applied sciences

MDPI

Article

Applications of Probabilistic Forecasting in Demand Response

María Carmen Ruiz-Abellón [1], Luis Alfredo Fernández-Jiménez [2], Antonio Guillamón [1] and Antonio Gabaldón [3,*]

[1] Department of Applied Mathematics and Statistics, Technical University of Cartagena, 30202 Cartagena, Spain; maricarmen.ruiz@upct.es (M.C.R.-A.); antonio.guillamon@upct.es (A.G.)
[2] Department of Electrical Engineering, University of La Rioja, 26004 Logroño, Spain; luisalfredo.fernandez@unirioja.es
[3] Power Systems Group, Technical University of Cartagena, 30202 Cartagena, Spain
* Correspondence: antonio.gabaldon@upct.es; Tel.: +34-968-338944

Abstract: Studies on probabilistic demand forecasting remain relatively limited compared with point forecasting, despite it being especially valuable for operational and planning purposes. This paper demonstrates different applications of probabilistic forecasting in demand response, together with Physical-Based Load Models. The first application shows how to determine the percentage of uncertainty potentially covered by demand response throughout a day, whereas the second application deals with obtaining the number of consumers that would be needed in demand response actions for a desired percentage of coverage. Finally, in the third application, a specific hourly strategy is proposed for demand response policies. These applications facilitate the aggregator's energy purchasing in the market and the planning of subsequent consumption adjustments through demand response to minimize deviations from the purchased energy. The approach is illustrated using hourly demand data from a small Spanish city, and two machine learning methods were used to produce a set of probabilistic forecasts and compare results: Linear Quantile Regression and a Quantile Regression Forest. Main results show a significant reduction in deviations from the purchased energy after applying the proposed strategy, especially in the case of the forecasting method being less accurate (40.95% and 33.62% of reduction, respectively).

Keywords: short-term load forecasting; probabilistic forecasting; linear quantile regression; quantile regression forest; demand response

check for updates

Citation: Ruiz-Abellón, M.C.; Fernández-Jiménez, L.A.; Guillamón, A.; Gabaldón, A. Applications of Probabilistic Forecasting in Demand Response. *Appl. Sci.* **2024**, *14*, 9716. https://doi.org/10.3390/app14219716

Academic Editor: Manuela Sechilariu

Received: 27 September 2024
Revised: 21 October 2024
Accepted: 22 October 2024
Published: 24 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The societal commitment to energy transition and the increasing integration of renewable energy sources into electrical grids have rendered the availability of accurate and reliable short-term electric power consumption forecasting models critical. These models enable stakeholders in the electrical system to make appropriate decisions to anticipate load fluctuations, optimize energy generation and distribution, and minimize operational and environmental costs.

The methods described in the international literature for short-term load forecasting can be categorized into two groups: deterministic or point forecast methods, and probabilistic methods. The distinction between these two categories is based on how they deal with uncertainty in predictions. In deterministic methods, the focus is on providing a single expected load value as the forecast. These methods do not explicitly account for uncertainty, and they offer the single point estimate of load value without information about the potential range of outcomes or the level of confidence in the prediction. On the other hand, probabilistic methods go further by not only providing a point estimate but also offering additional information that allows you to assess the uncertainty associated with that estimation. These methods provide a probability distribution or a set of probabilities

that describe the possible range of load values and the likelihood of each value occurring. This added information is valuable for making more informed decisions.

One of the decisions that these stakeholders can make focuses on the field of electric power distribution, particularly in actions known as demand response. Demand response (DR) is an integral component of modern electrical system management [1,2] contributing to the reduction in capacity requirements and costs, facilitating greater integration of renewable energy, and lowering overall system emissions [3]. As electrical systems become more complex and power generation more variable due to the inclusion of renewable sources, the ability to adjust electricity demand in real time has become a vital tool for maintaining the stability and efficiency of the grid. DR materializes in the adaptation of electricity consumption by end-users in response to external signals. These signals can be changes in electricity prices during periods of high demand or incentives offered by grid operators to reduce load during critical periods.

DR programs can be roughly classified into three groups based on who initiates the demand reduction action: (a) rate-based or price programs, where utilities vary electricity prices to motivate customers to adjust their consumption patterns; (b) incentive or event-based programs, where utilities or aggregators request or control customer load reductions through signals or commands; and (c) demand reduction bids, where customers initiate and offer load reductions to utilities at specified prices [4]. This paper focuses on programs classified in the second group, where utilities are expected to implement an efficient incentive system to adequately compensate customers who adjust their consumption in response to signals indicating the execution of DR actions. The initial step in compensating customers who have engaged in these actions involves assessing their actual level of commitment. This requires evaluating the degree to which they have adjusted their consumption patterns in response to the utility's signals. In this task, forecasting customers' load is essential, particularly in the short term, which involves predictions up to seven days in advance.

