

Predicción ordinal de rampas de viento usando *Echo State Networks* de complejidad reducida

M. Dorado-Moreno

Dpto. de Informática y Análisis Numérico
Universidad de Córdoba
Córdoba, España
manuel.dorado@uco.es

P. A. Gutiérrez

Dpto. de Informática y Análisis Numérico
Universidad de Córdoba
Córdoba, España
pagutierrez@uco.es

S. Salcedo-Sanz

Dpto. de Teoría de la Señal y Comunicaciones
Universidad de Alcalá
Alcalá de Henares, España
sancho.salcedo@uah.es

L. Prieto

Dpto. de Recursos Energéticos
Iberdrola
Madrid, España

C. Hervás-Martínez

Dpto. de Informática y Análisis Numérico
Universidad de Córdoba
Córdoba, España
chervas@uco.es

Resumen—Las *Renovables* son la fuente de energía que más ha crecido en los últimos años a nivel mundial. En particular, la energía eólica en Europa es actualmente la que tiene un mayor crecimiento, estando su capacidad de producción en la segunda posición, por detrás del gas natural. Existen una serie de problemas que complican la integración del recurso eólico en la red eléctrica. Uno de ellos es conocido como rampas de viento, que son incrementos o decrementos de gran magnitud en la velocidad del viento en un tiempo reducido. Estas rampas de viento pueden dañar las turbinas en los parques eólicos, así como reducir los ingresos generados a partir de la producción del parque. Actualmente, la mejor forma de afrontar este problema es predecir estas rampas de viento, de forma que se puedan parar las turbinas con suficiente antelación, evitando así los daños que puedan producirse. Para realizar esta predicción, se suelen utilizar modelos que puedan aprovechar la información temporal. Uno de los modelos más conocidos con estas características son las redes neuronales recurrentes. En este trabajo utilizaremos las conocidas como *Echo State Networks* (ESNs), las cuales han demostrado obtener un buen rendimiento en predicción de series temporales. En concreto, proponemos utilizar ESNs de complejidad reducida para afrontar un problema de predicción de rampas de viento en tres parques eólicos en España. A nivel metodológico, se comparan tres arquitecturas diferentes de red, dependiendo de la configuración de las conexiones de la capa de entrada con el *reservoir* o directamente con la capa de salida. Los resultados muestran que, por lo general, los mejores resultados son obtenidos por la estructura *Delay Line Reservoir with Feedback* (DLRB) y que el aumento en el rendimiento obtenido por la arquitectura de Doble *reservoir* con respecto a la arquitectura de Simple *reservoir* es mínima, y teniendo en cuenta el gran aumento de complejidad computacional de la arquitectura Doble, concluimos que los mejores resultados son obtenidos por la combinación de la estructura DLRB con la arquitectura Simple.

Index Terms—Echo state networks; Energía eólica; Clasificación ordinal; Rampas de viento; Redes neuronales recurrentes.

I. INTRODUCCIÓN

La naturaleza nos ofrece múltiples formas de generar energía de forma sostenible y libres de emisiones contaminantes. Este tipo de energías explotan recursos naturales renovables y actualmente son las que más están creciendo a nivel mundial. Algunas de las más conocidas son la energía solar, la eólica y la marina (mareomotriz, undimotriz, eólica *offshore*), así como sus combinaciones, aunque existen otras alternativas tales como la biomasa o la energía hidráulica. Nuestro trabajo se centrará en este caso en la energía eólica y, dentro de esta, en la producción en parques eólicos, los cuales, generan energía mediante turbinas eólicas de grandes dimensiones. El problema de la gran mayoría de los recursos renovables es que normalmente son intrínsecamente intermitentes, lo que dificulta la completa explotación del recurso, y su incorporación al *mix* energético en igualdad de condiciones con respecto a otros tipos de recursos no renovables. En el caso de la energía eólica, además de su intermitencia intrínseca, aparecen otro tipo de problemas en producción, relacionados con características específicas del recurso. En el caso de los parques eólicos, uno de los problemas más grave son las conocidas como *rampas de viento*, definidas como incrementos o decrementos en la velocidad del viento de gran magnitud en un corto periodo de tiempo. Estas rampas de viento pueden ser positivas, es decir, que se produce un incremento de la velocidad del viento, o negativas, cuando es un decremento. El efecto de las rampas positivas en un parque eólico es, principalmente, el posible daño que pueden causar a las turbinas existentes. Esto puede derivar en un aumento de los costes de mantenimiento del parque. En cuanto a las rampas negativas, su efecto fundamental es un decremento de producción de energía súbito, que puede acarrear problemas

Este trabajo ha sido desarrollado con la financiación de los proyectos TIN2017-85887-C2-1-P, TIN2017-85887-C2-2-P y TIN2017-90567-REDT del Ministerio de Economía y Competitividad de España (MINECO) y fondos FEDER. La investigación de Manuel Dorado-Moreno ha sido financiada por el programa predoctoral FPU (Ministerio de Educación y Ciencia) con referencia FPU15/00647. Los autores agradecen a *NVIDIA Corporation* la cesión de recursos computacionales a través del *GPU Grant Program*.



de abastecimiento si este tipo de sucesos no se predice con suficiente antelación.

