

Grado en Estadística y Empresa  
2021-2022

*Trabajo Fin de Grado*

# “Predicción probabilística de la producción eólica en Cádiz”

---

Javier Cárdenas Serres

Tutor

Ricardo Aler Mur

## DETECCIÓN DEL PLAGIO

La Universidad utiliza el programa **Turnitin Feedback Studio** para comparar la originalidad del trabajo entregado por cada estudiante con millones de recursos electrónicos y detecta aquellas partes del texto copiadas y pegadas. Copiar o plagiar en un TFG es considerado una **Falta Grave**, y puede conllevar la expulsión definitiva de la Universidad.



Esta obra se encuentra sujeta a la licencia Creative Commons **Reconocimiento - No Comercial - Sin Obra Derivada**

## **RESUMEN**

A diferencia de las predicciones puntuales, medir la incertidumbre de las predicciones con predicciones probabilísticas permite mejorar la toma de decisiones en procesos como la generación de energía eólica. En este estudio se hará predicción probabilística de la producción de energía eólica en la región de Cádiz, durante 2015-2018, mediante la predicción de cuantiles e intervalos de predicción. Para ello se comparará la bondad de distintos modelos lineales y no lineales (MLPs y LSTMs) con el objetivo de determinar cuál de ellos es el que mejor se ajusta al problema.

## ÍNDICE GENERAL

1. INTRODUCCIÓN. . . . .	1
2. LA PRODUCCIÓN Y PREDICCIÓN DE ENERGÍA EÓLICA . . . . .	2
2.1. Energía eólica en España . . . . .	2
2.2. Modelo numérico para la predicción meteorológica . . . . .	3
3. PREDICCIÓN PROBABILÍSTICA . . . . .	5
3.1. Función de pérdida Smooth Pinball . . . . .	6
3.2. De las predicciones a la distribución e intervalos . . . . .	7
3.3. Métricas de evaluación. . . . .	8
3.3.1. Quantile Score . . . . .	8
3.3.2. Average Coverage Error . . . . .	8
3.3.3. Interval Score. . . . .	9
4. METODOLOGÍA Y FUENTE DE DATOS . . . . .	10
4.0.1. Variable a predecir . . . . .	10
4.0.2. Variables de entrada . . . . .	10
4.0.3. Preproceso . . . . .	14
4.0.4. Intervalos a predecir . . . . .	14
4.1. Metodología . . . . .	14
4.1.1. Regresión cuantílica . . . . .	15
4.1.2. Perceptrón multicapa . . . . .	15
4.1.3. Long short-term memory . . . . .	16
4.1.4. Horizontes de predicción . . . . .	18
5. RESULTADOS . . . . .	19
5.1. Convergencia en el entrenamiento . . . . .	19
5.2. Resultados en el conjunto de validación . . . . .	20
5.3. Optimización de hiperparámetros. . . . .	22
5.3.1. Número de capas ocultas . . . . .	22
5.3.2. Número de observaciones hacia atrás . . . . .	23

5.4. Otras pruebas de interés . . . . .	25
5.4.1. Smooth Pinball vs Pinball . . . . .	25
5.4.2. Spread vs Non Spread . . . . .	26
6. CONCLUSIONES . . . . .	28
BIBLIOGRAFÍA . . . . .	29

# 1. INTRODUCCIÓN

Desde que se implantó por primera vez un aerogenerador en 1984, la energía eólica ha ido aumentando su capacidad productiva, hasta ser en 2021 la principal fuente de electricidad en España (REE, 2021).

Sin embargo, a pesar de ser una fuente de energía limpia y sostenible, las corrientes de aire son variables y poco estables, haciendo que la producción de energía eólica sea difícil de predecir e impidiendo establecer una oferta óptima en la subastas de electricidad que se llevan a cabo diariamente en España.

El principal objetivo de este trabajo es desarrollar un algoritmo que permita hacer predicciones probabilísticas sobre la producción de energía eólica que, al contrario que la predicción puntual, permita obtener la distribución de la producción futura esperada, aportando más información sobre el valor esperado y mejorando la toma de decisiones respecto a la cantidad a ofertar por las productoras de energía eólica.

Si bien hay diferentes maneras de hacer predicciones probabilísticas, en este trabajo en concreto, nos centraremos en la predicción cuantílica, de la que podremos sacar intervalos y estimar los posibles rangos en los que se debería encontrar la variable de predicción. Estas predicciones de generación de electricidad se harán para toda la región de Cádiz.

Entre los objetivos del trabajo está el probar distintos modelos de predicción cuantílica, lineales y no lineales (perceptrones multicapa o MLP y redes recurrentes LSTM). Además, se quiere estudiar algunas cuestiones, tales como si versiones suavizadas de la función de coste para las redes MLP y LSTM funcionan mejor o si añadir información adicional sobre la incertidumbre de predicciones meteorológicas mejora la calidad de los intervalos de predicción.

El trabajo está estructurado de la siguiente manera: En 2 se analiza la evolución y el papel de la energía eólica en España, además del sistema de predicción de las variables meteorológicas y el organismo encargado. En 3 se hace una descripción de la predicción probabilística, así como las funciones de pérdida y las métricas utilizadas. En la siguiente sección 4 se detalla la fuente de datos y la metodología utilizada para este trabajo. En 5 se comparan los resultados obtenidos por los diferentes algoritmos utilizados y se exponen los experimentos realizados. Por último, en 6 se resumen las principales conclusiones del estudio y el trabajo futuro.

## 2. LA PRODUCCIÓN Y PREDICCIÓN DE ENERGÍA EÓLICA

En este apartado veremos algunas de las características más importantes de la generación de energía eólica y algunos de los métodos utilizados para sus predicciones.

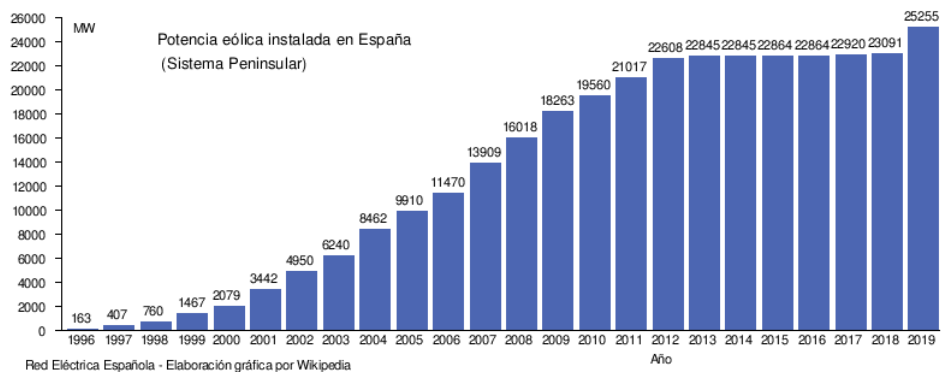
### 2.1. Energía eólica en España

La energía eólica es aquella que se obtiene a partir de las corrientes de aire. Actualmente esto se consigue a través de los aerogeneradores, transformando la energía cinética en energía mecánica. Es, además, una energía limpia, sostenible y se encuentra dentro de la categoría de energías renovables.

Desde que la cooperativa Ecotècnia instalase el primer aerogenerador en España en 1984 (Cerrillo, 2014; Fernández Munguía, 2016) como posible alternativa a la energía nuclear hasta 2012, la eólica ha tenido un crecimiento constante, como se puede ver en la Figura 2.1.

**Figura 2.1**

*Evolución de la potencia eólica instalada*

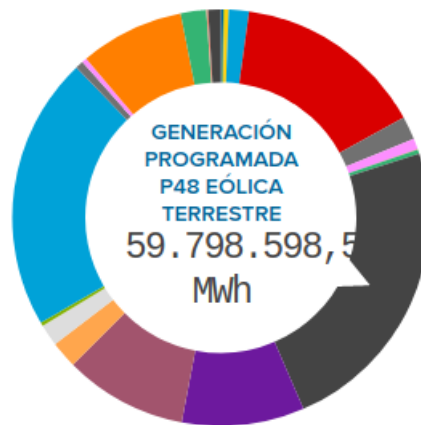


Fuente: Wikipedia

En 2021 la eólica pasó a ser la principal fuente de producción eléctrica en España con un porcentaje del 23 % sobre el total producido y representando mitad de la energía producida por las renovables, que fueron un 46 % en 2021, ver Figura 2.2 (REE, 2021).

**Figura 2.2**

*Generación de electricidad eólica respecto al resto de fuentes en 2021*



## 2.2. Modelo numérico para la predicción meteorológica

Una de las mayores ventajas de la energía eólica es que se trata de una energía renovable, no contaminante y además contribuye al desarrollo sostenible, reduciendo así el uso de combustibles fósiles.

Sin embargo, y a pesar de ser actualmente la principal fuente de energía en España, su producción es bastante impredecible e irregular ya que depende de las características temporales del momento y estas a su vez son muy difíciles de estimar. Esto afecta directamente en la toma de decisiones a la hora de saber cuánta cantidad de energía ofertar por las productoras de energía eólica en las subastas de electricidad que se llevan a cabo diariamente en España (OMIE, [s.f.](#)).

La única alternativa actual es tener buenos pronósticos meteorológicos y poder cubrir imprevistos de producción de energía con otras fuentes de respaldo (por ejemplo ciclos combinados en gas) en el caso que fuese necesario.

El matemático Lewis Fry Richardson fue la primera persona en proponer un modelo numérico para la predicción meteorológica (NWP, por sus siglas en inglés) en 1922 (Richardson, [1922](#)). No obstante, no fue hasta el inicio de los ordenadores en 1950 cuando se pudo hacer por primera vez las primeras predicciones meteorológicas utilizando métodos numéricos simples (Kalnay, [2003](#)).

