

TRABAJO DE FIN DE MÁSTER

# Comité de Expertos para combinación de fuentes de predicción.

---

Aplicación en series temporales  
eólicas.

**Yeray Gutiérrez Cedrés**

**Diciembre de 2012**

Tutores:

Mario Hernández Tejera

José Juan Cabrera Hernández

## Agradecimientos

En primer lugar quiero agradecer la predisposición y el apoyo mostrado por mis tutores, Mario y José Juan, que me han facilitado en gran medida la finalización de mis estudios de máster con este trabajo.

También agradecer la colaboración prestada por mi compañero Rafael Nebot, del que se puede aprender mucho, tanto a nivel profesional y académico como a nivel humano.

Dar las gracias también a mi amigo Abby Henríquez, que con su talento natural ha sabido escuchar, entender y proponer soluciones para muchas de las dudas que me han surgido durante el desarrollo de mi trabajo.

Finalmente, agradezco también el apoyo mostrado por mi familia, en especial por Elisa.

## Contenido

1. Introducción .....	1
2. Estado del arte.....	6
2.1 Métodos empíricos estadísticos.....	6
2.2 Regresión cuantílica.....	7
2.2.1 Bosques de regresión cuantílica (Quantile regression forests).....	7
2.2.2 Regresión cuantílica con B-Splines .....	9
2.3 Predicción de distribución .....	10
3. Metodología .....	11
3.1 Predicciones puntuales con modelo meteorológico .....	11
3.2 Método de Remuestreo Adaptado para la obtención de predicciones probabilistas.....	13
3.3 Modelo de Comité de Expertos .....	18
3.3.1 Máquinas de comité.....	18
3.3.2 Aplicación de las máquinas de comité a predicción probabilista.....	22
3.3.3 Método de entrenamiento de la Gating Network .....	24
4. Aplicación experimental .....	27
4.1 Datos de entrada.....	27
4.2 Experimentos realizados .....	30
5. Resultados.....	31
6. Conclusiones y trabajo futuro .....	39
7. Bibliografía.....	42

## 1. Introducción

La concienciación sobre el medio ambiente y las incertidumbres al respecto del suministro de combustibles fósiles están llevando a muchos países a replantearse su “mix” energético e invertir en el desarrollo de fuentes diversas de energías renovables. Por ejemplo, la Unión Europea con su iniciativa 20/20/20, pretende disminuir un 20% el consumo de energía primaria, mediante la mejora de la eficiencia energética, reducir las emisiones de GEI en un 20% (en comparación con 1990) y aumentar al 20% la contribución de las energías renovables.

La energía eólica, como fuente limpia y renovable, se ha desarrollado a gran escala durante la última década. La capacidad total mundial incrementó rápida y establemente desde el año 2000 al 2010, duplicándose cada tres años, como se muestra en la Figura 1.1.

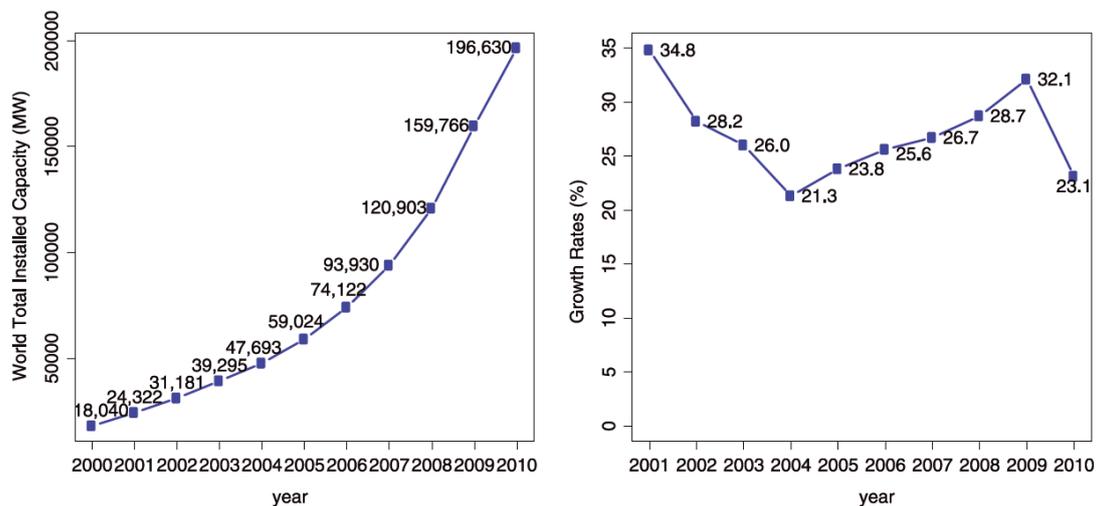


Figura 1.1. Tasa de crecimiento de la potencia total instalada a nivel mundial.

Es una de las fuentes renovables de más rápido crecimiento de la que se espera contribuya en gran medida a la reducción de la huella de carbón y el incremento de utilización de energías renovables en general. Se estima que a finales de 2012, el total acumulado de capacidad de generación de energía eléctrica por medio del viento llegue a los 273 GW [1] (Figura 1.2). La energía del viento es limpia, renovable y su explotación más económica en comparación con otras fuentes renovables.

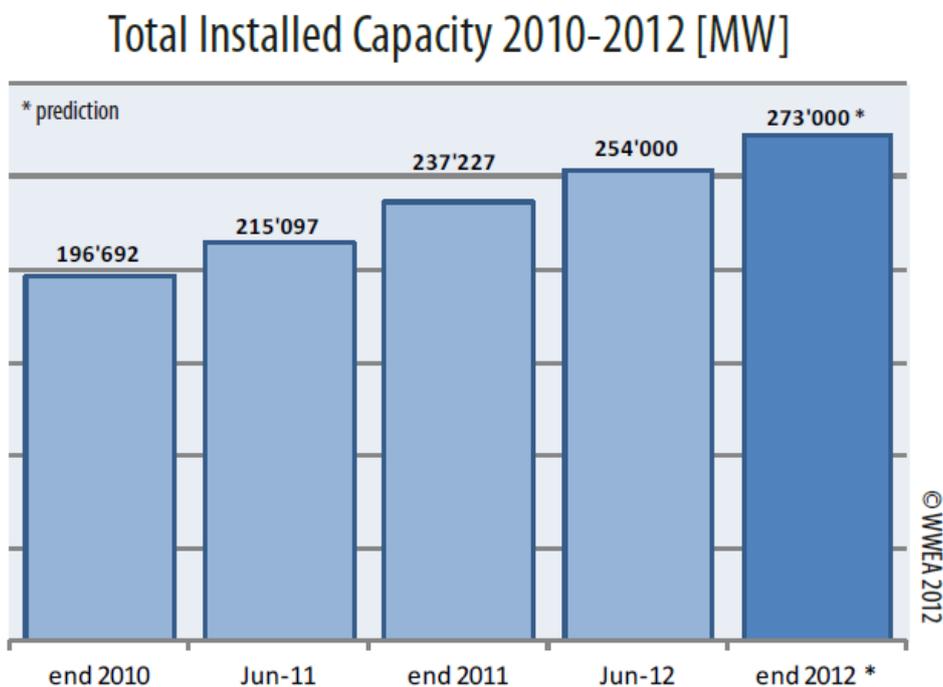


Figura 1.2: Capacidad de generación eólica mundial.

Según el Half Year Report 2012 de la World Wind Energy Association [1], China, Estados Unidos, Alemania, La India y España, representan el 74% de la capacidad total mundial (Figura 1.3).

## Total Installed Capacity 2011-2012 [MW]

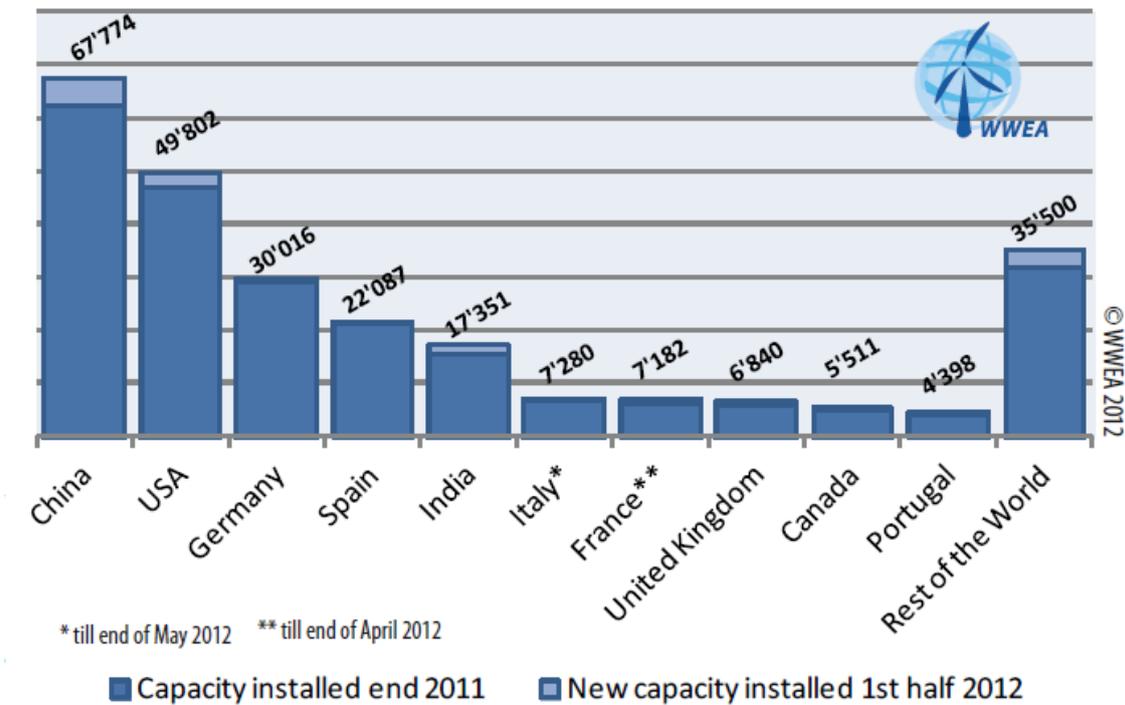


Figura 1.3: Capacidad de generación eólica por país.

