

Aplicación de la predicción probabilística en la respuesta de la demanda (SEIO 2023)

M. C. Ruiz-Abellón, L. A. Fernández-Jiménez, A. Gabaldón Marín, A. Guillamón Frutos

Proyectos TED2021-129722B-C32 y TED2021-129722B-C33 financiados por:



- 1 Introducción.
- 2 Evaluación de la calidad de las predicciones probabilísticas.
- 3 Método QRF (Quantile Regression Forest).
 - 3.1. Descripción del método QRF.
 - 3.2. Resultados del QRF para nuestros datos.
- 4 Aplicaciones a la DR (Demand Response).
 - 4.1. La DR y los modelos PBLM.
 - 4.2. Aplicación 1: determinar el potencial DR diario.
 - 4.3. Aplicación 2: determinar el número de usuarios necesario en acciones DR.

1. Introducción

- Predicción de la demanda: **predicción puntual** del consumo eléctrico en diferentes niveles de agregación. Aunque la literatura en este campo es muy amplia, es bastante reducida si nos referimos a la predicción probabilística.
- Uno de los motivos es la **dificultad para medir la calidad de las predicciones probabilísticas**. Las predicciones puntuales son sencillas de evaluar (desviaciones entre el valor predicho y el valor real).
- Las predicciones probabilísticas pueden proporcionarse en forma de **intervalos de predicción** para un nivel de cobertura prefijado (nivel de confianza), siendo preferibles los intervalos centrados. Fijar un nivel de cobertura adecuado es difícil, el rango 75%-85% es una buena solución de compromiso.
- Una predicción no paramétrica de la distribución se puede obtener proporcionando la **predicción de un conjunto de cuantiles**, a partir de los cuales se pueden construir intervalos de predicción a distintos niveles de confianza.

2. Evaluación de las predicciones probabilísticas

Wilks(2011) identifica las principales propiedades de las predicciones probabilísticas : *accuracy*, *reliability* y *sharpness*.

- **Accuracy (exactitud)**: se usan las mismas medidas que en predicción puntual,

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=h_1}^{t=h_2} (y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{N} \sum_{t=h_1}^{t=h_2} |y_t - \hat{y}_t|$$

donde y_t es el valor real, \hat{y}_t la predicción puntual, h_1 y h_2 las horas de inicio y fin, N el n° total de horas en el periodo de testeo.

2. Evaluación de las predicciones probabilísticas

- **Reliability (calibración):** las predicciones probabilísticas están calibradas si parecen provenir de la misma distribución que las observaciones. Herramientas:

- $PICP_\alpha$ (Prediction Interval Coverage Probability):

$$PI_{t,\alpha} = \left(\widehat{Q}_{t, \frac{1-\alpha}{2}}, \widehat{Q}_{t, \frac{1+\alpha}{2}} \right), \text{ intervalo de predicción}$$

$$PICP_\alpha = \frac{1}{N} \sum_{t=h_1}^{t=h_2} \mathbf{1}_{(y_t \in PI_{t,\alpha})}, \text{ donde } \mathbf{1}_{(\cdot)} \text{ es función indicador}$$

Para cada probabilidad α , debe ser $PICP_\alpha \simeq \alpha$.

- *Verification Rank Histogram:* Para cada observación y_t , se calcula la posición que ocupa relativa a los cuantiles predichos en dicho instante. Si el modelo está calibrado, todas las posiciones deben aparecer aproximadamente en la misma proporción, de manera que **el histograma debe ser plano.**

2. Evaluación de las predicciones probabilísticas

- **Sharpness (nitidez)**: es la habilidad de concentrar la información probabilística, es decir, está relacionada con la amplitud que tienen los intervalos de predicción calculados.

- $PINW_{t,\alpha}$ (Prediction Interval Normalized Width):

$$PINW_{t,\alpha} = \frac{\widehat{Q}_{t, \frac{1+\alpha}{2}} - \widehat{Q}_{t, \frac{1-\alpha}{2}}}{\frac{1}{N} \sum_{t=h_1}^{t=h_2} y_t}$$

- $PINAW_{\alpha}$ (Prediction Interval Normalized Average Width):

$$PINAW_{\alpha} = \frac{1}{N} \sum_{t=h_1}^{t=h_2} PINW_{t,\alpha}$$

Interesa minimizar $PINAW_{\alpha}$ (maximizar sharpness) manteniendo $PICP_{\alpha} \simeq \alpha$ (calibración).