Desde entonces, el avance de los ordenadores ha permitido que se mejoren las técnicas de predicción de variables meteorológicas y con ellas la precisión de dichas predicciones. Las previsiones meteorológicas son calculadas a partir de ecuaciones matemáticas de la física. Dada la complejidad de este sistema dinámico, las ecuaciones son no lineales y solo permiten sacar predicciones aproximadas de la realidad para breves lapsos de tiempo en el futuro. Existen diferentes tipos de modelos según el problema (modelos para planos

verticales o planos horizontales, globales o regionales, etc). Una vez se ha especificado el modelo a utilizar, se introducen las variables iniciales, normalmente son datos sacados de mediciones atmosféricas, y con ello predicen el siguiente estado, con el que luego van iterando  $n$  veces de manera recursiva en el tiempo hasta tener una previsión para un intervalo dado (Wiegand, 2015).

Como explicaremos en más detalle en 4, utilizaremos como fuente de información las predicciones meteorológicas del Centro Europeo de Previsiones Meteorológicas a Plazo Medio (ECMWF, por sus siglas en inglés).



### 3. PREDICCIÓN PROBABILÍSTICA

En esta sección se definirá la predicción probabilística y se detallarán algunos de las ventajas más importantes respecto a las predicciones puntuales.

La predicción puntual predice una única cantidad condicionada a los valores de los predictores escogidos. Como ejemplo un modelo construido con la fórmula habitual  $y = \beta_1 X_1 + \beta_2 X_2 + c$  está construido para predecir la media de la variable dependiente dados los predictores  $x_1$  y  $x_2$ . En cambio, en predicción probabilística, el objetivo es dar con la distribución de  $y$  (en la que lleva de manera implícita su media) a partir de los predictores  $x_1$  y  $x_2$ .

La predicción probabilista aporta más información permitiendo obtener la distribución de la producción futura y mejorando la toma de decisiones respecto a la cantidad a ofertar por las productoras de energía (Hatalis et al., 2017; Zhang et al., 2014).

Si bien existen diferentes maneras de hacer predicciones probabilísticas, en este trabajo nos centraremos en la predicción cuantílica, la cual estima cuantiles elegidos de la distribución de probabilidad de la variable dependiente, sin necesidad de suponer la forma de dicha distribución de probabilidad.

Fue en 1978 cuando, en busca de una alternativa robusta al estimador de mínimos cuadrados, Koenker y Bassett desarrollaron una nueva regresión lineal que tuviese en cuenta, en vez de la media, la mediana por ser un estimador robusto y no dependiente de los valores extremos que pueda existir en una muestra (Koenker y Bassett Jr, 1978).

A esta nueva regresión la llamaron la regresión cuantílica. La regresión cuantílica tiene como función dado el error 3.1, minimizar la función 3.2.

$$\xi = y - \hat{y} \quad (3.1)$$

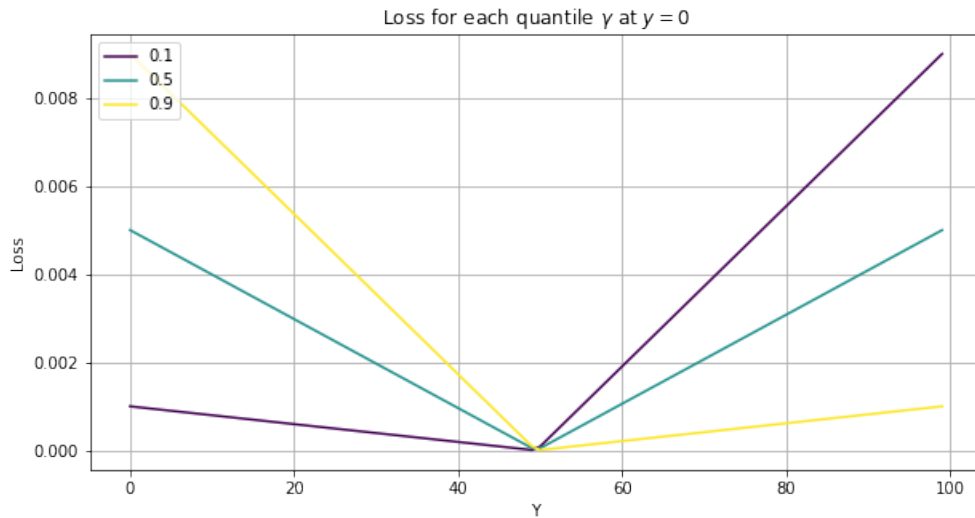
$$p_\tau(\xi) = \begin{cases} \tau\xi & \text{si } \xi \geq 0, \\ (\tau - 1)\xi & \text{si } \xi < 0. \end{cases} \quad (3.2)$$

Donde  $\tau$  es el cuantil que se quiere predecir. La ecuación 3.2, también se la suele llamar función de coste Pinball.

Como podemos ver en la Figura 3.1, la función penaliza más cuando el cuantil alto/bajo predicho se encuentra por debajo/encima de la variable observada para  $y = 0$  y los cuantiles 0,1 y 0,9. Si el modelo ha resultado efectivo, deberíamos encontrar el mismo número de variables observadas por encima del cuantil 0.5 que por debajo.

**Figura 3.1**

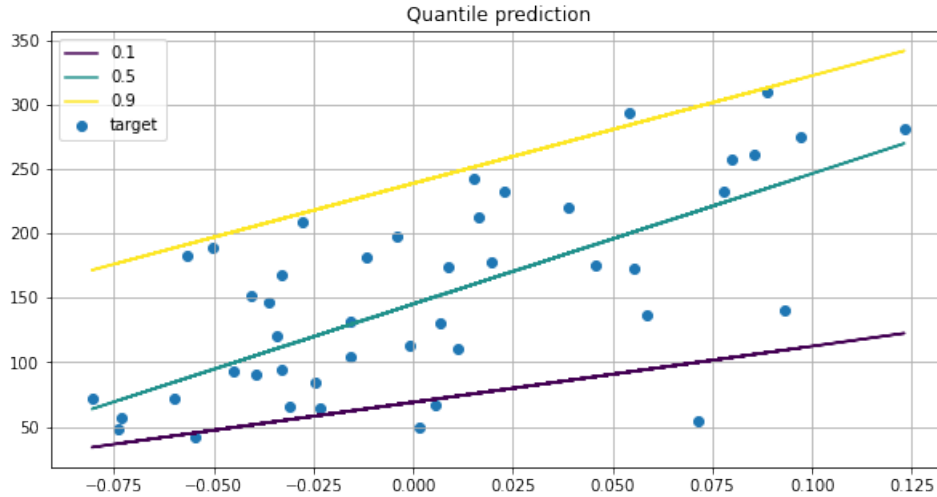
*Función de pérdida Pinball*



Y en la Figura 3.2 se muestra un ejemplo de cómo quedaría una predicción de la regresión cuantílica para los cuantiles 0,1, 0,5 y 0,9.

**Figura 3.2**

*Ejemplo de la predicción de cuantiles*



### 3.1. Función de pérdida Smooth Pinball

Uno de los problemas de la función de pérdida utilizada para la regresión cuantílica es que no es derivable en 0 y por tanto no es muy adecuada en modelos que utilizan el descenso de gradiente para optimizar los modelos, como por ejemplo los pesos de las redes neuronales. Por esta razón en Hatalis et al., [2017](#) se desarrolla una nueva función de coste que permite ser utilizada en modelos que utilicen el descenso de gradiente. Esta función de coste será la que utilicemos en los modelos del estudio.

La función de coste Smooth Pinball quedaría definida como:

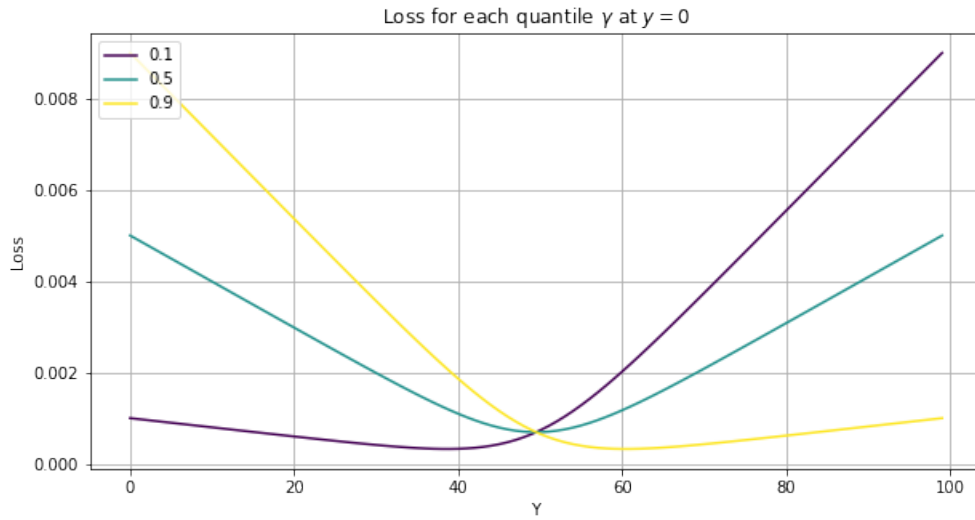
$$S_{\tau,\alpha}(\xi) = \tau\xi + \alpha \log(1 + e^{-\frac{\xi}{\alpha}}) \quad (3.3)$$

Donde  $\alpha > 0$  es el parámetro suavizado.

En la Figura 3.3 se muestra un ejemplo del coste obtenido por un cuantil y una predicción dada.

