de la siguiente expresión, [44]:

$$\text{RMSE}(\hat{f}) = \frac{\sum_{i=1}^n \varepsilon_i^2}{n} \quad (5.14)$$

donde:

$$\varepsilon_i = y_i - \hat{f}(x_i) \quad (5.15)$$

Se busca estimar una función que consiga representar el comportamiento de las observaciones. Por definición, [44], ningún modelo estimado se ajusta perfectamente al comportamiento de una variable aleatoria, ya que hay un error aleatorio asociado a cada observación, ε_i . Por tanto, el modelo estimado se aproxima localmente al promedio del valor que toma la variable aleatoria. Para el caso particular del modelo de la varianza de los errores, si se estima la media local de los errores al cuadrado se obtiene de forma indirecta el modelo de la varianza y, por tanto, de la ecuación (5.13), se define la varianza de la siguiente forma:

$$\text{Var}(e_{t+k|t}|\hat{p}_{t+k|t}) = E(e_{t+k|t}^2|\hat{p}_{t+k|t}) \quad (5.16)$$

Así, es posible modelar la varianza de los errores de la predicción, estimando un polinomio de cuarto orden por medio de la técnica de mínimos cuadrados, que se ajusta a los errores al cuadrado. De esta forma se puede estimar el modelo de la varianza asegurando que cada observación influye de igual forma en la función que se obtiene. Un ejemplo de esto se puede ver la figura 5.6, donde además se presenta el modelo de la varianza descrito en la sección 5.4.1.

En la figura 5.6 se representa con una línea continua la estimación propuesta anteriormente, donde se utilizan directamente las observaciones; con una línea negra y círculos se representa el primer modelo descrito, el cual se obtiene a partir de calcular la varianza de los errores por intervalos. Adicionalmente son representadas dos técnicas que serán explicadas en la sección 5.4.3.

En la revisión de los resultados se comprobó que, para algunos horizontes, la función estimada de la varianza toma valores negativos

para predicciones bajas, lo cual no es correcto según la definición de la varianza. Este problema es relevante ya que el parque opera con mucha frecuencia en este rango de producciones, por lo que no sería adecuado aplicar una aproximación. Ya que los modelos presentados anteriormente no garantizan una estimación fiable para la varianza de los errores en la predicción, se considerará la estimación no paramétrica mediante la regresión kernel, presentada a continuación.

5.4.3 Regresión kernel

La regresión kernel [44], en estadística, es considerada como una técnica no paramétrica utilizada para estimar la esperanza condicionada de una variable aleatoria. El objetivo es encontrar una relación no lineal entre Y (la variable dependiente) en términos de una variable X (la variable independiente, explicativa o regresor), de la siguiente forma:

$$Y = m(X), \quad (5.17)$$

Donde $m(\cdot)$ es una función cualquiera que relaciona a Y con X , determinada a partir del análisis empírico de las observaciones (conjunto de datos de entrenamiento) utilizadas en la estimación.

Tal como se dijo antes, partiendo del principio de que ningún modelo estimado se ajusta perfectamente a el comportamiento de una variable aleatoria, pero sí se aproxima localmente al promedio del valor que toma la variable, [44], se tiene que:

$$y_i = m(x_i) + \varepsilon_i \quad \forall \quad i = 1, \dots, n, \quad (5.18)$$

$$E(Y|X = x) = m(x) \quad (5.19)$$

La expresión $Y = m(X)$ no tiene que ajustarse perfectamente a cada observación i , que es afectada por una perturbación aleatoria ε_i . La ecuación (5.19), se refiere a que la relación se ajusta al promedio de las observaciones. El objetivo es que a partir de un conjunto finito de observaciones (y_i, x_i) , $i = 1, \dots, n$, se estima un modelo, $m(\cdot)$, que se ajusta a la esperanza condicional de una variable.

Por tanto, para cualquier regresión no paramétrica, la esperanza condicionada de una variable Y respecto a una variable X , se escribe como:

$$E(Y|X) = m(X) \quad (5.20)$$

La función estimada \hat{m} , es el resultado de la regresión no paramétrica conocida como el estimador Nadaraya-Watson [67, 101], y se define como,[44]:

$$\hat{m}_h(x) = \frac{1}{n} \sum_{i=1}^n W_{hi}(x) Y_i \quad (5.21)$$

Esta función se puede ver como promedio ponderado de la variable respuesta, ya que el sumatorio de los pesos debe cumplir con $\frac{1}{n} \sum_{i=1}^n W_{hi}(x) = 1$. Una descripción más detallada de la regresión kernel puede ser encontrada en [44].

Entonces, para modelar la varianza de los errores de la predicción de la ecuación (5.16) y (5.20), se tiene que:

$$\text{Var}(e_{t+k|t} | \hat{p}_{t+k|t}) = E(e_{t+k|t}^2 | \hat{p}_{t+k|t}) = \hat{m}(\hat{p}_{t+k|t}) \quad (5.22)$$

Tomando los datos de entrenamiento (en este caso, las predicciones y los respectivos errores en la predicción), se estima la función $\hat{m}(\hat{p}_{t+k|t})$ para cada horizonte de predicción k .

En la figura 5.6, se presenta una comparación de la varianza de los errores de la predicción eólica, estimada por medio de diferentes técnicas para el horizonte de predicción $k = 29$ h.

Los modelos de la varianza obtenidos por medio de la regresión kernel (para todos los horizontes), no toman valores negativos, garantizando de esta forma un resultado coherente para la varianza de los errores de la predicción, respecto a los modelos presentados anteriormente.

En la figura 5.6, se ha dibujado el resultado obtenido por medio de la regresión kernel (línea continua con puntos). Adicionalmente se ha dibujado el resultado obtenido al incorporar una ponderación de acuerdo a la antigüedad de las observaciones (línea discontinua). Una explicación detallada sobre los pesos o ponderaciones se desarrollará en la sección 6.3, donde también serán utilizados.

La regresión kernel tiene la capacidad de ajustar el modelo a los datos, y esta adaptación es útil para modelar los fenómenos estacionales de los errores. Adicionalmente es posible ponderar los datos por su antigüedad o relevancia. Estos aspectos se describirán con más detalle en la sección 6.3.1. Otra ventaja es la simple implementación del algoritmo que facilita su extensión, bien sea introduciendo pesos alternativos o diferentes parámetros de suavizado, [37].

Los resultados de los diferentes modelos son presentados en la figura 5.6. Se pueden apreciar diferencias considerables en la varianza de los errores (sobre todo para valores bajos de la predicción eólica $\hat{p}_{t+k|t} < 0,1$ pu) entre los modelos obtenidos mediante la técnica de mínimos cuadrados y la regresión kernel. Para verificar qué modelo conjunto (distribución beta más el modelo de la varianza de los

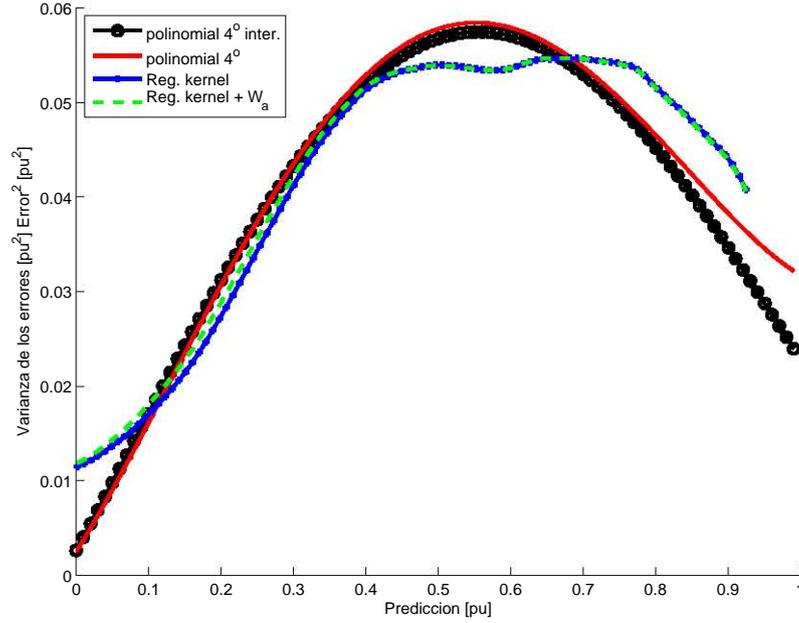


Figura 5.6: Comparación de diferentes técnicas para estimar la varianza de los errores en la predicción, $k = 29$ h.

errores) representa mejor la distribución marginal de los errores en la predicción, se utilizará la técnica descrita en la sección 4.3.1 como método de evaluación y comparación, presentada a continuación.

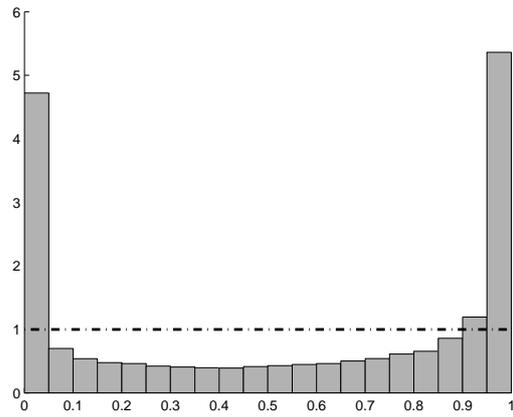
5.5 EVALUACIÓN DE LA FUNCIÓN DE DISTRIBUCIÓN

Los modelos conjuntos (distribución beta más el modelo de la varianza de los errores) son estimados a partir del conjunto de datos de entrenamiento. Posteriormente, por cada predicción del conjunto de datos de evaluación, $\hat{p}_{t+k|t}$, se obtiene el modelo de la función marginal para los errores en la predicción, donde \hat{F}_e^{-1} es la inversa de la CDF de los errores de la predicción. Una vez hecho esto, se convierte cada uno de los errores de la predicción ($e_{t+k|t} = p_{t+k} - \hat{p}_{t+k|t}$) al dominio uniforme, utilizando la inversa de la función de distribución de probabilidad, $F_e^{-1}(e_{t+k|t}) = U_{t+k|t}$.

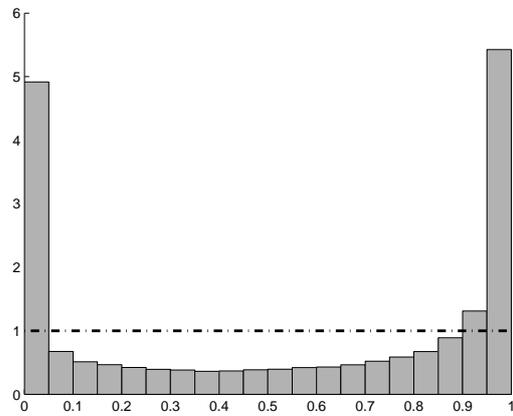
En la figura 5.7 se presentan los resultados del histograma PIT de los modelos, utilizando la distribución beta con los diferentes modelos para la varianza de los errores.

Todos los histogramas PIT obtenidos a partir de la distribución beta y mostrados en la figura 5.7 reflejan problemas en los modelos, ya que no se aproximan a una distribución uniforme, tal como se explicó en la sección 4.3.1 y figura 4.7. El RMSE del PIT y el CRPS no se calcularon, ya que no aportaba información adicional a este resultado que muestra claramente un problema en los modelos.

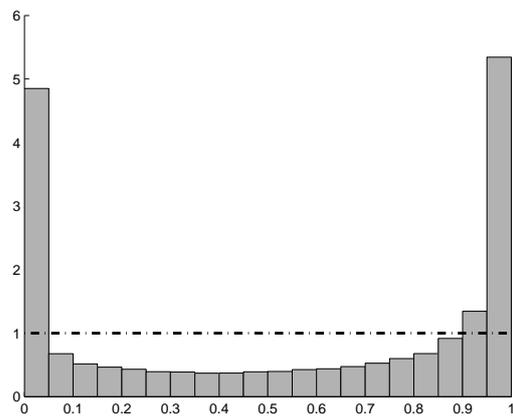
Al hacer un análisis de los resultados se observó que hay una gran concentración de producciones nulas para predicciones con valores



(a) Mínimos cuadrados.



(b) Regresión kernel.



(c) Regresión kernel con ponderación por la antigüedad en los datos.

Figura 5.7: Evaluación de la distribución beta, para diferentes técnicas de estimación de la varianza.

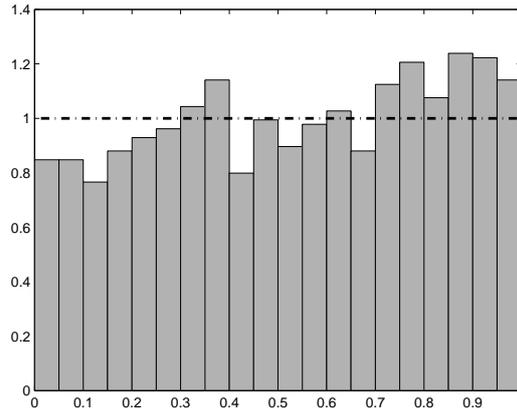


Figura 5.8: Evaluación de la distribución beta sin considerar las producciones nulas y sus respectivas predicciones. En este caso la varianza fue obtenida por medio de la técnica de mínimos cuadrados.

bajos, $\hat{p} \leq 0,2$. Para estos casos hay gran probabilidad de que la producción sea nula y la PDF estimada no consiga replicar el comportamiento extremo. Estos datos se ven reflejados en la zona izquierda de los histogramas PIT.

La concentración a la derecha de los histogramas PIT se podría justificar de forma análoga a la concentración de la izquierda, debida a producciones iguales a la potencia nominal del parque. Sin embargo, al revisar el datos de entrenamiento y evaluación no se encuentra un número considerable de observaciones con esta magnitud.

Haciendo una evaluación adicional se obtuvo el histograma PIT, eliminando las producciones nulas y sus respectivas predicciones, tanto del conjunto de datos de entrenamiento como de evaluación; el resultado del histograma PIT es mucho más aproximado a una distribución uniforme, figura 5.8. Con esto se explica que ambas concentraciones se deben a las producciones nulas, ya que introducen sesgos en las funciones marginales estimadas.

Uno de los inconvenientes de utilizar una función paramétrica es su limitada capacidad de adaptación a los datos reales, por lo que el modelo no es válido para todos los casos posibles. Como alternativa a la distribución beta se estimará la función de densidad por medio de una técnica no paramétrica. De esta forma se espera obtener una función de densidad que se adapte a las características de los datos.

MODELO LOCAL DE LA FUNCIÓN DE DENSIDAD POR MEDIO DE FUNCIONES KERNEL

Resumen

En este capítulo se presenta una técnica de modelado local para determinar la función de densidad de probabilidad de la producción eólica por medio de funciones kernel. El objetivo es tener un método alternativo a la distribución beta, que tiene limitaciones por ser una función paramétrica. Así, se busca una técnica no paramétrica que modele mejor el comportamiento de los errores de la producción. Finalmente el modelo es validado utilizando datos reales de producción eólica.

6.1 INTRODUCCIÓN

En estadística, la estimación de la función marginal de una variable aleatoria, mediante las funciones kernel, es clasificada como una técnica no paramétrica para obtener la PDF. Es decir, no supone ninguna hipótesis sobre la distribución que representaría los datos, sino que se estima la forma de la función del análisis empírico de los datos de entrenamiento.

La estimación no paramétrica de la PDF, mediante las funciones kernel, es considerada una técnica de suavizado, más conocida en inglés como Kernel Density Estimation (KDE). Para un modelo local, se parte de una ventana de datos de la forma $[\hat{p}_{t+k|t} - \delta_1, \hat{p}_{t+k|t} + \delta_2]$, donde el parámetro $\delta = \delta_1 + \delta_2$ es llamado «ancho de la ventana de datos». Éste determina el intervalo de los datos utilizados para el modelo local. Los datos seleccionados, $\{\hat{p}_{t+k|t,j} : j = 1, \dots, n\}$, se utilizan para calcular la PDF del modelo local en función del valor seleccionado $\hat{p}_{t+k|t,s}$. Un esquema de esta formulación se puede ver en la figura 6.1, donde los datos seleccionados están sobre el área sombreada.

Según los resultados presentados en [37], con un 20% de los datos más cercanos a $\hat{p}_{t+k|t,s}$, se obtiene una estimación fiable de la PDF. Por tanto, el ancho de la ventana de datos se determina de acuerdo a la cantidad de datos disponibles en el conjunto de entrenamiento. Según [37] se tomará un 20% de los datos más cercanos al valor $\hat{p}_{t+k|t,s}$, y por esta razón la magnitud de δ dependerá de la concentración de los datos. Para valores bajos de la predicción el parámetro δ será menor, comparado con la magnitud que tomará para valores altos. Tomando un 20% de los datos de entrenamiento, para un mo-

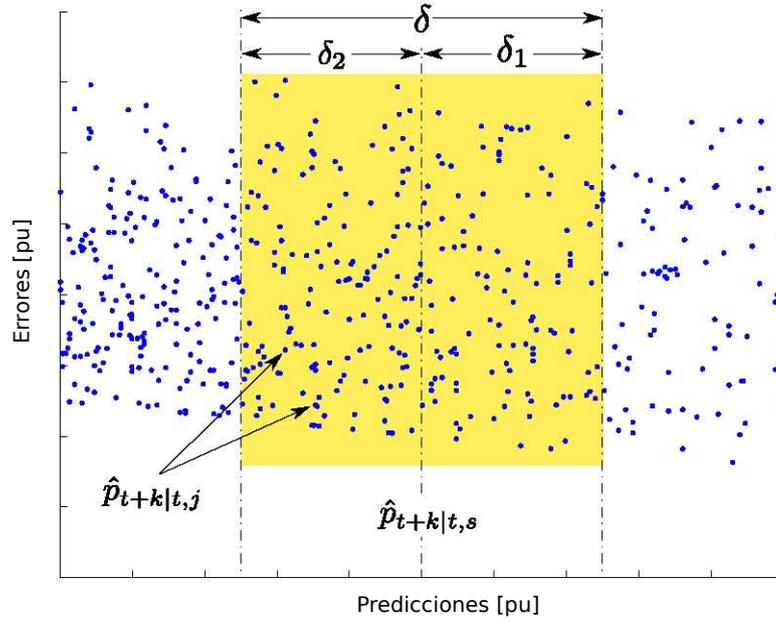


Figura 6.1: Ejemplo de los valores seleccionados según el ancho de la ventana de datos.

delo local, se garantiza que él método tiene los datos necesarios para estimar una PDF fiable en todo el rango de valores posibles que toma la predicción.

6.2 MODELO LOCAL DE LA PDF POR MEDIO DE FUNCIONES KERNEL

Sea (x_1, x_2, \dots, x_n) una serie de observaciones de una variable aleatoria, con una PDF desconocida f . Para estimar la forma de esta función en torno al punto x , por medio de KDE, se utiliza la siguiente expresión:

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x-x_j}{h}\right) = \frac{1}{nh} \sum_{j=1}^n K(u_j) \quad (6.1)$$

Donde $K(\cdot)$ es una función kernel que otorga pesos o pondera, y $h > 0$ es el parámetro de suavizado, también llamado «ancho de banda». En [44] se encuentra una explicación más detallada de la estimación de la densidad por medio de las funciones kernel.

Para el caso particular de estimar la PDF de los errores de la predicción, $\hat{f}_{e,k}$, por medio de KDE. La distancia relativa, u_j , de cada uno de los datos de entrenamiento seleccionados en la ventana de datos,

$\{\hat{p}_{t+k|t,j} : j = 1, \dots, n\}$, respecto al punto donde se busca el modelo local, $\hat{p}_{t+k|t,s}$, se mide usando la siguiente expresión:

$$u_j = \frac{\hat{p}_{t+k|t,j} - \hat{p}_{t+k|t,s}}{h} \quad \forall j \quad (6.2)$$

Hay diferentes formas de calcular el parámetro de suavizado, [44], pero suponen realizar hipótesis sobre la función a estimar, lo cual es un inconveniente a priori. En [37], se estima la PDF por medio de aplicar KDE a un conjunto de datos reales, donde solo se varía el valor del parámetro de suavizado. En las pruebas realizadas se demostró que la regla de Silverman, [44], se aproxima en la mayoría de los casos al valor que obtenía la mejor evaluación. Por tanto, de forma práctica se utiliza la regla de Silverman, expresada de la siguiente forma:

$$\hat{h} = 1,06 \min \left\{ \hat{\sigma}_X, \frac{R}{1,34} \right\} n^{-1/5} \quad (6.3)$$

Donde $R = Q_3 - Q_1 = X_{75\%} - X_{25\%}$, llamado «rango intercuartílico» (Q_1 y Q_3 son los cuartiles primero y tercero respectivamente), $\hat{\sigma}_X$ es la desviación estándar de la muestra, y n es el número de elementos en la muestra, tomados de la ventana de datos, figura 6.1.

6.2.1 Funciones kernel

Las funciones kernel son un método de ponderación o peso, utilizadas en técnicas de estimación no paramétrica. Estas funciones son usadas para estimar la PDF correspondiente de una variable aleatoria KDE; o en regresiones kernel, para estimar la esperanza condicional de una variable aleatoria.

Una función kernel $K(\cdot)$, debe garantizar que la función estimada \hat{f} corresponde a la PDF de la muestra analizada, por lo que debe cumplir la siguiente expresión:

$$\int_{-\infty}^{+\infty} K(u) \, du = 1 \quad (6.4)$$

Que a efectos prácticos, ya que se trata de observaciones, quedaría de la siguiente forma:

$$\sum_{j=1}^n K(u_j) = 1 \quad (6.5)$$

Además, para todas las observaciones de la muestra y sus respectivas distancias relativas se debe cumplir que $K(-u_j) = K(u_j)$.

En la figura 6.2 son dibujadas algunas funciones kernel. A continuación se presentan sus respectivas formulaciones.

$$\text{Epanechnikov} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{3}{4}(1 - u_j^2)^2 & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.6)$$

$$\text{Gaussiana} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_j^2} & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.7)$$

$$\text{Cuártica} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{15}{16}(1 - u_j^2)^2 & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.8)$$

$$\text{Triangular} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ (1 - |u_j|) & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.9)$$

$$\text{Tri-cúbica} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{70}{81}(1 - |u_j|^3)^3 & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.10)$$

$$\text{Tri-ponderada} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{35}{32}(1 - u_j^2)^3 & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.11)$$

$$\text{Uniforme} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{1}{2} & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.12)$$

$$\text{Coseno} \quad K(u_j) = \begin{cases} 0 & \text{Si } |u_j| > 1 \\ \frac{\pi}{4} \cos\left(\frac{\pi}{2}u_j\right) & \text{Si } |u_j| \leq 1 \end{cases} \quad (6.13)$$

En [44] y [37], a partir de un análisis matemático y de una serie de cálculos numéricos, se concluye que a efectos prácticos, la elección de la función kernel es irrelevante para la fiabilidad de la estimación. Un ejemplo de esto se puede ver en la figura 6.3, donde se generaron 10000 números aleatorios a partir de la distribución beta, con los parámetros $a = 2$ y $b = 5$, y posteriormente se estimó la función de densidad mediante KDE utilizando diferentes funciones kernel. Las curvas estimadas se sobrepone casi en su totalidad y solo hay diferencias menores, lo cual confirma la conclusión de los autores.

Por tanto se utilizará la función triangular por su simple implementación, además de que asigna valores nulos a todas las observaciones donde $|u_j| > 1$, lo que, a efectos de la implementación, hace más rápido el algoritmo.

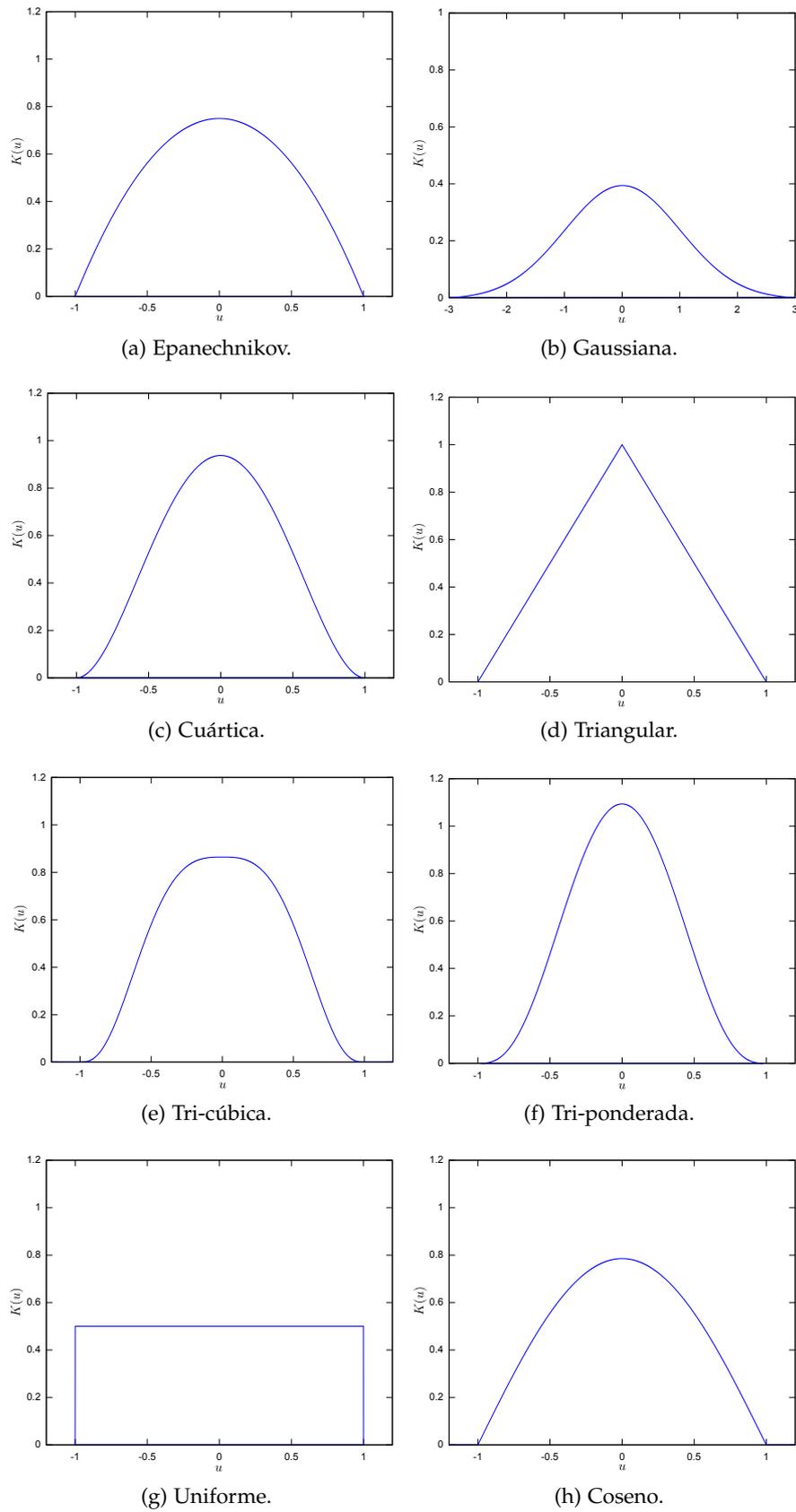


Figura 6.2: Funciones kernel.

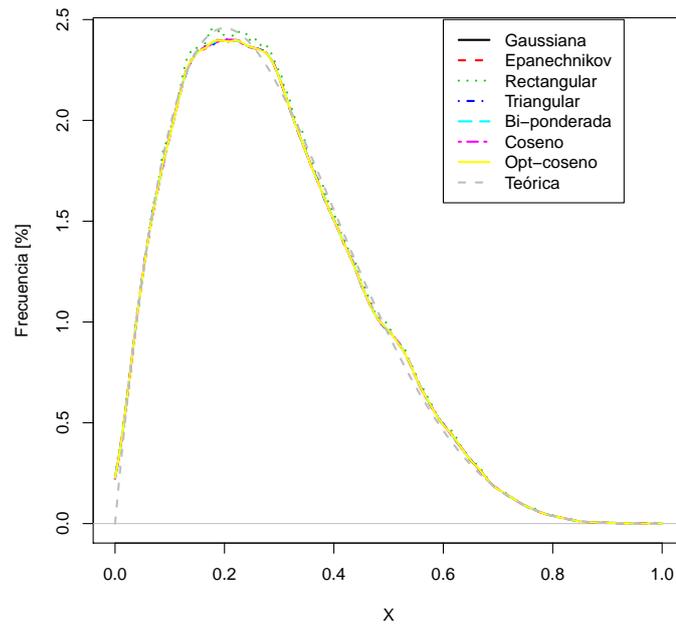


Figura 6.3: Estimación de la PDF usando diferentes funciones kernel. Función teórica: distribución beta, con los parámetros $\alpha = 2$ y $b = 5$.

6.3 FUNCIONES DE PONDERACIÓN

Además de las funciones kernel (que son funciones de ponderación para obtener la estimación de la PDF o de la regresión) se utilizan funciones de ponderación para dar relevancia a los datos históricos o de entrenamiento. Esta ponderación tiene como objetivo modelar fenómenos estacionales o dar más relevancia a los datos históricos que pueden describir mejor el fenómeno futuro.

6.3.1 Ponderación por antigüedad

La forma de la PDF puede variar a medida que llegan nuevos valores al modelo, sea por condiciones estacionales, porque varía la propia incertidumbre de las variables de entrada, por cambios en la precisión del modelo de predicción de potencia, por la instalación de nuevos parques eólicos en la región o por la sustitución de aerogeneradores en algún parque. En cualquiera de los casos, esto implicará variaciones en la forma de la PDF.

La ponderación por antigüedad da un nivel de adaptabilidad a la función estimada. Por tanto, se dará más peso o relevancia a las observaciones más recientes. Por esta razón se incluye una ponderación en función de su antigüedad, donde $\tau = 1$ representa la observación más reciente, y $\tau = n$ la observación más antigua. El parámetro λ es

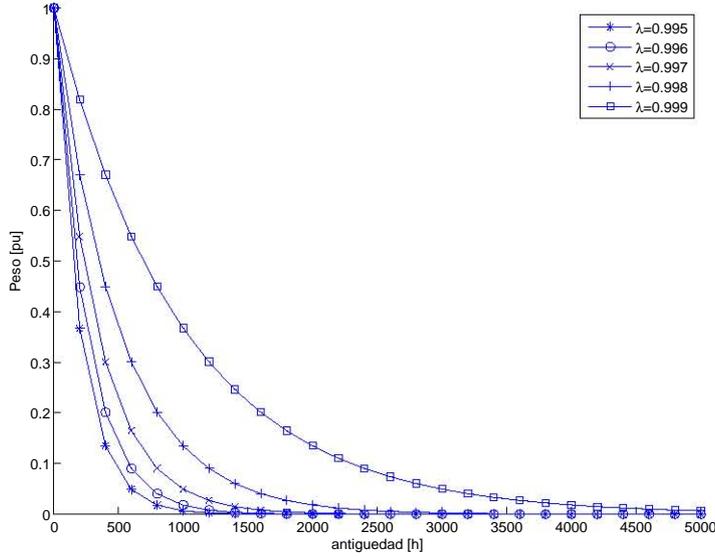


Figura 6.4: Ejemplos de la ponderación de una observación según el factor de olvido.

conocido como «factor de olvido», y de este valor depende la relevancia de la observación de acuerdo a la expresión λ^{τ_j} . En la figura 6.4 se dibujan algunos ejemplos. Para asegurar que la nueva estimación cumple lo expresado en la ecuación (6.4), los pesos son escalados para cada una de las observaciones de los datos seleccionados del modelo local, donde $j = 1, \dots, n$.

$$W_{\tau,j} = \frac{\lambda^{\tau_j}}{\sum_{l=1}^n \lambda^{\tau,l}} \quad \forall j \quad (6.14)$$

Entonces la estimación de la PDF, considerando una ponderación por antigüedad, quedaría de la siguiente forma:

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^n W_{\tau,j} K(u_j) \quad (6.15)$$

En la figura 6.4 se relaciona el factor de olvido y la antigüedad de la observación respecto a la ponderación que se asignaría a una observación concreta. Por ejemplo, un factor de olvido $\lambda = 1$ haría que todas las observaciones tuvieran la misma ponderación, y a medida que λ disminuye, las observaciones más antiguas pierden relevancia respecto a las más actuales, [37, 89].