The body of scientific literature on STLF models is extensive and diverse, although it is primarily focused on deterministic models—that is, models that do not provide information about uncertainty in prediction. Researchers have investigated a broad range of methodologies, including time series forecasting techniques such as linear regression, exponential smoothing or ARIMA [5], as well as more advanced approaches like artificial neural networks [6], support vector regression [7], random forests [8], and fuzzy logic [9]. The advancement of deep learning techniques has enabled the development of STLF models that significantly outperform traditional methods. These models have demonstrated their efficiency in identifying and modeling nonlinear patterns within historical load data, leading to enhanced predictive accuracy [10]. For a comprehensive description of the various techniques used in STL, we refer the reader to the exhaustive reviews published in the international literature [11,12].

In recent years, the development of probabilistic models for STLF has gained considerable attention from the scientific community, though the field is still in its early stages of development [13]. Probabilistic forecasting models provide predictive information that quantifies uncertainty by offering probability distributions over future quantities [14]. A widely adopted approach for obtaining probabilistic forecasts involves generating prediction intervals. A prediction interval defines the range within which the actual value of the dependent variable is expected to lie, based on a specified probability level [15]. This type of prediction is particularly valuable in decision-making processes, as it facilitates the minimization of risks.

On the one hand, researchers are trying to develop increasingly accurate STLF models that offer greater insight into the uncertainty of predictions. Among the factors influencing this accuracy, DR programs and actions have emerged as a significant source of uncertainty. On the other, DR programs fundamentally alter the STLF process [16], necessitating a combination of back-casting and two types of forecasting beyond the traditional single forecast based on historical load and weather data. The back-cast, using data from regular hours unaffected by these programs, helps utilities assess the impact of demand response

initiatives on load reduction. The first forecast follows traditional methodologies, predicting the load without DR activation. The second forecast, more intricate, utilizes back-cast insights to estimate load curtailment under DR conditions, informing utility decisions on program activation. The irregular activation of DR programs makes accurate estimation challenging, thus highlighting the importance of probabilistic forecasting in this domain as an emerging research focus.

The number of published works on probabilistic STLF models constitute only a small fraction compared to those focused on point (deterministic) forecasting models. An example of the application of a short-term probabilistic load forecasting model is provided in [17]. The authors propose a model based on a Quantile Regression Neural Network with a Triangle Kernel Function (QRNNT), enabling the non-parametric estimation of hourly probability density functions and the generation of prediction intervals. The proposed model is applied to three major urban areas in Canada, Singapore, and China, demonstrating a good probabilistic forecasting performance and surpassing traditional models in point prediction accuracy. Salinas et al. [18] propose a model based on deep learning techniques to generate prediction intervals. The authors apply the model to five datasets to evaluate its predictive performance, including a dataset of electricity consumption from 370 customers at 15 min intervals. The results demonstrate the superiority of these techniques over other approaches in the international literature, both for point predictions and prediction intervals. Xu et al. [13] propose an STLF model based on deep learning techniques, incorporating the novel inclusion of DR factors. The addition of these explanatory variables, particularly those related to electricity prices, significantly improves the model's predictive accuracy compared to other state-of-the-art models in a case study of regional load prediction in China. The authors acknowledge the limitations of their proposed model, noting that the variables related to DR are restricted to price factors and do not dynamically interact with the forecasted load.

Research published in the international literature focuses on developing optimal STLF models, both deterministic and probabilistic, and has recognized the need to integrate DR programs and actions into these models. However, how DR could be leveraged to improve the performance of STLF models has not yet been thoroughly analyzed. It is important to note that certain electricity market participants, such as retailers and aggregators, rely on the forecasts generated by STLF models to trade energy in the day-ahead market, aligning their bids with the expected demand of their customers for each of the 24 h of the following day. In this context, this article presents an example of how DR actions can first reduce the forecasting errors of STLF models, which provide the expected demand for the following day, thereby influencing the amount of energy purchased in the day-ahead market. This reduction in forecasting errors subsequently leads to a decrease in potential penalties for imbalances between the electricity purchased and consumed.

In recent years, several studies have addressed the integration of machine learning (ML) models with the potential activation of DR actions. In most cases, these models are designed for short-term load forecasting, while in other instances, they aim to directly adjust the actual load curve to match the initially predicted one. Table 1 provides a summary of relevant studies, detailing the ML techniques employed, the objectives pursued, and the advantages or limitations associated with each approach. The table also provides a brief summary of the characteristics of the method proposed in this article, which will be detailed in the following sections.