El archipiélago canario es una región con un alto potencial eólico, considerado en torno a los 3.600 MW [2].

Los beneficios de la energía eólica vienen acompañados de varios desafíos, alta variabilidad (intermitencia), predictibilidad limitada, limitaciones para el despacho y limitaciones para el almacenamiento. Al contrario de lo que ocurre con la generación mediante combustibles fósiles, la generación eólica no es completamente gestionable. Esto representa el principal escollo para la penetración de la energía eólica, ya que la incertidumbre asociada a la generación eólica pone obstáculos para su integración en el sistema eléctrico.

Por tanto, el objetivo de la predicción de generación eólica contribuye a una operación más segura y económica del sistema eléctrico. Tales predicciones proporcionan al operador del sistema (REE, Red Eléctrica de España) estimaciones sobre la energía que se producirá en el futuro, mitigando la naturaleza intermitente del viento. Las predicciones siempre contienen un error, ya que no pueden ser exactas. Sin embargo, dependiendo de la aplicación, la importancia de este error difiere en gran medida. En el área de la predicción de generación mediante energía eólica, el error en las predicciones se transforma en general en pérdidas económicas para los usuarios finales. Este problema es especialmente acuciante en el caso del sistema eléctrico de las Islas Canarias, por ser este un sistema aislado.

La meteorología energética tiene como meta la evaluación y predicción de la producción energética obtenida de fuentes renovables que se realiza a través del estudio de variables meteorológicas. La correcta evaluación y predicción de esta producción eléctrica, es de máximo interés para facilitar la operación de las redes eléctricas, dado el carácter no gestionable de las energías renovables. El principal beneficio de la meteorología energética, es la reducción de la dependencia de combustibles fósiles.

Tradicionalmente, las predicciones obtenidas a partir de modelos meteorológicos han sido deterministas, o puntuales, que no aportan información sobre la incertidumbre de la estimación. Actualmente, se están desarrollando modelos que dotan de información de incertidumbre a las predicciones puntuales, también llamadas predicciones probabilistas.

En múltiples ocasiones, es posible disponer de la predicción, puntual o probabilista, de más de un modelo meteorológico o del mismo modelo meteorológico con distinta parametrización. El objetivo de este Trabajo de Fin de Máster es el desarrollo de un modelo para la combinación de predicciones probabilistas, utilizando técnicas de aprendizaje máquina con el fin de obtener una única predicción probabilista a partir de éstas. En este caso, la experimentación del modelo desarrollado se realiza con predicciones de viento (y no de energía producida), ya que se cuenta con datos de velocidad del viento, tanto de predicciones puntuales obtenidas con un modelo meteorológico como de medidas reales.

Este Trabajo de Fin de Máster, está en relación con el Trabajo de Fin de Máster de D. Rafael Nebot Medina, titulado “Predicción con incertidumbre en meteorología energética. Aplicación en predicción eólica”, que comprende el desarrollo de un método para añadir información de incertidumbre a la predicción puntual.

## 2. Estado del arte

Debido a la naturaleza estocástica de la velocidad del viento, una predicción precisa de la generación a partir de energía eólica es prácticamente imposible de conseguir. Por consiguiente, se han estudiado e implementado diversas aproximaciones para tratar con la incertidumbre asociada a la integración de la generación eólica. Hasta hace poco, la mayoría de los métodos de predicción eólica eran deterministas, los cuales no pueden proveer información sobre la incertidumbre de los valores predichos, lo que representaba una limitación para su utilidad. Para atacar el problema de la debilidad de la predicción determinista, se han venido desarrollando métodos de predicción probabilista de generación eólica, que tratan con esta incertidumbre. Los métodos actuales de predicción de generación a partir de viento, se pueden clasificar en los siguientes: métodos empíricos estadísticos, métodos basados en regresión cuantílica y métodos de predicción de distribución.

### 2.1 Métodos empíricos estadísticos.

Este tipo de métodos de estimación de la función de densidad de probabilidad están basados en las predicciones puntuales de generación eólica. La variabilidad del error de predicción se estima comparando la generación predicha con la observación. Se considera que la no linealidad del proceso de generación eólica provoca que tenga muchas desventajas el asumir que el error de predicción pertenece a una familia paramétrica de distribuciones. Esto ha motivado el desarrollo de aproximaciones no paramétricas, que puede ser válido para métodos meteorológicos físicos o estadísticos. Una desventaja de estos métodos, es que los modelos específicos tienen que ser entrenados, lo que puede requerir un gran coste computacional. El método de Remuestreo Adaptado, cuyos resultados se utilizan como datos de entrada para el Modelo de Comité de Expertos implementado en este Trabajo de Fin de Máster, es un ejemplo de este tipo de aproximaciones.

## 2.2 Regresión cuantílica.

Los métodos de regresión cuantílica pretenden estimar los cuantiles especificados de la función de distribución de probabilidad de generación eólica. Esta aproximación tiene ventajas obvias, como la no asunción de una cierta distribución y la inclusión flexible de información predictiva. Sin embargo, su desventaja es que el modelo predefinido de regresión puede introducir errores de modelado, y la distribución resultante y la función de densidad de probabilidad son discontinuas. A continuación se describen dos métodos dos de los métodos basados en regresión cuantílica más comunes, el método basado en Bosques de Regresión Cuantílica y el método de Regresión cuantílica con B-Splines.

### 2.2.1 Bosques de regresión cuantílica (Quantile regression forests)

El método de bosques de regresión cuantílica es un método adaptado del método *random forest*, un método de clasificación basado en la realización de múltiples árboles de decisión sobre muestras de un conjunto de datos. El método base utilizado en árboles de regresión cuantílica es llamado método de árboles de clasificación y regresión (CART, del inglés Classification and Regression trees). La meta de CART es dividir una muestra de datos de manera que los nodos hijo sean menos heterogéneos que los nodos padre. Una vez que un árbol haya crecido, se puede extraer información a partir de esa estructura, lo que lo convierte en una herramienta para de análisis. El método *random forest* se ha diseñado para mejorar el CART debido a su inherente inestabilidad (un pequeño cambio en las variables de entrada puede provocar un gran cambio en el predictor).

Formalmente, al finalizar la construcción de un árbol de decisión  $\hat{T}_{(x)}$ , cada nodo hoja corresponde a un subespacio rectangular de las variables explicativas  $X$ . La predicción determinista de un árbol de decisión, dada la variable explicativa  $X_i=x$  se expresa en la ecuación (2.1)

$$\hat{T}(x, \theta) = \sum_{i=1}^N w_i(x, \theta) y_i \quad (2.1)$$

donde  $\theta$  representa los parámetros del árbol que definen la forma en la que el árbol crece,  $w_i(x, \theta)$  son pesos iguales a una constante positiva si  $X_i$  es clasificado en la misma hoja y 0 si no lo es. Las constantes positivas se escogen de tal modo que los pesos sumen uno.

Una forma de tratar la estabilidad de la predicción es generar varios modelos alternativos con pequeñas diferencias en las muestras de entradas o en el modelado. Estos métodos son llamados métodos de *ensemble* en estadística y comparten la misma filosofía como predicciones meteorológicas con ensembles a partir de NWP.

Una aproximación para generar ensembles de predictores con estructura de árbol es la llamada *bosques aleatorios* (Random Forests). Formalmente, un bosque aleatorio consiste en una colección de predictores árbol  $\hat{T}(x, \theta_k)$ , donde  $\theta_k$  son vectores aleatorios de parámetros independientes y distribuidos de manera idéntica, que determinan como el árbol crece. Los predictores, posteriormente, se combinan y proporcionan una aproximación de la media condicionada  $E(Y|X=x)$ . Una generalización de los bosques aleatorios son los bosques de regresión cuantílica, que proporcionan un método no paramétrico de estimación de cuantiles.

### 2.2.2 Regresión cuantílica con B-Splines

En este método, un cuantil se expresa como la combinación lineal de sus variables explicativas:

$$\hat{Q}(\tau, x) = \beta_0(\tau) + \sum_{i=1}^D \beta_i(\tau) x_i \quad (2.2)$$

donde  $\hat{Q}(\tau, x)$  es el cuantil estimado,  $\tau$  es el valor nominal del cuantil,  $D$  es el número de variables explicativas  $x_i$  ( $i=1, \dots, D$ ) considerado y  $\beta_i$  son los parámetros a estimar. Esta aproximación modela la relación entre el cuantil y sus variables explicativas como una combinación lineal de funciones base conocidas (por ejemplo, B-Splines)

$$\hat{Q}(\tau, x) = \alpha_0(\tau) + \sum_{i=1}^D \sum_{j=1}^{N_b} b_{ij}(x_i) \theta_{ij}(\tau) \quad (2.3)$$

donde  $b_{ij}(x_i)$  son las funciones base,  $N_b$  es el número de funciones base y  $\theta_{ij}(\tau)$  son coeficientes desconocidos que se calculan mediante métodos de programación lineal.

### 2.3 Predicción de distribución

Estos métodos son distintos que los métodos de regresión cuantílica, la predicción de la distribución no está limitada a los cuantiles predefinidos, sino que provee una función de densidad de probabilidad completa. Una aproximación a este tipo de métodos son los métodos de estimación de *kernel* (métodos KDE, Kernel Density Estimator).

