



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**MÉTODOS DE IDENTIFICACIÓN Y PREDICCIÓN DE RAMPAS
EN SISTEMAS ELÉCTRICOS CON GENERACIÓN
INTERMITENTE**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELECTRICISTA

TOMÁS ANDRÉS OGDEN ARAYA

PROFESOR GUÍA

LUIS SANTIAGO VARGAS DÍAZ

MIEMBROS DE LA COMISIÓN

MARCOS EDUARDO ORCHARD CONCHA

ERNESTO RICARDO HUBER JARA

SANTIAGO DE CHILE

2015

MÉTODOS DE IDENTIFICACIÓN Y PREDICCIÓN DE RAMPAS EN SISTEMAS ELÉCTRICOS CON GENERACIÓN INTERMITENTE

La tendencia en todos los países desarrollados y en vías de desarrollo en el mundo es al alza en cuanto al consumo de energía eléctrica. Esto lleva a buscar alternativas de recursos mediante los cuales suplir esta creciente demanda y que a su vez sean amigables con el medio ambiente. En de éste ámbito surgen las ERNC, y dentro de ellas el viento asoma como una fuente de gran potencial.

Si bien es cierto que el viento es una fuente limpia y gratuita de energía, no está libre de problemas. El hecho de que sea un recurso con un régimen aleatorio hace que su penetración en sistemas eléctricos presente ciertos desafíos. Dentro de estos desafíos están los asociados a las rampas de viento, aumentos o disminuciones repentinas de la potencia extraíble de este recurso. Estas rampas indudablemente son un problema ya que comprometen la operación óptima del sistema, desde el punto de vista técnico y económico.

En este trabajo se desarrolló una herramienta de predicción de rampas desde un punto de vista estadístico/probabilístico mediante redes Bayesianas. En base a datos correspondientes a la central Canela I se entrenó y validó el modelo. La validación se hizo mediante la comparación de los resultados obtenidos del modelo versus los datos reales. Se compararon distintos resultados mediante una función de puntaje asociada a los verdaderos positivos, falsos positivos y falsos negativos obtenidos para cada caso. Ésta función asigna puntaje entre 0 y 1, donde 1 corresponde a la predicción perfecta.

Inicialmente se probó el modelo para tiempos de anticipación de una, dos y tres horas. Los tres casos dieron resultados parecidos, con un puntaje que llega a 0,52. Como segundo caso se probó pronosticar rampas en una ventana de dos horas, obteniéndose en este caso un mejor resultado con un puntaje de 0,7. Finalmente éste se comparó con un modelo autorregresivo (AR), resultando mejor el modelo de red Bayesiana, con puntaje de 0,7 frente a 0,41 del modelo AR.

Como trabajo futuro se propone probar el método con datos de otras centrales, además de mejorar el modelo para incluir la estacionalidad del recurso.

A mis padres

*La vida no me pudo haber dado un mejor regalo
que el tenerlos a ustedes*

Agradecimientos

Primero y antes que todo tengo que agradecer a mis padres, mi ejemplo, mi pilar y mi alegría. Creo que las palabras no alcanzan para explicar, para expresar, lo agradecido que estoy de ustedes. La vida me bendijo desde mi primer día al tenerlos a ustedes como padres.

A mis abuelos, que de pequeño me han tratado con tanto cariño. Les agradezco infinitamente su preocupación, sus consejos y su ayuda. También a mi familia, tíos, primos, todos. Muchas gracias por el constante apoyo y palabras de ánimo.

También quiero agradecer al grupo, Maripoli+cuatro, mis compañeros de mil batallas, con los que compartimos las amarguras de las derrotas y las alegrías de los éxitos. Maripoli, Jaime, Ramiro, Pelao, sin ustedes esta etapa, la más feliz que he tenido en la vida, no hubiera sido ni remotamente cercana a lo buena que fue. Muchas gracias por cada sonrisa que me sacaron, por cada cerveza/piscola que compartimos, por todos los buenos momentos y por el apoyo que me dieron en los no tan buenos.

A mi amiga de la vida, Sofía. Tantos momentos que hemos compartido, desde que nos conocimos una tarde en el colegio hace ya tanto tiempo. Contigo he compartido tanto, tantas historias que se me vienen a la mente. Me has visto a través de todo este proceso, y en el mismo has sido un apoyo constante e incondicional a pesar de todo.

A la Camila. Te vine a conocer al final de esta etapa pero aun así no concibo terminarla si ti. Ha sido poco tiempo y a la vez parece tanto. Te agradezco todo el apoyo que me has dado este último año y toda la paciencia que me has tenido. Siempre has sido la sonrisa en los momentos de desánimo.

A la banda, Demolición, los que están y los que estuvieron. Que sería la vida sin un poco de música y sin buenos amigos.

A mis amigos del colegio, que nos conocemos hace tanto y aún estamos ahí. Muchas gracias por todos los buenos momentos compartidos.

A mi profesor guía, profesor Luís Vargas, por su apoyo y preocupación, fundamentales en el desarrollo de este trabajo.

También un agradecimiento grande a mi profesor co-guía, señor Marcos Orchard. Se agradece enormemente la buena disposición y la ayuda brindada. Indudablemente sin su cooperación este trabajo no hubiera llegado a este buen término.

Finalmente a todos con los que compartí algún momento, una conversación, un trago, risas, tantas buenas cosas. Me voy con los mejores recuerdos.

Tabla de contenido

1	Introducción	1
1.1	Motivación y descripción del problema.....	2
1.2	Objetivo general	3
1.3	Objetivos específicos	3
2	Revisión Bibliográfica.....	4
2.1	El recurso eólico.....	4
2.2	Generación Eólica en Chile.....	5
2.3	Variabilidad del viento	7
2.4	Rampas de potencia	8
2.5	Efectos de las rampas de viento en sistemas con generación eólica ...	9
2.6	Fenómeno de rampas en plantas de generación en Chile	11
2.7	Métodos para el pronóstico de la potencia eólica	13
2.8	Pronóstico de rampas	14
2.9	Marco matemático.....	15
2.9.1	Variables aleatorias y probabilidades condicionales.....	15
2.9.2	Redes Bayesianas.....	19
3	Metodología	26
3.1	Procesamiento de datos.....	26
3.2	Rampas.....	27
3.3	Construcción de la red Bayesiana.....	30
3.3.1	Hora del día:	30
3.3.2	Potencia generada:.....	30
3.3.3	Rampas anteriores	32
3.3.4	Red Bayesiana	32
3.3.5	Obtención de parámetros	33
3.4	Obtención de resultados	39
3.5	Sensibilidad frente a la magnitud de la rampa	40
3.6	Comparación de diferentes resultados.....	44
3.7	Comparación de métodos: Modelo autorregresivo (AR)	45
3.8	Diagrama de bloques	46
4	Resultados.....	47
4.1	Primer caso: rampas de magnitud 2 [MW]	47
4.1.1	Probabilidad de ocurrencia de rampa	47
4.1.2	Sensibilidad respecto al umbral.....	49
4.1.3	Incidencia de falsos negativos	57
4.1.4	Predicción en una ventana horaria	61

4.2	Segundo caso: rampas de magnitud 3 [MW]	65
4.2.1	Probabilidad de ocurrencia de las rampas.....	65
4.2.2	Sensibilidad respecto al umbral.....	67
4.2.3	Predicción en una ventana horaria	74
4.3	Comparación con modelo AR	76
5	Conclusiones y trabajo a futuro	79
5.1	Conclusiones.....	79
5.2	Trabajo futuro.....	81
6	Bibliografía.....	82
7	Anexos.....	85
7.1	Anexo A.....	85
7.2	Anexo B.....	86
7.3	Anexo C	87
7.4	Anexo D	89

Índice de Figuras

Figura 1. Direcciones globales del desplazamiento de los vientos.....	4
Figura 2. Porcentajes según tipo de generación en el SIC al año 2007.....	5
Figura 3. Porcentajes de generación de energía el año 2013 por tipo de central	7
Figura 4. Variabilidad del viento: (A) instantánea; (B) diaria; (C) estacional; (D) anual	8
Figura 5. Rampas de potencia de subida y de bajada	8
Figura 6. Parámetros relevantes en los eventos de rampas	9
Figura 7. Potencia despachada por la planta Canela I y II el 4 de junio del 2014	11
Figura 8. Potencia despachada por la planta Canela I y II el 5 de junio del 2014	12
Figura 9. Potencia despachada por la planta Canela I y II el 6 de junio del 2014	12
Figura 10. Potencia despachada por la planta Canela I y II el 7 de junio del 2014	13
Figura 11. Ilustración de un sistema de pronóstico del recurso eólico	13
Figura 12. Ilustración de dos eventos en que uno condiciona a otro.....	17
Figura 13. Representación de evento como nodo de una red.....	19
Figura 14. Red Bayesiana incluyendo fenómenos que afectan el evento estudiado.....	20
Figura 15. Posibles direcciones de los arcos entre dos nodos.....	24
Figura 16. Agrupación de datos para entrenamiento y prueba del modelo	27
Figura 17. Reducción de número de datos mediante promedios calculados cada una hora	27
Figura 18. Obtención de ocurrencia de rampas desde los datos originales: el gráfico superior corresponde a los datos intermedios, el gráfico central corresponde al primer procesamiento donde se aprecian las rampas con su magnitud y sentido, y el gráfico inferior refleja el resultado final del procesamiento, reduciendo las rampas a una variable binaria de tipo “ocurre o no ocurre”.....	29
Figura 19. Distribución horaria de rampas ocurridas durante noviembre y diciembre del 2012.....	30
Figura 20. Autocorrelación de los datos de la potencia generada. El eje LAG corresponde a los intervalos temporales considerados (representados por saltos de una hora cada uno).....	31
Figura 21. Autocorrelación de las rampas ocurridas. El eje LAG corresponde a los intervalos temporales considerados (representados por saltos de una hora cada uno)	32

Figura 22. Red Bayesiana empleada en el predictor.....	33
Figura 23. Diagrama de bloques de la metodología.....	46
Figura 24. Probabilidades de ocurrencia de rampas mayores a 2 [MW] para tiempo de anticipación de una hora y rampas realmente ocurridas	48
Figura 25. Probabilidades de ocurrencia de rampas mayores a 2 [MW] para tiempo de anticipación de dos horas y rampas realmente ocurridas.....	48
Figura 26. Probabilidades de ocurrencia de rampas mayores a 2 [MW] para tiempo de anticipación de tres horas y rampas realmente ocurridas.....	49
Figura 27. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora	51
Figura 28. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora. Debe considerarse el 100% como el total de rampas ocurridas en enero, las cuales fueron 92	51
Figura 29. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas	53
Figura 30. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas. El 100% son las 92 rampas ocurridas en enero	53
Figura 31. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas	55
Figura 32. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas. El 100% son las 92 rampas ocurridas en enero	55
Figura 33. Función Fscore para predicciones con una, dos y tres horas de anticipación	57
Figura 34. Esquema de predicción con dos horas de anticipación para una ventana horaria de dos horas.....	61
Figura 35. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas.....	63
Figura 36. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas. El 100% corresponde al número de ventanas de dos horas en el que si hubieron rampas, en este caso 150	63
Figura 37. Función Fscore para predicciones con una, dos y tres horas de anticipación, y empleando una ventana de predicción de dos horas	64
Figura 38. Probabilidades de ocurrencia de rampas mayores a 3 [MW] para tiempo de anticipación de una hora y rampas realmente ocurridas	65
Figura 39. Probabilidades de ocurrencia de rampas mayores a 3 [MW] para tiempo de anticipación de dos horas y rampas realmente ocurridas.....	66

Figura 40. Probabilidades de ocurrencia de rampas mayores a 3 [MW] para tiempo de anticipación de tres horas y rampas realmente ocurridas.....	66
Figura 41. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora y rampas de al menos 3 [MW].....	68
Figura 42. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora y rampas de al menos 3 [MW]. Debe considerarse el 100% como las rampas de esta magnitud ocurridas en enero, que fueron en total 4	68
Figura 43. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas y rampas de al menos 3 [MW].....	70
Figura 44. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas y rampas de al menos 3 [MW]. Debe considerarse el 100% como las rampas de esta magnitud ocurridas	70
Figura 45. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas y rampas de al menos 3 [MW].....	72
Figura 46. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas y rampas de al menos 3 [MW]. Debe considerarse el 100% como las rampas de esta magnitud ocurrida	72
Figura 47. Puntajes asociados a la función Fscore para los tres tiempos de anticipación usados.....	73
Figura 48. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas, considerando rampas de magnitud de al menos 3 [MW].....	75
Figura 49. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas, considerando rampas de magnitud de al menos 3 [MW]. El 100% corresponde al número de ventanas de dos horas en el que si hubieron rampas, 4 para este caso	75
Figura 50. Función Fscore para predicciones con una, dos y tres horas de anticipación, y empleando una ventana de predicción de dos horas, considerando rampas de al menos 3 [MW] de magnitud	76
Figura 51. Potencia real y pronosticada para el mes de enero empleando un modelo AR(3)	77
Figura 52. Puntajes del modelo de red Bayesiana para predicción en ventana de dos horas y para el modelo AR. Debe tenerse en cuenta que el puntaje asociado al modelo AR no depende del umbral, es un valor fijo, por ello en la gráfica se aprecia paralela al eje X.	78

Índice de Tablas

Tabla 1. Direcciones globales de los vientos.....	4
Tabla 2. Plantas eólicas del SIC al año 2014	6
Tabla 3. Generación de energía en el SIC el año 2013	6
Tabla 4. Rampas de noviembre, diciembre y enero.....	30
Tabla 5. Probabilidades de que la potencia generada esté en un rango específico	34
Tabla 6. Cálculo de probabilidad a priori de ocurrencia de rampas.....	35
Tabla 7. Probabilidad condicionada a la potencia para una anticipación de una hora	36
Tabla 8. Probabilidad condicionada a la potencia para una anticipación de dos horas	37
Tabla 9. Probabilidad condicionada a la potencia para una anticipación de tres horas	38
Tabla 10. Probabilidad condicionada a la hora del día.....	39
Tabla 11. Cálculo de probabilidad a priori de ocurrencia de rampas para rampas de magnitud 2[MW]	40
Tabla 12. Probabilidad condicionada a la hora del día para rampas de magnitud 3[MW].....	41
Tabla 13. Probabilidad condicionada a la potencia para una anticipación de una hora para rampas de magnitud 3[MW]	42
Tabla 14. Probabilidad condicionada a la potencia para una anticipación de dos horas para rampas de magnitud 3[MW]	43
Tabla 15. Probabilidad condicionada a la potencia para una anticipación de tres horas para rampas de magnitud 3[MW]	44
Tabla 16. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de una hora	50
Tabla 17. Verdaderos positivos, falsos positivos y falsos negativos obtenidos con un tiempo de anticipación de dos horas	52
Tabla 18. Verdaderos positivos, falsos positivos y falsos negativos obtenidos con un tiempo de anticipación de tres horas	54
Tabla 19. Falsos negativos separados por magnitud para tiempo de anticipación de una hora	58
Tabla 20. Falsos negativos separados por magnitud para tiempo de anticipación de dos horas.....	59
Tabla 21. Falsos negativos separados por magnitud para tiempo de anticipación de tres horas	60
Tabla 22. Verdaderos positivos, falsos positivos y falsos negativos obtenidos prediciendo en base a una probabilidad acumulada en una ventana de dos horas	62

Tabla 23. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de una hora, para rampas de magnitud mayor a 3 [MW].....	67
Tabla 24. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de dos horas, para rampas de magnitud mayor a 3 [MW].....	69
Tabla 25. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de tres horas, para rampas de magnitud mayor a 3 [MW].....	71
Tabla 26. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para rampas de al menos 3 [MW], prediciendo en base a una probabilidad acumulada en una ventana de dos horas	74
Tabla 27. Predicciones acertadas, falsos positivos y falsos negativos obtenidos con el modelo AR(3).....	77
Tabla 28. Precisión, Recall y puntaje Fscore obtenidos con el modelo AR(3) .	77

1 Introducción

Es un hecho que la tendencia mundial de la demanda de energía eléctrica en países desarrollados es creciente, por lo que cada año los sistemas eléctricos deben ampliarse y nuevas centrales generadoras son necesarias para poder satisfacer este aumento de la demanda. Es este también el caso de Chile, en el que el consumo eléctrico ha aumentado a una tasa de 6,7 % en los últimos 20 años. Esta demanda ha sido atendida mayoritariamente mediante centrales hidroeléctricas y termoeléctricas en base a petróleo, carbón o gas. Esto genera una gran dependencia de los recursos extranjeros ya que Chile no es un país productor de los combustibles fósiles requeridos [1].

Por otra parte, cada vez se ha hecho más importante el minimizar las emisiones de gases de efecto invernadero, emisiones que se deben en gran parte a la generación eléctrica mediante plantas de tipo termoeléctrico que emplean combustibles fósiles como parte de su proceso de generación. Es en este punto donde las energías renovables no convencionales están apareciendo como una fuente con gran potencial y cada vez más relevante para la generación de energía eléctrica.

En el caso de Chile, la energía renovable no convencional que ha tomado más fuerza en el último tiempo es la energía eólica. Su uso para la generación eléctrica comenzó el año 2007 con la puesta en marcha de la central Canela I. Esta planta se amplió y en el año 2009 entró en operación su segunda etapa, Canela II. Así hasta ahora se han construido y han entrado en operación un total de diez parques eólicos, los cuales tienen una potencia total instalada del orden de los 500 [MW]

Pese a que la energía eólica es limpia e ilimitada, tiene sus desventajas. El principal problema que presenta la generación eólica es que el viento es un recurso que presenta un comportamiento aleatorio, por lo que es imposible tener certeza de cuanta potencia instantánea se puede generar en base al viento. Parte del comportamiento aleatorio del viento es la aparición de fenómenos conocidos como rampas, en los que la velocidad del viento aumenta o disminuye bruscamente, por lo que la potencia que se puede generar en base a él también sufre cambios bruscos. Es por ello que disponer de una herramienta capaz de predecir la ocurrencia de estos fenómenos es imprescindible para la operación de un sistema eléctrico en el cual la energía eólica tenga un alto grado de penetración.

1.1 Motivación y descripción del problema

La energía eólica tiene un gran potencial para la generación de energía eléctrica. Tiene grandes ventajas como la de ser una fuente limpia de generación, con lo que se evita la emanación de gases de efecto invernadero. Además es un recurso prácticamente ilimitado, generado por procesos atmosféricos a grandes escalas que se dan en el planeta.

A pesar de tener sendas ventajas sobre otros recursos, el recurso eólico no está exento de inconvenientes. El principal problema de este recurso es que su comportamiento es aleatorio, por lo que un sistema en el que se presente una alta penetración eólica presentará un alto grado de incerteza de la potencia que será capaz de generar en cada momento.

Uno de los fenómenos característicos del viento y su comportamiento aleatorio es la aparición de rampas. Estas rampas son aumentos o disminuciones considerables del recurso eólico (de su velocidad) en periodos relativamente cortos de tiempo. Debido a que la potencia que se puede generar en base al viento depende directamente de su velocidad, estos fenómenos tienen gran repercusión en la operación de sistemas en los que esté presente la generación en base eólica. Es entonces imprescindible para la operación de un sistema eléctrico en el cuál esté presente la generación eólica contar con instrumentos que sean capaces de anticiparse a estos fenómenos con un margen de tiempo adecuado para que los operadores puedan generar acciones para así contrarrestar estos cambios intempestivos.

La motivación de este trabajo surge de la necesidad de contar con una herramienta que sea capaz de predecir la ocurrencia de fenómenos de rampas en el recurso eólico con cierto tiempo de anticipación suficiente para que los operadores puedan tomar medidas y atenuar los efectos que estos fenómenos puedan producir en la red eléctrica.

1.2 Objetivo general

El objetivo general de este trabajo es generar una herramienta capaz de calcular la probabilidad de ocurrencia de rampas en el recurso eólico y en base a dicha probabilidad tomar una decisión acerca de la ocurrencia de este fenómeno, y que además sea capaz de anticiparlos con un margen de tiempo suficiente para permitir tomar acciones de control a fin de mitigar los problemas asociados a la presencia de rampas de viento.

1.3 Objetivos específicos

Establecer el estado del arte de los métodos de pronóstico de rampas en el recurso eólico, reconociendo sus fortalezas y debilidades.

Crear un método capaz de calcular la probabilidad de ocurrencia de rampas del recurso eólico y discernir acerca de la ocurrencia de estos (decidir si se da la alarma o no), anticipándose en tres horas a la ocurrencia el fenómeno.

Validar el método en base a mediciones reales de la potencia generada por la central Canela I durante los meses de noviembre y diciembre del 2012 y enero del 2013.

2 Revisión Bibliográfica

2.1 El recurso eólico

En la superficie terrestre se produce el movimiento de grandes masas de aire. Dicho movimiento tiene su origen en tres distintos fenómenos:

- Radiación solar, la cual es desigual en la superficie terrestre, provocando diferencias de presión en la atmósfera.
- Rotación terrestre, la cual desplaza a las masas de aire por roce.
- Perturbaciones de la atmósfera terrestre, generando zonas de altas y bajas presiones que generan desplazamiento del aire.

Debido a la forma de la Tierra, la atmósfera terrestre se calienta de manera desigual, ya que recibe mucha más radiación en el ecuador que en los polos. Esta diferencia genera zonas de altas presiones en la zona de los polos y zonas de bajas presiones en el ecuador. Sin embargo los vientos no se mueven de sur a norte debido a la rotación terrestre y las fuerzas de Coriolis que esto genera. Así, a nivel mundial se aprecian corrientes de viento desde los trópicos (30° norte y sur) hacia el Ecuador denominados “alisios”. Adicionalmente se identifican vientos desde los trópicos hacia los círculos polares (60° norte y sur) denominados “contra-alisios”.

Tabla 1. Direcciones globales de los vientos

Latitud	Dirección
90-60°N	NE
60-30°N	SO
30-0°N	NE
0-30°S	SE
30-60°S	NO
60-90°S	SE

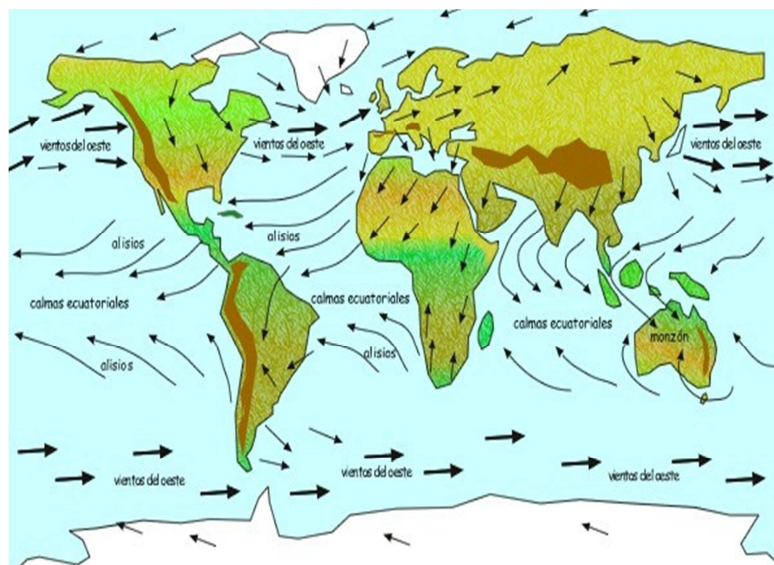


Figura 1. Direcciones globales del desplazamiento de los vientos

Los vientos descritos anteriormente son conocidos como vientos geostróficos. Estos vientos se encuentran a gran altura, desde mil metros sobre el nivel del suelo. Sin embargo los vientos aprovechables para una posible generación eólica son los que están a menores alturas (del orden de 100 metros sobre el suelo o menos). Estos vientos superficiales son influenciados en gran manera por los obstáculos presentes o la rugosidad de la superficie terrestre. Es por ello que los vientos en la superficie tendrán diferentes direcciones a los vientos geostróficos.

2.2 Generación Eólica en Chile

El sistema interconectado chileno que más presencia tiene de energía de origen eólico es el Sistema Interconectado Central (SIC). Este sistema interconectado es el más grande del país, atendiendo a cerca del 90% de la población. Se extiende desde la ciudad de Taltal por el norte hasta la Isla Grande de Chiloé por el sur. El SIC tenía al 2007 una potencia instalada de 9228,2 [MW], la cual es generada con una combinación de recursos renovables y no renovables entre los que se cuentan centrales hidráulicas de embalse y de pasada, centrales térmicas (carbón, fuel, diesel y ciclo combinado de gas natural), generación mediante recurso eólico. Los porcentajes son mostrados en la Figura 2 [1]:

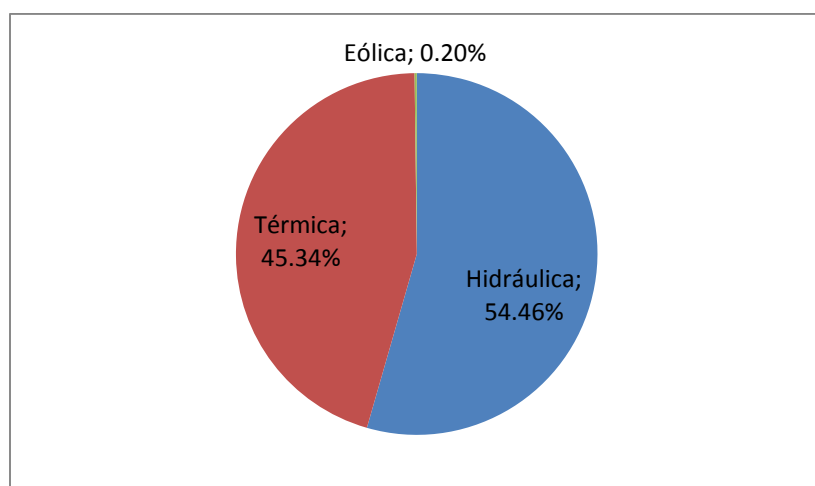


Figura 2. Porcentajes según tipo de generación en el SIC al año 2007

Si bien y tal como se mostró en la Figura 2 la generación eléctrica en base al viento aun es un porcentaje pequeño en comparación a la utilización de otros recursos, la tendencia es al alza en el uso del viento como potencial recurso para la generación de energía eléctrica. Desde la instalación del primer parque eólico en el SIC el año 2007 – parque Canela I – se han creado nueve nuevos parques, además de la ampliación del parque Canela con una segunda planta llamada Canela II. Así, desde la creación del parque eólico Canela I, con el cual se tenía una potencia instalada de origen eólico en el SIC de 18,15 [MW], este número ha crecido bastante con la instalación de otros parques, llegando actualmente a una potencia instalada de 480 [MW]. Las centrales actualmente en operación en el SIC se muestran en la Tabla 2:

Tabla 2. Plantas eólicas del SIC al año 2014

Planta	Ubicación (comuna)	Potencia instalada [MW]	Generación media anual [GWh]	Factor de planta [%]	Año de entrada en servicio
Canela I [2]	Los Vilos	18,15	47,14	29,65	2007
Canela II [2]	Los Vilos	60	140,5	26,73	2009
Lebu [3]	Lebu	3,5	8	26,09	2008
Totoral [4]	Canela	46	103	25,56	2010
Monte Redondo [5]	Ovalle	48	100	23,78	2009
Punta Colorada [6]	La Higuera	20	52,6	33,08	2011
Ucuquer [7]	Litueche	7,2	22	34,88	2012
Talinay [8]	Ovalle	90	200	25,37	2013
Negrete [9]	Negrete	36	85	26,95	2013
Arrayán [10]	Ovalle	115	280	27,29	2014
Los Cururos [11]	Ovalle	109,6	Entrada en operación reciente		2014
San Pedro [12]	Dalcahue	36	-	-	-

Es así que, pese a que aún es un porcentaje pequeño en cuanto a la generación total del SIC, la energía eólica poco a poco empieza a tener mayor relevancia. Se puede ver en la Tabla 3 la cantidad de energía total generada en el SIC el año 2013, donde como se aprecia en la Figura 3 el aporte de plantas eólicas fue de poco más del 1% del total anual [13].

Tabla 3. Generación de energía en el SIC el año 2013

Tipo de central	Generación total anual [GWh]
Pasada	1.0295,8
Térmicas	30.838,6
Embalse	9.222,3
Eólicas	546,8
Solar	2,7
Total	50.906,4

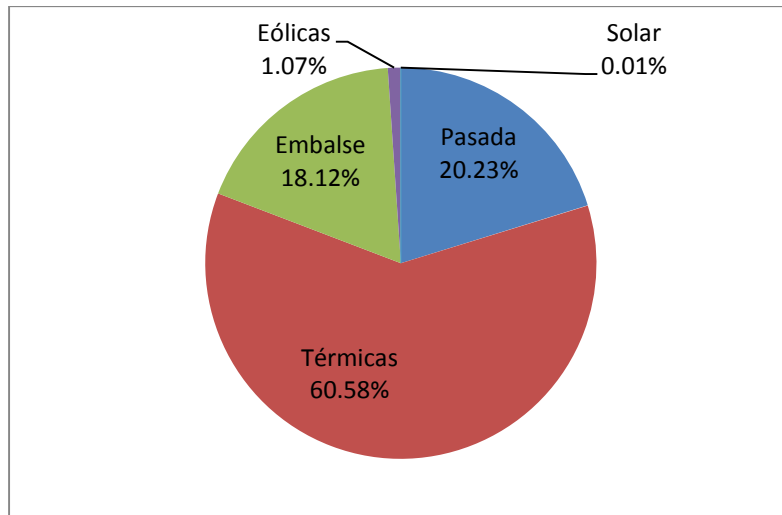


Figura 3. Porcentajes de generación de energía el año 2013 por tipo de central

2.3 Variabilidad del viento

La posibilidad de generar energía a partir del viento está condicionada por la velocidad que tiene este, la cual se puede definir en diferentes horizontes de tiempo:

- Variabilidad Instantánea: El viento varía constantemente debido a las fluctuaciones climáticas y a la geografía local y los obstáculos. Estas variaciones están en el rango de segundos a minutos.
- Variabilidad diaria: En general el viento es más veloz en el día que en la noche (este es el caso general, pero hay casos en que se da lo contrario, como sucede en España) debido principalmente a las diferencias de temperatura, diferencias que se acentúan en el día por la presencia del Sol.
- Variabilidad estacional: El viento también sufre variaciones de velocidad dependiendo de la estación del año. Es por ello que en zonas más templadas los vientos en invierno suelen ser más fuertes que en verano dado que los gradientes de temperatura son mayores.
- Variabilidad entre años: Las condiciones eólicas pueden variar de un año a otro, aunque generalmente estas variaciones no son muy grandes.

Estas características se pueden apreciar en la Figura 4.