6.3.2 Ponderación según el valor de la predicción

Para tener en cuenta el nivel de potencia prevista, se añade una ponderación $W_{\hat{p}}$ que hace más relevantes los datos históricos cercanos al

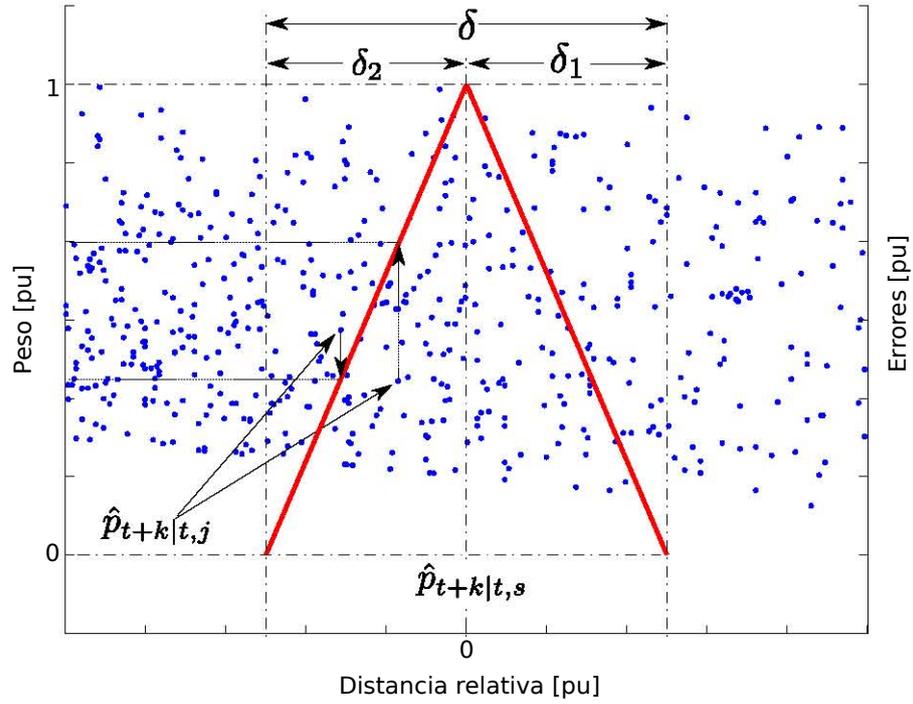


Figura 6.5: Ponderación según la distancia relativa.

valor previsto. De forma similar a la distancia relativa de las observaciones, descrita en la ecuación (6.2), se determinan las ponderaciones utilizando nuevamente una función kernel, quedando de la siguiente forma:

$$W_{\hat{p},j} = \frac{K(u_{\hat{p},j})}{\sum_{l=1}^n K(u_{\hat{p},l})} \quad (6.16)$$

Donde:

$$u_{\hat{p},j} = \frac{\hat{p}_{t+k|t,j} - \hat{p}_{t+k|t,s}}{\delta/2} \quad (6.17)$$

Según (6.2), una observación con una distancia relativa igual a cero obtendría la máxima relevancia según la ponderación. A medida que la distancia relativa se hace mayor, la ponderación disminuiría, haciendo que la observación influya menos en la estimación del modelo. En la figura 6.5 se muestra un esquema de esta formulación.

De forma semejante a el procedimiento de las funciones kernel, presentado en la figura 6.1, las observaciones que están por fuera de la ventana de datos no serán consideradas en los cálculos.

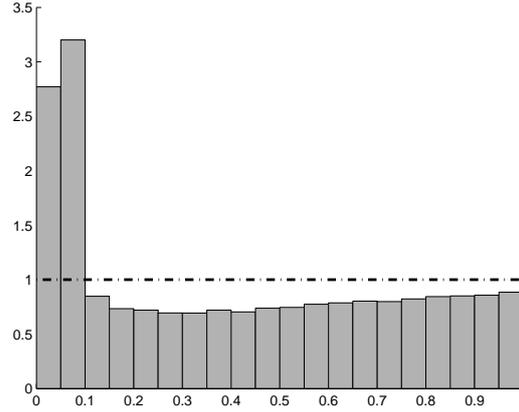


Figura 6.6: Evaluación de la función de distribución estimada mediante funciones kernel.

Finalmente, la estimación de la PDF, considerando una ponderación por antigüedad y por el valor de la potencia prevista, quedaría de la siguiente forma:

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^n W_{\tau,j} W_{\hat{p},j} K(u_j) \tag{6.18}$$

6.4 EVALUACIÓN DE LA FUNCIÓN DE DISTRIBUCIÓN

A partir de la ecuación (6.15) se estima el modelo local de los errores en la predicción eólica. Se ajusta el factor de olvido a $\lambda = 0,999$, que es un valor comúnmente utilizado para este tipo de cálculos, [88]. Según [37], se toma un 20% de las observaciones del conjunto de datos de entrenamiento (en este caso, predicciones y sus respectivos errores) más cercanas al punto escogido para el modelo local.

Por cada predicción del conjunto de datos de evaluación, $\hat{p}_{t+k|t}$, se obtiene el modelo de la función marginal para los errores en la predicción, donde \hat{F}_e^{-1} es la inversa de la CDF de los errores de la predicción. Una vez hecho esto, se convierte cada uno de los errores de la predicción, $e_{t+k|t} = p_{t+k} - \hat{p}_{t+k|t}$, al dominio uniforme utilizando la inversa de la función de distribución de probabilidad, $F_e^{-1}(e_{t+k|t}) = U_{t+k|t}$.

En la figura 6.6 se presentan los resultados del histograma PIT de la distribución estimada mediante las funciones kernel. El RMSE del PIT y el CRPS no se calcularon, ya que no aportaban información adicional a este resultado, que muestra claramente un problema en el modelo.

Al igual que los histogramas PIT de la distribución beta mostrados en la figura 5.7, se aprecia un problema en el modelo en la figura 6.6. Sin embargo, hay una clara diferencia, ya que solo tiene una concen-

tración a la izquierda, debida a la considerable cantidad de producciones nulas para predicciones con valores bajos, $\hat{p} \leq 0,2$. Para modelar los errores en la predicción hay que considerar las producciones nulas de una forma explícita, puesto que los modelos de la distribución beta y la estimación con funciones kernel no consiguen modelar de forma general el comportamiento de los errores.

La regresión logística que se presenta en el capítulo siguiente tiene el objetivo de explicar los errores debidos a las producciones nulas como una variable de respuesta cualitativa. Modelando de forma independiente los errores debidos a las producciones nulas se pretende realizar un modelo conjunto que logre describir el comportamiento de los errores en la predicción.

REGRESIÓN LOGÍSTICA PARA MODELAR EL ERROR EN LA PREDICCIÓN, CUANDO LA PRODUCCIÓN ES NULA

Resumen

Debido a la incapacidad de modelar correctamente los errores en la predicción en el caso de las producciones nulas (utilizando la distribución beta y la función de distribución estimada por medio de KDE), se hace necesario considerar y modelar de forma independiente este suceso particular para poder obtener un modelo fiable. Con la regresión logística se pueden modelar los errores debidos a las producciones nulas como una variable de respuesta cualitativa en función de la predicción. En este capítulo se presenta una breve explicación sobre la regresión logística y cómo se aplicará para estimar la probabilidad de que ocurra una producción nula. Una vez estimada esta probabilidad se utilizará para reconstruir el modelo completo de los errores en la predicción eólica, donde se considerará todo el intervalo posible. Finalmente el modelo se valida utilizando datos reales de producción.

7.1 INTRODUCCIÓN

Los histogramas PIT de la distribución beta, mostrados en la figura 5.7, reflejan problemas en el modelo al no estar uniformemente distribuidos. Al hacer un análisis de los resultados, se determinó que hay una concentración de producciones nulas que introduce errores en el modelo, dando como resultado concentraciones en los extremos del histograma PIT. Este problema ya sido tratado en la sección 5.5.

Algo semejante ocurre en los histogramas PIT para las distribuciones estimadas con funciones kernel, figura 6.6. En este caso, es apreciable una concentración a la izquierda del histograma PIT debida a las producciones cero. Este problema ya sido tratado en la sección 6.4.

Para solucionar estos problemas se plantea una alternativa, ya que este suceso no logra ser modelado por las funciones de densidad propuestas en los capítulos anteriores. En [51], las producciones nulas son modeladas como una discontinuidad y son introducidas como una parte discreta en la distribución que modela la producción eólica. En este caso la probabilidad es calculada como la proporción de producciones nulas, dada la potencia prevista.

El procedimiento propuesto en este capítulo propone dividir el modelo en dos partes. La primera se encargará de modelar los errores

de la predicción debidos a producciones mayores que cero, utilizando las técnicas explicadas en los capítulos 5 y 6. Y la segunda parte se encargará de modelar los errores debidos a las producciones nulas por medio de la regresión logística, que es un método mucho más apropiado que una función de masa de probabilidad, presentada en [51].

La regresión logística tiene como objetivo modelar la probabilidad de un suceso cualitativo en función de otros factores explicativos que puede ser continuos y/o discretos. El problema consiste en estimar la probabilidad de que ocurra una producción nula. Entonces, usando como variable explicativa el valor de la predicción a un horizonte de tiempo determinado, se quiere estimar la probabilidad de que la producción real sea igual a cero, $P(p_{t+k} = 0 | \hat{p}_{t+k|t})$. Una vez estimada esta probabilidad, será utilizada para obtener la CDF que modele correctamente el comportamiento de los errores de la predicción eólica.

7.2 FORMULACIÓN

En la teoría de la probabilidad y la estadística, la regresión logística (a veces llamada «modelo logístico» o «modelo logit») es un modelo de regresión que se utiliza para estimar la probabilidad de ocurrencia de una variable con respuesta cualitativa en función de un conjunto de variables explicativas, que puede ser continuo y/o discreto.

En muchos problemas se desea explicar el comportamiento de un fenómeno (variable respuesta) en función de un conjunto de entradas (variables explicativas). La variable respuesta, \mathcal{Y} , es cualitativa y tomará valores de $\mathcal{Y} = 1$ si el suceso de interés ocurre, y $\mathcal{Y} = 0$ en el caso contrario. Para obtener el modelo se dispone de un conjunto de variables explicativas denotadas por $\mathcal{X} = (\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_n)$.

El modelo logístico asume una relación suave entre el vector de variables explicativas, \mathcal{Y} , y la probabilidad de ocurrencia del suceso de interés, \mathcal{P} .

$$\mathcal{P}_i = \frac{1}{1 + e^{-\mathcal{B}^T \mathcal{X}_i}} \quad (7.1)$$

con:

$$\mathcal{B}^T \mathcal{X}_i = \mathcal{B}_0 + \mathcal{B}_1 \mathcal{X}_{1,i} + \mathcal{B}_2 \mathcal{X}_{2,i} + \dots + \mathcal{B}_n \mathcal{X}_{n,i} \quad (7.2)$$

Donde \mathcal{B}_0 es la «intercepción», y $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$ son los «coeficientes de la regresión» de $\mathcal{X}_{1,i}, \mathcal{X}_{2,i}, \dots, \mathcal{X}_{n,i}$ respectivamente.

Una vez se tiene el vector de variables respuesta \mathcal{Y} y el vector o matriz con las variables explicativas \mathcal{X} , se estiman los coeficientes de la regresión $\hat{\mathcal{B}}$ por medio del método de máxima verosimilitud (ML). Una explicación más detallada de la regresión logística y de cómo estimar sus coeficientes se puede encontrar en [73].

7.2.1 Estimación de los coeficientes de la regresión

A partir del conjunto de datos de entrenamiento se calcula la variable respuesta, que tomará valores de $\mathcal{Y} = 1$ si la producción es nula, $p_{t+k} = 0$, y en el caso contrario, $\mathcal{Y} = 0$, si la producción es mayor que cero, $p_{t+k} > 0$. Para estimar los modelos se ha considerado como valor nulo a toda producción menor que 0,001 pu.

Como primera aproximación (modelo 1) para el vector de variables explicativas se utilizó $\mathcal{X}_i = \hat{p}_{t+k|t}$, y como segunda aproximación (modelo 2) se tiene $\mathcal{X}_i = (\hat{p}_{t+k|t}, \hat{p}_{t+k|t}^2)$.

Siguiendo el procedimiento para determinar los coeficientes de la regresión $\hat{\mathcal{B}}$ se aplica el método de máxima verosimilitud (ML), tal como se propone en [73]. En la figura 7.1 se muestran los modelos estimados para $P(p_{t+k} = 0 | \hat{p}_{t+k|t})$ a partir de las dos opciones presentadas para el vector de variables explicativas. En ambos ejemplos, el modelo estimado es para el horizonte de predicción $k = 20$ h.

Para evaluar las dos opciones presentadas se utiliza una de las medidas más usuales para determinar el ajuste global de la regresión logística, [73], llamado «coeficiente de determinación». Para el caso particular de la regresión logística se define como:

$$R^2 = 1 - \frac{L(\hat{\mathcal{B}})}{L(\hat{\mathcal{B}}_0)} \quad (7.3)$$

Donde el numerador es la desviación total (verosimilitud) del modelo estimado, y el denominador, la desviación (verosimilitud) en el modelo que incluye sólo la constante, [73]. Si el ajuste del modelo estimado es perfecto, entonces $R^2 = 1$; esto quiere decir que todas las observaciones con $\mathcal{Y} = 1$ obtienen una probabilidad del 100%, y todas las observaciones con $\mathcal{Y} = 0$ obtienen una probabilidad de 0%. Entonces la desviación total del modelo es cero ($L(\hat{\mathcal{B}}) = 0$). Por el contrario, si las variables explicativas no influyen nada en la estimación del modelo, la desviación con las variables explicativas será igual a la desviación sin ellas; por tanto, $L(\hat{\mathcal{B}}) = L(\hat{\mathcal{B}}_0)$ y entonces $R^2 = 0$.

En la figura 7.2 se muestra el resultando obtenido del modelo estimado con los datos de entrenamiento, y evaluado con el conjunto de datos de evaluación. Se puede observar que el modelo 2 consigue un mejor desempeño en todos los horizontes de predicción, ya que el coeficiente de determinación toma un valor mayor que el modelo 1. Por esta razón se utilizará el modelo 2 para realizar el modelo de los errores en la predicción.

7.3 INTEGRACIÓN DE LOS MODELOS

Del conjunto de datos de entrenamiento se toman las producciones mayores que cero, $p_{t+k} > 0$, con sus respectivas perdiciones. Con este

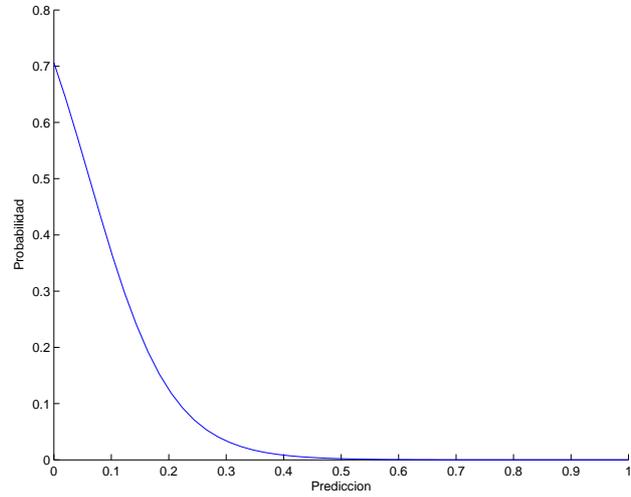
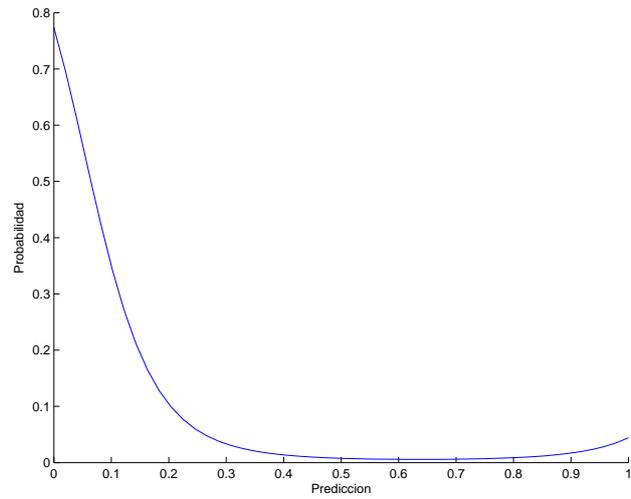
(a) Modelo 1, $\mathcal{X}_i = \hat{p}_{t+k|t}$.(b) Modelo 2, $\mathcal{X}_i = (\hat{p}_{t+k|t}, \hat{p}_{t+k|t}^2)$.

Figura 7.1: Ejemplos de los modelos estimados para la probabilidad de las producciones nulas usando la regresión logística, horizonte de predicción $k = 20$ h.

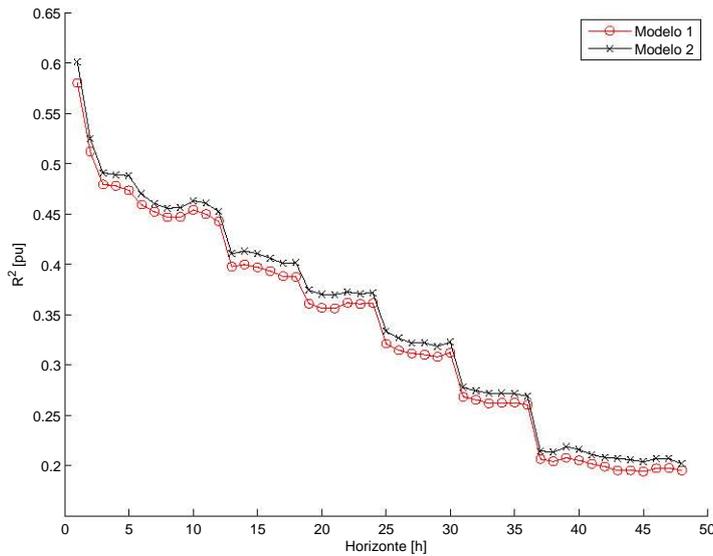


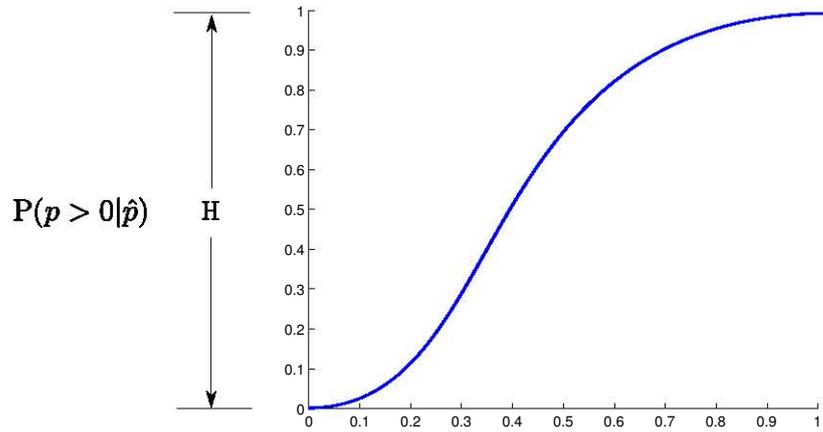
Figura 7.2: Evaluación de los modelos de la probabilidad de que ocurra una predicción nula.

subconjunto de datos de entrenamiento son calculadas nuevamente las funciones marginales de los errores de la predicción, descritas en los capítulos 5 y 6. De esta forma, las nuevas funciones marginales no son influenciadas por las producciones nulas. Un ejemplo de esto se puede ver en la figura 7.3a. Estas funciones marginales estimadas son escaladas respecto a su altura, de acuerdo con la probabilidad estimada por medio de la regresión logística, tal y como se ve en la figura 7.3b. El procedimiento consiste en escalar la altura de la función marginal y desplazarla hacia arriba; el espacio resultante en la parte inferior modela los errores en la predicción, debidos a las producciones nulas, ya que se reserva una distancia proporcional a la probabilidad de que éstas ocurran. En la parte superior son modelados los errores de la predicción debidos a producciones mayores que cero, usando las funciones marginales (descritas en los capítulos 5 y 6) que son estimadas con el subconjunto de datos entrenamiento, donde no son consideradas las producciones nulas.

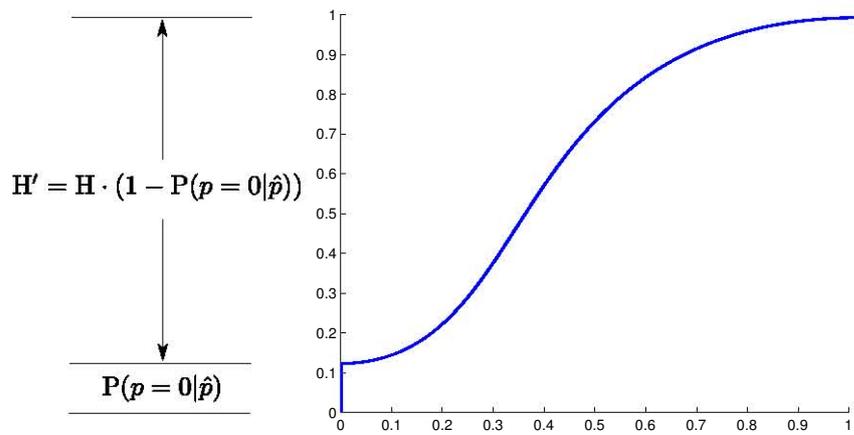
Con la CDF resultante es posible modelar los errores de la predicción eólica en todo el intervalo posible, desde producciones nulas hasta la potencia máxima del parque.

7.4 EVALUACIÓN DE LA FUNCIÓN DE DISTRIBUCIÓN

Por cada predicción del conjunto de datos de evaluación, $\hat{p}_{t+k|t}$, se obtienen los modelos de la función marginal de los errores en la predicción eólica. En este caso se estimaron las distribuciones utilizando la distribución beta con los dos modelos de varianza de los errores (mínimos cuadrados y regresión kernel) y la distribución estimada



(a) Marginal para producciones mayores que cero.



(b) Marginal para todo el intervalo de producciones.

Figura 7.3: Escalado de la función de distribución de probabilidad.

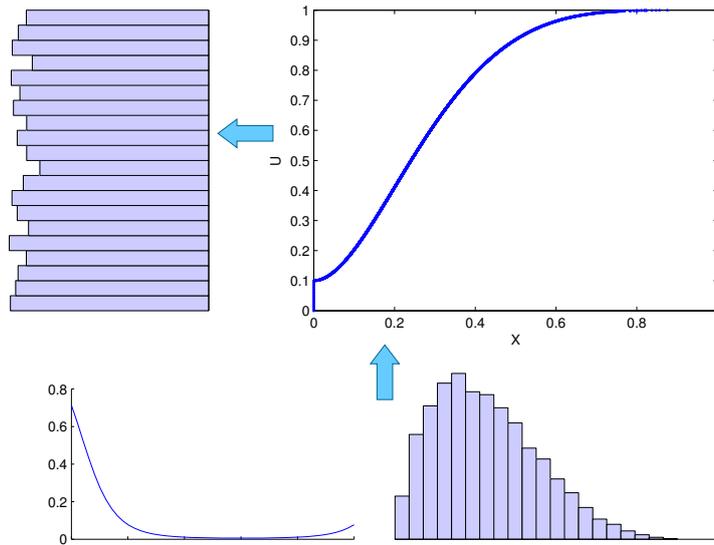


Figura 7.4: Esquema de evaluación de la función de distribución con la regresión logística.

mediante las funciones kernel. Estas funciones son estimadas utilizando el subconjunto de datos de entrenamiento donde no se consideraron las producciones nulas.

Luego por cada predicción del conjunto de datos de evaluación se estimó la probabilidad de que ocurriera una producción nula, utilizando el modelo de la regresión logística. Con el valor obtenido se reconstruyeron las CDF, tal y como se muestra en la figura 7.3b.

Una vez hecho esto, se convierte cada uno de los errores de la predicción del conjunto de datos de evaluación ($e_{t+k|t} = p_{t+k} - \hat{p}_{t+k|t}$) al dominio uniforme, utilizando la inversa de la función de distribución de probabilidad de los modelos presentados anteriormente ($F_e^{-1}(e_{t+k|t}) = u_{t+k|t}$).

En la figura 7.6 se presentan los resultados del histograma PIT para las diferentes funciones de distribución propuestas. Distribución beta: donde la varianza se estima por medio de la regresión polinómica, figura 7.6a; donde la varianza se estima usando una regresión kernel, figura 7.6b. Luego se presenta el histograma PIT de la función de distribución estimada, utilizando las funciones kernel, figura 7.6c.

Para evaluar la función de distribución se utiliza el RMSE del histograma PIT, tal y como se describe en la sección 4.3. Los resultados son presentados en la tabla 7.1, donde se puede apreciar que la distribución estimada mediante funciones kernel es la que tiene un mejor ajuste entre las opciones presentadas. Cabe recordar que en este tipo de evaluación se están considerando todos los horizontes de predicción al mismo tiempo. Para hacer una evaluación discriminando cada horizonte y comparando los modelos entre si, se utiliza la técnica conocida en inglés como Continuous Ranked Probability Score (CRPS), [42].

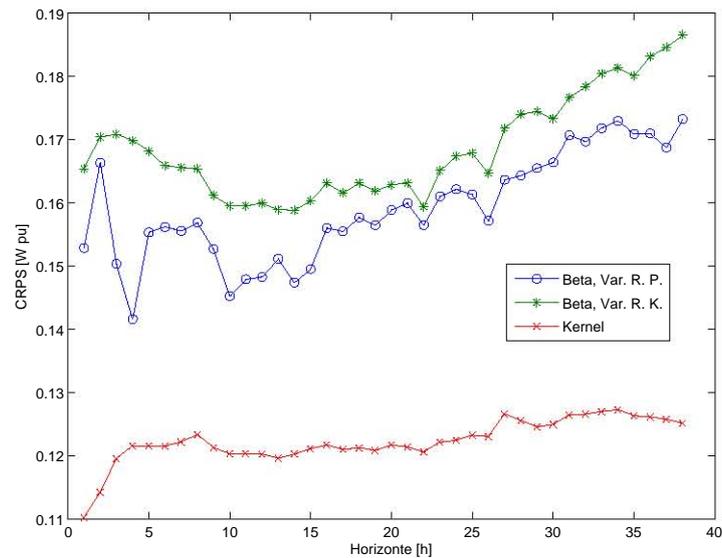


Figura 7.5: Comparación de los modelos propuestos para las funciones marginales, utilizando el criterio de evaluación CRPS.

El CRPS se calcula para evaluar y comparar los modelos presentados. La principal ventaja del CRPS es que es sensible a la forma de la función de distribución (sección 4.3.2). Los resultados obtenidos a través del CRPS se muestran en la figura 7.5. Con los resultados obtenidos se ratifica el desempeño de la distribución estimada mediante funciones kernel, comparada con las dos versiones presentadas de la distribución beta. Tal se explica en la sección 4.3.2, si el modelo estimado representara perfectamente la variable se obtendría un valor cero. Esto quiere decir que cuanto más bajo sea el valor del CRPS, será mejor el modelo que representa la variable.

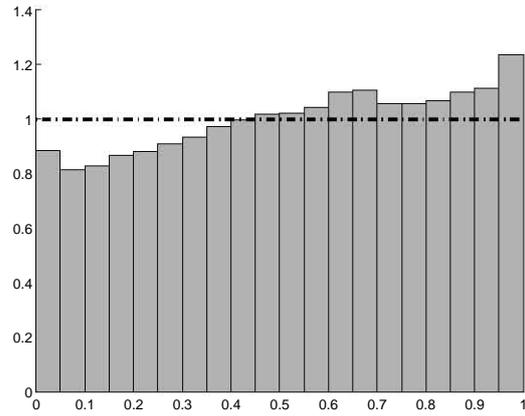
Tabla 7.1: Resumen de la evaluación de las diferentes funciones de distribución propuestas.

Tipo de función de distribución	RMSE _{PIT}
Distribución beta, regresión polinómica	0,1088
Distribución beta, regresión kernel	0,1301
Distribución estimada mediante funciones kernel	0,0430

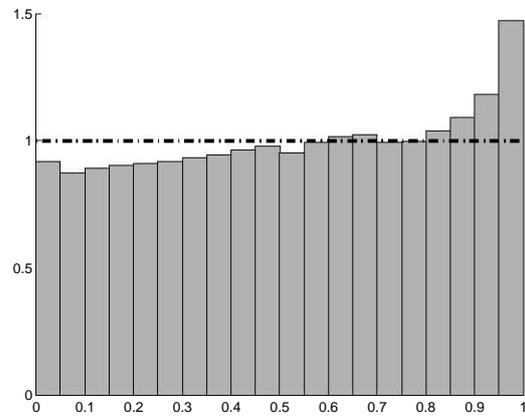
7.5 CONCLUSIONES

Dados los resultados de las evaluaciones presentadas, se utilizará la función de distribución obtenida por medio de las funciones kernel y la regresión logística.

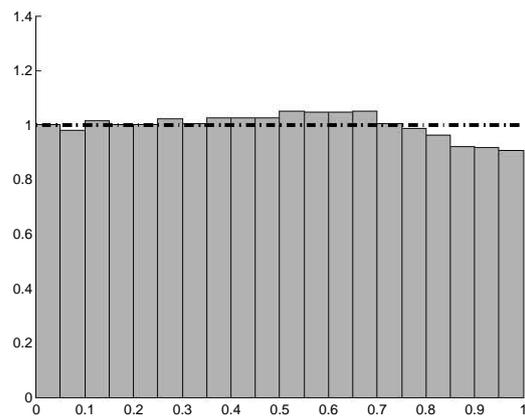
El modelo utilizando de la distribución beta tiene problemas de ajuste, pero éstos pueden deberse a errores en la estimación de los



(a) Distribución beta, con la varianza estimada por medio de la regresión polinómica.



(b) Distribución beta, con la varianza estimada por medio de la regresión kernel.



(c) Función de distribución estimada mediante funciones kernel.

Figura 7.6: Evaluación las funciones de distribución propuestas.

parámetros. Sin embargo, es una aproximación simple que puede ser de interés para algunas aplicaciones en modelos regionales, por lo que se considerará para trabajos futuros.

En el área del modelado de los errores de la predicción eólica, el modelo conjunto mediante la regresión logística y una función de distribución, paramétrica o no paramétrica, es un aporte claro de la tesis, ya que no se ha encontrado literatura especializada donde se utilice un método semejante al presentado aquí.

Otro aporte de la tesis es la posibilidad de simular los errores de la predicción eólica para todo el intervalo de producción, desde producciones nulas hasta la potencia máxima del parque, con la CDF obtenida. En el capítulo 8 se explicará con más detalle el proceso de simulación de las trayectorias posibles de producciones eólicas mediante las funciones cópula.

ESTIMACIÓN DE LA DEPENDENCIA DE LA FUNCIÓN CÓPULA NORMAL

Resumen

En este capítulo se presentan diferentes opciones para calcular la estructura de dependencia de la función cópula Normal. Además de la estructura de dependencia estacionaria, son presentadas dos opciones adaptativas para la estimación de este parámetro, siendo más apropiado ya que considera el comportamiento estacional de la producción eólica. Una vez estimada la estructura de dependencia de la cópula normal, son formulados los modelos temporal y temporal-espacial para simular trayectorias de producción eólica.

8.1 INTRODUCCIÓN

La correlación lineal es usada comúnmente como medida de la dependencia entre variables aleatorias. Sin embargo, no es una medida general sobre la dependencia, sino que da una idea de la relación lineal entre las variables aleatorias. La correlación lineal representa correctamente la estructura de dependencia solo si el problema está representado por una distribución elíptica, sea ésta la normal multivariante o la t-student multivariante, ya que la estructura de su dependencia es lineal, [66].

El uso del coeficiente de correlación lineal como una medida general de la dependencia, en problemas donde las variables aleatorias tienen relaciones no lineales, puede llevar a conclusiones incorrectas. Por esta razón son consideradas formas alternativas para calcular la estructura de dependencia entre las variables aleatorias.

Cuando la dependencia es estimada en el dominio original de las variables aleatorias, se deben utilizar los coeficientes de correlación de Kendall o Spearman, denominados como Rank Correlation (RC).