**Figura 3.3**

*Función de pérdida Smooth Pinball*

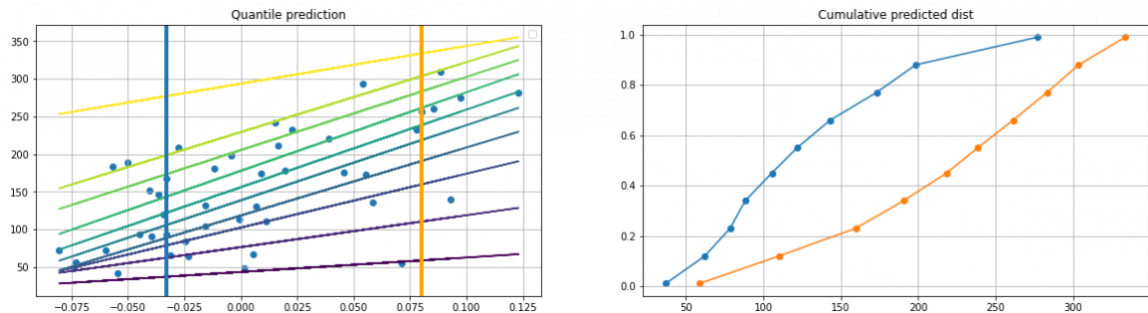


### 3.2. De las predicciones a la distribución e intervalos

Juntando las predicciones cuantílicas podemos generar una función de distribución acumulada de la variable a predecir como aparece en la Figura 3.4. Además, una vez tenemos los cuantiles, podremos generar intervalos de predicción. Como ejemplo, si predecimos el cuantil 0.025 y 0.975 tendremos una cota inferior y una cota superior donde deberían encontrarse los valores reales de la variable dependiente el 95 % de las veces.

**Figura 3.4**

*Función de distribución acumulada a partir de las predicciones de cuantiles*



### 3.3. Métricas de evaluación

Para las predicciones de los modelos será importante tener una serie de métricas que permitan evaluar correctamente cada modelo. Para ello utilizaremos métricas que no solo evalúen las estimaciones de los cuantiles sino también los intervalos predichos a partir de esos cuantiles. Las métricas utilizadas son las mismas que en el estudio Hatalis et al., 2017. Estas son:

#### 3.3.1. Quantile Score

Esta primera métrica permite evaluar la bondad de las estimaciones de los cuantiles. Utilizando la función de pérdida Pinball para todas las muestras obtenemos el quantile score (QS) y será la métrica principal que utilizaremos para evaluar los modelos. Cuanto menor sea esta métrica, mejores serán las predicciones.

$$QS = \sum_{t=1}^N \sum_{m=1}^M p_{\tau_m}(y_t - \hat{q}_t^{\tau_m}) \quad (3.4)$$

Donde  $N$  es el número de observaciones,  $M$  el número de cuantiles, y la variable observada,  $p_{\tau}$  el valor obtenido en la función smooth pinball para ese cuantil dado y  $\hat{q}_t^{\tau}$  el valor predicho para ese cuantil.

Por tanto la métrica  $QS$  nos da el resultado medio de la función smooth pinball para todos los cuantiles predichos y todas las variables observadas.

#### 3.3.2. Average Coverage Error

Por otro lado, también necesitaremos evaluar la bondad de los intervalos creados a partir de las predicciones de los cuantiles. Para ello se utilizarán tres métricas ACE, IS y Sharpness.

La métrica ACE, mide cuan certeros son los intervalos predichos según las variables observadas. Como ejemplo, si utilizásemos un intervalo del 80 %, idealmente querríamos que los datos observados estuvieran fuera de dicho intervalo un 20 % de las veces.

A continuación se muestra cual es la función que determina dicha métrica. Cuanto más cerca de 0 esté el valor de la métrica ACE, mejores serán los resultados del modelo.

$$PICP_i = \frac{1}{N} \sum_{t=1}^N \mathbb{1}_{y_t \in I_t^{\beta_i}(x_t)} \quad (3.5)$$

$$ACE = 100 \sum_{i=1}^{\frac{M}{2}} |PICP_i - (1 - \beta_i)| \quad (3.6)$$

Donde  $PICP_i$  es el *prediction interval coverage probability*,  $N$  es el número de observaciones,  $M$  el número de cuantiles,  $I_t^\beta$  es el intervalo de predicción  $[l_t^\beta, u_t^\beta]$  y  $(1 - \beta) = \tau_u - \tau_l = PINC$  es la cobertura nominal del intervalo de predicción, es decir el total que teóricamente queríamos cubrir.

De esta manera podemos ver que  $PICP$  nos da el ratio del número de veces que la variable dependiente observada se encuentra en el intervalo predicho entre el número de variables observadas, por tanto  $PICP \in [0, 1]$ .

Por otro lado la métrica  $ACE$  nos muestra la distancia porcentual media entre el número de veces que la variable observada aparece fuera del intervalo menos el número esperado de veces que debería aparecer según ese intervalo, para todos los intervalos de predicción estimados.

### 3.3.3. Interval Score

El problema de las métricas de fiabilidad es que en ocasiones se podrían obtener buenos resultados con las coberturas requeridas  $PINC$  simplemente haciendo los intervalos más anchos, no obstante se quiere que los intervalos de predicción cubran lo requerido y a la vez sean lo más estrechos posible. La métrica Interval Score (IS) y *Sharpness* permiten evaluar al modelo según la anchura de sus intervalos, dado que se quiere. A continuación en 3.7 y 3.8 se muestran las funciones necesarias para calcular dichas métricas.

$$Sharpness = \frac{2}{NM} \sum_{t=1}^N \sum_{i=1}^{\frac{M}{2}} (u_t^{\beta_i} - l_t^{\beta_i}) \quad (3.7)$$

$$IS = Sharpness + \frac{2}{\beta_i} (l_t^{\beta_i} - y_t) \mathbb{1}(y_t < l_t^{\beta_i}) + \frac{2}{\beta_i} (y_t - u_t^{\beta_i}) \mathbb{1}(y_t > u_t^{\beta_i}) \quad (3.8)$$

Donde  $N$  es el número de variables observadas en la muestra,  $M$  el total de cuantiles y por tanto  $\frac{M}{2}$  el número de intervalos a predecir,  $l_t^{\beta_i}$  y  $u_t^{\beta_i}$  representan el valor predicho para el cuantil inferior y superior del intervalo  $i$ , respectivamente.

Por tanto *sharpness* es la anchura media de los intervalos de predicción y  $IS$  es la anchura media de los intervalos de predicción más una penalización cuando el valor observado está fuera del intervalo de predicción (por encima o por debajo del intervalo) que además es proporcional a la distancia que el valor esté del límite del intervalo.

Si bien, a la hora de evaluar los resultados nos centraremos en la métrica  $IS$  más que en el *sharpness*, se incluirá también esta última en las tablas a modo informativo. Por orden de relevancia, las métricas más importantes serán  $QS$ ,  $IS$ ,  $ACE$  y *Sharpness* y este será el orden que siga la evaluación y comparación de los resultados entre los distintos modelos en la Sección 5.

## 4. METODOLOGÍA Y FUENTE DE DATOS

En esta sección se detallará la información y la fuente de los datos a utilizar en este estudio. Además se describe el preproceso utilizado para limpiar los datos hasta que son introducidos dentro de los modelos. Por último, en esta sección también veremos los diferentes modelos que se van a utilizar para hacer las predicciones probabilísticas.

### 4.0.1. Variable a predecir

La variable a predecir es la generación de energía eólica horaria en la región de Cádiz desde 2015 a 2018. La razón de utilizar este intervalo es que, durante este periodo, la capacidad instalada de eólica se mantuvo constante y por tanto no es necesario hacer ningún tipo de ajuste en los datos.

ESIOS es el sistema de información desarrollado por Red Eléctrica que contiene la información relevante sobre el mercado eléctrico en España y de donde sacaremos los datos de la variable a predecir.

### 4.0.2. Variables de entrada

Los predictores se han sacado del Centro Europeo de Previsiones Meteorológicas (ECMWF, por sus siglas en inglés) y son:

- *10 y 100m u-component of wind*: La velocidad, en metros por segundo, del movimiento del aire en dirección este a 10 y 100m sobre la superficie terrestre.
- *10 y 100m v-component of wind*: La velocidad, en metros por segundo, del movimiento del aire en dirección norte a 10 y 100m sobre la superficie terrestre.
- *2m temperature*: Es la temperatura del aire a dos metros sobre el nivel del mar.
- *Surface pressure*: La presión de la atmósfera en la superficie terrestre.
- *Norma de la velocidad a 10m y 100m*: Es una variable calculada a partir de las variables u-component y v-component of wind. Se calcula de la siguiente forma, donde  $i = 10, 100$ :

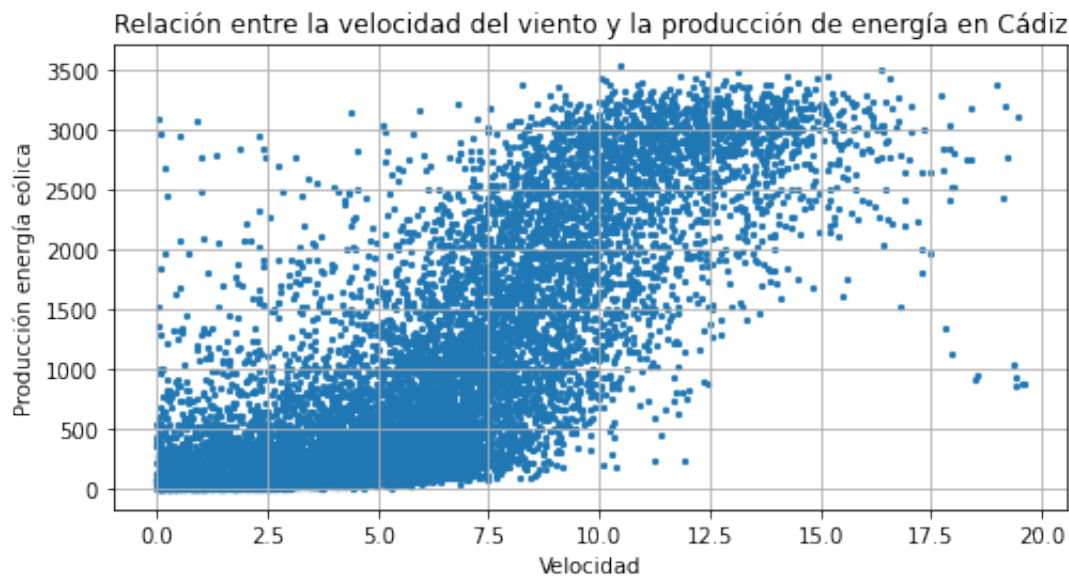
$$n_i = \sqrt{u_i^2 + v_i^2}$$

- *Hora, día y mes*: Se han añadido 3 variables temporales por que la información temporal puede tener influencia en el comportamiento del viento.