Table 1. Studies exploring the integration of ML techniques with the activation of DR strategies.

Reference	ML Technique	Specific Objective	Pros/Cons
[19]	Deep neural network and reinforcement learning	Improved network stability and reliability	Calculation of optimal incentive rates/ Possible customer mismatch
[20]	Whale optimization algorithm	Fitting the actual load curve to the forecasted load curve	Cost optimization for electricity retailers/ No predefined load forecasting model. Only day-ahead optimization
[21]	Stacked autoencoder neural network	Support for decision-making processes for day-ahead dynamic pricing	Decision support tool for retail electric providers. It considers off-customers with different behaviors/ It is essential to incorporate prosumer customers
[22]	K-Nearest Neighbor (kNN)	Shift demand to periods of low demand	Calculation of optimal incentive rates/ Challenges arising from user non-participation
[23]	Heuristic and genetic algorithm	Development of an Intelligent Electric Power Management System	It aims to reach the lowest daily cost of electricity/ It does not account for energy storage
[24]	Temporal Convolution Neural Network	PV power forecast for battery energy storage management	Payback period reduction for users/ Potential users are entire buildings
[25]	Multi-Agent Proximal Policy Optimization	Frequency regulation	Control of residential air conditioners/ Real-time control without prior planning
[26]	Ensemble with neural network, random forest, support vector machine, and extreme gradient boosting	Assessing the energy flexibility of residential buildings	Useful for energy aggregators Updated forecasts one hour in advance/ No evaluation of uncertainty associated with the prediction
This study	Quantile regression models	Fitting the actual load curve to the forecasted load curve	Probabilistic approach with uncertainty evaluation Day-ahead and one-hour-ahead forecasts One-hour-ahead fitting

As shown in Table 1, ML models have a broad range of applications in the activation of DR strategies, each with specific objectives. Nevertheless, they all share the general goals of reducing peak load, ensuring grid stability, lowering costs, and facilitating the integration of renewable energy into electricity grids. However, few studies focus on developing models for load aggregators to plan DR actions [20,26]. These studies aim to maximize the benefits of aggregators while reducing electricity purchasing costs for their customers, but do not consider the assessment of uncertainty in load forecasts. The uncertainty assessment is of particular interest in the case of DR action-planning decisions. This article aims to fill this gap with a probabilistic approach.

The main contributions of this article are linked to the use of probabilistic rather than point prediction and its applications in DR. The contributions can be listed as follows:

1. Given a set of consumers and various load types, such as Electric Heating and Water Heaters, we show how to determine the maximum percentage of predictive uncertainty that can be mitigated by DR actions. This approach enables the aggregator to estimate, at least one day in advance, the likelihood of successfully adjusting the final consumption through DR to meet the purchased energy, thereby establishing the desired balance between supply and demand.
2. In this context, we also demonstrate how to determine the number of consumers required to cover a predetermined level of predictive uncertainty. This will enable the aggregator to plan, at least one day in advance, which loads and how many should be available to participate in DR actions to effectively address the specified uncertainty.
3. Finally, we propose a specific DR strategy based on the calculation of conditional load quantiles in the very short term which enables aggregators to more accurately align the actual consumption with the purchased energy.

The paper is organized as follows. Section 2 describes the two probabilistic forecasting methods selected in the research and the dataset used to illustrate the proposed approach, as well as how to evaluate the accuracy of point and probabilistic predictions. Additionally, it develops the methodology used to simulate the effect of demand response throughout the behavior of elemental and aggregated loads. Section 3 deals with the introduction of the three main applications proposed in this paper and the analysis of the experimental results. The main findings and recommendations of the study are presented in Section 4.

2. Materials and Methods

This section consists of 5 subsections. In Section 2.1, we describe in detail the real dataset used to apply the proposed methodology. In Section 2.2, we highlight the different features that must be considered in the context of a probabilistic forecast, and we introduce indicators to evaluate the quality of the probabilistic predictions. Sections 2.3 and 2.4 deal with the two forecasting methods used in this paper: Linear Quantile Regression and the Quantile Regression Forest, respectively. Finally, in Section 2.5 we introduce the physical models used to simulate load behavior and their utility in demand response (elemental and aggregate levels).

2.1. Dataset

The proposed method is illustrated using real data from a small city in Spain, with a population of around 4500 inhabitants. For the last decade, Spanish national databases (2011) account for around 1700 main and 650 secondary residences in this city and its surroundings. This allows us to estimate 2–3 inhabitants per household, which enables us to assume, to some extent, customer requirements (for instance, daily hot water demand or tank size according to inhabitants). Power ratings of commonly used household appliances such as electric heaters (EHs) and Water Heaters (WHs) are considered around 1.5–2.0 kW and around 1.0–2.0 kW respectively. This work has considered also that more than 50% of customers could engage a DR pilot with EH and WH appliances in this geographical area.