La correlación de Kendall, τ , es una medida no paramétrica de la dependencia estadística, y mide la relación que existe entre las variables aleatorias. Sea $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ un conjunto de observaciones de dos variables aleatorias X e Y , respectivamente, de tal manera que todos los valores x_i e y_i son únicos. Para cualquier par de observaciones (x_i, y_i) y (x_j, y_j) se dice que son concordantes si $x_i > x_j$ e $y_i > y_j$ o si $x_i < x_j$ e $y_i < y_j$. Son discordantes si $x_i > x_j$ e $y_i < y_j$ o si $x_i < x_j$ e $y_i > y_j$. Si $x_i = x_j$ o $y_i = y_j$, la pareja no es

concordante ni discordante. La correlación de Kendall se calcula de la siguiente forma:

$$\tau = \frac{\text{npc} - \text{npd}}{\frac{1}{2}n(n-1)} \quad (8.1)$$

Donde npc es el número de parejas concordantes y npd es el número de parejas discordantes.

La correlación de Spearman, ρ_r , es una medida no paramétrica de la dependencia estadística, y mide la relación que existe entre las variables aleatorias X e Y , si estas pudieran ser explicadas por una función monótona, creciente o decreciente.

$$\rho_r = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2 \sum_i^n (y_i - \bar{y})^2}} \quad (8.2)$$

Los coeficientes de correlación de Spearman o Kendall son invariantes frente a transformaciones estadísticas de las variables aleatorias.

Al usar el dominio uniforme de las variables aleatorias, se elimina el efecto de las funciones marginales. Para variables aleatorias uniformemente distribuidas, la correlación lineal y la correlación de Spearman son iguales $\rho_r = \rho$, [71], por tanto, la medida de dependencia, para un grupo de variables aleatorias (X_1, X_2, \dots, X_d) , con sus respectivas función de distribución de probabilidad, F_i , se puede definir de la siguiente forma:

$$\rho_r(X_1, X_2, \dots, X_d) = \rho(F_1(X_1), F_2(X_2), \dots, F_d(X_d)) \quad (8.3)$$

La correlación de Spearman es invariante frente a transformaciones estadísticas de las variables aleatorias, [71], a partir de la ecuación (8.3) se puede obtener la estructura de dependencia de la función cópula Normal. Pero primero es necesario transformar las variables aleatorias al dominio uniforme.

8.2 TRANSFORMACIÓN AL DOMINIO UNIFORME

Tal y como se describe en la sección 4.2, los datos del caso de estudio corresponden a un parque eólico, donde se busca construir un modelo temporal para simular posibles trayectorias de producción eólica.

A partir del conjunto de datos de entrenamiento, se obtienen las función de distribución de probabilidad de los errores para cada horizonte de predicción, $F_{e,k}$. Utilizando estas funciones, se transforman al dominio uniforme los errores de la predicción para cada horizonte, $e_{t+k|t}$, pertenecientes al conjunto de datos de evaluación. Este nuevo conjunto de datos se definen de la siguiente forma:

$$\mathbf{U}_t = (U_1, \dots, U_h)^T = (F_{e,1}(e_{t+1|t}), \dots, F_{e,d}(e_{t+h|t}))^T \quad (8.4)$$

Donde la notación $t + k|t$ se refiere a las predicciones hechas para la hora $t + k$ en la hora t , y $k = [1, 2, \dots, h]$ se refiere al horizonte de predicción.

El conjunto de datos de entrenamiento es utilizado para estimar las funciones de densidad. Luego, el conjunto de datos de evaluación es convertido al dominio uniforme, que se denomina como \mathbf{U}_t . En este caso se adoptó la forma de funcionamiento de una aplicación en tiempo real, donde las nuevas observaciones de la potencia producida pasan a formar parte del conjunto de datos de entrenamiento del modelo.

Los métodos descritos en la sección 3.4 para estimar los parámetros de las funciones cópula (entre ellos, la estructura de dependencia) se orientan a modelos estacionarios, por esta razón son descartados y se busca un método de estimación que permita ponderar los cambios de la variable de entrada a medida que éstos se integran en los datos de entrenamiento. El objetivo es que la estructura de dependencia modele los fenómenos estacionales, las variaciones de la propia incertidumbre de las variables de entrada, los cambios en la precisión del modelo de predicción de potencia, la instalación de nuevos parques eólicos en la región y la sustitución de aerogeneradores en algún parque. En cualquiera de los casos, esto implicará que la estructura de dependencia debe ser actualizada y deberá ajustarse los cambios de la variable de entrada.

8.3 CÁLCULO ESTACIONARIO

Como primera aproximación se presenta el cálculo estacionario para la matriz de covarianza:

$$\Sigma_t = (\mathbf{u}_t \mathbf{u}_t^T) \quad (8.5)$$

Esta opción tiene el defecto de no ser adaptativa frente a los datos que alimentan el modelo. Cuando el conjunto de datos almacenados o históricos es considerablemente grande, hay un escaso efecto de los datos más recientes.

En otras palabras, la capacidad de adaptación es muy baja respecto a la entrada de nuevas observaciones. Por esta razón no es posible modelar correctamente la dinámica de la dependencia a medida que llegan datos nuevos.

8.4 ESTIMACIÓN RECURSIVA CON FACTOR DE OLVIDO EXPONENCIAL

La formulación estacionaria es modificada para introducir un factor de olvido (ecuación (8.6)) que pondera las observaciones más recientes y les da la mayor relevancia; a medida que las observaciones se

hacen antiguas respecto al presente, éstas pierden relevancia; en la sección 6.3.1 es explicado este factor. En [72, 76] los autores utilizan esta aproximación para hacer una estimación adaptativa de la estructura de dependencia.

$$\Sigma_t = \lambda \Sigma_{t-1} + (1 - \lambda)(\mathbf{u}_t \mathbf{u}_t^T) \quad (8.6)$$

Donde λ es el factor de olvido en la estimación. El factor de olvido puede tomar valores dentro del intervalo $0 < \lambda < 1$, de modo que un valor pequeño de λ hará que los datos más actuales tengan una mayor influencia en el valor estimado. En el caso contrario, si el valor de λ es cercano a 1, se asemejará a la formulación estacionaria, donde todos los datos tienen la misma ponderación.

El problema de este tipo de formulación es que necesita un conjunto de datos de entrenamiento lo suficientemente grande para que el valor final no se vea afectado por una inicialización inadecuada. En [76] se sugiere inicializar Σ_0 con la matriz identidad, lo cual podría provocar una estimación incorrecta cuando se cuenta con pocos datos en el conjunto de entrenamiento, ya que persistiría la influencia del valor inicial. Otro inconveniente de este tipo de procedimientos es la posibilidad de que la matriz obtenida no sea semidefinida positiva y simétrica, que es un requisito para la matriz que representa la estructura de dependencia en los modelos que utilizan la familia de funciones cópula elípticas, [71].

8.5 ESTIMACIÓN RECURSIVA CON MEDIA MÓVIL PONDERADA EXPONENCIALMENTE

El método de estimación recursiva de media móvil ponderada exponencialmente, conocido en inglés como Exponentially Weighted Moving Average (EWMA), garantiza una correcta estimación, incluso desde las primeras iteraciones, ya que cada observación puede ser ponderada respecto a la cantidad de observaciones pasadas o que se encuentran en el conjunto de datos de entrenamiento, [89, 106]. Otra de las ventajas de este método es que no necesita ser inicializado, por lo que se pueden tener estimaciones correctas desde las primeras iteraciones.

En este caso el factor de olvido debe garantizar la correcta adaptación a los datos. La memoria asintótica tendrá que ser lo suficientemente pequeña como para detectar con adecuada rapidez los cambios en la estructura de dependencia de los datos entrantes, y lo suficientemente elevada para evitar una gran variabilidad en los valores estimados.

Cada vez que en un instante t llega un nuevo vector \mathbf{U}_t , la matriz de covarianza se actualiza de la siguiente forma:

$$\hat{\Sigma}_{t|t-k} = \frac{\hat{\mathbf{U}}_{t|t-k}}{N_t}, \quad (8.7)$$

Donde:

$$\hat{\mathbf{U}}_{t|t-k} = \left(\mathbf{u}_{t|t-k} \mathbf{u}_{t|t-k}^T + \lambda \hat{\mathbf{U}}_{t|t-k-1} \right) \quad (8.8)$$

$$N_t = 1 + \lambda N_{t-1} \quad (8.9)$$

La mayoría de aplicaciones utilizan un factor de olvido que típicamente se encuentra dentro del rango $0,995 \leq \lambda \leq 0,999$, [37, 89].

8.6 ESTRUCTURA DE DEPENDENCIA PARA LA CÓPULA NORMAL

En la figura 8.1 son presentados ejemplos del cálculo de las tres alternativas para la estructura de la dependencia (matriz de covarianza) para el caso de un modelo temporal de un parque eólico, donde se obtiene la dependencia entre los diferentes horizontes. El factor de olvido se ajusta a $\lambda = 0,999$, un valor utilizado para este tipo de aplicaciones [37, 89, 106], tanto para la estimación con factor de olvido exponencial como para la media móvil ponderada exponencialmente.

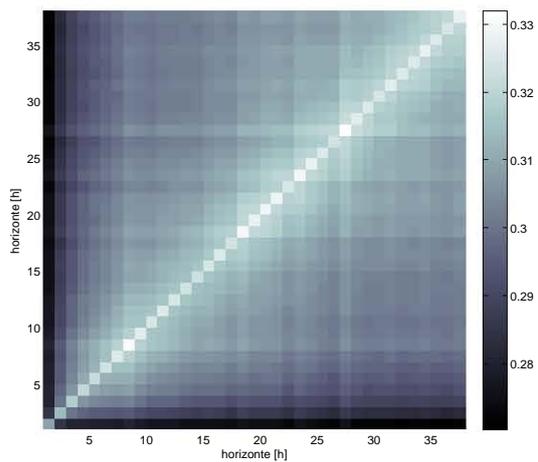
Las estimaciones de la estructura de dependencia, utilizando un factor de olvido y la EWMA, tienden a ser casi iguales cuando se han procesado muchas observaciones, tal como se ve en las figuras 8.1b y 8.1b. Sin embargo, se utilizará la EWMA, ya que garantiza una correcta estimación desde las primeras iteraciones.

8.7 MODELO TEMPORAL

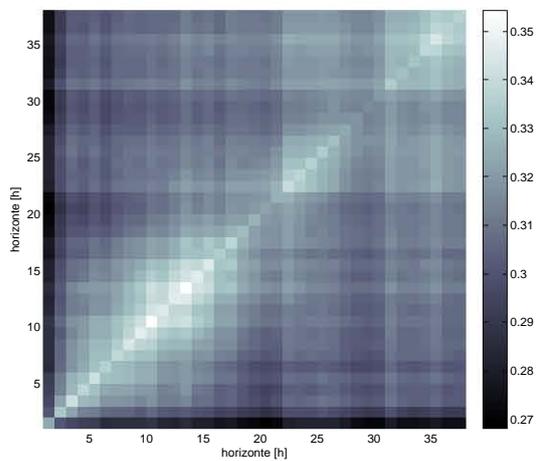
Un modelo temporal para un parque eólico se refiere a un modelo multivariate que simulará el comportamiento de los errores en la predicción a futuro, y a su vez, también simulará las trayectorias posibles de la producción eólica, en este caso utilizando la cópula normal.

Para construir el modelo con la cópula normal, se debe tener en el instante t un conjunto de datos de entrenamiento, con el que se obtienen las función de distribución de probabilidad univariantes por cada horizonte de predicción $F_k(e_{t+k|t}|\hat{p}_{t+k|t})$ y la estructura de dependencia Σ_t .

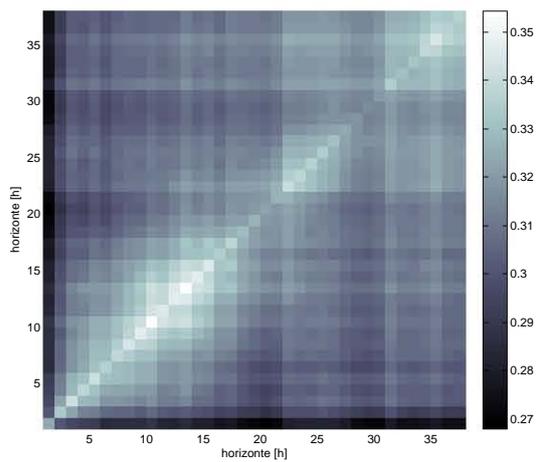
A continuación se describen los pasos necesarios para construir el modelo temporal para un parque eólico por medio de la cópula normal:



(a) Estacionaria.



(b) Factor de olvido exponencial.



(c) Media móvil ponderada exponencialmente.

Figura 8.1: Matriz de covarianzas, diferentes técnicas de estimación.

1. Datos de entrada para el modelo temporal de un parque eólico.
 - Registros históricos de las producciones y predicciones a diferentes horizontes temporales.
 - Calculo de los errores en la predicción, ecuación (4.3).
2. Construcción del modelo basado en la cópula normal.
 - Función marginal para los errores en la predicción.
 - Modelo estimado pero medio de las funciones Kernel y la regresión logística, para cada horizonte de predicción, $F_k(e_{t+k|t}|\hat{p}_{t+k|t})$, capítulo 7.
 - Estructura de dependencia para los errores en la predicción, Σ_t .
 - Conversión de los errores en la predicción al dominio uniforme, ecuación (8.4).
 - Calculo de la matriz de covarianza por medio del método conocido como Exponentially Weighted Moving Average (EWMA), ecuación (8.7).

8.7.1 Simulación de trayectorias

En el instante t se cuenta con un conjunto de datos de entrenamiento, con el que ha sido obtenido el modelo temporal utilizando la cópula normal. Además se cuenta con la predicción de la potencia eólica futura, $\hat{p}_{t+h|t}$, para los horizontes $k = 1, \dots, 38$ horas (se toma como referencia la regulación española, donde los agentes deben informar al Operador del Mercado (OM) las estimación de la producción de un día (24 horas) 14 horas antes, [5]). Como ejemplo, se simulan 10000 trayectorias de los posibles errores en la predicción, $e_{t+k|t}^{(s)}$.

Las trayectorias simuladas de los errores en la predicción eólica para el período futuro, $\mathbf{p}_{t+k|t}^{(s)}$, son obtenidas por medido de la cópula normal.

$$\mathbf{e}_{t+k|t}^{(s)} = \left[e_{t+1|t}^{(s)}, \dots, e_{t+h|t}^{(s)} \right] = C_{\Sigma_t} \left(F_1(e_{t+1|t}|\hat{p}_{t+1|t}), \dots, F_h(e_{t+h|t}|\hat{p}_{t+h|t}) \right) \quad (8.10)$$

Sumando la predicción a cada una de las posibles trayectorias simuladas se obtienen las posibles trayectorias de producción eólica.

$$\mathbf{p}_{t+k|t}^{(s)} = \mathbf{e}_{t+k|t}^{(s)} + \hat{\mathbf{p}}_{t+k|t} = \left[e_{t+1|t}^{(s)} + \hat{p}_{t+1|t}^{(s)}, \dots, e_{t+h|t}^{(s)} + \hat{p}_{t+h|t}^{(s)} \right] \quad (8.11)$$

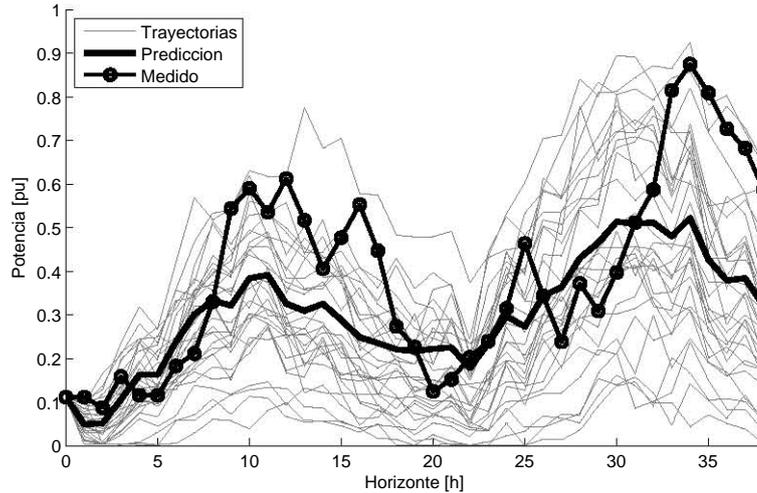


Figura 8.2: Trayectorias simuladas de producción eólica, modelo temporal.

A continuación se describen los pasos necesarios para simular trayectorias de producción eólica utilizando el modelo temporal para un parque eólico construido por medio de la cópula normal:

1. Con el modelo temporal de un parque eólico, son simuladas las trayectorias de los errores en la predicción eólica para el período futuro, condicionadas a la potencia prevista $\hat{p}_{t+h|t}$, ecuación (8.10).
2. Las trayectorias de producción eólica son obtenidas al sumar la predicción a cada una de las trayectorias simuladas, ecuación (8.11).

Con este conjunto de simulaciones es posible cuantificar la incertidumbre de la producción. El método propuesto para obtener las trayectorias simuladas es rápido (del orden de segundos), por lo que para otro tipo de análisis el número de escenarios podría ser incrementado. En la siguiente parte de la tesis serán utilizadas las trayectorias simuladas para obtener la operación óptima de una central hidroeléctrica reversible que cubre los desvíos en la producción de un productor eólico.

Basados en las trayectorias simuladas, se calculan los intervalos de la predicción y los posibles desvíos de energía. En la figura 8.2 son dibujadas sólo treinta del total de las trayectorias simuladas, $\mathbf{p}_{t+k|t}^{(s)}$, la predicción de la potencia eólica $\hat{p}_{t+k|t}$, y la producción real, p_{t+k} . Cada una de las trayectorias se modela de acuerdo a la estructura de dependencia, donde cada punto de la trayectoria depende de sus valores anteriores; además están condicionadas por el valor de la predicción $\hat{p}_{t+k|t}$ y el horizonte k .

En la figura 8.3 se representan con bandas los intervalos de predicción obtenidos a partir de todas las trayectorias simuladas. Adicional-

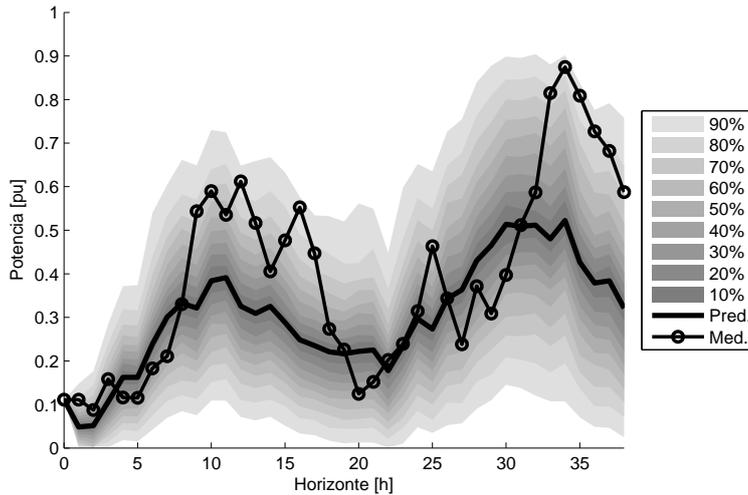


Figura 8.3: Intervalos de predicción basados en la simulación de trayectorias, modelo temporal.

mente se dibuja la previsión de potencia $\hat{p}_{t+h|t}$ y la producción real p_{t+k} . Cada banda representa la probabilidad que tiene la producción de potencia eólica futura de estar dentro de ésta. La banda de mayor área corresponde al 90 % y sobre ésta, conteniéndola, está la banda del 80 %, un poco más estrecha y oscura. Continúa de esta forma hasta llegar a la banda del 10 %, la más estrecha y oscura de todas ellas. Esto quiere decir que cuanto menor sea el área de la banda, menor será la probabilidad de que la producción de potencia eólica esté dentro de ésta.

8.7.2 Desvíos de energía

Con las trayectorias simuladas es posible estimar los desvíos en términos de energía. A partir de esta información se puede obtener la cantidad de energía necesaria para cubrir los desvíos entre el programa previsto de un parque eólico y la energía que es finalmente generada. Cuando no se cuenta con una herramienta para estimar esta información la cantidad de energía puede ser dimensionada o calculada de forma heurística, con la intención de evitar problemas en la operación o en el mercado. Sin embargo, estas medidas conservadoras hacen que los recursos energéticos no sean aprovechados de una forma adecuada y, en algunos casos, se desperdician.

Para estimar el desvío de energía respecto a la previsión, primero se obtienen las magnitudes de los desvíos y la frecuencia con que estos ocurren. Luego son estimados los desvíos como la integral de la diferencia entre la predicción de la potencia eólica y cada uno de las trayectorias simuladas (ecuación (8.12)). Para este cálculo son consideradas solamente las últimas 24 horas del horizonte simulado ($k_1 = 15$

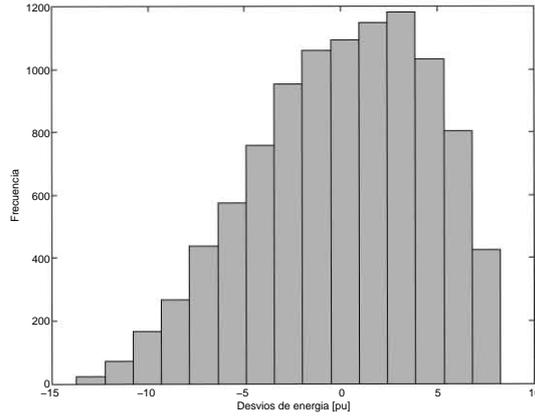


Figura 8.4: Desvíos de energía para un período de 24h

y $k_2 = 38$), tomando como referencia la regulación del mercado en España, donde son realizadas las ofertas. La figura 8.4 muestra el histograma con los posibles desvíos de energía. Con esta información es posible estimar la cantidad de energía a reservar en función de las trayectorias simuladas para un día de operación de un parque eólico.

$$E^{(s)} = \int_{k_1}^{k_2} \left(p_{t+k|t}^{(s)} - \hat{p}_{t+k|t} \right) dk \quad \forall s \quad (8.12)$$

Esta información será utilizada para obtener la operación óptima de una central hidroeléctrica reversible que cubra los desvíos en la producción de un productor eólico.

8.8 MODELO ESPACIAL Y TEMPORAL

La formulación del problema temporal de un parque eólico se extiende para modelar el comportamiento de varios parques situados en una misma región, ya que su producción guarda dependencia debido al viento.

Las trayectorias simuladas de producción de potencia eólica para el período futuro, $p_{t+h|t}^{(s,d)}$, son obtenidas usando la cópula normal, la cual se representa de la siguiente forma:

$$\begin{aligned} p_{t+k|t}^{(s,d)} &= e_{t+k|t}^{(s,d)} + \hat{p}_{t+k|t}^{(d)} = \\ &C_{\Sigma_t} \left(F_1^{(1)}(e_{t+1|t} | \hat{p}_{t+1|t}), \dots, F_h^{(1)}(e_{t+h|t} | \hat{p}_{t+h|t}), \dots, \right. \\ &\quad \left. F_1^{(q)}(e_{t+1|t} | \hat{p}_{t+1|t}), \dots, F_h^{(q)}(e_{t+h|t} | \hat{p}_{t+h|t}) \right) + \hat{p}_{t+k|t}^{(d)} \quad (8.13) \end{aligned}$$

Donde el superíndice (s, d) se refiere a que son simulaciones y $d = [1, 2, \dots, q]$ es el número del parque del modelo espacial-temporal.

A continuación se describen los pasos necesarios para construir el modelo temporal-espacial de un grupo de parques eólicos por medio de la cópula normal y la simulación de trayectorias de producción:

1. Datos de entrada para el modelo temporal-espacial de un grupo de parques eólicos.
 - Registros históricos de las producciones y predicciones a diferentes horizontes temporales de cada parque.
 - Calculo de los errores en la predicción, ecuación (4.3).
2. Construcción del modelo basado en la cópula normal.
 - Función marginal para los errores en la predicción.
 - Modelo estimado pero medio de las funciones Kernel y la regresión logística, para cada horizonte de predicción y parque, $F_k^{(d)}(e_{t+k|t}|\hat{p}_{t+k|t})$, capítulo 7.
 - Estructura de dependencia para los errores en la predicción y cada parque, Σ_t .
 - Conversión de los errores en la predicción al dominio uniforme de cada parque, ecuación (8.4).
 - Calculo de la matriz de covarianza por medio del método conocido como Exponentially Weighted Moving Average (EWMA), ecuación (8.7).
3. Simulación de trayectorias de producción futura.
 - Con el modelo temporal-espacial de un grupo de parques eólicos, son simuladas las trayectorias de los errores en la predicción eólica para el período futuro, condicionadas a la potencia prevista $\hat{p}_{t+k|t}^{(d)}$.
 - Las trayectorias de producción eólica son obtenidas al sumar la predicción a cada una de las trayectorias simuladas, ecuación (8.13).

En la figura 8.5 se presenta un ejemplo de la estructura de dependencia de dos parques eólicos, y se puede ver que es simplemente una extensión del modelo temporal de uno de ellos. Tomando como referencia el modelo temporal de un parque, se unieron los datos de dos parques para crear un conjunto donde las primeras $[1, \dots, h]$ columnas corresponden a los datos de un parque para todos sus horizontes; las siguientes $[h + 1, \dots, 2 \cdot h]$ columnas corresponden al segundo parque. Este modelo puede ser ampliado para incluir más parques añadiendo las columnas respectivas.

En este ejemplo los parques eólicos están separados una distancia de 5,1 [km] y están situados en el noroeste de España. Los datos de las predicciones y producciones son obtenidos con un sistema de predicción a corto plazo. Esta herramienta obtiene la predicción por medio de modelos estadísticos que utilizan como datos de entrada predicciones meteorológicas de la velocidad y dirección del viento, además de las telemidas de potencia generadas del parque. Una

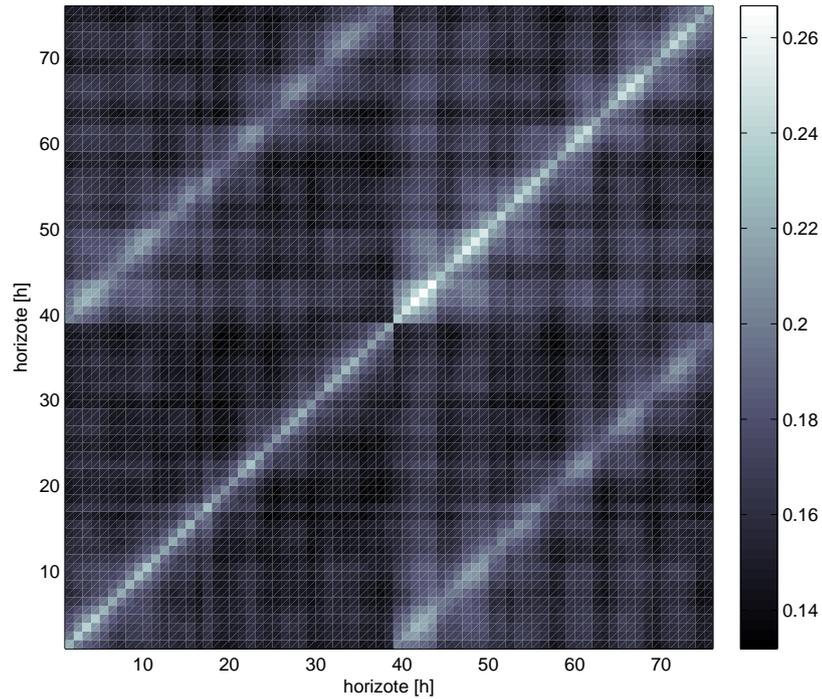


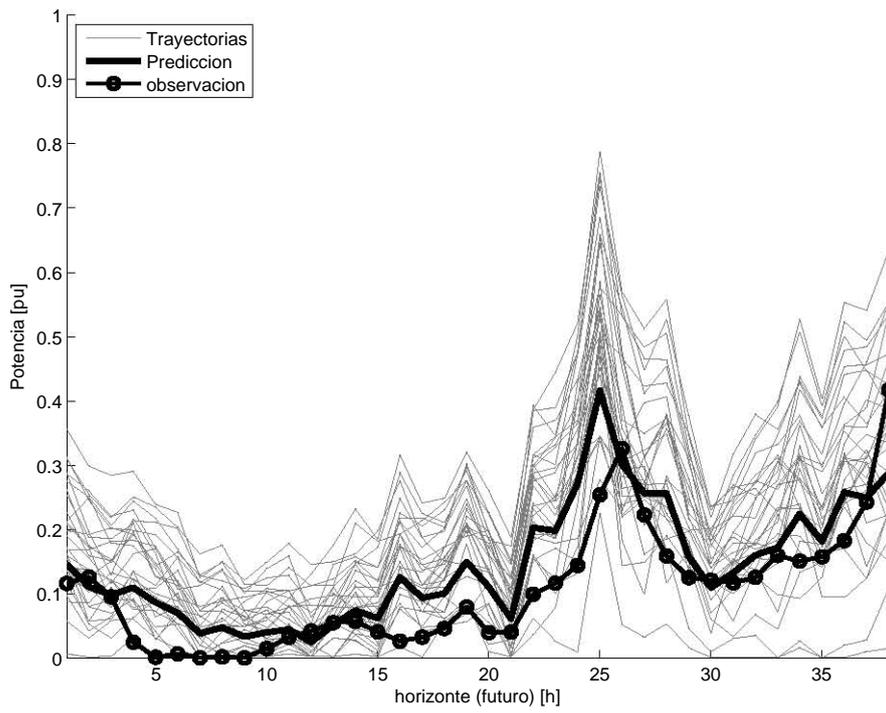
Figura 8.5: Estructura de dependencia modelo espacial-temporal

descripción más detallada de los algoritmos de esta herramienta de predicción se encuentra en [88, 90].

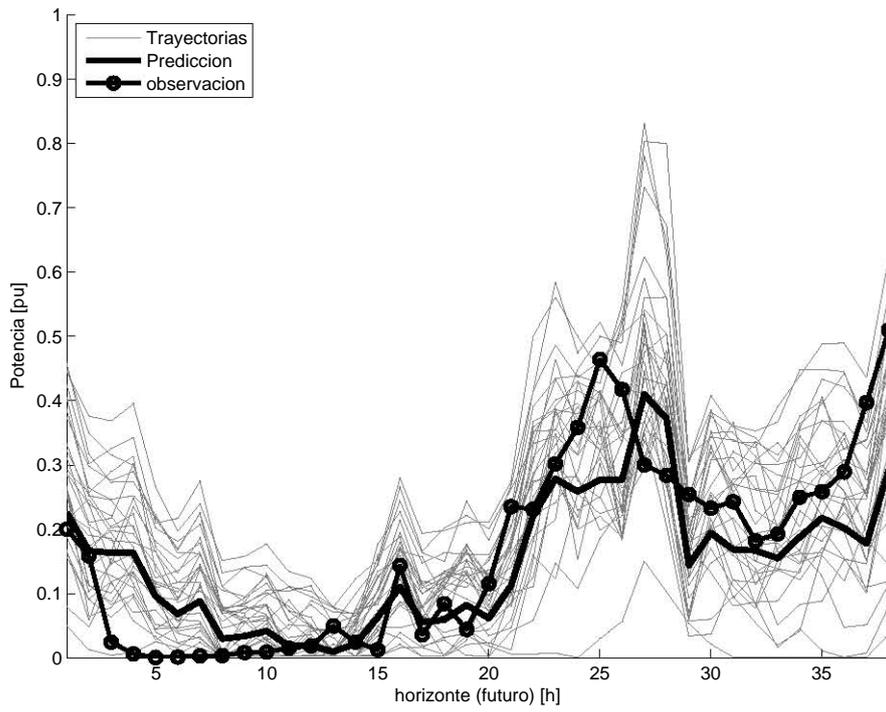
Para el modelo temporal-espacial se cuenta con 2812 medidas de producción eólica y sus respectivas predicciones para cada uno de los parques.

