

# CNN's en predicción de renovables

P. Acuaviva, C. González-Guillén, M. Jaenada, R. Ruperto, I.  
Villanueva

30 de marzo de 2023

# El problema

- Uno de los objetivos principales de REE es casar producción y demanda instantáneamente en todo momento.
- Demanda, producción eólica y producción fotovoltaica son aleatorias. El resto de las producciones (nuclear, gas, hidroeléctrica...) son ajustables.
- Demanda se predice sobre todo con técnicas de series temporales (con influencia de la temperatura). No la trabajamos.
- Eólica y fotovoltaica dependen de la meteorología. Las previsiones meteorológicas tienen limitaciones.
- Hay dos incertidumbres encadenadas: la meteorología (no la trabajamos) y la traducción de la meteorología en producción eléctrica.

# El problema

- Contexto: España es, de entre los países grandes, el que más proporción de renovables tiene en la mezcla eléctrica. Por ello, REE es uno de los líderes mundiales en gestión de renovables.
- Para facilitar el casamiento constante entre producción y demanda, REE publica horariamente predicciones de demanda, producción eólica y producción solar para las  $n$  horas futuras ( $> 72$ ).
- REE elabora estas predicciones a partir de modelos propios, junto con predicciones que compra a empresas proveedoras.
- REE ha utilizado hasta ahora modelos ARIMA, Redes Neuronales (shallow), SVM's...

# Los datos

- Las previsiones meteorológicas vienen en forma de NWP, Numerical Weather Predictions.
- Cada 6 horas, el ECMWF u otras agencias elaboran un conjunto de predicciones para múltiples variables meteorológicas a distintas alturas en un mallado cuadrado de  $0,1^\circ$  (aproximadamente 10km).
- Las predicciones asignan un valor a cada una de esas variables para cada una de las 240 horas siguientes al momento de la predicción.

# Los datos

- Para las producciones **eólicas** se usan como variables las componentes  $u$ ,  $v$  del viento a 100m (la altura media del eje de los molinos es 90m, y creciendo). Datos horarios.
- Tenemos además las producciones horarias de la fuente a predecir (parque, región peninsular, península, isla canaria).
- 8760 horas por año. Típicamente tenemos alrededor de 5 años, no siempre homogéneos.
- Observacionalmente: cuantos más años, mejor. A partir de 4 años el error disminuye poco.

# Los datos

- Para **fotovoltaica** se usan: radiación solar directa, difusa y total, radiación clear sky (determinista), temperatura en superficie, viento a 10 m, zenith y azimuth (deterministas).
- En fotovoltaica, tenemos muchos menos datos.
  - ▶ El boom fotovoltaico está siendo ahora
  - ▶ De noche no hay luz!
  - ▶ Las redes se fijan sobre todo en el mediodía. El amanecer y el crepúsculo “informan” menos.
  - ▶ A cambio, parece que los escenarios solares son mucho menos variados que los eólicos.

# Nuestro enfoque

- Las NWP de un mapa alrededor del parque eólico, o alrededor de la península, son formalmente idénticas a una foto: un rectángulo de píxeles con dos canales por pixel.
- Así pues, usamos CNN's. Mantienen la estructura local de la información.

# Nuestras arquitecturas

- Arquitecturas de 3 capas de convolución, seguidas de pooling, seguidas de una capa fully connected.
- Para las primeras horas (hasta la hora 7 o así) incluimos las últimas producciones disponibles en la capa fully connected, quizás vía LSTM.
- Arquitecturas mucho más profundas (ResNet34) tardan más en entrenar pero no mejoran resultados.
- Main problem: error en el input.
- Entrenamos para minimizar MAE. Estándar en energía.
- Importante regularizar. Cientos de miles de parámetros.

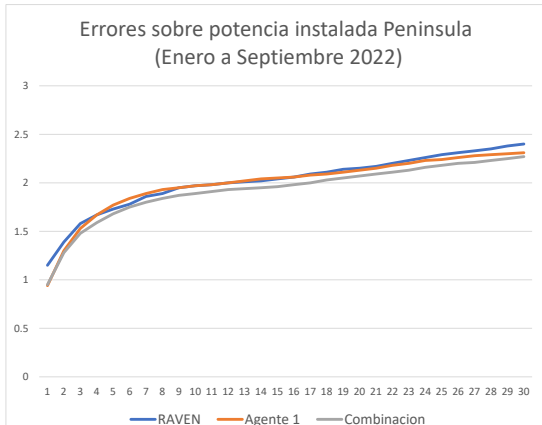


# Problemas específicos

- El “eje de datos” es el tiempo. Lleva información. Tenemos una serie temporal.
  - ▶ Tratamiento de vídeo (sin persistencia)
  - ▶ LSTM,... (persistencia)
  - ▶ Atención a la partición train/val/test.
- Falta de datos. Regression  $\implies$  no es fácil data augmentation
- Incertidumbre en los inputs. No tenemos acceso a los vientos reales. Ensembles, poor-man's ensembles.
- Saturación de la RAM. TFRecords + batches.