Muy diversos problemas relacionados con energías renovables han sido abordados mediante técnicas de aprendizaje automático, por ejemplo, en energía solar [2], undimotriz [8] o energía eólica [4]–[6]. En aprendizaje automático, uno de los modelos más conocidos a la hora de tratar con series temporales y realizar predicciones son las redes neuronales recurrentes [12]. La diferencia con las redes neuronales convencionales es la inclusión de ciclos entre sus neuronas, es decir, se permiten conexiones de una neurona consigo misma y con neuronas situadas en las capas anteriores o en su misma capa. De cualquier forma, al incrementar el número de capas (o neuronas) de las redes recurrentes para incrementar su capacidad de cómputo, las redes sufren un problema conocido como desvanecimiento del gradiente [12], debido al cual, al ir enlazando las derivadas a través de los ciclos, estas tienden a cero, por lo que no aportan información al gradiente e impiden la actualización de los pesos. Una de las propuestas más aceptadas para solventar este problema son las *echo state networks* (ESNs) [11], las cuales tienen una capa oculta conocida como *reservoir* en la que se encuentran todos los ciclos de los enlaces entre las neuronas, inicializados de forma aleatoria. Este *reservoir* está totalmente conectado con las entradas y las salidas, y estas últimas conexiones son las únicas que se entrenan. Así, se evita el problema de la tendencia del gradiente a 0, ya que no es necesario utilizar las derivadas para entrenar los pesos del *reservoir*.

Una de las dificultades asociadas a las ESNs es su naturaleza estocástica, ya que parte de su rendimiento depende del azar. Para solventar este problema, en este trabajo vamos a utilizar las ESNs de complejidad reducida propuestas en [14], las cuales establecen los enlaces entre las neuronas siguiendo un patrón razonable y además los inicializan de forma determinista, pudiendo así, justificar su correcto funcionamiento. Además proponemos tres arquitecturas distintas, siguiendo el trabajo realizado en [5], para comprobar las distintas formas en las que el *reservoir* afecta al resultado del modelo dependiendo de las entradas que se conecten a él. Es importante destacar que, debido al orden natural que muestran las distintas categorías a predecir (rampa negativa, ausencia de rampa y rampa positiva), el problema se aborda desde la perspectiva de la clasificación ordinal [10]. Por último, cabe destacar que para este trabajo se utilizarán dos fuentes de datos, a partir de las cuales extraeremos las variables de entrada utilizadas por las ESNs para la predicción de las rampas de viento. La primera incluye mediciones recogidas por los sensores de tres parques eólicos situados en España (ver Figura 1), mientras que la segunda se corresponde a datos generados a través de modelos físicos y matemáticos que son conocidos como datos de reanálisis [7]. Estos datos de reanálisis son muy fiables y se calculan a lo largo de todo el mundo cada 0,125 grados (en latitud y longitud) y cada 6 horas, alcanzando así una buena resolución tanto espacial como temporal.

En la Sección II se expondrán las características extraídas de las dos bases de datos y se explicará la generación de la

base de datos conjunta. Las distintas arquitecturas propuestas para el modelado se expondrán en la Sección III, justo antes de explicar el diseño experimental en la Sección IV. Para concluir, en la Sección V, se mostrarán los resultados obtenidos y se realizará una discusión de los mismos. La Sección VI expondrá las conclusiones obtenidas tras este trabajo.

II. BASE DE DATOS

En esta sección se explican las características de las fuentes de información utilizadas para resolver el problema de la predicción de rampas de viento, las transformaciones llevadas a cabo y la unión de ambas fuentes de información. La primera fuente de información corresponde a medidas de la velocidad de viento, obtenidas cada hora en tres parques eólicos situados en España, como se puede observar en la Figura 1. Calcularemos las rampas de viento como valores objetivo a ser predichos utilizando distintas variables predictivas. La segunda fuente de información de la que obtendremos variables predictivas es el proyecto de reanálisis ERA-Interim [7], que almacena información sobre el clima cada 6 horas. Estos datos se calculan utilizando modelos físicos, es decir, que no dependen de ningún sensor que pueda generar datos perdidos, por lo que se pueden estimar valores futuros de estos modelos para predecir las rampas de viento.