The raw dataset for this residential area contains the hourly electricity demand of the whole city, as well as some climate variables we describe below. Data range from 1 October 2008 to 31 March 2011; the first two years were used as the training sample and the last six months as the test dataset. The climate variables considered in the dataset include the hourly forecasts of the variables shown in Table 2.

Table 2. Forecasted weather variables.

Variable	Meaning
temp	Temperature at 2 m (K)
swflux	Surface downwelling shortwave flux (W/m^2)
mod	Wind module at 10 m (m/s)
rh	Relative humidity at 2 m (0 to 1)
cft	Total cloud cover (0 to 1)
prec	Accumulated rainfall (kg/m^2)
snow	Accumulated large scale snowfall (kg/m^2)

These forecasts were obtained from Meteogalicia, the regional weather service for the region of Galicia, Spain [27]. Meteogalicia uses the numerical prediction model Weather Research and Forecast (WRF) [28], enabling it to provide hourly forecasts of a set of meteorological variables for an analysis grid covering the entire Iberian Peninsula. The points in the analysis grid are spaced approximately 12 km apart. For each day in the training and testing dataset, the forecasts of the variables listed in Table 2 correspond to the four nearest points of the analysis grid to the city's location, with forecasting horizons from 1 to 48 h downloaded. The final hourly forecasts of the weather variables for the city were obtained through a weighted average of the forecasts for the four nearest points, using the inverse of the distance from the city center to each point as the weighting factor.

According to [29], around 42.9% of the residential consumption corresponds to space heating and 17.9% to water heating. In this work, it is supposed that all the users are of the residential type with homogeneous characteristics in their controlled loads (the total of Electric Heating and Water Heaters). In addition, the comfort conditions established are as follows: from 17 °C to 22 °C for the internal room temperature and at least 35 °C for the hot water supply.

2.2. Evaluation of the Probabilistic Forecast

Point predictions (deterministic approach) are easy to evaluate by measuring the deviations between the real demand and the forecast. However, assessing probabilistic predictions is a difficult task. For example, [30] identifies some relevant properties of a probabilistic forecast: accuracy, calibration (reliability), and sharpness.

Accuracy refers to the same concept as in point prediction, thus the same measurements such as *MAE* or *RMSE* can be used, defined as follows:

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

$$MAE = \frac{1}{N}\sum_{t=h_1}^{t=h_2}|y_t - \hat{y}_t| \tag{2}$$

where y_t is the real value, $\hat{y}_t$ is the point forecast, h_1 and h_2 the boundaries hours of the test period, and N the overall number of hours in that period.

It is said that a probabilistic prediction is calibrated if it seems to come from the same distribution as the observations. Two useful tools to check this property are the Prediction Interval Coverage Probability (*PICP*) and the Verification Rank Histogram (VRH).

Recall that probabilistic predictions can be provided by means of prediction intervals. For a given coverage probability, α in (0, 1), we look for an interval that will contain the future prediction with probability α. In this context, the predicted quantiles play an important role in obtaining those prediction intervals. For example, the prediction interval at time "t" with probability α in (0, 1) is given by the following:

$$PI_{t,\,\alpha} = \left(\hat{Q}_{t,\frac{1-\alpha}{2}}, \hat{Q}_{t,\frac{1+\alpha}{2}}\right) \tag{3}$$

where $\hat{Q}_{t,\beta}$ denotes the predicted quantile a, time "t", with lower-tail probability β.

For the test dataset and given a fixed α in (0, 1), we can count the number of observations (real demand of the test dataset) that lay inside the prediction intervals at level α, and this quantity is used to calculate the *PICP*, which is given by the following:

$$PICP_\alpha = \frac{1}{N}\sum_{t=h_1}^{t=h_2}\mathbf{1}_{(y_t \in PI_{t,\,\alpha})} \tag{4}$$

where $\mathbf{1}_{(.)}$ denotes the indicator function.

Note that $PICP_\alpha \approx \alpha$ for calibrated models, so a linear equation of the form $y = x$ must follow when plotting *PICP*s against α.

An alternative way to contrast whether probabilistic forecasts are calibrated is through the *VRH*. For each observation y_t in the test dataset, we compute its rank related to the predicted quantiles at that time "t". Under calibration, all the positions must appear with a similar frequency. Therefore, the corresponding histogram (VRH) should be approximately flat. It is important to highlight that a flat VRH is a necessary but not sufficient condition for calibration (see [31]).