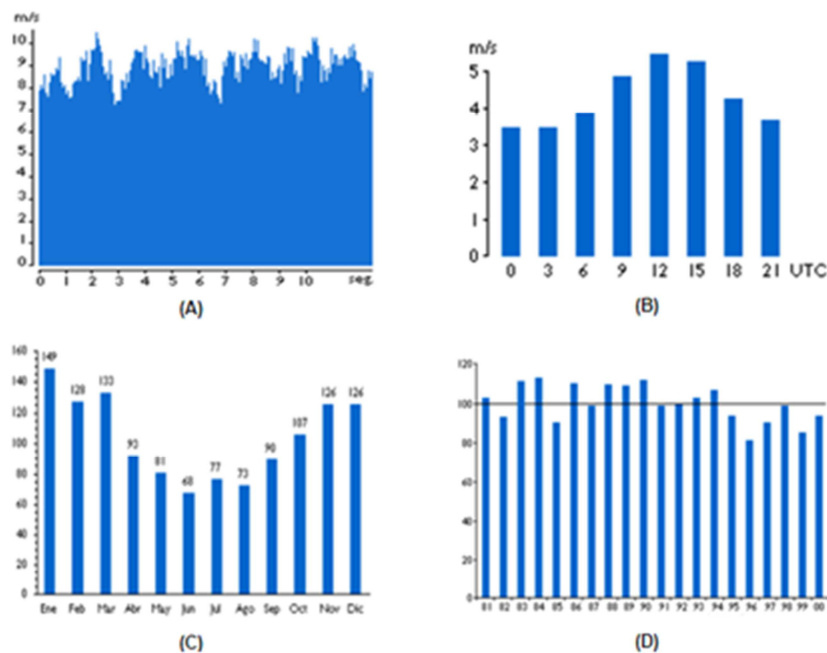


Figura 4. Variabilidad del viento: (A) instantánea; (B) diaria; (C) estacional; (D) anual

Estas variaciones en la velocidad del viento tienen directa repercusión en la potencia que se puede extraer del viento para la generación eléctrica. Esto es de especial importancia en la operación de los sistemas eléctricos. Fluctuaciones muy grandes en la potencia que se genera pueden llevar a que la generación de potencia no se mantenga equilibrada con la demanda de la misma, generando así inestabilidad en el sistema. Es por ello que un punto sumamente importante en la operación de sistemas con alta penetración de generación eólica es el pronóstico del recurso eólico.

2.4 Rampas de potencia

En cuanto a la operación del sistema eléctrico, un punto crítico a tener en consideración son las fluctuaciones repentinas del recurso eólico. Es aquí donde se reconocen fenómenos conocidos como “rampas”. Una rampa es un cambio significativo en la salida de un aerogenerador en un tiempo relativamente breve.

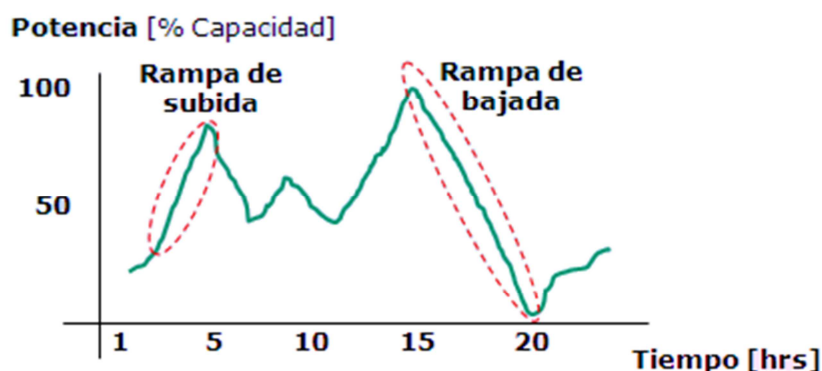


Figura 5. Rampas de potencia de subida y de bajada

Un evento de rampa queda determinado por los siguientes parámetros:

- Sentido: La rampa es ascendente (subida) o descendente (bajada),
- Magnitud: Diferencia de potencia dado por la rampa (ΔP) en función de la potencia instalada o a algún otro parámetro relevante,
- Tiempo: Duración temporal (Δt) del evento rampa.

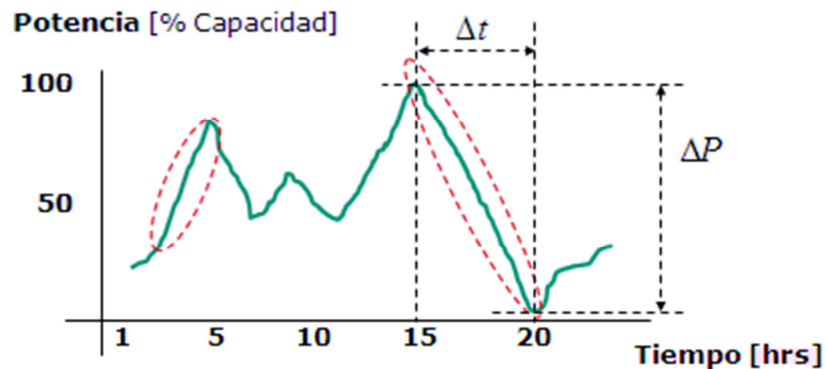


Figura 6. Parámetros relevantes en los eventos de rampas

2.5 Efectos de las rampas de viento en sistemas con generación eólica

Como se mencionó, una rampa de viento es un aumento o disminución repentina de la velocidad del viento, y por consiguiente de la potencia extraíble de éste. Por consiguiente el efecto de una rampa, ya sea de subida o bajada, significa un cambio brusco en la potencia generada en el sistema.

Para un funcionamiento seguro y dentro de los parámetros de calidad en un sistema eléctrico es importante que la generación de potencia esté equilibrada con el consumo. Un desequilibrio en estas dos variables tiene un efecto directo en la frecuencia del sistema. Al haber un exceso de generación las máquinas tienden a acelerarse, por lo que la frecuencia del sistema aumenta. Caso contrario, al tener menor generación que consumos, el déficit se tiende a compensar extrayendo potencia mecánica del giro de las máquinas, lo que gradualmente las desacelera y por consiguiente se reduce la frecuencia del sistema.

Dado que una rampa de viento tiene asociado un cambio brusco e inesperado en la potencia generada, la consecuencia inevitable es tener un desequilibrio generación-consumo, y por ello la consecuencia explicada. Sin embargo, dado este y otros problemas, existen en los sistemas eléctricos protocolos de regulación para regresar al sistema a un punto de operación seguro, económico y dentro de los estándares de calidad. Existen tres tipos de regulación [14]:

- Regulación primaria: actúa de forma rápida, segundos después de aparecido la perturbación. Su función es mantener un equilibrio en el sistema, compensando los desbalances de potencia que generan los

cambios en la frecuencia. Sin embargo no es capaz de llevar al sistema a una operación con frecuencia nominal.

- Regulación secundaria: Actúa pocos minutos después del desequilibrio de potencia. Es un control centralizado, realizado mediante un control manual o mediante un AGC ("Automatic Generation Control", control de generación automático). Su función es llevar al sistema a un punto de operación dentro del rango admisible para la frecuencia.
- Regulación terciaria: actúa entre 10 a 15 minutos pasado el desequilibrio. Es un control manual orientado a devolver al sistema a un punto de operación técnica y económicamente óptimo.

Si bien es cierto que existen estos protocolos de regulación frente a contingencias, lo ideal es operar el sistema de forma segura, en un estado óptimo del punto de vista económico y técnico. Es por ello que en la medida que se puedan prever la ocurrencia de contingencias para tomar acciones preventivas en vez de paliativas se aseguraría una operación más estable del sistema.

Otro punto a tenerse en cuenta frente a las rampas es que, si bien tienen repercusión en la operación técnicamente óptima del sistema, también representan una fuente de energía que de poder ser utilizada representaría una fuente económica para la generación eléctrica. Considérese el siguiente caso, supóngase que en cierto momento el sistema eléctrico está funcionando en su óptimo técnico y económico, con la demanda bien cubierta por la generación. Si en dicho momento comenzara una ráfaga de viento, con el aumento de la potencia asociado, esta potencia no puede ser ingresada inmediatamente al sistema ya que el consumo no aumentará tan rápidamente para asimilar este exceso de energía instantánea generada (de hecho, el consumo no depende de la generación, tiene un comportamiento propio). Por su parte las centrales hídricas y sobre todo térmicas no son capaces de disminuir lo suficientemente rápido su generación para compensar el aumento dado por la potencia repentina de la ráfaga de viento. Es por ello que en el intervalo en que el resto de las centrales disminuye su generación para poder ingresar la potencia extra del viento, la energía asociada se pierde. Poder anticiparse a las rampas de viento permitiría una coordinación de la generación para poder aprovechar al máximo este recurso, ahorrando así otros recursos (agua en centrales hídricas, carbón, gas, etc.)

Es por estas razones que resulta importante contar con una herramienta de predicción de las rampas de la potencia del viento, para así lograr una operación del sistema eléctrico segura y técnica y económicamente óptima.

2.6 Fenómeno de rampas en plantas de generación en Chile

Las rampas de potencia en el recurso eólico son un fenómeno que se da a nivel mundial, ya que en general el viento es un recurso que resulta muy cambiante y muchos de estos cambios pueden ocurrir de manera brusca. Sin embargo no en todos los lugares se da de la misma manera, ya que depende tanto de las condiciones climáticas locales como de la topografía de cada zona.

El fenómeno de rampas de potencia del viento se da también en plantas presentes en el territorio nacional. Estos se deben a los factores descritos en la Sección 2.3, además de la demanda instantánea que debe satisfacer el sistema y que determina la potencia que debe ser despachada por cada planta en cada momento.

Como ejemplo real del fenómeno de ocurrencia de rampas se muestran las figuras Figura 7, Figura 8, Figura 9 y Figura 10. Las figuras mencionadas corresponden a la potencia generada por las centrales Canela I y II para los días 4, 5, 6 y 7 de junio del 2014. Se aprecia que a lo largo del día se presentan rampas en la potencia generada por las centrales, rampas tanto de subida como de bajada, las que tienen diferentes deltas de potencia y de tiempo, y que además ocurren en diferentes momentos del día [13].

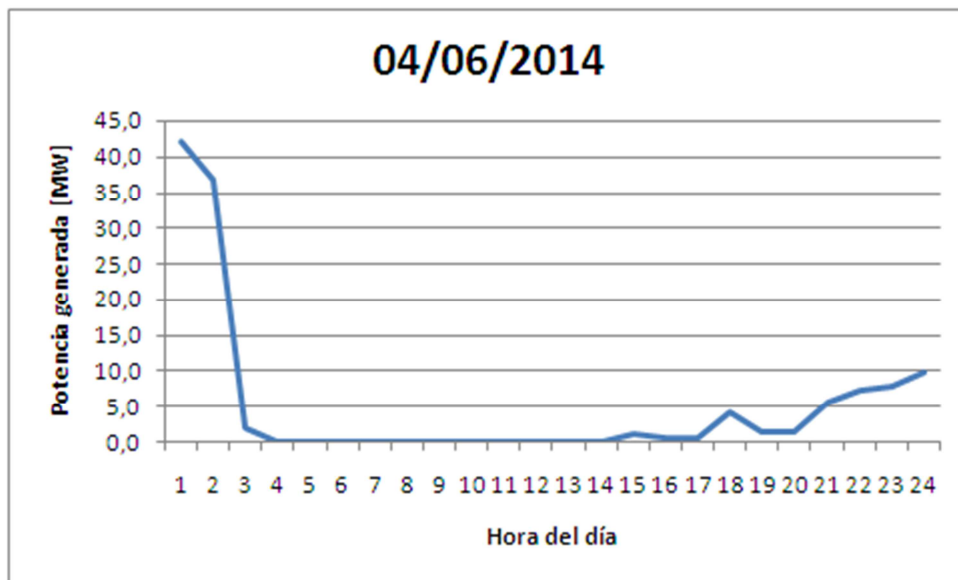


Figura 7. Potencia despachada por la planta Canela I y II el 4 de junio del 2014

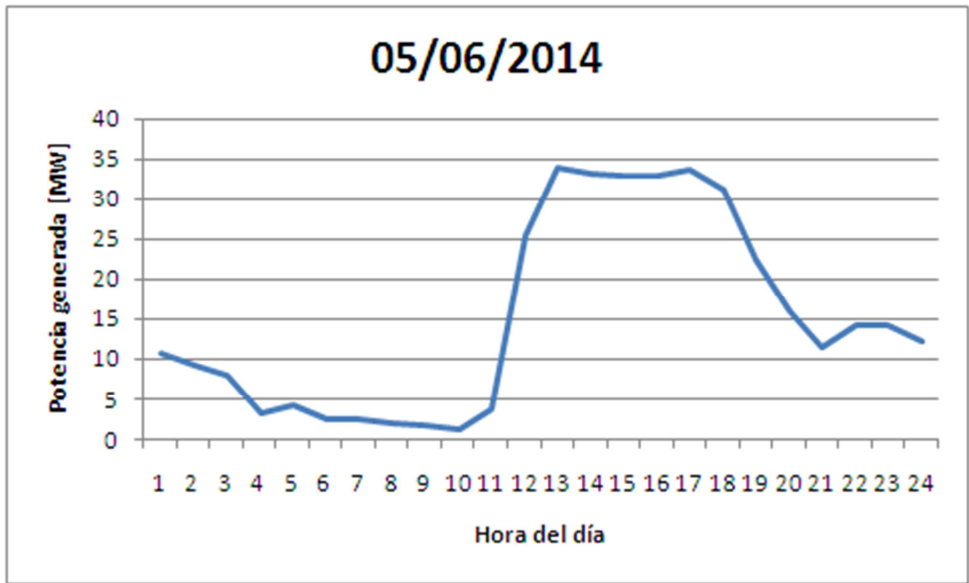


Figura 8. Potencia despachada por la planta Canela I y II el 5 de junio del 2014

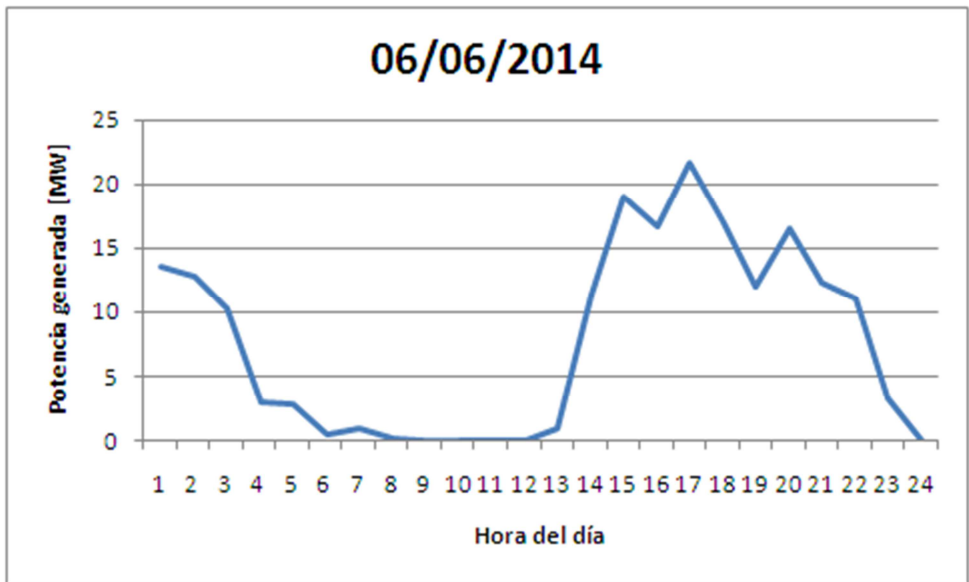


Figura 9. Potencia despachada por la planta Canela I y II el 6 de junio del 2014



Figura 10. Potencia despachada por la planta Canela I y II el 7 de junio del 2014

2.7 Métodos para el pronóstico de la potencia eólica

Existen dos enfoques en el desarrollo de métodos para el pronóstico del recurso eólico: el enfoque físico y el estadístico. El enfoque físico es el que describe la relación entre las condiciones atmosféricas, la topografía de la zona, la velocidad del viento y la potencia que es capaz de entregar la planta eólica. En contraparte, el enfoque estadístico es aquel que realiza estimaciones estadísticas de la relación entre toda la información de entrada que resulte relevante en el sistema con la potencia que puede generar la planta en su salida, sin considerar todas las condiciones de viento, atmosféricas o de topografía local. Los modelos actuales de pronóstico del recurso eólico usualmente son una combinación de ambos enfoques, como se muestra en la Figura 11:

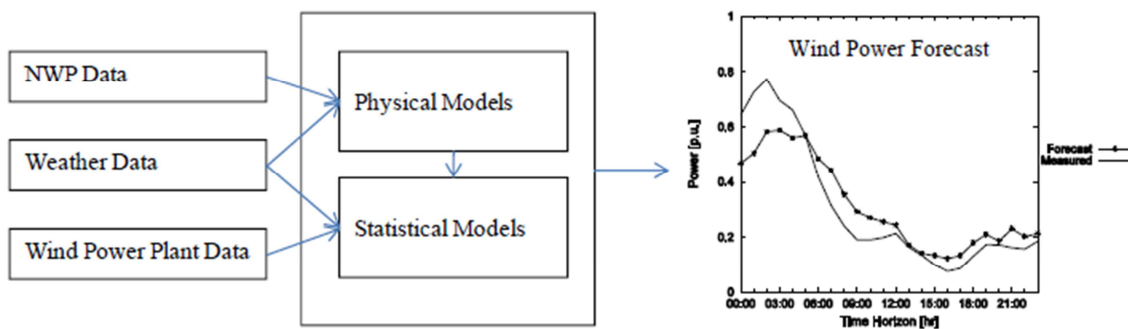


Figura 11. Ilustración de un sistema de pronóstico del recurso eólico

Como se mostró en la Figura 11, un sistema de pronóstico del recurso eólico necesita como entrada información de diferentes fuentes, incluyendo resultados de las predicciones numéricas del clima (NWP: Numerical Weather Prediction), que en términos simples corresponden a predicciones climáticas basadas en modelos matemáticos de la atmósfera y los océanos. Además también se necesitan mediciones meteorológicas locales, información que describa el

estado en tiempo real de las plantas eólicas, y finalmente información adicional acerca de las características específicas de las plantas eólicas y el terreno cercano a la ubicación de la misma.

El desempeño del sistema de pronóstico y la precisión del pronóstico mismo dependen en gran medida de poseer buenos pronósticos NWP (buenos en el sentido de que sean precisos), la complejidad del terreno y de contar con la información en tiempo real del clima y el viento presente en las cercanías de la planta eólica. Debido a esto, puede que se tengan errores no menores entre los errores resultantes del sistema de pronóstico pueden ser muy diferentes dependiendo de la zona en la que la planta se ubique [15].

Existen en el mundo países donde la energía eólica es un recurso importante para la generación eléctrica de sus respectivos sistemas eléctricos. Por ejemplo se puede citar a Dinamarca, que al año 2008 tenía instalado un total de 12 parques eólicos que sumaban una potencia instalada de 3.189 [MW] [16]. También se puede tomar el caso de Alemania, con sus 31.300 [MW] de potencia eólica instalada al año 2012 [17]. En estos países existe gran interés en el desarrollo de sistemas de predicción de la potencia del viento dada la alta penetración que este tipo de energía tiene en sus sistemas eléctricos. Es así como se han implementado algunos sistemas como por ejemplo el sistema PREVIENTO en Alemania, sistema desarrollado según el enfoque físico del problema. En Dinamarca por su parte también se han desarrollado sistemas de predicción, como es el caso del sistema conocido como PREDIKTOR, el que también se desarrolló según el enfoque físico [18].

2.8 Pronóstico de rampas

Como se mencionó, existen sistemas que están siendo empleados en el mundo para el pronóstico de la potencia eólica. Sin embargo la mayoría de estos sistemas están enfocados en realizar pronósticos para horizontes de tiempo más grandes, cuyo fin es conocer la potencia misma que se podrá obtener en el futuro y así poder realizar la programación del despacho de potencia. Es decir, estos sistemas están enfocados en la programación de la generación. Sin embargo, sistemas enfocados en el pronóstico de rampas no han sido desarrollados, o están aún en etapas de pruebas y perfeccionamiento. A pesar de ello, existe un interés creciente en el pronóstico de estos fenómenos debido a las ventajas asociadas a una herramienta capaz de hacer esto. Es así como varios autores han trabajado en el desarrollo de sistemas capaces de realizar pronósticos de rampas de la potencia del viento, desde diferentes enfoques.

Por ejemplo, investigadores de Garrad Hassan (GH) han realizado investigaciones que estudian metodologías para predecir la ocurrencia de rampas, incluyéndose la incertidumbre temporal asociada a estos eventos [19]. En este trabajo los autores definen una rampa como el cambio en la potencia de salida mayor a un 50% de la potencia nominal en un periodo de cuatro horas o menos. En el sistema implementado por estos autores se combinan *múltiples entradas de NWP, procesamiento estadístico y algoritmos adaptativos*.

Otro enfoque mediante el cual se aborda el problema es el de *minería de datos*. Por ejemplo Zheng y Kusiak combinan selección de características con cinco

algoritmos de minería de datos para el estudio de un predictor de 10 a 60 minutos de adelanto [20]. Para el estudio los autores usan datos de series de tiempo de potencia obtenidas de sistemas SCADA que controlaban cien turbinas. Para el predictor fueron usados la velocidad promedio viento, su desviación estándar y sus valores máximo y mínimo, y también emplean la potencia eólica medida del parque y la tasa de rampas. Dentro de la metodología empleada se utiliza un boosting tree para seleccionar las características más interesantes, para luego entrenar con esta característica los cinco algoritmos de minería de datos empleados: Perceptron Multicapa, Support Vector Machines (SVM), Random Forest, Árboles de Clasificación y Regresión, y Regresión de Paso.

Otra metodología es la presentada por Bossavy et al. [21], la cual identifica rampas mapeando la serie de tiempo asociada a la potencia del viento a una señal resultante de computar el promedio de las diferencias temporales de potencia. En base a esto los autores proponen dos metodologías de pronóstico enfocados en predecir la potencia del viento y el tiempo de ocurrencia de las rampas. El primer método propuesto usa información del espacio propuesto (de la serie de potencia original al transformado mediante los promedios de las diferencias temporales de la potencia), la intensidad de las rampas y la información respecto al tiempo de ocurrencia de las rampas. El otro método está basado en traducir la señal de cada uno de los miembros de un arreglo de curvas de potencia y en base a esto definir un intervalo de confianza para la localización temporal de las rampas.

También se ha propuesto el empleo de series de potencia para el pronóstico de rampas. Por ejemplo, se ha propuesto el mismo modelo ARIMA (“auto regressive integrated mobile average”, en inglés) para realizar pronósticos [22]. Autores como H. Mi et al. [23] han usado esta técnica para el pronóstico de rampas empleando métodos de suavizado exponencial. Otros, como Sevlian y Rajagopal [24] consideran los efectos estacionales asociados a la variable. Sin embargo, la potencia del viento tiene incertidumbre a diferentes escalas de tiempo, por lo que para cada escala considerada debe emplearse un enfoque diferente de series de tiempo.

2.9 Marco matemático

Previo a la explicación del método de predicción de rampas de la potencia eólica desarrollado en esta memoria se explicarán las herramientas matemáticas empleadas.

2.9.1 Variables aleatorias y probabilidades condicionales

El mundo está lleno de procesos que resultan de interés y que son estudiados para poder entender y muchas veces predecir su comportamiento. Es así como se realizan modelos de dichos fenómenos con los que los interesados en el área pueden entender la ocurrencia de ciertos eventos, predecir eventos al futuro, etc. Lamentablemente muchos de los fenómenos que suceden en el universo están sujetos a demasiadas variables que los afectan, las cuales difícilmente pueden ser identificadas y caracterizadas en su totalidad. Es así como los resultados predichos por los modelos pueden diferir, en mayor o menor medida, de los resultados reales que se observan. En casos en que

estas variables afecten mucho al fenómeno estudiado y sea muy difícil caracterizarlas individualmente es que se opta por trabajar el problema desde un punto de vista probabilístico, en los que la variable de interés deja de ser determinística y pasa a ser lo que se conoce como una variable aleatoria, un tipo de variable que asocia a cada evento una probabilidad de que este ocurra, por lo que lo que se obtiene a fin de cuentas de los modelos es la probabilidad de ocurrencia de ciertos fenómenos. Así, finalmente se busca conocer la probabilidad de ocurrencia de cierto evento, habiendo cierto grado de incertidumbre en lo que realmente ocurrirá.

2.9.1.1 Probabilidades condicionales

Cualquier fenómeno o evento representado por una variable aleatoria está, a pesar de la incertidumbre intrínseca que tiene asociada, relacionado con otros fenómenos que pueden afectarlo y facilitar su ocurrencia o, caso contrario, dificultarla. Entonces, al estudiar un fenómeno particular, si se tiene la evidencia de que ha ocurrido cierto evento que se relaciona con el fenómeno estudiado, las probabilidades asociadas van a cambiar. Por ejemplo, es difícil saber si una persona es fumadora simplemente viendo una foto de la persona, pero si fuera posible analizarla un tiempo y se observa que la persona tose mucho da una evidencia de que la hipótesis: “el individuo es fumador” tiene más chance de ser cierta. En este caso el evento: “el sujeto tose mucho” condiciona el evento que se está estudiando.

Más formalmente, dados dos eventos A y B, la probabilidad condicionada $P(A|B)$ corresponde a la probabilidad de que el evento A ocurra sabiendo que ha ocurrido el evento B. Esto se lee como la probabilidad de A dado B [25].

No tiene por qué haber una relación causal entre ambos eventos, ni temporal tampoco. A puede causar B o viceversa, o inhibirlo; además A puede preceder a B temporalmente hablando o ser al revés.

Definición (Probabilidad condicional):

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y dos eventos A y $B \in \mathcal{F}$, con $P(B) > 0$, se define la probabilidad condicional $P(A|B)$ como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.1)$$

Dicha probabilidad puede ser interpretada como, en el espacio en el que todos los eventos B se cumplen, la fracción de eventos A que están dentro de éste espacio. Así, tomando por ejemplo que el evento A sea tener fiebre y el evento B sea padecer de amigdalitis, en este caso la probabilidad $P(A|B)$ corresponde a la probabilidad de presentar fiebre cuando se está enfermo de amigdalitis. Gráficamente, observando la Figura 12, si se interpreta el espacio completo como el espacio de todos los posibles eventos, el espacio A representa todos los eventos en los que se presenta fiebre y el espacio B todos los eventos en los que se padece de amigdalitis. Así, la zona roja corresponde a todos los casos en los que se presenta fiebre y además se padece de amigdalitis, $A \cap B$. Como se sabe que el evento B “el individuo está enfermo de amigdalitis” el universo completo de casos se reduce a estos, ya que se posee la evidencia de aquello. Por otra parte, dado que se produce el evento B, necesariamente la

probabilidad de que el evento A “el individuo tiene fiebre” se reduce a los casos en que además de padecer fiebre se tiene que el individuo padece amigdalitis. Así, la nueva probabilidad corresponde a la intersección $A \cap B$ (zona roja de la Figura) dentro de todo el universo donde se cumplen los eventos B. Por ello la probabilidad $P(A|B)$ queda dada por la Ecuación 3.1

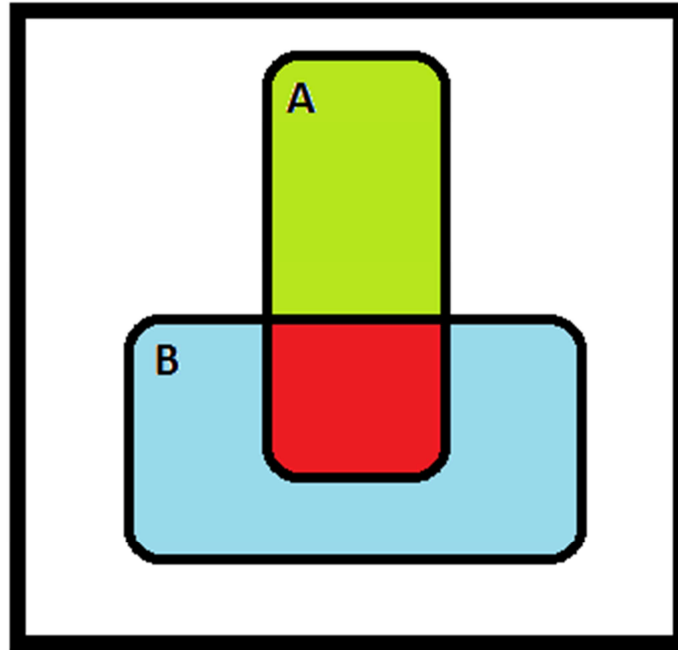


Figura 12. Ilustración de dos eventos en que uno condiciona a otro

Propiedades:

1. $P(A|B) + P(\bar{A}|B) = 1.$
2. $B \subseteq A \rightarrow P(A|B) = 1.$

2.9.1.2 Independencia de sucesos

Puede ser que la ocurrencia de un evento no influya en la ocurrencia de otro y viceversa. En estos casos se dice que las variables aleatorias son independientes.

Definición (Independencia de variables aleatorias):

Dos variables aleatorias A y B son independientes si y sólo si:

$$P(A, B) = P(A)P(B). \tag{2.2}$$

Lo anterior corresponde a la probabilidad de que ocurra A y B juntos. Es decir, dos variables son independientes si su probabilidad conjunta puede ser expresada como el producto de sus probabilidades individuales. Equivalentemente:

$$P(A|B) = P(A),$$

$$P(B|A) = P(B).$$

2.9.1.3 Teorema de Bayes

Las probabilidades condicionales tienen muchos usos en el cálculo de probabilidades cuando se posee cierta evidencia de la ocurrencia de fenómenos que pueden afectar el evento que se estudia. En casos en los que se dispone de evidencia que se actualiza constantemente es que conviene usar el enfoque Bayesiano, basado en el teorema de Bayes. Este teorema, que fue desarrollado por el reverendo Thomas Bayes en el siglo XVIII, es una extensión de lo explicado anteriormente acerca de probabilidad condicional.

Comúnmente se inicia el análisis de probabilidades con una asignación inicial de probabilidad, la probabilidad a priori. Cuando se tiene nueva evidencia que pueda afectar al evento en cuestión se procede a calcular la probabilidad revisada o a posteriori, la cual corresponde a la probabilidad considerando la nueva evidencia de la que se dispone. El teorema de Bayes permite calcular las probabilidades a posteriori, según muestra la Ecuación 2.3 [26]:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2.3)$$

Dónde:

$P(A)$:	Probabilidad a priori
$P(B A)$:	Verosimilitud
$P(B)$:	Probabilidad total del evento condicionante
$P(A B)$:	Probabilidad a posteriori condicionada a B

2.9.1.4 Clasificador Bayesiano ingenuo (Naive Bayes)

Un evento en particular puede ser afectado por varios fenómenos observables, en términos probabilísticos, estaría condicionado a varias otras variables. Considérese un evento A que se ve afectado por otros n eventos F_1, F_2, \dots, F_n . En este caso, la probabilidad a posteriori para A considerando la evidencia de los eventos F_i ($i=1, \dots, n$) se puede calcular en base al teorema de Bayes como muestra la Ecuación 2.4:

$$P(A|F_1, \dots, F_n) = \frac{P(F_1, \dots, F_n|A)P(A)}{P(F_1, \dots, F_n)}. \quad (2.4)$$

Como se aprecia, la probabilidad a posteriori para A en función de la evidencia depende de su probabilidad a priori $P(A)$, de la probabilidad conjunta de la evidencia $P(F_1, \dots, F_n)$ y de la probabilidad condicional $P(F_1, \dots, F_n|A)$.

El clasificador Bayesiano ingenuo considera que cada uno de los eventos que condiciona al evento estudiado lo hace de manera independiente uno de otro, es decir, son variables independientes. Esta hipótesis simplificatoria es por la que el clasificador recibe el apelativo de “ingenuo” (“naive”, en inglés) [27].