En la figura 8.6 son presentados sólo treinta del total de las trayectorias simuladas para un ejemplo de un modelo espacial-temporal de dos parques ($\mathbf{p}_{t+h|t}^{(s, d)}$), la predicción de la potencia eólica de cada parque ($\hat{p}_{t+h|t}$) y la producción real de cada uno (p_{t+h}). Cada trayectoria se modela de acuerdo a la estructura de dependencia, donde cada punto de una trayectoria depende de sus valores anteriores y está condicionado por el valor de la predicción $\hat{p}_{t+h|t}$ y el horizonte k . La diferencia del modelo temporal con el modelo espacial-temporal, radica en la estimación de la estructura de dependencia, que relaciona los errores de cada parque.

En la figura 8.7 encontramos los intervalos de predicción obtenidos a partir de todas las trayectorias simuladas de cada parque (representadas como bandas). Adicionalmente es dibujada la previsión de potencia $\hat{p}_{t+h|t}$ y la producción real p_{t+h} .

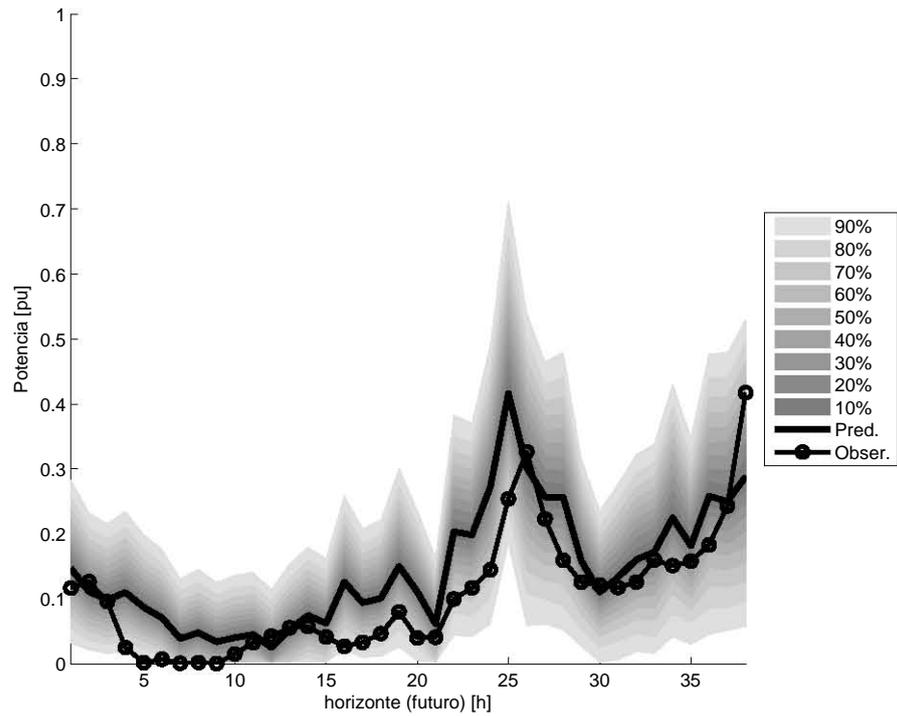


(a) Parque 1.

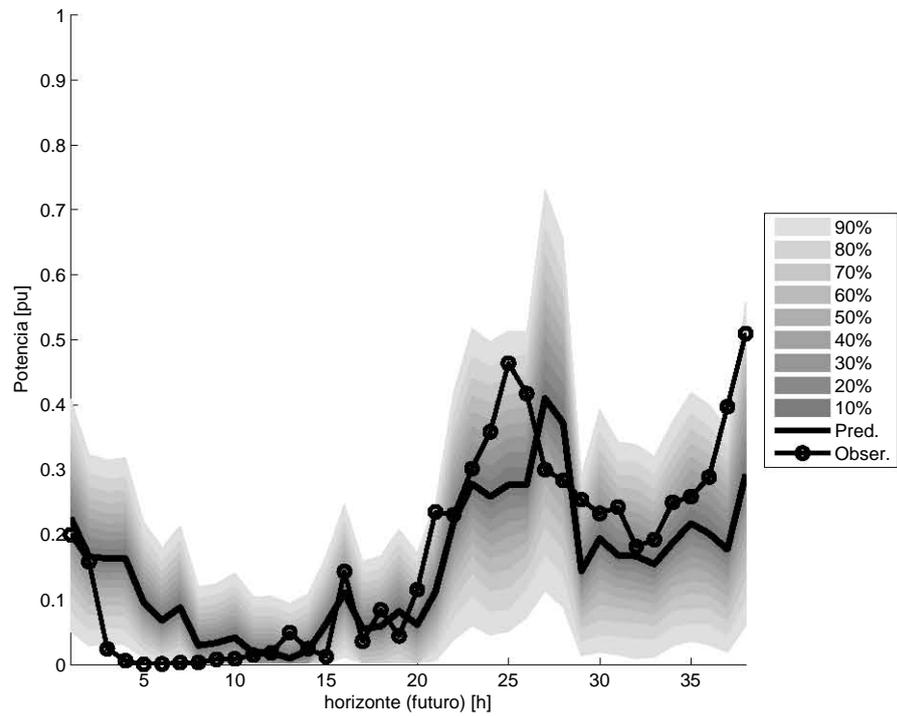


(b) Parque 2.

Figura 8.6: Trayectorias simuladas de producción eólica, modelo temporal-espacial.



(a) Parque 1.



(b) Parque 2.

Figura 8.7: Intervalos de predicción basados en la simulación de trayectorias, modelo temporal-espacial.

8.9 CONCLUSIONES

El método presentado permite la estimación de los posibles errores en la predicción de potencia eólica a través de trayectorias simuladas por medio de la cópula normal.

La principal ventaja del método es que cada punto en las trayectorias depende de los valores de los puntos previos y del valor de la predicción de la potencia eólica, recreando de esta forma la naturaleza de la variable original. Esto hace posible modelar el comportamiento no lineal y la asimetría de la producción de potencia eólica.

El cálculo de las trayectorias simuladas es muy prometedor, ya que permite la obtención de los posibles desvíos de energía y los intervalos de predicción, facilitando la comprensión del comportamiento de la producción de la energía eólica y la relación con su predicción.

La información conseguida es útil para los centros de control de energías renovables y para calcular el funcionamiento óptimo de los sistemas con capacidad de almacenamiento, además de servir para resolver otros problemas de la operación de sistemas eléctricos.

En la siguiente parte del documento son presentados dos tipos de problemas en sistemas de energía eléctrica, donde se utilizan las trayectorias simuladas y los intervalos de predicción.

Parte III

APLICACIONES A SISTEMAS ELÉCTRICOS

En esta parte de la tesis se presentan casos en los que se busca mejorar la integración de la generación eólica, por medio de la operación conjunta de centrales hidroeléctricas reversibles. Los casos presentados se modelan como problemas de optimización, donde la incertidumbre de la predicción eólica es utilizada como dato de entrada en los modelos.

En el primer caso, se presenta un problema de optimización donde se busca maximizar los ingresos de un productor eólico y reducir los desvíos en la producción mediante la operación conjunta con una central hidroeléctrica reversible. Se analizan los costos y los ingresos de los programas propuestos, al operar de forma independiente y conjunta.

Finalmente se representan diversas medidas de compensación por parte del Operador del Sistema, modeladas como problemas de optimización. En estos casos, el modelo del sistema eléctrico es representado explícitamente y se considera la generación de un grupo de parques eólicos. Se analizarán los resultados para determinar los aspectos positivos y negativos de las alternativas presentadas.

OPERACIÓN CONJUNTA DE UNA CENTRAL HIDROELÉCTRICA REVERSIBLE Y UN PARQUE EÓLICO

Resumen

En este capítulo se presenta y desarrolla una propuesta para el uso del almacenamiento de energía por medio de una central hidroeléctrica reversible. El objetivo es maximizar los beneficios al mitigar los desvíos de un productor eólico debido a los errores de la predicción de la potencia eólica.

Primero se plantea un problema para obtener el programa óptimo que maximiza los beneficios de una central hidroeléctrica reversible, que funciona de forma independiente. Luego el problema original es modificado para introducir los posibles desvíos de potencia y las reservas necesarias de energía, para asegurar el cumplimiento del programa del productor eólico que ha sido informado con antelación. El programa resultante es puesto a prueba analizando la producción real de la central.

Finalmente son calculados y analizados los ingresos de los programas propuestos para los agentes, al operar de forma independiente y conjunta. Para este análisis son considerados diferentes perfiles típicos de precios y producciones.

9.1 INTRODUCCIÓN

En el mercado español (ver descripción en el anexo [A](#)) los productores eólicos con potencias instaladas superiores o iguales a los 10 MW y menores a 50 MW, pueden realizar ofertas al mercado. Generalmente, esta oferta se hace para el día siguiente, y debe ser en bloques horarios (desde la hora 00:00 h hasta la hora 24:00 h). El mercado cierra el plazo para la presentación de las ofertas a las 10:00 h del día anterior, lo que agrega 14 horas de anticipación a las 24 horas de programación. Debido a esto, los productores eólicos están expuestos a tener errores en su programa ofertado.

Según [36], los errores pueden estar entre el 30-50 % de la energía ofertada, con respecto a la energía real al momento de despacharla. Este error se reduce acudiendo a los mercados intradiarios, para recomprar o vender la energía, quedando entre el 15-25 %. Finalmente, los desvíos entre la energía programada y la energía entregada al sistema son cubiertos a través de los mercados de reserva.

La estrategia más común para un productor eólico es ofertar la energía prevista para cada hora; este dato es obtenido por medio de las herramientas de predicción de las que dispone. Sin embargo, si contara con una central hidroeléctrica reversible asociada podría almacenar energía, llevandola a un embalse para luego producirla en períodos donde se obtenga el máximo beneficio de cuerdo con la diferencia entre los ingresos en el mercado y los costos variables. Este concepto, llamado «central virtual» [14, 24, 77], tiene el objetivo de maximizar los ingresos de forma conjunta en un entorno de mercado, minimizando los desvíos debidos a los errores en la predicciones de potencia eólica. Para una central térmica, los costos variables dependen de los costos del combustible, operación y mantenimiento, mientras que para una central hidroeléctrica reversible y un parque eólico los costos son menores, ya que sus insumos son agua y viento. Para este tipo de casos los costos variables son insignificantes. Finalmente, el objetivo es maximizar los ingresos de la operación conjunta en el mercado.

9.2 DEFINICIÓN DEL PROBLEMA

Para encontrar el programa de operación óptima conjunta de una central hidroeléctrica reversible y un generador eólico se parte del planteamiento individual de la operación de una central hidroeléctrica reversible. Luego se modifica su programa original, con el objetivo de cubrir los posibles desvíos del productor eólico y, de esta forma, obtener un programa conjunto de operación. Esto significa realizar una oferta conjunta al OM para el mercado diario. La oferta conjunta solo es posible si ambos agentes, el productor eólico y la central hidroeléctrica reversible, están dentro del *Régimen Especial* del mercado de electricidad español, regulado por el Real Decreto 661/2007 de 25 de mayo.

En el presente problema se analiza la programación futura a corto plazo de la central hidroeléctrica reversible, por lo que se necesita una estimación del precio futuro de la energía eléctrica. Para este caso, se considera una predicción perfecta del precio de la energía en el mercado, y como fuente de incertidumbre solo se introduce la relativa a los errores en la predicción. La predicción perfecta del precio no es posible; sin embargo, la predicción a corto plazo del precio de la energía tiene una menor incertidumbre comparada con la producción de potencia eólica, por lo que no se considerará en este estudio [22, 36].

Los datos del parque eólico (predicción y producción eólica) que se utilizan en este problema, son los mismos que fueron presentados en el capítulo 4. Algunos casos que representan producciones típicas del conjunto de datos de evaluación son seleccionados para el estudio.

Teniendo como referencia la regulación española, los agentes presentan las ofertas al OM, que usualmente es el valor de la predicción de la producción eólica de un día (24 horas), presentada 14 horas antes [5]. Por esta razón en el planteamiento del problema de optimización solo se consideran las últimas 24 horas de programación, es decir, el mercado diario. El mercado intradiario no se analiza en el presente análisis.

Para los embalses de la central hidroeléctrica reversible se asume que no hay aportaciones externas que sean considerables o que afecten el nivel de los embalses. En el problema de optimización no se modelan los costos debidos a las paradas y arranques de la central.

9.3 OPERACIÓN ÓPTIMA DE UNA CENTRAL HIDROELÉCTRICA REVERSIBLE

El objetivo de este problema es conseguir el programa óptimo que maximice los ingresos de una central hidroeléctrica reversible que opera de forma independiente, dada la predicción de los precios de la energía para el día siguiente en el mercado diario. La central bombea agua para almacenar energía en un embalse, en períodos donde el precio de la energía en el mercado es bajo, y luego opera como un generador hidráulico, descargando el agua almacenada en períodos donde el precio de la energía es alto [18, 19, 49].

$$\text{máx} \sum_{i=1}^n (c_i \bar{P}_{Hi} - c_{Pi} \bar{P}_{Pi}) \tag{9.1}$$

sujeto a:

$$E_{i+1} = E_i + t \left(\eta_P \bar{P}_{Pi} - \frac{\bar{P}_{Hi}}{\eta_H} \right) \tag{9.2}$$

$$E_1 = E_1^{\text{esp}} \tag{9.3}$$

$$E_{n+1} = E_{n+1}^{\text{esp}} \tag{9.4}$$

$$0 \leq \bar{P}_{Hi} \leq P_H^u \tag{9.5}$$

$$\bar{P}_{Hi} \leq \eta_H \frac{E_i}{t} \tag{9.6}$$

$$0 \leq \bar{P}_{Pi} \leq P_P^u \tag{9.7}$$

$$0 \leq E_i \leq E^u \tag{9.8}$$

$$\forall \quad i = 1, \dots, n$$

Donde:

\bar{P}_{Hi}	Potencia programada, generación hidroeléctrica en el período i . [pu]
\bar{P}_{Pi}	Potencia programada, consumo bombeo en el período i . [pu]
c_i	Precio horario de la energía en el mercado diario en el período i . [€/ (pu·h)]
c_{Pi}	Precio horario de la energía cuando la central bombea en el período i . [€/ (pu·h)]
E_i	Energía útil almacenada en el embalse en el período i . [pu·h]
$E_1^{esp} \wedge E_{n+1}^{esp}$	Niveles especificados de energía almacenada en el embalse. [pu·h]
E^u	Límite superior de la energía almacenada en el embalse. [pu·h]
η_H	Eficiencia de la central hidroeléctrica. [%]
η_P	Eficiencia de la central de bombeo. [%]
P_H^u	Límite superior de operación de la central operando como generador. [pu]
P_P^u	Límite superior de operación de la central operando como bomba. [pu]
t	Duración de cada período. [h]
n	Número de períodos.

En la notación se marca la programación anticipada con una barra en la parte superior de las variables « $\bar{\cdot}$ ». La ausencia de la barra indica el valor real de la variable después de llevarse a cabo el programa de operación.

Con la función objetivo (9.1) se busca maximizar los ingresos de una central hidroeléctrica reversible que opera de forma independiente, minimizando el costo de operación durante el bombeo y maximizando los ingresos al vender la energía almacenada en períodos de precios altos. Cuando la central bombea, el precio horario de la energía será considerado como el precio marcado del mercado, más al costo interno de operación de la central.

En la ecuación (9.2) se calcula el nivel de energía almacenada para el siguiente período, el cual se modifica por la operación de bombeo o de generación. El nivel inicial del embalse es conocido, y el nivel final se ajustará de acuerdo a la necesidad del día siguiente. Para todas los problemas modelados, el nivel final será igual al nivel inicial $E_1^{esp} = E_{n+1}^{esp} = 0,8$ pu·h (ecuaciones (9.3) y (9.4) respectivamente).

La cantidad de potencia inyectada a la red está determinada por la capacidad física del generador (ecuación (9.5)). Adicionalmente, la central no puede generar más energía que la disponible en el embalse

(ecuación (9.6)). La cantidad de potencia consumida durante el bombeo también está determinada por la capacidad física de la maquina (ecuación (9.7)). Finalmente, el embalse está limitado por su capacidad física (ecuación (9.8)).

Todos los cálculos en este capítulo se realizaron con MATLAB[®], en un ordenador con procesador de doble núcleo a 3 GHz con 3 GB de memoria.

El problema de optimización definido en las ecuaciones (9.1)-(9.8) es resuelto utilizando el Método de Puntos Interiores descrito en [48].

Como caso de estudio son tomados los datos de una central hidroeléctrica reversible con la que se espera cubrir los desvíos en la predicción de un productor eólico (tabla 9.1). La capacidad nominal del parque eólico, 30 MW, se toma como potencia base para llevar los valores a por unidad.

E_u pu	P_H^u pu	P_P^u pu	η_H [%]	η_P [%]
8,00	1,09	1,34	88	92

Tabla 9.1: Datos técnicos de la central hidroeléctrica reversible.

9.3.1 Resultados

El resultado del problema definido en las ecuaciones (9.1)-(9.8) es mostrado en la figura 9.1, donde se puede apreciar la operación de la central como bomba, \bar{P}_{Pi} ; la operación de la central como generador hidráulico, \bar{P}_{Hi} ; el nivel de energía almacenada en el embalse, E_i , y el precio de la energía en el mercado según la predicción, c_i . El precio de la energía es decisivo en el resultado del programa óptimo. Un perfil diferente del precio de la energía significaría un programa óptimo diferente.

La operación obtenida comienza en la primera hora, sin realizar ninguna actuación sobre la central. Esto se debe a que está a la espera de precios bajos antes de comenzar a operar como bomba. Al comienzo, el nivel de la energía almacenada en el embalse es E_1^{esp} . Cuando los precios son lo suficientemente bajos, la central comienza a operar como bomba hasta que se llena o consigue la energía almacenada óptima, de la hora 2 al final de la hora 7, para luego vender en los períodos de precios altos. Luego, de las horas 8 a 17, la central espera hasta que los precios son lo suficientemente altos para vender la energía almacenada y así maximizar los ingresos. En el período entre la hora 18 y el final de la hora 23, la central hidroeléctrica reversible opera como generador y vende la energía almacenada al mercado, hasta que llega al valor final requerido de energía en el embalse, E_{n+1}^{esp} . En

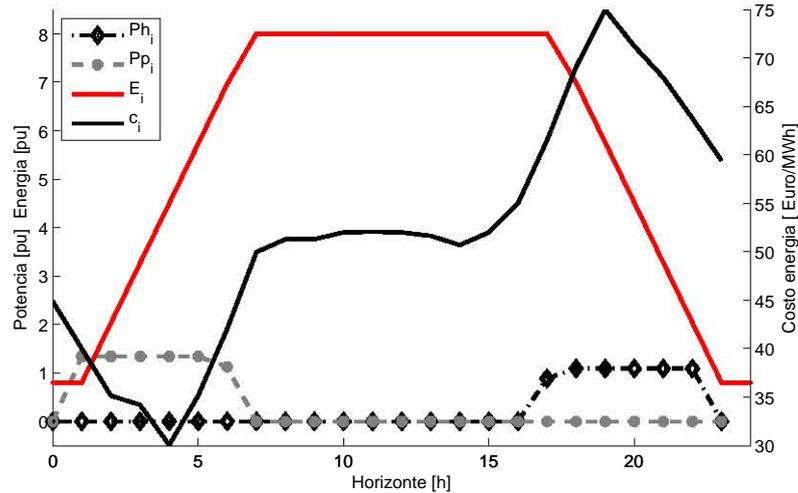


Figura 9.1: Operación óptima para una central hidroeléctrica reversible.

la última hora de programa no se realiza ninguna actuación sobre la central, y el nivel de energía en el embalse no se modifica.

9.4 OPERACIÓN CONJUNTA DE UNA CENTRAL HIDROELÉCTRICA REVERSIBLE Y UN PRODUCTOR EÓLICO

El objetivo de este problema es conseguir el programa óptimo que maximice los ingresos de una central hidroeléctrica reversible que opera de forma conjunta para cubrir los desvíos de un productor eólico, dada la predicción de los precios de la energía para el día siguiente en el mercado diario. Como en el caso anterior, la central bombea agua para almacenar energía en el embalse en períodos donde el precio de la energía en el mercado es bajo, y luego opera como un generador hidráulico, descargando el agua almacenada en períodos donde el precio de la energía es alto. Durante todo el programa de operación, la central debe ajustar su operación para cubrir los desvíos del productor eólico, reduciendo o aumentando la producción hidroeléctrica o el consumo debido al bombeo. De esta forma, desde el punto de vista del OS, el productor eólico siempre seguiría el programa de producción informado con anterioridad.

Este problema supone una adaptación a la formulación original de la operación óptima de una central hidroeléctrica reversible (ecuaciones (9.1)-(9.8)). Para el nuevo problema es necesario conocer las magnitudes de los posibles desvíos del productor eólico y la estimación de la energía necesaria para cubrir los desvíos en un intervalo determinado. Esta información no se obtiene de forma directa con las herramientas de predicción; por esta razón son utilizados los registros históricos de producción y predicción de potencia eólicos, para realizar un análisis estadístico donde se simulan posibles trayectorias

de producción para un horizonte a corto plazo. Con estas trayectorias se puede cuantificar la incertidumbre en la predicción, y de esta forma tener una estimación realista de la magnitud de los posibles desvíos de potencia y de energía para un período de corto plazo. En la parte II de este documento se desarrolla y explica el método utilizado para llegar a estos resultados. En la parte II, la predicción de potencia eólica se denominó $\hat{p}_{t+k|t}$. Para evitar confusiones de índices y notación en el planteamiento del problema de optimización, en este capítulo se denominará la misma variable como \hat{P}_{Wi} .

Sea \hat{P}_{Wi} la predicción de potencia eólica, se consideran las estimaciones de los posibles desvíos de potencia para cada hora en el futuro a corto plazo P_{Wi}^M y P_{Wi}^m , que representan los desvíos máximos y mínimos respectivamente, según los cálculos realizados en la sección 8.7.

9.4.1 Desvíos de potencia

Cuando el productor eólico genera menos potencia de la informada en su programa de operación, la central hidroeléctrica (que funciona como generador) debe reservar cierta cantidad de potencia para cubrir el desvío, por lo que la potencia máxima que generará en cada período será $\bar{P}_{Hi} \leq P_H^u - P_{Wi}^m$, donde P_{Wi}^m representa el desvío esperado de subproducción del parque eólico en el período i . Para el caso contrario, cuando el parque eólico produce más potencia de la informada en su programa de operación, la central que opera como bomba cubre este desvío consumiendo la sobreproducción. Entonces la potencia máxima que consumirá en cada período no será mayor que $\bar{P}_{Pi} \leq P_P^u - P_{Wi}^M$, donde P_{Wi}^M representa el desvío esperado de sobreproducción del parque eólico en el período i . Durante el período de inactividad se debe comportar de forma análoga, donde también será necesaria la actuación de la central para cubrir los desvíos.

Para determinar los niveles de potencia a reservar se utiliza el método propuesto en la sección 8.7, donde a partir de las posibles trayectorias de producción son determinados los valores correspondientes a los percentiles (medidas de posición no central) 5% y 95% para cada intervalo i . De esta forma se puede asociar un nivel de cobertura que, para este caso, correspondería al 90%, lo que quiere decir que se garantiza una cobertura del 90% de los casos, según los datos simulados, que a su vez están basados en los registros históricos del generador eólico.

9.4.2 Reservas de energía

Para compensar las sobreproducciones del productor eólico respecto a su predicción se debe reservar una capacidad de almacenamiento en el embalse en cada período i , especialmente en los periodos de

inactividad. Esto quiere decir que no se debe llenar el embalse en su totalidad hasta que comienza la operación como generador, ya que a partir de este momento la compensación se realiza modificando la potencia generada de la central.

También se debe reservar una cantidad de energía mínima almacenada al comienzo de la operación de la central, cuando aún está inactiva. Este mínimo de energía almacenada sirve para compensar los desvíos del productor eólico cuando éste no consigue producir la potencia informada con antelación a la acción de bombeo de la central. Después del período de generación también se debe asegurar un mínimo de energía almacenada para compensar los desvíos hasta el final del día.

Tomando como referencia el programa obtenido del problema descrito en las ecuaciones (9.1)-(9.8), se pueden definir los siguientes estados de la central: *a)* períodos de inactividad: primera hora del programa, de la hora 8 al final de la hora 17 y en la última hora del programa; *b)* período de bombeo: de la hora 2 al final de la hora 7; *c)* período de descarga: de la hora 18 al final de la hora 23. Partiendo de este programa son tomados los tiempos de cada uno de los períodos como una aproximación para calcular las reservas de energía necesarias en los periodos de inactividad, ya que no se conocen los períodos de actuación real de la central con antelación a la solución del problema. Posteriormente, las reservas de energía son calculadas considerando estos intervalos de tiempo como posibles desvíos según la simulación de trayectorias de producción, sección 8.7.2.

La cantidad de energía que se debe reservar en el período de inactividad (entre el bombeo y la descarga) se calcula para cada hora, comenzando por $E_{i=8}^R$, donde la reserva considera los desvíos posibles del período comprendido entre la hora 8 y el final de la hora 17 utilizando la ecuación (8.12), con los parámetros $k_1 = 8$ y $k_2 = 17$. Para el período $i = 9$, los parámetros de la ecuación (8.12) cambiarán de la siguiente forma: $k_1 = 9$ y $k_2 = 17$. Continuarán así hasta llegar al intervalo $i = 17$, donde los parámetros serán $k_1 = 17$ y $k_2 = 17$. Cuando la central está operando como bomba (entre la hora 2 y el final de la hora 7) la capacidad de almacenamiento de energía será ajustada al valor máximo del intervalo comprendido entre la hora 8 y el final de la hora 17. De igual forma será al comienzo del ciclo de operación de la central, cuando está inactiva (hora 1). Para el intervalo comprendido entre la hora 18 y el final de la hora 24 la central opera como generador, por lo que no es necesario considerar los límites en la capacidad de almacenamiento de energía. En este periodo se asigna un valor nulo.

De forma análoga se determina la energía mínima inicial E_1^{esp} utilizando la ecuación (8.12), donde son calculados los posibles errores de energía para el primer período de inactividad, antes de que comience la acción de bombeo. Después de la acción de bombeo se calcula

la energía mínima necesaria para cubrir los posibles desvíos del productor eólico hasta el final del día. De forma semejante se utiliza la ecuación (8.12) con los parámetros $k_1 = 8$ y $k_2 = 24$ para calcular la energía mínima que se debe almacenar en el embalse, $E_{i=8}^S$. De forma semejante, para el período $i = 9$ los parámetros de la ecuación (8.12) cambian de la siguiente forma: $k_1 = 9$ y $k_2 = 24$; continúan así hasta llegar al intervalo $i = 24$, donde los parámetros son $k_1 = 24$ y $k_2 = 24$.

Al igual que en caso de los desvíos de potencia, el nivel de cobertura es ajustado al 90 % de los casos simulados utilizando las medidas de posición no central o percentiles. Esto aplica tanto para los desvíos positivos (de sobreproducción, donde es necesario reservar una capacidad de almacenamiento de energía) como para los negativos (de subproducción, donde se debe mantener una cantidad de energía mínima almacenada).

9.4.3 Resultados de la operación conjunta

El problema para obtener el programa de operación conjunta de una central hidroeléctrica reversible que cubre los desvíos de un productor eólico se describe en las ecuaciones (9.9)-(9.16). Para este problema, la potencia del productor eólico, así como los desvíos de potencia y de energía son considerados en por unidad, tomando como potencia base la capacidad nominal del parque eólico, 30 MW.

$$\text{máx} \sum_{i=1}^n (c_i \bar{P}_{Hi} - c_{Pi} \bar{P}_{Pi}) \quad (9.9)$$

sujeto a:

$$E_{i+1} = E_i + t \left(\eta_P \bar{P}_{Pi} - \frac{\bar{P}_{Hi}}{\eta_H} \right) \quad (9.10)$$

$$E_1 = E_1^{\text{esp}} \quad (9.11)$$

$$E_{n+1} = E_{n+1}^{\text{esp}} \quad (9.12)$$

$$0 \leq \bar{P}_{Hi} \leq (P_H^u - P_{Wi}^m) \quad (9.13)$$

$$\bar{P}_{Hi} \leq \eta_H \frac{E_i}{t} \quad (9.14)$$

$$0 \leq \bar{P}_{Pi} \leq (P_P^u - P_{Wi}^M) \quad (9.15)$$

$$E_i^S \leq E_i \leq (E^u - E_i^R) \quad (9.16)$$

$$\forall \quad i = 1, \dots, n$$

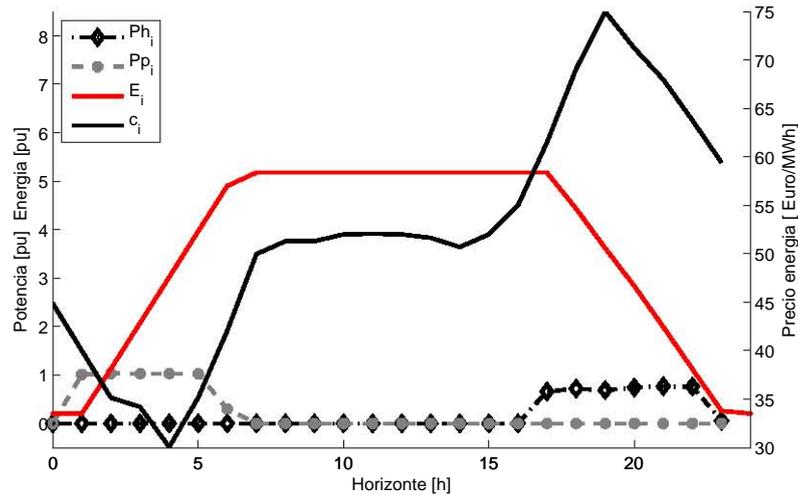


Figura 9.2: Operación óptima para una central hidroeléctrica reversible que cubre los errores de un productor eólico.

Donde:

$P_{Wi}^M \wedge P_{Wi}^m$ Desvíos esperados de sobre y sub producción del parque eólico en el período i . [pu]

E_i^R Capacidad de almacenamiento de energía adicional a reservar en el período i . [pu·h]

E_i^S Capacidad de energía almacenada mínima que se debe mantener en el período i . [pu·h]

En la figura 9.2 se puede apreciar la operación de la central como bomba (\bar{P}_{Hi}), la operación de la central como generador hidráulico (\bar{P}_{Hi}), el nivel de energía almacenada en el embalse (E_i) y el precio de la energía en el mercado según la predicción (c_i). Al comparar con la figura 9.1, se observa una reducción en la energía almacenada en el embalse y en los niveles a los que opera la central como bomba y como generador. Esto se debe a las restricciones impuestas por los desvíos esperados del productor eólico, y a la necesidad de reservar una cierta capacidad de energía en el embalse.

El programa de operación conseguido cubrirá el 90 % de los casos posibles, según las simulaciones realizadas en la sección 8.7. El nuevo programa da la posibilidad al productor eólico de controlar el riesgo desde el punto de vista técnico y económico; esto representa una ventaja estratégica frente a otros competidores, disminuyendo el riesgo al que está expuesto.

En los cálculos hechos, no se están cubriendo todos los posibles casos sorteados en las simulaciones, ya que se tomaron el 90 % de los mismos. Sin embargo, si se aumentara la cobertura hasta un 95 % se necesitaría reservar un 18 % más de espacio en el embalse para los errores de sobreproducción, y un 26 % más de energía almacenada para los errores de subproducción, limitando así la operación de la

central; incluso podría hacer inviable el programa de cobertura, ya que las dimensiones actuales del embalse no serían suficientes para cubrir los posibles desvíos.

9.5 OPERACIÓN REAL DE LA CENTRAL HIDROELÉCTRICA REVERSIBLE CUBRIENDO LOS DESVÍOS DEL PRODUCTOR EÓLICO

La medida propuesta se evalúa usando los datos reales de la producción eólica para el día considerado y adaptando la producción de la central hidroeléctrica reversible para cubrir los desvíos.

En la figura 9.3a se puede ver la diferencia de energía almacenada resultante. Para este caso particular los desvíos de energía son predominantemente de sobreproducción. Sin embargo, también tuvieron lugar con algunos errores de subproducción. La sobreproducción fue aprovechada almacenando energía en el embalse desde el comienzo de la operación y hasta el final de la hora 5. En las horas 11 y 24 se pueden observar las operaciones de bombeo debidas a las sobreproducciones de potencia eólica. De forma contraria, hay una reducción en la operación de bombeo entre la hora 6 y el final de la hora 7, cuando ocurrieron errores de subproducción (figura 9.3b).