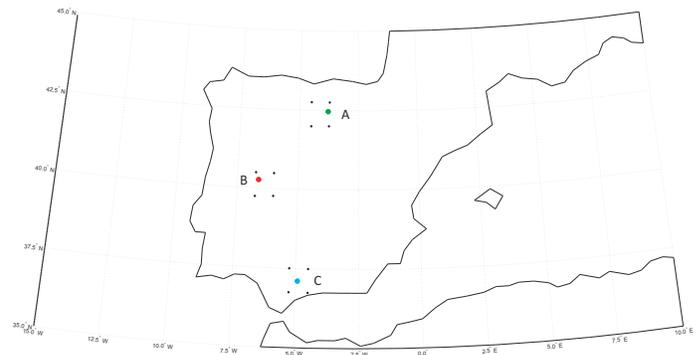


Figura 1: Localización de las tres parques eólicos (A, B y C) y de los nodos de reanálisis considerados en las cercanías de cada parque.

II-A. Definición de Rampa de Viento

La función $S_t : \mathbb{R}^k \rightarrow \mathbb{R}$ será la función evaluada para decidir si se ha producido una rampa de viento, o no, en un determinado periodo de tiempo, donde k será el número de características a considerar. Hay muchas definiciones de S_t [9] y todas de ellas incluyen la producción de energía (P_t) como un criterio del parque eólico o la turbina de viento a considerar. En este artículo utilizamos la siguiente definición:

$$S_t = P_t - P_{t-\Delta t_r},$$

donde Δt_r es el intervalo de tiempo considerado para estudiar la rampa (6 horas en nuestro caso, de acuerdo con la frecuencia de los datos de reanálisis).

Utilizando S_t , podemos definir el problema de clasificación definiendo un umbral (S_0) para discretizar la variable objetivo de la siguiente forma:

$$y_t = \begin{cases} C_{NR}, & \text{si } S_t \leq -S_0, \\ C_{NoR}, & \text{si } -S_0 < S_t < S_0, \\ C_{PR}, & \text{si } S_t \geq S_0. \end{cases}$$

donde $\{C_{NR}, C_{NoR}, C_{PR}\}$ son las diferentes categorías de eventos a detectar, es decir, rampas negativas (NRs), no rampas (NoRs) y rampas positivas (PRs).

En nuestro caso, S_0 se ha definido como un porcentaje de la capacidad de producción de energía del parque eólico (en concreto, un 50% siguiendo [9]). La predicción de rampas de viento también implica un vector de variables predictivas: utilizaremos datos de reanálisis climatológicos como datos de entrada (\mathbf{z}) (definidos en la siguiente sección), junto con el valor de la velocidad de viento medido por los sensores del parque eólico en el instante anterior al que queremos predecir.

II-B. Datos de reanálisis

Para cada una de los tres parques, tenemos 48 predictores, que corresponden a 12 variables por cada nodo de reanálisis (considerando los 4 nodos más cercanos a la parque eólico, ver Figura 1). Estos nodos están situados cada 15 kilómetros (0.125 grados) en todo el mundo y en esas localizaciones se calculan las variables de reanálisis utilizando modelos físicos que estiman medidas de variables meteorológicas a muy diferentes alturas. Entre las mismas, tenemos la velocidad de viento, presión y temperatura del aire. Para evitar trabajar con tantos datos, que en muchos casos estarán altamente correlados, por lo que introducirán ruido al modelo, realizamos una media ponderada por la distancia de cada nodo de reanálisis al centro del parque eólico. De esta forma reducimos el número de predictores de reanálisis a 12, sin perder la información relativa de cada nodo. Así, en primer lugar calculamos la distancia de cada nodo de reanálisis al parque eólico mediante la distancia de Haversine:

$$d(p_0, p_j) = \arccos(\sin(lat_0) \cdot \sin(lat_j) \cdot \cos(lon_0 - lon_j) + \cos(lat_0) \cdot \cos(lat_j)),$$

donde p_0 es la localización del parque eólico, p_j la localización de cada nodo de reanálisis, y lat y lon serán la latitud y longitud de los puntos, respectivamente. Una vez que la distancia de cada uno de los nodos de reanálisis (los cuatro puntos negros que rodean a cada parque eólico, ver de nuevo Figura 1) al parque eólico ha sido calculada, estas distancias se invierten y normalizan, considerando que cuanto más corta sea la distancia, más grande será el peso de la información de ese nodo de reanálisis en la media ponderada:

$$w_i = \frac{\sum_{j=1}^4 d(p_0, p_j)}{d(p_0, p_i)}, \quad i = 1, \dots, 4. \quad (1)$$

Después de calcular estos pesos, se utilizan para obtener una media ponderada de cada una de las 12 variables:

$$\bar{z}_i = \sum_{j=1}^4 w_j z_{i,j} \quad i \in \{0, 1, \dots, 11\}$$

siendo i el índice de cada una de las variables de reanálisis, j cada uno de los nodos de reanálisis y w_j el peso correspondiente, calculado en la Ecuación (1).