Si se aplica la hipótesis simplificatoria del clasificador Bayesiano ingenuo la Ecuación 2.4 se puede desarrollar un poco más, recordando que la probabilidad conjunta de sucesos independientes es la multiplicación de sus probabilidades individuales:

$$P(F_1, \dots, F_n) = P(F_1)P(F_2) \dots P(F_n) = \prod_{i=1}^n P(F_i). \quad (2.5)$$

Y

$$P(F_1, \dots, F_n|A) = P(F_1|A) \dots P(F_1|A)P(F_2|A) \dots P(F_n|A) = \prod_{i=1}^n P(F_i|A). \quad (2.6)$$

Entonces, bajo esta nueva hipótesis, la Ecuación 2.4 se transforma en la Ecuación 2.7:

$$P(A|F_1, \dots, F_n) = \frac{P(A) \prod_{i=1}^n P(F_i|A)}{\prod_{i=1}^n P(F_i)}. \quad (2.7)$$

2.9.2 Redes Bayesianas

Antes de proceder a definir formalmente lo que es una red Bayesiana se explicará en términos sencillos lo que esta herramienta matemática representa y para lo que puede ser empleada.

La base de estas redes Bayesianas son ciertos eventos de los cuales no se posee un conocimiento determinístico, por lo que deben ser tratados como variables aleatorias. Estos eventos están representados por nodos dentro de la red que se procede a construir. Sin embargo, a pesar de la aleatoriedad de estos eventos puede existir cierta dependencia entre unos y otros, por lo que la ocurrencia o inexistencia de uno puede cambiar la probabilidad de ocurrencia de alguno otro. Así es como se va construyendo una red de nodos y aristas, donde cada nodo representa algún fenómeno o evento en particular y las aristas señalan la dependencia o influencia de algunos eventos con y hacia otros.

Para dejar más claro el punto anterior tómesese el siguiente ejemplo. Considérese que se quiere conocer la probabilidad de que llueva en cierto día cualquiera en el año. Por experiencia se sabe que en esto influyen muchos factores, como la estación del año (o el mes), la presencia o ausencia de nubes o viento, entre otras cosas. A priori, sin tener información alguna, la probabilidad de que llueva se podría estimar en una primera instancia, por ejemplo, de forma estadística, haciendo uso de los datos de los que se tiene registro de uno o varios años simplemente contando el número de veces que llueva en el total de días de los que se tienen registro. Así, si estadísticamente se tiene que llovió veinticinco días en un año se podría decir que la probabilidad de que llueva es de veinticinco en trescientos sesenta y cinco. Con esto se tiene un nodo dentro de la red.



Figura 13. Representación de evento como nodo de una red

Sin embargo, sólo con esta información, el evento: "hoy va a llover" posee una gran incertidumbre, ya que el no poseer información más exacta acerca de los fenómenos u otros eventos que puedan afectar al evento de interés solo se conoce las llamadas probabilidades "a priori", correspondientes a la probabilidad de ocurrencia del evento por sí solo. Ahora, considérese que se tiene información extra que tiene influencia en el evento de interés. La ocurrencia de algún otro evento vinculado al que se está estudiando cambia la probabilidad original por el peso que éste significa sobre el evento estudiado. Siguiendo con el ejemplo, sea por ejemplo que se considera la época del año o el estado de nubosidad del día. Claramente este conocimiento altera la probabilidad de que el evento analizado ocurra o no, es mucho más probable que llueva en un día nublado de invierno que en un día de pleno verano. Siguiendo esta línea, la red debe ser ampliada para incluir estos dos nuevos eventos.

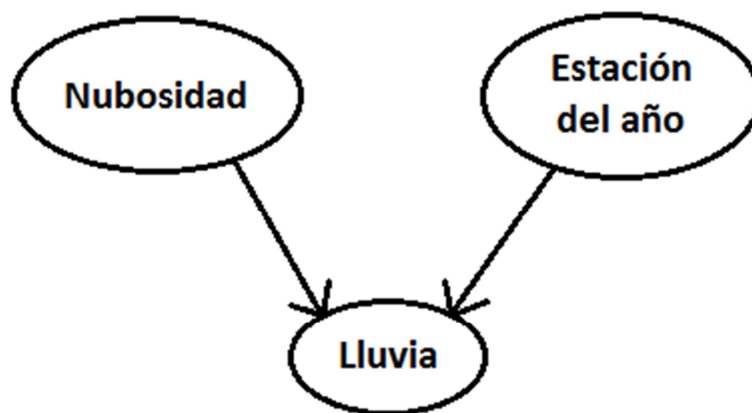


Figura 14. Red Bayesiana incluyendo fenómenos que afectan el evento estudiado

Como se aprecia en la Figura 14 se ha incluido en la red la nubosidad del día y la estación del año ya que tiene directa influencia sobre la posibilidad de que llueva. Así, el conocimiento de alguna de estas variables influye en la probabilidad del evento de interés. La probabilidad obtenida después de considerar la evidencia de la que se dispone se llama probabilidad "a posteriori", la cual es mucho más confiable que la probabilidad a priori ya que se consideró la evidencia de la que se dispone y que puede afectar el evento de interés. Este es el motivo por el cual estas redes sean llamadas Bayesianas, ya que el enfoque Bayesiano considera que las probabilidades de ocurrencia de un evento se ven afectadas por otros eventos que tienen influencia en él y que lo condicionan y los consideran como evidencia. Esto se condice con la idea intuitiva de que el tener cierta evidencia nos da mayor seguridad acerca de la ocurrencia de ciertos fenómenos.

2.9.2.1 Definición formal de red Bayesiana

Antes de definir formalmente una red Bayesiana se definirán algunos conceptos de teoría de grafos y de la teoría de probabilidad [28].

2.9.2.2 Definiciones previas

- **Arco:** Es un par ordenado (X,Y) . Esto representa una arista que une los nodos X e Y . Este arco puede ser dirigido, apuntando de un nodo a otro, en cuyo caso queda representado por una flecha.
- **Grafo dirigido:** Es un par $G=(N,A)$ donde N es un conjunto de nodos y A un conjunto de arcos no orientados (esto es, pares no ordenados (X,Y)) definidos sobre los nodos.
- **Camino:** Es una secuencia ordenada de nodos (X_1, \dots, X_r) tal que $\forall j=1, \dots, r-1$, ó bien el arco $X_j \rightarrow X_{j+1} \in A$ ó bien el arco $X_{j+1} \rightarrow X_j \in A$.
- **Camino Dirigido:** Es una secuencia ordenada de nodos (X_1, \dots, X_r) tal que $\forall j=1, \dots, r-1$ el arco $X_j \rightarrow X_{j+1} \in A$.
- **Ciclo:** es un camino no dirigido que empieza en un nodo X y termina en el mismo nodo X .
- **Grafo acíclico:** es un grafo que no contiene ciclos.
- **Padre:** X es un padre de Y si y solo si existe un arco $X \rightarrow Y$. Se dice también que Y es **hijo** de X . Al conjunto de los padres de X se les representa como $pa(X)$, y al de los hijos de X como $s(X)$.
- **Antepasado o ascendiente:** X es un antepasado ascendiente de Z si y sólo si existe un camino dirigido de X a Z .
- **Conjunto ancestral:** el conjunto ancestral de un nodo X es un conjunto que contiene a X y a todos sus antepasados.
- **Descendiente:** Z es un descendiente de X si y sólo si X es un antepasado de Z . El conjunto de los descendientes de X será denotado en adelante por $de(X)$.
- **Variable proposicional:** es una variable aleatoria que toma un conjunto exhaustivo y excluyente de valores. Usualmente se denota con letras mayúsculas, por ejemplo X , y los un valores que la variable puede tomar se denotan con la letra minúscula, x .
- **Variables independientes:** Dos variables X e Y son independientes si se tiene que la probabilidad condicional $P(X/Y)=P(X)$. De esta definición se tiene una caracterización de la independencia que puede ser utilizada como definición alternativa, se dice que X e Y son variables independientes si y sólo si $P(X,Y)=P(x)P(Y)$.
- **Variables independientes dado una tercera variable:** Dos variables X e Y son independientes dada una tercera variable Z si se tiene que $P(X/Y,Z)=P(X/Y)$.

2.9.2.3 Definición de red Bayesiana

Una red Bayesiana es:

- Un conjunto de variables proposicionales V .
- Un conjunto de relaciones binarias definidas sobre las variables de V , E .
- Una distribución de probabilidad conjunta sobre las variables de V , P .

tales que:

- (V,E) forman un gráfico acíclico, conexo y dirigido G .
- (G,P) cumplen las *hipótesis de independencia condicional*, también llamadas *de separación direccional*, las que son enunciadas a continuación:

2.9.2.4 Hipótesis de independencia condicional

Un grafo acíclico conexo y dirigido $G=(V,E)$ y una distribución de probabilidad conjunta P definida sobre las variables del grafo se dice que cumplen las hipótesis de independencia condicional si para toda variable X de V se tiene que el conjunto de padres directos de X , denotados como $pa(X)$ separa condicionalmente a X de otro nodo Y de la red que no sea X , ni sus descendientes ni sus padres.

$$\forall X \in V, \forall Y \in V - \{X \cup de(X) \cup pa(x)\}, P(X|pa(X), Y) = P(X|pa(X)),$$

donde, como se mencionó anteriormente, $pa(X)$ denota a los padres de X y $de(X)$ denota a los descendientes de X .

2.9.2.5 Factorización de la probabilidad

Dada una red Bayesiana, su distribución de probabilidad puede expresarse como:

$$P(x_1, \dots, x_n) = \prod_i P(x_i|pa(x_i)). \quad (2.8)$$

Demostración:

Supóngase una ordenación de las variables en la que los padres de cada nodo aparezcan después de él, esto es, en la ordenación de las variables (X_1, \dots, X_n) se tiene que X_{j+1}, \dots, X_n corresponden a los padres de X_j (esto corresponde sólo a una ordenación de las variables, por lo que al hacer esto no se pierde generalidad). Dado esto, se tiene:

$$P(x_1, \dots, x_n) = \prod_i P(x_i|x_{i+1}, \dots, x_n).$$

Pero dado el orden que se dio previamente a las variables, el conjunto (X_{i+1}, \dots, X_n) corresponde al de todos los padres de X_i , y por lo tanto, la separación direccional dice que:

$$P(x_i|x_{i+1}, \dots, x_n) = P(x_i|pa(x_i)) \quad \blacksquare$$

La importancia de este teorema es que permite describir una red Bayesiana a partir de la probabilidad conjunta de cada nodo (o la probabilidad a priori en caso de los nodos sin padres) en lugar de dar la probabilidad conjunta, la cual requiere un número de parámetros que crece exponencialmente en el número de nodos y que además plantea el problema de verificar la separación direccional.

2.9.2.6 Modelación de problemas con redes Bayesianas

Habiendo definido previamente el concepto de redes Bayesianas se procederá a explicar el proceso de modelado de problemas de la vida real utilizando dicho enfoque.

2.9.2.7 Identificación de variables

Como primera cosa es importante realizar un estudio acerca del problema que se quiere modelar para tener el grado máximo de conocimiento y comprensión de éste. Esto se traduce en que en muchos casos reales será necesaria la experiencia de expertos en el área.

Una vez que se conoce suficientemente bien el problema, el paso siguiente es identificar las variables que son relevantes, para así centrarse sólo en las que son de real interés para afrontar el problema.

Dadas las diferentes variables, estas pueden ser agrupadas en ciertas clases. Teniendo estas variables en mente al momento del modelado del problema se puede simplificar el proceso. Estas clases son:

Variables Objetivo

Son las variables empleadas para modelar los objetos o eventos de interés, es decir, aquellos sobre los cuales se busca razonar. Éstas son variables que no son observables directamente.

Variables de Observación

Corresponden a variables empleadas para modelar las formas indirectas mediante las cuales se puede conocer las variables objetivo. También llamadas variables de evidencia.

Factores

Éstas son las variables empleadas para modelar los fenómenos que afectan a las variables objetivo. También denominadas variables de contexto.

2.9.2.8 Estados y valores

Las variables también pueden ser clasificadas según su escala de medición:

Variables cualitativas:

Son las variables que expresan distintas cualidades, características o modalidades. Este tipo de variables pueden ser dicotómicas, representando dos características como podría ser el sexo de una persona, o politómicas cuando representan más de dos posibles características.

Variables cuantitativas:

Son variables que se expresan mediante cantidades numéricas. Estas variables además pueden ser discretas o continuas.

Frecuentemente y para simplificar el estudio de un problema es conveniente representar un fenómeno continuo mediante valores discretos. Para esto se puede definir un conjunto de intervalos de valores, agrupando todas las variables que están dentro de ese intervalo como un solo valor discreto.

La definición de variable proposicional tiene suma importancia a la hora de modelar un problema con una red Bayesiana ya que los nodos de la red son exactamente variables proposicionales y por lo tanto deben tomar un conjunto exhaustivo y excluyente de valores.

2.9.2.9 Estructura

Habiendo definido las variables el paso siguiente para la construcción de un modelo es definir su estructura. Para ello se debe conectar variables mediante arcos. Como se mencionó anteriormente, en las redes Bayesianas los arcos son dirigidos, por lo que la dirección en la que apuntan tiene directa relación en cómo se influyen las variables entre sí. Cambiar la dirección de un arco cambia su significado. Si entre dos variables no existe un arco que las una significa que éstas no tienen una dependencia directa, si no que a lo sumo se relacionan mediante otras variables.

La presencia de un arco indica una relación causal entre dos variables. Por ejemplo, la presencia o ausencia de una enfermedad tiene influencia en que en los resultados de las pruebas o test médicos den positivos o negativos (ver Figura 15). Una alternativa a la dirección causal es la dirección de diagnóstico (Figura 15). Esta dirección surge de inferir la presencia o ausencia del elemento que puede causar un fenómeno que puede ser observado, como por ejemplo el observar que el paciente presenta fiebre puede llevar a la conclusión de que padece de gripe. En este último caso la variable de observación precede a la variable objetivo. Utilizar el tipo de relaciones recién mencionadas no es en sí un error, pero añade dificultad y puede desencadenar en que sean usadas de manera incorrecta.

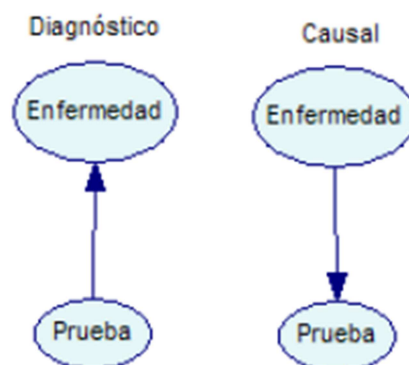


Figura 15. Posibles direcciones de los arcos entre dos nodos

El marco de trabajo de las redes Bayesianas no impone que los arcos sean necesariamente construidos siempre en la dirección causal. Según Pearl, "los patrones de independencia plasmados en un grafo acíclico dirigido son típicos de organizaciones causales" (Pearl, 1998). También Druzel y Simon explican que "se pueden usar muchos modelos equivalentes para representar al mismo sistema, pero se prefieren los modelos que utilicen relaciones causales ya que minimizan el número de arcos en el grafo, aumentando la claridad de los modelos y ofreciendo ventajas computacionales" (Druzel y Simon, 1993).

2.9.2.10 Parámetros

El paso final en el modelado es especificar los parámetros. Esto corresponde a proporcionar las probabilidades a priori de todos los nodos raíz y las probabilidades condicionales de los nodos que tengan padres.

Existen varias alternativas para la obtención de los parámetros de la red, entre las cuales se tienen las siguientes:

- Especificación directa de los parámetros, normalmente contando con la ayuda de expertos. Este proceso es usualmente costoso.
- Aprendizaje a partir de bases de datos. Para esto es imprescindible la existencia de dichas bases de datos. Con estas es posible determinar los parámetros de forma estadística.
- Combinar especificación y aprendizaje, en la que se pueda contar con expertos que ayuden a especificar la estructura, y por otro lado aprender los parámetros en base a los datos y finalmente disponer de expertos nuevamente para supervisar el modelo obtenido.

3 Metodología

En este capítulo se detalla la metodología e implementación de un método de predicción de rampas en la potencia eólica en base a la herramienta matemática conocida como redes Bayesianas. El horizonte de anticipación de la predicción frente al dato real que se empleará es de una, dos y tres horas, presentándose los resultados para los tres casos. También se prueba realizando una predicción en una ventana de tiempo. Cabe destacar que la herramienta que se desarrolla en este trabajo pretende ser una herramienta que recibe datos en línea, por lo que hora a hora se actualizan las probabilidades en las que se basa el pronóstico.

Los datos con los que se trabaja son los correspondientes a la potencia generada por la central Canela I en los meses de noviembre y diciembre del año 2012 y de enero del año 2013. La razón de emplear estos tres meses es que el viento tiene una importante componente estacionaria dentro del año, es decir, los vientos promedio en verano son diferentes a los de invierno. Es por esto que se emplean esos tres meses que corresponde más o menos al periodo de verano. Estos datos en un intervalo de muestreo de 5 minutos, desde el día 8 de noviembre al 31 de enero, por lo que sumaban un total de 25.632 datos. Estos datos fueron procesados de modo de obtener los promedios horarios y trabajar en base a estos.

A continuación se procede a explicar la creación del modelo, desde el procesamiento de los datos, pasando por la verificación de las correlaciones de la potencia de salida de la planta con diferentes variables para la construcción de la red Bayesiana, la obtención de los parámetros necesarios para la construcción de la red y finalmente la obtención de las probabilidades de que sucedan rampas para ciertas horas específicas en el mes de enero, a fin de corroborar la efectividad y fiabilidad del método.

3.1 Procesamiento de datos

El primer paso para la creación del método de predicción es el procesamiento de los datos de los que se dispone. Como se mencionó, se trabajará con datos correspondientes a los meses noviembre y diciembre del 2012 y enero del 2013, los cuales están tomados con un intervalo de muestreo de cinco minutos.

Primero, se separa la data en dos grupos, con uno de los cuales se entrenará el modelo, obteniendo los parámetros necesarios, y con el otro se procede a la validación del mismo, desde el punto de vista de la sensibilidad y la especificidad del predictor. Dentro del grupo uno, correspondiente al grupo de entrenamiento, se utilizan los datos de noviembre y diciembre del 2012, mientras que para la evaluación se emplean los datos correspondientes a enero del 2013.

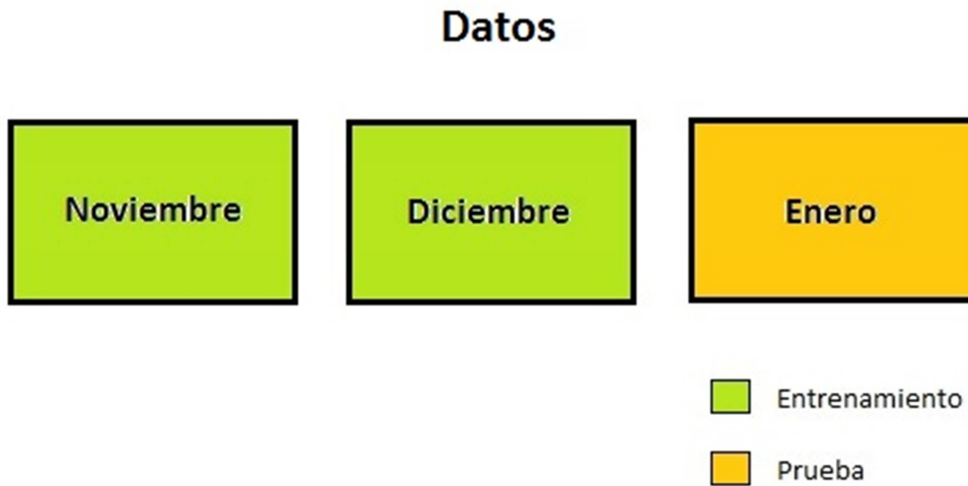


Figura 16. Agrupación de datos para entrenamiento y prueba del modelo

Como se menciona más adelante, se considera que un buen intervalo de tiempo para considerar que ha ocurrido una rampa es una hora. Es por ello que los datos, que corresponden a mediciones cada cinco minutos, se deben promediar en intervalos de una hora, para trabajar finalmente con estos promedios horarios. Así, los datos originales deben promediarse en grupos de 12 datos (ya que con un periodo de muestreo de 5 minutos se tendrán 12 datos para cada hora).

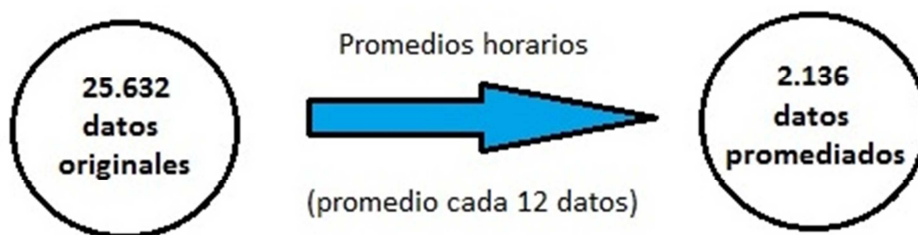


Figura 17. Reducción de número de datos mediante promedios calculados cada una hora

3.2 Rampas

Un paso importante es la definición de lo que se considerará una rampa. Existen diferentes definiciones de rampas, pero de estas se debe escoger una en base a la que se trabajará. En este caso se escogió la definición que dice que una rampa sucede cuando hay un cambio en la potencia que excede un umbral en un rango de tiempo determinado [22]:

$$|P_{t+1} - P_t| \geq P_{umbral}$$

Según algunos autores (por ejemplo Potter et al.) buen umbral para definir que un evento es en verdad una rampa es de al menos un 10% de la potencia total de la planta entre dos horas consecutivas [29]. Bajo esta definición el intervalo considerado es de una hora, ya para el umbral se tiene que el 10% de la

potencia instalada es 1,815 [MW], pero por simplicidad se considerara un umbral un poco mayor de 2 [MW].

El método creado se desarrolla bajo un punto de vista de la detección de ocurrencia de rampas, independiente de la magnitud o dirección de estas. Es así como se crea una variable binaria, la cual toma dos valores en función de si ocurre o no una rampa. Esta variable queda definida de la siguiente forma:

$$R(t) = \begin{cases} 0 & \text{si no hay un evento rampa en } t \\ 1 & \text{si hay un evento rampa en } t \end{cases}$$

Bajo esta definición se procede a encontrar las rampas ocurridas en los meses correspondientes a los datos usados para, por una parte, poder obtener los parámetros del modelo, y por otra poder contrastar los resultados del modelo con datos reales para corroborar que tan bueno es. Esto se muestra en la Figura 18:

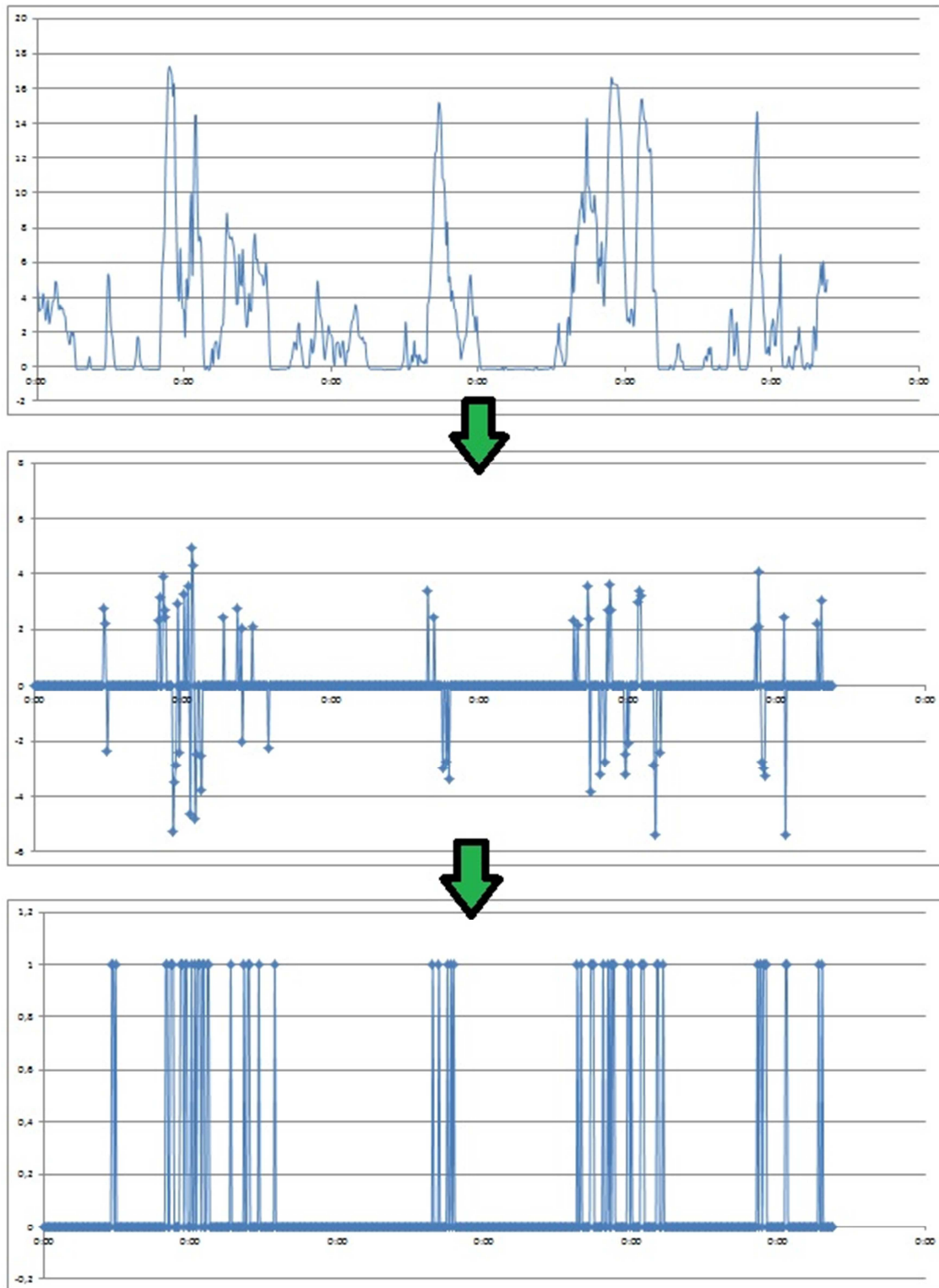


Figura 18. Obtención de ocurrencia de rampas desde los datos originales: el gráfico superior corresponde a los datos intermedios, el gráfico central corresponde al primer procesamiento donde se aprecian las rampas con su magnitud y sentido, y el gráfico inferior refleja el resultado final del procesamiento, reduciendo las rampas a una variable binaria de tipo “ocurre o no ocurre”.

Con esto se comprobó que en noviembre ocurrieron un total de 63 rampas, en diciembre ocurrieron 108 y en enero 92.

Tabla 4. Rampas de noviembre, diciembre y enero.

Mes	Rampas
Noviembre	63
Diciembre	108
Enero	92

3.3 Construcción de la red Bayesiana

El siguiente paso es la construcción de la red Bayesiana. Para esto es necesario definir cuáles son las variables que están correlacionadas con la ocurrencia de las rampas. En este caso, dada la información de la que se dispone, se pueden distinguir que podría formar parte de la red Bayesiana la hora del día (como una variable “tiempo”), la misma potencia generada o las rampas ocurridas hasta el momento.

3.3.1 Hora del día:

Primero se estudia la incidencia de la hora del día en las rampas de potencia. Como se muestra en la Figura 19 las rampas tienden a ocurrir más frecuentemente en ciertas horas específicas del día. Por esto último es que se considera que la variable tiempo (u hora del día más específicamente) debe formar parte de la red Bayesiana a construir.

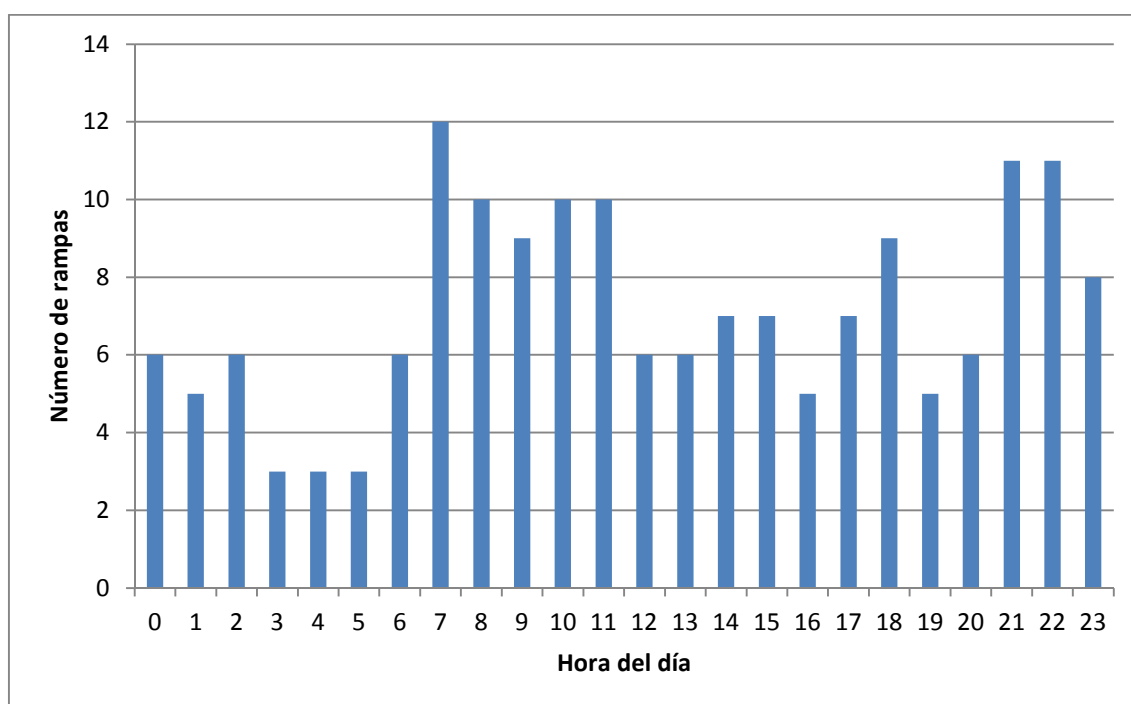


Figura 19. Distribución horaria de rampas ocurridas durante noviembre y diciembre del 2012

3.3.2 Potencia generada:

Otra variable que podría estar correlacionada con la ocurrencia de rampas es la misma potencia que se ha generado anteriormente. Para verificar la incidencia de esta variable se emplea la autocorrelación estadística. Entonces se calcula y grafica la autocorrelación de la potencia generada, la cual se muestra en la Figura 20:

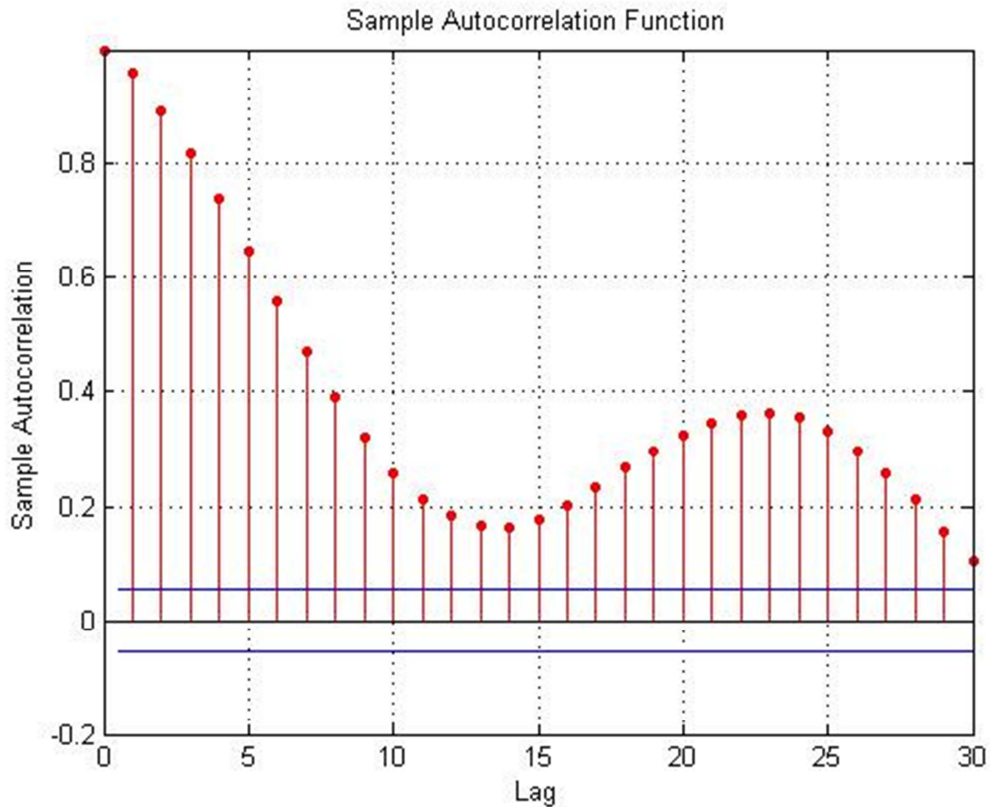


Figura 20. Autocorrelación de los datos de la potencia generada. El eje LAG corresponde a los intervalos temporales considerados (representados por saltos de una hora cada uno)

Como se aprecia, existe la correlación entre los valores de potencia es alto sobre todo entre valores cercanos en el tiempo (como se utilizan los promedios horarios, el eje Lag de la Figura corresponde a un intervalo de horas para los que se calcula la correlación entre los datos). Ahora, como el fenómeno de rampas está directamente relacionado con las potencias generadas en cada hora y como esta cambia, dada la correlación mostrada en la Figura 20 se concluye que también debe incluirse la variable “potencia generada” a la red Bayesiana.