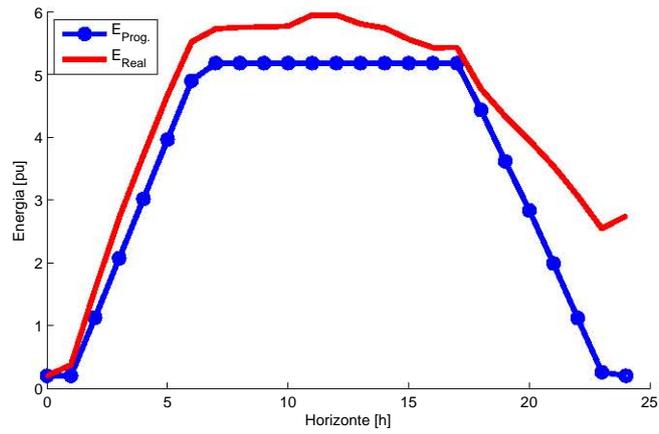
En la figura 9.3c se puede ver la adaptación en la operación de la central cuando opera como generador; en este caso, cubre los errores de subproducción entre la hora 12 y el final de la hora 16. Luego reduce la producción de energía debido a la sobreproducción eólica, en el período comprendido entre la hora 17 y el final de la hora 23.

En este caso se verifica la correcta programación de la central hidroeléctrica reversible, capaz de compensar los errores reales producidos en el horizonte programado.

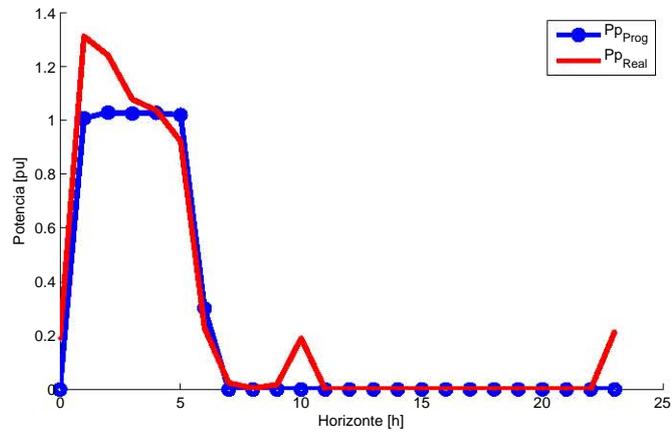
9.6 COSTOS DE LA PROGRAMACIÓN ANTICIPADA Y DE LA OPERACIÓN REAL

El costo se calcula teniendo en cuenta las dimensiones del productor eólico y de la central hidroeléctrica reversible. Tanto los precios de la energía como los de la reserva se tomaron del mercado ibérico [78] el día 01/08/2009. Las mediciones y las predicciones corresponden a la misma fecha.

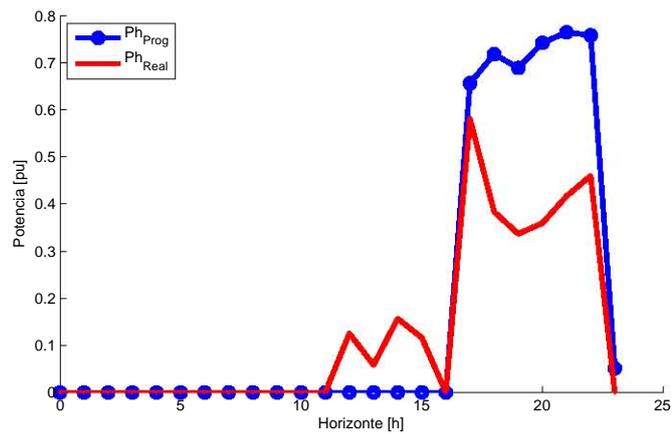
Los ingresos de la central hidroeléctrica reversible operando de forma independiente (primer caso planteado, sección 9.3) son de 4375 [€]. En el segundo caso (sección 9.4), los ingresos de la central hidroeléctrica reversible que opera de forma conjunta con el productor eólico son de 3093 [€], lo que supone un 29,3 % menos. La diferencia de estas dos cantidades es llamada «costo de oportunidad», ya que es la cantidad de dinero que deja de ganar la central hidroeléctrica reversible al asumir en su nuevo programa los posibles desvíos del productor eólico.



(a) Nivel de la energía en el embalse.



(b) Potencia consumida por el bombeo.



(c) Potencia generada.

Figura 9.3: Comparación del programa obtenido con antelación y la operación real

Los ingresos debidos al nuevo programa de la central hidroeléctrica reversible son menores debido a que no opera a su capacidad máxima. Sin embargo, los ingresos reales deben ser calculados después de la operación real de la central, a partir de la potencia generada por el productor eólico y el programa ajustado de la central hidroeléctrica reversible.

Los ingresos reales de los agentes después de la operación se calculan teniendo en cuenta el precio de reserva en el mercado. En el presente caso, estos precios son tomados del MIBEL [78], y corresponden a una proporción del precio horario de la energía del mismo día.

$$\sum_{i=1}^n (c_i \hat{P}_{Wi} + c_i d) \quad (9.17)$$

Con:

$$d = \begin{cases} +c_i^{\text{sell}} (P_{Wi} - \hat{P}_{Wi}) & \text{Si } P_{Wi} > \hat{P}_{Wi}, \\ -c_i^{\text{buy}} (\hat{P}_{Wi} - P_{Wi}) & \text{Si } P_{Wi} < \hat{P}_{Wi} \end{cases} \quad (9.18)$$

Donde:

- P_{Wi} Potencia producida por el productor eólico en el período i . [pu]
- \hat{P}_{Wi} Predicción eólica en el período i . [pu]
- c_i^{sell} Precio de venta de la energía de reserva, proporción del precio de la energía en el mercado diario en el período i . [%]
- c_i^{buy} Precio de compra de la energía de reserva, proporción del precio de la energía en el mercado diario en el período i . [%]

Finalmente, los ingresos reales de la central hidroeléctrica reversible y el productor eólico se calculan de forma conjunta a partir de la oferta única de los dos productores presentada al OM y de la producción real, considerando los costos por acción en el mercado de reservas.

$$\sum_{i=1}^n (c_i \bar{P}_{WHPi} + c_i d) \quad (9.19)$$

Con:

$$d = \begin{cases} +c_i^{\text{sell}} (P_{WHPi} - \bar{P}_{WHPi}) & \text{Si } P_{WHPi} > \bar{P}_{WHPi}, \\ -c_i^{\text{buy}} (\bar{P}_{WHPi} - P_{WHPi}) & \text{Si } P_{WHPi} < \bar{P}_{WHPi} \end{cases} \quad (9.20)$$

$$\bar{P}_{WHPi} = \hat{P}_{Wi} + \bar{P}_{Hi} - \bar{P}_{Pi} \quad (9.21)$$

$$P_{WHPi} = P_{Wi} + P_{Hi} - P_{Pi} \quad (9.22)$$

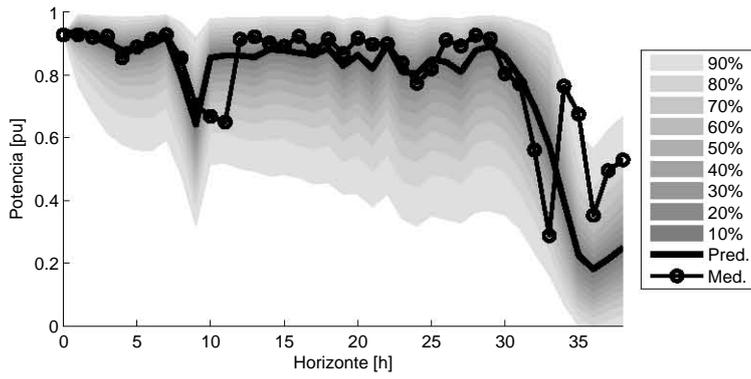
Donde:

P_{Hi} Potencia producida por la central hidroeléctrica en el período i . [pu]

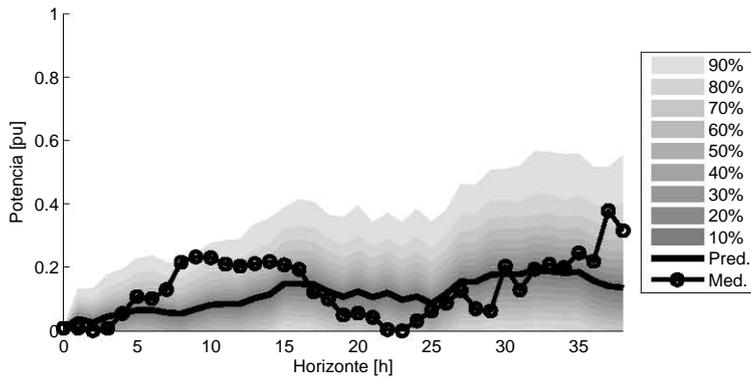
P_{Pi} Potencia consumida por la central operando como bomba en el período i . [pu]

El problema analizado en la sección 9.4 se considera un caso de producción de potencia media con un perfil de precios de un día de semana típico. Adicionalmente se analizan y compararan diferentes perfiles de precios de la energía para un sábado y un domingo, tal como se ve en la figuras 9.5a y 9.6a. Estos perfiles de precios corresponden a la misma semana del caso inicial (días 8, 10 y 11 de enero del 2009), considerada como un periodo de demanda alta, típico de invierno. Además son considerados dos casos de producción, alta y baja, tal como se ve en las figuras 9.4a y 9.4b respectivamente. Estos casos son seleccionados teniendo en cuenta solo el nivel de potencia producida durante el día y no corresponden a las fechas de los perfiles del precio de la energía. Así, se busca ilustrar las condiciones típicas. En todos los casos se compararán los ingresos antes y después de la operación, tanto de forma independiente como conjunta. Los resultados son presentados en al tabla 9.2.

Con el mismo conjunto de casos de producción eólica (medio, alto y bajo) se analizan los ingresos antes y después de la operación, para un conjunto diferente de precios de la energía en una semana típica de verano (un período de demanda baja). Los datos corresponden a los días 13, 15 y 16 de agosto de 2009 (día de semana, sábado y domingo respectivamente), y son presentados en las figuras 9.5b y 9.6b. Los ingresos se comparan antes y después de la operación real, de forma independiente y conjunta. Los resultados se muestran en la tabla 9.3.

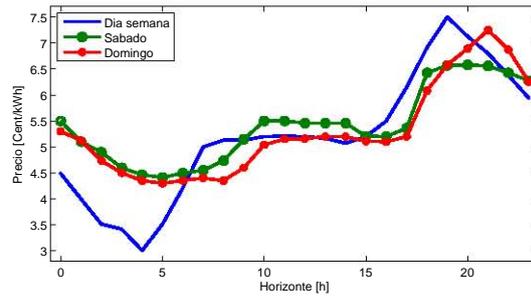


(a) Caso de producción alta.

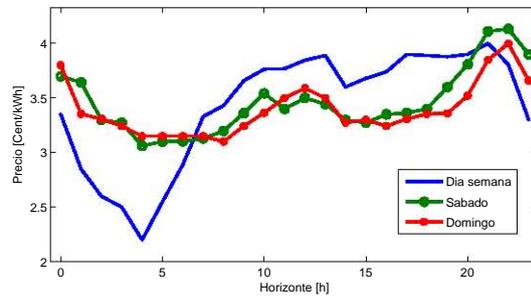


(b) Caso de producción baja.

Figura 9.4: Intervalos de predicción basados en la simulación de trayectorias

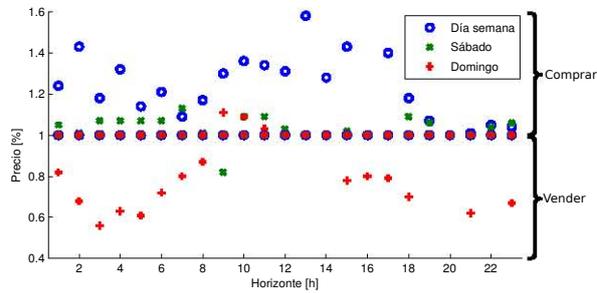


(a) Invierno.

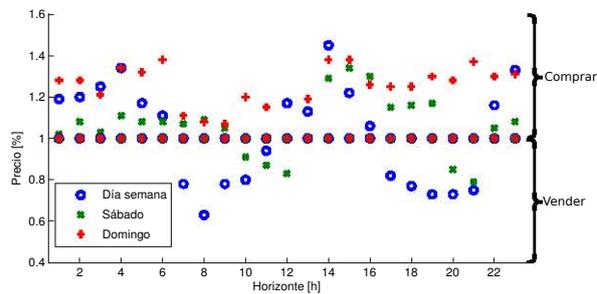


(b) Verano.

Figura 9.5: Precio de la energía para días típicos en una semana.



(a) Invierno.



(b) Verano.

Figura 9.6: Precio de los desvíos de la energía para días típicos en una semana.

Día	Semana			Sábado			Domingo		
	Media	Alta	Baja	Media	Alta	Baja	Media	Alta	Baja
Producción									
Ingresos hidráulico (I)	4375	4375	4375	1446	1446	1446	2209	2209	2209
Ingresos eólico (I)	17888	27441	5263	18348	29316	5696	17333	28115	5542
Ingresos totales (I)	22263	31816	9638	19794	30762	7142	19542	30324	7750
Ingresos hidráulico (C)	3093	3658	3654	1006	1087	1208	1485	1641	1880
Costo de oportunidad	-1282	-717	-721	-440	-358	-237	-723	-568	-329
	-29,3 %	-16,4 %	-16,5 %	-30,4 %	-24,8 %	-16,4 %	-32,7 %	-25,7 %	-14,9 %
Ingresos reales (C)	24448	38711	17309	24104	36988	16069	22896	35838	15938
Diferencia de los ingresos	2185	6895	7671	4310	6226	8928	3354	5514	8188
	9,8 %	21,7 %	79,6 %	21,8 %	20,2 %	125,0 %	17,2 %	18,2 %	105,6 %

Tabla 9.2: Costos de operación de la central hidroeléctrica reversible y el productor eólico, invierno.

Día	Semana			Sábado			Domingo		
	Media	Alta	Baja	Media	Alta	Baja	Media	Alta	Baja
Producción									
Ingresos hidráulico (I)	1231	1231	1231	157	157	157	31	31	31
Ingresos eólico (I)	10817	18374	3364	11183	18734	3489	10943	18561	3316
Ingresos totales (I)	12048	19605	4595	11340	18891	3646	10974	18592	3347
Ingresos hidráulico (C)	889	1086	1008	-52	-123	135	-311	-435	-39
Costo de oportunidad	-342	-145	-223	-209	-281	-22	-342	-466	-70
	-27,8 %	-11,8 %	-18,1 %	-133,0 %	-178,4 %	-14,3 %	-1111,6 %	-1514,4 %	-227,4 %
Ingresos reales (C)	15307	25609	10754	14007	23206	5533	14476	24195	6002
Diferencia de los ingresos	3259	6004	6159	2667	4315	1887	3502	5603	2655
	27,1 %	30,6 %	134,0 %	23,5 %	22,8 %	51,8 %	31,9 %	30,1 %	79,3 %

Tabla 9.3: Costos de operación de la central hidroeléctrica reversible y el productor eólico, verano.

Las tablas 9.2 y 9.3 tienen nueve columnas agrupadas en tres clases. Cada una de estas clases representa un día de semana, un sábado y un domingo respectivamente. Para cada día hay tres tipos de producción: media, alta y baja. En las tablas se marca con «I» la operación independiente de los agentes, y con «C» la operación conjunta.

Los «Ingresos hidráulico (I)» representan la ganancia que la central hidroeléctrica reversible consigue de forma independiente, gracias al programa óptimo.

Los «Ingresos eólico (I)» representan los ingresos que tiene el productor eólico al final de la operación real, funcionando de forma independiente. La ganancia es calculada por medio de la ecuación (9.17). Los «Ingresos totales (I)» son la suma de los ingresos de los dos agentes operando de forma independiente.

El «Ingresos hidráulico (C)» representa la ganancia que la central hidroeléctrica reversible obtiene al cubrir en su programa los desvíos del productor eólico. Con este valor y con los valores de «Ingresos hidráulico (I)» son calculados los «Costos de oportunidad» para la central hidroeléctrica reversible, que representan la pérdida de ésta al modificar su punto de operación. Este valor es dado en Euros (€) y en porcentaje.

El «Ingresos reales (C)» son los ingresos que obtienen la central hidroeléctrica reversible y el productor eólico operando de forma conjunta. Este valor es calculado con la ecuación (9.19) después que ha terminado el día analizado.

Finalmente «Diferencia ingresos» es la relación que hay entre los ingresos totales de la central hidroeléctrica reversible y el productor eólico, operando de forma independiente y conjunta. Este valor es dado tanto en euros como en porcentajes.

Los resultados conseguidos son muy prometedores ya que, para los casos analizados, los ingresos totales de la central hidroeléctrica reversible y el productor eólico operando de forma conjunta son positivos. Esto quiere decir que se obtuvo un ingreso extra al cubrir los errores. De manera adicional, los costos de oportunidad son menores que los ingresos totales; esto quiere decir que se pueden cubrir las expectativas económicas de los agentes al obtener ingresos mayores.

En la tabla 9.3, el campo «Ingresos hidráulico (C)» toma valores negativos ocasionalmente. Esto significa que, en estos casos, es más rentable comprar energía y almacenarla en forma de agua, garantizando las reservas suficientes para cubrir los desvíos del productor eólico. Los resultados dependen del precio de la energía y, en estos casos, los precios corresponden a un perfil típico de verano donde no hay grandes diferencias entre los valores de los periodos de punta y valle, figura 9.5b. Si comparamos este caso con el de invierno, donde las diferencias entre los precios de los periodos de valle y punta son más pronunciados, resulta más difícil garantizar un beneficio en la operación combinada (figura 9.5a). En estos casos las diferencias

en los ingresos están marcadas por los precios de los desvíos, figura 9.6b.

Sin embargo, la operación conjunta es más rentable que la operación independiente en todos los casos simulados. Los ingresos obtenidos por la operación conjunta, finalmente, se deberán distribuir entre el productor eólico y la central hidroeléctrica reversible.

9.7 CONCLUSIONES

El método propuesto permite al productor eólico controlar el riesgo al que está expuesto debido a la incertidumbre en la predicción de la potencia eólica, evitando los desvíos en la producción.

La acción de la central hidroeléctrica reversible es utilizada para minimizar los desvíos debidos a los errores en la previsión de potencia eólica, y la energía necesaria en el embalse está directamente relacionada con la magnitud de los errores. Por esta razón es necesario evaluar, en un largo plazo, la operación de la central hidroeléctrica reversible: para determinar la viabilidad y rentabilidad de su uso y evitar en la subutilización de la misma.

ANÁLISIS DE LOS DESVÍOS EN LA GENERACIÓN EÓLICA EN LOS SISTEMAS DE ENERGÍA ELÉCTRICA

Resumen

En la gestión y operación de los sistemas de energía eléctrica con gran porcentaje de generación eólica, además es necesario considerar los desvíos debidos a los errores en la predicción de potencia. Para compensar los desvíos en el balance de potencia, el Operador del Sistema (OS) reserva algunas centrales de generación para suministrar la energía necesaria, en caso de que la generación sea diferente a la demanda. Actualmente este proceso de compensación se realiza sin considerar dónde se encuentran las centrales que asumen los desvíos y dónde se han producido.

Utilizando un sistema de energía eléctrica tipo, que representa de forma adecuada la complejidad de un sistema real, son simulados diferentes casos en los que se analiza cómo son compensados los desvíos debido a la generación eólica. Los resultados serán analizados para determinar qué efectos producen las diferentes opciones de compensación en el sistema eléctrico.

10.1 INTRODUCCIÓN

Uno de los criterios que asegura la fiabilidad y estabilidad de los sistemas de potencia es mantener los niveles de tensión dentro de un rango admisible. Usualmente el rango de tensiones se mantiene dentro del $\pm 5\%$ del valor por unidad, lo que en condiciones normales del sistema asegura el suministro de energía. El nivel de tensión depende de la potencia reactiva, lo cual significa que para controlar los niveles de tensión se debe operar la potencia reactiva generada o inyectada por elementos pasivos y/o activos.

La energía eólica se caracteriza por la incertidumbre que existe en su producción, por lo cual es catalogada como una fuente no gestionable. La potencia que proviene de fuentes intermitentes y se inyecta al sistema puede tener impactos positivos o negativos, que no son fáciles de conocer a priori. Es necesario hacer estudios que modelen las fuentes intermitentes y su dependencia, a la vez que es considerado el modelo del sistema de energía eléctrica, de modo que sea posible estimar el impacto de las fuentes no gestionables.

En la práctica, si no ocurre un acontecimiento particular, toda la generación eólica es inyectada en el sistema. Por tanto, algunos gene-

radores conectados a la red deben estar disponibles para ajustar su producción, de modo que se cumpla el balance entre la generación y la demanda del sistema. Recordemos que siempre existen errores entre la predicción y la producción real de la energía eólica.

La topología de la red, el punto de conexión de los generadores y la dependencia entre las fuentes intermitentes, entre otros, son factores que influyen en el funcionamiento del sistema. Es posible que se den casos donde el sentido de los flujos de potencia puede cambiar debido a los errores en las predicciones. Esto se vería reflejado en las pérdidas del sistema o llevaría la tensión fuera de los niveles deseados.

Por estas razones es necesario estudiar el efecto que tiene la integración de la generación eólica, considerando también las tensiones en los nudos y la potencia generada.

10.2 DESCRIPCIÓN DEL CASO

Para tener una referencia o punto de comparación es necesario utilizar un modelo del sistema de energía eléctrica con características fiables, de forma que los resultados puedan ser reproducidos y tomados como referencia para interpretar resultados de otros modelos de mayor dimensión y complejidad.

10.2.1 *Sistema de potencia*

El caso tipo, llamado «IEEE 30 Bus Test Case», representa una porción del sistema eléctrico de los EE.UU., localizado en la región del medio oeste del país, y que pertenece a la empresa «American Electric Power System». Los datos fueron tomados en diciembre de 1961 por Iraj Dabbagchi de AEP y fueron introducidos en «IEEE Common Data Format» por Rich Christie, de la Universidad de Washington, en agosto de 1993 [95]. Este caso fue modificado basado en [2] donde se ajustaron los datos de la red. Las funciones de costo de los generadores y las áreas del sistema son tomados de [34]. Finalmente el caso fue incluido como un ejemplo en MATPOWER [108]. Para el trabajo desarrollado en esta tesis, el caso se modificó nuevamente basado en [105], donde son agregados dos generadores que representan la producción eólica en los nudos 19 y 14. Además se conecta otro generador en los nudos 15 o 21, que representa una central hidroeléctrica reversible (figura 10.1).

Como cambios menores, se agrega una carga de $2,4+j1$ MVA en el nudo 11, ya que no tenía conectado ningún elemento. Para mejorar los perfiles de tensión en el sistema, se modifica la carga conectada en el nudo 8, pasando de $30+j30$ MVA a $30+j10$ MVA.

Los parques eólicos conectados en los nudos 19 y 14 tienen una capacidad de 20 MW. La central hidroeléctrica reversible tienen una

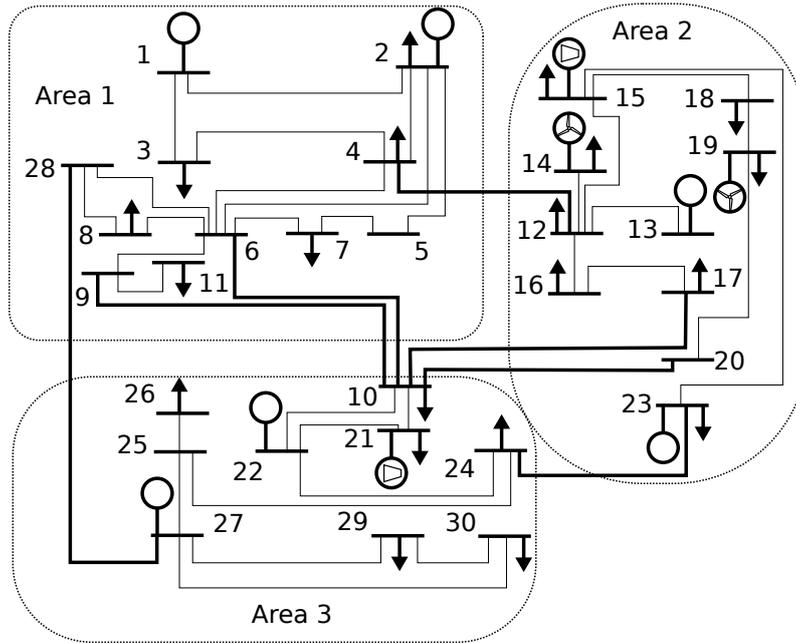


Figura 10.1: «IEEE 30 Bus Test Case», donde se agregó generación eólica en los nudos 19 y 14, además de centrales hidroeléctricas reversibles en los nudos 15 y 21.

capacidad de 30 MW. Está conectada en unos casos en el nudo 15, y en otros, en el nudo 21 (esto será indicado de forma explícita en los problemas).

El resto de generadores del sistema mostrado en la figura 10.1 están caracterizados por el tipo de operación.

GENERACIÓN TRADICIONAL: son los generadores que no van a variar su producción durante las diferentes medidas de compensación (unidades en los nudos 2, 13 y 27).

RESERVAS: son los generadores que reservan una cantidad de energía para cubrir los desvíos (unidades en los nudos 1, 23 y 22).

En el anexo B se presentan los datos del sistema utilizado para obtener los resultados de este capítulo.

10.2.2 MATPOWER como herramienta de simulación

MATPOWER [108, 112] es un paquete originalmente diseñado para MATLAB®, [60], pero actualmente también es compatible con GNU Octave, [27]. La herramienta está orientada principalmente a calcular el Flujo de Potencia (FP) y el Flujo de Potencia Óptimo (FPO) en sistemas de energía eléctrica. MATPOWER está concebido como una herramienta de simulación para investigadores y educadores, ya que es fácil de utilizar y modificar [110]. Además está diseñado para dar

el mejor rendimiento posible, siguiendo unos criterios de desarrollo que hacen fácil entender y modificar el código. Otra de sus ventajas es su documentación detallada [111] y una lista de correo moderada por el Doctor Ray D. Zimmerman, investigador senior de la Universidad de Cornell, donde se resuelven todo tipo de inquietudes o problemas relativos al uso de MATPOWER [109]. Este programa es software libre y está licenciado bajo GNU GPL versión 3.0 [35].

Otros autores [1, 21, 53, 56, 91, 107] han utilizado o basado sus propios desarrollos en MATPOWER. Según [108], al 2009, se tenía un registro de más de 20×10^3 descargas de MATPOWER, donde el 50% tenían fines educativos, el 43% estaban orientados a la investigación, y un 7% para uso en la industria y otros fines. Estos datos garantizan que es una herramienta ampliamente utilizada y probada. Al estar licenciada bajo la GNU GPL versión 3.0 es posible acceder a todo su código, permitiendo verificar sus algoritmos y, por tanto, la calidad y fiabilidad de los resultados. Esta es la principal ventaja si comparamos un programa libre con un programa para el mismo fin, pero de código cerrado, donde no es posible verificar los algoritmos utilizados, [61].

En MATPOWER se utiliza la formulación clásica del flujo de potencia, donde se puede calcular el estado de un sistema eléctrico con el siguiente conjunto de ecuaciones, [16, 111]:

Equilibrio de potencia activa:

$$P_{Gi} - P_{Di} = V_i \sum_{k=1}^n y_{i,k} V_k \cos(\delta_i - \delta_k - \theta_{i,k}) \quad (10.1)$$

Equilibrio de potencia reactiva:

$$Q_{Gi} - Q_{Di} = V_i \sum_{k=1}^n y_{i,k} V_k \sin(\delta_i - \delta_k - \theta_{i,k}) \quad (10.2)$$

Los módulos de la tensión en cada nudo, V_i , deberán estar dentro de un límite inferior V_i^{\min} y un límite superior V_i^{\max} :

$$V_i^{\min} < V_i < V_i^{\max} \quad (10.3)$$

De igual forma, para el argumento de la tensión en cada nudo δ_i , deberá estar dentro de un límite inferior δ_i^{\min} y un límite superior δ_i^{\max} :

$$\delta_i^{\min} < \delta_i < \delta_i^{\max} \quad (10.4)$$

Cada central, por sus características constructivas, puede inyectar una cantidad de potencia activa en el nudo i dentro de un límite inferior $P_{Gi}^{\text{mín}}$ y un límite superior $P_{Gi}^{\text{máx}}$:

$$P_{Gi}^{\text{mín}} < P_{Gi} < P_{Gi}^{\text{máx}} \quad (10.5)$$

De igual manera, las unidades pueden inyectar una cantidad de potencia reactiva en el nudo i dentro de un límite inferior $Q_{Gi}^{\text{mín}}$ y un límite superior $Q_{Gi}^{\text{máx}}$:

$$Q_{Gi}^{\text{mín}} < Q_{Gi} < Q_{Gi}^{\text{máx}} \quad (10.6)$$

El flujo de potencia aparente que circula por las líneas de transmisión se obtiene a partir del módulo y argumento de las tensiones de los nudos en ambos extremos de las mismas y esta limitado por su capacidad nominal:

$$|S_i(V_{i,k}, \delta_{i,k})| - S_{\text{máx}} \leq 0 \quad (10.7)$$

$$|S_k(V_{k,i}, \delta_{k,i})| - S_{\text{máx}} \leq 0 \quad (10.8)$$

Con el conjunto de ecuaciones anteriores se modela el sistema de energía eléctrica, donde las variables restantes están en función del módulo y argumento de la tensión y las potencias generadas por cada central.

10.2.3 Simulaciones

Como programa de partida se calcula un Flujo de Potencia Óptimo (FPO) que minimiza el costo de la generación de acuerdo a las características de las centrales, [34]. Este procedimiento sirve para obtener de forma anticipada el programa de la operación.

FPO_o Flujo de Potencia Óptimo donde se minimiza el costo de la generación, programa de partida.

Para visualizar los efectos de los desvíos en la predicción eólica se propone un ejemplo: durante la operación en tiempo real de los parques eólicos, situados en los nudos 14 y 19, ocurre un desvío en la generación eólica respecto al programa de partida: se pasa de una expectativa de 3,2 MW a una producción real de 0,0 MW, y de 2,9 MW a una producción real de 0,0 MW respectivamente. El desvío total eólico es de 6,1 MW. Para mantener el balance entre la demanda del sistema y la generación, son aplicadas las siguientes medidas:

- FP_p Flujo de Potencia, el desvío de la generación eólica es cubierto de forma proporcional por las centrales dedicadas a la acción de reserva (unidades en los nudos 1, 23 y 22). En este caso las centrales hidroeléctricas reversibles no operan.
- FP₁₅ Flujo de Potencia, el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 15, y las pérdidas del sistema son asumidas por el generador en el nudo 1.
- FP₂₁ Flujo de Potencia, el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 21, y las pérdidas del sistema son asumidas por el generador en el nudo 1.
- FPO Flujo de Potencia Óptimo, el desvío de la generación eólica es cubierto por las centrales dedicadas a la acción de reserva (unidades en los nudos 1, 23 y 22), minimizando el costo de la generación. En este caso las centrales hidroeléctricas reversibles no operan.
- FPO₁₅ Flujo de Potencia Óptimo, el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 15, y las pérdidas del sistema son asumidas por las centrales dedicadas a la acción de reserva, minimizando el costo de la generación.
- FPO₂₁ Flujo de Potencia Óptimo, el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 21, y las pérdidas del sistema son asumidas por las centrales dedicadas a la acción de reserva, minimizando el costo de la generación.
- TIES Minimización de los intercambios entre áreas donde el desvío de la generación eólica es cubierto por las centrales dedicadas a la acción de reserva (unidades en los nudos 1, 23 y 22). En este caso las centrales hidroeléctricas reversibles no operan.
- TIES₁₅ Minimización de los intercambios entre áreas donde el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 15, y las pérdidas del sistema son asumidas por las centrales dedicadas a la acción de reserva.
- TIES₂₁ Minimización de los intercambios entre áreas donde el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 21, y las pérdidas del sistema son asumidas por las centrales dedicadas a la acción de reserva.