El estimador de densidad de Kernel obtiene una estimación a partir de muestras de datos, haciendo corresponder a cada muestra individual una función que representa su contribución a la densidad. La distribución es, posteriormente, obtenida sumando todas estas contribuciones. Formalmente, el estimador de kernel d-dimensional viene dado por:

$$\hat{f}(x) = \frac{1}{N|\mathbf{H}|} \sum_{i=1}^N K(\mathbf{H}^{-1}(x - x_i)) \quad (2.4)$$

Donde  $x$  es el punto de evaluación,  $x_i$ ,  $i=1, \dots, N$ , son las muestras,  $\mathbf{H}$  es una matriz de  $d \times d$  que controla el suavizado de la estimación y  $K$  es la función de kernel elegida adecuadamente.

### 3. Metodología

En este apartado se explica la metodología utilizada durante el desarrollo del trabajo. Se comienza con la introducción a las características y modelos de obtención de predicciones puntuales, para posteriormente presentar el método de obtención de predicciones probabilistas que sirve, a su vez, como datos de entrada para el modelo de combinación de fuentes de predicción que se persigue conseguir.

#### 3.1 Predicciones puntuales con modelo meteorológico

Las series temporales de velocidad del viento  $\{v_t\}_{t \in T}$  se pueden definir como una serie temporal *no estacionaria* y *no lineal*. La característica de ser una serie temporal no estacionaria implica que su comportamiento, que puede ser descrito por sus momentos (media, desviación estándar, etc) evoluciona con el tiempo. La no linealidad significa que esas series temporales exhiben características que no pueden ser explicadas por modelos lineales.

La predicción del valor  $v$  en un cierto instante  $t$  para un horizonte de predicción  $t+k$  se denota como  $\hat{p}_{t+k/t}$ . Los métodos de predicción son procedimientos que proporcionar predicciones a partir el conocimiento que se tiene sobre el pasado y el presente. El valor  $\hat{p}_{t+k/t}$  es una predicción puntual, en el sentido de que es un único valor, en contraposición a una predicción probabilista para la que rangos de posibles valores se asocian a una probabilidad. Las predicciones puntuales se obtienen mediante modelos meteorológicos que pueden clasificarse en físicos y estadísticos.

Los modelos físicos proporcionan predicción en nodos específicos de una malla que cubre un área. Debido a que los parques eólicos no están situados exactamente en estos nodos, es necesario extrapolar estas predicciones. La idea principal de la aproximación de modelo meteorológico físico es la de refinar un NWP (del inglés Numerical Weather Prediction) para determinar el viento alrededor de un parque eólico a la altura del generador, utilizando consideraciones físicas acerca del terreno, como la orografía. Mientras que los modelos NWP típicamente tienen resoluciones horizontales de entre 5 y 10 km, los modelos a mesoescala pueden llegar hasta una resolución de 500 m.

La metodología estadística se basa en establecer relaciones entre valores históricos de viento y medidas, intentando ajustar un modelo particular a la muestra de datos usada para entrenarlo, lo que se convierte en un problema de optimización del error. Este tipo de métodos son, en general, métodos de predicción de series temporales que pueden aplicarse en este área.

Partiendo de predicciones puntuales, en el caso de este trabajo predicciones puntuales de velocidad del viento proporcionadas por un modelo físico, se pretende obtener una predicción probabilista.

### 3.2 Método de Remuestreo Adaptado para la obtención de predicciones probabilistas

Se denota  $y_t$  como el viento medido en un tiempo  $t$  que corresponde a la realización de la variable aleatoria  $Y_t$ . Se puede escribir  $f_t$  y  $F_t$  como la función de densidad de probabilidad y función de distribución, respectivamente. Asumiendo que  $F_t$  es una función estrictamente creciente, el cuantil  $q_t^{(\alpha)}$  con proporción  $\alpha \in [0,1]$  de la variable aleatoria  $Y_t$  queda definido como el valor mínimo de  $x$  tal que

$$P(Y_t < x) = \alpha \quad (3.1)$$

o de manera equivalente

$$q_t^{(\alpha)} = F_t^{-1}(\alpha) \quad (3.2)$$

Por tanto, el cuantil predicho  $\hat{q}_t^{(\alpha)} = F_t^{-1}(\alpha)$  con proporción nominal  $\alpha$  es una estimación de  $q_{t+k}^{(\alpha)}$  que se produce en tiempo  $t$  para un horizonte de predicción  $t+k$ , dada la información  $\Omega_t$  disponible en el instante  $t$ .

El primer paso consiste en calcular los errores en los que ha incurrido el modelo de predicción puntual en el pasado. Se define una ventana (un cierto número de horas) para almacenar esos errores. Cuando se realiza una nueva predicción, se compara con las pasadas en ese determinado momento (consideramos predicciones horarias, aunque el método es aplicable para distintas resoluciones temporales). El tamaño  $n$  de la ventana determina el tamaño de las muestras de error.

En el instante  $t$ , se define una muestra  $S_{t,k}$  para cada horizonte de predicción, dado que la incertidumbre en ésta varía de manera significativa con este horizonte temporal. Se define  $\Omega_{t,k}$ , ecuación (3.3), como el conjunto de los errores de predicción pasados asociados al horizonte  $k$  hasta el instante  $t$  actual.

$$\Omega_{t,k} = \{\varepsilon_{t-i+k|t-i}, i \in \mathbb{N}, i \geq k\} \quad (3.3)$$

donde  $\varepsilon_{t-i+k|t-i}$  es el error de predicción normalizado con respecto a la predicción puntual  $\hat{y}_{t-i+k|t-i}$ . Por tanto, una muestra de error  $S_{t,k}$  que contenga las últimas  $k$  predicciones puntuales en el instante  $t$  viene dada por la ecuación. (3.4)

$$S_{t,k} = \{\varepsilon_i \in \Omega_{t,k}, i = 1, \dots, n\} \quad (3.4)$$

Si una determinada muestra está completa cuando quiere añadirse un nuevo valor de error, se descarta el valor de error más antiguo, remplazándose por el más reciente. Esta idea viene dada por el carácter no estacionario de los errores de predicción de viento.

La distribución empírica de errores  $\hat{f}_{t,k}^\varepsilon$ , en el instante  $t$  con horizonte  $k$ , se define como una distribución discreta que asigna una probabilidad  $1/n$  para cada elemento de  $S_{t,k}$ . Se puede demostrar que  $\hat{f}_{t,k}^\varepsilon$  es la estimación no paramétrica de máxima verosimilitud (maximum likelihood) de la verdadera función de distribución del error  $f_{t,k}^\varepsilon$ . Por tanto, la verosimilitud se maximiza sin realizar ninguna asunción sobre el carácter paramétrico del proceso de error. Como consecuencia, cualquier parámetro  $\hat{\theta}(\hat{f}_{t,k}^\varepsilon)$  que se estime a partir de  $\hat{f}_{t,k}^\varepsilon$  es

la estimación no paramétrica de máxima verosimilitud del parámetro  $\theta(f_{t,k}^\epsilon)$ . La función de distribución  $\hat{F}_{t,k}^\epsilon(x)$  da la fracción de errores menores o iguales a  $x$ .

$$\hat{F}_{t,k}^\epsilon(x) = \frac{1}{n} \#\{\epsilon_i\} \quad (3.5)$$

En la ecuación (3.5) “#” es el operador matemático de cardinalidad, que devuelve el número de elementos del conjunto sobre el que es aplicado. En esta aproximación empírica se asume que la incertidumbre futura puede expresarse a partir del comportamiento observado reciente del modelo de predicción puntual. Esto se extrapola a la distribución empírica  $\hat{f}_{t,k}^\epsilon$  del error, que puede verse como una estimación de los errores de distribución asociados a la predicción puntual  $\hat{y}_{t+k|t}$ .

Para el caso de la predicción de viento, las distribuciones de probabilidad del error pueden verse afectadas por una serie de variables influyentes. Podemos definir  $c_{t,k}$  en el instante  $t$  con horizonte  $k$  como un conjunto de valores que se consideran variables influyentes. Denotamos con  $v_l$  la variable influyente  $l$ -ésima (considerando  $L$  variables influyentes,  $l=1, \dots, L$ ) y  $v_{t,k}^l$  su valor en el instante  $t$  con horizonte de predicción  $k$ . Por ejemplo, podría definirse  $v_{t,k}^1 = \hat{y}_{t+k|t}$  y  $v_{t,k}^2 = \hat{u}_{t+k|t}$  como dos variables influyentes para la predicción del viento siendo éstas presión y temperatura. Se asume que las variables influyentes tienen unos mínimos y unos máximos, es decir, pertenecen a cierto intervalo determinado por la física y, por tanto, pueden ser normalizadas, como puede verse en la ecuación (3.6).

$$v_{t,k}^l \in V_l = [0,1] \forall l, t, k \quad (3.6)$$

Hay que tener en cuenta que los errores de predicción también están normalizados en el intervalo  $[-1,1]$ .

Una condición de predicción en el instante  $t$  para un horizonte de predicción  $t+k$  se define de manera unívoca con la asociación de cada una de las  $L$  variables influyentes, como puede verse en la ecuación (3.7).

$$c_{t,k} = \{v_{t,k}^1, v_{t,k}^2, \dots, v_{t,k}^L\}, c_{t,k} \in C = V_1 \times \dots \times V_L \quad (3.7)$$

Donde  $C$  es el conjunto de posibles condiciones de predicción en un instante  $t$  para un horizonte de predicción  $k$ . Posteriormente, se hace corresponder  $C$  con un número finito de subconjuntos que están asociados a distintos tipos de características de distribuciones de error de predicción. Para ello, se consideran  $J_l$  rangos de posibles valores de cada variable influyente  $v_l (l=1, \dots, L)$  y se define como  $V_l^{j_l}$  como el conjunto de  $V_l$  que contiene los valores de las variables influyentes en el rango  $j_l$ -ésimo. En la ecuación (3.8) se muestra  $V_l$  como la unión de todos sus subconjuntos

$$V_l = V_l^1 \cup V_l^2 \cup \dots \cup V_l^{j_l}, \forall l \quad (3.8)$$

tal que todos estos subconjuntos sean disjuntos.

$$V_l^i \cap V_l^j = \emptyset, \forall l, i, j \ i \neq j \quad (3.9)$$

Los conjuntos de posibles valores para las variables influyentes se han dividido en varios subconjuntos asociados a diferentes características de la distribución de la predicción del error.