Algo que debe considerarse es que la potencia generada puede tomar cualquier valor (dentro de los rangos admisibles de generación de la planta), es decir, es una variable continua. Ahora, como se está trabajando de un punto de vista estadístico es más conveniente trabajar con variables discretas, en las que se pueda contar cierta cantidad de ocurrencias de un evento en particular en el total de casos. Esto no puede hacerse con variables continuas, ya que es difícil que en distintos momentos en que se sense la potencia el número obtenido sea el mismo (probabilísticamente hablando la probabilidad de obtener un resultado específico para una variable aleatoria continua es teóricamente cero). Es por ello que se trabaja en base a histogramas de potencia, en los cuales se consideran rangos de potencia dentro de los cuales se clasifica la potencia generada en cierta hora dada. Para este caso se consideraron rangos de 1 [MW] de potencia, empezando del rango 0-1 [MW] hasta el rango 17-18 [MW], ya que la potencia nunca supera este valor.

3.3.3 Rampas anteriores

Se investiga también la posible correlación de ocurrencia de rampas con otros eventos de rampas que hayan ocurrido en el pasado. Igualmente que para el caso analizado de la potencia, se emplea la autocorrelación estadística de esta variable, la cual se muestra en la Figura 21:

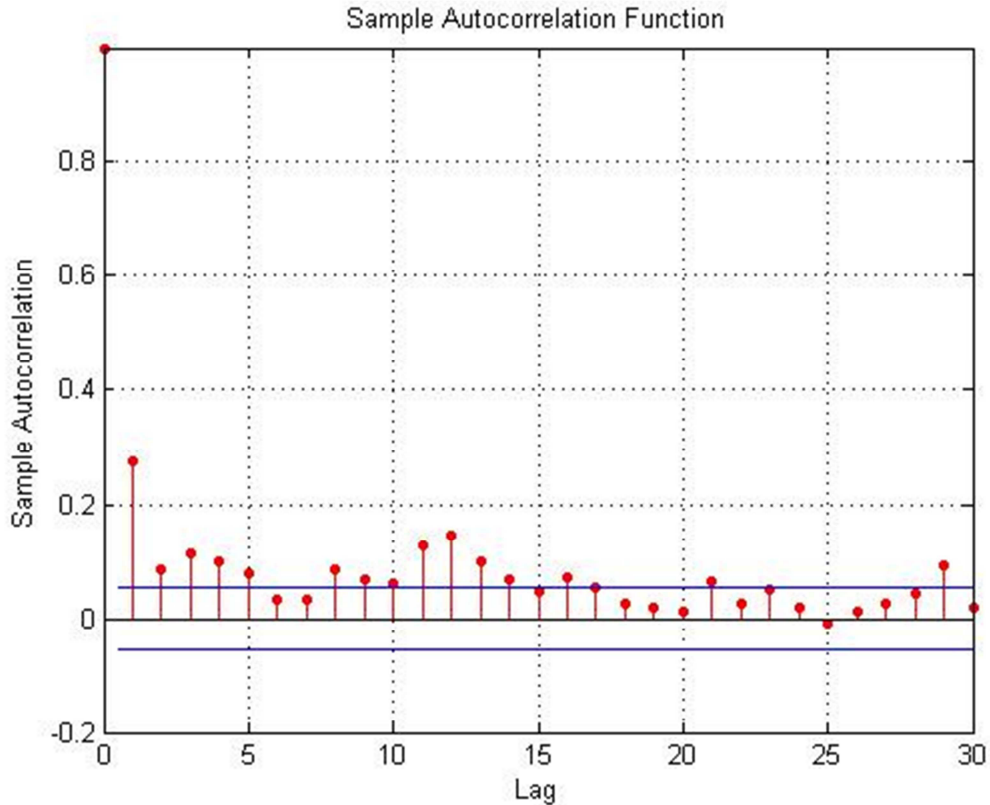


Figura 21. Autocorrelación de las rampas ocurridas. El eje LAG corresponde a los intervalos temporales considerados (representados por saltos de una hora cada uno)

Como se aprecia en la Figura, la autocorrelación es pequeña. Tiene el valor más alto con los eventos ocurridos en el instante de tiempo (hora) inmediatamente anterior, disminuyendo cuando el intervalo se hace más grande. Aun así el valor de la correlación es pequeño, por lo que se descarta el empleo de esta variable en la red Bayesiana.

3.3.4 Red Bayesiana

Considerando las variables que están más correlacionadas con la ocurrencia de rampas se crea la red Bayesiana sobre la que se trabajará. En este caso se consideran tres nodos, uno correspondiente al evento "rampa". Además de dicho nodo se agregan otros dos, los que serán los nodos padres, correspondientes a potencia generada y a hora del día. La red Bayesiana entonces queda como muestra la Figura 22:

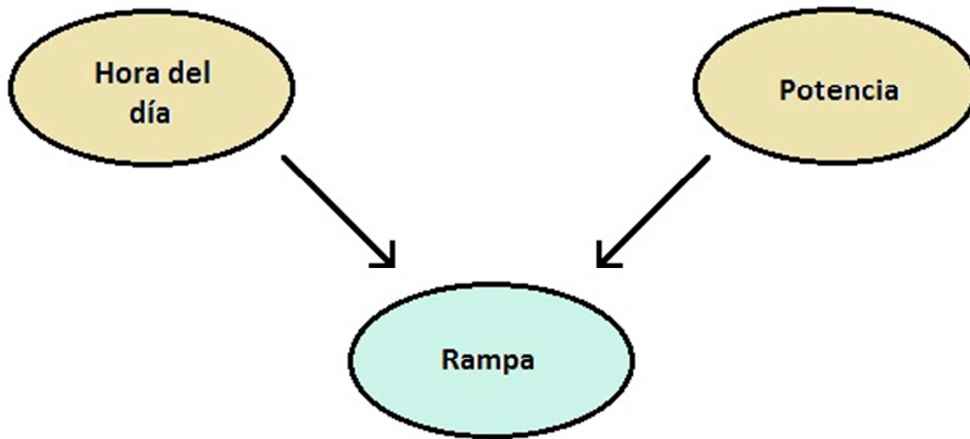


Figura 22. Red Bayesiana empleada en el predictor

Esta es la red Bayesiana que se emplea, por lo que la predicción se basará en el cálculo de la probabilidad de ocurrencia de rampas condicionado a la hora del día en que ésta puede ocurrir y a la potencia generada previamente.

3.3.5 Obtención de parámetros

Una vez definida la red Bayesiana que se emplea se prosigue a calcular los parámetros necesarios dentro del modelo.

Dada la red Bayesiana, los parámetros necesarios para hacer los cálculos y obtener resultados son la probabilidad total correspondiente a la hora del día (quizás parece un poco extraño ver esta variable como aleatoria, pero debe pensarse desde el punto de vista en que no se tiene acceso a saber qué hora del día es pero se puede inferir de alguna otra evidencia) y de la potencia que se genera en cierto momento del día, y la probabilidad de ocurrencia de rampas condicionada al valor que haya tomado la potencia en un momento dado del día y la hora del día en que se esté prediciendo la potencia.

3.3.5.1 Hora del día

Lo primero es obtener la probabilidad de que sea cierta hora del día, es decir, que la variable aleatoria H tome el valor $H=h$ con $h=0, \dots, 23$. Para esto se considera que el día tiene 24 horas, por lo que la probabilidad es simplemente uno entre 24 opciones:

$$P(H = h) = \frac{1}{24} \quad \text{con } h = 0, \dots, 23.$$

3.3.5.2 Potencia generada

Para calcular la probabilidad de que se genere una potencia dada se realizaron los siguientes pasos. Primero, como se explicó previamente, se separó las potencias generadas en rangos de 1 [MW] cada uno, desde el rango 0-1 [MW] hasta el rango 17-18 [MW], y se un cuenta cuantas veces la potencia generada estuvo en cada uno de estos rangos. Luego, el cálculo de la probabilidad de que se genere una potencia que esté en uno de estos rangos se realiza de forma estadística con el conteo asociado a cada rango dividido por la cantidad total de casos (es decir, la cantidad total de datos promediados de los que se dispone para el entrenamiento del modelo).

Las probabilidades encontradas para la potencia generada se muestran en la Tabla 5.

Tabla 5. Probabilidades de que la potencia generada esté en un rango específico

Rango de potencia	Número de veces que se da potencia en ese rango	Probabilidad P(D=d)
0 – 1	567	0,4073
1 – 2	125	0,0898
2 – 3	104	0,0747
3 – 4	99	0,0711
4 – 5	76	0,0546
5 – 6	77	0,0553
6 – 7	42	0,0302
7 – 8	58	0,0417
8 – 9	38	0,0273
9 – 10	33	0,0237
10 – 11	31	0,0223
11 – 12	33	0,0237
12 – 13	22	0,0158
13 – 14	24	0,0172
14 – 15	22	0,0158
15 – 16	15	0,0108
16 – 17	21	0,0151
17 – 18	5	0,0036
Total	1392	1,0000

3.3.5.3 Probabilidad condicionada de ocurrencia de rampas

Dada la red Bayesiana se necesita conocer la probabilidad de que ocurra una rampa dado que se está evaluando dicha probabilidad en una hora dada y dado el dato de potencia que se tuvo en el periodo de tiempo precedente que se consideró para la predicción (una, dos o tres horas). Para calcular estas probabilidades se puede recurrir a la fórmula dada por la Ecuación 3.1:

$$P(R|P, H) = \frac{P(R, P, H)}{P(R, H)}. \quad (3.1)$$

Ahora, como hipótesis simplificatoria se consideró la hipótesis asociada a Naive Bayes, por lo que se asume que las variables de los nodos padres del nodo asociado al evento rampa son independientes. Considerando esto y aplicando el teorema de Bayes sobre la probabilidad mostrada en la Ecuación 3.1 se llega a la ecuación que será empleada finalmente para el cálculo de las probabilidades, mostrada a continuación:

$$P(R|P, H) = \frac{P(R)P(P|R)P(H|R)}{P(P)P(H)}. \quad (3.2)$$

Entonces, con la formulación dada por la Ecuación 3.2 son necesarios los parámetros: probabilidad a priori para la ocurrencia de rampas $P(R)$, probabilidad de que la potencia esté en un rango dado condicionado a la ocurrencia de rampas $P(P|R)$ y la probabilidad de estar en una hora específica del día condicionado a la ocurrencia de rampas $P(H|R)$. Las probabilidades totales $P(H)$ y $P(P)$ fueron calculadas previamente.

El parámetro $P(R)$ es posible de calcular mediante un conteo de casos. Como la variable es binaria, hay dos posibles valores, $P(R)$ y $P(\neg R)$, y en este caso solo interesa el primero. Entonces la forma de calcularlo será estadísticamente haciendo un conteo de cuantos eventos rampa hubo dentro de todos los datos que se disponen. Estos datos se muestran en la Tabla 6:

Tabla 6. Cálculo de probabilidad a priori de ocurrencia de rampas

Número total de datos	1392
Cantidad de ocurrencia de rampas	171
P(R)	0,1228

Así se obtiene que la probabilidad a priori de ocurrencia de rampas es $P(R)=0,1228$.

Para el cálculo de las probabilidades condicionadas se utiliza la fórmula dada por la Ecuación 2.1. Así, las expresiones para la probabilidad $P(P|R)$ y $P(H|R)$ vienen dadas como sigue:

$$P(P|R) = \frac{P(P, R)}{P(R)},$$

$$P(H|R) = \frac{P(H, R)}{P(R)}.$$

Entonces el cálculo de estos parámetros se realiza de forma estadística. En el primer caso se cuentan las veces en que se dio una rampa dado que la potencia en el instante de tiempo considerado anterior (que puede ser una, dos o tres horas) estuvo en cierto rango, y para el segundo caso se cuentan cuantas rampas ocurrieron en cada diferente hora del día, y en ambos casos se divide dicho número por el número total de rampas.

En el caso de estar condicionando con respecto a la potencia, como se mencionó se evalúan tres tiempos de anticipación, una hora, dos horas y tres horas. Por simplicidad se utilizará solo un valor de potencia para la predicción, no un conjunto de datos, es decir, se considerara sólo la potencia que se generó una, dos o tres horas atrás según la ventana de predicción que se esté evaluando, y no el acumulado en este periodo. Las probabilidades condicionales en los tres casos son diferentes ya que se está condicionando frente a potencias correspondientes a diferentes instantes de tiempo para un mismo instante en que se evalúa la presencia de rampas, por lo que el

parámetro $P(P|R)$ debe calcularse para los tres casos distintos. Las probabilidades son mostradas en las siguientes tablas:

- Probabilidad condicionada a la potencia para una hora de anticipación:

Tabla 7. Probabilidad condicionada a la potencia para una anticipación de una hora

Rango de potencia [MW]	Dato en rango y ocurre rampa	$P(P=p R=1)$
0 a 1	8	0,0468
1 a 2	5	0,0292
2 a 3	10	0,0585
3 a 4	11	0,0643
4 a 5	16	0,0936
5 a 6	23	0,1345
6 a 7	14	0,0819
7 a 8	18	0,1053
8 a 9	12	0,0702
9 a 10	10	0,0585
10 a 11	12	0,0702
11 a 12	12	0,0702
12 a 13	3	0,0175
13 a 14	6	0,0351
14 a 15	8	0,0468
15 a 16	0	0,0000
16 a 17	2	0,0117
17 a 18	0	0,0000

- Probabilidad condicionada a la potencia para dos horas de anticipación:

Tabla 8. Probabilidad condicionada a la potencia para una anticipación de dos horas

Rango de potencia [MW]	Dato en rango y ocurre rampa	$P(P=p R=1)$
0 a 1	19	0,1111
1 a 2	1	0,0058
2 a 3	12	0,0702
3 a 4	21	0,1228
4 a 5	11	0,0643
5 a 6	16	0,0936
6 a 7	10	0,0585
7 a 8	16	0,0936
8 a 9	10	0,0585
9 a 10	11	0,0643
10 a 11	9	0,0526
11 a 12	4	0,0234
12 a 13	7	0,0409
13 a 14	6	0,0351
14 a 15	8	0,0468
15 a 16	5	0,0292
16 a 17	5	0,0292
17 a 18	0	0,0000

- Probabilidad condicionada a la potencia para tres horas de anticipación:

Tabla 9. Probabilidad condicionada a la potencia para una anticipación de tres horas

Rango de potencia [MW]	Dato en rango y ocurre rampa	$P(P=p R=1)$
0 a 1	22	0,1287
1 a 2	6	0,0351
2 a 3	19	0,1111
3 a 4	20	0,1170
4 a 5	10	0,0585
5 a 6	11	0,0643
6 a 7	7	0,0409
7 a 8	13	0,0760
8 a 9	11	0,0643
9 a 10	5	0,0292
10 a 11	8	0,0468
11 a 12	10	0,0585
12 a 13	5	0,0292
13 a 14	6	0,0351
14 a 15	6	0,0351
15 a 16	7	0,0409
16 a 17	3	0,0175
17 a 18	2	0,0117

Para el condicionante de la hora del día, el concepto es calcular la probabilidad de que haya una rampa en una hora específica del día dado que es efectivamente esa hora. Esto quizás suena un poco redundante, pero como se vio anteriormente la hora del día tiene incidencia en la ocurrencia de rampas ya que estas tienden a aparecer más a ciertas horas del día. Entonces, siguiendo la misma línea, para calcular el parámetro $P(H|R)$ se deben contar cuantas rampas ocurren para cada hora del día y dividir por el total de rampas ocurridas. Este conteo más el cálculo del parámetro asociado se muestra en la tabla siguiente:

Tabla 10. Probabilidad condicionada a la hora del día

Hora del día	Rampas ocurridas en cada hora	$P(H=h R=1)$
0	6	0,0351
1	5	0,0292
2	6	0,0351
3	3	0,0175
4	3	0,0175
5	3	0,0175
6	6	0,0351
7	12	0,0702
8	10	0,0585
9	9	0,0526
10	10	0,0585
11	10	0,0585
12	6	0,0351
13	6	0,0351
14	7	0,0409
15	7	0,0409
16	5	0,0292
17	7	0,0409
18	9	0,0526
19	5	0,0292
20	6	0,0351
21	11	0,0643
22	11	0,0643
23	8	0,0468

3.4 Obtención de resultados

Con los parámetros ya obtenidos se procede a obtener resultados, los cuales se contrastan con los datos reales para corroborar que tan eficaz es el método en la predicción de rampas. Para hacer esto se emplean los datos de los que se dispone correspondientes a enero del 2013. La forma en que se realizará esto es la siguiente:

- I. Primero se obtendrán las probabilidades de ocurrencia de rampas en base a los datos. Para esto se considera cada dato como si fuera el que se está obteniendo en línea en ese instante, y con estos datos se realizará el cálculo de la probabilidad de que suceda una rampa en los distintos espacios de tiempo de predicción que se consideran.

- II. Para decidir si se predice la ocurrencia de una rampa se introduce un nuevo parámetro que corresponde a un umbral de probabilidad, sobre el cual se tomará la decisión. Si la probabilidad calculada supera este umbral se dará una predicción de rampa positivo, esto es, se predice que va a ocurrir una rampa en el momento correspondiente a la ventana de predicción. Resumido en una fórmula es como sigue:

$$P(R|P, H) \geq P_{umbral} \Rightarrow \text{Se predice que ocurrirá una rampa.}$$

- III. Una vez hecha la predicción con todos los datos se procede a obtener la ocurrencia real de rampas, en base a los mismos datos. Con esto se obtiene finalmente cuantas buenas predicciones se realizaron, y se puede saber también cuántos falsos negativos y falsos positivos se obtuvieron. Como todos estos números dependen de la probabilidad umbral que se emplee, se obtendrán curvas de estos para diferentes valores de dicho umbral.

3.5 Sensibilidad frente a la magnitud de la rampa

Como complemento al trabajo se evaluará la sensibilidad del método predictor creado frente a la magnitud considerada para las rampas. Se pretende evaluar la efectividad del método para una definición de rampa considerando una magnitud mayor. En este caso se emplea una magnitud de 3 [MW] para las rampas, es decir, se considera que hubo una rampa cuando la potencia eólica cambia su magnitud en 3 [MW] en un lapso de tiempo de una hora. Para fines de obtención de resultados hay que calcular nuevamente los parámetros empelados en la red Bayesiana. No se detallará el proceso ya que este fue detallado en el punto 3.3.5. Los nuevos parámetros, correspondientes a la nueva magnitud de rampa considerada, se muestran a continuación:

- **Probabilidad a priori de ocurrencia de rampas P(R):**

Tabla 11. Cálculo de probabilidad a priori de ocurrencia de rampas para rampas de magnitud 2[MW]

Número total de datos	1392
Cantidad de ocurrencia de rampas	70
P(R)	0,05029

- Probabilidad condicionada a la hora del día:

Tabla 12. Probabilidad condicionada a la hora del día para rampas de magnitud 3[MW]

Hora del día	Rampas ocurridas en cada hora	$P(H=h R=1)$
0	4	0,0234
1	2	0,0117
2	1	0,0058
3	1	0,0058
4	1	0,0058
5	1	0,0058
6	2	0,0117
7	4	0,0234
8	4	0,0234
9	3	0,0175
10	5	0,0292
11	4	0,0234
12	1	0,0058
13	1	0,0058
14	3	0,0175
15	4	0,0234
16	1	0,0058
17	4	0,0234
18	4	0,0234
19	3	0,0175
20	2	0,0234
21	3	0,0117
22	6	0,0058
23	6	0,0058

- Probabilidad condicionada a la potencia generada a los distintos instantes de anticipación considerados:

Tabla 13. Probabilidad condicionada a la potencia para una anticipación de una hora para rampas de magnitud 3[MW]

Rango de potencia [MW]	Dato en rango y ocurre rampa	$P(P=p R=1)$
0 a 1	8	0,1143
1 a 2	0	0,0000
2 a 3	2	0,0286
3 a 4	10	0,1429
4 a 5	2	0,0286
5 a 6	7	0,1000
6 a 7	5	0,0714
7 a 8	3	0,0429
8 a 9	5	0,0714
9 a 10	5	0,0714
10 a 11	1	0,0143
11 a 12	7	0,1000
12 a 13	5	0,0714
13 a 14	3	0,0429
14 a 15	3	0,0429
15 a 16	2	0,0286
16 a 17	2	0,0286
17 a 18	0	0,0000

Tabla 14. Probabilidad condicionada a la potencia para una anticipación de dos horas para rampas de magnitud 3[MW]

Rango de potencia [MW]	Dato en rango y ocurre rampa	P(P=p R=1)
0 a 1	9	0,1286
1 a 2	0	0,0000
2 a 3	4	0,0571
3 a 4	9	0,1286
4 a 5	5	0,0714
5 a 6	6	0,0857
6 a 7	1	0,0143
7 a 8	3	0,0429
8 a 9	5	0,0714
9 a 10	4	0,0571
10 a 11	5	0,0714
11 a 12	0	0,0000
12 a 13	4	0,0571
13 a 14	3	0,0429
14 a 15	6	0,0857
15 a 16	3	0,0429
16 a 17	3	0,0429
17 a 18	0	0,0000

Tabla 15. Probabilidad condicionada a la potencia para una anticipación de tres horas para rampas de magnitud 3[MW]

Rango de potencia [MW]	Dato en rango y ocurre rampa	P(P=p R=1)
0 a 1	11	0,1571
1 a 2	2	0,0286
2 a 3	9	0,1286
3 a 4	6	0,0857
4 a 5	5	0,0714
5 a 6	2	0,0286
6 a 7	0	0,0000
7 a 8	4	0,0571
8 a 9	5	0,0714
9 a 10	2	0,0286
10 a 11	3	0,0429
11 a 12	4	0,0571
12 a 13	2	0,0286
13 a 14	3	0,0429
14 a 15	5	0,0714
15 a 16	4	0,0571
16 a 17	2	0,0286
17 a 18	1	0,0143

3.6 Comparación de diferentes resultados

Como se mencionó en el apartado 3.4, se consideraron tres tiempos de anticipación para la predicción de rampas. La finalidad de esto es comparar los resultados obtenidos en los tres casos y discernir si alguno de los casos estudiados es mejor que los demás. Para realizar esta comparación se utiliza la función llamada F_{score} , la cual depende de las predicciones acertadas, los falsos positivos y los falsos negativos. La función F_{score} asocia un puntaje entre 0 y 1, donde 1 corresponde a una predicción perfecta, mientras que un puntaje igual a cero indica que no se logró detectar ningún evento [30]. La función F_{score} a su vez se calcula en base a otros dos indicadores, la Precisión y el Recall, los cuales se definen en función del número de verdaderos positivos (VP), falsos positivos (FP) y falsos negativos (FN) tal como se muestra a continuación:

$$Precisión = \frac{VP}{(VP + FP)},$$

la cual corresponde a la fracción de eventos que fueron predichos y que realmente ocurrieron, y

$$Recall = \frac{VP}{VP + FN},$$

que es la fracción de eventos que ocurrieron y que fueron correctamente predichos.

En base a la Precisión y el Recall se define la función F_{score} , la cual corresponde a la media armónica entre ambos indicadores:

$$F_{score} = 2 \cdot \frac{Presición \cdot Recall}{Presición + Recall}$$

Luego, empleando la función definida anteriormente se obtendrán las curvas para F_{score} y se compararán, siendo el mejor caso el que presente puntajes mayores.

Cabe mencionar que la función F_{score} da igual peso a los falsos positivos que a los falsos negativos. En un caso real, los falsos negativos son mucho más indeseables que los falsos positivos, y por ello tienen más peso en un indicador más aplicado a la realidad. Es por ello que el indicador dado por la función F_{score} es solo una referencia, pero finalmente es el usuario el que debe decidir los pesos que se le asociará a los falsos negativos y falsos positivos para un indicador más acorde a sus necesidades. Aun así, en general cualquier indicador tendrá un comportamiento similar, independiente de los pesos (la forma de las curvas será parecida, solo cambiarán las escalas) por lo que a priori el indicador propuesto sirve para fines de comparación.

3.7 Comparación de métodos: Modelo autorregresivo (AR)

Finalmente, se obtendrán resultados con otro método de predicción a fin de establecer una comparación entre dos métodos diferentes.

Como se observa en la Figura 20 la autocorrelación de la potencia generada en el instante actual con la generada en instantes previos es alta. Es por esto que a modo de comparación se utilizará un modelo autorregresivo. La razón de emplear este modelo es que es un modelo simple pero a pesar de esto es capaz de capturar el fenómeno asociado a la influencia que tienen valores pasados de la variable en su valor actual, en este caso la potencia generada.

Un proceso autorregresivo corresponde a un tipo de proceso en que el valor actual de la variable corresponde a una combinación lineal de las variables pasadas. Más específicamente, un proceso autorregresivo de orden p es aquel que puede ser representado mediante la Ecuación 3.3:

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t \quad (3.3)$$

Donde:

- Y_t : valor de la variable en el instante t
- Φ_i : parámetros que definen el modelo
- ϵ_t : término asociado al error. Corresponde a ruido blanco (proceso estocástico de media 0 y varianza σ^2)

Para este caso se escogió usar un modelo de orden $p=3$. Para estimar los parámetros se empleó el código de Matlab® mostrado en el Anexo C. Éste

código fue creado por Harry L. Hurd y se encuentra de forma gratuita en internet [31]. Con este código se obtienen los parámetros del modelo, mostrado a continuación:

$$\begin{aligned}\phi_1 &= 1,805 \\ \phi_2 &= -0,9118 \\ \phi_3 &= -0,5714\end{aligned}$$

Finalmente, el código empleado para la predicción mediante el modelo AR se muestra en el Anexo D.

3.8 Diagrama de bloques

En la Figura 23 se presenta el diagrama de bloques resumiendo la metodología.

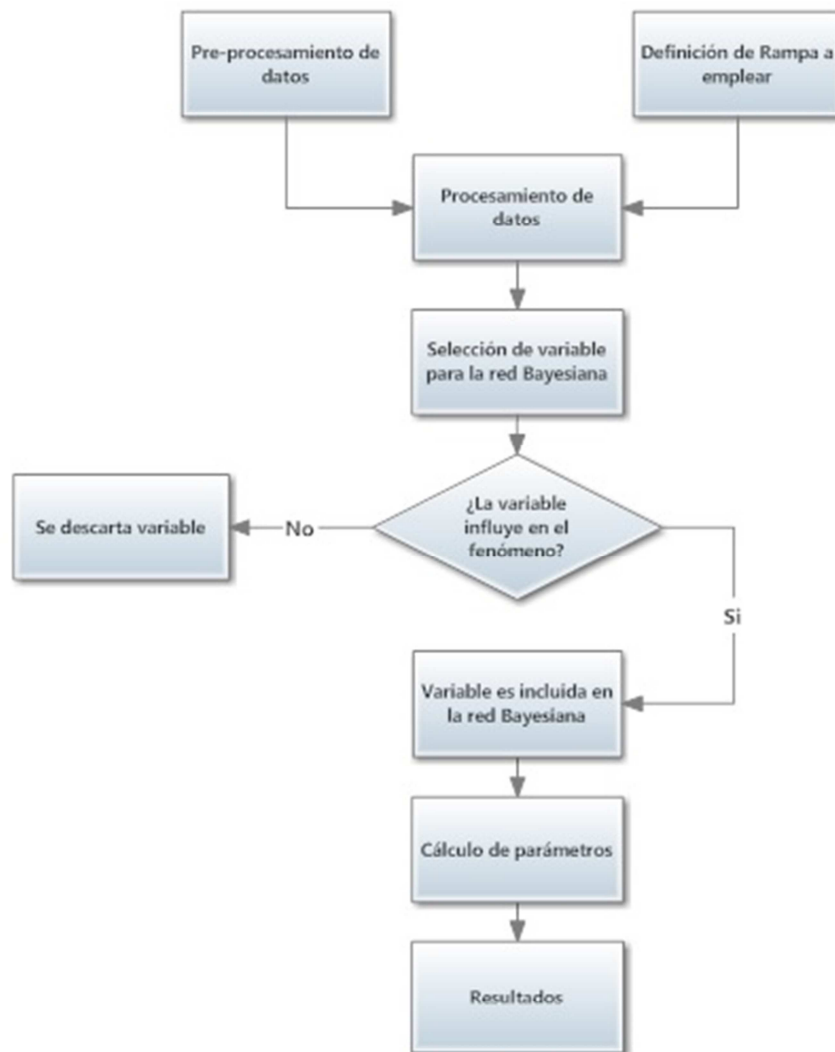


Figura 23. Diagrama de bloques de la metodología

4 Resultados

En este capítulo se procede a exponer y analizar los resultados obtenidos durante este trabajo. El modelo implementado está basado en un enfoque probabilístico condicionando a la evidencia de la que se dispone, en este caso la hora del día a la que se está analizando la posible aparición de una rampa en la potencia, y los valores de potencia generados previamente al momento en el que se está realizando la predicción. Los resultados se evalúan contrastando con los eventos realmente ocurridos mediante una serie de indicadores.

El software empleado para la realización de este trabajo fue MATLAB® (R2010a). El cálculo de parámetros y las posteriores predicciones fueron obtenidos mediante programas realizados en este software. Estos programas fueron desarrollados íntegramente y son mostrados en la sección Anexos.

Se obtuvieron una serie de resultados, los cuales se muestran a continuación. Primero se obtiene la probabilidad de ocurrencia de rampas para el mes de enero del 2013 y se contrasta con las rampas que realmente ocurrieron durante dicho mes. Después se muestra el comportamiento de los resultados obtenidos en función del parámetro “probabilidad umbral” que se escoja, en términos de falsos negativos, falsos positivos y predicciones acertadas obtenidas. Estos son mostrados como curvas en función del parámetro, en forma del conteo total y también de manera porcentual en base a la cantidad total de rampas de enero. Finalmente y en busca de mejorar los resultados se realizó la predicción considerando una ventana de tiempo en vez de predecir para una hora en específico, y los resultados también son analizados en relación a falsos negativos, falsos positivos y predicciones acertadas obtenidas, creándose para esto las mismas curvas anteriores en función de la probabilidad umbral escogida.