- R Minimización de las pérdidas activas y reactivas, donde el desvío de la generación eólica es cubierto por las centrales dedicadas a la acción de reserva (unidades en los nudos 1, 23 y 22). En este caso las centrales hidroeléctricas reversibles no operan.
- R₁₅ Minimización de las pérdidas activas y reactivas, donde el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 15, y las pérdidas del sistema son asumidas por las centrales dedicadas a la acción de reserva.
- R₂₁ Minimización de las pérdidas activas y reactivas, donde el desvío de la generación eólica es cubierto por la central hidroeléctrica reversible que está conectada en el nudo 21, y las pérdidas del sistema son asumidas por las centrales dedicadas a la acción de reserva.

Utilizando MATPOWER como herramienta base se programaron funciones nuevas para obtener los resultados referentes a la «minimización de los intercambios entre áreas» (TIES) y la «minimización de las pérdidas activas y reactivas» (R), los cuales se encuentran disponibles en [47].

10.3 RESULTADOS

Primero se presentan los valores de potencia generada activa, reactiva y aparente por cada área del sistema (tabla 10.1). Las pérdidas del sistema son discriminadas en activa, reactiva y aparente, y son presentadas en la tabla 10.2. También se indica el costo marginal de la generación del sistema, donde se ha considerado un costo cero para la producción eólica y para la producción de la central hidroeléctrica reversible.

En las tablas presentadas, el término «FPO» se refiere a los resultados obtenidos a través de un proceso de optimización, donde el objetivo es minimizar el costo de la generación, ecuación (10.9). En el problema de optimización se modela el sistema de energía eléctrica considerando las ecuaciones (10.1)-(10.8), modelo no lineal conocido como «modelo en AC», donde se consideran las restricciones relativas a los niveles de tensión y las capacidades de las líneas y de los generadores, [111]. Para estos casos se utilizó el Método de Puntos Interiores (MPI) descrito en [48, 108]. Las funciones de costo de los generadores son de orden cuadrático y están especificadas de acuerdo al formato de MATPOWER [111], en el anexo B.

$$\text{mín} \sum_{i=1}^{n_g} f_{Pi}(P_{Gi}) \quad (10.9)$$

Donde:

n_g Número de generadores en el sistema eléctrico.

f_{P_i} Función de costo del generador i .

Los valores encabezados con «FP» se obtuvieron por medio del método de Newton [92], donde se encontró una solución que satisfacía el conjunto de ecuaciones que modelan el sistema de energía eléctrica: (10.1) y (10.2). En el FP no se consideran las restricciones relativas a los niveles de tensión y a las capacidades de las líneas y de los generadores.

Los valores encabezados con «TIES» se obtuvieron a través de un proceso de optimización, donde se minimizan los intercambios de potencia entre áreas. En estos casos se utilizó MATPOWER, reemplazando la función objetivo por defecto. En el proceso de optimización se minimiza el resultado del sumatorio de la diferencia al cuadrado de las potencias activas y reactivas que circulan por las líneas de transmisión que interconectan las diferentes áreas del sistema respecto al caso inicial «FPO₀», [47]. En la figura 10.1 las líneas que interconectan las áreas tienen un trazo más grueso.

$$\text{mín} \sum_{i \in n_{\text{TIES}}} (P_{0i} - P_i)^2 + (Q_{0i} - Q_i)^2 \quad (10.10)$$

Donde:

n_{TIES} Conjunto que contiene los índices de los nodos donde hay una línea de transmisión que interconecta dos áreas del sistema.

P_{0i} Potencia activa en el nudo i del caso FPO₀.

Q_{0i} Potencia reactiva en el nudo i del caso FPO₀.

Finalmente, el término «R» se refiere a los valores que se obtuvieron a través de un proceso de optimización, donde el objetivo era minimizar las pérdidas activas y reactivas del sistema. Para esto se modificaron las funciones de costo de las centrales (nodos 1, 22 y 23), por una misma función de costo lineal. De esta forma el FPO minimiza de forma indirecta las pérdidas del sistema, ya que la única manera de reducir los costos es mediante la reducción de las pérdidas del sistema, [47].

Los términos con el subíndice «15» y «21» refieren a los casos donde la central hidroeléctrica reversible está operativa en los nudos 15 o 21 para compensar el desvío de los generadores eólicos. En ambos casos solo está operativa una de las centrales, pero nunca lo están las dos al mismo tiempo.

En la tabla 10.3 se presentan las potencias activa, reactiva y aparente que van de un área a otra. La pérdida de cada línea es despreciable respecto a la potencia que circula por ésta misma. Por tal motivo solo se muestra el flujo de potencia en un solo extremo de cada línea. En

Caso	Área	MW	MVAR	MVA	MW _T	MVAR _T	MVA _T
FPO ₀	1	77,67	16,89	79,48	194,97	78,66	210,24
	2	55,42	24,81	60,72			
	3	61,89	36,96	72,08			
FP _P	1	79,73	16,91	81,51	195,05	78,89	210,40
	2	51,43	25,32	57,33			
	3	63,88	36,66	73,65			
FP ₁₅	1	77,74	17,17	79,62	195,05	78,88	210,40
	2	55,42	24,04	60,41			
	3	61,89	37,67	72,45			
FP ₂₁	1	77,67	17,47	79,61	194,98	78,64	210,24
	2	40,44	28,38	49,41			
	3	76,87	32,79	83,57			
FPO	1	80,50	17,61	82,40	195,08	78,90	210,44
	2	51,74	24,66	57,31			
	3	62,85	36,63	72,75			
FPO ₁₅	1	77,80	16,66	79,56	195,04	78,91	210,40
	2	55,33	25,19	60,80			
	3	61,91	37,06	72,16			
FPO ₂₁	1	77,68	15,55	79,22	194,97	78,70	210,25
	2	40,66	26,15	48,34			
	3	76,62	37,01	85,09			
TIES	1	77,80	16,84	79,60	195,01	78,84	210,34
	2	53,66	24,89	59,15			
	3	63,55	37,11	73,59			
TIES ₁₅	1	77,74	17,04	79,58	195,07	78,99	210,46
	2	54,91	25,07	60,36			
	3	62,43	36,88	72,51			
TIES ₂₁	1	78,06	16,68	79,82	194,99	78,85	210,33
	2	51,54	24,70	57,15			
	3	65,39	37,47	75,37			
R	1	58,46	21,63	62,33	194,59	77,09	209,30
	2	55,55	22,22	59,83			
	3	80,58	33,24	87,17			
R ₁₅	1	55,95	21,43	59,92	194,57	77,07	209,27
	2	57,95	21,65	61,86			
	3	80,66	33,98	87,53			
R ₂₁	1	56,70	20,23	60,20	194,58	77,05	209,28
	2	49,64	21,58	54,12			
	3	88,25	35,24	95,03			

Tabla 10.1: Potencia activa, reactiva y aparente generada por área y sistema.

Caso	MW	MVAR	MVA	Costo generación [€]
FPO ₀	2,11	8,80	9,05	563,61
FP _P	2,18	9,03	9,29	582,18
FP ₁₅	2,19	9,02	9,28	563,83
FP ₂₁	2,12	8,78	9,03	563,63
FPO	2,22	9,09	9,36	582,11
FPO ₁₅	2,18	9,02	9,28	563,81
FPO ₂₁	2,11	8,78	9,03	563,59
TIES	2,14	8,97	9,22	583,44
TIES ₁₅	2,21	9,11	9,37	563,97
TIES ₂₁	2,13	8,94	9,20	575,73
R	1,69	7,70	7,89	612,90
R ₁₅	1,67	7,69	7,87	596,44
R ₂₁	1,68	7,69	7,87	583,76

Tabla 10.2: Pérdidas en el sistema y costo de la generación.

la columna de la derecha, marcada con «FO», está el resultado del sumatorio de la diferencia al cuadrado de las potencias activas y reactivas que circulan por las líneas de transmisión que interconectan las diferentes áreas del sistema respecto al caso inicial «FPO₀».

En la figura 10.2 se muestran los niveles de tensión en cada uno de los nudos del sistema. Además se representan con líneas punteadas los límites utilizados en los procesos de optimización.

En las figuras 10.3 y 10.4 están representadas las potencias activa y reactiva generadas por cada central.

10.3.1 *Análisis*

Los valores correspondientes a «FPO₀» se refieren al programa obtenido de forma anticipada a la operación real, donde considerando los costes marginales de cada una de las centrales, se busca asegurar la energía necesaria para cubrir la demanda del sistema al menor costo.

Los resultados marcados con «FP_P» han obtenido los resultados por medio de un Flujo de Potencia. En este caso, el desvío de la producción eólica es cubierto de forma proporcional por las centrales en los nudos 1, 22 y 23, donde cada uno pertenece a un área del sistema. La central conectada al nudo 1 tiene la labor de asumir las pérdidas del sistema (denominado nudo oscilante). Esta solución se asemeja al caso español, ya que cuando se presenta un desvío, el Operador del Sistema (OS) envía consignas proporcionales a los agentes que están participando en el mercado de reservas. En la tabla 10.1 se puede observar que, después del desvío (6,1 MW en el área 2), las potencias generadas por área aumentan al rededor de los 2 MW para cubrirlo

Caso	Áreas	MW	MVAR	MVA	FO
FPO ₀	A1->A2	-1,70	-2,96	3,41	0,00
	A1->A3	-2,99	-8,27	8,79	
	A3->A2	6,00	4,16	7,30	
FP _P	A1->A2	-1,76	-1,45	2,28	6,98
	A1->A3	-3,06	-7,74	8,33	
	A3->A2	5,61	6,67	8,72	
FP ₁₅	A1->A2	-1,50	-3,23	3,56	1,85
	A1->A3	-2,94	-7,92	8,45	
	A3->A2	6,73	4,50	8,10	
FP ₂₁	A1->A2	-1,30	0,64	1,45	69,97
	A1->A3	-2,73	-11,85	12,16	
	A3->A2	2,18	15,61	15,76	
FPO	A1->A2	-1,39	-1,34	1,93	7,60
	A1->A3	-2,73	-7,11	7,62	
	A3->A2	5,90	6,27	8,61	
FPO ₁₅	A1->A2	-1,87	-3,15	3,66	1,61
	A1->A3	-3,10	-7,95	8,54	
	A3->A2	5,96	4,50	7,47	
FPO ₂₁	A1->A2	-1,59	0,48	1,66	63,08
	A1->A3	-4,41	-11,67	12,48	
	A3->A2	4,67	15,52	16,21	
TIES	A1->A2	-1,63	-2,44	2,93	4,51
	A1->A3	-3,14	-8,65	9,21	
	A3->A2	5,98	5,44	8,09	
TIES ₁₅	A1->A2	-1,74	-3,10	3,56	1,48
	A1->A3	-2,94	-8,09	8,60	
	A3->A2	5,94	4,88	7,69	
TIES ₂₁	A1->A2	-1,34	-1,22	1,81	18,28
	A1->A3	-3,60	-9,61	10,26	
	A3->A2	5,91	6,37	8,69	
R	A1->A2	1,17	-8,24	8,32	135,46
	A1->A3	0,32	-21,96	21,97	
	A3->A2	5,54	9,26	10,79	
R ₁₅	A1->A2	1,16	-9,89	9,96	160,50
	A1->A3	0,21	-22,80	22,80	
	A3->A2	6,17	8,51	10,51	
R ₂₁	A1->A2	1,31	-7,06	7,18	191,45
	A1->A3	-1,10	-24,88	24,90	
	A3->A2	6,03	14,01	15,25	

Tabla 10.3: Intercambios de potencia entre áreas.

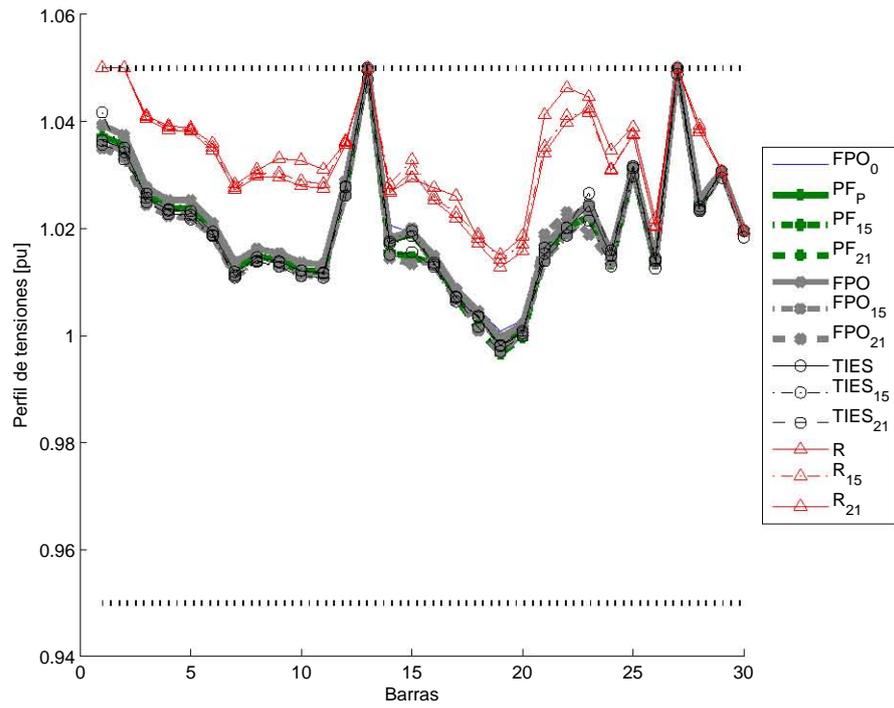


Figura 10.2: Tensiones en los nudos.

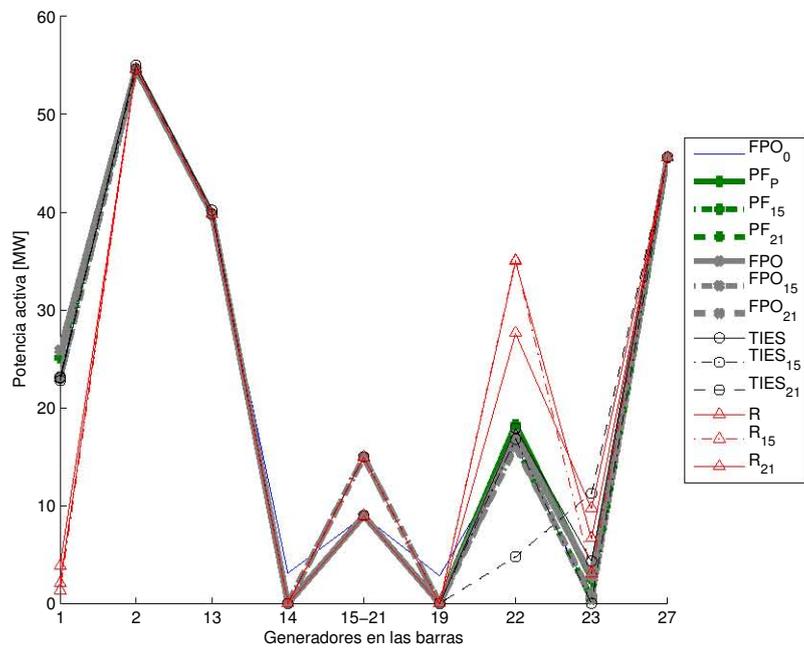


Figura 10.3: Potencia activa inyectada por los generadores.

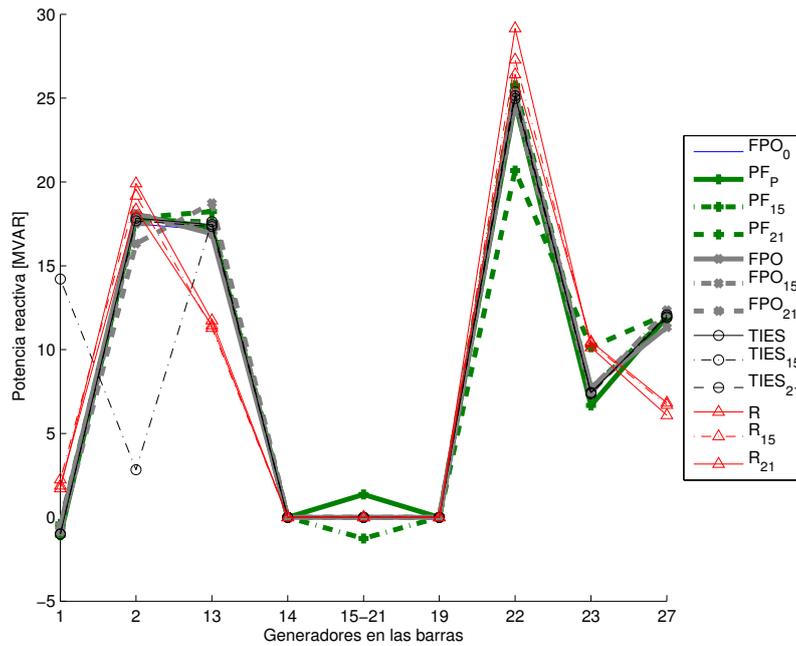


Figura 10.4: Potencia reactiva inyectada por los generadores.

en las áreas 1 y 3. En los casos denotados por «FP₁₅» y «FP₂₁», el desvío ha sido asumido por la central hidroeléctrica reversible, y las pérdidas por la central en el nudo 1. Al comparar el resultado de la «FP₁₅» con el caso inicial, solo se observa un cambio mínimo que se puede apreciar en las potencias totales, columnas 6-8 de la tabla 10.1. Para el caso «FP₂₁», donde la central hidroeléctrica reversible está conectada en el nudo 21, ésta incrementa su producción para asumir los desvíos de los parques eólicos, lo cual se ve reflejado en la potencia producida por área. Para saber cuál de los casos ofrece algún beneficio, se comparan las pérdidas del sistema (tabla 10.2) y se observa que el caso «FP₂₁» tiene pérdidas menores respecto a los otros dos casos. Este valor se debe a que hay un intercambio de potencia entre las áreas, que resulta en menores pérdidas al transportar electricidad (tabla 10.3). La principal desventaja del FP es que no considera las restricciones relativas a los niveles de tensión, a las capacidades de las líneas y a las de los generadores.

Los resultados marcados con «FPO» han sido obtenidos por medio de un proceso de optimización donde se minimiza el costo de la producción, basado en las características de las centrales conectadas a los nudos 1, 22 y 23. De los casos marcados con «FPO» se puede decir que el que ofrece un mejor desempeño, en términos de costo y potencia generada, es el FPO₂₁, ya que según los datos de las tablas 10.1 y 10.2 se cumple la demanda y el costo es menor, comparado con los otros casos de este tipo. Para comparar el costo de la operación de forma general se debe considerar que la potencia generada por los parques eólicos y por la central hidroeléctrica reversible se ha puesto

a un valor cero. Además el sistema debe estar bajo las mismas condiciones, por lo que se deben comparar los valores por grupos: primer grupo FP_P , FPO , $TIES$ y R ; segundo grupo FP_{15} , FPO_{15} , $TIES_{15}$ y R_{15} ; y el tercer grupo FP_{21} , FPO_{21} , $TIES_{21}$ y R_{21} . Según los datos de la tabla 10.2, para los resultados marcados con «FPO» siempre se obtiene la operación más barata.

Los valores obtenidos con la medida denominada «TIES» se pueden verificar en la tabla 10.3. Para este caso, el objetivo es minimizar el resultado del sumatorio de la diferencia al cuadrado de las potencias activas y reactivas que circulan por las líneas de transmisión que interconectan las diferentes áreas del sistema respecto al caso inicial FPO_0 . Este valor ha sido calculado para todas las medidas, de forma que se puedan comparar; el valor se presenta en la columna de la derecha de la tabla 10.3, encabezado con el término «FO». Ya que se compara con el caso inicial FPO_0 , el valor «FO», para el mismo, es cero. Para este grupo de casos, si el objetivo es mantener el estado del sistema lo más parecido al caso inicial FPO_0 , la mejor opción es compensar lo más cerca posible de donde han ocurrido los desvíos. Con los resultados de la tabla 10.3 se verifica esta afirmación, donde el caso marcado con $TIES_{15}$ es el que obtiene el mínimo valor de «FO». Esto significa que se acerca más al caso inicial después de que ocurre el desvío en los parques eólicos. En este caso la compensación se ha realizado en la misma área donde ocurrieron los desvíos. Al igual que para los otros casos, los valores obtenidos con la medida denominada «TIES» se deben comparar de forma general para que estén bajo las mismas condiciones. El resultado de esta medida se puede verificar con el valor de la columna encabezada por «FO», tabla 10.3, donde se puede comprobar que, para los grupos bajo las mismas condiciones, la medida «TIES» consigue el menor valor de la potencia intercambiada entre áreas. Este tipo de medida es interesante en el contexto de diferentes reglas de mercado por área. El objetivo es mantener lo más cerca posible los niveles de potencia intercambiados previamente, ya que no solo pueden estar relacionados a factores económicos, sino también aspectos técnicos.

Los resultados marcados con «R» tienen como objetivo minimizar las pérdidas, tanto activas como reactivas, en el sistema. Entre los casos marcados con «R» en la tabla 10.2 no se puede identificar cuál ofrece un mejor desempeño en términos de pérdidas, ya que los valores para R_{15} y R_{21} obtienen el mismo valor, menos para la potencia activa. Según los datos de la tabla 10.1, la generación total del caso R_{15} es ligeramente menor que la del caso R_{21} . Esto significa que con este caso se cumple la demanda y se obtiene la solución que minimiza las pérdidas activas y reactivas del sistema.

Con los resultados presentados en la tabla 10.2 se puede constatar de forma general que, para los grupos bajo las mismas condiciones, los casos marcados con «R» consiguen las menores pérdidas en com-

paración con el resto de casos presentados. En un contexto donde la operación de las reservas es pagada a un mismo precio a todos los agentes que participan, se podrían minimizar las pérdidas del sistema al realizar las consignas de acuerdo con los resultados obtenidos con esta medida.

Los resultados presentados en las tablas 10.1, 10.2 y 10.3, y en la gráficas 10.2, 10.3 y 10.4, se refieren a un caso posible de dos parques eólicos, a un horizonte de predicción $k = 28$ h. Este ejemplo fue tomado del modelo espacial-temporal presentado en la sección 8.8, donde los parques eólicos, situados en el noroeste de España, están separados por una distancia de 5,1km. Con este modelo se calculan 10×10^3 trayectorias posibles de producción eólica para los dos parques. Las trayectorias se emplearán para realizar un Flujo de Potencia Probabilístico (FPP) y obtener el comportamiento de algunas variables del sistema; de esta forma se podrán comparar las diferentes medidas de compensación del desvío eólico.

10.4 FLUJO DE POTENCIA PROBABILÍSTICO

El Flujo de Potencia Probabilístico (FPP) se puede llevar a cabo utilizando un enfoque numérico o un enfoque analítico. En el enfoque numérico se utiliza el método de Monte Carlo: se parte de una muestra aleatoria de las variables y parámetros de entrada del modelo del sistema de energía eléctrica y luego se realiza un análisis determinista para cada valor, de modo que se obtiene el mismo número de valores para las variables respuestas que se quieren conocer. El enfoque analítico parte del estudio de las variables y parámetros de entrada, que son modelados como expresiones matemáticas (por ejemplo, la función de densidad); los resultados que se obtienen también están expresados en términos matemáticos, [20].

El método de Monte Carlo ha sido aplicado en diversos campos de la ciencia. Esta técnica debe su nombre a los casinos situados en Monte Carlo, Mónaco. Una simulación efectuada con este método utiliza una muestra de números aleatorios para modelar algún tipo de proceso. La técnica funciona especialmente bien cuando se conocen las probabilidades de las variables de entrada, pero se desconoce la probabilidad de las variables de salida (la naturaleza de ciertos problemas hace difícil determinar de forma analítica el comportamiento de las variables de salida). Autores como [55, 96, 99, 100, 107] han utilizado un enfoque numérico para analizar la incertidumbre en los sistemas de energía eléctrica. Una de las ventajas del enfoque numérico es que no hay pérdida de fiabilidad en los modelos, comparados con las simplificaciones matemáticas que se hacen en los enfoques analíticos, [20].

Actualmente gran parte del tiempo de cálculo de algunos de los ordenadores más rápidos del mundo se dedica a realizar simulaciones

por medio del método de Monte Carlo, ya que hace posible describir algunas de las leyes fundamentales de la física, [6].

El principal inconveniente del método de Monte Carlo es la necesidad de un gran número de simulaciones para obtener un resultado fiable, lo que se traduce en largas esperas mientras se realizan los cálculos [20]. En [46] se comparan los tiempos de cálculo y precisión de un programa comercial, llamado PSS®E [83], funcionando en un ordenador corriente que utiliza una unidad central de procesamiento (en inglés, Central Processing Unit (CPU)), y un prototipo que utiliza una unidad de procesamiento gráfico (en inglés, Graphics Processing Units (GPU)) que coopera con una CPU. En [46] se desarrollaron simulaciones de estabilidad transitoria sobre diferentes sistemas de prueba. El sistema de mayor dimensión (número de nudos: 1248; número de generadores: 320; número de líneas: 1244) consiguió un desempeño 344,8 veces más rápido que la aproximación tradicional con PSS®E; 1m44,4s del prototipo que utiliza la GPU-CPU contra 10 h el sistema tradicional que utiliza solo una CPU.

En [84] se presenta un prototipo basada en MATLAB que explota las características de las GPU para el cálculo numérico en el análisis de imágenes con fines médicos. Los resultados obtenidos muestran un rendimiento 1,7 veces más rápido en los tiempos de cálculo, comparado con una CPU empleada para la misma aplicación. Con base en los casos presentados en [46, 84] se podría plantear la posibilidad de modificar MATPOWER, que es un programa libre, para aprovechar los beneficios que ofrecen las GPU. Emplear este tipo de procesadores podrían reducir considerablemente los tiempos de cálculo en las simulaciones del método de Monte Carlo, incluso en problemas de mayor dimensión.

10.4.1 Simulación

Tomando como referencia el ejemplo presentado en la sección 10.3, se pretende realizar un Flujo de Potencia Probabilístico (FPP) por medio del método de Monte Carlo, donde a partir del modelo espacial-temporal presentado en la sección 8.8, se dispone de 10×10^3 trayectorias posibles de producción eólica para dos parques, y se considerará el horizonte de predicción $k = 28$ h. Con estos valores se realizarán los análisis numéricos para obtener el comportamiento de algunas variables del sistema de energía eléctrica; de esta forma se pueden comparar las diferentes medidas presentadas en la sección 10.2.3.

En las figuras 10.5, 10.6 y 10.7 se representa, con un diagrama de caja, la generación total por cada área activa, reactiva y aparente.

En la figura 10.8 se representa con un diagrama de caja el comportamiento de las centrales de generación que cubren los errores de la predicción eólica.

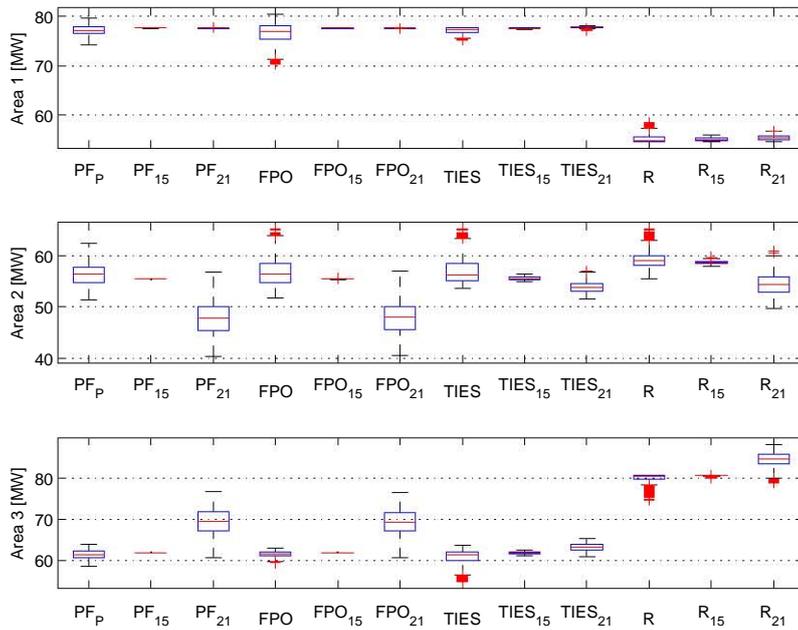


Figura 10.5: Potencia activa generada por área.

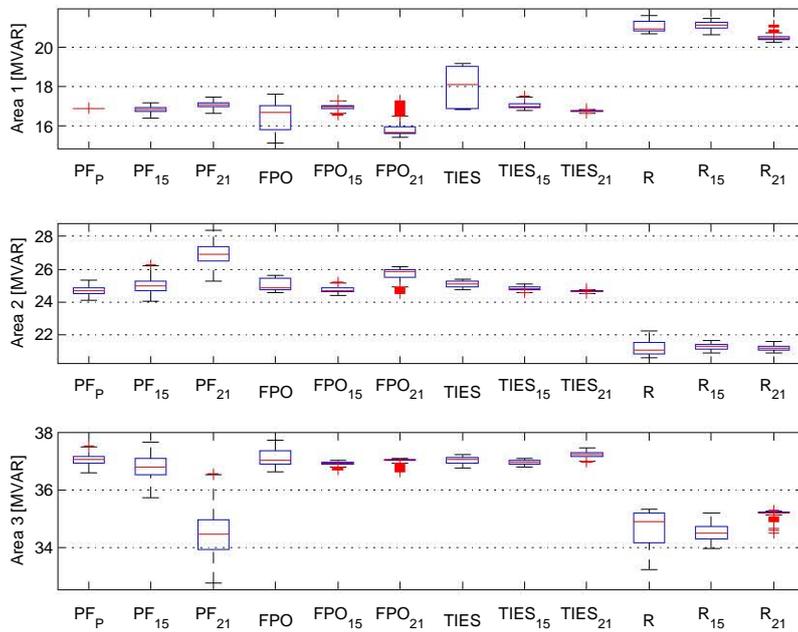


Figura 10.6: Potencia reactiva generada por área.

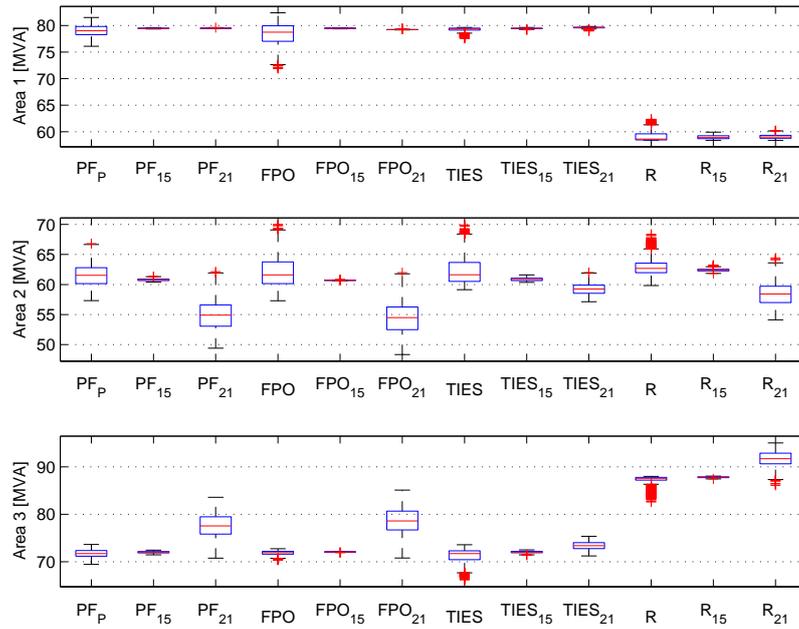


Figura 10.7: Potencia aparente generada por área.