Esta clasificación descrita previamente es la base para un método empírico e independiente de la distribución. La elección de variables influyentes, así como la división de los conjuntos de posibles valores en varios subconjuntos con diferentes características de distribuciones de error de predicción, son el resultado de la experiencia que se tenga en este proceso.

El método de Remuestreo Adaptado define distribuciones condicionales de probabilidad, dada cierta condición de predicción  $c_{t,k}$ . Permite obtener una estimación de los cuantiles de la predicción (y por lo tanto la distribución completa) a partir de una combinación de distribuciones de errores pasados separadas en diferentes clases, utilizando un clasificador difuso. Este clasificador permite introducir conocimiento sobre las situaciones que se presentan en el problema que se esté contemplando, ya sea predicción de viento o radiación, o de potencia eólica o fotovoltaica.

### 3.3 Modelo de Comité de Expertos

Se explica a continuación el Modelo de Comité de Expertos, que realizará la combinación de fuentes de predicción o predicciones probabilistas que son el resultado del Método de Remuestreo Adaptado.

#### 3.3.1 Máquinas de comité

Una tarea computacional compleja puede resolverse de acuerdo al principio “divide y vencerás”, dividiéndola en una serie de tareas que son computacionalmente menos costosas. En el área del aprendizaje supervisado, la simplicidad computacional se alcanza con la distribución de las tareas de aprendizaje entre un número de **expertos**, combinando posteriormente el conocimiento adquirido por cada uno de ellos.

Las máquinas de comité pueden ser clasificadas en dos categorías [3]:

- Estructuras estáticas: El mecanismo de combinación de los expertos no incluye a la señal de entrada. De ahí su denominación de “estática”.
  - Ensemble averaging: Las salidas de los distintos expertos se combinan linealmente.
  - Boosting: Combinar algoritmos de aprendizaje débiles para obtener uno de mayor precisión.

- Estructuras dinámicas: En esta clase de máquinas de comité, la señal de entrada sí influye en el comportamiento del modelo de combinación.
  - Mezcla de Expertos: Las respuestas de cada uno de los expertos se combina de manera no lineal a través de una *Gating Network*.
  - Mezcla Jerárquica de Expertos: Las respuestas de cada uno de los expertos se combina de manera no lineal a través de una jerarquía de *Gating Networks*.

Se presta especial atención al modelo de Mezcla de Expertos, que será el utilizado para la experimentación con combinación de predicciones probabilistas. Este modelo se presenta en la Figura 3.1. Consiste en K módulos supervisados llamados *expertos* y una unidad que interviene en la combinación de estos llamada *gating network*. Se asume que diferentes expertos trabajan mejor en diferentes regiones del espacio de entrada.

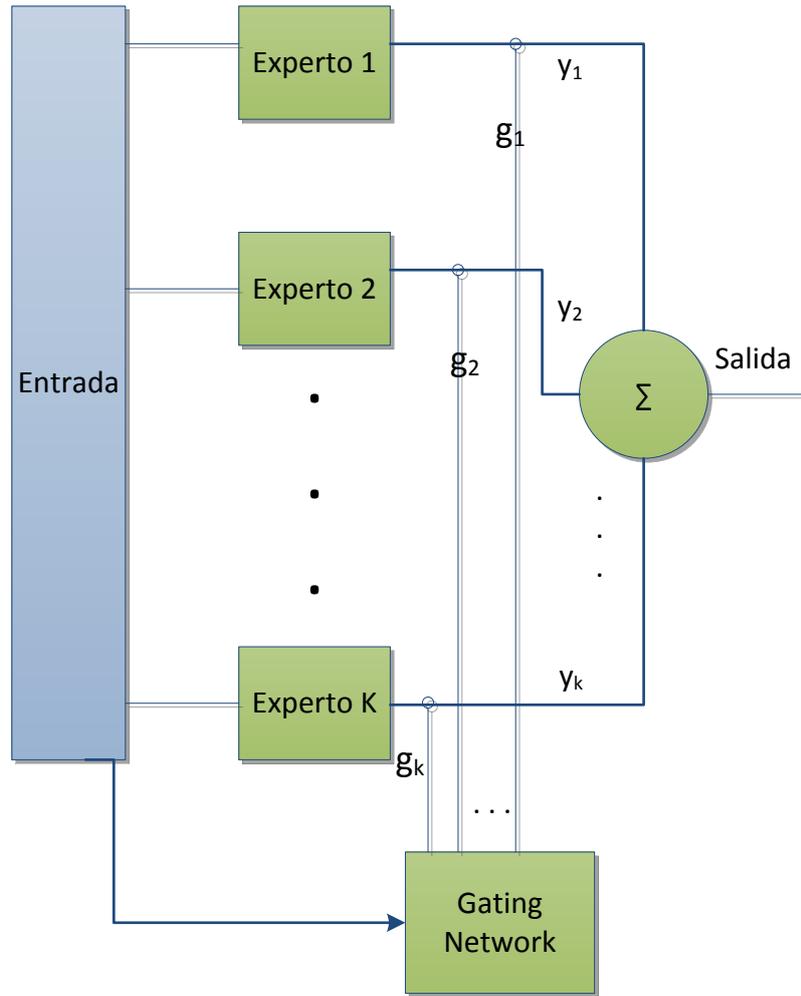


Figura 3.1: Diagrama de bloques del modelo de Mezcla de Expertos.

La Gating Network consiste en una capa de  $K$  neuronas con cada neurona asignada a un experto específico. La Figura 3.2 muestra la arquitectura de la Gating Network, mientras que la Figura 3.3 representa el grafo de señal de una de sus neuronas  $k$ .

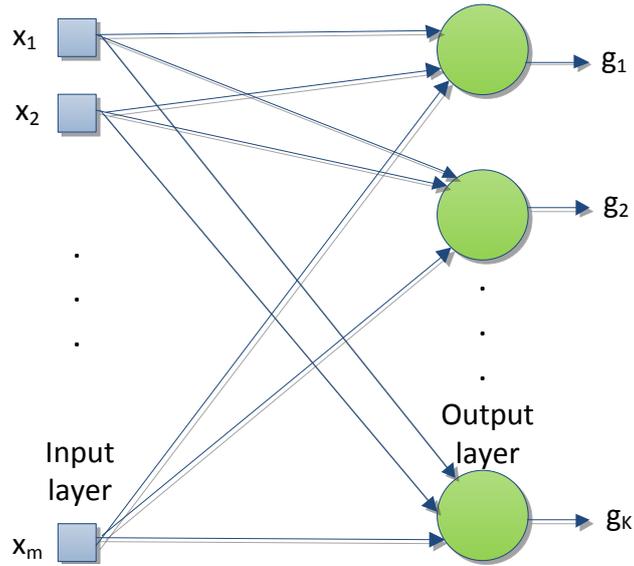


Figura 3.2: Capa de neuronas softmax para la Gating Network.

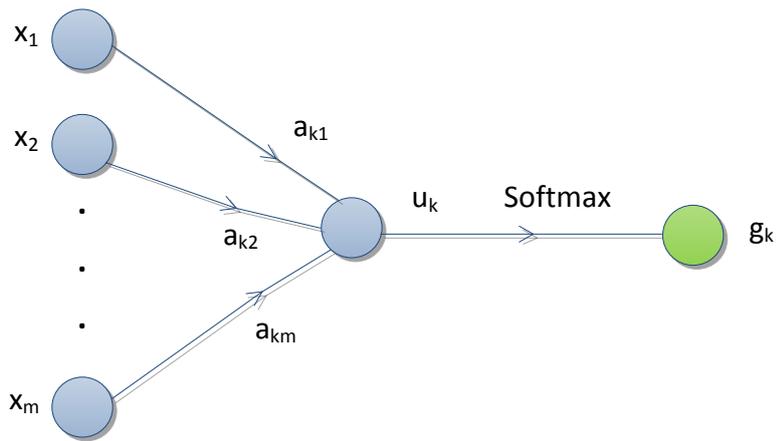


Figura 3.3: Gráfico de flujo de señal de una neurona softmax.

Las neuronas de la Gating Network son no lineales, con su función de activación definida en la ecuación (3.10).

$$g_k = \frac{\exp(u_k)}{\sum_{j=1}^K \exp(u_j)}, k = 1, 2, \dots, K \quad (3.10)$$

$$u_k = \mathbf{a}_k^T \mathbf{x}, k = 1, 2, \dots, K \quad (3.11)$$

La transformación exponencial normalizada de la ecuación (3.10) puede verse como una generalización de múltiples entradas de la función logística.

### 3.3.2 Aplicación de las máquinas de comité a predicción probabilista

A partir del modelo de “Mezcla de Expertos” podemos diseñar un modelo de combinación de predicciones probabilistas, que hayan sido previamente obtenidas a partir de predicciones puntuales. En este caso, cada experto es una predicción probabilista proporcionada por el Método de Remuestreo Adaptado y produce una combinación de éstas para producir una predicción probabilista resultante.

Como puede verse en la Figura 3.4, se considera que las entradas al modelo son N expertos, entendidos como N predicciones probabilistas. Por supuesto, estos expertos pueden provenir de distintos modelos meteorológicos, siempre y cuando tengan en común tanto el horizonte de predicción como la resolución (por ejemplo, horaria). En este caso, se muestra el modelo en su fase de entrenamiento, ya que el Remuestreo Adaptado requiere contar con las observaciones para proporcionar su resultado.