4.1 Primer caso: rampas de magnitud 2 [MW]

Los primeros resultados que se obtienen consideran rampas de magnitud 2[MW]. Cabe recordar que en los meses correspondientes al periodo de prueba, noviembre y diciembre del 2012, se contabilizaron 171 rampas, mientras que en el periodo de prueba, enero del 2013, hubo 92 rampas. La serie de resultados obtenidos se muestran en las subsecciones siguientes.

4.1.1 Probabilidad de ocurrencia de rampa

Lo primero en obtenerse fueron las probabilidades de ocurrencia de rampas para cada hora de enero. Se consideraron tres ventanas de predicción, con una, dos y tres horas de anticipación, a fin de comparar resultados, y estos se comparan con la ocurrencia real de rampas en enero.

- Tiempo de anticipación de una hora

Los resultados obtenidos empleando una ventana de predicción de una hora se muestran en la Figura 24:

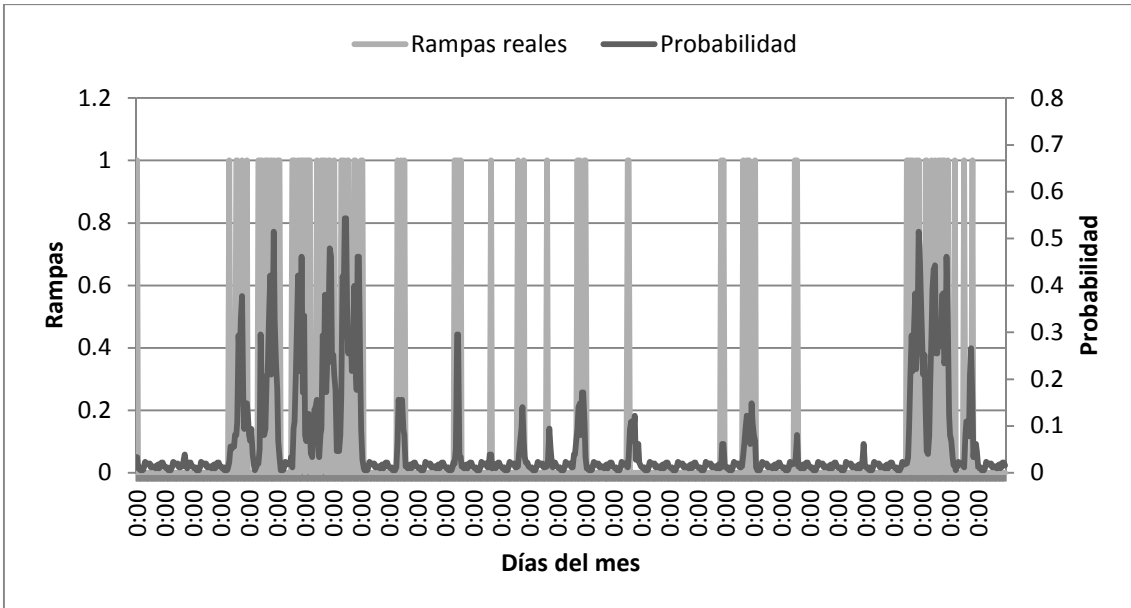


Figura 24. Probabilidades de ocurrencia de rampas mayores a 2 [MW] para tiempo de anticipación de una hora y rampas realmente ocurridas

- **Tiempo de anticipación de dos horas**

Para la ventana de predicción de dos horas, los resultados obtenidos son los mostrados en la Figura 25

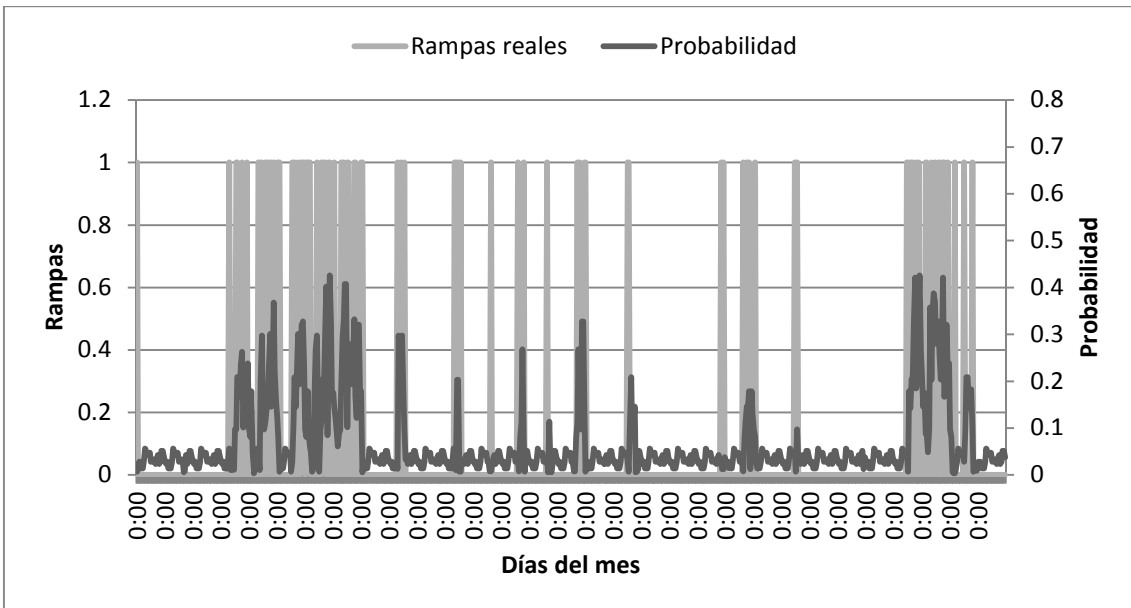


Figura 25. Probabilidades de ocurrencia de rampas mayores a 2 [MW] para tiempo de anticipación de dos horas y rampas realmente ocurridas

- **Tiempo de anticipación de tres horas**

Finalmente, los resultados obtenidos al aplicar una ventana de predicción de tres horas son los mostrados en la

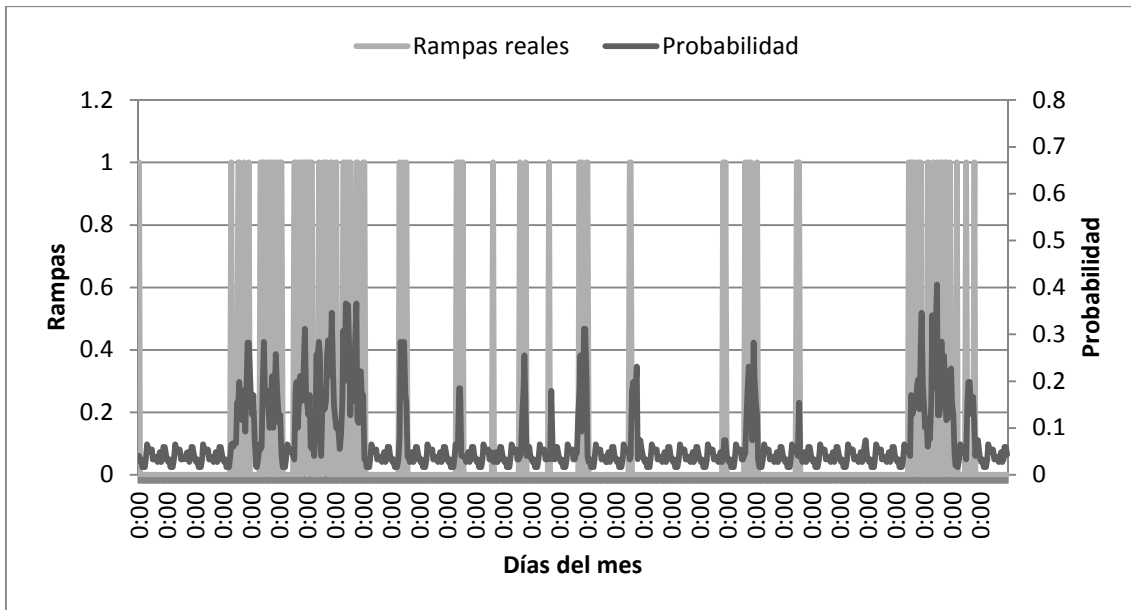


Figura 26. Probabilidades de ocurrencia de rampas mayores a 2 [MW] para tiempo de anticipación de tres horas y rampas realmente ocurridas

Como se aprecia en las Figuras anteriores, la probabilidad calculada para la ocurrencia de rampas tiende a subir notoriamente cuando la densidad de rampas que realmente ocurrieron aumenta, desde el punto de vista temporal. Esto a priori es bueno ya que cuando tienden a aparecer más rampas la probabilidad calculada para las tres ventanas aumenta notoriamente, lo que estaría asociado a una buena capacidad de predicción.

En cuanto a los valores de probabilidad obtenidos, se tiene que en el caso de considerar un tiempo de predicción de una hora los valores máximos de probabilidad obtenidos sobrepasan el 60%, mientras que para una anticipación de dos horas dicha probabilidad está entre el 40 y el 60%, y para una anticipación de tres horas los máximos obtenidos para la probabilidad están entre un 35% y un 40%. Esto se explica porque mientras más cercano se encuentra el instante de tiempo para el que se realiza la predicción, más certeza se tiene acerca de lo que puede ocurrir.

4.1.2 Sensibilidad respecto al umbral

Se realiza un análisis de sensibilidad de los resultados respecto al umbral de decisión, en función de la cantidad de predicciones acertadas, falsos positivos y falsos negativos se obtienen para distintos valores de dicho umbral. A continuación se muestran estos resultados, considerando las tres ventanas de predicción.

- **Tiempo de anticipación de una hora**

Tabla 16. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de una hora

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,01	514	0	92	5,587	0,000	1,000
0,02	196	11	81	2,130	0,120	0,880
0,03	128	13	79	1,391	0,141	0,859
0,04	114	21	71	1,239	0,228	0,772
0,05	108	22	70	1,174	0,239	0,761
0,06	103	24	68	1,120	0,261	0,739
0,07	90	30	62	0,978	0,326	0,674
0,08	88	30	62	0,957	0,326	0,674
0,09	75	34	58	0,815	0,370	0,630
0,10	66	41	51	0,717	0,446	0,554
0,11	62	44	48	0,674	0,478	0,522
0,12	62	44	48	0,674	0,478	0,522
0,13	59	47	45	0,641	0,511	0,489
0,14	57	48	44	0,620	0,522	0,478
0,15	50	51	41	0,543	0,554	0,446
0,16	48	53	39	0,522	0,576	0,424
0,17	48	54	38	0,522	0,587	0,413
0,18	43	56	36	0,467	0,609	0,391
0,19	42	56	36	0,457	0,609	0,391
0,20	42	56	36	0,457	0,609	0,391

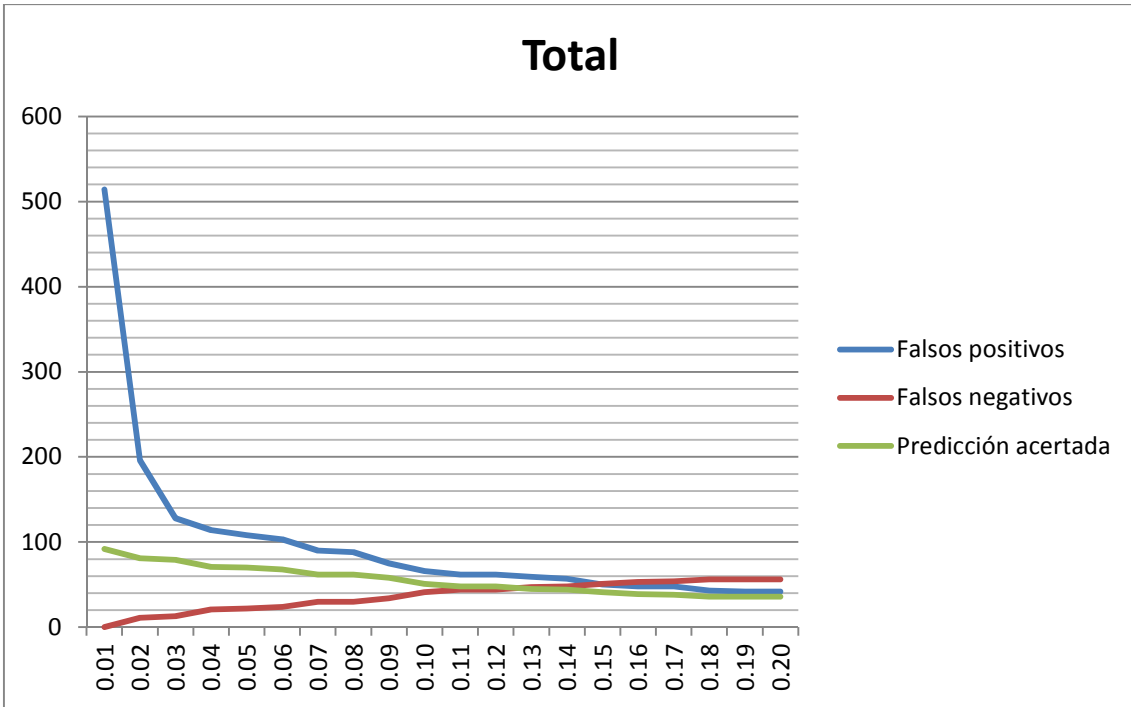


Figura 27. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora

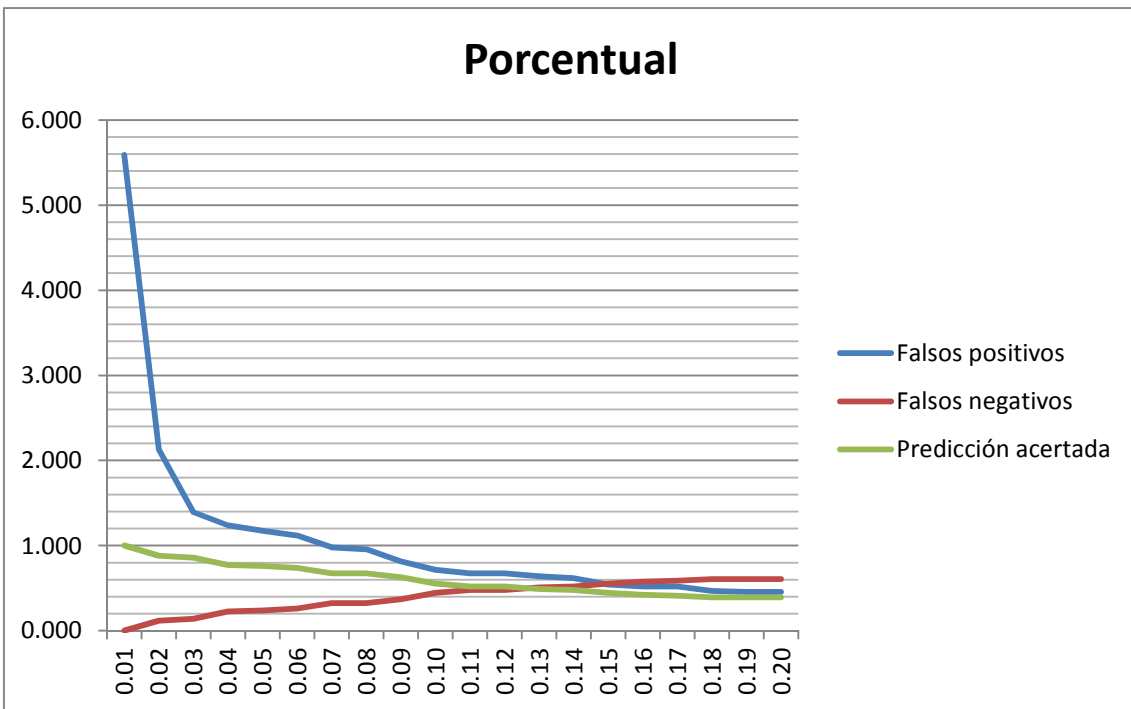


Figura 28. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora. Debe considerarse el 100% como el total de rampas ocurridas en enero, las cuales fueron 92

- **Tiempo de anticipación de dos horas**

Tabla 17. Verdaderos positivos, falsos positivos y falsos negativos obtenidos con un tiempo de anticipación de dos horas

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,01	631	10	82	6,859	0,109	0,891
0,02	539	14	78	5,859	0,152	0,848
0,03	351	18	74	3,815	0,196	0,804
0,04	277	20	72	3,011	0,217	0,783
0,05	162	24	68	1,761	0,261	0,739
0,06	98	25	67	1,065	0,272	0,728
0,07	97	26	66	1,054	0,283	0,717
0,08	97	26	66	1,054	0,283	0,717
0,09	88	33	59	0,957	0,359	0,641
0,10	76	36	56	0,826	0,391	0,609
0,11	75	39	53	0,815	0,424	0,576
0,12	70	43	49	0,761	0,467	0,533
0,13	67	45	47	0,728	0,489	0,511
0,14	67	45	47	0,728	0,489	0,511
0,15	61	51	41	0,663	0,554	0,446
0,16	61	51	41	0,663	0,554	0,446
0,17	55	54	38	0,598	0,587	0,413
0,18	46	58	34	0,500	0,630	0,370
0,19	45	59	33	0,489	0,641	0,359
0,20	43	62	30	0,467	0,674	0,326

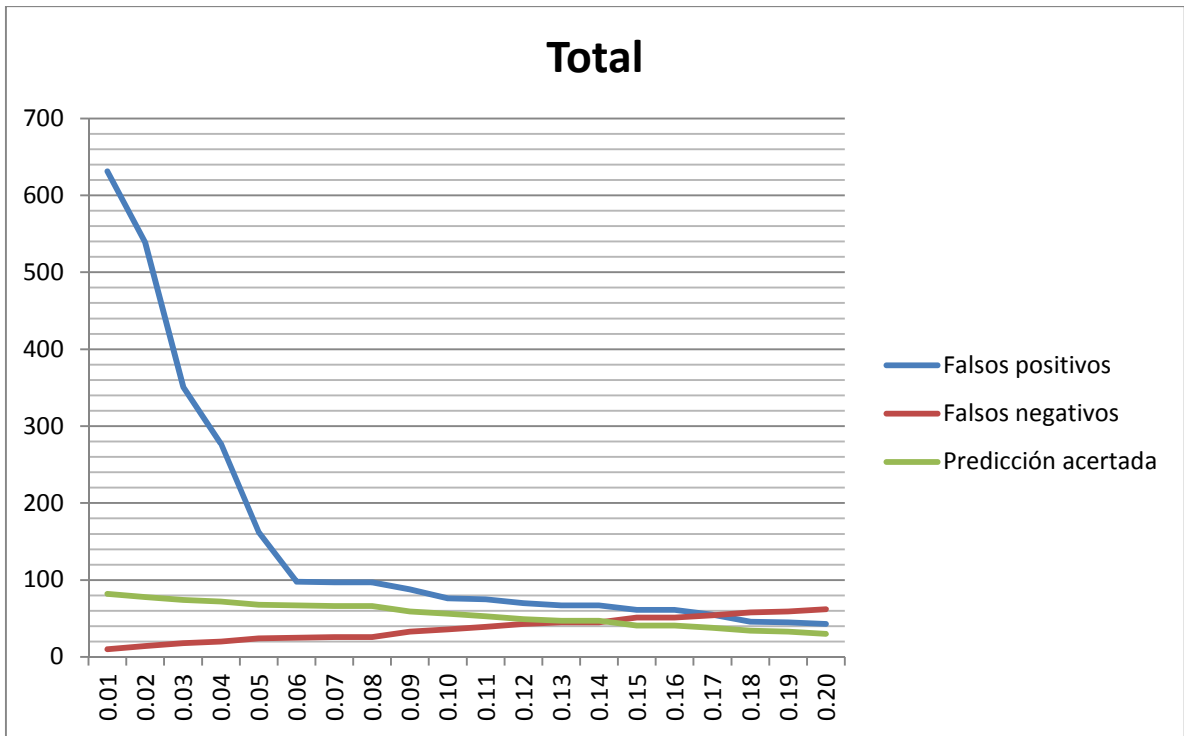


Figura 29. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas

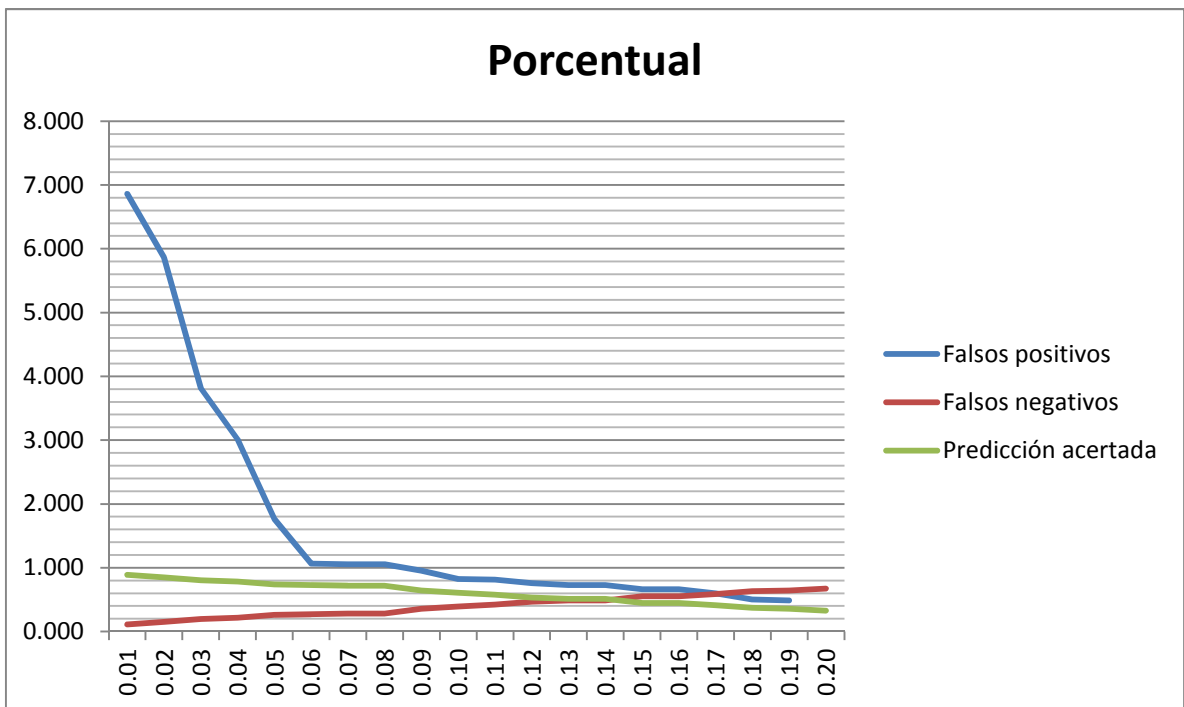


Figura 30. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas. El 100% son las 92 rampas ocurridas en enero

- **Tiempo de anticipación de tres horas:**

Tabla 18. Verdaderos positivos, falsos positivos y falsos negativos obtenidos con un tiempo de anticipación de tres horas

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,01	652	0	92	7,087	0,000	1,000
0,02	572	0	92	6,217	0,000	1,000
0,03	512	1	91	5,565	0,011	0,989
0,04	325	8	84	3,533	0,087	0,913
0,05	248	16	76	2,696	0,174	0,826
0,06	138	21	71	1,500	0,228	0,772
0,07	107	26	66	1,163	0,283	0,717
0,08	98	27	65	1,065	0,293	0,707
0,09	95	29	63	1,033	0,315	0,685
0,10	94	31	61	1,022	0,337	0,663
0,11	92	34	58	1,000	0,370	0,630
0,12	89	37	55	0,967	0,402	0,598
0,13	77	44	48	0,837	0,478	0,522
0,14	77	44	48	0,837	0,478	0,522
0,15	67	47	45	0,728	0,511	0,489
0,16	53	52	40	0,576	0,565	0,435
0,17	49	53	39	0,533	0,576	0,424
0,18	40	60	32	0,435	0,652	0,348
0,19	36	64	28	0,391	0,696	0,304
0,20	32	67	25	0,348	0,728	0,272

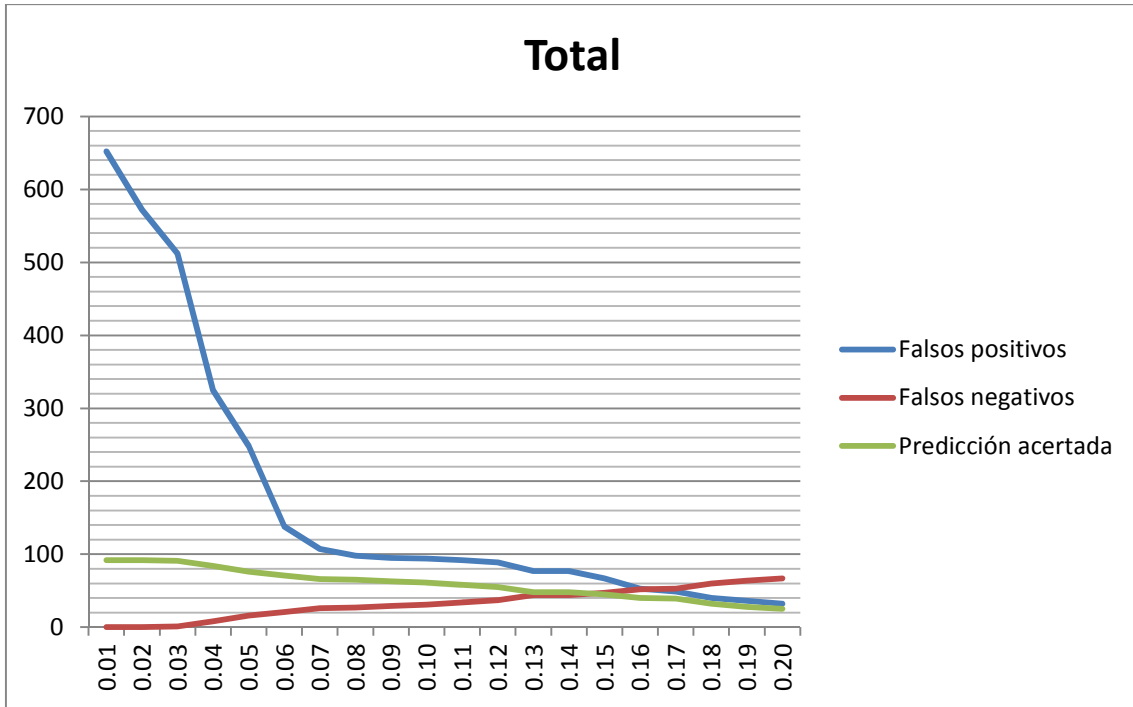


Figura 31. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas

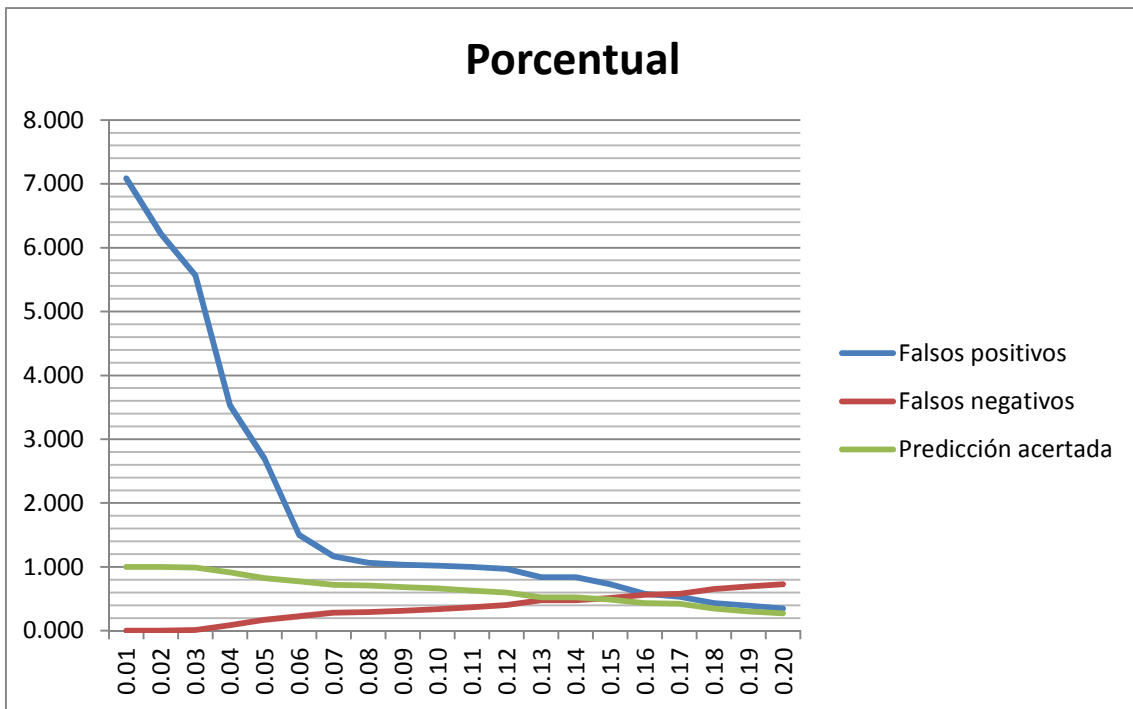


Figura 32. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas. El 100% son las 92 rampas ocurridas en enero

Como se aprecia en las figuras y tablas mostradas previamente, los resultados son altamente sensibles a la variación del parámetro probabilidad umbral. Con un criterio más bien relajado, esto es, escogiendo un parámetro umbral menor,

se tiene una tasa de predicciones acertadas alta, pudiendo detectar el 100% de las rampas sucedidas en caso de ventanas de predicción de una y tres horas, y 89% de predicción acertada en caso de una ventana de dos horas. Sin embargo al aplicar un criterio más relajado se tiene el problema de que aparecen muchos falsos positivos, los cuales corresponden a predicciones positivas de rampa cuando en verdad no ocurren. Como se ve, para parámetros de umbral de alrededor de $P_{\text{umbral}}=0,01$ se llega a tasas de falsos positivos de sobre el 500% (sobre 660% en caso de la ventanas de predicción de dos y tres horas).

Al usar un criterio más estricto en términos de la probabilidad umbral empleada (entiéndase aumentar dicho parámetro) se logra disminuir bastante el número de falsos negativos, llegando a tasas porcentuales de 45% en caso de ventanas de predicción de una y dos horas, y a un 35% en caso de tres horas para un umbral $P_{\text{umbral}}=0,02$. Sin embargo aquí se produce un trade-off entre dicho indicador y el número de predicciones correctas. Así, al aumentar el umbral escogido la tasa de predicciones correctas disminuye, desde un 100% para ventanas de predicción una y tres horas y 89% para la ventana de dos horas con un umbral $P_{\text{umbral}}=0,02$, hasta valores de predicción acertada menores a 40% para un umbral $P_{\text{umbral}}=0,2$ (en caso de la ventana de predicción de tres horas el valor disminuye a poco más del 27%).

Se aprecia entonces que simplemente variando el parámetro umbral empleado no se mejoran sustancialmente los resultados, ya que si bien al aumentar el valor de este se disminuyen los falsos negativos que existen también esto disminuye el desempeño del predictor en términos de predicciones acertadas.