The property of sharpness is related to the width of the prediction intervals. The widths can be analyzed using boxplots; however, it is preferable to use objective indices.

Koshravi et al. [32] propose computing the Prediction Interval Normalized Width (*PINW*) and the Prediction Interval Normalized Average Width (*PINAW*), given by the following:

$$PINW_{t,\,\alpha} = \frac{\hat{Q}_{t,\frac{1+\alpha}{2}} - \hat{Q}_{t,\frac{1-\alpha}{2}}}{\frac{1}{N}\sum_{t=h_1}^{t=h_2} y_t} \tag{5}$$

$$PINAW_{\alpha} = \frac{1}{N}\sum_{t=h_1}^{t=h_2} PINW_{t,\,\alpha} \tag{6}$$

The lower the *PINAW*, the higher the sharpness. However, sharpness can be increased at the cost of losing calibration. Thus, the aim is to minimize the *PINAW* while keeping calibration.

The Continuous Rank Probability Score (*CRPS*) is a measure that can be used to evaluate jointly the three properties (accuracy, calibration, and sharpness) [33]. It is widely used in weather forecasting, and it reduces to *MAE* for deterministic forecasts. The lower the *CRPS*, the better the probabilistic forecast.

$$CRPS_t = \int_{-\infty}^{+\infty} \left[\hat{F}(z) - 1_{(z \geq y_t)}\right]^2 dz \tag{7}$$

$$CRPS = \frac{1}{N}\sum_{t=h_1}^{t=h_2} CRPS_t \tag{8}$$

where $\hat{F}(z)$ denotes the estimated distribution function by means of the probabilistic forecasting.

2.3. Lineal Quantile Regression

Quantile regression involves the estimation of parameters of functions that establish the relationship between the quantiles of the dependent variable and the corresponding explanatory variables. These functions can take various forms, and if they are linear, the model is referred to as a Linear Quantile Regression (LQR) model. In [34], the authors present a method for fitting the LQR model. This method allows the model to estimate the θth quantile of the dependent variable, with $0 < \theta < 1$, based on a linear relationship. Such a relationship is expressed in (9), where $L_\theta(t)$ is the θth conditional quantile of the load variable for the hour t; β_θ is the vector of parameters (coefficients); $\boldsymbol{x}(t)$ is the vector of explanatory variables for such a moment; and ε is the error term.

$$L_\theta(t) = \beta_\theta \boldsymbol{x}(t) + \varepsilon \tag{9}$$

The parameter vector β_θ can be estimated by minimizing the loss function $\rho_\theta(z)$, known as the pinball loss function, which is defined by Equation (10). This estimation process yields the expected values of the parameter vector, $\hat{\beta}_\theta$, for the θth quantile, as shown in Equation (11), where i represents any hour in the fitting period (training dataset) and N_f is the number of data used to fit the model. The expected value of the θt conditional quantile of the hourly load at hour t, $\hat{L}_\theta(t)$, can be obtained using Equation (12).

$$\rho_\theta(z) = \begin{cases} \theta z & if\ z \geq 0 \\ (1-\theta)z & if\ z < 0 \end{cases} \tag{10}$$

$$\hat{\beta}_\theta = \underset{\beta}{\operatorname{argmin}}\left[\sum_{i=1}^{N_f} \rho_\theta(L(i) - \beta \boldsymbol{x}(i))\right] \tag{11}$$

$$\hat{L}_\theta(t) = \hat{\beta}_\theta \boldsymbol{x}(t) \tag{12}$$

It is important to note that each conditional quantile has its own parameter vector, denoted as β_θ. When the vectors for different quantiles are computed separately, the

possibility of encountering crossing quantiles arises. This occurs when the estimated quantiles at different levels do not maintain the expected ordering, i.e., when a lower quantile exhibits higher values than a higher quantile at specific points, thus violating the expected relationship between quantiles. This can lead to inconsistencies in the quantile regression model. The issue of quantile crossing and its solution have been addressed in the literature [35], involving the rearrangement of quantiles for each hour to prevent crossing.

2.4. Quantile Regression Forest

The Quantile Regression Forest (*QRF*) is an adaptation of the *RF* algorithm, considered an effective method to estimate the mean of the response variable given the values of the predictors. It was firstly introduced in [36], which shows that the *RF* not only provides relevant information about the conditional mean of the response variable, but also about the conditional distribution.