III. ARQUITECTURAS PROPUESTAS

En este artículo proponemos una modificación de los modelos utilizados en [5]: en lugar de realizar una clasificación binaria, vamos a resolver un problema de clasificación ordinal para tres clases. Por otra parte, vamos a modificar la estructura del *reservoir* basándonos en las distintas estructuras propuestas en [14], las cuales reducen la complejidad del *reservoir* además de eliminar la aleatoriedad en la inicialización de los mismos, sin reducir de forma considerable el rendimiento del modelo. Gracias a estas estructuras de *reservoir* de complejidad reducida, se construyen los ciclos entre las neuronas del *reservoir* de forma determinista, además de otorgarle a la red una capacidad de memoria óptima para cada problema. Un esquema de las tres estructuras de *reservoir* puede analizarse en la Figura 2. En la capa de salida utilizaremos un modelo de regresión logística ordinal basado en umbrales [13], el cual proyecta los patrones en una dimensión y optimiza el valor de los umbrales para separar las distintas clases.

A continuación describiremos las arquitecturas propuestas para resolver la predicción de rampas de viento, que exploran distintas formas de combinar los valores pasados de la velocidad de viento y los datos de reanálisis procedentes del proyecto ERA-Interim. Proponemos tres arquitecturas, la primera (ver Figura 3a) tiene un único *reservoir* cuya entrada es la velocidad de viento recogida en cada parque eólico en el instante anterior al que se quiere predecir, de forma que el resto de variables de reanálisis se utilizarán directamente como entradas a la capa de salida, sin ser procesadas por ningún *reservoir*. La segunda arquitectura (ver Figura 3b) dispone de dos *reservoir* independientes, uno para la velocidad del viento y otro cuyas entradas serán todas las variables de reanálisis. Por último, proponemos una tercera arquitectura (ver Figura 3c) en la que solo disponemos de un *reservoir* cuyas entradas englobarán tanto la velocidad del viento como el resto de variables de reanálisis. Con estas tres arquitecturas, estudiaremos la capacidad de cómputo del *reservoir* así como su utilidad para cada tipo de variables.

En la capa de entrada, incluimos los vectores de entrada con la velocidad de viento y las 12 variables de reanálisis en los instantes t (para la primera) y $t + 1$ (para las demás), respectivamente. El uso de \mathbf{z}_{t+1} en la capa de entrada para predecir y_{t+1} es posible, como se ha mencionado en la sección II, debido a que estos datos se calculan mediante modelos físicos que permiten estimarlos de forma fiable 6 horas después del instante actual.

La metodología propuesta para entrenar los distintos modelos propuestos es la siguiente:

1. Crear un *reservoir* de tamaño M , conectando sus neuronas según las restricciones de cada tipo de *reservoir* (DLR, DLRB y SCR) tal y como se indica en [14].
2. Recoger todos los estados del *reservoir*. Para ello, se alimenta el *reservoir* desde el instante $t = 1$ hasta el