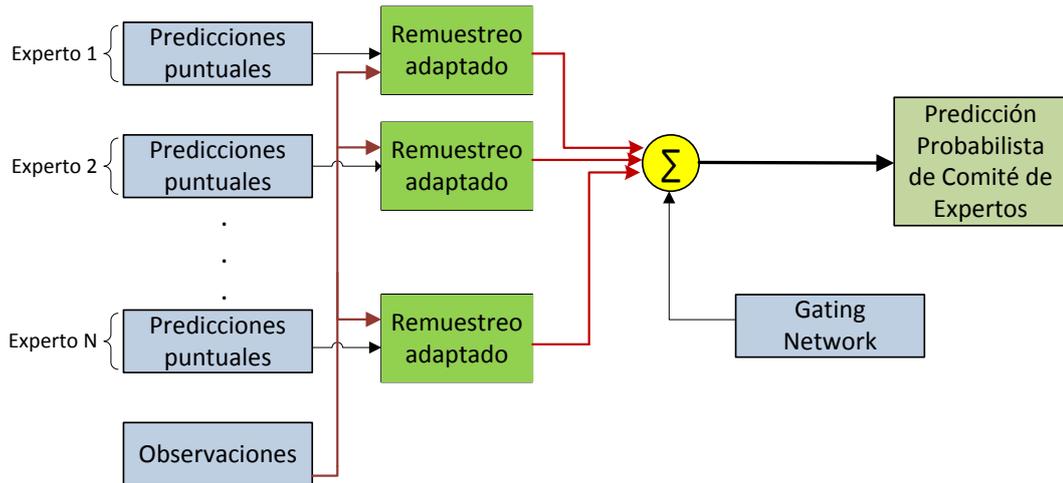


Figura 3.4: A partir de las predicciones puntuales dadas por distintos expertos, se realiza una combinación ponderada con pesos dados por la Gating Network.

Una vez halladas las predicciones probabilistas (los expertos) se realiza la combinación. La Gating Network es la encargada de proporcionar los pesos para la combinación ponderada de los expertos. Siguiendo el planteamiento en [3], la Gating Network se modela como una red neuronal cuya capa de salida se una función logística.

Una vez la red neuronal determine los pesos adecuados para la cada uno de los expertos, la combinación se realiza mediante la técnica de bootstrapping. En un primer paso se obtienen muestras de cada uno de los expertos, que se combinan con un remuestreo con reemplazo. A partir de la muestra resultado del remuestreo con reemplazo, se obtiene su función de distribución de probabilidad (se obtiene el histograma y, posteriormente, se halla la distribución acumulada). Este proceso se hace iterativamente para obtener un promedio de la función de distribución de probabilidad que se obtiene en cada iteración, dada la naturaleza aleatoria del algoritmo de obtención de muestras de cada función de distribución de probabilidad correspondiente a cada experto.

### 3.3.3 Método de entrenamiento de la Gating Network

La parametrización de este modelo pasa por definir, para cada caso de interés, una serie de pesos con los cuales poder entrenar a la Gating Network, de manera que sea capaz de proporcionar pesos adecuados (Figura 3.5).

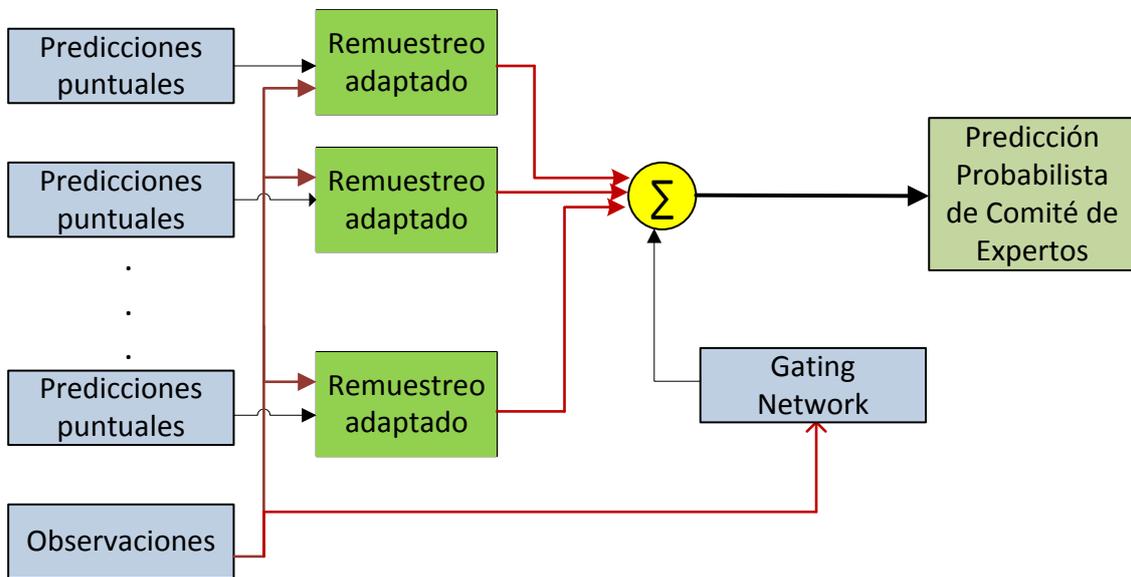


Figura 3.5: Entrenamiento de la Gating Network.

En este trabajo, se propone un método para el entrenamiento de la Gating Network, de manera que proporcione los pesos adecuados para realizar la combinación, aplicando un algoritmo genético que optimiza una figura de mérito.

A partir de varias muestras de predicciones puntuales proporcionadas por distintos expertos (modelos meteorológicos) y sus observaciones correspondientes puede obtenerse, para cada combinación, un conjunto de pesos que proporcione una salida combinada tal que minimice una determinada figura de mérito.

Las combinaciones de predicciones puntuales y pesos correspondientes hallados por el algoritmo genético constituyen un conjunto de datos de muestra para el entrenamiento y test de una red neuronal.

Se ha diseñado una figura de mérito (**Figura 3.6**) tal que a partir de la distribución de probabilidad y la realización de la variable (velocidad del viento en este caso) u observación  $v$ , genere un intervalo  $w=[W_L, W_R]$  tal que  $W_L=v-\Delta v$  y  $W_R=v+\Delta v$ , de modo que el ancho de este intervalo ( $w$ ) sea tal que el área de la función de densidad de probabilidad comprendida en él sea igual a un cierto valor  $\sigma$  (por ejemplo, un 50% del área total). Al comparar dos funciones de densidad de probabilidad, aquella que necesite un tamaño de intervalo  $w$  menor para alcanzar ese valor de área  $\sigma$ , será considerada como “más precisa”.

El algoritmo genético, tiene como variables de diseño los pesos que servirán para obtener una combinación de expertos. El resultado de esa combinación será evaluado con la figura de mérito, que será la función a optimizar. Por tanto, dadas dos combinaciones de pesos, la que permita una combinación que minimice la figura de mérito, será considerada la mejor de las dos combinaciones.

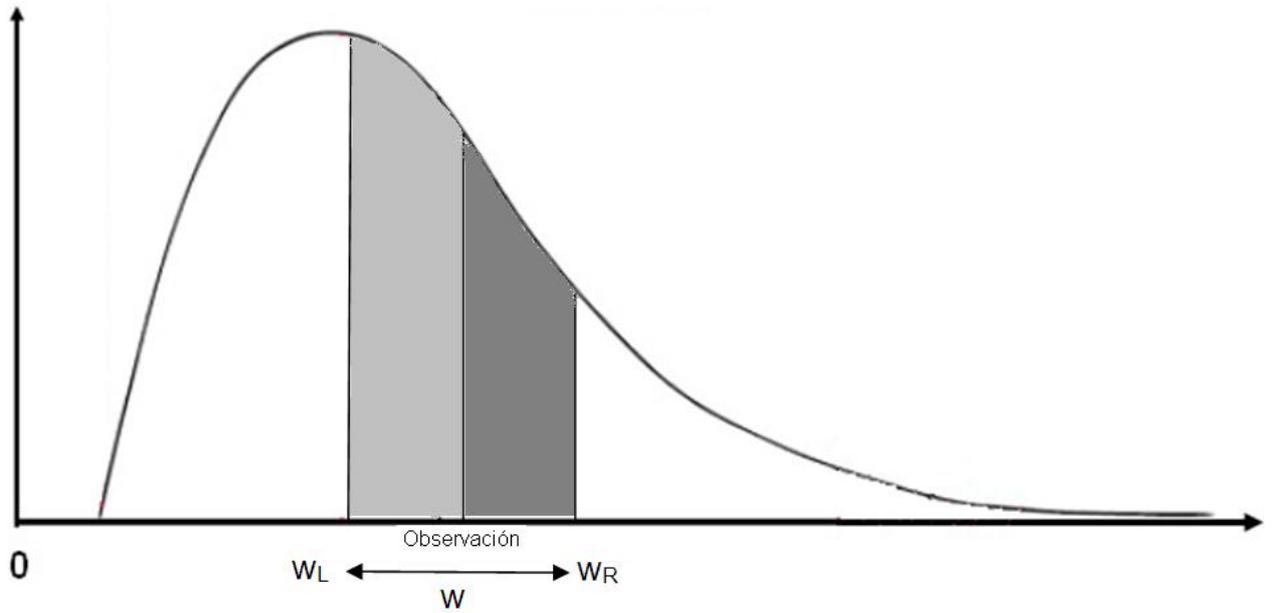


Figura 3.6: Figura de mérito.

Una vez obtenidos los pesos para un cierto número de combinaciones predicción-observación, se puede entrenar la Gating Network para que proporcione los pesos adecuados cuando el modelo esté en operación (Figura 3.7).