Con fines comparativos se obtiene la función F_{score} para los tres casos, la cual se muestra en la Figura 33:

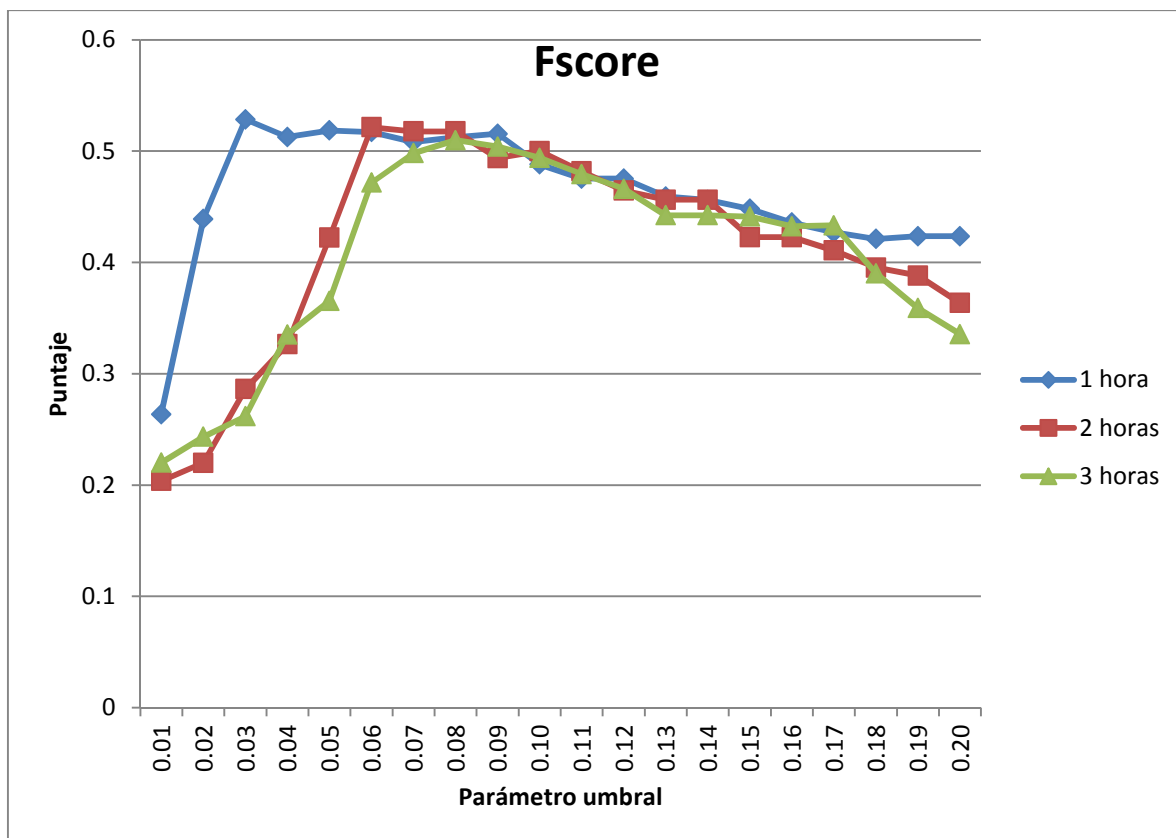


Figura 33. Función Fscore para predicciones con una, dos y tres horas de anticipación

Como se puede apreciar, para umbrales pequeños, menores al 6% de probabilidad, el puntaje asociado a la predicción con una hora de antelación es mayor a los otros dos casos, mientras que en el rango de 6% a 18% los tres casos presentan puntajes similares, y finalmente para umbrales mayores a 18% el puntaje asociado a la predicción con una hora de anticipación vuelve a superar a los otros dos. Según esto entonces, utilizando la función F_{score} la elección entre los tres casos tratados depende del umbral escogido. Ahora, como se discutió en la sección Metodología, el usuario debe escoger el umbral según los pesos que desee darle a los falsos positivos y a los falsos negativos, y esto será lo que finalmente lleve a decidir cuál de las tres opciones de tiempos de antelación se escogería.

4.1.3 Incidencia de falsos negativos

Como se ha discutido, en términos de seguridad en la operación del sistema son los falsos negativos los que colocarían al sistema en estados de operación más riesgosos, y por lo tanto son el problema de detección más indeseable. Sin embargo estos falsos negativos pueden tener diferentes impactos sobre el sistema. En este caso, el impacto que una eventual rampa en la potencia de viento que no fue detectada tiene directa relación con su magnitud. Es así como los falsos negativos pueden tener un mayor o menor impacto.

En el caso específico estudiado se indaga acerca de la incidencia que tendrían los falsos negativos en el sistema. Para esto se separaron en rangos según la potencia. Así, se diferenciaron tres clases: rampas entre 2 [MW] y 3 [MW] de potencia, rampas entre 3 [MW] y 4 [MW] y finalmente rampas mayores a

4[MW]. Recordando que la potencia instalada de la central es de 18,15 [MW], en términos de porcentaje sobre este total los rangos corresponden a [11%; 16,5%), [16,5%; 22%) y [22%; inf). Los resultados se muestran a continuación, en función del parámetro umbral y para los tres tiempos de anticipación considerados:

- **Tiempo de anticipación de una hora**

Tabla 19. Falsos negativos separados por magnitud para tiempo de anticipación de una hora

Parámetro	Falsos Negativos	FN ente 2 y 3 MW	FN entre 3 y 4 MW	FN mayores a 4 MW
0,01	0	0	0	0
0,02	11	11	0	0
0,03	13	13	0	0
0,04	21	21	0	0
0,05	22	22	0	0
0,06	24	24	0	0
0,07	30	30	0	0
0,08	30	30	0	0
0,09	34	34	0	0
0,10	41	41	0	0
0,11	44	44	0	0
0,12	44	44	0	0
0,13	47	47	0	0
0,14	48	48	0	0
0,15	51	51	0	0
0,16	53	53	0	0
0,17	54	53	1	0
0,18	56	55	1	0
0,19	56	55	1	0
0,20	56	55	1	0

- **Tiempo de anticipación de dos horas**

Tabla 20. Falsos negativos separados por magnitud para tiempo de anticipación de dos horas

Parámetro	Falsos Negativos	FN ente 2 y 3 MW	FN entre 3 y 4 MW	FN mayores a 4 MW
0,01	10	10	0	0
0,02	14	14	0	0
0,03	18	18	0	0
0,04	20	20	0	0
0,05	24	24	0	0
0,06	25	25	0	0
0,07	26	26	0	0
0,08	26	26	0	0
0,09	33	33	0	0
0,10	36	36	0	0
0,11	39	39	0	0
0,12	43	43	0	0
0,13	45	45	0	0
0,14	45	45	0	0
0,15	51	51	0	0
0,16	51	51	0	0
0,17	54	54	0	0
0,18	58	58	0	0
0,19	59	59	0	0
0,20	62	61	1	0

- **Tiempo de anticipación de tres horas**

Tabla 21. Falsos negativos separados por magnitud para tiempo de anticipación de tres horas

Parámetro	Falsos Negativos	FN ente 2 y 3 MW	FN entre 3 y 4 MW	FN mayores a 4 MW
0,01	0	0	0	0
0,02	0	0	0	0
0,03	1	1	0	0
0,04	8	8	0	0
0,05	16	16	0	0
0,06	21	21	0	0
0,07	26	26	0	0
0,08	27	27	0	0
0,09	29	29	0	0
0,10	31	31	0	0
0,11	34	34	0	0
0,12	37	37	0	0
0,13	44	44	0	0
0,14	44	44	0	0
0,15	47	47	0	0
0,16	52	52	0	0
0,17	53	53	0	0
0,18	60	60	0	0
0,19	64	63	1	0
0,20	67	66	1	0

Se aprecia entonces en las tablas mostradas anteriormente que si bien el número de falsos positivos aumenta de manera considerable al usar valores más estrictos para el umbral (esto es, más altos) el valor de la magnitud de la rampa que no está siendo detectada no es muy alto, estando la mayoría en el rango de 2-3 [MW]. En los tres casos, a lo sumo una rampa de las que no son correctamente pronosticadas está en el rango de 3-4 [MW], y ninguna en rangos mayores. Esto indica que si bien hay un número alto de falsos negativos para umbrales más grandes, estos no serían tan graves dado la magnitud de potencia que representan.

4.1.4 Predicción en una ventana horaria

Como se vio, un problema con la realización de predicciones es la alta tasa de falsos positivos que se obtienen. Una de las razones por las que se obtienen tantos falsos positivos es que muchas predicciones de ocurrencia rampas que se realizan están desfasadas en tiempo, vale decir, ciertas predicciones de ocurrencia de rampas prevén la aparición de una rampa en cierta hora t del día, pero la rampa se dio o bien en la hora anterior o en la hora posterior. Para evitar este tipo de problemas se propone una mejora al sistema de predicción desarrollado. Ya que muchas de las predicciones de ocurrencia de rampas están con cierto desfase respecto a la real ocurrencia del fenómeno se estudia la efectividad del método predictor usando una ventana de predicción en vez de realizar predicciones para horas específicas del día. Así, la predicción corresponde a la posible aparición de una rampa en una ventana u rango horario.

Para la realización de la predicción en base a la probabilidad acumulada, y por consiguiente en base a un rango horario, se emplearon las probabilidades obtenidas para la ventana de predicción de tres horas y se empleó un rango de dos horas, como lo muestra la Figura 34. Con esto es posible poder predecir con dos horas de anticipación la aparición de una rampa en un rango de dos horas.

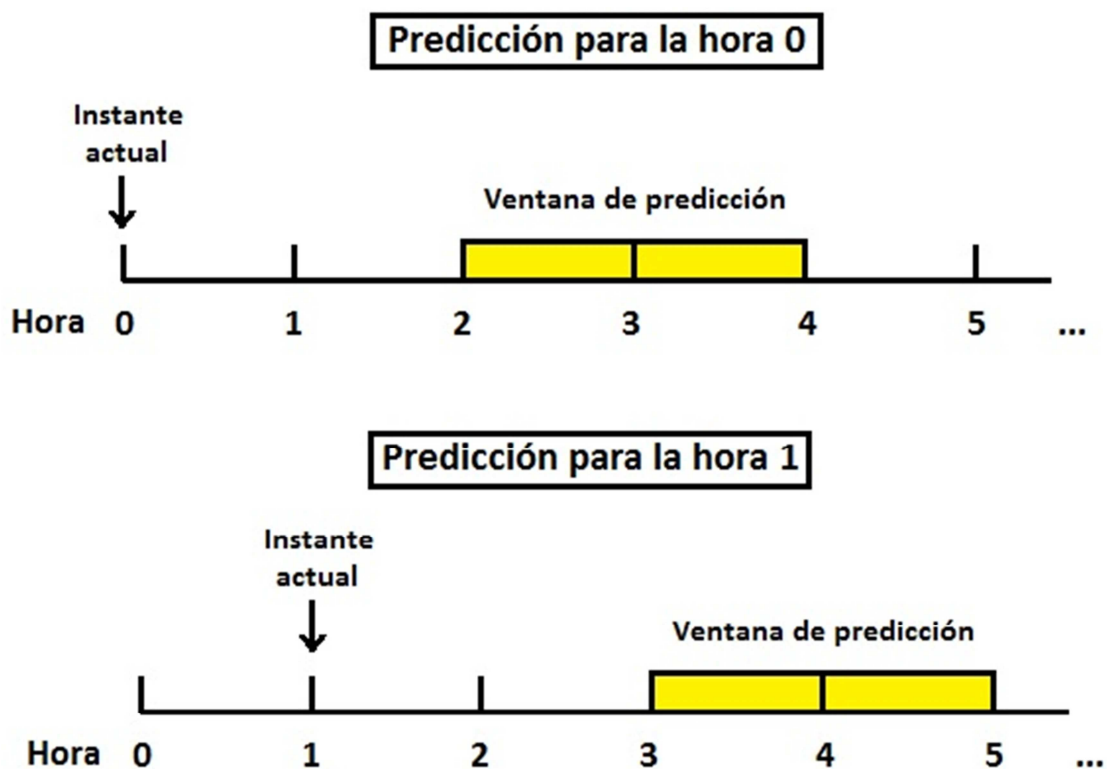


Figura 34. Esquema de predicción con dos horas de anticipación para una ventana horaria de dos horas

Es importante tener claro que para cada hora específica la potencia que condiciona la probabilidad que se calcula es la que se generó tres horas antes, pero esto no corresponde para la ventana completa. Vale decir, si se quiere predecir la probabilidad de ocurrencia de una rampa en una ventana que va de

las 3 a las 5 por ejemplo, tal como muestra el segundo dibujo de la Figura 34, primero se calcula la probabilidad de cada hora dentro del rango condicionado a la potencia generada tres horas antes, es decir, para la probabilidad de ocurrencia de una rampa a las 3 se utiliza como condicionante la potencia generada a las 0 horas, pero para la probabilidad obtenida para las 4 se condiciona según la potencia que se genera a la 1. Luego, teniendo ambas se obtiene el acumulado.

Los resultados obtenidos son mostrados en la Tabla 22, la Figura 35 y la Figura 36:

Tabla 22. Verdaderos positivos, falsos positivos y falsos negativos obtenidos prediciendo en base a una probabilidad acumulada en una ventana de dos horas

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,04	541	1	149	3,607	0,007	0,993
0,048	541	2	148	3,607	0,013	0,987
0,056	488	2	148	3,253	0,013	0,987
0,064	426	4	146	2,840	0,027	0,973
0,072	356	9	141	2,373	0,060	0,940
0,08	298	14	136	1,987	0,093	0,907
0,088	260	19	131	1,733	0,127	0,873
0,096	239	22	128	1,593	0,147	0,853
0,104	143	23	127	0,953	0,153	0,847
0,112	121	26	124	0,807	0,173	0,827
0,12	75	33	117	0,500	0,220	0,780
0,128	74	33	117	0,493	0,220	0,780
0,136	68	36	114	0,453	0,240	0,760
0,144	66	36	114	0,440	0,240	0,760

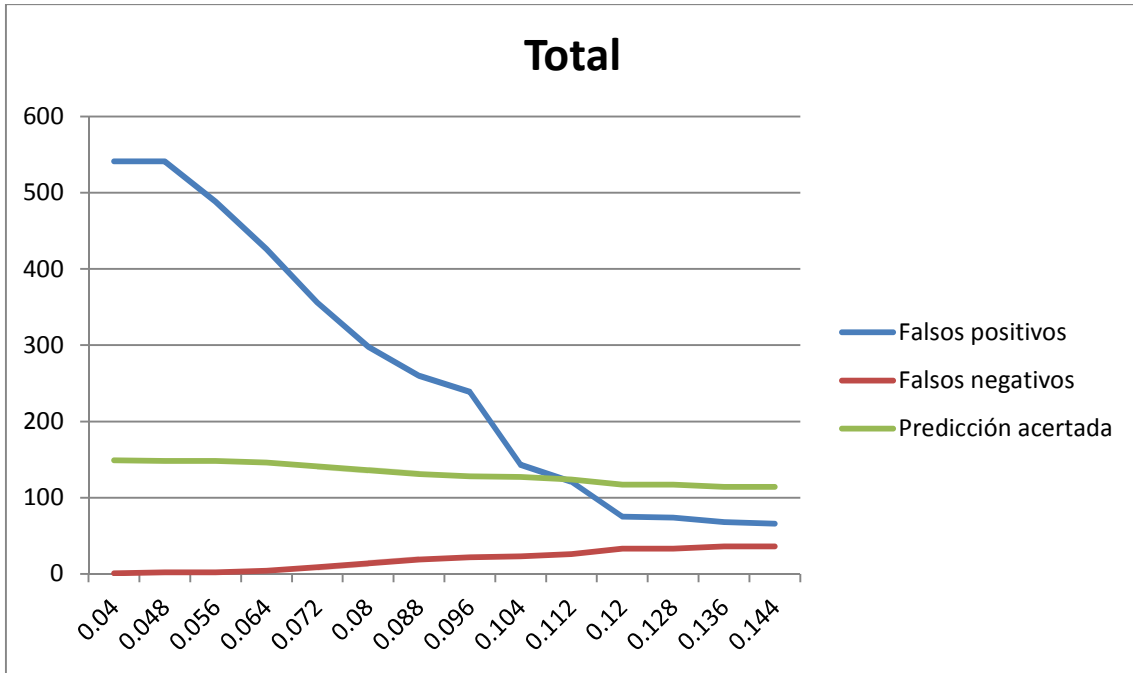


Figura 35. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas

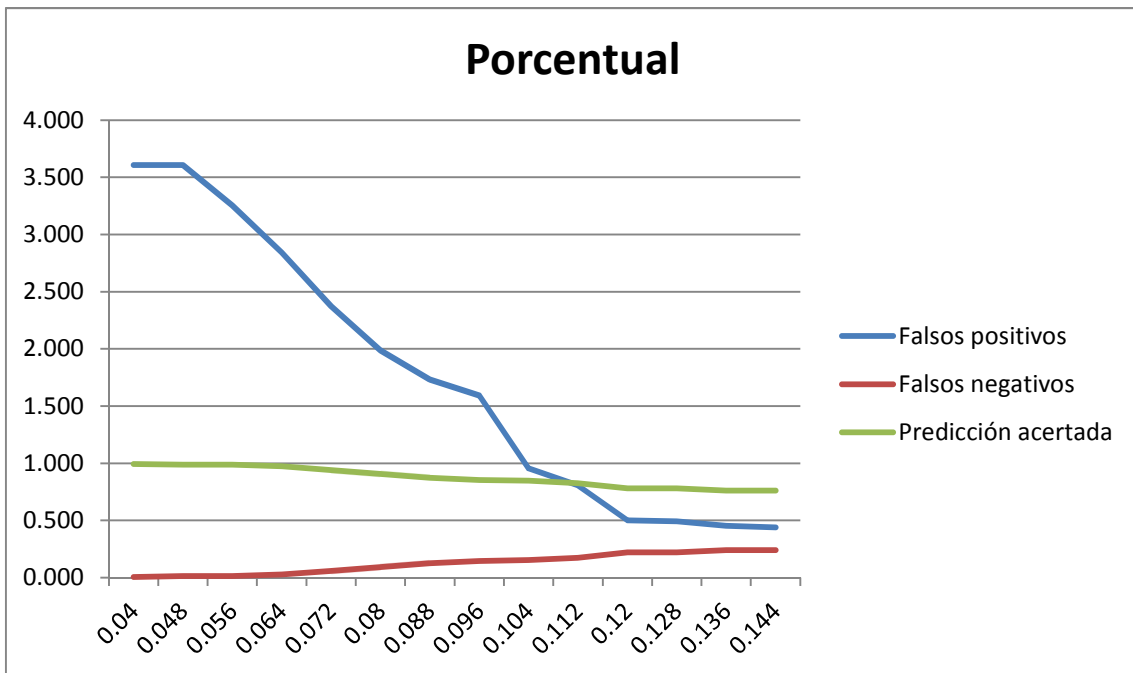


Figura 36. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas. El 100% corresponde al número de ventanas de dos horas en el que si hubieron rampas, en este caso 150

Se aprecia que al realizar la predicción sobre una ventana horaria en base a probabilidades acumuladas los resultados mejoran bastante. Se aprecia por ejemplo que al escoger un umbral $P_{umbral}=0,04$ la tasa de predicción acertada es de un 99% mientras que la tasa de falsos positivos es de poco más de un 350%. Si bien en primera instancia es alta esta tasa, es considerablemente más baja que la que se obtuvo realizando la predicción para horarios específicos, en que dicho valor aumentaba casi al doble.

Se ve claramente también que al usar un criterio más riguroso aumentando el parámetro umbral la tasa de falsos positivos disminuye en gran medida mientras que la tasa de detecciones acertadas no se ve afectada en la misma medida. Así, cuando el umbral escogido es $P_{umbral}=0,144$ la tasa de falsos positivos disminuye a un 44%, mientras que la tasa de detección acertada no baja del 76%.

Hay que recalcar que, si bien los resultados obtenidos para este caso son mucho mejores que los anteriores, no desaparece el inconveniente de tener un trade-off entre la tasa de falsos positivos y la tasa de detecciones acertadas. Aun así, el hecho de aumentar el umbral afecta en mucha mayor medida a la tasa de falsos positivos que la tasa de detecciones acertadas, habiendo valores para dicho umbral en que estas dos tasas alcanzan valores aceptables. Ejemplo de esto es usar un umbral $P_{umbral}=0,104$, para el cual se obtiene una tasa de falsos positivos de un 95% mientras que las predicciones acertadas alcanzan casi el 85%.

A modo de comparación entre los casos estudiados se muestra en la Figura 37 la función F_{score} para tiempos de anticipación de una, dos y tres horas, y además la función asociada a emplear la predicción en una ventana horaria de dos horas.

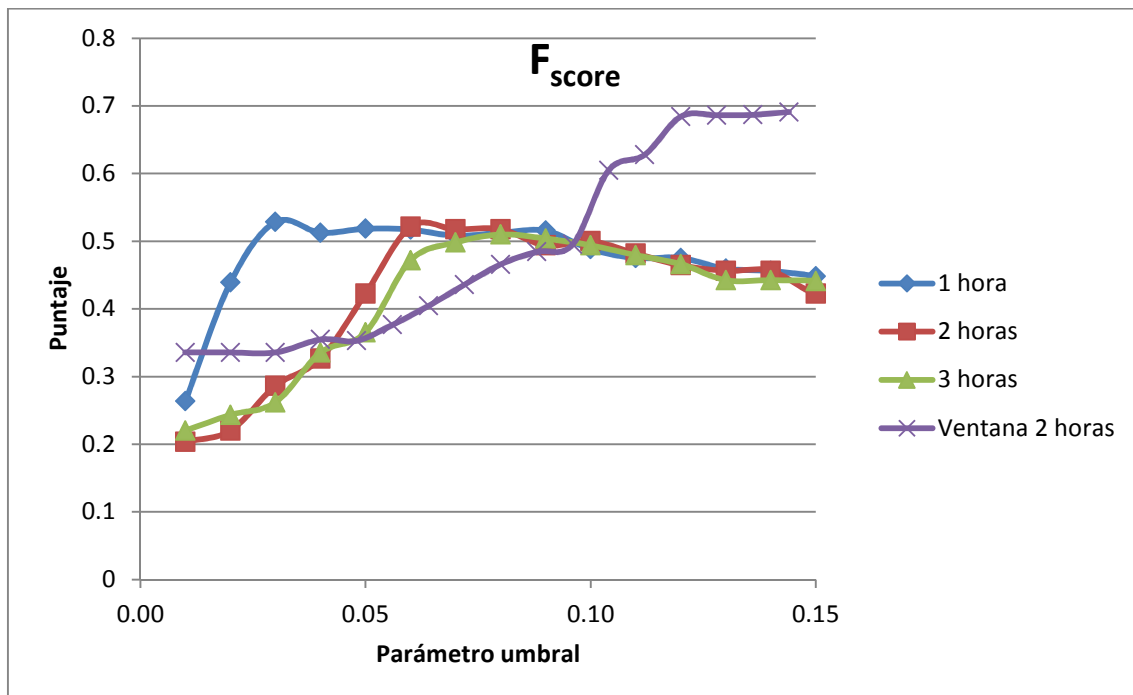


Figura 37. Función Fscore para predicciones con una, dos y tres horas de anticipación, y empleando una ventana de predicción de dos horas

Como se aprecia, para umbrales a 10% la función F_{score} asociada a la ventana de predicción no es mejor que los casos estudiados previamente. Sin embargo para umbrales superiores al valor mencionado el puntaje para la ventana es claramente mayor a los otros tres casos. Ahora, teniendo en cuenta la información dada por la Tabla 22 y sus gráficos asociados, en términos de falsos positivos, falsos negativos y predicciones acertadas el desempeño del

predicor es mejor utilizando umbrales mayores a 10%, por lo que se concluye que es mucho mejor considerar una ventana de predicción de dos horas.

4.2 Segundo caso: rampas de magnitud 3 [MW]

Siguiendo el mismo procedimiento empleado para consideración de rampas de 2 [MW] de magnitud, se procede a obtener resultados pero considerando rampas más grandes, de mínimo 3 [MW] de magnitud. Cabe mencionar que, dada la magnitud mayor considerada para la potencia de las rampas, el número de estas disminuye. Es así como para este caso, en los meses del periodo de entrenamiento se contabilizaron 70 rampas, mientras que en el mes del periodo de prueba se contabilizaron solo 4. Dado esto, en esta subsección se evaluará el desempeño del predicor en situaciones de pronóstico de eventos muy raros. Los resultados obtenidos son mostrados en las siguientes páginas.

4.2.1 Probabilidad de ocurrencia de las rampas

Tal como se hizo previamente, el primer resultado obtenido es la probabilidad de ocurrencia de las rampas, y esta se contrasta con las rampas realmente ocurridas durante enero. Esto se realiza para los tres tiempos de anticipación considerados: una, dos y tres horas.

- Tiempo de anticipación de una hora

Resultados mostrados en la Figura 38

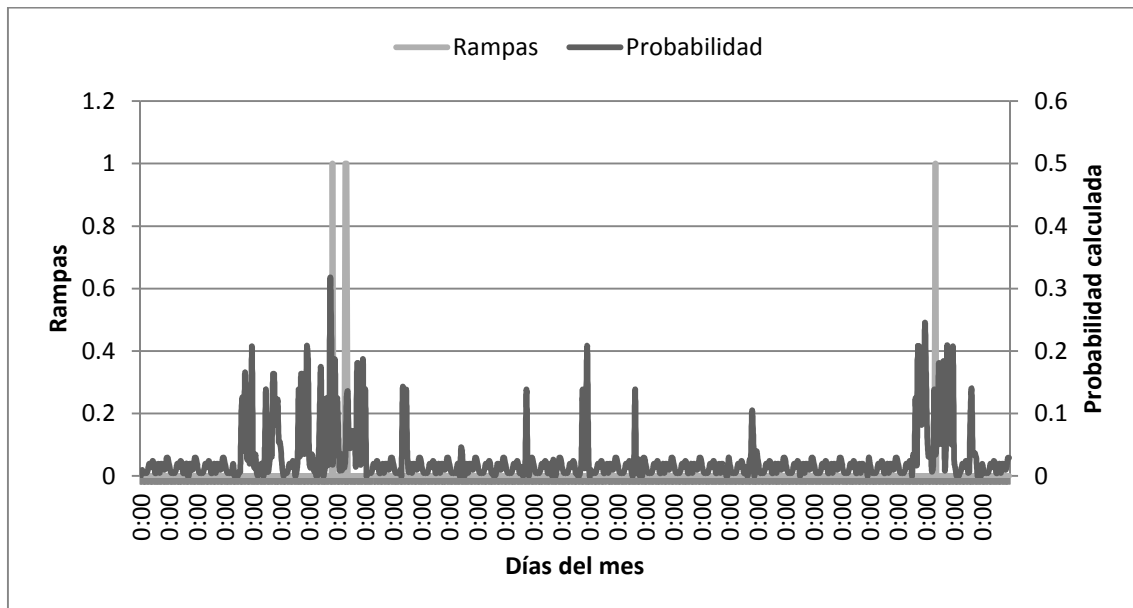


Figura 38. Probabilidades de ocurrencia de rampas mayores a 3 [MW] para tiempo de anticipación de una hora y rampas realmente ocurridas

- Tiempo de anticipación de dos horas

Resultados mostrados en la Figura 39

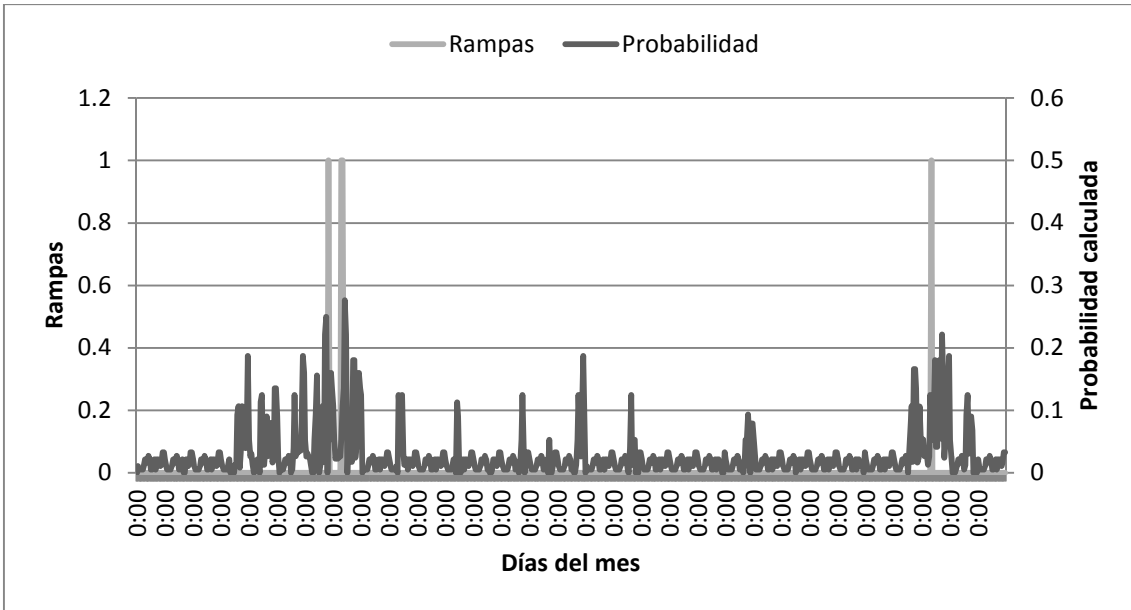


Figura 39. Probabilidades de ocurrencia de rampas mayores a 3 [MW] para tiempo de anticipación de dos horas y rampas realmente ocurridas

- **Tiempo de anticipación de tres horas**

Resultados mostrados en la Figura 40

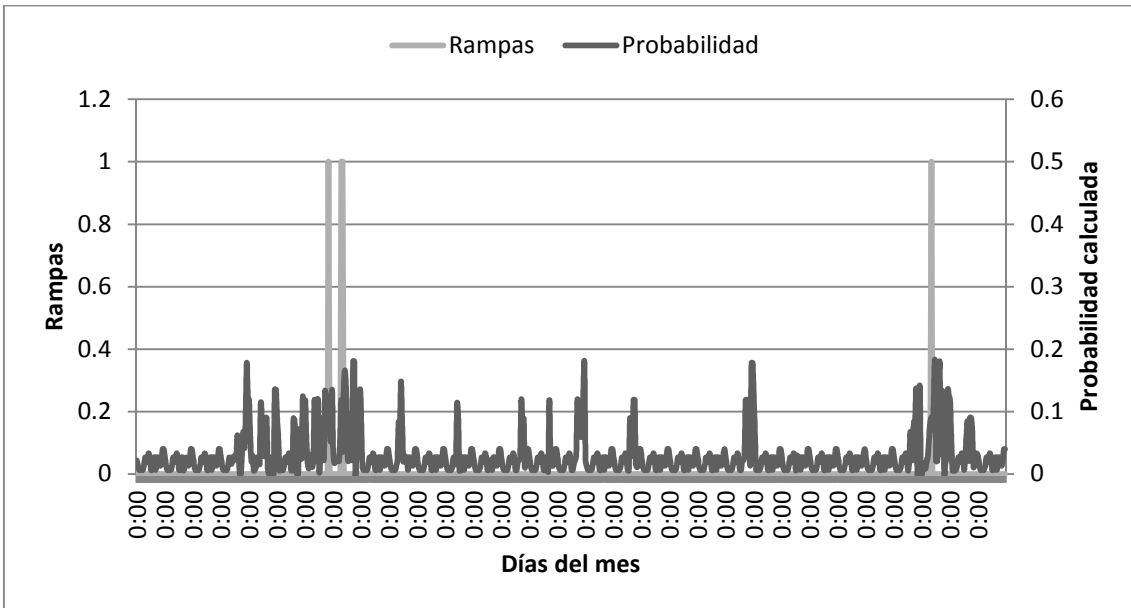


Figura 40. Probabilidades de ocurrencia de rampas mayores a 3 [MW] para tiempo de anticipación de tres horas y rampas realmente ocurridas

En los tres casos se observa que en la ventana de tiempo cercana a los momentos en que ocurren las rampas la probabilidad de ocurrencia tiende a aumentar. Esto a priori auspicia que en los tres casos se podrá tener una tasa de predicción acertada alta. Sin embargo, dado que en este caso la ocurrencia de rampas es rara, la probabilidad en muchos casos aumenta en instantes de

tiempo donde no se aprecian rampas, por lo que para este caso se espera que la tasa de falsos positivos sea grande.