2. Evaluación de las predicciones probabilísticas

- **CRPS (Continuos Ranked Probability Scores)**: es una medida que contiene información simultánea del “accuracy”, “reliability” y “sharpness”. Se reduce al MAE en el caso de predicciones puntuales.

$$CRPS_t = \int_{-\infty}^{+\infty} \left[\hat{F}(z) - \mathbf{1}_{(z \geq y_t)} \right]^2 dz$$

$$CRPS = \frac{1}{N} \sum_{t=h_1}^{t=h_2} CRPS_t$$

Interesan valores bajos del CRPS.

3.1. Descripción del método QRF

- **Quantile Regression Forest (QRF)** es una adaptación del algoritmo de Random Forest (RF), éste último reconocido como un método eficiente para estimar la media de la variable respuesta a partir de valores dados de los predictores o regresores.
- Propuesto por Meinshausen (2006), muestra que **RF también proporciona información relevante acerca de la distribución condicionada** de la variable respuesta, no solo sobre la media.
- El análisis de regresión estándar tiene como objetivo estimar la media condicionada de la variable respuesta para valores dados de los predictores, minimizando para ello una función de coste que se corresponde con **el error cuadrático medio**:

$$E(Y | X = x) = \arg \min_z E \left\{ (Y - z)^2 | X = x \right\}$$

3.1. Descripción del método QRF

- Más informativa que la media condicionada, es la **distribución condicionada de la variable respuesta**:

$$F(y | X = x) = \Pr(Y \leq y | X = x)$$

- El cuantil α de la respuesta Y condicionado a $(X = x)$ viene dado por:

$$Q_\alpha(x) = \inf \{y : F(y | X = x) \geq \alpha\}$$

- El cálculo de los cuantiles permite obtener **intervalos de predicción**:

$$PI(x) = (Q_{0.05}(x), Q_{0.95}(x))$$

3.1. Descripción del método QRF

- La regresión estándar utiliza como función de coste a minimizar el error cuadrático medio, mientras que la **regresión cuantil** utiliza la media de la **quantile loss function**:

$$L_{\alpha}(y, q) = \begin{cases} \alpha |y - q| & y > q \\ (1 - \alpha) |y - q| & y \leq q \end{cases}$$

$$Q_{\alpha}(x) = \arg \min_q E \{L_{\alpha}(y, q) | X = x\}$$

3.1. Descripción del método QRF

- **QRF no minimiza la función de coste anterior**, sino que utiliza el algoritmo RF para estimar la distribución condicionada de la variable respuesta, y a partir de ésta, se estiman los cuantiles.
- **¿Cómo funciona Random Forest?** Para cada input ($X = x$), se obtiene un conjunto de pesos $\{w_i(x), i : 1, 2, \dots, n\}$ de manera que la **estimación de la media de** ($Y | X = x$) viene dada por la media ponderada (con dichos pesos) de las observaciones muestrales $(y_i)_{i=1,2,\dots,n}$ usadas en el conjunto de entrenamiento:

$$RF: \hat{\mu}(x) = \sum_{i=1}^n w_i(x) \cdot y_i$$

3.1. Descripción del método QRF

- **¿Cómo funciona QRF?** Para cada input ($X = x$), **se estima la distribución de ($Y | X = x$)** como la distribución ponderada (**usando los pesos de RF**) de las observaciones muestrales.

$$QRF: \hat{F}(y | X = x) = \sum_{i=1}^n w_i(x) \cdot \mathbf{1}_{(y_i \leq y)}$$

- Los *hiperparámetros* a seleccionar para QRF son los de RF:
 - *n*tree: número de árboles.
 - *m*try: número de predictores que intervienen en la construcción de cada árbol.
 - *min.node.size*: número mínimo de observaciones en cada nodo.