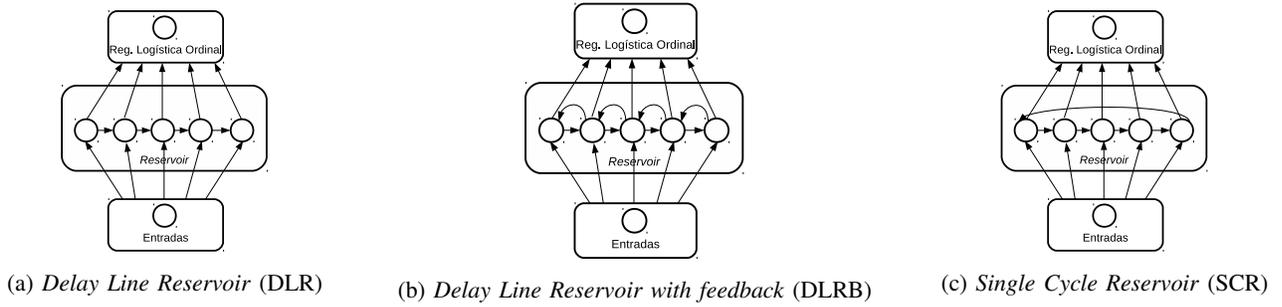
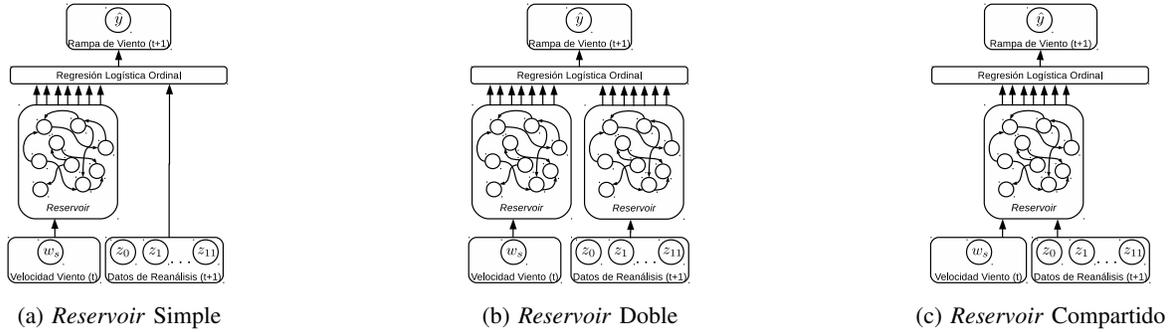

 Figura 2: Estructuras de *reservoir* consideradas


Figura 3: Distintas arquitecturas de red propuestas

instante $t = M$, de forma tal que todas las conexiones del *reservoir* hayan recibido una señal, permitiendo obtener el vector \mathbf{x}_t completo.

3. Combinar los estados del *reservoir* (\mathbf{x}_t) y/o las variables de reanálisis (\mathbf{z}_{t+1}).
4. Calcular los pesos de salida entrenando un modelo de regresión logística ordinal, proyectando los patrones sobre una recta y estableciendo los umbrales para distinguir cada una de las tres clases (C_{NR} , C_{NoR} y C_{PR}).

Una vez que la red ha sido entrenada, se puede utilizar para la predicción en tiempo real de rampas de viento, descartando el patrón correspondiente a $t = 0$ ya que no habría información de instantes anteriores de tiempo para poder predecirlo.

IV. DISEÑO EXPERIMENTAL

En esta sección se describen los experimentos que se han llevado a cabo para comparar las distintas arquitecturas propuestas en la Sección III.

IV-A. Métricas de evaluación

Hay muchas métricas que pueden ser utilizadas para evaluar clasificadores ordinales. Las más comunes en aprendizaje automático son el Error Absoluto Medio (*MAE*) y el Error Medio Cero-uno (*MZE*) [10], siendo $MZE = 1 - Acc$, donde *Acc* es la precisión o la proporción de patrones bien clasificados. De cualquier modo, estas métricas podrían no ser las mejores opciones, por ejemplo, cuando medimos el rendimiento en bases de datos desequilibradas (como es nuestro caso, ver Tabla I) [1], y/o cuando los costes de diferentes

errores varían notablemente. Es por ello que, para poder evaluar correctamente el rendimiento de los clasificadores, hemos incluido la media geométrica de las sensibilidades (*GMS*), que es una métrica más estricta a la hora de penalizar la mala clasificación de las clases minoritarias, siendo 0 cuando una de las clases está mal clasificada por completo. Así, hemos considerado 3 métricas para evaluar estos modelos:

- La proporción de patrones bien clasificados (*CCR*) que se define como:

$$CCR = \frac{1}{N} \sum_{i=1}^N (I(y_i^* = y_i)),$$

donde $I(\cdot)$ es la función de pérdida cero-uno, y_i es la salida deseada para el patrón \mathbf{x}_i , y_i^* es la predicción del modelo y N es el número de patrones del conjunto de entrenamiento o *test* en la base de datos. Los valores del *CCR* varían de 0 a 1 y representa el rendimiento global de la tarea de clasificación.

- Las sensibilidades de cada clase representan la habilidad del modelo para predecir correctamente cada tipo de evento:

$$S_{NR} = \frac{CC_{NR}}{N_{NR}}, \quad S_{NoR} = \frac{CC_{NoR}}{N_{NoR}}, \quad S_{PR} = \frac{CC_{PR}}{N_{PR}},$$

donde CC_{NR} , CC_{NoR} y CC_{PR} es el número de eventos de tipo *NR*, *NoR* y *PR* bien clasificados y N_{NR} , N_{NoR} y N_{PR} ($N_{NR} + N_{NoR} + N_{PR} = N$) es el número total de eventos de tipo *NR*, *NoR* y *PR*. La media geométrica de las sensibilidades (*GMS*) se define como:

$$GMS = \sqrt[3]{S_{NR} \cdot S_{NoR} \cdot S_{PR}}.$$

Incluimos esta métrica debido al alto desequilibrio de la base de datos, de forma que los clasificadores que no se centren en una de las clases sean fácilmente detectables, ya que su GMS será cercano a 0.