It is well known that standard regression aims to estimate the conditional mean of the response variable. For that, the cost function corresponding to the expectation of the squared error is minimized:

$$E(Y|X = x) = \text{arg}min_z E\left\{(Y - z)^2 \middle| X = x\right\} \tag{13}$$

However, estimating the conditional distribution of the response variable is more informative than just focusing on the mean, even in the form of the conditional quantiles, given respectively by the following:

$$F(y|X = x) = P(Y \leq y|X = x) \tag{14}$$

$$Q_\alpha(x) = \inf\{y \ : F(y|X = x) \geq \alpha\} \tag{15}$$

Note that we can obtain prediction intervals for a given coverage level $\alpha \in (0,\ 1)$ from the estimation of the conditional quantiles, so this will be our goal.

It is important to highlight that the standard regression minimizes the mean squared error whereas the quantile regression minimizes the mean of the quantile loss function. Therefore, the function to minimize is as follows:

$$Q_\alpha(x) = \text{arg}min_q E\{L_\alpha(y, q)|X = x\} \tag{16}$$

where the quantile loss function is:

$$L_\alpha(y, q)| = \begin{cases} \alpha|y - q| & y > q \\ (1 - \alpha)|y - q| & y \leq q \end{cases} \tag{17}$$

Nevertheless, the *QRF* does not minimize the quantile loss function. Instead, the *RF* is employed to estimate the conditional distribution of the response. It results in a non-parametric method to estimate the quantiles of the response for given values of the predictors, which is efficient in the case of high-dimension datasets.

Recall that in RFs, for each input ($X = x$), the estimation of the response conditional mean is given by a weighted average of the sample observations, as follows:

$$RF: \ \hat{E}(Y|X = x) = \hat{\mu}(x) = \sum_{i=1}^{n} w_i(x) y_i \tag{18}$$

In the case of QRFs, for each input ($X = x$), the estimation of the response conditional distribution is given by the weighted distribution of the sample observations, using exactly the same weights provided by the *RF* algorithm:

$$QRF: \ \hat{F}(y|X=x) = \sum_{i=1}^{n} w_i(x)\mathbf{1}_{(y_i \leq y)} \tag{19}$$

Therefore, the hyperparameters to be tunned are the same as in RFs—that is, the number of trees (ntree), the number of predictors to be considered in each split (mtry), and the minimum number of elements in a node (min.node.size).

2.5. Demand Response and Physically Based Load Models

The methodology proposed for modeling responsive loads is Physical-Based Load Models (PBLMs). These models have been used to evaluate demand, especially demand response, storage and generation for different kind of devices, from electrical transportation (electromotive units [37] and coach internal loads [38]) to storage (batteries and double-layer/super capacitors [39,40]) and main loads from residential [41,42] and commercial load segments [43]. They can be considered as white models because they use a set of equations based on physical laws and phenomena. Nevertheless, some authors classify PBLM approaches as grey models [44], because they are simpler than the detailed models used in thermal engineering to define energy requirements for buildings or specific dwellings. The reason is that, for many applications, PBLMs still need further aggregation, and it is necessary to balance accuracy and computing and engineering complexities (or "costs") due to aggregation duties. Moreover, this kind of load model is compatible with other electrical equivalent models for distributed energy resources, which are also physically based [45]. In the DR area, and for small customer segments, PBLMs usually use from one to four RC networks to build an electrical–thermal equivalent that simulates the physical processes (energy conversion, heat conduction, convection, and radiation) between the load and its environment. In our case, the equations usually are presented in a space-system format [46], and are solved using scripts from software tools such as Matlab Control (version R2024b) or Simulink Toolboxes (version R2024b) [47]. In next subsections, the characteristics of elemental and aggregated models will be revisited. A more detailed description of those models and further references can be found in [46].

2.5.1. Elemental Models

To manage the computational complexity of the aggregation processes through the consideration of several hundreds of elemental/individual models, physical processes (in our case heat conversion and transport) must be simplified to some extent to only account for main-state variables which can be of interest for evaluating and performing load management and demand response. Some examples of these basic models are depicted in Figure 1. Firstly, Figure 1a illustrates a basic third-order model for Electric Heating (EH) and the connection with each electrical component of the model (notice that it is a thermal-electrical equivalent), which is the physical appliance and its "environment". The picture shows various heat sources, some indicated with yellow arrows (like solar radiation) and others in magenta, representing internal heat generation loads (such as other appliances, inhabitants, or ventilation leaks in the dwelling). Heat sources are modeled through current sources. In the equivalent, voltage nodes (in green) represent temperatures (state variables in PBLMs) inside the dwelling, walls, ceilings, or ground (according to the relevance of heat exchanges through other dwellings or flats downstairs and upstairs (Figure 1 on the left)). Capacitances (red/yellow triangle) represent the specific heat capacities of envelope building materials, and air and furniture to a minor extent. Resistors (arrow in brown) represent the heat conduction/convection coefficients that model heat fluxes (losses or gains according to seasons) through walls and windows. Heat exchange through solar radiation (current sources H_{SW} and H_W on the top) can be considered or simplified. It depends on the relevance of these inputs for each specific dwelling (for instance, north or south orientation) and season (winter or summer).