3.2. Resultados del QRF para nuestros datos

● Descripción de nuestros datos:

- Consumo eléctrico horario para el conjunto de usuarios de una localidad riojana.
- Datos de entrenamiento: desde 01/10/2008 hasta 30/09/2010.
- Datos zona test: desde 01/10/2010 hasta 31/03/2011.

● Predictores (total 55) y horizonte de predicción:

- Horizonte: 48 horas debido al funcionamiento del Mercado Eléctrico.
- Calculamos las predicciones de los cuantiles 0.01, 0.02, ..., 0.99.
- *Consumos con retardos* de 48, 72, ..., 168 horas.
- *VARIABLES CLIMÁTICAS*: temperatura, radiación, velocidad viento, humedad, fracción_total, precipitación y nieve. Todas ellas se predicen también con horizonte de 48h usando el modelo de predicción numérica meteorológica WRF.
- *VARIABLES CALENDARIO*: 42 variables dummy indicando si es festivo, horario de verano, hora del día, día de la semana y mes del año.

3.2. Resultados del QRF para nuestros datos

- **Selección hiperparámetros:**

- Hacemos 10-folds Cross Validation usando el CRPS como medida de ajuste a minimizar.
- Creamos una rejilla con diferentes valores de "mtry" y "min.node.size", fijando "ntree=1000" suficientemente grande.
- Hiperparámetros óptimos: *ntree=1000*, *min.node.size=5* y *mtry=27*.

Calidad de las predicciones probabilísticas en zona test:

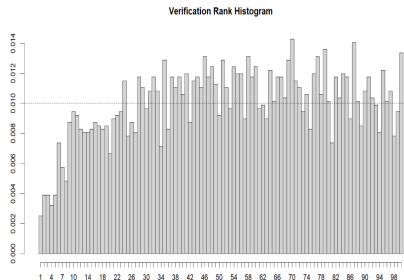
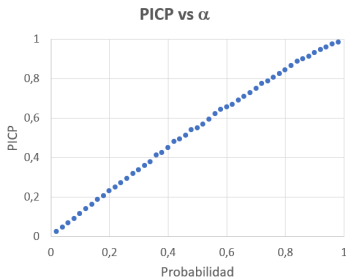
- **Accuracy:** consideramos como predicción puntual el cuantil 50%.

Consumo medio zona test = 3633.2 kW

RMSE = 213.3 kW, *MAE* = 154.8 kW

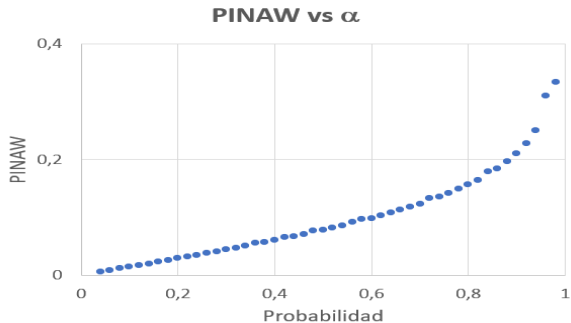
3.2. Resultados del QRF para nuestros datos

- **Reliability** (calibration):



3.2. Resultados del QRF para nuestros datos

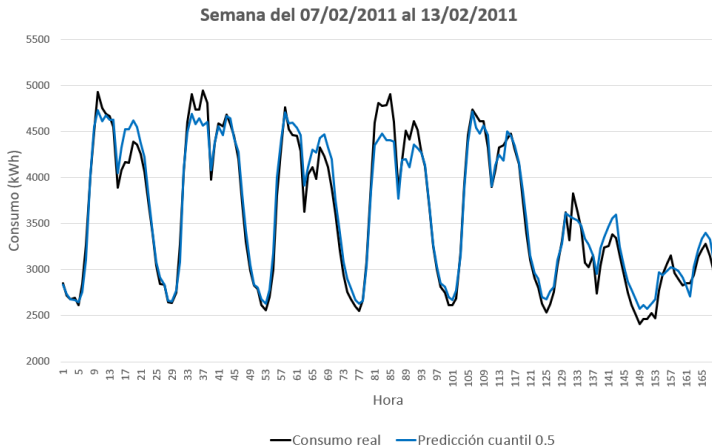
- **Sharpness:**



- **CRPS:** 112.3 kW.