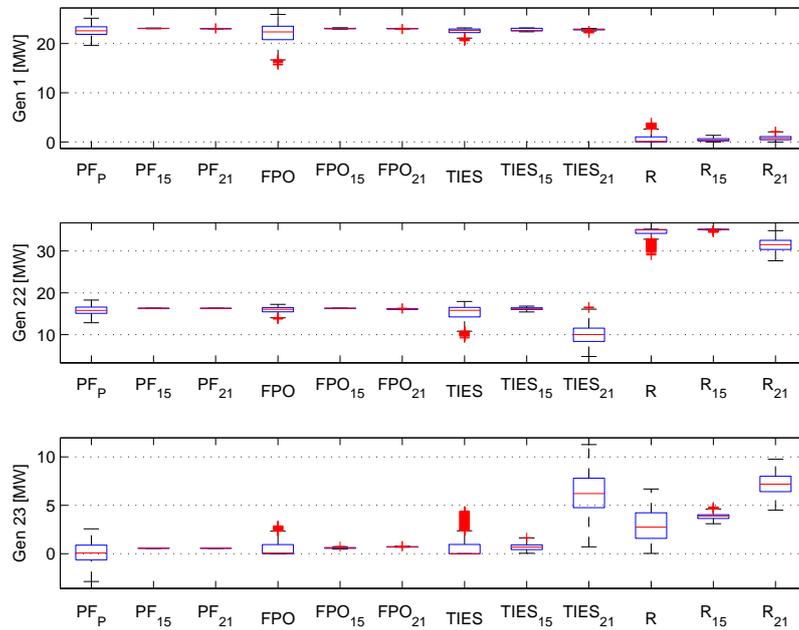
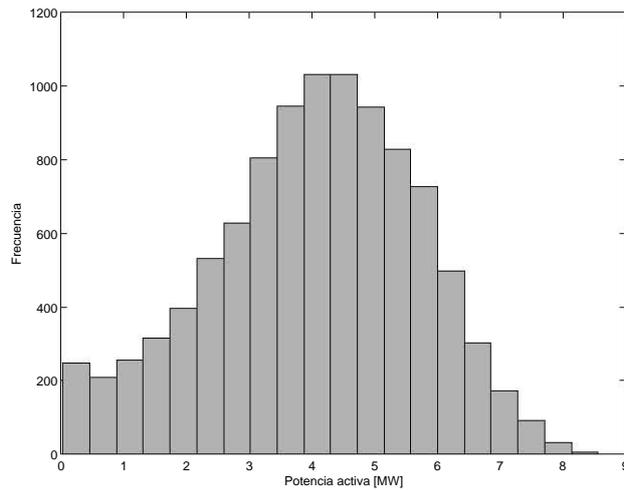
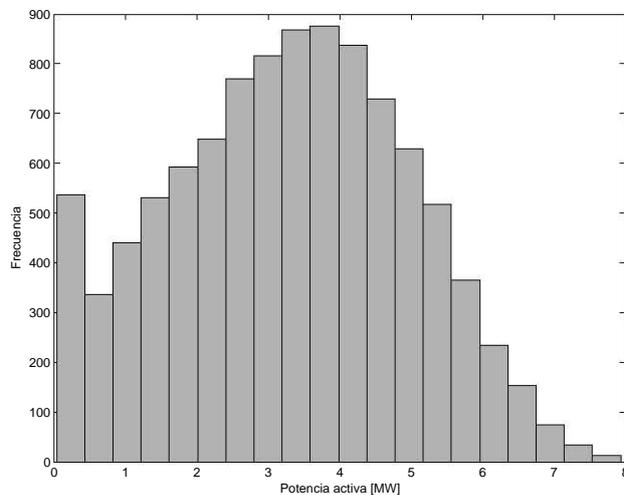


Figura 10.8: Potencia activa generada por unidad, acciones de compensación.



(a) PDF de la potencia producida por el parque eólico en el nudo 14, para el horizonte $k = 28$ h.



(b) PDF de la potencia producida por el parque eólico en el nudo 19, para el horizonte $k = 28$ h.

Figura 10.9: Potencia activa generada, método de Monte Carlo.

En las figuras 10.9a y 10.9b se puede observar el comportamiento esperado de la producción de los parques eólicos ubicados en los nudos 14 y 19. La forma de la PDF se caracteriza por una concentración de producciones nulas. Este comportamiento fue modelado con la regresión logística, sección 7, donde se consideraron y modelaron, de forma independiente, los errores en la predicción cuando la producción era nula.

En las figuras 10.10, 10.11 y 10.12 se presentan los flujos de potencia que se intercambian entre áreas. Las pérdidas de las líneas son despreciables en comparación con la potencia que circula por éstas. Por tal razón solo se muestra el flujo en un sentido.

En la figura 10.13, se pueden apreciar las pérdidas del sistema para cada una de las medidas planteadas.

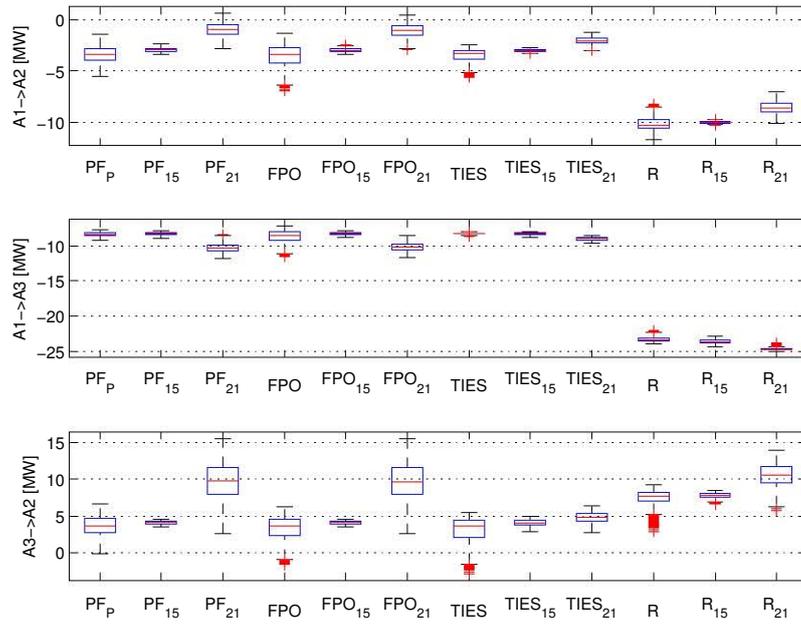


Figura 10.10: Intercambios de potencia activa entre áreas.

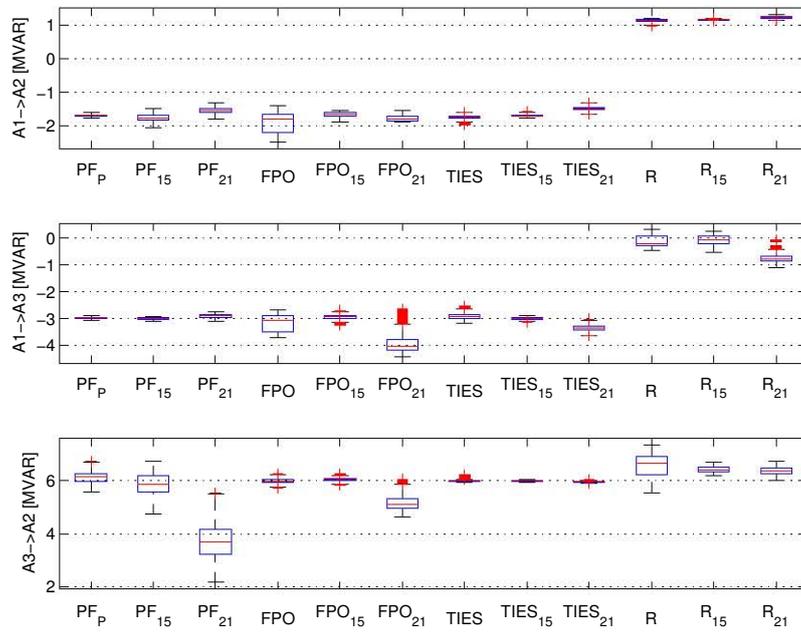


Figura 10.11: Intercambios de potencia reactiva entre áreas.

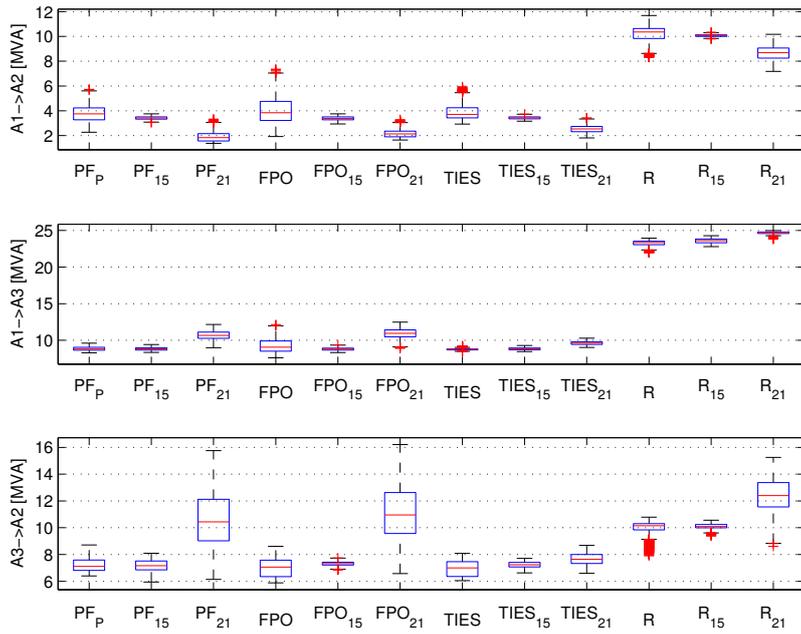


Figura 10.12: Intercambios de potencia aparente entre áreas.

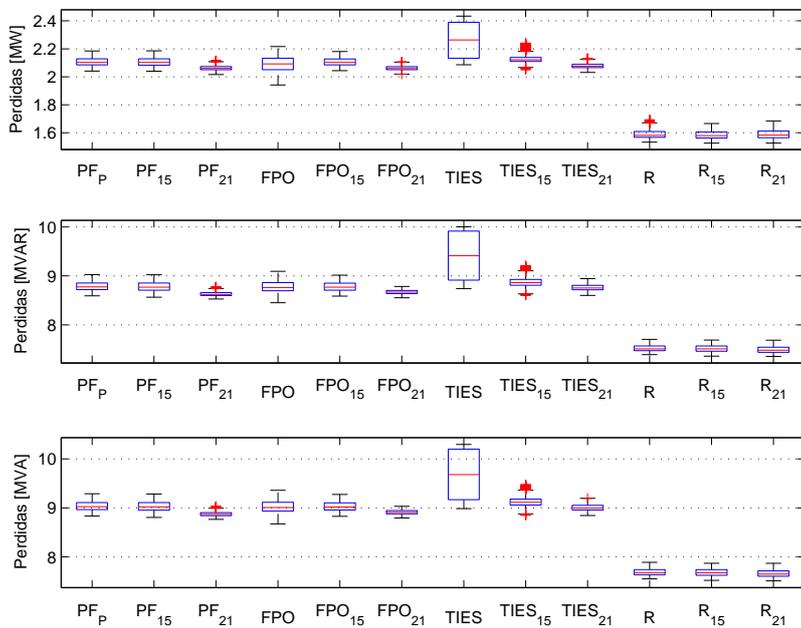


Figura 10.13: Perdidas en el sistema.

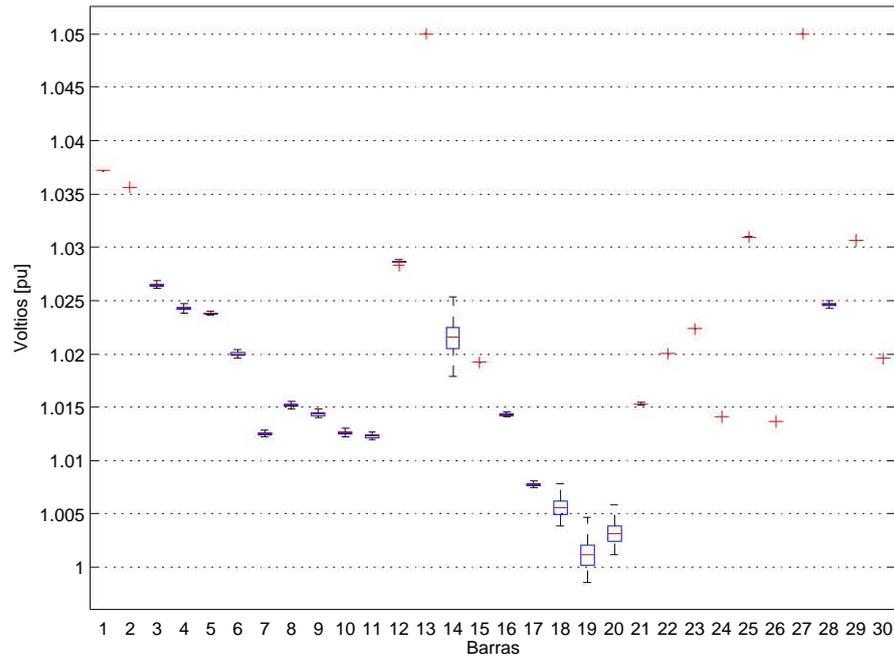


Figura 10.14: Tensiones en el sistema, caso FPO.

Como información adicional, se presentan las tensiones en los nudos (solo para la medida FPO en la figura 10.14).

Los resultados presentados en las figuras 10.5 - 10.14, son realizados utilizando MATLAB en un ordenador con doble CPU de 3,00 GHz y 3 Gb de memoria RAM. Cada uno de los 10×10^3 escenarios de producción se utilizó para calcular las 12 medidas, por tanto, se calcularon 90×10^3 flujos de potencia óptimos y 30×10^3 flujos de potencia, para un total de 120×10^3 casos evaluados, que tardaron en total 2h20m aproximadamente.

10.4.2 Análisis

En el análisis del estado de un sistema de energía eléctrica se parte tradicionalmente de un grupo de escenarios, considerando la generación eólica como fuente de incertidumbre: un escenario de poca producción, uno de producción media y otro de producción alta, de modo que se obtienen tres resultados con los que se espera explicar el comportamiento del sistema de energía eléctrica. Con un Flujo de Potencia Probabilístico (FPP), por medio del método de Monte Carlo, se puede obtener el comportamiento del sistema de energía eléctrica de acuerdo a las características de las variables de entrada y un proceso de cálculo determinista. Esto permite estimar cuál es la probabilidad de una variable de salida dadas las probabilidades de las variables de entrada, en este caso, la producción eólica.

En las figuras 10.5, 10.6 y 10.7, los resultados marcados con «FP» se han obtenido de un Flujo de Potencia. Este tipo de análisis se realiza

tradicionalmente, pero no garantiza que los niveles de tensión, las capacidades de las líneas y las de los generadores estén dentro de sus límites. Utilizando un proceso de optimización se evita este tipo de problemas, a la vez que se puede perseguir un objetivo que ofrezca un beneficio adicional a la correcta operación del sistema.

Los resultados marcados con «FPO», «TIES» y «R» han sido obtenidos por medio de un proceso de optimización. Para comparar o decidir cuál de las medidas es más apropiada se debe definir un criterio que tenga en cuenta los tres estados del sistema: sin acción de almacenamiento, con almacenamiento en el nudo 15 o con almacenamiento en el nudo 21.

Los valores obtenidos con la medida denominada «FPO», garantizan que se cubre la demanda del sistema con las centrales de generación más baratas, sin embargo esta medida no garantiza que el sistema opera de una forma adecuada respecto a las pérdidas. Con estos resultados se podría determinar cuáles son las líneas que limitan la potencia inyectada en la red, o en cuáles puntos del sistema es apropiado instalar un elemento pasivo para compensar la potencia reactiva o incluso la instalación de una central hidroeléctrica reversible.

Los valores obtenidos con la medida denominada «TIES», se pueden verificar en las figuras 10.10, 10.11 y 10.12. Con estos resultados se puede determinar para todas las medidas las respectivas PDF de las potencias intercambiadas entre áreas.

Los resultados marcados con «R» tienen como objetivo minimizar las pérdidas activas y reactivas en el sistema. En la figura 10.13 se puede constatar que para esta medida se consiguen las menores pérdidas en el sistema, en comparación con el resto de casos.

Para los casos con almacenamiento en el nudo 15 se consiguen los resultados con menor varianza (gráficas 10.5-10.13), donde los diagramas de caja son los más estrechos. Este resultado se debe a que el desvío es cubierto en la misma área, de modo que las centrales 1, 22 y 23, disponibles para este mismo fin, apenas tiene que modificar su punto de operación.

Por otro lado, para el caso con almacenamiento en el nudo 21 se obtienen, en la mayoría de los casos, los resultados con mayor varianza (gráficas 10.5-10.13), donde los diagramas de caja son mucho más anchos, comparados con el resto de casos. Este resultado se debe a que el desvío ocurrido en el área 1 es cubierto en el área 3, de modo que las centrales 1, 22 y 23 deben ajustar su punto de operación para garantizar el balance entre generación y demanda.

En la figura 10.14 se obtiene el comportamiento de la tensión en los nudos para la medida FPO. Además de la tensión se puede obtener el comportamiento de cualquier otra variable del sistema de energía eléctrica y, de esta forma, comparar las diferentes medidas presentadas.

Part IV

CONCLUSIONS AND FUTURE WORK

CONCLUSIONS

A new modeling approach that permits the generation of probable trajectories of future wind power production for the probabilistic analysis of power systems is presented. The main goals of this thesis are the development of a tool to generate trajectories of future wind power production (part II of the thesis) for a short term horizon (one or two days in advance) and their use to obtain a joint operation of a hydro-pump plant cooperating with a wind farm (chapter 9) and a development of a probabilistic power flow through the Monte Carlo simulation for the power system analysis (chapter 10). These applications are designed to help in the integration of wind energy into power systems.

In part II of the thesis, a new approach to deal with the uncertain behavior of the stochastic sources is introduced; a method to simulate trajectories of wind power production is proposed. The main feature of the proposed method is that the simulated trajectories have similar properties as the wind power production. In particular, they are able to reproduce the non-linear and non-gaussian behavior of these. The simulated trajectories are obtained through the method proposed in a common computer in a few seconds, making it possible to use it in real-time applications.

In order to investigate and to model the stochastic dependence, the copula theory has been presented in the chapter 3. The gaussian copula has been used to represent the errors of the wind power forecast in two cases, temporal and spatial dependence.

Based on the copula modeling, the procedure is divided in two parts: the modeling of the forecast errors as an unidimensional marginal distributions and the estimation of the multi-dimensional stochastic dependence structure. The marginal distributions represent the stochastic behavior of each wind farm at each specific horizon, and the stochastic dependence structure defines the dependence between the stochastic sources at each horizon or location.

The models were obtained from real data, predictions and productions (chapter 4), reproducing the characteristics of original random variables. For univariate marginal functions, was considered a parametric model based on the beta distribution, and the Kernel Density Estimation (KDE) for the local model, a nonparametric method, chapters 5 and 6 respectively.

In both models, parametric and nonparametric, were implemented adaptive estimates of the parameters. The new data of the wind farm are included and weighted according to their relevance and age, giv-

ing the model an adaptive feature, replying the seasonal behavior of wind production.

The two models, parametric and nonparametric, failed to model the behavior of the forecast errors in the particular case of zero productions, as shown in sections 5.5 and 6.4 respectively. This problem was solved using the logistic regression, a method used to estimate the probability of a qualitative response variable. The behavior of the errors in the wind power forecasting in the case of zero productions were modeled by logistic regression. Finally, a model for the marginal function was build with logistic regression and a distribution function (beta distribution and KDE), chapter 7.

For the marginal distribution, the model using Kernel Density Estimation and logistic regression performs better than the model using the beta distribution and logistic regression. The results of this evaluation are presented in section 7.4.

Finally, for the copula model, the dependence structure was obtained, the variables were transformed to the uniform domain throw the PIT using the CDF model obtained by the Kernel Density Estimation and logistic regression. Thus, eliminating the influence of the original domain of the variables. These results are presented in the chapter 8.

The calculation of the simulated trajectories using the copula functions is very promising, because possible energy imbalance and prediction error intervals can be obtained, making easy to understand the behavior of the wind energy production and the relationship with its prediction.

In the third part of the thesis, chapter 9, the statistical information obtained from the simulated trajectories makes possible to calculate the joint optimal operation of a hydro-pump plant and a wind power producer. The wind power production uncertainty is included in an optimization problem that use a hydro-pump plant in an optimal way, reducing the imbalance costs and optimizing the benefit of the two agents.

The hydro-pump plant and the wind power producer participate jointly in the market. In this way, the utilities may reduce the risk due to the uncertainty in the wind power prediction and in the prices of the reserve market. Eventually, the imbalances in the wind power production could be fully avoided by the action of the hydro-pump plant. In all the results presented in the chapter 9, the joint optimal operation achieved more benefits than the independent operation.

In the chapter 9 two procedures are presented as an optimization problems and the estimates of the wind power uncertainty are used to obtain the optimal operation of a hydro-pump plant cooperating with a wind power producer. The results do not guarantee more benefits for all possible cases, but certainly have advantages, increasing the action of the utility into the energy market. There are also advan-

tages in the power system operation, because the utility could also use the hydro-pump plant to control the voltage, frequency and to store energy, making the system safer and enhancing the integration of renewable sources. Extensions of these strategies could also be applied in the next generation networks, called Smart Grids.

Different actions to compensate wind power uncertainties in power system are compared in the chapter 10. The approach was tested on a modified "IEEE 30 Bus Test Case", modeling the complexity of a small real system.

The method proposed in chapter 10 is valid to perform the probabilistic load flow of an electrical power system through Monte Carlo simulation where wind power uncertainty and their interdependence are considered. One of the goals of the presented method is the ability to determine the Probability Density Function (PDF) for all the output variables in the power system, for example, the charge of the lines, the losses of the system, the use of storage devices, etc. Both wind power and storage influence in the power systems can be observed just by comparing the output variables (generation, area interchange flows, system losses and storage use) of the Monte Carlo Simulation in the power systems simulations.

The deterministic forecasts of wind generation is a valuable input to a number of decision-making problems, related with power system operation and trading of wind generation into the markets. However, they do not provide any information on the interdependence of forecasts errors, when it is necessary to quantify the possible deviations in terms of power and energy. This information is required to obtain the optimal operation of wind power farms in combination with conventional generation and storage, or for trading in energy markets. Indeed, for such problems, the appropriate input concerning the interdependence of forecasts errors must consist of possible trajectories, that provide the whole information about the uncertainty in the periods. In this theses a suitable method to generate trajectories of possible wind power from deterministic wind power forecasts is presented. The trajectories were used to obtain the optimal operation of a wind farm cooperating with a hydro-pump plant. Also, they were used to deal with the operation of power systems with high wind power penetration, considering the uncertainty of short term wind power production, the spatial dependence and the power system model.

11.1 ORIGINAL CONTRIBUTIONS

The main original contribution of this work are the following:

- The local model of forecast errors by using Kernel Density Estimation (KDE) and logistic regression.
- The adaptive estimation of the marginal function and dependence structure of the wind production.
- A method to simulate trajectories of possible wind power productions based on the historical data (Observations and forecasts).
- A method to obtain the optimal operation of a hydro-pump plant cooperating with a wind farm.

11.2 FUTURE WORK

The work presented in this thesis sets the theoretical background for the analysis of wind power production and hydro-pump storage in power systems. The future work involves possible applications and extensions of this work.

11.2.1 *Enhancing the integration of stochastic energy sources in large power system, as the Iberian system.*

The integration of stochastic generation in large power systems needs the use of techniques able to deal with the stochastic generation spread over the network. The method presented on this document, can be the start point for this kind of analysis. But first, it is necessary to develop prototypes and evaluate their performance with real data.

One of the main problems associated with the power system analyses is to obtain reliable models where the stochastic sources are identified. The power system model must replicate the topology, behavior and complexity of the real system.

With all the stochastic sources identified into the power system model, and using real data to obtain the multivariate model and to calculate the trajectories or scenarios of the stochastic sources, could be performed a probabilistic analysis providing a reliable approach to understand the impact of stochastic generation into the network.

The intraday market must be introduced into the model to get more accurate results, thus following more closely the current rules of the Iberian market, which are well suited to the high penetration of intermittent generation.

11.2.2 *Stochastic model.*

The dependence structure of the random variables will be one of the main topics to improve the model, because it refers to the relative behavior of the random variables. For this, its role is central in the calculation of the properties of the sum of a multitude of random variables; even in the case of the same marginal distributions, different dependence structures lead to totally different distribution of their sum, around the same mean value.

The other aspect to be considered is the applicability of different copula families to the power system uncertainty analysis.

11.2.3 *Impact of large-scale storage.*

It is necessary to evaluate the effect of the use of large-scale storage in large power systems with stochastic sources. This aspect can be assessed by using a similar approach as the presented in this thesis. It would be also interesting to test storage use for different horizons: hours, days, weeks to evaluate its performance.

11.2.4 *Speed up simulations with a GPU*

The development of a prototype using MATPOWER to obtain the power flow and the optimal power flow, using a Graphics Processing Units (GPU) to reduce calculation times in the Monte Carlo method may be implemented.

11.2.5 *Planing evaluation.*

For planning studies it is required to evaluate the viability and profitability of the joint operation of storage and stochastic energy sources, focusing in the size and the cost of installation for both. With a modification and extension of the procedure presented in this thesis, this information could be obtained.

11.2.6 *Unit Commitment and wind farms.*

In power systems with high levels of wind power, the integration of these stochastic sources causes technical and financial impacts on the operation of the existing electricity system, due to the stochastic behavior and the forecast errors of the wind power. The stochastic wind power must be included into the unit commitment problem for the power systems analysis, the effects of the stochastic sources on the unit commitment and dispatch of power systems are relevant in the cost of the operation and must be studied in detail.

Parte V

APÉNDICES

MERCADO DE ELECTRICIDAD

El mercado de electricidad en España está definido por las transacciones derivadas de la participación de los agentes en las sesiones de los mercados diario e intradiario. Los agentes del mercado son aquellas empresas habilitadas para actuar en el mercado de producción como vendedores y compradores de electricidad.

El OMIE [70] realiza una subasta para cada una de las horas del día siguiente, 14 horas antes de que éste comience. Esta subasta es denominada como *mercado diario*, donde los agentes del mercado presentan ofertas de venta y compra de energía eléctrica.

En el mercado diario los agentes vendedores son todas las unidades de generación disponibles que no estén comprometidas con un contrato bilateral físico, además los agentes externos registrados como vendedores. Los agentes compradores son los distribuidores, comercializadores, consumidores cualificados y agentes externos registrados como compradores.

En el mercado diario, en primer lugar, son adquiridas las ofertas de energía más baratas, seguidas por las más caras hasta que se cubre la demanda proyectada. La última oferta que cubre la demanda proyectada, marca el precio marginal de la energía para esa hora, por lo que todas las demás ofertas son pagadas a ese precio.

Después de que se cierra el mercado diario y la información sobre los contratos bilaterales físicos ejecutados, el operador del sistema evalúa la viabilidad técnica del programa de funcionamiento de las unidades de producción para garantizar la seguridad y fiabilidad del suministro en la red de transporte. Si el resultado de la casación del mercado diario más las ejecuciones de los contratos bilaterales físicos lleva a una operación donde se supera la capacidad máxima de intercambio entre los sistemas eléctricos, o los requisitos de seguridad no son cumplidos, el procedimiento de *solución de restricciones técnicas* modifica el programa original hasta llegar a una solución viable.

Por medio del *mercado intradiario* se negocian cantidades adicionales de energía para efectuar el balance entre oferta y demanda. Este proceso está a cargo del operador del sistema, para el cual cuenta con la ayuda del operador del mercado. El mercado intradiario se estructura en seis sesiones durante el día, en éstas son realizados los ajustes sobre el programa diario viable, mediante la presentación de ofertas de venta y adquisición de energía eléctrica. Al finalizar cada una de estas casaciones se estudian nuevamente las restricciones técnicas del sistema, y se aplica el procedimiento de solución de restricciones técnicas hasta alcanzar un nuevo programa viable.

Los *servicios complementarios* y el procedimiento de gestión de desvíos tienen por objeto que el suministro de energía eléctrica se produzca en las condiciones de calidad, fiabilidad y seguridad que están establecidas y que se verifique de forma permanente el equilibrio entre la generación y la demanda. Los servicios complementarios de banda de regulación son incorporados al programa diario viable por el operador del sistema con posterioridad al mercado diario. Una vez celebrada cada una de las sesiones del mercado intradiario, el operador del sistema realiza la gestión en tiempo real mediante la utilización de servicios complementarios y el procedimiento de gestión de desvíos.

Según la documentación de MATPOWER [111], anexo B, son presentados a continuación los valores del caso «IEEE 30 Bus Test Case».

El caso tipo, llamado «IEEE 30 Bus Test Case», representa una porción del sistema eléctrico de los EE.UU. localizado en la región del medio oeste del país, y que pertenece a la empresa «American Electric Power System». Los datos fueron tomados en diciembre de 1961 por Iraj Dabbagchi de AEP y fueron introducidos en «IEEE Common Data Format» por Rich Christie, de la Universidad de Washington, en agosto de 1993 [95].

Este caso fue modificado y adaptado utilizando datos de [2, 34], y luego fue incluido como un caso tipo en MATPOWER [108]. Basado en [105] se agregan dos generadores que representan la producción eólica en los nudos 19 y 14. Además se conecta otro generador en el nudo 15 o 21, que representa una central hidroeléctrica reversible (figura 10.1).

Como cambios menores, se agrega una carga de $2,4+j1$ MVA en el nudo 11, ya que no tenía conectado ningún elemento. Para mejorar los perfiles de tensión en el sistema, se modifica la carga conectada en el nudo 8, pasando de $30+j30$ MVA a $30+j10$ MVA.

Los parques eólicos conectados en los nudos 19 y 14, tienen una capacidad de 20 MW. La central hidroeléctrica reversible tienen una capacidad de 30 MW. Está conectada en unos casos en el nudo 15, y en otros, en el nudo 21.