En la Figura 4.1 se muestra la relación entre la velocidad del viento y la producción de energía eólica. Es una gráfica muy común de encontrar en estudios relacionados con la producción de energía eólica (Hernandez et al., 2016; Norbu et al., 2021) y nos muestra que existe una relación no lineal entre la velocidad del viento y la producción de energía eólica. Esto es, llega un momento que aunque sople más el viento los aerogeneradores no tienen mayor capacidad productiva (esto es debido a que los aerogeneradores se desconectan a velocidades altas para evitar daños), lo mismo cuando hay muy poco viento, donde da igual que el viento sople ligeramente o no lo haga en absoluto.

**Figura 4.1**

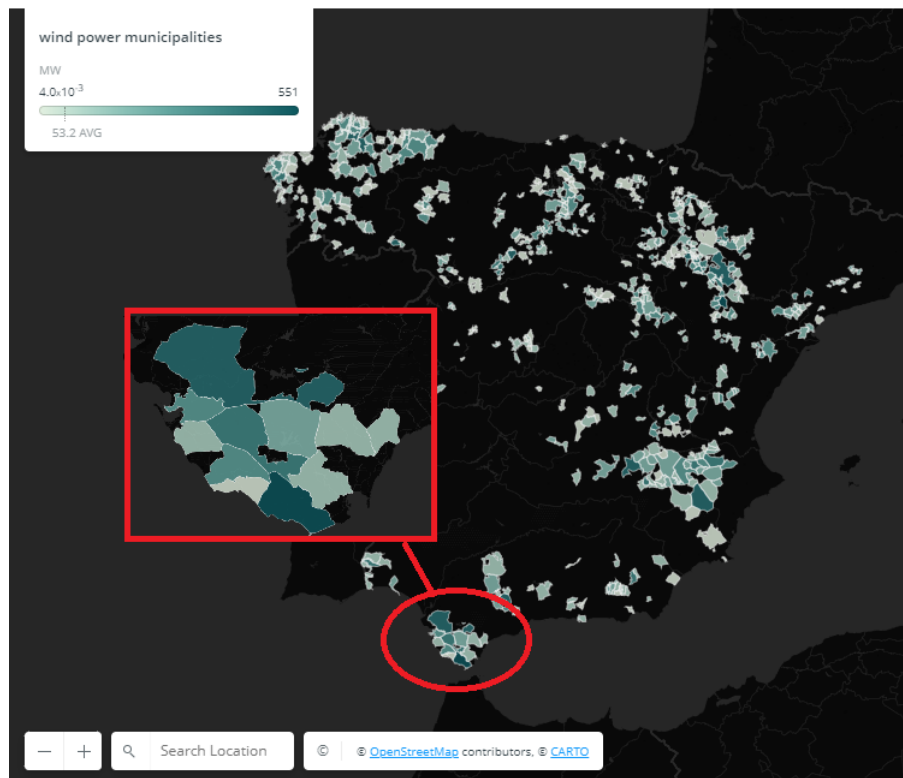
*Relación entre la velocidad del viento y la energía producida*



ECMWF nos da los valores (reales y estimados) de estas variables por cuadrantes de 30km. Teniendo esto en cuenta, para la región de Cádiz hay 9 cuadrantes por lo que tendremos 9 estimaciones de las variables mencionadas para cada intervalo de tiempo. Esto hace un total de 75 variables de entrada  $((9 \times 8) + 3 = 75)$ . Las coordenadas introducidas en ECMWF son: [37.1, -6.45, 36,-5] (Norte, oeste, sur y este, respectivamente). En la figura 4.2 se muestra una imagen de la región concreta a estudiar, así como una representación del número de instalaciones de aerogeneradores en la región.

**Figura 4.2**

*Instalaciones eólicas por municipios*



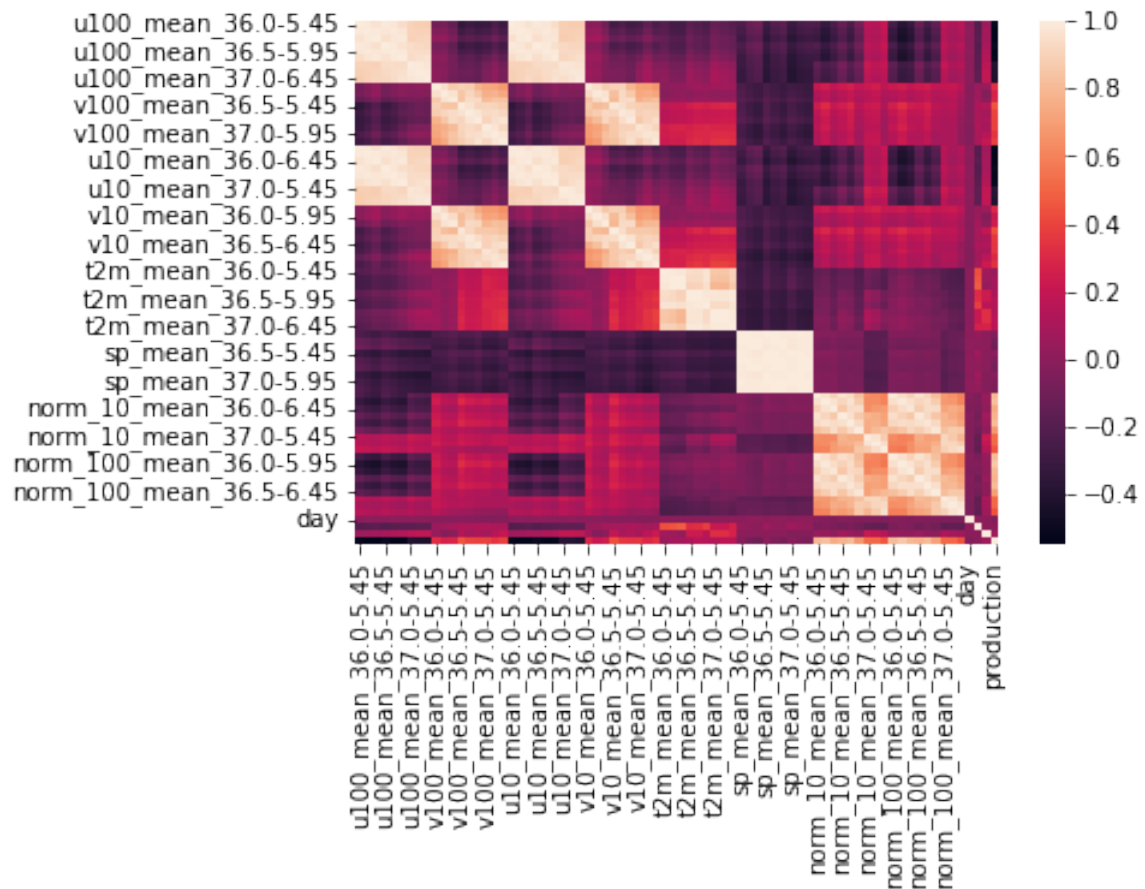
Fuente: ESIOS

En la Figura 4.3 se muestra un mapa de correlaciones entre los predictores a utilizar en el estudio. Como podemos ver, las variables de presión en la superficie están negativamente correlacionadas con las velocidades del viento. También vemos que las velocidades del viento en dirección norte y las velocidades del viento en dirección muestran una correlación negativa.



**Figura 4.3**

*Gráfica de correlaciones entre las variables de entrada*



## Mean y Spread

Recordemos que las variables que se encuentran en el ECMWF son el resultado de un ensemble de predicciones utilizando el modelo físico matemático, y dentro de este ensemble tenemos una predicción básica (utilizando los datos iniciales obtenidos con los medidores atmosféricos) y el resto predicciones obtenidas añadiendo ligeras perturbaciones a los datos iniciales. De esta manera tenemos un ensemble de modelos donde podemos comparar la variabilidad de las predicciones según ligeros cambios en las variables iniciales(ECMWF, [s.f.](#)).

Actualmente ECMWF permite obtener tanto la media como la desviación de las predicciones de sus variables meteorológicas.

Las 75 variables mencionadas anteriormente solo son asumiendo que estamos utilizando la media de las predicciones. No obstante, dado que se quiere hacer predicción probabilística, se podría pensar en añadirse también las variables de desviación, lo que añadiría 36 variables nuevas ( $9 * 6 = 36$ , quitando las dos normas) al total de variables de entrada.

En principio la combinación de ambos predictores (tanto la media como la desviación

de la combinación de los modelos) parece ser interesante, así que se ha decidido en este trabajo también dedicar un apartado (ver Sección 5.4.2) a la comprobación y al análisis de la bondad de esta posibilidad.

#### 4.0.3. Preproceso

Como hemos mencionado anteriormente, los datos de producción eólica sacados de ESIOS son cada hora, mientras que las variables que se utilizan para predecir el viento son cada tres. Por este motivo, se han agrupado los datos de la producción eólica de ESIOS cada tres horas, sumando durante ese intervalo toda la energía generada.

Una vez hemos sincronizado las variables de entrada con la variable a predecir, haremos una separación entre los datos de entrenamiento, los datos de validación y los datos de test (80 %/10 %/10 % respectivamente).

Cabe mencionar que se ha utilizado la función *MinMaxScaler* de *Scikit-learn* (ver 4.1) para que los datos se encuentre entre 0 y 1 y así acelerar y mejorar el entrenamiento de la red (Huang et al., 2020).

$$X^* = \frac{X - \text{mín}(X)}{\text{máx}(X) - \text{mín}(X)} \quad (4.1)$$

#### 4.0.4. Intervalos a predecir

Si bien se han probado a predecir diferentes rangos, finalmente en este trabajo se van a predecir los siguientes intervalos con coberturas de: 95 %, 90 %, 80 % y 70 %

Para hacer la predicción de dichos intervalos, se van a predecir los siguiente cuantiles: 0,025; 0,05; 0,1; 0,15; 0,85; 0,9; 0,95 y 0,975.