3.2. Resultados del QRF para nuestros datos

- Ejemplo de consumo real y predicción puntual:



4.1. La DR y los modelos PBLM

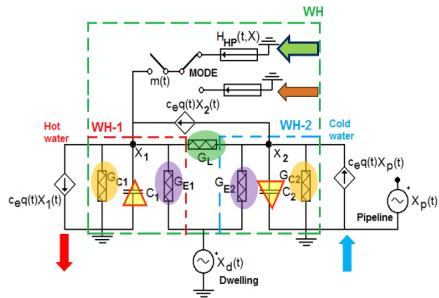
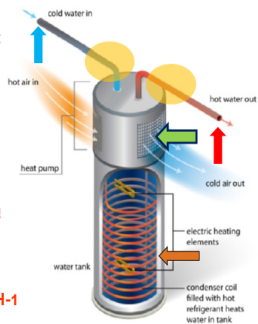
- **Demand Response (DR)** : se refiere al cambio en la demanda que realizan los clientes (usuarios) de las redes eléctricas con el fin de contribuir al necesario equilibrio entre oferta y demanda del sistema.
- Los participantes en DR reciben incentivos económicos.
- Las redes inteligentes y los sistemas de almacenamiento de energía hacen de la DR una fuente importante de flexibilidad, especialmente necesaria con el aumento de las energías renovables (mayor incertidumbre).

4.1. La DR y los modelos PBLM

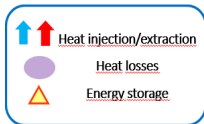
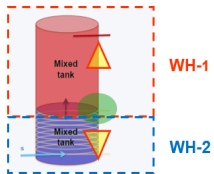
- **PBML (Physically-Based Load Model):** modelos obtenidos a partir del comportamiento físico de típicas cargas eléctricas residenciales (aire acondicionado, calefacción, termo de agua, etc).
- Estos modelos se basan en equilibrios energéticos entre el aire interno, los elementos constructivos de la vivienda, los aparatos acondicionadores y el entorno externo a través de un sistema de ecuaciones de espacio-estado.
- La utilidad principal de estos modelos dinámicos es simular el comportamiento de la carga bajo acciones de control o programas de gestión de dicha carga.

4.1. La DR y los modelos PBLM

Figure Source:
NREL, USA



Tank detail: stratification



PBLM termo agua caliente

4.1. La DR y los modelos PBLM

● **Nuestro escenario:**

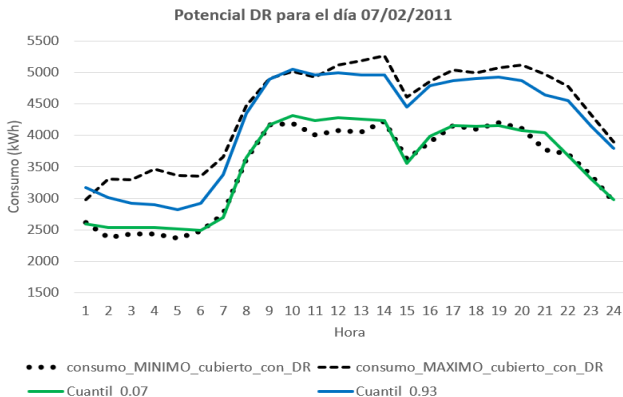
- Datos de consumo eléctrico horario de una localidad riojana de unos 4500 habitantes.
- Supondremos controlable por un agregador el 50% de la carga correspondiente a **calefacción eléctrica** y el 100% de la carga correspondiente a **termo de agua caliente**, así como que las características de las viviendas son similares. En nuestro caso: 1430 termos y 750 aparatos de calefacción controlables (unas 1500 viviendas).
- La calefacción representa el 42.9% del consumo total y el termo el 17.9% (datos de IDAE).
- Condiciones de confort a cumplir: (17°C , 22°C) para temperatura de estancias y al menos 35°C para salida de agua.