Cuadro de texto B.1: IEEE 30

```

function mpc = caso.tesis

%% MATPOWER Case Format : Version 2
mpc.version = '2';

%%----- Power Flow Data -----%%
%% system MVA base
mpc.baseMVA = 100;

%% bus data
% bus_i type Pd Qd Gs Bs area Vm Va baseKV zone Vmax Vmin lam.P lam.Q mu_Vmax mu_Vmin
mpc.bus = [
1 3 0 0 0 0 1 1.0222582 0 135 1 1.05 0.95 0.0100 -0 0 0;
2 2 21.7 12.7 0 0 1 1.0221587 -0.041471317 135 1 1.1 0.95 0.0100 -0 0 0;
3 1 2.4 0 1.2 0 1 1.0162283 -0.87884859 135 1 1.05 0.95 0.0101 0 0 0;
4 1 7.6 1.6 0 0 1 1.0153336 -1.0130055 135 1 1.05 0.95 0.0101 0 0 0;
5 1 0 0 0 0.19 1 1.0115542 -1.1356308 135 1 1.05 0.95 0.0101 0 0 0;
6 1 0 0 0 0 1 1.0094929 -1.2644705 135 1 1.05 0.95 0.0101 0 0 0;
7 1 22.8 10.9 0 0 1 1.001223 -1.7270792 135 1 1.05 0.95 0.0102 0.0001 0 0;
8 1 30 10 0 0 1 0.9982394 -1.6438422 135 1 1.05 0.95 0.0111 0.0010 0 0;
9 1 0 0 0 0 1 1.0165587 -1.4812132 135 1 1.05 0.95 0.0102 0.0001 0 0;
10 1 5.8 2 0 0 3 1.0202655 -1.5935462 135 1 1.05 0.95 0.0103 0.0001 0 0;
11 2 2.4 1 0 0 1 1.0165587 -1.4812132 135 1 1.05 0.95 0.0102 0.0001 0 0;
12 1 11.2 7.5 0 0 2 1.0433037 -0.30912711 135 1 1.05 0.95 0.0101 0 0 0;
13 2 0 0 0 0 2 1.0832474 2.5310679 135 1 1.1 0.95 0.0101 0 0 0;
14 1 6.2 1.6 0 0 2 1.0320971 -0.91965787 135 1 1.05 0.95 0.0102 0 0 0;
15 2 8.2 2.5 0 0 2 1.0332574 -0.7816813 135 1 1.05 0.95 0.0102 0 0 0;
16 1 3.5 1.8 0 0 2 1.026259 -1.134033 135 1 1.05 0.95 0.0102 0.0001 0 0;
17 1 9 5.8 0 0 2 1.0167916 -1.6651527 135 1 1.05 0.95 0.0103 0.0001 0 0;
18 2 3.2 0.9 0 0 2 1.0160388 -1.7760821 135 1 1.05 0.95 0.0104 0.0001 0 0;
19 1 9.5 3.4 0 0 2 1.009504 -2.1836227 135 1 1.05 0.95 0.0105 0.0001 0 0;
20 1 2.2 0.7 0 0 2 1.011331 -2.0907602 135 1 1.05 0.95 0.0104 0.0001 0 0;
21 1 17.5 11.2 0 0 3 1.026747 -1.5051463 135 1 1.05 0.95 0.0103 0 0 0;
22 2 0 0 0 0 3 1.0325637 -1.3571232 135 1 1.1 0.95 0.0103 0 0 0;
23 2 3.2 1.6 0 0 2 1.0473624 0.20617267 135 1 1.1 0.95 0.0100 -0 0 0;
24 1 8.7 6.7 0 0.04 3 1.0328266 -0.38186709 135 1 1.05 0.95 0.0102 0.0001 0 0;
25 1 0 0 0 0 3 1.0499999 1.4846384 135 1 1.05 0.95 0.0101 -0 0.0438 0;
26 1 3.5 2.3 0 0 3 1.0330439 1.0858293 135 1 1.05 0.95 0.0103 0.0001 0 0;
27 2 0 0 0 0 3 1.0689387 2.8769798 135 1 1.1 0.95 0.0100 0 0 0;
28 1 0 0 0 0 1 1.0160552 -0.99522254 135 1 1.05 0.95 0.0103 0.0002 0 0;
29 1 2.4 0.9 0 0 3 1.0499867 1.7422645 135 1 1.05 0.95 0.0102 0.0001 0 0;
30 2 10.6 1.9 0 0 3 1.0391 0.94874963 135 1 1.05 0.95 0.0104 0.0001 0 0;

```

```

];
%% generator data
% bus Pg Qg Qmax Qmin Vg mBase status Pmax Pmin Pc1 Pc2 Qc1min Qc1max Qc2min Qc2max ramp_agc ramp_10 ramp_30 ramp_q apf mu_Pmax mu_Pmin mu_Qmax mu_Qmin
mpc.gen = [
1 9.84297 -1.8073 20 -1 1.0222582 100 1 80 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
23 19.2 6.82942 40 -10 1.0473624 100 1 30 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
22 21.59 29.0162 62.5 -15 1.0325637 100 1 50 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
2 54.8325 14.4897 60 -20 1.0221587 100 1 55 54.6 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
13 40 31.898 44.7 -15 1.0832474 100 1 40.2 39.8 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
27 45.8116 20.3874 48.7 -15 1.0689387 100 1 46 45.6 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
19 10 0 0 0 1.0689387 100 1 10 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
14 12 0 0 0 1.0689387 100 1 12 12 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
15 5 0 0 0 1.0689387 100 1 5 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
];

%% branch data
% fbus tbus r x b rateA rateB rateC ratio angle status angmin angmax Pf Qf Pt Qt mu_Sf mu_St mu_angmin mu_angmax
mpc.branch = [
1 2 0.02 0.06 0.03 130 130 130 0 0 1 -360 360 1.1855 -1.7926 -1.1852 -1.3413 0 0 0 0;
1 3 0.05 0.19 0.02 130 130 130 0 0 1 -360 360 8.6575 -0.0147 -8.6211 -1.9248 0 0 0 0;
2 4 0.06 0.17 0.02 65 65 65 0 0 1 -360 360 10.5201 -0.5663 -10.456 -1.3290 0 0 0 0;
3 4 0.01 0.04 0 130 130 130 0 0 1 -360 360 6.2211 0.7248 -6.2173 -0.7096 0 0 0 0;
2 5 0.05 0.2 0.02 130 130 130 0 0 1 -360 360 10.5888 1.8220 -10.531 -3.6597 0 0 0 0;
2 6 0.06 0.18 0.02 65 65 65 0 0 1 -360 360 13.2088 1.8753 -13.103 -3.6239 0 0 0 0;
4 6 0.01 0.04 0 90 90 90 0 0 1 -360 360 14.0788 11.3305 -14.047 -11.204 0 0 0 0;
5 7 0.05 0.12 0.01 70 70 70 0 0 1 -360 360 10.5312 3.8541 -10.467 -4.7145 0 0 0 0;
6 7 0.03 0.08 0.01 130 130 130 0 0 1 -360 360 12.3875 5.3219 -12.332 -6.1855 0 0 0 0;
6 8 0.01 0.04 0 32 32 32 0 0 1 -360 360 22.3951 22.8573 -22.294 -22.455 0.0015 0 0 0;
6 9 0 0.21 0 65 65 65 0 0 1 -360 360 1.8486 -3.3931 -1.8486 3.4239 0 0 0 0;
6 10 0 0.56 0 32 32 32 0 0 1 -360 360 1.0563 -1.9389 -1.0563 1.9657 0 0 0 0;
9 11 0 0.21 0 65 65 65 0 0 1 -360 360 0 0 0 -0 0 0 0;
9 10 0 0.11 0 65 65 65 0 0 1 -360 360 1.8486 -3.4239 -1.8486 3.4400 0 0 0 0;
4 12 0 0.26 0 65 65 65 0 0 1 -360 360 -5.0051 -10.892 5.0051 11.2543 0 0 0 0;
12 13 0 0.14 0 65 65 65 0 0 1 -360 360 -39.999 -28.775 39.9998 31.8980 0 0 0 0;
12 14 0.12 0.26 0 32 32 32 0 0 1 -360 360 5.3580 2.0475 -5.3217 -1.9689 0 0 0 0;
12 15 0.07 0.13 0 32 32 32 0 0 1 -360 360 8.6792 3.4173 -8.6233 -3.3134 0 0 0 0;
12 16 0.09 0.2 0 32 32 32 0 0 1 -360 360 9.7575 4.5560 -9.6616 -4.3429 0 0 0 0;
14 15 0.22 0.2 0 16 16 16 0 0 1 -360 360 -0.8783 0.3689 0.8802 -0.3672 0 0 0 0;
16 17 0.08 0.19 0 16 16 16 0 0 1 -360 360 6.1616 2.5429 -6.1279 -2.4627 0 0 0 0;
15 18 0.11 0.22 0 16 16 16 0 0 1 -360 360 9.8888 3.2144 -9.7774 -2.9916 0 0 0 0;
18 19 0.06 0.13 0 16 16 16 0 0 1 -360 360 6.5774 2.0916 -6.5497 -2.0316 0 0 0 0;
19 20 0.03 0.07 0 32 32 32 0 0 1 -360 360 -2.9503 -1.3684 2.9534 1.3756 0 0 0 0;
10 20 0.09 0.21 0 32 32 32 0 0 1 -360 360 5.1806 2.1390 -5.1534 -2.0756 0 0 0 0;
10 17 0.03 0.08 0 32 32 32 0 0 1 -360 360 2.8777 3.3523 -2.8721 -3.3373 0 0 0 0;
];

```

```

10 21 0.03 0.07 0 32 32 32 0 0 1 -360 360 -5.3704 -7.1434 5.3934 7.1971 0 0 0 0;
10 22 0.07 0.15 0 32 32 32 0 0 1 -360 360 -5.5830 -5.7535 5.6262 5.8461 0 0 0 0;
21 22 0.01 0.02 0 32 32 32 0 0 1 -360 360 -22.894 -18.397 22.9752 18.5608 0 0 0 0;
15 23 0.1 0.2 0 16 16 16 0 0 1 -360 360 -10.346 -2.0338 10.4498 2.2421 0 0 0 0;
22 24 0.12 0.18 0 16 16 16 0 0 1 -360 360 -7.0115 4.6093 7.0907 -4.4904 0 0 0 0;
23 24 0.13 0.27 0 16 16 16 0 0 1 -360 360 5.5502 2.9874 -5.5031 -2.8896 0 0 0 0;
24 25 0.19 0.33 0 16 16 16 0 0 1 -360 360 -10.288 0.7227 10.4770 -0.3937 0 0 0 0;
25 26 0.25 0.38 0 16 16 16 0 0 1 -360 360 3.5411 2.3625 -3.5000 -2.3000 0 0 0 0;
25 27 0.11 0.21 0 16 16 16 0 0 1 -360 360 -14.018 -1.9688 14.2180 2.3505 0 0 0 0;
28 27 0 0.4 0 65 65 65 0 0 1 -360 360 -18.336 -12.813 18.3364 14.7521 0 0 0 0;
27 29 0.22 0.42 0 16 16 16 0 0 1 -360 360 6.1568 1.6509 -6.0786 -1.5015 0 0 0 0;
27 30 0.32 0.6 0 16 16 16 0 0 1 -360 360 7.1004 1.6339 -6.9517 -1.3552 0 0 0 0;
29 30 0.24 0.45 0 16 16 16 0 0 1 -360 360 3.6786 0.6015 -3.6483 -0.5448 0 0 0 0;
8 28 0.06 0.2 0.02 32 32 32 0 0 1 -360 360 -7.7053 -7.5446 7.7669 5.7210 0 0 0 0;
6 28 0.02 0.06 0.01 32 32 32 0 0 1 -360 360 -10.537 -8.0194 10.5695 7.0923 0 0 0 0;
];

%%----- OPF Data -----%%
%% area data
% area refbus
mpc.areas = [
1 8;
2 23;
3 26;
];

%% generator cost data
% 1 startup shutdown n x1 y1 ... xn yn
% 2 startup shutdown n c(n-1) ... c0
mpc.gencost = [
2 0 0 3 0.02 2 0;
2 0 0 3 0.025 3 0;
2 0 0 3 0.0625 1 0;
2 0 0 3 0.0175 1.75 0;
2 0 0 3 0.025 3 0;
2 0 0 3 0.00834 3.25 0;
2 0 0 3 0 0 0;
2 0 0 3 0 0 0;
2 0 0 3 0 0 0;
];

```

BIBLIOGRAFÍA

- [1] Akorede, M. F. y Hizam, H. Teaching power system analysis courses using MATPOWER. En *2009 International Conference on Engineering Education (ICEED)*, páginas 45–51. IEEE, Kuala Lumpur, Malaysia, diciembre 2009.
- [2] Alsac, O. y Stott, B. Optimal load flow with Steady-State security. *IEEE Transactions on Power Apparatus and Systems*, PAS-93(3):745–751, 1974.
- [3] Anagnostopoulos, J. y Papantonis, D. Simulation and size optimization of a pumped-storage power plant for the recovery of wind-farms rejected energy. *Renewable Energy*, 33(7):1685–1694, julio 2008.
- [4] Anagnostopoulos, J. S. y Papantonis, D. E. Pumping station design for a pumped-storage wind-hydro power plant. *Energy Conversion and Management*, 48(11):3009–3017, 2007.
- [5] Angarita, J. M. y Usaola, J. G. Combining hydro-generation and wind energy biddings and operation on electricity spot markets. *Electric Power Systems Research*, 77(5-6):393–400, 2007.
- [6] Billinton, R. y Li, W. *Reliability assessment of electric power systems using Monte Carlo methods*. Springer, 1994.
- [7] Bludszuweit, H., Dominguez-Navarro, J., y Llombart, A. Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems*, 23(3):983–991, 2008.
- [8] Bludszuweit, H. y Dominguez-Navarro, J. A. A probabilistic method for energy storage sizing based on wind power forecast uncertainty. *IEEE Transactions on Power Systems*, 26(3):1651–1658, agosto 2011.
- [9] Bofinger, S., Luig, A., y Beyer, H. G. Qualification of wind power forecasts. En *Proceedings of the Global Wind Power Conference*. Paris, France, abril 2002.
- [10] Bremnes, J. B. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, 9(1-2):3–11, 2006.
- [11] Brier, G. W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, enero 1950.

- [12] Brown, P., Lopes, J. P., y Matos, M. Optimization of pumped storage capacity in an isolated power system with large renewable penetration. *IEEE Transactions on Power Systems*, 23(2):523–531, 2008.
- [13] Bueno, C. y Carta, J. Wind powered pumped hydro storage systems, a means of increasing the penetration of renewable energy in the Canary Islands. *Renewable and Sustainable Energy Reviews*, 10(4):312–340, agosto 2006.
- [14] Caldon, R., Patria, A. R., y Turri, R. Optimisation algorithm for a virtual power plant operation. En *Universities Power Engineering Conference, 2004. UPEC 2004. 39th International*, tomo 3, páginas 1058–1062. IEEE, Bristol, Reino Unido, septiembre 2004.
- [15] Carney, M. y Cunningham, P. Evaluating density forecasting models. Informe técnico, Trinity College Dublin, Department of Computer Science, mayo 2006.
URL <http://www.tara.tcd.ie/handle/2262/13502>
- [16] Castillo, E., Conejo, A. J., Pedregal, P., García, R., y Alguacil, N. *Formulación y resolución de modelos de programación matemática en Ingeniería*. Universidad de Castilla-La Mancha Escuela Técnica Superior de Ingenieros Industriales, febrero 2002.
- [17] Castronuovo, E., Usaola, J., y Jaramillo Duque, Á. Delegated dispatching of wind farms: an optimal approach considering continuous control and interruption capabilities. *Wind Energy*, 12(4):332–347, septiembre 2008.
- [18] Castronuovo, E. D. y Lopes, J. A. P. On the optimization of the daily operation of a wind-hydro power plant. *IEEE Transactions on Power Systems*, 19(3):1599–1606, 2004.
- [19] Castronuovo, E. D. y Lopes, J. A. P. Optimal operation and hydro storage sizing of a wind-hydro power plant. *International Journal of Electrical Power & Energy Systems*, 26(10):771–778, diciembre 2004.
- [20] Chen, P., Chen, Z., y Bak-Jensen, B. Probabilistic load flow: A review. En *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, páginas 1586–1591. IEEE, Nanjuing, China, abril 2008.
- [21] Cole, S. y Belmans, R. MatDyn, a new Matlab-Based toolbox for power system dynamic simulation. *IEEE Transactions on Power Systems*, 26(3):1129–1136, agosto 2011.
- [22] Conejo, A., Plazas, M., Espinola, R., y Molina, A. Day-ahead electricity price forecasting using the wavelet transform and

- ARIMA models. *IEEE Transactions on Power Systems*, 20(2):1035–1042, 2005.
- [23] Cook, R. D. y Weisberg, S. *Residuals and Influence in Regression*. Chapman and Hall, 1982.
- [24] Costa, L. M., Bourry, F., Juban, J., y Kariniotakis, G. Management of energy storage coordinated with wind power under electricity market conditions. En *Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems*, páginas 1–8. IEEE, Rincon, Puerto Rico, mayo 2008.
- [25] Daneshi, H. y Daneshi, A. Price forecasting in deregulated electricity markets - a bibliographical survey. En *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008*, páginas 657–661. IEEE, Nanjuing, China, abril 2008.
- [26] Doherty, R. y O'Malley, M. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Transactions on Power Systems*, 20(2):587–595, 2005.
- [27] Eaton, J. W., Bateman, D., y Hauberg, S. *GNU Octave Manual*. Network Theory Ltd., 3 edición, octubre 2008.
- [28] Epstein, E. S. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8:985–987, diciembre 1969.
- [29] European Commission. *Green Paper, Towards a European strategy for the security of energy supply*. Office for Official Publications of the European Communities, 2000.
URL http://ec.europa.eu/energy/green-paper-energy-supply/doc/green_paper_energy_supply_en.pdf
- [30] European Wind Energy Association (EWEA). EWEA annual report 2010. Informe técnico, European Wind Energy Association (EWEA), junio 2011.
URL <http://www.ewea.org/>
- [31] Fabbri, A., GomezSanRoman, T., RivierAbbad, J., y Mendez-Quezada, V. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Transactions on Power Systems*, 20(3):1440–1446, agosto 2005.
- [32] Fang, K., Kotz, S., y Ng, K. W. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, 1 edición, noviembre 1989.
- [33] Farhangi, H. The path of the smart grid. *IEEE Power and Energy Magazine*, 8(1):18–28, febrero 2010.

- [34] Ferrero, R., Shahidehpour, S., y Ramesh, V. Transaction analysis in deregulated power systems using game theory. *IEEE Transactions on Power Systems*, 12(3):1340–1347, 1997.
- [35] Free Software Foundation. GNU general public license versión 3.0. Free Software Foundation. Visitado el 15 de noviembre de 2011.
URL <http://www.gnu.org/licenses/gpl.txt>
- [36] García-González, J., de la Muela, R., Santos, L., y Gonzalez, A. Stochastic joint optimization of wind generation and Pumped-Storage units in an electricity market. *IEEE Transactions on Power Systems*, 23(2):460–468, 2008.
- [37] García Lobo, M. *Métodos de predicción de la generación agragada de energía eólica*. Tesis de doctorado, Universidad Carlos III de Madrid. Departamento de Ingeniería Eléctrica, Leganés, España, noviembre 2010.
URL <http://hdl.handle.net/10016/10275>
- [38] Google. Google maps. Visitado el 10 de septiembre de 2011.
URL <http://maps.google.com/>
- [39] Grothe, O. y Schnieders, J. Spatial dependence in wind and optimal wind power allocation: A copula-based analysis. *Energy Policy*, 39(9):4742–4754, septiembre 2011.
- [40] Hahn, H., Meyer-Nieberg, S., y Pickl, S. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 199(3):902–907, diciembre 2009.
- [41] Hamilton, J. D. Historical oil shocks. *National Bureau of Economic Research Working Paper Series*, No. 16790, 2011.
- [42] Hersbach, H. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15:559–570, octubre 2000.
- [43] Holttinen, H. y Ikäheimo, J. Wind prediction and bids in denmark. Informe técnico, VTT Energia, junio 2007.
- [44] Härdle, W. *Nonparametric and semiparametric models*. Springer, mayo 2004.
- [45] Iberdrola S.A. Iberdrola. Visitado el 20 de septiembre de 2011.
URL <http://www.iberdrola.es/>
- [46] Jalili-Marandi, V. y Dinavahi, V. Large-scale transient stability simulation on graphics processing units. En *IEEE Power & Energy Society General Meeting, 2009. PES '09*, páginas 1–6. IEEE, Calgary, Canada, julio 2009.

- [47] Jaramillo Duque, Á. Cosas para compartir. Visitado el 11 de noviembre de 2011.
URL <http://electronica.uc3m.es/~ajaramil/cosas.html>
- [48] Jaramillo Duque, Á. *Operación óptima de una planta hidráulica reversible que cubre los desvíos en la predicción de un productor eólico*. Tesis de máster, Departamento de Ingeniería Eléctrica, Universidad Carlos III de Madrid, Leganés, España, septiembre 2009.
URL <http://hdl.handle.net/10016/11761>
- [49] Jaramillo Duque, Á., Castronuovo, E. D., Sánchez, I., y Usaola, J. Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer. *Electric Power Systems Research*, 81(9):1767–1777, septiembre 2011.
- [50] Jaramillo Duque, Á., Sánchez, I., Castronuovo, E., y Usaola, J. Simulating scenarios and prediction intervals in wind power forecasting with the beta distribution. En *11th. Spanish-Portuguese Conference on Electrical Engineering (11CHLIE)*, página 5. Zaragoza, Spain, junio 2009.
- [51] Juban, J., Siebert, N., y Kariniotakis, G. N. Probabilistic short-term wind power forecasting for the optimal management of wind generation. En *Power Tech, 2007 IEEE*, páginas 683–688. IEEE, Lausanne, Suiza, julio 2007.
- [52] Koepfel, G. y Korpås, M. Increasing the network in-feed accuracy of wind turbines with energy storage devices. En *The Sixht World Energy System Conference*. Torino, Italia, julio 2006.
- [53] Kristiansen, T. y Wangensteen, I. Effect of losses on area prices in the norwegian electricity market. *International Journal of Electrical Power & Energy Systems*, 28(1):43–47, enero 2006.
- [54] Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., y Yan, Z. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915–920, mayo 2009.
- [55] Leite da Silva, A. M., Chaves de Resende, L., da Fonseca Manso, L. A., y Miranda, V. Composite reliability assessment based on Monte Carlo simulation and artificial neural networks. *IEEE Transactions on Power Systems*, 22(3):1202–1209, agosto 2007.
- [56] Lincoln, R., Galloway, S., y Burt, G. Open source, agent-based energy market simulation with python. En *2009 6th International Conference on the European Energy Market*, páginas 1–5. Leuven, Bélgica, mayo 2009. doi:10.1109/EEM.2009.5207125.
- [57] Lindsay Anderson, C. y Zimmerman, R. Wind output forecasts and scenario analysis for stochastic multiperiod optimum

- power flow. Power Systems Engineering Research Center (PSERC), noviembre 2011. Visitado el 20 de noviembre de 2011.
URL <http://mediasite.engr.wisc.edu/Mediasite/Viewer/?peid=c8c36aab10ac40b092c33e4a68fa20c31d>
- [58] Luickx, P. J., Delarue, E. D., y D'haeseleer, W. D. The examination of different energy storage methods for wind power integration. Artículo en preparación WP EN2009-001, Katholieke Universiteit Leuven, Energy Institute, 2009.
URL <http://www.mech.kuleuven.be/energy/>
- [59] Matevosyan, J. y Söder, L. Short-term hydropower planning coordinated with wind power in areas with congestion problems. *Wind Energy*, 10(3):195–208, 2007.
- [60] MATLAB. *version 7.11 (R2010b)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [61] Milano, F. y Vanfretti, L. State of the art and future of OSS for power systems. En *IEEE Power & Energy Society General Meeting, 2009. PES '09*, páginas 1–7. IEEE, Calgary, Canada, julio 2009.
- [62] Ministerio de Industria, Turismo y Comercio de España. Planificación de los sectores de electricidad y gas 2008-2016. Informe técnico, Gobierno de España, mayo 2008.
URL <http://www.mityc.es/>
- [63] Ministerio de Industria, Turismo y Comercio de España. Plan de acción nacional de energías renovables 2011-2020. Informe técnico, Gobierno de España, junio 2010.
URL <http://www.mityc.es/>
- [64] Mohd, A., Ortjohann, E., Schmelter, A., Hamsic, N., y Morton, D. Challenges in integrating distributed energy storage systems into future smart grid. En *IEEE International Symposium on Industrial Electronics, 2008. ISIE 2008*, páginas 1627–1632. IEEE, Cambridge, Reino Unido, junio 2008.
- [65] Momoh, J. A. Smart grid design for efficient and flexible power networks operation and control. En *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, páginas 1–8. IEEE, Seattle, EE.UU., marzo 2009.
- [66] Muhaisen, O. S. H. *Simulación basada en la teoría de valor extremo y copula para estimación de caudales en sistema de sanitaria*. Tesis de doctorado, Departamento de Ingeniería Civil, Universidad de Granada, Granada, España, 2006.
URL <http://digibug.ugr.es/bitstream/10481/1045/1/1618399x.pdf>

- [67] Nadaraya, E. A. On estimating regression. *Theory of Probability and its Applications*, 9(1):141, 1964.
- [68] Nelsen, R. B. *An Introduction to Copulas*. Springer, New York, USA, 1 edición, octubre 1998.
- [69] Ngala, G., Alkali, B., y Aji, M. Viability of wind energy as a power generation source in Maiduguri, Borno state, Nigeria. *Renewable Energy*, 32(13):2242–2246, octubre 2007.
- [70] OMIE. Mercado de electricidad | OMIE. Visitado el 13 de enero de 2012.
URL <http://www.omie.es/inicio/mercados-y-productos>
- [71] Papaefthymiou, G. y Kurowicka, D. Using copulas for modeling stochastic dependence in power system uncertainty analysis. *IEEE Transactions on Power Systems*, 24(1):40–49, 2009.
- [72] Papaefthymiou, G. y Pinson, P. Modeling of spatial dependence in wind power forecast uncertainty. En *Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems*, páginas 1–9. Rincon, Puerto Rico, mayo 2008.
- [73] Peña, D. *Regresión y diseño de experimentos*. Alianza, Madrid, 2002.
- [74] Pinson, P., George, P., Bernd, K., y Jody, V. Dynamic sizing of energy storage for hedging wind power forecast uncertainty. En *IEEE Power & Energy Society General Meeting*. Calgary, Canada, julio 2009.
- [75] Pinson, P. y Kariniotakis, G. N. Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. En *Power Tech Conference Proceedings, 2003 IEEE Bologna*, tomo 2. IEEE, Bologna, Italia, junio 2003.
- [76] Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., y Klöckl, B. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.
- [77] Pudjianto, D., Ramsay, C., y Strbac, G. Virtual power plant and system integration of distributed energy resources. *IET Renewable Power Generation*, 1(1):10–16, marzo 2007.
- [78] Red Eléctrica de España, S. A. Sistema de información del operador del sistema. Visitado el 15 de febrero de 2010.
URL <http://www.esios.ree.es/web-publica/>
- [79] Red Eléctrica de España (REE). Informe del sistema eléctrico español en 2010. Informe técnico, Red Eléctrica de España (REE),

- España, julio 2011.
URL <http://www.ree.es/>
- [80] Rey Porto, M., Garde, R., Martínez Campillo, R., Remke, S., Hendriks, M., Ribera Fumaz, J. M., Peña Alzola, R., y Fernández, R. S. Guía del almacenamiento de energía. Informe técnico, Dirección General de Industria, Energía y Minas y la Fundación de la Energía de la Comunidad de Madrid, octubre 2011.
URL <http://www.fenercom.com>
- [81] Rodríguez Amenedo, J. L., Burgos Díaz, J. C., y Arnalte Gómez, S., editores. *Sistemas eólicos de producción de energía eléctrica*. Editorial Rueda, 2003.
- [82] Shafiee, S. y Topal, E. When will fossil fuel reserves be diminished? *Energy Policy*, 37(1):181–189, enero 2009.
- [83] Siemens. Siemens - PSS®E. Visitado el 11 de diciembre de 2011.
URL <http://www.energy.siemens.com/>
- [84] Simek, V. y Asn, R. R. GPU acceleration of 2D-DWT image compression in MATLAB with CUDA. En *Second UKSIM European Symposium on Computer Modeling and Simulation, 2008*. EMS '08, páginas 274–277. IEEE, Liverpool, Reino Unido, septiembre 2008.
- [85] Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–23, 1959.
- [86] Spanjer, A. Russian gas price reform and the EU–Russia gas relationship: Incentives, consequences and european security of supply. *Energy Policy*, 35(5):2889–2898, mayo 2007.
- [87] Summaries of EU legislation. Green paper on the security of energy supply. Summaries of EU legislation, febrero 2007. Visitado el 20 de noviembre de 2011.
URL http://europa.eu/legislation_summaries/energy/external_dimension_enlargement/l27037_en.htm
- [88] Sánchez, I. Short-term prediction of wind energy production. *International Journal of Forecasting*, 22(1):43–56, 2006.
- [89] Sánchez, I. Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, 24(4):679–693, octubre 2008.
- [90] Sánchez, I., Usaola, J., Ravelo, O., Velasco, C., Domínguez, J., Lobo, M. G., González, G., Soto, F., Diaz-Guerra, B., y Alonso, M. Sipleólico-a wind power prediction system based on flexible combination of dynamic models, application to the spanish power system. En *World Wind Energy Conference*. Berlin, Germany, 2002.

- [91] Thomas, R. J., Murillo-Sanchez, C. E., y Zimmerman, R. D. An advanced security constrained OPF that produces correct market-based pricing. En *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, páginas 1–6. IEEE, Pittsburgh, EE.UU., julio 2008.
- [92] Tinney, W. F. y Hart, C. E. Power flow solution by Newton's method. *IEEE Transactions on Power Apparatus and Systems*, PAS-86(11):1449–1460, noviembre 1967.
- [93] Tsikalakis, A., Katsigiannis, Y., Georgilakis, P., y Hatziargyriou, N. Determining and exploiting the distribution function of wind power forecasting error for the economic operation of autonomous power systems. En *Power Engineering Society General Meeting, 2006. IEEE*, página 8 pp. Montreal, Canada, junio 2006.
- [94] Ummels, B. C., Pelgrum, E., y Kling, W. L. Integration of large-scale wind power and use of energy storage in the Netherlands' electricity supply. *IET Renewable Power Generation*, 2(1):34–46, marzo 2008.
- [95] University of Washington. Power systems test case archive. Visitado el 5 de noviembre de 2010.
URL <http://www.ee.washington.edu/research/pstca/>
- [96] Usaola, J. Probabilistic load flow with wind production uncertainty using cumulants and Cornish-Fisher expansion. *International Journal of Electrical Power & Energy Systems*, 31(9):474–481, octubre 2009.
- [97] Valizadeh Haghi, H., Tavakoli Bina, M., Golkar, M., y Moghaddas-Tafreshi, S. Using copulas for analysis of large datasets in renewable distributed generation: PV and wind power integration in Iran. *Renewable Energy*, 35(9):1991–2000, septiembre 2010.
- [98] Vieira, F. y Ramos, H. Hybrid solution and pump-storage optimization in water supply system efficiency: A case study. *Energy Policy*, 36(11):4142–4148, noviembre 2008.
- [99] Villanueva, D., Pazos, J. L., y Feijoo, A. Probabilistic load flow including wind power generation. *IEEE Transactions on Power Systems*, 26(3):1659–1667, agosto 2011.
- [100] Vlachogiannis, J. G. Probabilistic constrained load flow considering integration of wind power generation and electric vehicles. *IEEE Transactions on Power Systems*, 24(4):1808–1817, noviembre 2009.

- [101] Watson, G. S. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.
- [102] World Wind Energy Association (WWEA). WWEA half-year report 2011. Informe técnico, World Wind Energy Association (WWEA), julio 2011.
URL <http://wwindea.org/>
- [103] Yan, J. Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21, octubre 2007.
URL <http://www.jstatsoft.org/v21/i04>
- [104] Yang, B., Makarov, Y., Desteese, J., Viswanathan, V., Nyeng, P., McManus, B., y Pease, J. On the use of energy storage technologies for regulation services in electric power systems with significant penetration of wind energy. En *2008 5th International Conference on the European Electricity Market*, páginas 1–6. Lisboa, Portugal, 2008.
- [105] Yang, K., Garba, A., Tan, C., y Lo, K. The impact of the wind generation on reactive power requirement and voltage profile. En *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, páginas 866–871. Nanjuing, China, abril 2008.
- [106] Zapata Londoño, W., Uribe Segura, J. J., Molina Castro, J. D., Jaramillo Duque, A., y Villada Duque, F. *Derivados financieros en el mercado eléctrico colombiano*. Imprenta Universidad de Antioquia, 1º edición, febrero 2010.
- [107] Zhang, H. y Li, P. Probabilistic analysis for optimal power flow under uncertainty. *IET Generation, Transmission & Distribution*, 4(5):553–561, mayo 2010.
- [108] Zimmerman, R., Murillo-Sánchez, C., y Thomas, R. MATPOWER: Steady-State operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, 2011.
- [109] Zimmerman, R. D. MATPOWER mail list. Visitado el 22 de junio de 2011.
URL <http://www.mail-archive.com/matpower-l@list.cornell.edu/>
- [110] Zimmerman, R. D., Murillo-Sanchez, C. E., y Thomas, R. J. MATPOWER's extensible optimal power flow architecture. En *IEEE Power & Energy Society General Meeting, 2009. PES '09*, páginas 1–7. IEEE, Calgary, Canada, julio 2009.
- [111] Zimmerman, R. D. y Murillo-Sánchez, C. E. *MATPOWER 4.1 User's Manual*. Power Systems Engineering Research Center

(PSERC), diciembre 2011.

URL <http://www.pserc.cornell.edu/matpower/manual.pdf>

- [112] Zimmerman, R. D., Murillo-Sánchez, C. E., y Gan, D. MATPOWER matlab power system simulation package - home page. Power Systems Engineering Research Center (PSERC). Visitado el 1 de diciembre de 2011.

URL <http://www.pserc.cornell.edu/matpower/>

COLOFÓN

Este documento fue editado con el estilo tipográfico `classicthesis` desarrollado por André Miede. El estilo fue inspirado por Robert Bringhurst del libro «*The Elements of Typographic Style*». El estilo tipográfico `classicthesis` está disponible tanto para \LaTeX como para \LyX :

<http://code.google.com/p/classicthesis/>

Algunos usuarios felices de `classicthesis` envían una postal al autor, una colección de postales recibidas hasta el momento se puede ver en el siguiente enlace:

<http://postcards.miede.de/>

Versión Final a 17 de mayo de 2012 (`classicthesis` version 4.0).