### 4.1. Metodología

En este apartado detallaremos los algoritmos utilizados para hacer las predicciones.

De menor a mayor complejidad tendremos los siguientes modelos:

- **Regresión cuantílica:** este será nuestro algoritmo de referencia y veremos si podemos mejorarlo. Se trata de un modelo simple, lineal. A partir de este modelo podremos ver si el aumento en la complejidad de los algoritmos es justificable.
- **Perceptrón multicapa:** Como primera acercamiento a las redes neuronales y a algoritmos no lineales, se probará una red con múltiples capas.
- **LSTM:** Por último, se utilizará una red recurrente, en concreto la LSTM por su idoneidad para el problema a resolver.

Por simplicidad se ha probado primero una única versión de cada algoritmo (sin hacer ningún cambio en los hiperparámetros). Después de comparar los resultados entre los tres tipos de modelo se ha procedido a hacer la optimización de los hiperparámetros del mejor de ellos.

#### 4.1.1. Regresión cuantílica

Como hemos mencionado anteriormente, la regresión cuantílica fue desarrollada por Koenker y Basset con el objetivo de evitar hacer asunciones sobre la distribución de los errores de una regresión lineal (Koenker y Bassett Jr, 1978).

Este modelo será el más sencillo que utilicemos y lo utilizaremos como base para comparar los modelos más complejos y evaluar si finalmente el aumento de complejidad en el modelo está justificado y respaldado por una mejora en las métricas de evaluación con respecto al modelo de regresión lineal cuantílica.

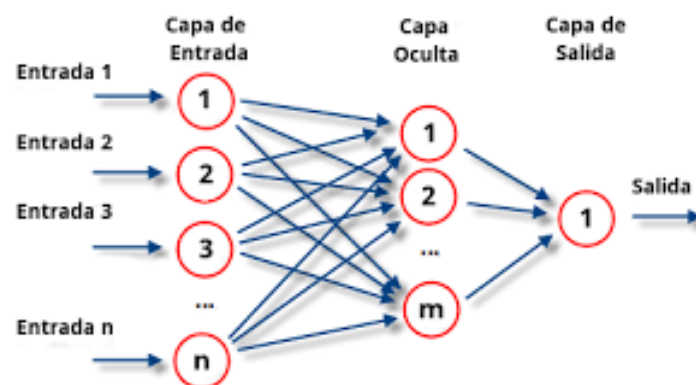
Al igual que los siguientes dos modelos, se ha desarrollado este algoritmo en *PyTorch* y se ha utilizado el descenso de gradiente para conseguir los pesos óptimos.

#### 4.1.2. Perceptrón multicapa

El perceptrón multicapa es un tipo de red neuronal formado al menos por tres capas: La capa de entrada, la capa oculta y la capa de salida, como se puede venir en la siguiente Figura:

**Figura 4.4**

*Ejemplo de estructura de un perceptrón multicapa*



Fuente: Wikipedia

A diferencia del modelo anterior, el perceptrón multicapa es un algoritmo más complejo y puede utilizarse para datos no lineales que, como hemos visto en la Figura 4.1, podría ser el caso.

Este algoritmo tiene los siguientes hiperparámetros:

- Número de capas ocultas: Para este trabajo se ha fijado a 2.
- Número de neuronas por capa: El número de neuronas por cada capa serán 50 y 20 respectivamente. La razón de que se haya fijado a 50 el número de neuronas en la primera capa oculta, teniendo en cuenta que las variables de entrada son 75 es para ver si la red es capaz de reducir las dimensiones sin perder la información y al mismo tiempo evitando tener que hacer un PCA o un algoritmo similar para reducir la dimensionalidad de las variables de entrada.
- Optimizador a utilizar: Se ha utilizado el algoritmo de *Adam*, con un learning rate de 0.001. Y utilizando el resto de parámetros que da por defecto la función de optimización de *PyTorch*.

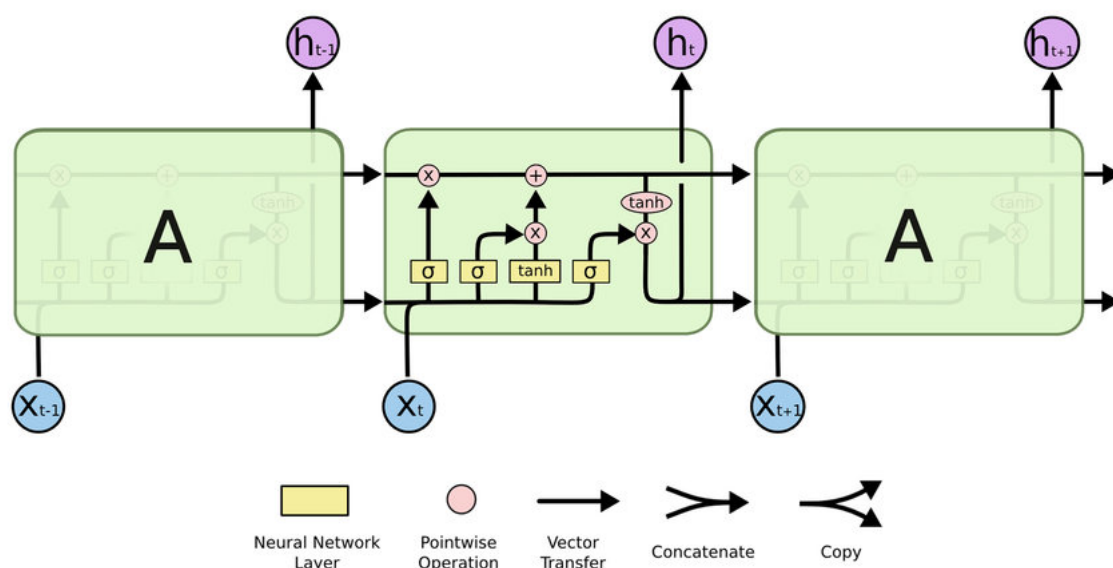
#### 4.1.3. Long short-term memory

Por último utilizaremos un algoritmo más complejo, la LSTM. Las LSTM forman parte del conjunto de redes neuronales recurrentes. Fueron introducidas por primera vez por Hochreiter y Schmidhuber en 1997 (Hochreiter y Schmidhuber, 1997). Este tipo de red recurrente fue desarrollada para resolver el problema que tenía las redes de este tipo con el desvanecimiento o la explosión del gradiente en la retropropagación. Y, al contrario que sus homólogas anteriores, funciona bien en series temporales donde la relación entre variables se da en periodos temporales desconocidos.

Se ha decidido introducir este algoritmo en el trabajo dada su capacidad de aprender relaciones a largo plazo, especialmente en problemas de predicción secuencial.

**Figura 4.5**

*Ejemplo de estructura de una LSTM*



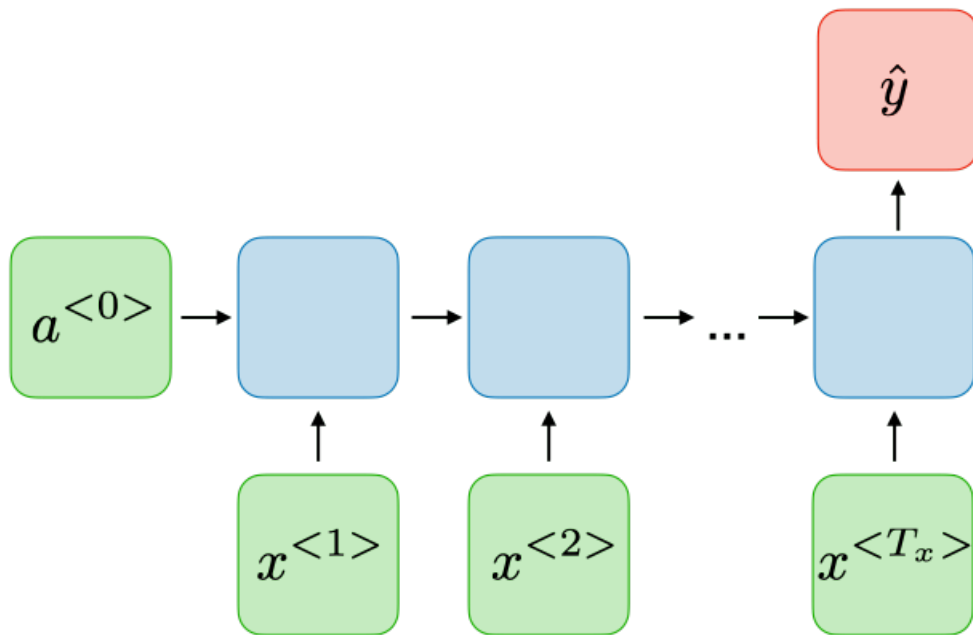
Fuente: Colah's web

El enfoque que se ha realizado en este estudio es el many-to-one Thakur, 2022, donde el modelo recibe una secuencia de datos (en este caso un histórico de los últimos  $t$  días de los  $n$  predictores) y predice una única salida (en este caso será el/los cuantiles a predecir).

En la Figura 4.6 se muestra un ejemplo de este enfoque, donde  $x^i$  representa un vector con el número de predictores del día  $i$ .

**Figura 4.6**

*Ejemplo de enfoque many-to-one*



Fuente: Stanford.edu

Como en los algoritmos anteriores, en la LSTM, por simplicidad, solo se probará una única combinación de hiperparámetros y en el caso de que sea el algoritmo que mejores resultados obtenga, se hará una optimización de hiperparámetros.

- Número de capas ocultas: Para este trabajo se ha fijado a 2.
- Número de estados ocultos: Se ha fijado al mismo número que las variables de entrada.
- Optimizador a utilizar: Se ha utilizado el algoritmo de *Adam*, con un learning rate de 0.0001 y las mismas betas que se ponen por defecto en *PyTorch*.
- Número de observaciones hacia atrás: Otra de las variables importantes a tener en cuenta en la LSTM es el número de datos que le pasamos hacia atrás. En la primera prueba se ha decidido fijar a 6, esto es, utilizará los últimos 6 datos observados de cada uno de los inputs para predecir la producción. Como nuestros datos son cada 3 horas, miraremos los valores de las variables meteorológicas 18h hacia atrás.