- La media del error absoluto medio ($AMAE$) [1] es la media del error de clasificación MAE en cada clase, donde MAE es la media de la desviación absoluta entre la clase predicha y la clase real. Esta métrica es capaz de mitigar el efecto de las distribuciones de clase desequilibradas y se define como:

$$AMAE = \frac{1}{J} \sum_{j=1}^J MAE_j,$$

donde

$$MAE_j = \frac{1}{N_j} \sum_{i=1}^{N_j} |\mathcal{O}(y_i) - \mathcal{O}(y_i^*)|,$$

donde $1 \leq j \leq J$, $\mathcal{O}(C_{NR}) = 1$, $\mathcal{O}(C_{NoR}) = 2$, $\mathcal{O}(C_{PR}) = 3$. Los valores de MAE van desde 0 hasta $J - 1$, al igual que los de $AMAE$. En nuestro caso tendremos $J = 3$.

IV-B. Diseño Experimental

Los tres parques eólicos de la Figura 1 se han utilizado en la comparación de resultados de las distintas arquitecturas propuestas. Para evaluar los resultados, las tres bases de datos se han dividido de la misma forma: los últimos 365 días son utilizados para el conjunto de test y el resto de la base de datos para el conjunto de entrenamiento. Todas las bases de datos comienzan el 2/3/2002 y finalizan el 29/10/2012. Con esta partición de los datos, los patrones por clase de cada una de las tres bases de datos se muestra en la Tabla I, especificando el tipo de evento recogido (rampa negativa, NR, no rampa, NoR, y rampa positiva, PR).

Tabla I: Número de patrones por clase de las distintas bases de datos consideradas en la experimentación

Parque	Conjunto	#NR	#NoR	#PR
A	Entrenamiento	753	12469	886
	Test	67	1288	105
B	Entrenamiento	1161	11804	1074
	Test	117	1220	123
C	Entrenamiento	661	12768	679
	Test	58	1340	62

Las diferentes arquitecturas presentadas en la sección III han sido comparadas entre si, comparando además las distintas estructuras internas del *reservoir* de acuerdo a [14]. Deseamos identificar la arquitectura con mejor rendimiento y comprobar si el *reservoir* de complejidad reducida es suficiente para nuestro problema.

Debido al desequilibrio del problema, realizamos un sobre-muestreo mediante la metodología SMOTE [3] a las salidas del *reservoir*, tal y como se explica y justifica en [4]. Para

cada clase minoritaria (C_{NR} y C_{PR}), generaremos tantos patrones sintéticos como se indique mediante una proporción del número de patrones de la clase mayoritaria (en nuestro caso, un 60% de los patrones de la clase mayoritaria), evitando así obtener clasificadores triviales. El sobre-muestreo se realiza únicamente sobre el conjunto de entrenamiento.

Si no se controlan los pesos de la regresión logística ordinal, estos pueden llegar a ser muy grandes, sobre-ajustando la red o impidiendo clasificar los patrones de las clase minoritarias. Para paliar este efecto, incluimos un término de regularización, que obliga a ajustar un parámetro α que controla su importancia. Este parámetro ha sido ajustado mediante una validación cruzada interna de tipo 5-fold sobre el conjunto de entrenamiento. Los valores explorados han sido $\alpha \in \{2^{-5}, 2^4, \dots, 2^{-1}\}$. La selección final del mejor valor se realiza en base a maximizar la mínima sensibilidad, es decir, $MS = \min\{S_{NR}, S_{NoR}, S_{PR}\}$.

El resto de parámetros se han configurado de la siguiente forma: el número de neuronas del *reservoir* es $M = 50$, asumiendo que es un tamaño suficiente para abordar el problema sin suponer un coste computacional demasiado alto. Los valores de los enlaces del *reservoir* se establecen según una distribución uniforme entre los valores $[-0,9, 0,9]$.

V. RESULTADOS

Esta sección expone y discute los resultados obtenidos por las distintas arquitecturas de red y las diferentes estructuras de *reservoir* propuestas. En la Tabla II, se pueden observar los resultados de las tres arquitecturas con el *reservoir* de estructura DLR, en la Tabla III los resultados con la estructura DLRB y en la Tabla IV los resultados obtenidos usando la estructura SCR (ver Figuras 2a, 2b y 2c, respectivamente).