4.2.2 Sensibilidad respecto al umbral

Para este caso también se realiza un análisis de sensibilidad frente al umbral escogido, bajo las mismas consideraciones tomadas para el caso de rampas de al menos 2 [MW]. Así entonces se obtienen el número de predicciones acertadas, falsos positivos y falsos negativos. Se muestran a continuación los resultados:

- Tiempo de anticipación de una hora

Tabla 23. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de una hora, para rampas de magnitud mayor a 3 [MW]

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,10	53	3	1	13,250	0,750	0,250
0,11	46	3	1	11,500	0,750	0,250
0,12	46	3	1	11,500	0,750	0,250
0,13	37	4	0	9,250	1,000	0,000
0,14	25	4	0	6,250	1,000	0,000
0,15	25	4	0	6,250	1,000	0,000
0,16	24	4	0	6,000	1,000	0,000
0,17	18	4	0	4,500	1,000	0,000
0,18	17	4	0	4,250	1,000	0,000
0,19	10	4	0	2,500	1,000	0,000
0,20	10	4	0	2,500	1,000	0,000
0,21	3	4	0	0,750	1,000	0,000
0,22	2	4	0	0,500	1,000	0,000
0,23	2	4	0	0,500	1,000	0,000
0,24	2	4	0	0,500	1,000	0,000
0,25	1	4	0	0,250	1,000	0,000
0,26	1	4	0	0,250	1,000	0,000
0,27	1	4	0	0,250	1,000	0,000
0,28	1	4	0	0,250	1,000	0,000
0,29	1	4	0	0,250	1,000	0,000
0,30	1	4	0	0,250	1,000	0,000

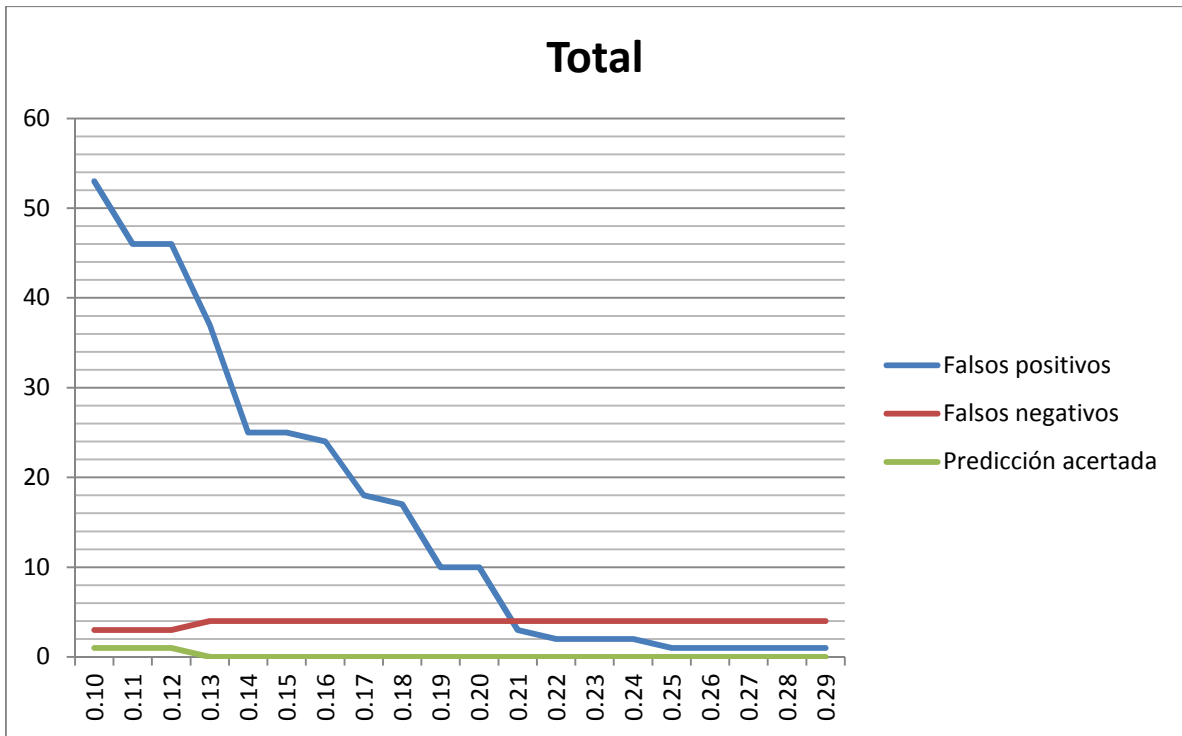


Figura 41. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora y rampas de al menos 3 [MW]

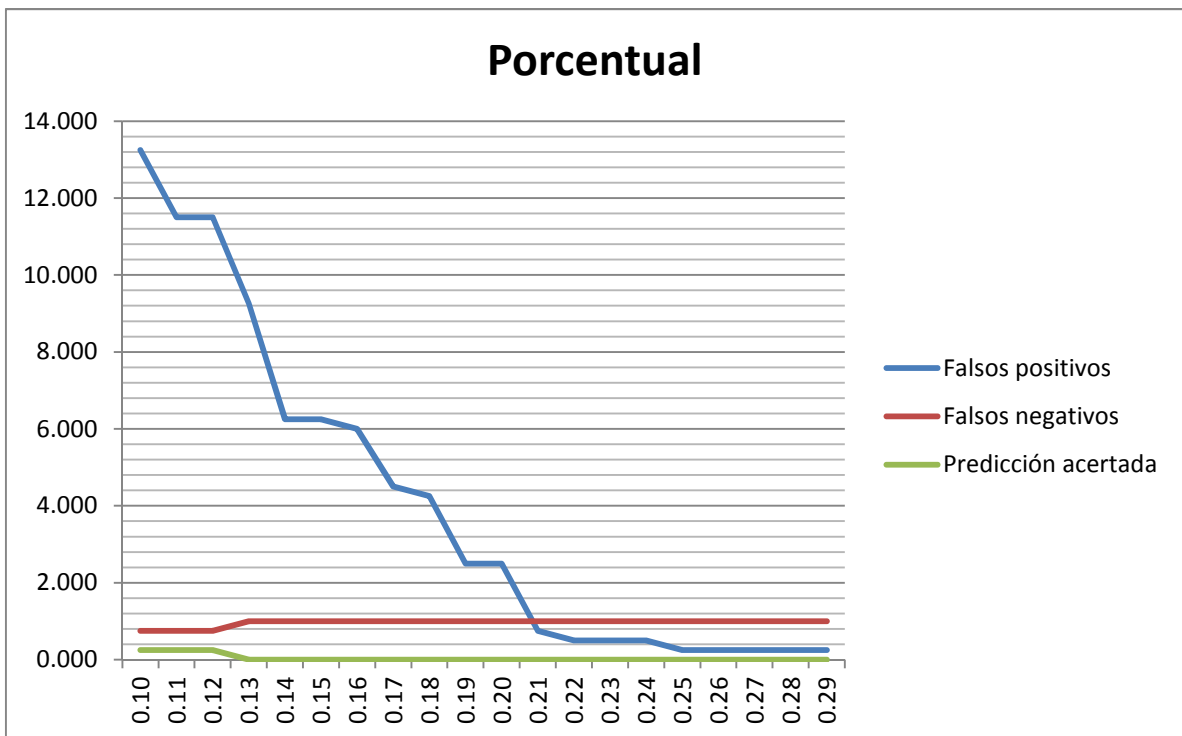


Figura 42. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de una hora y rampas de al menos 3 [MW]. Debe considerarse el 100% como las rampas de esta magnitud ocurridas en enero, que fueron en total 4

- **Tiempo de anticipación de dos horas**

Tabla 24. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de dos horas, para rampas de magnitud mayor a 3 [MW]

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,10	46	2	2	11,500	0,500	0,500
0,11	40	3	1	10,000	0,750	0,250
0,12	38	4	0	9,500	1,000	0,000
0,13	27	4	0	6,750	1,000	0,000
0,14	20	4	0	5,000	1,000	0,000
0,15	20	4	0	5,000	1,000	0,000
0,16	19	4	0	4,750	1,000	0,000
0,17	13	4	0	3,250	1,000	0,000
0,18	13	4	0	3,250	1,000	0,000
0,19	5	4	0	1,250	1,000	0,000
0,20	5	4	0	1,250	1,000	0,000
0,21	5	4	0	1,250	1,000	0,000
0,22	5	4	0	1,250	1,000	0,000
0,23	2	4	0	0,500	1,000	0,000
0,24	2	4	0	0,500	1,000	0,000
0,25	1	4	0	0,250	1,000	0,000
0,26	1	4	0	0,250	1,000	0,000
0,27	1	4	0	0,250	1,000	0,000
0,28	0	4	0	0,000	1,000	0,000
0,29	0	4	0	0,000	1,000	0,000
0,30	0	4	0	0,000	1,000	0,000

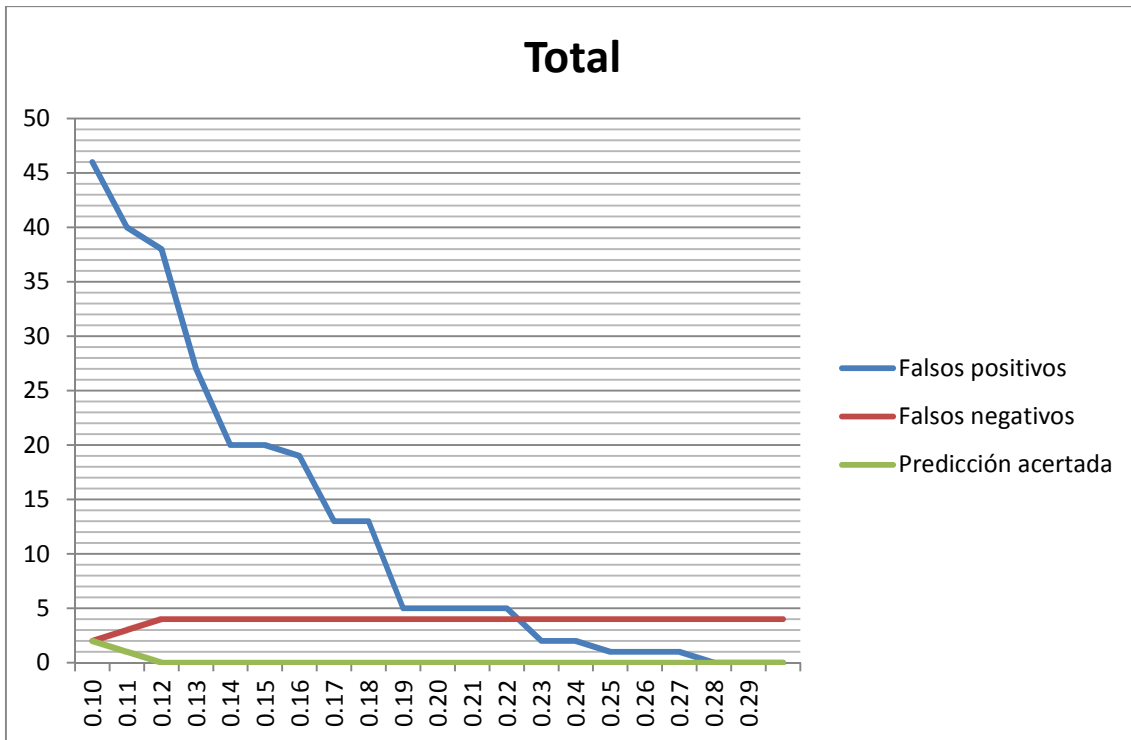


Figura 43. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas y rampas de al menos 3 [MW]

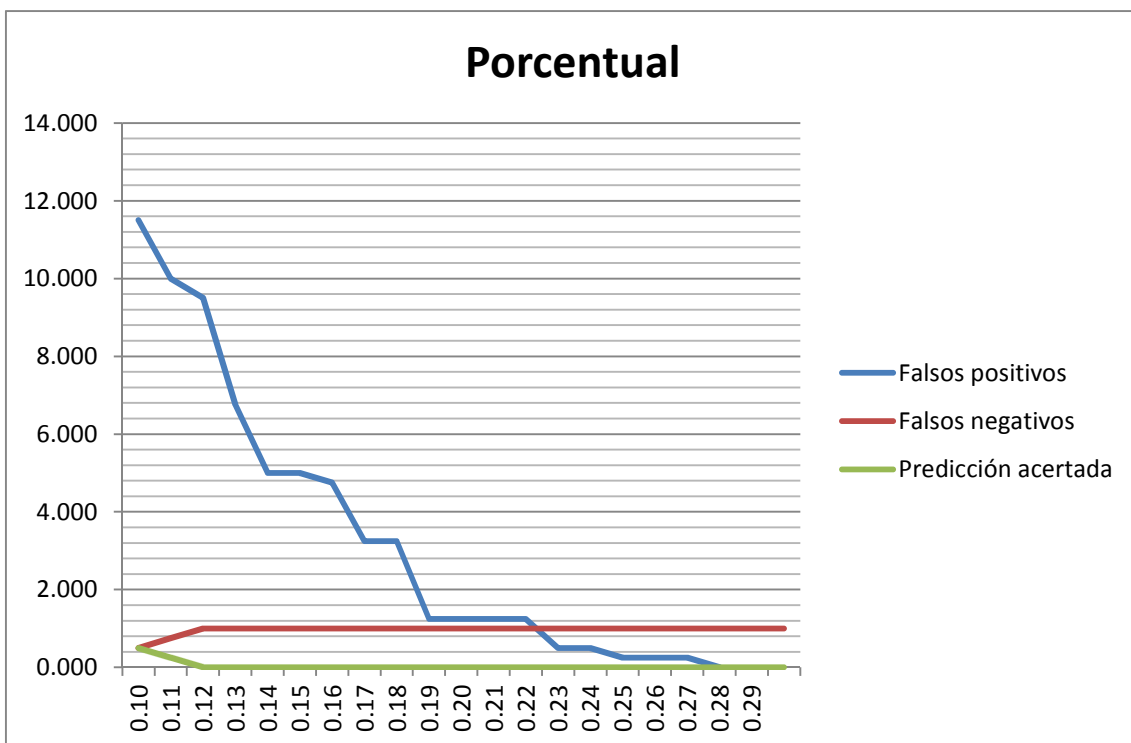


Figura 44. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de dos horas y rampas de al menos 3 [MW]. Debe considerarse el 100% como las rampas de esta magnitud ocurridas

- **Tiempo de anticipación de tres horas**

Tabla 25. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para un tiempo de anticipación de tres horas, para rampas de magnitud mayor a 3 [MW]

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,10	44	3	1	11,000	0,750	0,250
0,11	43	3	1	10,750	0,750	0,250
0,12	27	4	0	6,750	1,000	0,000
0,13	21	4	0	5,250	1,000	0,000
0,14	11	4	0	2,750	1,000	0,000
0,15	9	4	0	2,250	1,000	0,000
0,16	9	4	0	2,250	1,000	0,000
0,17	8	4	0	2,000	1,000	0,000
0,18	4	4	0	1,000	1,000	0,000
0,19	0	4	0	0,000	1,000	0,000
0,20	0	4	0	0,000	1,000	0,000
0,21	0	4	0	0,000	1,000	0,000
0,22	0	4	0	0,000	1,000	0,000
0,23	0	4	0	0,000	1,000	0,000
0,24	0	4	0	0,000	1,000	0,000
0,25	0	4	0	0,000	1,000	0,000
0,26	0	4	0	0,000	1,000	0,000
0,27	0	4	0	0,000	1,000	0,000
0,28	0	4	0	0,000	1,000	0,000
0,29	0	4	0	0,000	1,000	0,000
0,30	0	4	0	0,000	1,000	0,000

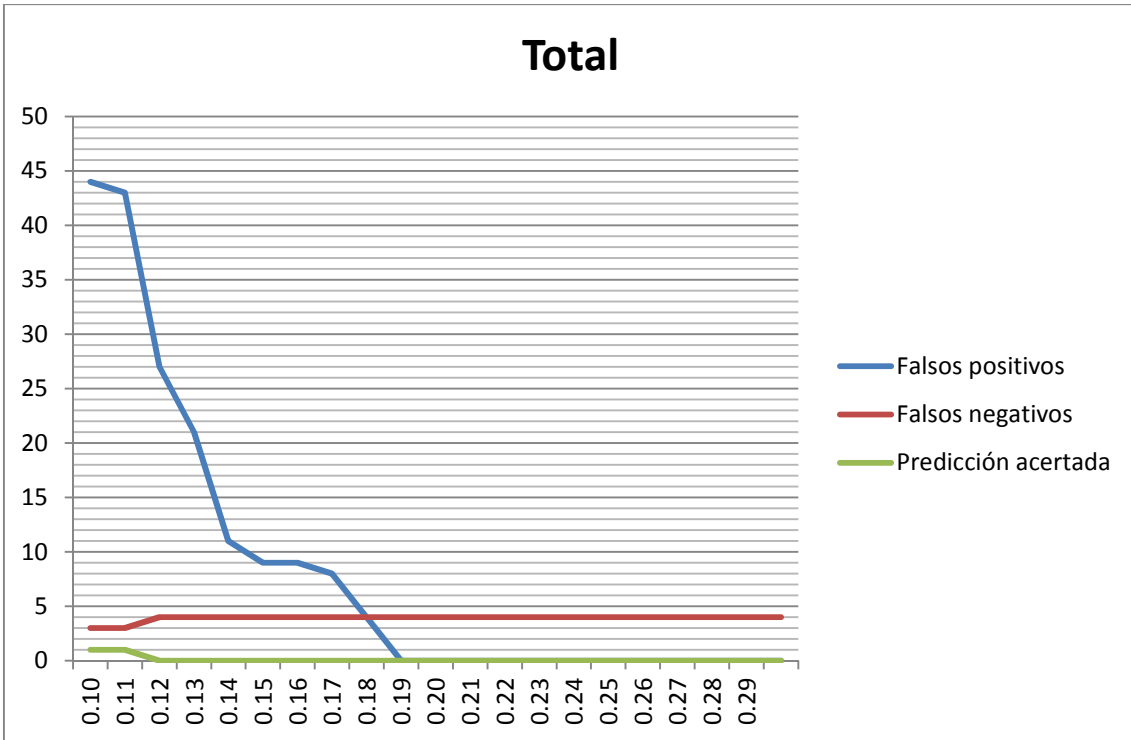


Figura 45. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas y rampas de al menos 3 [MW]

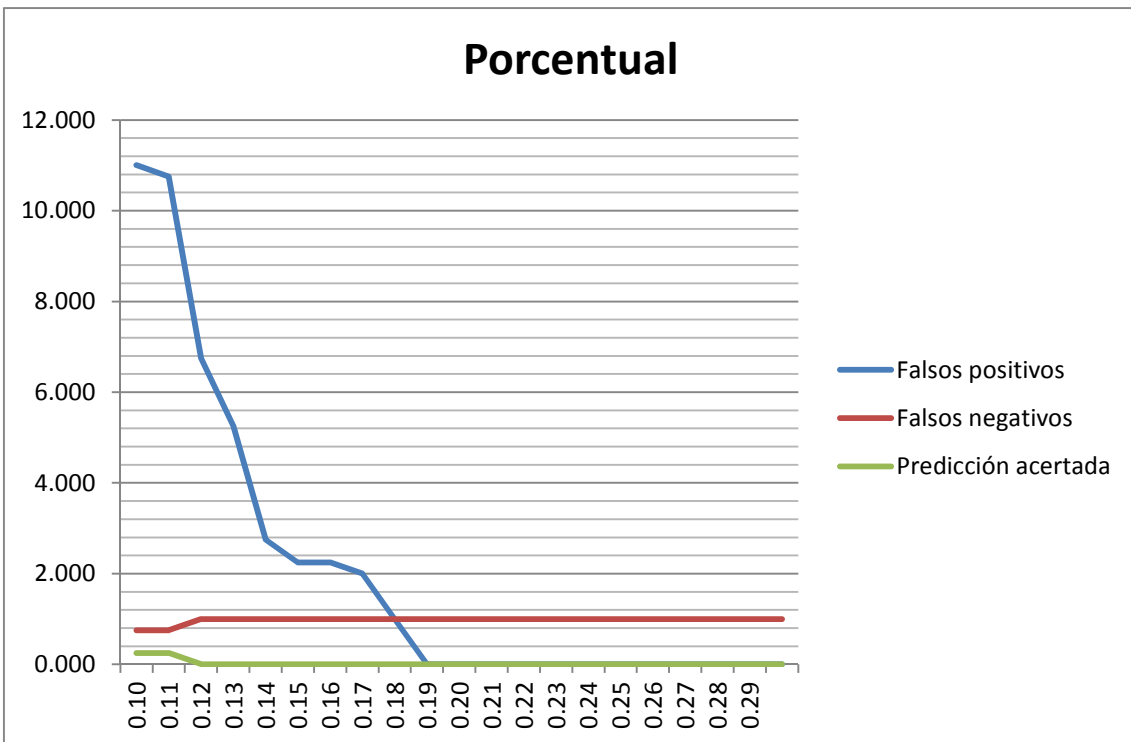


Figura 46. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para un tiempo de anticipación de tres horas y rampas de al menos 3 [MW]. Debe considerarse el 100% como las rampas de esta magnitud ocurrida

En este caso se vuelve a apreciar una alta sensibilidad frente al umbral escogido, viéndose afectado sobre todo la tasa de falsos positivos, la cual decae prácticamente de forma exponencial al aumentar el umbral escogido. Sin embargo la tasa de predicción acertada no es buena. En el mejor de los casos, empleando un tiempo de anticipación de una hora, la tasa de predicción acertada es de un 75% como máximo, decayendo al aumentar el umbral. Esto se explica ya que en el mes de enero las rampas de magnitud mayor o igual a 3 [MW] fueron solo cuatro, es decir, durante enero fue un evento muy raro, por lo que es mucho más difícil poder realizar la predicción con éxito.

Se obtuvo también los puntajes asociados a la función F_{score} para cada tiempo de anticipación empleado a fin de realizar una comparación. Los resultados se muestran en la

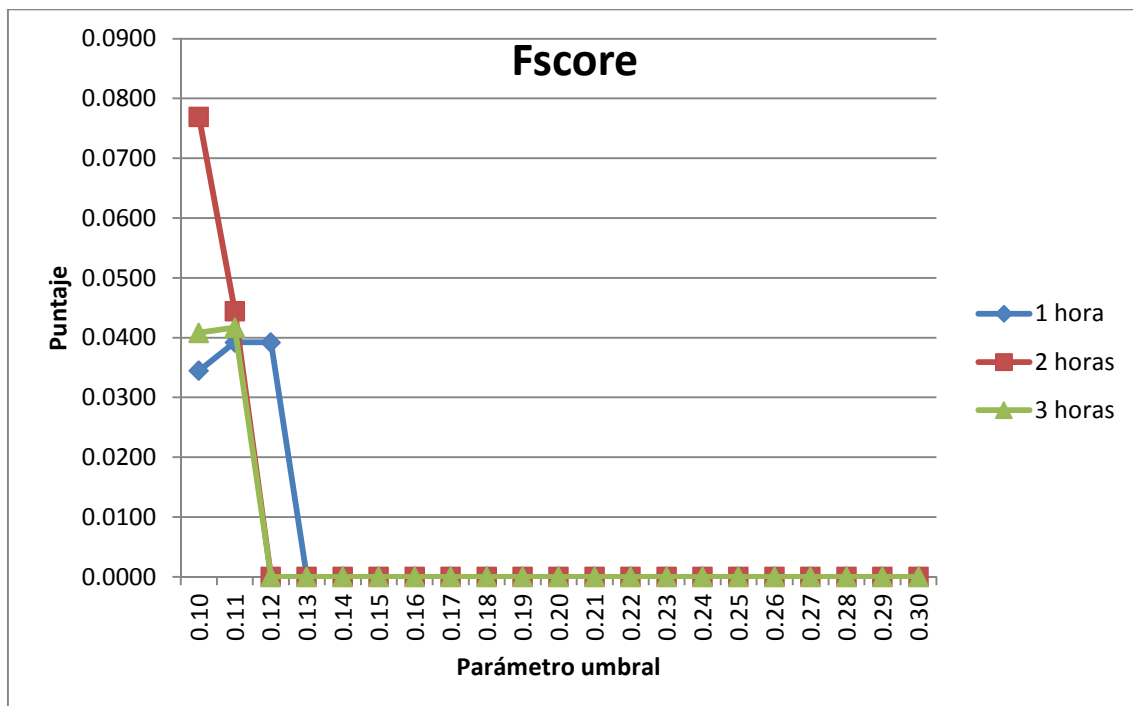


Figura 47. Puntajes asociados a la función Fscore para los tres tiempos de anticipación usados

Como se ve, en los tres casos el puntaje es mayor para parámetros pequeños. Sin embargo el valor del puntaje cae a cero al aumentar un poco el parámetro. Esto se da porque en caso de pronosticar eventos muy raros (como fueron las rampas de magnitud mayor o igual a 3 [MW] en enero) la eficacia del método no es buena.

4.2.3 Predicción en una ventana horaria

También se estudia el caso de predecir rampas en una ventana horaria de dos horas para las rampas de magnitud igual o mayor a 3 [MW], tal como se realizó anteriormente. Los resultados son mostrados a continuación:

Tabla 26. Verdaderos positivos, falsos positivos y falsos negativos obtenidos para rampas de al menos 3 [MW], prediciendo en base a una probabilidad acumulada en una ventana de dos horas

Parámetro umbral	Totales			Porcentual		
	Falsos positivos	Falsos negativos	Predicción acertada	Falsos positivos	Falsos negativos	Predicción acertada
0,10	105	0	4	26,250	0,000	1,000
0,11	97	0	4	24,250	0,000	1,000
0,12	90	0	4	22,500	0,000	1,000
0,13	82	0	4	20,500	0,000	1,000
0,14	78	1	3	19,500	0,250	0,750
0,15	66	1	3	16,500	0,250	0,750
0,16	62	2	2	15,500	0,500	0,500
0,17	57	3	1	14,250	0,750	0,250
0,18	43	4	0	10,750	1,000	0,000
0,19	35	4	0	8,750	1,000	0,000
0,20	31	4	0	7,750	1,000	0,000
0,21	26	4	0	6,500	1,000	0,000
0,22	21	4	0	5,250	1,000	0,000
0,23	19	4	0	4,750	1,000	0,000
0,24	16	4	0	4,000	1,000	0,000
0,25	12	4	0	3,000	1,000	0,000
0,26	10	4	0	2,500	1,000	0,000
0,27	9	4	0	2,250	1,000	0,000
0,28	8	4	0	2,000	1,000	0,000
0,29	8	4	0	2,000	1,000	0,000
0,30	5	4	0	1,250	1,000	0,000

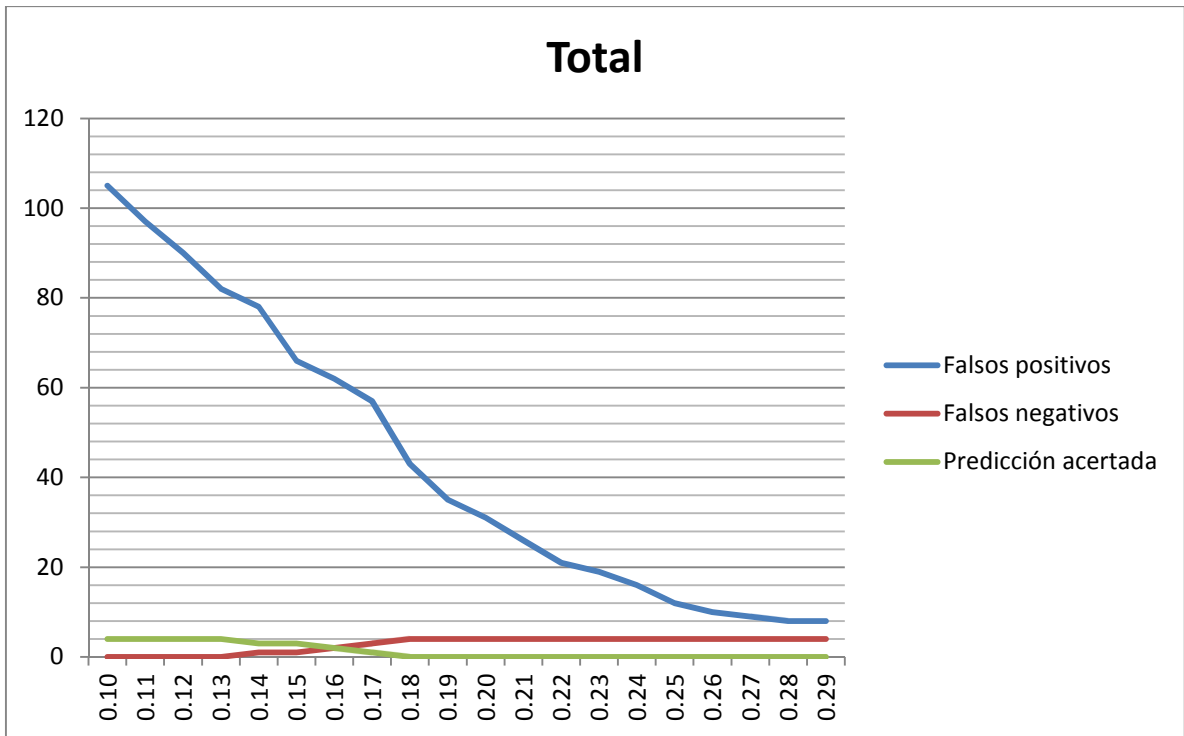


Figura 48. Gráfica de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas, considerando rampas de magnitud de al menos 3 [MW]

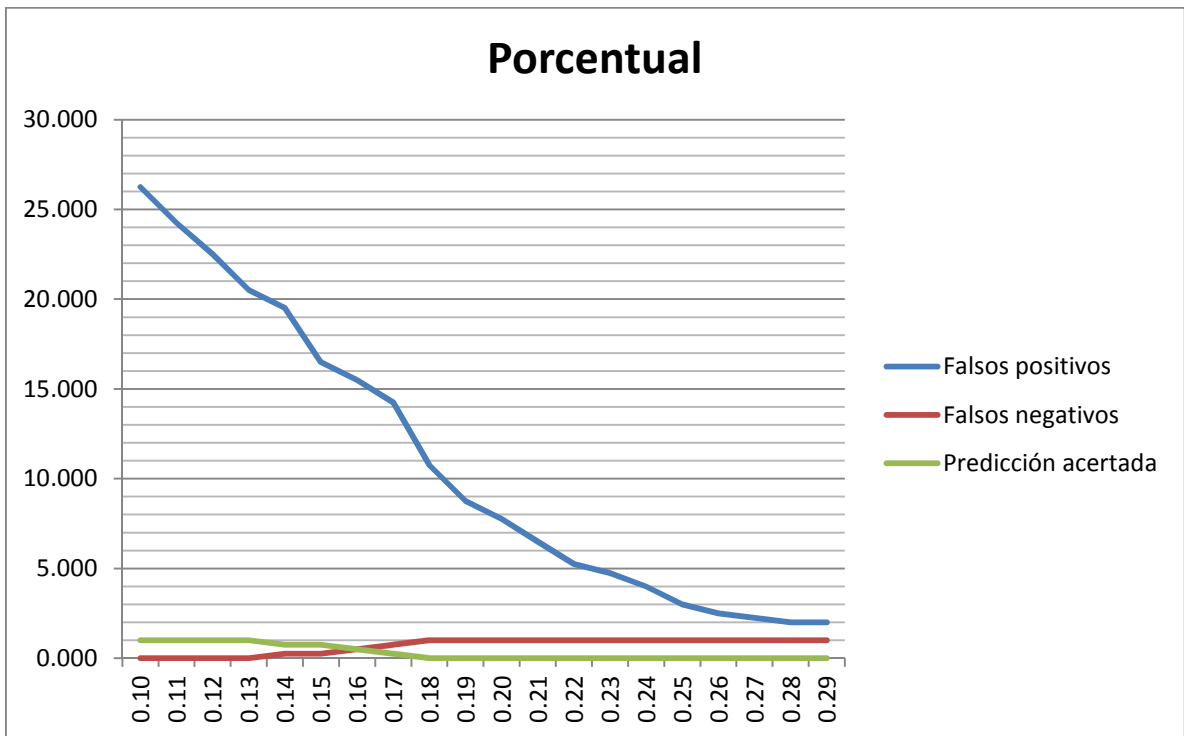


Figura 49. Gráfica porcentual de verdaderos positivos, falsos positivos y falsos negativos en función del parámetro umbral para probabilidad acumulada en ventana de dos horas, considerando rampas de magnitud de al menos 3 [MW]. El 100% corresponde al número de ventanas de dos horas en el que si hubieron rampas, 4 para este caso

Como se ve de la tabla y las figuras, la tasa de predicción mejora en este caso, pudiéndose predecir para umbrales bajos el 100% de las rampas. Sin embargo se sigue teniendo el problema de los falsos positivos, el cual es más notorio para este caso.