# Comparativa de errores

- Errores sobre potencia instalada. La producción media es el 22 % de la instalada.
- Error típico de un parque a 24 h: 9 %.



# Industrialización

- REE ha tenido nuestro modelo en pruebas más de un año y lo están poniendo en producción.
- En 2022 hemos elaborado un predictor para cada uno de los 1200 parques eólicos de Península y Canarias. Aprox 6000 horas de GPU.
- Nos estamos convirtiendo en empresa intentando vender el servicio.

# Transfer learning en eólica

- Se entrenan las redes en una región.
- Preentrenadas, se sobreentrenan en los parques
- Reduce a la mitad el tiempo de entrenamiento, no reduce el error.

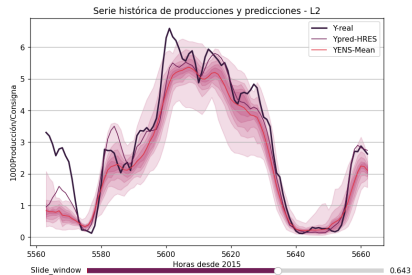
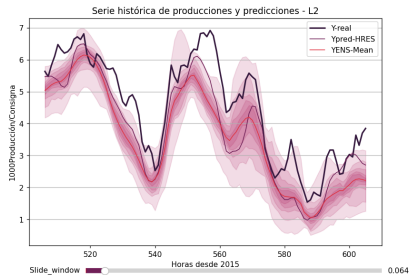
# Ensembles

- Tiempos de descarga infernales.
- Tenemos 5 ensembles desde 2015-2022.
- Tenemos los 50 ensembles en 2018

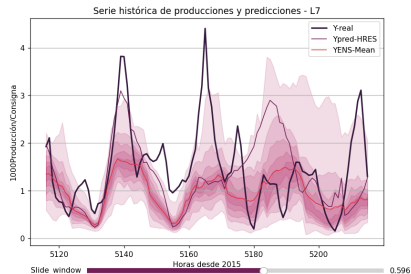
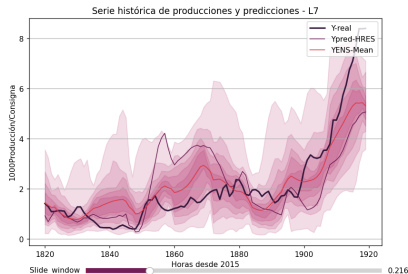
	Entrenamiento HRES	Entrenamiento HRES + 5 ENS
16 HORAS	4.59 %	4.48 %
24 HORAS	4.67 %	4.60 %
40 HORAS	4.94 %	4.86 %

**Cuadro:** Error de predicción eólica en Mancha Oriental con distintos datos para entrenamiento.

# Ensembles



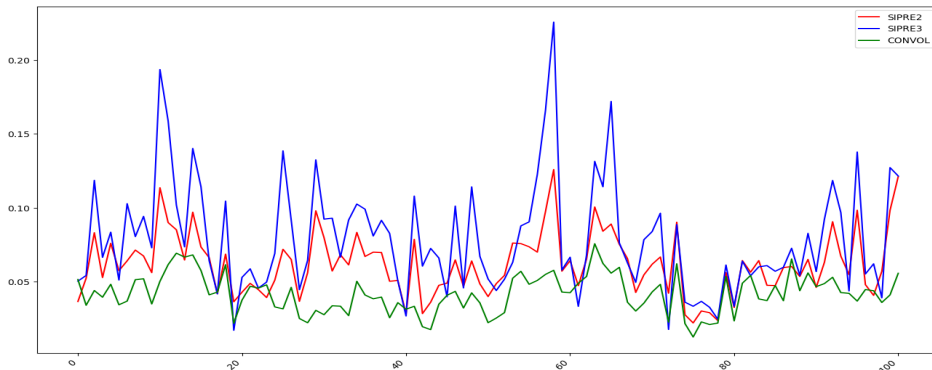
# Ensembles



- Menos trabajo hasta ahora.
- Muchos menos datos.
- Arquitecturas.
  - ▶ Redes neuronales totalmente conectadas
  - ▶ Redes neuronales de convolución: Estas redes pretenden captar características locales de las variables meteorológicas.
  - ▶ Redes neuronales recurrentes (LSTM): Estas redes pretenden incluir características de serie temporal de la propia predicción.



# Solar



**Figura:** Comparativa de MAE semanal entre los modelos de SIPRESOLAR y CONVOL

Errores en la central solar fotovoltaica La Mula (Murcia).

- Error “naive”: 7.1 %
- Error regresor lineal: 6.9 %
- Error Redes Neuronales: 4.4 %