4.2. Aplicación 1: determinar el potencial DR diario

- **Objetivo:** determinar la banda de incertidumbre (a través de los cuantiles) que podemos cubrir con DR para un día concreto y de manera que se cumplan unos estándares de confort.
- **Procedimiento:**
 - Seleccionamos un día concreto "D" objeto de estudio.
 - Obtenemos las predicciones probabilísticas para las 24 horas del día "D" usando QRF y horizonte 48h (condiciones Mercado Eléctrico).
 - Usaremos los valores del cuantil 0.5 cada hora como curva base del día "D" (energía a comprar).
 - Con los modelos PBLM, simulamos las subidas y bajadas máximas de energía soportados por la agragación para el día "D".
 - Seleccionamos los cuantiles de la predicción probabilística más próximos a la flexibilidad máxima calculada en la etapa anterior.

4.2. Aplicación 1: determinar el potencial DR diario

- **Día 07/02/2011: el potencial DR permitirá cubrir al menos un 86% de incertidumbre**, intervalo $(\hat{Q}_{0.07}, \hat{Q}_{0.93})$, pudiendo aumentar en algunos tramos. Hay una alta probabilidad mantener el consumo final próximo a la curva base de energía solicitada gracias a DR.

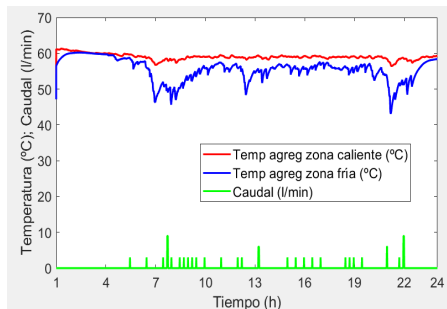
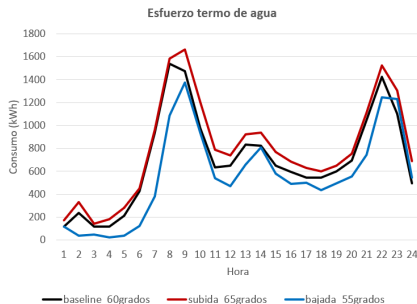


4.3. Aplicación 2: determinar el número de usuarios necesario en DR

- **Objetivo:** para un porcentaje de incertidumbre prefijado (p.e. 70%), determinar el número de usuarios de la agregación que deben participar en acciones DR con el fin de garantizar que podemos cubrir dicho nivel de incertidumbre.
- **Procedimiento:**
 - Seleccionamos un día concreto "D" objeto de estudio.
 - Obtenemos las predicciones probabilísticas para las 24 horas del día "D" usando QRF y horizonte 48h. Usaremos los valores del cuantil 0.5 cada hora como curva base del día "D" (energía a comprar).
 - Consideramos los cuantiles predichos correspondientes a la incertidumbre fijada: $\hat{Q}_{h,0.15}$ y $\hat{Q}_{h,0.85}$, $h = 1, 2, \dots, 24$.
 - Calculamos la energía a bajar con $\hat{Q}_{h,0.85} - \hat{Q}_{h,0.5}$ y energía a subir con $\hat{Q}_{h,0.5} - \hat{Q}_{h,0.15}$.
 - Con los modelos PBLM, calculamos el número de usuarios necesario para subir o bajar las cantidades anteriores.

4.3. Aplicación 2: determinar el número de usuarios necesario en DR

- **Día 07/02/2011, fijamos incertidumbre del 70%:**
 - ¿Cuántos usuarios necesitaremos para acciones DR?
730 termos de agua y 383 aparatos de calefacción.
 - ¿Esfuerzo del termo? ¿Se cumplen condiciones de confort?



- Chu, Y., Coimbra, C., 2017. Short-term probabilistic forecasts for Direct Normal Irradiance. *Renewable Energy* 101, 526-536.
- Meinshausen, N., 2006. Quantile Regression Forests. *Journal of Machine Learning Research* 7, 983-999.
- Verbois, H., Rusydi, A., Thiery, A., 2018. Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting. *Solar Energy* 173, 313–327.
- Wilks, D.S., 2011. *Statistical Methods in the Atmospheric Sciences*. Academic Press.

**Proyectos TED2021-129722B-C32 y TED2021-129722B-C33
financiados por MCIN/AEI /10.13039/501100011033 y por
la Unión Europea NextGenerationEU/ PRTR**