Finalmente se calcula el puntaje la función F_{score} para este caso y se compara con los casos anteriores. Esto se muestra en la Figura 50.

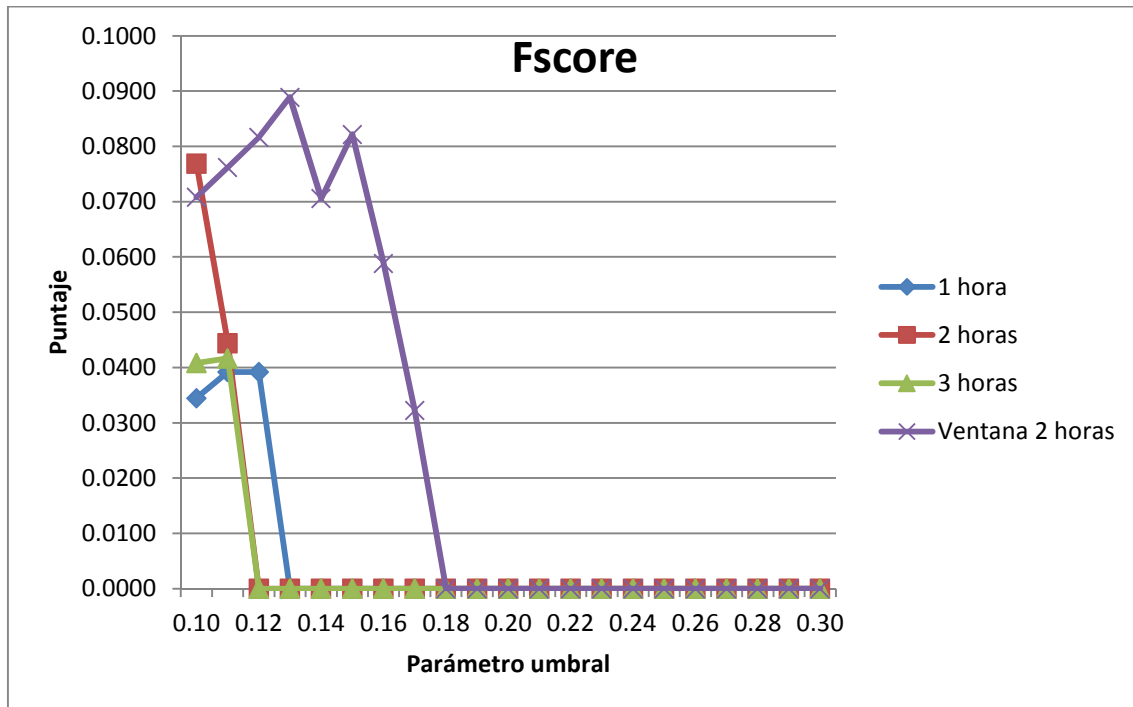


Figura 50. Función Fscore para predicciones con una, dos y tres horas de anticipación, y empleando una ventana de predicción de dos horas, considerando rampas de al menos 3 [MW] de magnitud

Como se aprecia, para umbrales pequeños el puntaje para la predicción en una ventana horaria es mejor que los otros tres casos, pero no es demasiado mejor. Mejora un poco debido a que es capaz de pronosticar correctamente todas las rampas para umbrales pequeños, pero para estos mismos umbrales la tasa de falsos positivos se eleva mucho, o que a su vez le resta puntaje y el resultado final es un puntaje que no es mucho mejor que los que se obtuvieron para los casos previos. También se ve que el puntaje decae a cero para un umbral igual a $P_{umbral}=0,18$. Esto porque por sobre ese umbral el predictor no es capaz de realizar un pronóstico acertado para ninguna rampa, teniéndose una tasa de predicción acertada nula.

4.3 Comparación con modelo AR

Finalmente se procede a obtener los resultados mediante el modelo autorregresivo. Primero se obtuvo la curva de potencia pronosticada para el mes de enero y se contrastó con la potencia real. Este resultado se aprecia en la Figura 51:

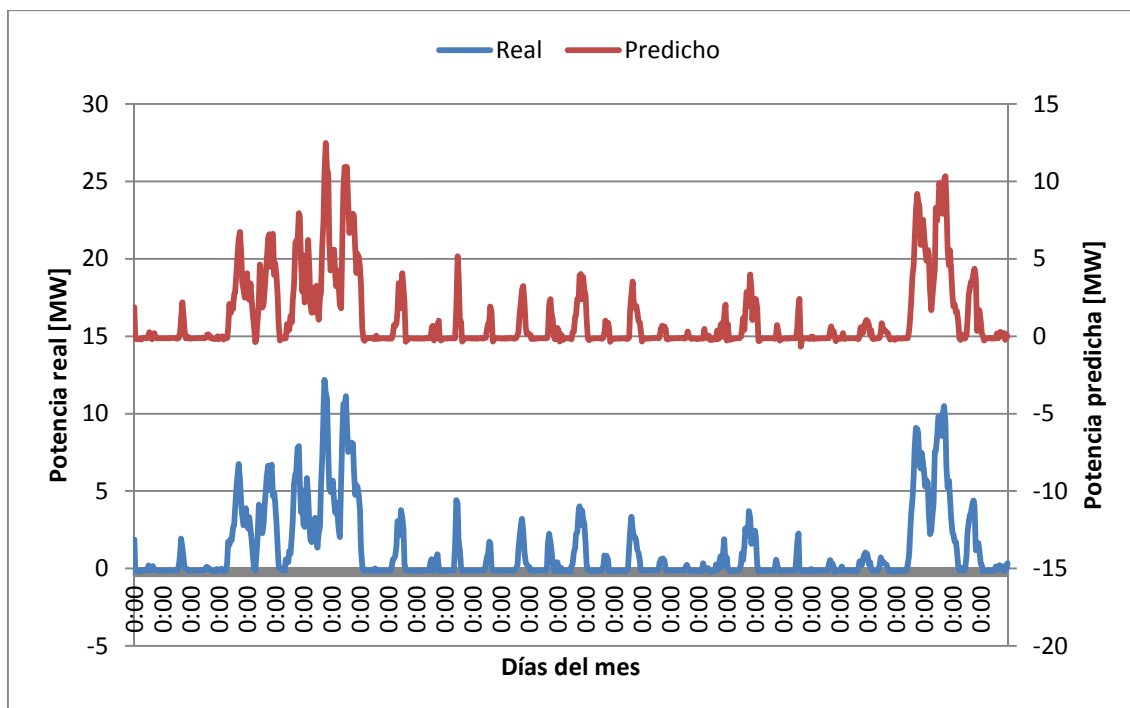


Figura 51. Potencia real y pronosticada para el mes de enero empleando un modelo AR(3)

Como se aprecia, la predicción es bastante buena ya que la curva de potencia pronosticada es bastante parecida a la curva real. Sin embargo esta es una estimación de la potencia generada, pero hay que ver cómo se comporta en el caso de interés, que es la predicción de rampas. Así se procede a obtener el número de predicciones acertadas, falsos positivos y falsos negativos al emplear este método, para finalmente obtener el puntaje asociado a la función F_{score} que obtiene este método. Los resultados se muestran a continuación:

Tabla 27. Predicciones acertadas, falsos positivos y falsos negativos obtenidos con el modelo AR(3)

Predicciones acertadas	43
Falsos positivos	73
Falsos negativos	43

Luego, la Precisión, el Recall y el F_{score} se muestran en la Tabla 28:

Tabla 28. Precisión, Recall y puntaje Fscore obtenidos con el modelo AR(3)

Precision	0,37068
Recall	0,46739
Fscore	0,41346

Este resultado se compara con el mejor caso obtenido empleando el modelo propuesto en base redes Bayesianas, el cual corresponde a la predicción en una ventana horaria de dos horas. Los resultados comparados se muestran en la Figura 52:

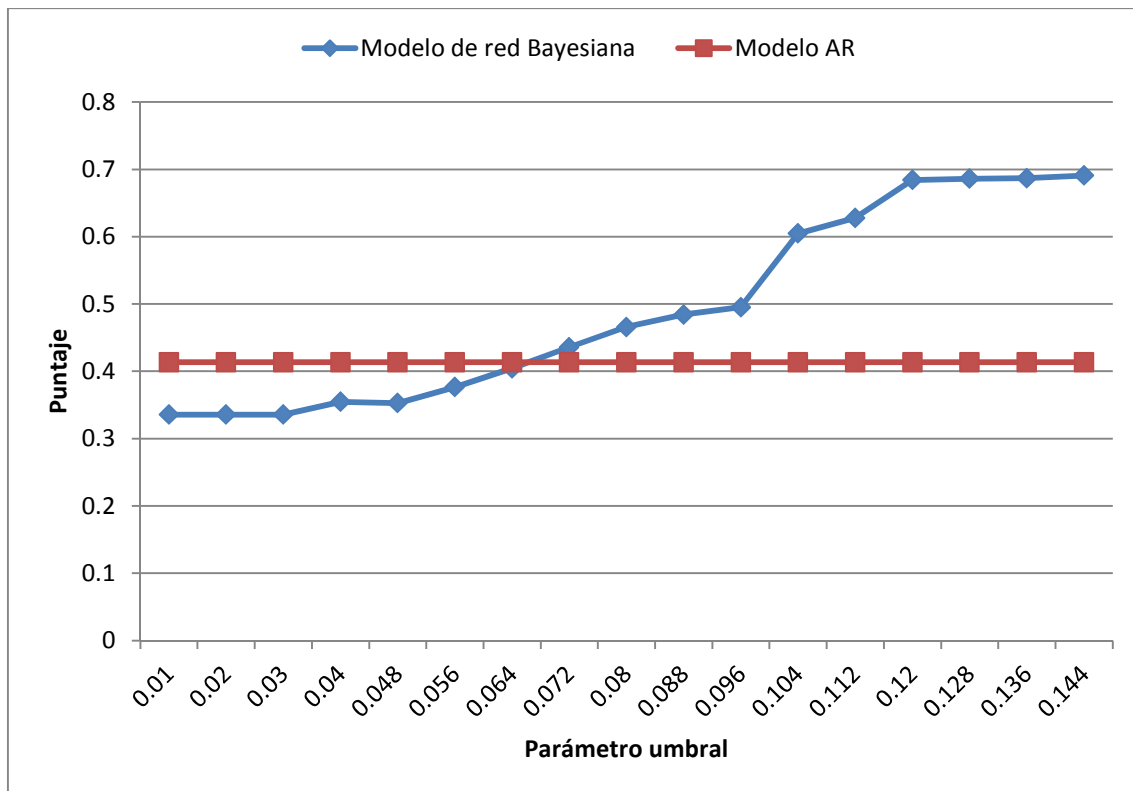


Figura 52. Puntajes del modelo de red Bayesiana para predicción en ventana de dos horas y para el modelo AR. Debe tenerse en cuenta que el puntaje asociado al modelo AR no depende del umbral, es un valor fijo, por ello en la gráfica se aprecia paralela al eje X.

Se aprecia entonces que para valores más bajos del umbral, menores a 0,06, el modelo AR presenta un puntaje mayor. Ahora, hay que tener en cuenta que el mejor desempeño del modelo en base a la red Bayesiana construida se obtenía para valores de parámetros mayores, mayor a 0,1, y en este caso se aprecia que para esos valores el puntaje asociado al método de la red Bayesiana es mayor.

Entonces, como se aprecia, la red Bayesiana es mejor en términos de predecir el momento en que ocurrirá la rampa, en ese sentido es mejor que el modelo autorregresivo. Sin embargo, tal como muestra la Figura 51, el modelo autorregresivo es capaz de predecir con bastante exactitud la magnitud de la potencia, cosa que el modelo en base a la red Bayesiana no es capaz de hacer. Dado que por un lado la red Bayesiana no es capaz de predecir la magnitud ni el sentido de la rampa (solo predice la ocurrencia de este fenómeno) pero si puede predecir de manera más exacta el momento en el que ocurrirá la rampa, y por otro lado el modelo autorregresivo no es tan bueno prediciendo el momento de ocurrencia de las rampas, pero si es bueno para predecir la magnitud que tendrá la potencia que se puede generar en base al viento, una buena solución (más completa por lo demás, capaz de entregar más información) para atacar el desafío asociado a las rampas es emplear un método combinado que se base en ambos métodos. Esto garantizaría una predicción mucho más precisa y con más información relevante (no solo el momento en el que ocurrirá una rampa, sino que también su magnitud y sentido).

5 Conclusiones y trabajo a futuro

Finalmente se presentan las conclusiones y el trabajo futuro

5.1 Conclusiones

La dependencia a nivel país que existe de los combustibles fósiles extranjeros para la generación de energía eléctrica, sumado a los problemas ambientales que este tipo de generación acarrea, ha puesto en la palestra la necesidad de contar con fuentes de energía alternativas, posibles de explotar dentro del mismo territorio nacional y que puedan generar energía de manera limpia y sustentable. Dentro de las posibles opciones una que ha tomado fuerza es el recurso del viento. Pese a ser un recurso con gran potencial, presenta algunos inconvenientes tales como su aleatoriedad, y es por lo mismo que es de suma importancia contar con herramientas capaces de predecir el comportamiento del viento para que, en conjunto con sistemas de control, garanticen una generación segura para los sistemas eléctricos.

Uno de los fenómenos limitantes para una mayor penetración de la energía eólica en los sistemas eléctricos es que el viento presenta fenómenos de rampas, bruscas subidas o caídas de la potencia asociada al viento en cortos periodos de tiempo. Sin un buen sistema de predicción de estos fenómenos, el cuál puede ser una entrada de un sistema de control que tome medidas para aminorar los efectos, una mayor penetración de la energía eólica no es factible ya que pone en riesgo la estabilidad del sistema eléctrico.

En términos generales se cumplieron los objetivos de este trabajo. Se implementó un sistema de predicción y se probó su efectividad empleando tiempos de anticipación de una, dos y tres horas, corroborando su eficacia para cada una de estos tres intervalos. Este sistema se creó utilizando un enfoque probabilístico-estadístico, en base a mediciones de las que se disponía. Además se estudió el comportamiento del sistema predictor para dos magnitudes de rampas.

Los resultados obtenidos en primera instancia no fueron los esperados, ya que si bien se lograban obtener buenas tasas de predicción, estas estaban irremediablemente asociadas a altas tasas de falsos positivos. Usar criterios más estrictos para predecir la aparición de rampas disminuye la tasa de falsos positivos, pero también disminuye la tasa de predicciones acertadas. Para mejorar los resultados se probó realizar la predicción en una ventana horaria en vez de realizarla para una hora en específico. En conclusión, el sistema de predicción implementado presenta resultados mucho mejores al intentar predecir la ocurrencia de una ventana horaria, empleándose para esto la probabilidad de aparición de rampas acumulada para ese periodo.

Se corroboró que en caso de predecir eventos muy raros el sistema no era muy eficaz. Esto se vio al emplear rampas de magnitud 3 [MW] como parte del modelo (en vez de las rampas de 2 [MW] consideradas inicialmente) las cuales, al ser de mayor potencia, eran a su vez mas escasas en el mes correspondiente al periodo de validación. Para este caso los resultados no fueron buenos ni siquiera al emplear la predicción en una ventana horaria. La causa de esto es que dado que estos eventos son más raros, estadísticamente

son más escasos, por lo que la probabilidad calculada en base al modelo será menor y el pronóstico correcto se hace más difícil.

Cuando se comparó el resultado obtenido con el modelo basado en redes Bayesianas y el modelo autorregresivo se observó que el primero tenía mejores resultados en cuanto a la predicción de la ocurrencia de los eventos. Sin embargo el modelo autorregresivo es bueno para extrapolar la serie de potencia y poder pronosticar los valores de potencia eólica generada para el futuro. Entonces, dado las ventajas y debilidades de cada modelo, un modelo conjunto podría dar resultados mucho más exactos.

Dado que la manera de entrenar el modelo es estadísticamente en mediante bases de datos, contar con una buena base de datos es imprescindible. Ya que los parámetros se calcularon estadísticamente en base a conteos de sucesos sobre un total, el hecho tener bases de datos más grandes da la ventaja de conocer más eventos sucedidos en ciertas condiciones, aparte de dar una idea del comportamiento de la variable, que si bien es aleatoria, tiene a presentar ciertos comportamientos. Bases de datos más grandes y mejores garantizan una mejor estimación de los parámetros necesarios en la creación del modelo. En el caso trabajado en esta memoria, los resultados se podrían mejorar sustancialmente si se hubiese dispuesto de bases de datos más amplias.

El enfoque probabilístico empleado en la creación del sistema de predicción tiene gran potencial en la predicción de ocurrencia de fenómenos aleatorios, pero para tener buenos resultados empleando este enfoque es imprescindible contar con amplias y buenas bases de datos, a fin de poder caracterizar de mejor manera la variable aleatoria en cuestión.

Finalmente, el sistema de predicción desarrollado no sería difícil de implementar y ser incluido dentro de un esquema de control. Dado que todos los parámetros necesarios son calculados con anterioridad, el cálculo de la probabilidad sólo depende de la hora del día para el que se realiza la predicción y la potencia que se generó, por lo que el cálculo asociado se reduce a operaciones matemáticas simples, y luego la toma de decisión en base a un umbral preestablecido.

A pesar de haber obtenido buenos resultados, estos se pueden mejorar empleando otras entradas o ampliando el modelo. El conocimiento de la variables físicas que influyen en el fenómeno puede ser una entrada importante dentro del modelo. Generar un modelo híbrido, que mezcle metodologías de un modelo estadístico con un modelo físico, podría llevar a obtener una herramienta de predicción muy exacta. Dentro de un gran modelo híbrido el método planteado puede representar la parte estadística que ayude a obtener mejores resultados.

Debe tomarse en cuenta que este trabajo se basó en bases de datos de la potencia generada por la central Canela I. Para centrales ubicadas en otras locaciones los parámetros serán distintos dado que el viento presenta comportamientos que dependen a zona geográfica.

5.2 Trabajo futuro

En base a lo trabajado se proponen los siguientes trabajos a futuro:

Entrenar el método empleando bases de datos más grandes a fin de corroborar la posible mejora de los resultados.

Probar la eficacia del método empleando datos de otras centrales. Idealmente usando bases de datos correspondientes a centrales ubicadas en lugares diferentes, como puede ser la zona centro o sur del país, a fin de probar que tan efectivo es aplicado a centrales ubicadas en zonas en las que el viento pueda tener un comportamiento diferente.

Desarrollar el método que incluya la estacionalidad que presenta el viento a lo largo del año. Para esto se proponen dos alternativas: asociar varios meses a ciertas estaciones del año (meses de verano, otoño, invierno y primavera), calculando diferentes parámetros para el modelo según la estación del año; o bien hacer cálculos de parámetros para cada mes del año y realizar las predicciones según el mes del año en el que se esté. Para esto es recomendable contar con bases de datos de varios años para englobar de mejor manera el comportamiento de las variables.

Ampliar el desarrollo del método de manera que en la predicción de la ocurrencia de rampas se incluya la magnitud y el sentido de estas, no solo la ocurrencia de éstas.

Probar el desempeño del método en predicción asociada a otras fuentes de generación intermitente, como por ejemplo la energía solar.

6 Bibliografía

- [1] Comisión Nacional de Energías. (2014, Abril) Sistema interconectado central (SIC). [Online]. <http://www.cne.cl/energias/electricidad/sistemas-electricos/343-sic>
- [2] Endesa. (Abril, 2014) Parque Eólico Canela. [Online]. www.endesa.cl/es/conocenos/nuestronegocio/canela/Documents/index.htm.
- [3] K. Torres. (2011, Noviembre) Parque Eólico Lebu. [Online]. es.scribd.com/doc/73075495/Parque-Eolico-Lebu.
- [4] SN Power. (2014, Abril) Parque Eólico Totoral. [Online]. [SNPower. www.snpower.cl/proyectos-y-centrales/parque-eolico-total/](http://www.snpower.cl/proyectos-y-centrales/parque-eolico-total/).
- [5] GDF Suez. (2014, Abril) Parque Eólico Monte Redondo. [Online]. <http://www.eolicamonteredondo.cl/el-parque-y-su-funcionamiento/>
- [6] Sociedad Nacional de Minería SONAMI. (2012, Septiembre) SONAMI. [Online]. <http://www.sonami.cl/files/presentaciones/543/08.-%20PARQUE%20EOLICO%20PUNTA%20COLORADA.pdf>
- [7] Gobierno de Chile. (2011, Octubre) Declaración de Impacto Ambiental "Parque Eólico Ucuquer". [Online]. <http://seia.sea.gob.cl/documentos/documento.php?idDocumento=6207936>
- [8] Barrick. (2012, Agosto) Parque Eólico Punta Colorada. [Online]. barricksudamerica.com/parque-eolico-punta-colorada/barrick/2012-11-28/101849.html
- [9] Gobierno de Chile. (2014, Julio) Declaración de Impacto Ambiental Proyecto "Parque Eólico Negrete". [Online]. http://seia.sea.gob.cl/expediente/ficha/fichaPrincipal.php?modo=ficha&id_expediente=2128951039
- [10] Electro Industrias. (2012, Agosto) Parque Eólico El Arrayás. [Online]. <http://www.emb.cl/electroindustria/articulo.mvc?xid=1899>
- [11] Grupo Editorial Editec. (2014, Octubre) Electricidad, la revista energética de Chile. [Online]. <http://www.revistaei.cl/tag/parque-eolico-los-cururos/>
- [12] SysteP Ingeniería y Diseños S.A. (2014, Enero) Estadísticas. [Online]. http://www.systeP.cl/?page_id=1203
- [13] CDEC-SIC. (2014, Abril) Operación Real. [Online]. <http://www.cdecsic.cl/informes-y-documentos/fichas/operacion-real/>
- [14] A. Cárdenas, Metodología para el Cálculo de Reserva en Giro y Desconexión de Cargas en Sistemas Eléctricos, 2007, Memoria para

optar al título de Ingeniero Civil Electricista, Universidad de Chile.

- [15] A. Botterud and W. Jianhung, "Wind Power Forecasting in U.S. Electrical Markets," *Electricity Journal*, vol. 23, no. 3, pp. 71-82, Abril 2010.
- [16] José Santamarta. (2009, Agosto) Revista Eólica y del Vehículo Eléctrico (reve). [Online]. <http://www.evwind.com/2009/08/15/energia-eolica-en-dinamarca-las-empresas-eolicas-danesas-exportaron-por-5-700-millones-de-euros-en-2008-por-jose-santamarta/>
- [17] ABO WIND. (2012, Enero) Alemania. [Online]. <http://www.abowind.com/es/energia-eolica/alemania.html>
- [18] M. Lange and U. Focken. (2005, Enero) Research Gate. [Online]. http://www.researchgate.net/publication/237677945_State-of-the-Art-in-Wind-Power-Prediction-in-Germany-and-International-Developments
- [19] B. Greaves, J. Collins, J. Parker, and A. Tindal. (2009) Temporal Forecast Uncertainty for Ramp Events. [Online]. <http://www.dis.anl.gov/pubs/69166.pdf>
- [20] Z. Haiyang and A. Kusiak. (2009) Prediction of Wind Farm Power Ramp Rates: A Data-Mining Approach. [Online]. <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=F059A2846BEB18944A852D2B683D01B0?doi=10.1.1.150.730&rep=rep1&type=pdf>
- [21] A. Bossavy, R. Girard, and G. Kariniotakis. (2010, Abril) Forecasting Uncertainty Related to Ramps of Wind Power Production. [Online]. http://proceedings.ewea.org/ewec2010/allfiles2/141_EWEC2010presentation.pdf
- [22] T. Ouyang, X. Zha, and L. Qin. (2013, Julio) Scientific Research. [Online]. <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=38042#.VLmtgUeG-5g>
- [23] Y. H. Mi and et al, "Prediction of Wind Power Generation," in *Awareness Science and Technology (iCAST), 3rd International conference*, 2011.
- [24] R. Sevlian and R. Rajagopal, "Wind power ramps: Detection and statistics," in *Power and Energy Society General Meeting, 2012 IEEE*, San Diego, CA, 2012, pp. 1-8.
- [25] H. Mendoza and G. Bautista. (2002, Marzo) Probabilidad y Estadística. [Online]. <http://www.virtual.unal.edu.co/cursos/ciencias/2001065/index.html>
- [26] Vitutor. (2012) Probabilidad Condicionada. [Online]. http://www.vitutor.com/pro/2/a_12.html

- [27] T. M. Mitchell. (2010, Enero) Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression. [Online]. <http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>
- [28] E. Millán, "Redes Bayesianas," Universidad de Málaga, Málaga,.
- [29] C. W. Potter, E. Gritti, and B. Nijsen, "Potential benefits of a dedicated probabilistic rapid ramp event forecast tool," in *Power Systems Conference and Exposition*, Seattle, WA, 2009, pp. 1-5.
- [30] C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang. (2010, Diciembre) A Survey of Power Ramp Forecasting. [Online]. <http://www.dis.anl.gov/pubs/69166.pdf>
- [31] H.L. Hurd. (2002, Febrero) MATLAB scripts and programs. [Online]. <http://www.stat.unc.edu/faculty/hurd/stat185Data/progdoc.html>

7 Anexos

7.1 Anexo A

Programa implementado para el cálculo de la probabilidad condicionada a la hora del día:

```
% datos(:,1) hora
% datos(:,2) potencia
% datos(:,3) rampas

cont=zeros(1,24);

for k=0:1:23
for j=1:1:length(datos)
if datos(j,1)==k && datos(j,3)==1

cont(k+1)=cont(k+1)+1;

end
end
end

cont=cont';
```


7.2 Anexo B

Programa implementado para el cálculo de la probabilidad condicionada a la potencia generada en el instante anterior correspondiente:

```
% datos(:,1) hora
% datos(:,2) potencia
% datos(:,3) rampas

cont=zeros(1,18);

for k=1:1:18
for j=2:1:length(datos)
if datos(j-1,2)==k && datos(j,3)==1

cont(k)=cont(k)+1;

end
end
end

cont=cont';
```

7.3 Anexo C

Programa empleado para el cálculo de los coeficientes del modelo AR

```
%ARMA Matlab

function [mbest,minaic,pbest,qbest]=armabat(x,pvec,qvec)
%
% [mbest,minaic,pbest,qbest]=armabat(x,pvec,qvec)
%
% This functions computes ARMA (p,q) models for
% (p,q) in (pvec x qvec ); it returns the best according to AIC
% where AIC has been modified to account for fixed parameters
%
% x = input data
% pvec = vector of p's ; set pvec=[0] for no AR
% qvec = vector of q's; set qvec=0[] for no MA
% mbest = matlab model structure for the best one found using the matlab aic
% the model parameters can be retrieved from mbest
% minaic = value of the aic at the min
% pbest = best value of p found
% qbest = best value of q found

% now estimate ARMA model; ARMAX is AX=BU + Ce so identify phi with A and theta with C
nx=length(x);
np=length(pvec);
nq=length(qvec);
aicsave=-99*ones(np,nq);
fpesave=-99*ones(np,nq);
minaic=1e+6;
for pp=1:np
    p=pvec(pp);
    for qq=1:nq
        q=qvec(qq);
        if p+q ~=0
            orders=[p q];

            m=armax(x,orders); % m is a structure with the model stuff in it
            resids=pe(m,x); % pe returns the prediction errors
            nres=length(resids);
            % next compute the Ljung - Box statistic and P-values
            deltak=floor(nrho/10)+1;
            kvec=[p+q+1:deltak:p+q+1+4*deltak];

            rhores=acf(resids,100);
            nrho=length(rhores);

            for kk=1:5
                Qsum=0;
                for j=2:kvec(kk)+1
                    Qsum=Qsum+(rhores(j).^2)/(nx-j);
                end
                Qsum=nx*(nx-1)*Qsum;
                ljpv(kk)=1-chi2cdf(Qsum,kvec(kk)-p-q); % df=kvec(kk)-p-q
            end
            aicsave(pp,qq)=aic(m);
            fpesave(pp,qq)=fpe(m);
            AICTEST=1;
        end
    end
end
```

```

if AICTEST
    nval = m.EstimationInfo.DataLength;
    vval = m.EstimationInfo.LossFcn;
    totalpars=m.na+m.nc;
    netpars=totalpars-length(m.fix);
    aicval=log(vval)+2*netpars/nval;
    disp(sprintf('AIC : %g NEWAIC : %g',aicsave(pp,qq),aicval));
end
if aicsave(pp,qq) < minaic
    minaic=aicsave(pp,qq); % save the min
    pbest=p;
    qbest=q;
    mbest=m;
end
disp(sprintf('(p,q)=(%d,%d) aic=%g fpe=%g',p,q,aicsave(pp,qq),fpesave(pp,qq)));
QQ=[kvec:ljpgv];
disp(sprintf('Ljung-Box P-values : '));
disp(sprintf(' K=%d P-v=%6.3e \n',QQ(:)));
end
end
end

```

7.4 Anexo D

Programa empleado para realizar el pronóstico de potencia con el modelo AR

```
% Inicialización
estenero=zeros(length(enero),1);
e=zeros(length(enero),1);

estenero(1)=enero(1);
estenero(2)=enero(2);
estenero(3)=enero(3);

% Coeficientes modelo ARMA(1,1)
phi1=1.252;
phi2=-0.2513;
phi3=-0.03395;

% Estimación
for i=4:1:length(enero)
    e(i)=normrnd(0,11); %parte estocástica
    estenero(i)=phi1*enero(i-1)+phi2*enero(i-2)+phi3*enero(i-3)+e(i);
end

for i=4:1:length(estenero)
    estenero(i)=estenero(i);
end

plot(estenero)
```