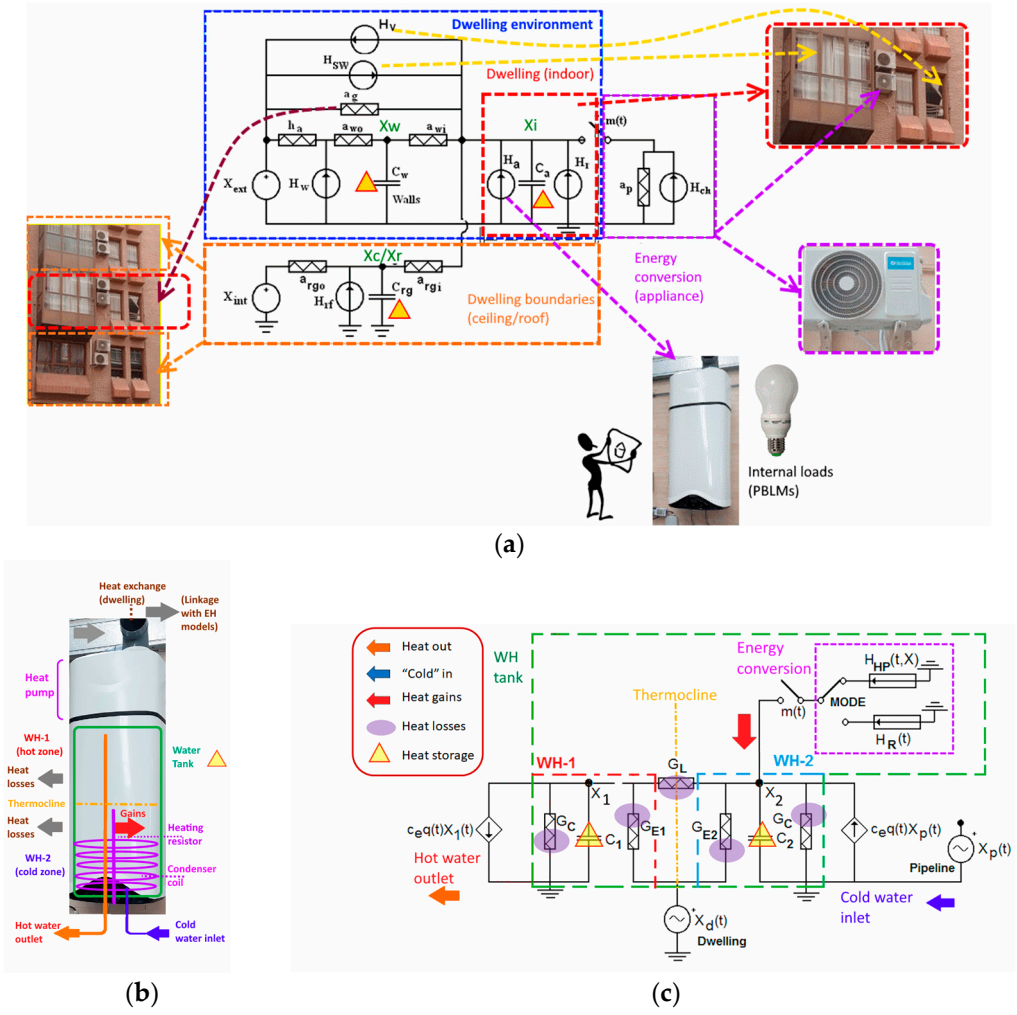


Figure 1. Elemental PBLMs and their environment: (**a**) Electric Heating (EH) in flats; (**b**) water heating (WH) appliance; (**c**) WH Electrical–Thermal equivalent.

Figure 1b,c presents the physical device and the elemental equivalent RC network for Water Heaters (WHs). Heat Pump (HPWHs) and Electrical Resistor Water Heaters (ERWHs) can be described by the same PBLM network. In the elemental HPWH model, heating is driven by two independent heat sources: a main source (the compressor of the Heat Pump) and an ancillary source, the heating resistor (Figure 1b). Operation settings between the resistor and Heat Pump can be considered through the "MODE" switch (Figure 1c). Obviously, the model can be adapted for ERWH loads by setting $H_{HP} = 0$. Heat exchanges in the water tank are modeled in two ways: conventional current sources for energy conversion (in red in Figure 1c), and two dependent current sources representing the water inlet ("cold" input to tank) and outlet ("heat" output, the load service). In the equivalent, node voltages are again temperatures (state variables) in the different zones of the water tank (WH1-2, Figure 1c), justified by water stratification inside the tank (hot water at the top and cold water at the bottom of reservoir). Finally, electrical capacitances C_1 and C_2 represent the heat storage capacity (red/yellow triangles) of each tank zone (i.e., WH1 and WH-2), while resistors model heat losses through conduction/convection between the boundaries of the tank zones (thermocline) and between the tank and its surroundings (pipelines, dwelling).