Tabla II: Resultados obtenidos con las tres arquitecturas propuestas con estructura DLR

Parque	Modelo	GMS	AMAE	CCR	MS
A	Simple	<i>0,6607</i>	<i>0,3485</i>	0,7212	<i>0,5671</i>
	Doble	0,6951	0,3060	0,7411	0,5820
	Compartido	0,3056	0,6207	<i>0,7328</i>	0,1791
B	Simple	0,6394	0,3850	<i>0,7082</i>	0,5811
	Doble	<i>0,6311</i>	<i>0,3903</i>	0,7000	<i>0,5726</i>
	Compartido	0,3185	0,5921	0,7630	0,0813
C	Simple	0,6344	0,3768	0,7383	<i>0,5689</i>
	Doble	<i>0,6227</i>	<i>0,3931</i>	<i>0,7452</i>	0,5862
	Compartido	0,2443	0,6598	0,7636	0,0967

El mejor resultado se muestra en negrita y el segundo mejor en cursiva

Tal y como se observa en la Tabla II, la arquitectura simple gana en dos de los tres parques eólicos, la doble gana en uno y la compartida obtiene los peores resultados. Como se ha comentado anteriormente en la sección IV-A, se obtiene un alto valor de CCR pero a coste de un valor muy bajo de GMS , clasificando incorrectamente las clases minoritarias.

Por el contrario, en la Tabla III, la arquitectura con *reservoir* doble gana en dos de los tres parques eólicos, mientras que la arquitectura con un único *reservoir* para la velocidad de



Tabla III: Resultados obtenidos con las tres arquitecturas propuestas con estructura DLRB

Parque	Modelo	GMS	MAAE	CCR	MS
A	Simple	<i>0,6715</i>	0,3389	0,7294	<i>0,5970</i>
	Doble	0,6971	0,3057	<i>0,7397</i>	0,6268
	Compartido	0,3630	0,5484	0,7863	0,1343
B	Simple	0,6397	0,3847	<i>0,7089</i>	0,5726
	Doble	<i>0,6352</i>	<i>0,3912</i>	0,7006	<i>0,5470</i>
	Compartido	0,1648	0,6634	0,7821	0,0427
C	Simple	<i>0,6290</i>	<i>0,3871</i>	0,7376	<i>0,5645</i>
	Doble	0,6437	0,3733	<i>0,7486</i>	0,5862
	Compartido	0,1922	0,6454	0,8445	0,0645

El mejor resultado se muestra en negrita y el segundo mejor en cursiva

viento obtiene el segundo resultado en estos dos parques y el mejor resultado en el restante. El mal comportamiento de la arquitectura con *reservoir* compartido se repite.

Tabla IV: Resultados obtenidos con las tres arquitecturas propuestas con estructura SCR

Parque	Modelo	GMS	MAAE	CCR	MS
A	Simple	<i>0,6607</i>	0,3485	0,7212	<i>0,5671</i>
	Doble	0,6951	0,3060	0,7411	0,5970
	Compartido	0,3056	0,6207	<i>0,7328</i>	0,1492
B	Simple	0,6394	0,3850	<i>0,7082</i>	0,5726
	Doble	<i>0,6311</i>	<i>0,3903</i>	0,7000	<i>0,5641</i>
	Compartido	0,3185	0,5921	0,7630	0,1623
C	Simple	0,6344	0,3768	<i>0,7383</i>	0,5689
	Doble	<i>0,6227</i>	<i>0,3931</i>	<i>0,7383</i>	<i>0,5517</i>
	Compartido	0,2443	0,6598	0,7636	0,1290

El mejor resultado se muestra en negrita y el segundo mejor en cursiva

Por último, los resultados obtenidos con la estructura SCR, que podemos observar en la Tabla IV, siguen la misma dirección que los obtenidos con la estructura DLR, lo que indica que en dos de las tres bases de datos el modelo simple obtiene los mejores resultados.

Comparando las tres tablas, la estructura de *reservoir* que mejor rendimiento obtiene para la predicción de rampas de viento es la DLRB. A su vez, la arquitectura de *reservoir* doble solo mejora los resultados para la estructura DLRB (y no para las otras dos). Si consideramos el incremento de complejidad que se introduce en el entrenamiento de la regresión logística ordinal (62 entradas para el modelo simple frente a 100 entradas para el modelo doble), podemos afirmar que la arquitectura de *reservoir* simple es la más adecuada para este problema.

VI. CONCLUSIONES

Este artículo propone y evalúa tres arquitecturas distintas de redes recurrentes, y cada una de estas tres arquitecturas se crea con una estructura diferente de *reservoir* de complejidad reducida. Estas propuestas se utilizan para realizar una predicción ordinal en tres clases de rampas de viento, donde también se considera el alto nivel de desequilibrio de la base de datos.