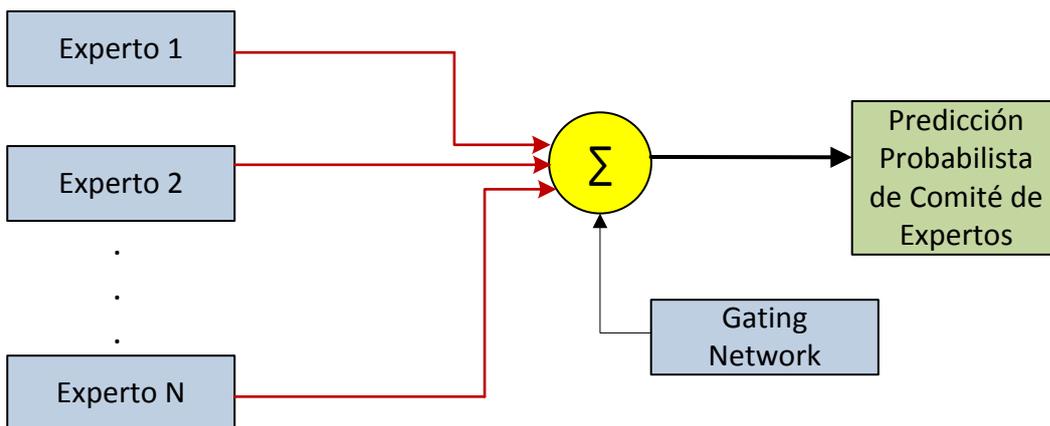


Figura 3.7: Modelo de Comité de Expertos en operación.

## 4. Aplicación experimental

### 4.1 Datos de entrada

Como datos de entrada al Modelo de Comité de Expertos, se tienen cuatro predicciones probabilistas. Estas predicciones probabilistas se han obtenido con la aplicación del Método de Remuestreo Adaptado a partir de predicciones puntuales o discretas de un modelo meteorológico basado en la física de la atmósfera aplicado al archipiélago Canario.

En este caso, se dispone de una serie de predicciones puntuales que van desde el julio de 2009 hasta octubre de 2011 del modelo meteorológico MM5 (acrónimo de Fifth-Generation Penn State/NCAR Mesoscale Model) desarrollado en colaboración entre la Universidad de Pennsylvania State (Penn State) y el National Centre for Atmospheric Research (NCAR). Es un modelo de área limitada, tanto hidrostático como no hidrostático, que utiliza coordenadas sigma para el seguimiento del terreno diseñado para predicciones a mesoescala y microescala. MM5 es la quinta generación de este modelo. Cada día a las 06:00 se ejecuta el modelo meteorológico con un horizonte de predicción de 120 horas (cinco días), con resolución horaria.

	+0-24h	+24-48h	+48-72h	+72-96h	+96-120h
Día D-4					Predicciones puntuales D
Día D-3				Predicciones puntuales D	
Día D-2			Predicciones puntuales D		
Día D-1		Predicciones puntuales D			
Día D	Predicciones puntuales D				

Figura 4.1: Predicciones del modelo meteorológico (MM5) durante cinco días. Cada una de ellas predice con horizonte de +120 horas.

Las predicciones obtenidas por el modelo meteorológico MM5 para el día D, equivalen a las obtenidas por la ejecución del modelo correspondiente al día anterior (D-1) en su horizonte de predicción +24, como puede verse en la Figura 4.1.

	+24-48h	+48-72h	+72-96h	+96-120h
Día D-3				Predicciones puntuales D+1
Día D-2			Predicciones puntuales D+1	
Día D-1		Predicciones puntuales D+1		
Día D	Predicciones puntuales D+1			

Figura 4.2: Predicciones del modelo meteorológico (MM5) a partir del segundo día (desde el horizonte +24 al +120).

En operación, la predicción realizada el mismo día para el mismo día no es útil y se descarta, debido a que la operadora del sistema eléctrico (Red Eléctrica de España) necesita la predicción de producción energética con antelación. En la Figura 4.2 se ilustran las predicciones producidas por el modelo meteorológico MM5 que son útiles para REE. Se muestran los cuatro intervalos de 24 horas que son coincidentes en el tiempo.

Con estos cuatro conjuntos de predicciones puntuales (con resolución horaria) para un mismo día D+1, se puede considerar que se tienen cuatro resultados de un mismo modelo meteorológico con distintas condiciones iniciales, ya que cada una de estas ejecuciones de MM5 posee datos cada vez más cercanos al día D+1, siendo el día justo anterior (D) el que posee las condiciones iniciales más cercanas.

En total se tiene una muestra de 8270 predicciones probabilistas con su observación correspondiente.

## 4.2 Experimentos realizados

Se han realizado diversos experimentos para comprobar el correcto funcionamiento de este método de combinación o mezcla de expertos.

En primer lugar se ha evaluado el comportamiento del modelo mediante una Gating Network que establece un peso máximo para cada uno de los expertos y un peso mínimo para el resto. De esta manera, podemos garantizar el correcto funcionamiento del framework, que se completa con la evaluación del rendimiento de la predicción, como se podrá ver en el capítulo 5.

Posteriormente, a partir del conjunto de datos disponible, se realiza un particionado de datos (2-fold cross-validation) en un conjunto de train y uno de test. El conjunto de train contiene el 70% de los datos, mientras que el conjunto de test contiene el 30% restante. Con el conjunto de train se obtienen pesos correspondientes mediante el algoritmo genético, que posteriormente servirán para entrenar la red neuronal que constituye la Gating Network. Una vez entrenada la Gating Network, se procede a la evaluación del comportamiento del modelo con el conjunto de test. El criterio de evaluación del rendimiento de los modelos se describe en el siguiente capítulo.

## 5. Resultados

Una posibilidad de comparar el rendimiento entre varios métodos y, además, de cuantificar la mejora de una aproximación avanzada con respecto a una de referencia, que son el resultado de consideraciones simples y no de esfuerzos de modelado. Probablemente el modelo de referencia más común utilizado en predicción eólica, o en meteorología en general, es el modelo de persistencia. Este simple predictor plantea que la velocidad de viento futura será la misma que el último valor medido.

$$\hat{p}_{t+k/t}^p = p_t \quad (5.1)$$

A pesar de su aparente simplicidad, esta referencia puede ser difícil de vencer para un horizonte de predicción cercano. Una generalización del método de persistencia es remplazar el último valor medido por la media de los últimos  $n$  valores medidos.

$$\hat{p}_{t+k/t}^{MA,n} = \frac{1}{n} \sum_{i=0}^{n-1} p_{t-i} \quad (5.2)$$

Asintóticamente (a medida que  $n$  tiende a infinito), tienden a la media global (también conocida como la media climatológica).

$$\hat{p}_{t+k/t}^0 = \bar{p}_t \quad (5.3)$$

donde  $\bar{p}_t$  es la media de todas las observaciones de viento disponibles hasta el instante  $t$ . Esto también puede verse como un modelo de referencia, y viene dado por todas las observaciones disponibles, dando como resultado una única predicción probabilista. Debido a su carácter estático su rendimiento puede ser muy pobre en el horizonte de predicción cercano. Sin embargo, para horizontes de predicción lejanos, su *skill* es mucho mejor que el de persistencia. El modelo de referencia más común para predicción probabilista de variables y procesos meteorológicos es el climatológico, por tanto será el modelo de referencia utilizado para la evaluación del rendimiento del Modelo de Comité de Expertos.

Por tanto, a la hora de comparar el rendimiento de dos modelos, podemos cuantificar la mejora de cada uno de esos modelos con respecto a un criterio de un modelo con respecto a uno de referencia. Esta mejora (llamada *skill score*) corresponde a la reducción del error que alcanza un determinado método, para una determinada medida de error. Se define en la ecuación (5.4).

$$Imp_Y^{ref}(k) = \frac{Y^{ref}(k) - Y(k)}{Y^{ref}(k)} \quad (5.4)$$

donde  $Y$  se considera el criterio de evaluación.  $Y^{ref}(k)$  denota su valor para el modelo de referencia y  $Y(k)$  para el modelo cuyo rendimiento se quiere evaluar, para el horizonte  $k$ . Esta mejora normalmente se multiplica por cien y se expresa como un porcentaje de mejora con respecto a la aproximación de referencia. Los valores de  $Imp_Y^{ref}$  son siempre menores que el 100% (ya que esta mejora significaría que las predicciones son perfectas), pero pueden ser menores del 0% si el método evaluado es peor que la referencia.

Las predicciones probabilistas suelen evaluarse en base a su nitidez [4] [5] [6] [7]. La nitidez o *sharpness* es un atributo de la predicción probabilista, independiente de la observación. Una posible definición de la nitidez se la habilidad de la predicción de desviarse de las probabilidades climatológicas. Las predicciones que no se desvían demasiado de la referencia climatológica, exhiben un bajo valor de nitidez. La nitidez está relacionada con la concentración de probabilidades, esto es, la habilidad del modelo para distinguir situaciones específicas. Este framework de evaluación, por tanto, está basado en maximizar la nitidez.

El criterio utilizado para evaluar la nitidez, es el utilizado en [4], que proporciona una *puntuación* para una distribución  $\hat{f}_{t+k}$  dada y su realización correspondiente  $y_{t+k}$ , viene dado en la ecuación (5.5).

$$S_c(\hat{f}_{t+k|t}, y_{t+k}) = \sum_{i=1}^m (\xi^{\alpha_i} - \alpha_i) (y_{t+k} - \hat{q}_{t+k|t}^{(\alpha_i)}) \quad (5.5)$$

Donde  $\xi^{\alpha_i}$  es una variable indicadora que es 1 cuando el la observación está opr debajo del cuantil  $\hat{q}_{t+k|t}^{(\alpha_i)}$  y 0 en caso contrario. En un conjunto de N series de predicción, se aplica ese criterio de puntuación para un cierto horizonte de predicción  $k$ .

$$S_{C_K} = \frac{1}{N} \sum_{t=1}^N S_c(\hat{f}_{t+k}, y_{t+k}) \quad (5.6)$$

Para evaluar el *skill score* de las distribuciones de probabilidad basándonos en el criterio anterior, se utiliza el modelo de referencia climatológico (ecuación (5.7)).

$$SS_{C_K} = 100 \frac{S_{C_K^0} - S_{C_K}}{S_{C_K^0}} \quad (5.7)$$

El primer resultado que se muestra en la Figura 5.1 muestra el Skill Score del modelo de Comité de Expertos frente al Skill Score individual de cada uno de los expertos. En este caso, como era de esperar, el Skill Score del modelo en estudio es muy similar al del primer experto.