#### 4.1.4. Horizontes de predicción

Como hemos mencionado, las variables predictoras son en si mismas variable que ECMWF predice el día anterior.

Los modelos que acabamos de mencionar estimarán en  $t_0$  cada día a las 00:00 varios cuantiles para varios horizontes futuros ( $t_0 + 3h, t_0 + 6h, \dots$ ) usando como variables predictoras variables meteorológicas predichas para esos mismos horizontes, ver 4.2.

$$q_{t_0+3h} = f(X_{1,t_0+3h}, X_{2,t_0+3h}, \dots, X_{n,t_0+3h}) \quad (4.2)$$

Donde  $X_{i,t}$  es el predictor  $i$  (variable meteorológica predicha por ECMWF) para el tiempo futuro  $t$ .

## 5. RESULTADOS

En esta sección se entrenarán los modelos mencionados (lineal, MLP y LSTM) (sección 5.1), se analizarán sus resultados en el conjunto de validación para elegir un modelo y ajustar sus hiper-parámetros (secciones 5.2 y 5.23). Finalmente, se realizarán algunos experimentos de interés (sección 5.4, para comprobar si la sustitución de la función de coste normal por una suavizada mejora los resultados (pinball loss vs. smooth pinball loss) o si añadir información adicional proporcionada por las predicciones meteorológicas sobre su incertidumbre, mejora la calidad de los intervalos de predicción.

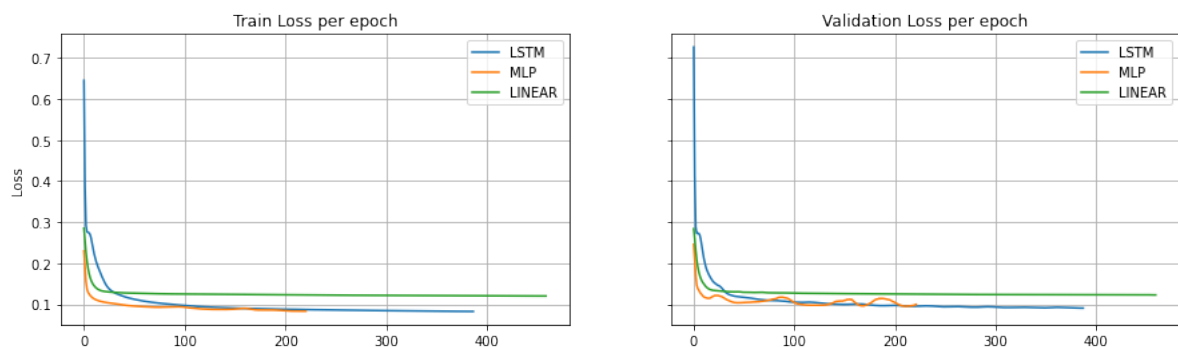
### 5.1. Convergencia en el entrenamiento

En la Figura 5.1 se muestran los resultados obtenidos por los modelos en el entrenamiento (gráfica superior) y en el conjunto de validación (imagen inferior). El eje Y representa la pérdida obtenida por el modelo en cada iteración (eje X). Las tres líneas representan los tres modelos utilizados mencionados en la Sección 4.1, estos son: LSTM, MLP (perceptrón multicapa) y LINEAR (regresión cuantílica).

Los tres modelos convergen en el entrenamiento. Como podemos ver, dependiendo del modelo utilizado el número de iteraciones durante el entrenamiento es mayor o menor. Esto se debe a que se ha utilizado la técnica de *early stopping* (Prechelt, 1998). De esta manera, el entrenamiento se para en el momento en el que el coste en el conjunto de validación empieza a aumentar durante un número determinado de epochs (que se ha fijado a 100).

**Figura 5.1**

*Evolución de coste por cada iteración en el conjunto de entrenamiento y validación*



Otra de las conclusiones que podemos sacar de la gráfica anterior, es que el entrenamiento los algoritmos MLP y LSTM parecen obtener mejores resultados que la regresión cuantílica, que se estabiliza en un coste superior, indicando que pueden existir relaciones

no lineales entre las variables de entrada y la variable de salida, como se vio en la figura 4.1.

## 5.2. Resultados en el conjunto de validación

A continuación se muestran los resultados obtenidos en el conjunto de validación. Antes de comenzar a analizar los resultados, cabe destacar que durante esta prueba no se hacen comprobaciones en el conjunto de test ya que serán las últimas en hacerse cuando hayamos concluido con qué modelo quedarnos.

Como podemos ver en la Tabla 5.1, el modelo LSTM es el que mejores resultados obtiene en todas las métricas (exceptuando la métrica ACE). Esto no es de sorprender si tenemos en cuenta que se trata de un algoritmo específicamente diseñado para secuencias de datos y funciona muy bien en series temporales (Hochreiter y Schmidhuber, 1997).

La métrica *QS* nos indica cual es el modelo que tiene el error de predicción escalado por cuantil más bajo, en este caso es la LSTM. También, los resultados obtenidos en la métrica *IS* junto con la métrica *sharpness* indican que el modelo tiene los intervalos de predicción más estrechos y al mismo tiempo acierta más en dichos intervalos es la LSTM. Y por último viendo el valor del *ACE* podemos decir que el modelo que suele acercarse más a la cobertura nominal de la predicción del intervalo (PINC) es el MLP.

Por otro lado, en términos generales, siguiendo en la línea con lo que veíamos en el apartado anterior, los dos modelos no lineales (LSTM y MLP) obtienen mejores resultados en el conjunto de validación que la regresión cuantílica, indicando que, efectivamente, pueden existir relaciones no lineales entre los predictores y la variable dependiente.

Métrica	LSTM	MLP	Regresión cuantílica
QS	0.087	0.095	0.116
IS	0.302	0.328	0.391
Sharpness	0.197	0.202	0.306
ACE	1.549	0.895	1.460

**Tabla 5.1**

*Resultados de los modelos LSTM, MLP y Regresión cuantílica en el conjunto de validación*

A continuación en las Figuras 5.2, 5.3 y 5.4 se muestran las predicciones de los modelos en el conjunto de validación.

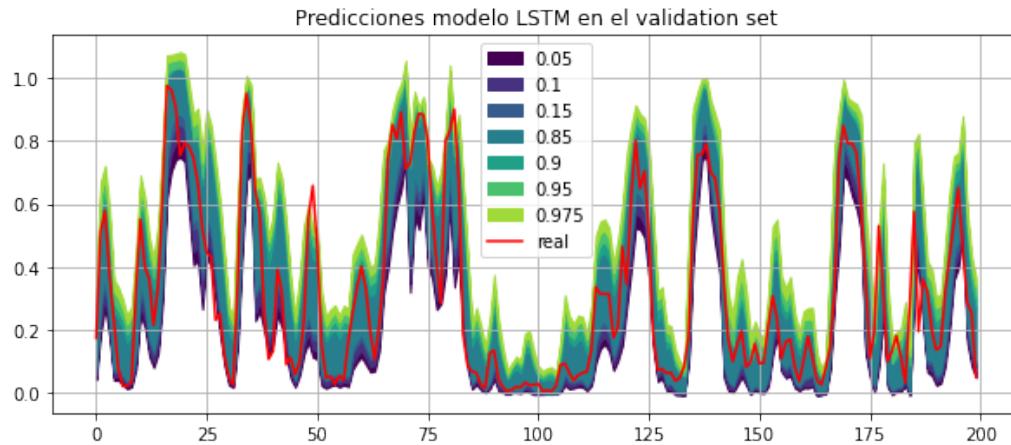
A modo de ejemplo en la primera Figura 5.2 se muestran las predicciones de los cuantiles del estudio. El intervalo de un cuantil a otro tiene un color distinto para poder diferenciar entre los distintos cuantiles. La línea roja, por otro lado, representa el valor observado de la variable dependiente. Continuando con este ejemplo, podríamos considerar que la LSTM hace buenas predicciones para el cuantil 0.9 si la variable observada (en rojo) es mayor que la variable predicha por el modelo (en verde) el 10 % de las veces.



Cabe destacar que para los tres modelos la variable observada (línea roja) suele encontrarse siempre dentro de los intervalos predichos, indicando un cierto grado de fiabilidad en los tres modelos y en concreto en la LSTM, que cuenta con los intervalos más estrechos entre los tres.