Las arquitecturas propuestas cambian la forma en que se procesan los datos de entrada, que son una combinación de la velocidad del tiempo medida en el parque eólico y 12 variables de reanálisis. Por una parte tenemos solo la velocidad de viento (medida en cada parque eólico) procesada por el *reservoir* mientras que todas las variables de reanálisis se introducen directamente en la regresión logística ordinal. Otra de las arquitecturas dispone de dos estructuras de *reservoir*, una para procesar la velocidad del viento y otra para procesar las variables de reanálisis. La última dispone de un único *reservoir* para procesar todas las entradas.

Para evitar los modelos triviales (que clasifican todo en la clase mayoritaria), hemos aplicado sobre-muestreo a las activaciones del *reservoir*, mejorando así los resultados. Podemos concluir que el modelo con un único *reservoir* para la velocidad de viento obtiene el mejor rendimiento, siendo pocos los casos en los que la arquitectura de doble *reservoir* funciona mejor que la simple. Con respecto a las estructuras de *reservoir* de complejidad reducida que han sido comparadas, los mejores resultados se obtienen con la estructura *Delay Line Reservoir with feedback*.

REFERENCIAS

- [1] S. Baccianella, A. Esuli, F. Sebastiani, Evaluation measures for ordinal regression, in: Proc. of the 9th Int. Conf. on Intelligent Systems Design and Apps, pp. 283-287 (2009).
- [2] S. Basterrech and T. Buriánek, "Solar irradiance estimation using the Echo State Network and the flexible neural tree," *Intelligent data analysis and its Applications, Volume 1*, Springer, pp. 475-484 (2014).
- [3] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research* vol. 16, pp. 321-357 (2002).
- [4] M. Dorado-Moreno, A.M. Durán-Rosal, D. Guijo-Rubio, P.A. Gutiérrez, L. Prieto, S. Salcedo-Sanz and C. Hervás-Martínez, "Multiclass Prediction of Wind Power Ramp Events Combining Reservoir Computing and Support Vector Machines," *Conference of the Spanish Association for Artificial Intelligence, Lecture Notes in Computer Science*, vol. 9868, pp. 300-309 (2016).
- [5] M. Dorado-Moreno, L. Cornejo-Bueno, P.A. Gutiérrez, L. Prieto, S. Salcedo-Sanz and C. Hervás-Martínez, "Combining *reservoir* computing and over-sampling for ordinal wind power ramp prediction," *International Work-conference on Artificial Neural Networks, Lecture Notes in Computer Science*, vol. 10305, pp. 708-719 (2017).
- [6] M. Dorado-Moreno, L. Cornejo-Bueno, P.A. Gutiérrez, L. Prieto, C. Hervás-Martínez and S. Salcedo-Sanz, "Robust estimation of wind power ramp events with *reservoir* computing," *Renewable Energy*, vol. 111, pp. 428-437 (2017).
- [7] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli et al., "The ERA-Interim reanalysis: configuration and performance of the data assimilation system," *Quart. J. of the Royal Met. Society*, vol. 137, pp. 553-597 (2011).
- [8] J. C. Fernandez, S. Salcedo-Sanz, P. A. Gutiérrez, E. Alexandre y C. Hervás-Martínez. "Significant wave height and energy flux range forecast with machine learning classifiers," *Engineering Applications of Artificial Intelligence*, vol. 43, pp. 44-53 (2015).
- [9] C. Gallego-Castillo, A. Cuerva-Tejero and O. López-García, "A review on the recent history of wind power ramp forecasting," *Renewable and Sustainable Energy Rev.*, vol. 52, pp. 1148-1157, (2015).
- [10] Pedro Antonio Gutiérrez, María Pérez-Ortiz, Javier Sánchez-Monedero, Francisco Fernández-Navarro, and César Hervás-Martínez, "Ordinal regression methods: survey and experimental study," *IEEE Trans. on Knowledge and Data Engineering*, vol.28, pp. 127-146 (2016)
- [11] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks," *GMD Report 148*, German National Research Center for Information Technology, pp. 1-43, (2001).



- [12] M. Lukosevicius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127-149 (2009).
- [13] P. McCullagh, "Regression Modelos for Ordinal data," *J. of the Royal Statistic Society*, vol. 42, no. 2, pp. 109-142 (1980).
- [14] A. Rodan and P. Tiño, "Minimum Complexity Echo State Network," *IEEE Trans. Neural Networks*, vol. 22, no. 1, pp. 131-144 (2011).