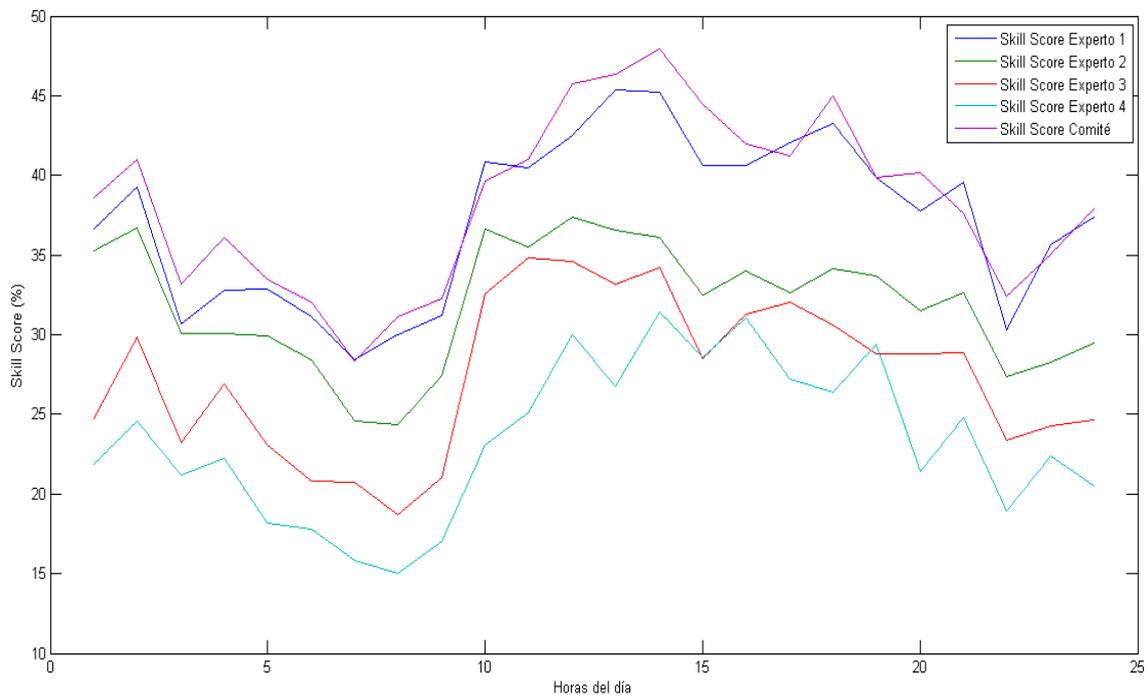


Figura 5.1: Gating Network configurada para que el Experto 1 obtenga un peso mayor.

Las figuras Figura 5.2, Figura 5.3 y Figura 5.4 muestran, respectivamente, el *skill score* cuya Gating Network favorece a los expertos 2, 3 y 4. Como en el resultado anterior correspondiente al primer experto, el *skill score* es similar a los *skill score* individuales de cada uno de los expertos.

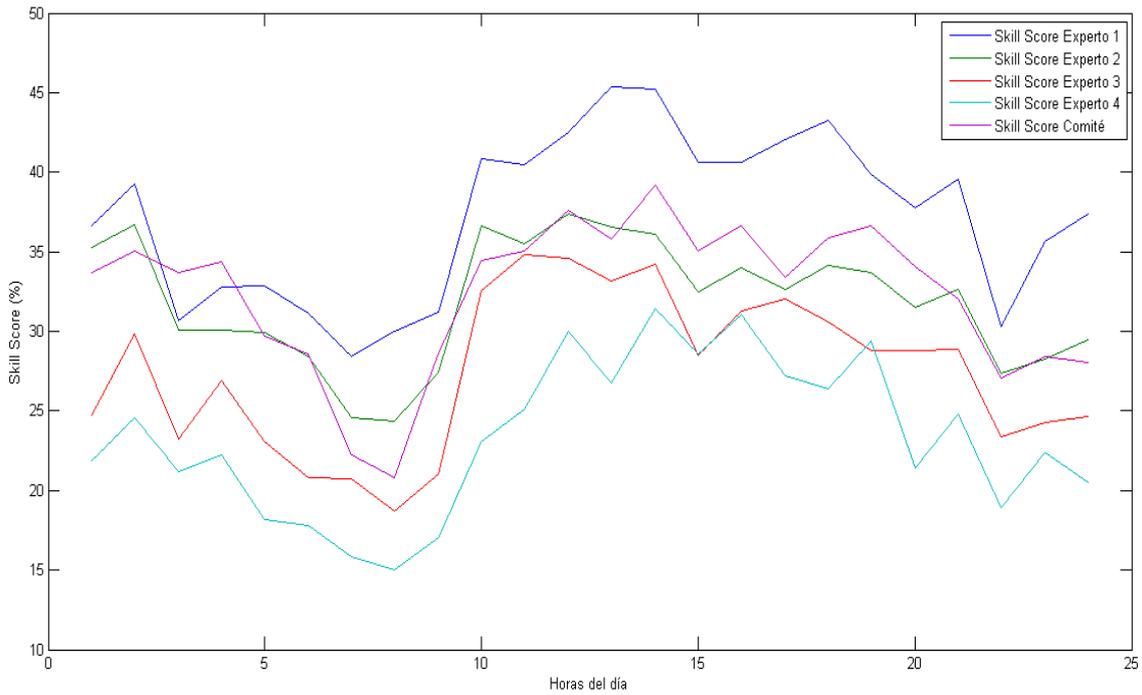


Figura 5.2: Gating Network configurada para que el Experto 2 obtenga un peso mayor.

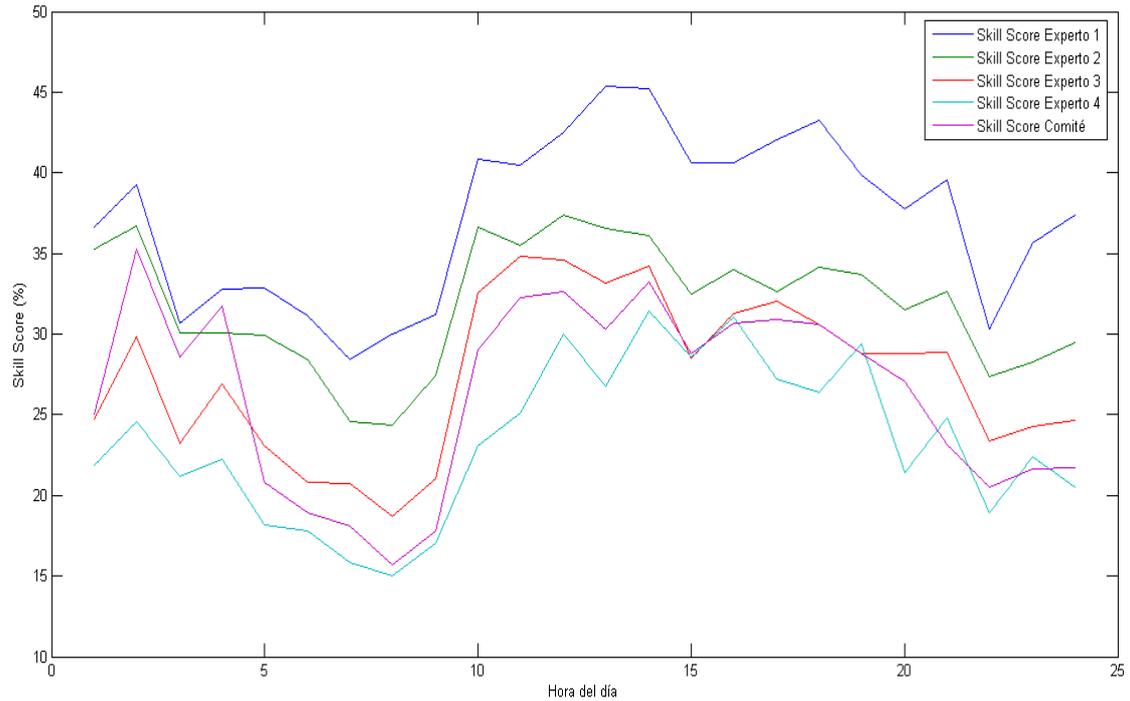


Figura 5.3: Gating Network configurada para que el Experto 3 obtenga un peso mayor.