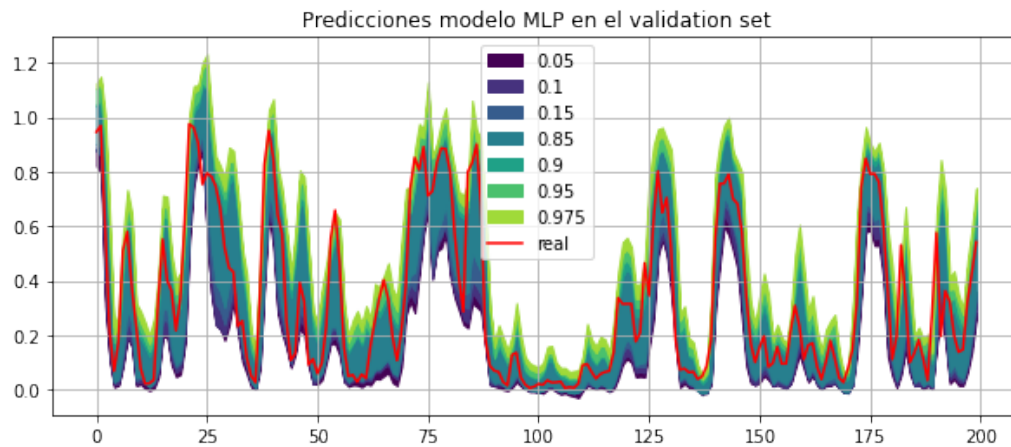
**Figura 5.2**

*Predicciones del modelo LSTM en el conjunto de validación*



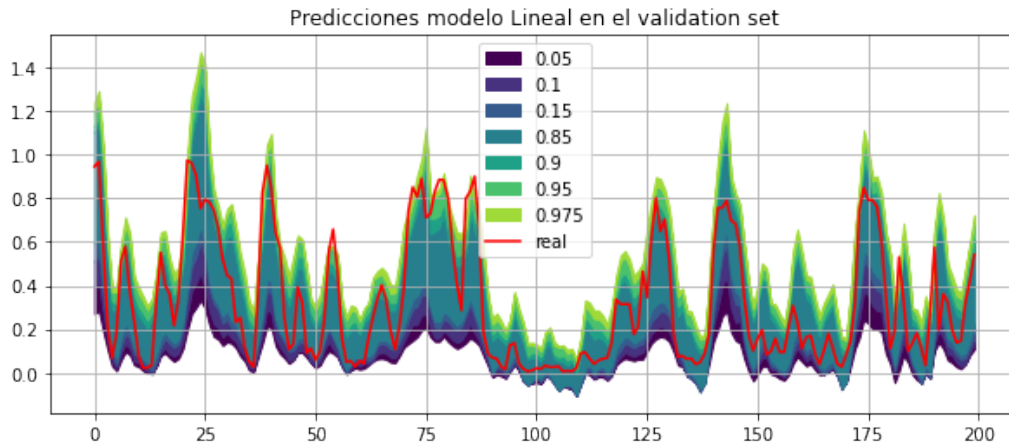
**Figura 5.3**

*Predicciones del modelo MLP en el conjunto de validación*



**Figura 5.4**

*Predicciones del modelo lineal en el conjunto de validación*



A la vista de los resultados, se ha decidido seguir el análisis con el modelo LSTM y realizar el ajuste de hiperparámetros únicamente para ese modelo.

### 5.3. Optimización de hiperparámetros

En este apartado haremos una optimización de los hiperparámetros de la red y veremos qué combinación de hiperparámetros es la que mejores resultados obtiene.

Debido al alto coste de entrenamiento, solo se van a tratar de optimizar dos parámetros:

- *El número de capas ocultas:* Se probará 1, 2 y 3 capas ocultas.
- *El número de observaciones hacia atrás:* 4, 6 y 8 observaciones. Esto observaría las 12h, 18h y 24h hacia atrás respectivamente.

#### 5.3.1. Número de capas ocultas

El número de capas ocultas de la LSTM aumenta proporcionalmente la complejidad del modelo y esto podría llevar a una mejora las predicciones de los intervalos de predicción. En el anterior análisis se había hecho la prueba utilizando por defecto 2 capas ocultas en la LSTM. Para este análisis utilizaremos tres modelos distintos uno con 1 capa, otro con 2 y un último con 3. De esta manera podremos ver si el aumento de complejidad está respaldado por una mejora en las predicciones o si por el contrario es mejor utilizar un modelo simple.

Como se muestra en la Tabla 5.2, el modelo que mejores métricas tiene (de nuevo, sin contrar con la métrica ACE) es la LSTM con dos capas ocultas, seguido del modelo con 3 capas ocultas y que tiene el mejor resultado en la métrica ACE. Esto nos indica que

no se obtiene ninguna mejora sustancial utilizando un modelo más complejo y, al mismo tiempo, se empeora si quisiéramos utilizar un modelo más simple.

Se puede ver, no obstante, como la LSTM con una capa obtiene unos resultados similares a los obtenidos con 3 capas.

Métrica	LSTM 1 capa	LSTM 2 capas	LSTM 3 capas
QS	0.090	0.083	0.086
IS	0.313	0.292	0.301
Sharpness	0.200	0.192	0.194
ACE	1.052	1.355	0.452

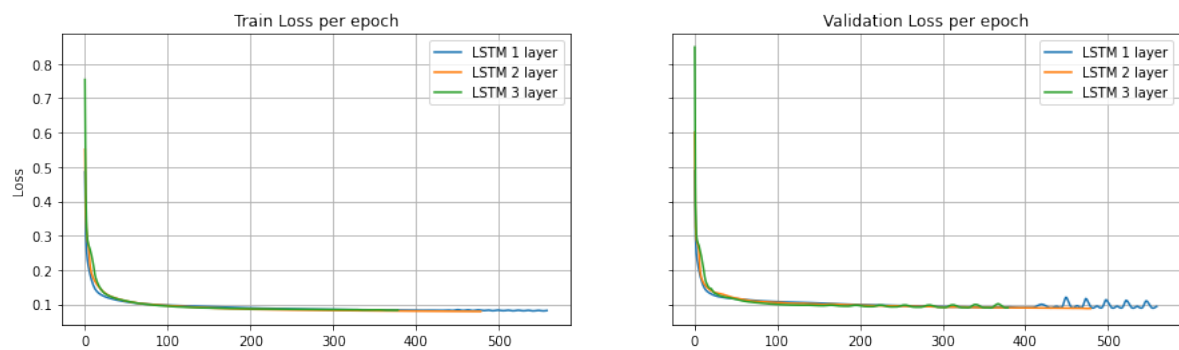
**Tabla 5.2**

*Resultados del modelo LSTM con 1, 2 y 3 capas en el conjunto de validación*

Además, en la Figura 5.5 vemos que la evolución del coste tanto en el conjunto de entrenamiento como en el de validación es muy parecida para las tres tipologías de LSTM.

**Figura 5.5**

*Evolución del coste con la LSTM con 1, 2 y 3 capas ocultas en el conjunto de entrenamiento y validación*



A la vista de los resultados, nos decantamos por utilizar una LSTM con 2 capas ocultas y será esta versión la que utilicemos para el siguiente apartado de optimización de hiperparámetros.

### 5.3.2. Número de observaciones hacia atrás

Una vez tenemos el número de capas que vamos a utilizar en la LSTM, también trataremos de optimizar el hiperparámetro *número de observaciones hacia atrás* que se le pasará al modelo.

Como hemos visto en el apartado de metodología, la LSTM tiene la capacidad de recibir una secuencia de los predictores para después hacer las predicciones. Esto significa que, a diferencia de los otros modelos que solo utilizan los predictores  $X_t$  en tiempo  $t$  para predecir la variable dependiente  $Y_{t+1}$  en  $t + 1$  (siendo  $X$  un vector de dimensión  $75 \times 1$ , la LSTM es capaz de introducir (sin aumentar el número de pesos)  $n$  observaciones hacia

atrás de los predictores, esto es  $(X_t, X_{t-1}, \dots, X_{t-n})$  para predecir la variable dependiente  $Y_{t+1}$ .

Teniendo esto en mente, en este apartado trataremos de analizar los resultados que se obtienen por el modelo utilizando 4, 6 y 8 observaciones hacia atrás, esto nos permitiría coger secuencias de 12h, 18h y 24h, respectivamente.

Como se muestra en la Tabla 5.3, los resultados son bastante parecidos entre utilizar 8, 6 o 4 observaciones hacia atrás (los datos de las 24h, 18h y 12h anteriores agrupados en tiempos de 3h, respectivamente). Por simplicidad, viendo que añadir más variables del pasado no aporta valor, se ha decidido utilizar únicamente 4 observaciones hacia atrás.

Métrica	LSTM 8 obs	LSTM 6 obs	LSTM 4 obs
QS	0.084	0.083	0.082
IS	0.296	0.294	0.292
Sharpness	0.184	0.180	0.180
ACE	1.690	2.634	1.025

**Tabla 5.3**

*Resultados del modelo LSTM con 4, 6 y 8 observaciones hacia atrás*

Nuestro modelo final por tanto será una LSTM con dos capas ocultas y que tenga 4 observaciones hacia atrás de los predictores.

En la siguiente Tabla 5.4 se muestran los resultados obtenidos por el modelo final. Como podemos ver, los resultados son similares a los observados en el conjunto de validación, confirmando a su vez que el modelo no está sobreajustado.

Métrica	LSTM 4 obs
QS	0.085
IS	0.293
Sharpness	0.192
ACE	1.123

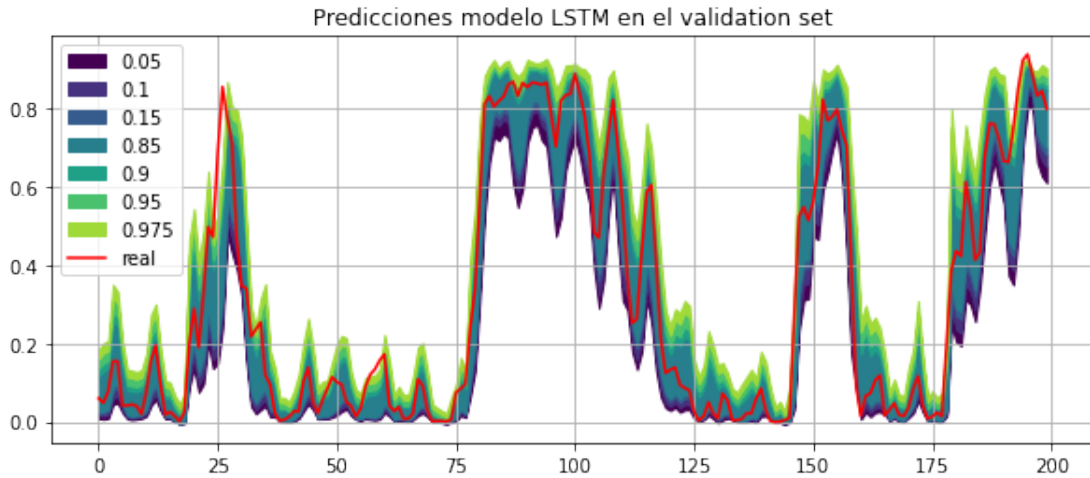
**Tabla 5.4**

*Resultados del modelo final LSTM con 4 observaciones hacia atrás*

En la siguiente Figura 5.6 se muestran las predicciones del modelo final (LSTM con dos capas y 4 observaciones hacia atrás) en el conjunto del test.