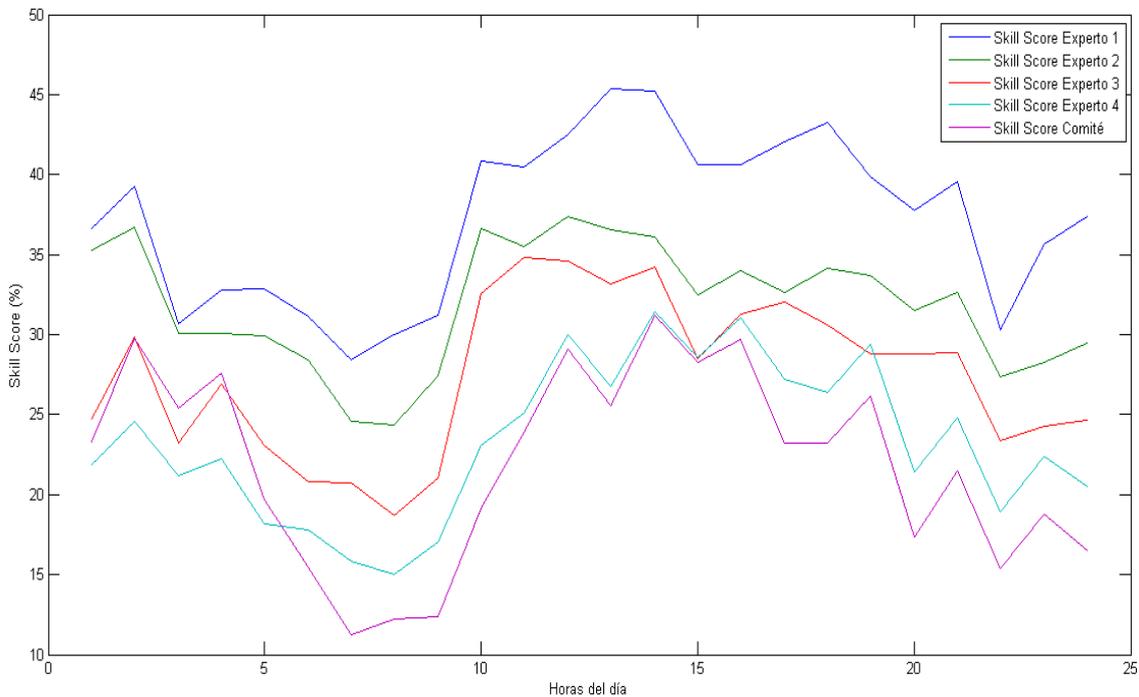


Figura 5.4: Gating Network configurada para que el Experto 4 obtenga un peso mayor.

Por tanto, el *skill score* con respecto al modelo de referencia climatológico representado en la Figura 5.1 es el que obtiene un mejor resultado. Esto es lógico si tenemos en cuenta que el Experto 1 es el que tiene unas condiciones iniciales más cercanas al instante temporal predicho. Con estos resultados mostrados en las figuras 5.1, 5.2, 5.3 y 5.4 se puede concluir que el modelo de mezcla se comporta como se espera, realizando la combinación correcta.

Esto permite plantear su utilización con otras combinaciones de pesos que puedan proporcionarse desde la Gating Network con un entrenamiento previo adecuado para un caso concreto. Como ejemplo, se muestra el resultado del Modelo de Comité de Expertos para el caso de entrenamiento de la Gating Network con pesos hallados con un algoritmo genético, como se ha explicado en la sección 3.3.3. En la figura Figura 5.5 se muestra el Skill Score para este caso. Puede comprobarse se obtiene un resultado parecido al del Experto 1, lo cual es algo lógico, ya que, con los datos disponibles, el Experto 1 será mejor que el resto de los expertos en la inmensa mayoría de los casos de muestra, puesto que las predicciones puntuales obtenidas por el modelo meteorológico MM5 para el Experto 1 parten de condiciones iniciales más recientes.

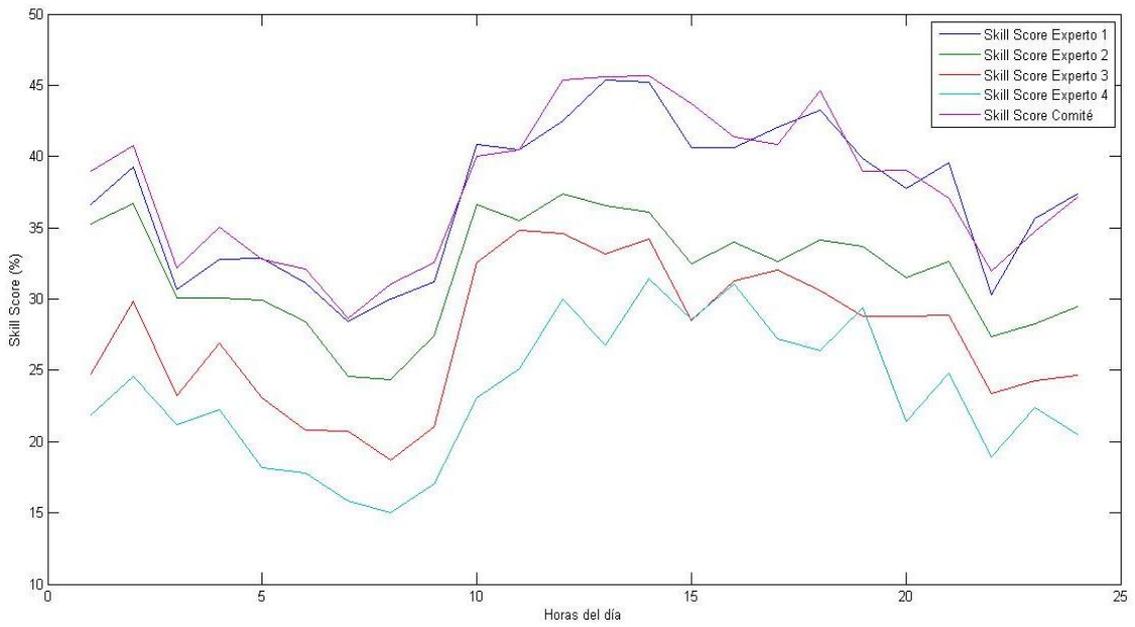


Figura 5.5. Gating Network entrenada con los pesos obtenidos mediante un algoritmo genético.

## 6. Conclusiones y trabajo futuro

Este trabajo se planteó en un primer momento como un complemento a otro Trabajo de Fin de Máster que se realizaría de manera paralela por D. Rafael Nebot Medina, que lleva por título “Predicción con incertidumbre en meteorología energética. Aplicación en predicción eólica”. En él, se implementaría el Método de Remuestreo Adaptado, basado en la tesis de Pierre Pinson [8], cuyo objetivo es obtener predicciones probabilistas de velocidad del viento. Este método integra un modelo de inferencia difusa que permite introducir conocimiento experto sobre las características de la incertidumbre predictiva. Partiendo de una serie larga de predicciones puntuales obtenidas por un modelo físico (MM5), mediante el Método de Remuestreo Adaptado se obtuvieron predicciones probabilistas que sirvieron como entrada al Modelo de Comité de Expertos, que se ha implementado en este trabajo.

La idea de poder realizar la combinación o mezcla de predicciones probabilistas surge del planteamiento de poder obtener predicciones probabilistas a partir de distintos modelos meteorológicos físicos (WRF o MM5) o bien de cualquier modelo de predicción puntual. La intención es aprovechar las mejores características de cada uno de los modelos meteorológicos de cuyos resultados para predicción se dispongan, ya que el rendimiento de cada uno de ellos puede ser muy distinto en función de las condiciones bajo las que se ejecutan.

La idea que se ha desarrollado con el Modelo de Comité de Expertos es una aproximación a la utilización de técnicas de Machine Learning para la combinación de distribuciones de probabilidad. Si bien al revisar la literatura, hay métodos que intentan este objetivo, son en gran parte analíticos y puramente estadísticos.

Una de las principales aportaciones de este trabajo es precisamente la utilización de herramientas de aprendizaje automático para la obtención de esta combinación de predicciones probabilistas, siendo algo novedoso en este ámbito la utilización de Máquinas de Comité.

Debido a las características de los datos disponibles para la experimentación con el nuevo modelo, que son predicciones para un mismo día lanzadas en hasta cuatro días anteriores, no se ha podido realizar un experimento con distintos modelos meteorológicos ejecutados el mismo día y, por tanto, partiendo de las mismas condiciones iniciales. En condiciones como ésta, en la que se trabaje con predicciones probabilistas con un mismo nivel de confianza, es cuando este modelo gana en utilidad.

Por tanto, se ha logra obtener un modelo para la combinación de predicciones probabilistas aplicado a la predicción eólica. La parametrización de este modelo pasa por definir, para cada caso de interés, una serie de pesos con los cuales poder entrenarlo, de modo que puede haber un refuerzo con conocimiento experto para la obtención de estos pesos. En este trabajo se ha propuesto un método basado en algoritmos genéticos para el entrenamiento del modelo.

Como trabajo futuro, se plantea la mejora del método de entrenamiento de la Gating Network, con variantes del algoritmo genético u otra forma de entrenamiento que permita manejar distintos niveles de confianza, por ejemplo, predicciones con distintos niveles de confianza, expresadas de manera difusa.

También es conveniente la aplicación de la predicción probabilista directamente a la producción eólica, ya que esta información es la interesante para el operador del sistema eléctrico y su incorporación a sistemas estocásticos de operación con una consecuente mejora de la eficiencia energética.

## 7. Bibliografía

- [1] "www.wwindea.org," WWEA, World Wind Energy Half Year Report 2012, 2012. [Online].
- [2] "Plan Energético de Canarias Revisión 2006-2015".
- [3] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice Hall, 1999.
- [4] P. Pinson and G. Kariniotakis, "Conditional Prediction Intervals of Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 25, no. 4, 2010.
- [5] Gneiting T. and A. Raftery, "Strictly proper scoring rules, prediction, and estimation.," *Journal of the American Statistical Association*, no. 102, pp. 359-378, 2007.
- [6] D. S. Wilks, *Statistical methods in Atmospheric Sciences*, Second ed., Academic Press, 2006.
- [7] T. Gneiting, F. Balabdaoui and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," no. 69, pp. 243-268, 2007.
- [8] P. Pinson, *Estimation of the uncertainty in wind power forecasting, PhD Dissertation*, Ecole des Mines de Paris, 2006.
- [9] P. Pinson, H. A. Nielsen, J. K. Moller, H. Madsen and G. N. Kariniotakis, "Non-parametric Probabilistic Forecasts of Wind Power: Required Properties and Evaluation," *Wind Energy*, no. 10, pp. 497-516, 2007.
- [10] J. Juban, L. Fugon and G. Karionitakis, "Probabilistic short-term wind power forecasting based on kernel density estimators".
- [11] R. Ranjan, *Combining and Evaluating Probabilistic Forecasts*, 2009.