**Figura 5.6**

*Predicciones del modelo final en el conjunto del test*



#### 5.4. Otras pruebas de interés

Además de las pruebas mencionadas hasta ahora, se han realizado una serie de comprobaciones y análisis que se consideraba necesario para evaluar la bondad y la capacidad de los algoritmos propuestos.

##### 5.4.1. Smooth Pinball vs Pinball

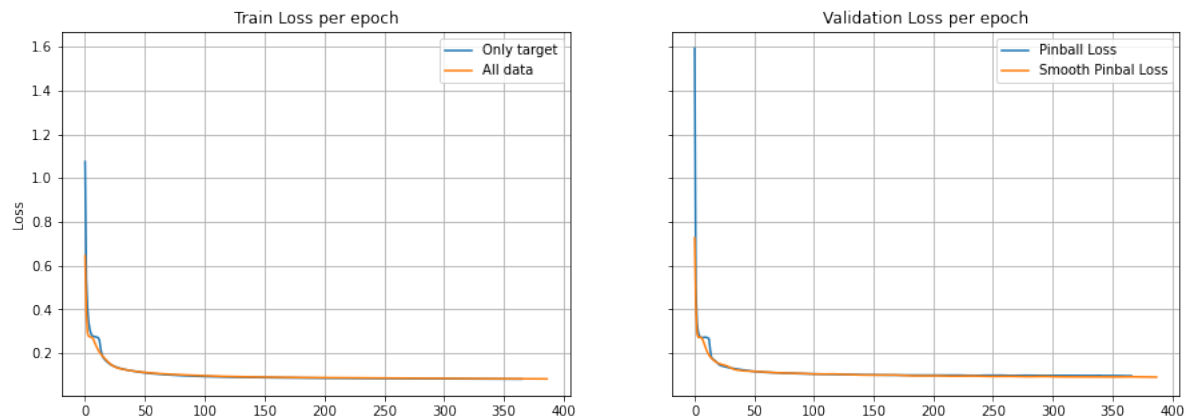
En un primer lugar, se ha tratado de comparar las diferencias en los resultados que puede ocasionar utilizar una función de pérdida u otra. Como se ha mencionado, la función de pérdida Pinball no es diferenciable en 0, por lo que podría dar problemas en el algoritmo de *backpropagation* a la hora de actualizar los pesos.

A continuación se muestran los resultados obtenidos por el algoritmo LSTM (se ha decidido utilizar únicamente este algoritmo porque era el que mejores resultados obtenía).

En primer lugar, de la Figura 5.7 podemos ver que, a pesar de que la evolución del coste obtenido el entrenamiento es muy similar utilizando ambas funciones, la convergencia de la red utilizando la función suavizada pinball es más estable.

**Figura 5.7**

*Evolución del coste del entrenamiento función Pinball vs Pinball suavizada*



Además, viendo la Tabla 5.5 podemos ver que la función Pinball suavizada obtiene mejores resultados en todas las métricas en comparación con su homóloga no diferenciable Pinball. A la vista de estos resultados y la gráfica del entrenamiento, podemos concluir que la función de coste Smooth Pinball mejora tanto el entrenamiento como los resultados obtenidos por los modelos.

Métrica	Pinball	Smooth Pinball
QS	0.089	0.087
IS	0.311	0.302
Sharpness	0.204	0.197
ACE	2.022	1.549

**Tabla 5.5**

*Pinball suavizado vs Pinball*

### 5.4.2. Spread vs Non Spread

Como hemos mencionado en el apartado 4.0.2, trataremos de analizar si existe alguna mejora en los resultados del modelo si utilizamos, además del estimador de la media de los modelos, el spread, esto es, su desviación.

En principio la desviación entre modelos podría ser interesante para el algoritmo ya que permite añadir información sobre la incertidumbre de los modelos, ampliando el intervalo cuantílico cuando existe una mayor discrepancia entre las predicciones de las variables de los modelos.

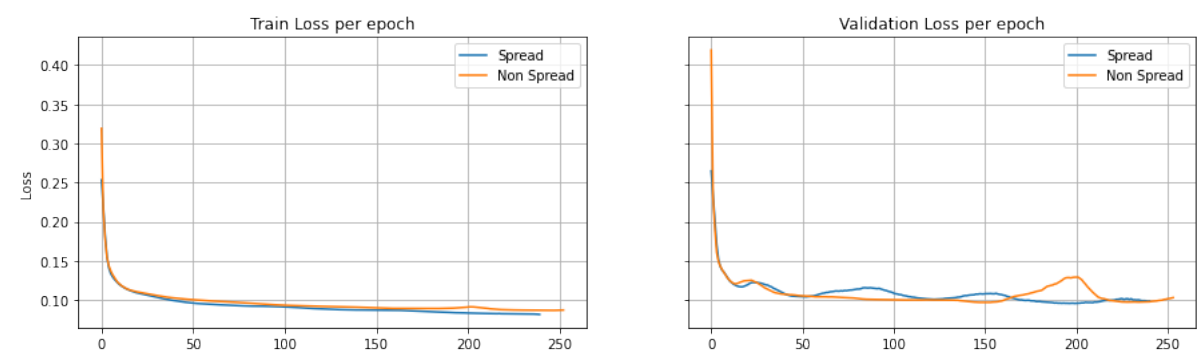
Debido al elevado número de variables (superando los 100 variables de entrada) se ha decidido realizar esta prueba con el algoritmo MLP que permite captar relaciones no lineales pero es menos costoso computacionalmente que la LSTM.

Como se muestra en la Tabla 5.6, los resultados muestran que exista cierta mejora

al usar el spread, pero no muy grande (exceptuando el caso de la métrica *ACE*). Por otro lado, el entrenamiento parece muy similar en ambos casos, siendo menos estable en algunas ocasiones en uno u otro.

**Figura 5.8**

*Evolución del coste del entrenamiento utilizando el spread vs non spread*



Métrica	Spread	Non Spread
QS	0.094	0.098
IS	0.326	0.335
Sharpness	0.203	0.220
ACE	1.789	1.164

**Tabla 5.6**

*Resultados utilizando o no utilizando el spread en los predictores*

## 6. CONCLUSIONES

En 2021 la eólica fue la principal fuente de energía eléctrica en España. Sin embargo, su integración en la red eléctrica requiere de buenas predicciones a un día vista. Típicamente, dichas predicciones son puntuales pero en este trabajo se quiere ir un paso más allá y ofrecer estimaciones de la incertidumbre de la predicción.

Por ello, en este trabajo se han realizado predicciones probabilísticas de la generación de energía eólica en la región de Cádiz durante los años 2015-2018 con el fin de tener una cantidad mayor de información a la hora de tomar decisiones respecto a la cantidad a ofertar al día siguiente en la subasta eléctrica.

Para hacer estas predicciones, se han utilizado distintos modelos (lineal, perceptrón multicapa y LSTM), comparando sus resultados con distintas métricas de evaluación. Todos los modelos utilizan como entradas predicciones de variables meteorológicas y ofrecen a la salida un intervalo de predicción de la producción eléctrica, a varios horizontes de predicción del día siguiente. El modelo que mejores resultados obtiene es la LSTM utilizando una dos capa oculta y utilizando secuencias de 4 días hacia atrás.

Además, se ha comprobado que la función de coste Smooth Pinball obtiene mejores resultados que la función de coste Pinball y que el spread de las predicciones del ensemble de modelos de ECMWF ofrece también una ligera mejora.

Futuros estudios podrían comprobar la bondad de las redes LSTM utilizadas en este trabajo, para otras regiones de la península.



## BIBLIOGRAFÍA

- Cerrillo, A. (2014). La eólica se ha convertido en la primera fuente de electricidad en España. *La Vanguardia*.
- ECMWF. (s.f.). Modelling and Prediction. *ECMWF Modelling and Prediction*.
- Fernández Munguía, S. (2016). Un poco de historia. El primer aerogenerador de España. *Diario Renovables*.
- Hatalis, K., Lamadrid, A. J., Scheinberg, K. y Kishore, S. (2017). Smooth pinball neural network for probabilistic forecasting of wind power. *arXiv preprint arXiv:1710.01720*.
- Hernandez, W., Méndez, A., Maldonado-Correa, J. L. y Balleteros, F. (2016). Modeling of a robust confidence band for the power curve of a wind turbine. *Sensors*, 16(12), 2080.
- Hochreiter, S. y Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L. y Shao, L. (2020). Normalization techniques in training dnns: Methodology, analysis and application. *arXiv preprint arXiv:2009.12836*.
- Kalnay, E. (2003). Historical overview of numerical weather prediction. *Handbook of Weather, Climate, and Water: Dynamics, Climate, Physical Meteorology, Weather Systems, and Measurements*, 95-115.
- Koenker, R. y Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.
- Norbu, S., Couraud, B., Robu, V., Andoni, M. y Flynn, D. (2021). Modelling the redistribution of benefits from joint investments in community energy projects. *Applied Energy*, 287, 116575.
- OMIE. (s.f.). Funcionamiento del mercado diario. *OMIE*.
- Prechelt, L. (1998). Early stopping-but when? *Neural Networks: Tricks of the trade* (pp. 55-69). Springer.
- REE. (2021). La eólica se convierte en la principal fuente de generación de energía eléctrica en España en 2021. *Red Eléctrica de España*. <https://www.ree.es/es/sala-de-prensa/actualidad/nota-de-prensa/2021/12/la-eolica-se-convierte-en-la-principal-fuente-de-generacion-de-energia-electrica-en-espana-en-2021>
- Richardson, L. F. (1922). *Weather Prediction by Numerical Process*.
- Thakur, A. (2022). LSTM RNN in Keras: Examples of One-to-Many, Many-to-One Many-to-Many. *wandb.ai*.
- Wiegand, B. (2015). Introduction to Numerical Weather Prediction. *Fakultät für Ingenieurwissenschaften SystemTechnikLabor (STL)*.
- Zhang, Y., Wang, J. y Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32, 255-270.