It is important to underline that, in some cases, there is potential feedback between the operation of a class of load (end-use) and others in the same dwelling, and it can be considered with this PBLM approach. For instance, HPWHs cool the surrounding of the load (Figure 1b, heat exchange) and this can affect EH performance because the internal

load changes (Figure 1a, at the bottom): it can support cooling in summer, or it can increase heat "losses" in winter. This is not considered in our approach because WHs are outside the dwelling considered for EH, as is usual in flats. More details on PBLMs can be found in [42], and validation databases are available on the authors' website [46].

2.5.2. Aggregated Models

For a practical participation of the load flexibility potential in electricity markets or services, flexibility must reach some thresholds (usually, some hundreds of kW), and this involves load aggregation. Literature usually uses two alternatives: Monte-Carlo and Stochastic Equations. Monte-Carlo is the older and most common option [48], and it has been selected for this work. Another possibility was first discussed by Malhamé et al. in [49]. Authors use probability density functions obtained through the resolution of a Stochastic Partial Differential Equation but for a one-zone water tank (this elemental model is only accurate if the appliance performs near its steady state). If the elemental model of the WH tank requires several zones, this method becomes very complex because it involves higher order equations and becomes computationally more expensive.

Two aggregated load groups are simulated for testing: one for EH loads and another for WH loads. To consider the actual diverse characteristics of different end-uses within a large population of loads (perhaps thousands of appliances and customers) and simplify the aggregation methodology, a quasi-homogeneous control group (QHCG) is established for each end-use. The QHCG involves the consideration of different load characteristics (e.g., elemental model parameters, performance, or the size of each appliance) and customer uses (e.g., small variations in thermostat settings, water and heating usage patterns). As an example, the tank capacity of WH loads in the QHCG has been considered a uniform distribution between 80 L and 110 L (a standard size for 2–3 inhabitants, see Section 2.1). The electrical parameters can be found in manufacturers' catalogs and by considering the size of loads in a specific country (e.g., in Spain, the average size of a WH is around 90 L).

For those interested in PBLM aggregation methodologies, additional models, validation databases, individual end-use profiles for residential customers, and some Matlab scripts with elemental models for solving PBLMs can be found in [46].

3. Results

This section gathers the results of the proposed methodology. The first subsection summarizes the comparative results obtained with the LQR and *QRF* methods and for horizons of 48 h and 1 h, including the evaluation of the probabilistic predictions. The rest of the subsections are dealing with each one of the applications developed in the paper: how to quantify the uncertainty covered by daily DR actions, determining the number of loads needed, and proposing an hourly strategy for those DR actions.

3.1. Probabilistic Predictions

This section depicts the probabilistic predictions obtained with the two methods described above (LQR and the *QRF*), with the double objective of comparing accuracy between the methods and evaluating the effect on the applications for demand response.

Two different prediction models were developed for each of the two methods. The first model of each method was designed to predict the demand value for each hour of the following day (forecasting horizon of 48 h); they are denoted by LQR48 and QRF48, respectively. The second model of each method was designed to provide the demand value with a prediction horizon of 1 h; they are denoted by LQR1 and QRF1, respectively. The first model compromises a total of 55 predictors, whereas the second model involves a total of 79 predictors, both detailed in Tables 3 and 4, respectively. In all cases, the 99 quantiles corresponding to probabilities 0.01, 0.02, …, 0.99 were estimated. The coefficients for each model were adjusted using the training dataset, including all available explanatory variables. Once the models were developed, they were applied for the forecast of the hourly load using the data from the testing dataset.

Table 3. List of predictors used in the LRQ and *QRF* methods with a prediction horizon of 48 h.

Weather variables (7 predictors)	temp, swflux, mod, rh, cft, prec and snow
Historical loads (6 predictors)	Actual loads with lags 48, 72, 96, 120, 144, 168.
Calendar variables (42 predictors)	Eleven dummies for the month, six dummies for the weekday, twenty-three dummies for the day hour, one dummy for holidays, and one dummy for daylight saving time.

Table 4. List of predictors used in the LRQ and *QRF* methods with a prediction horizon of 1 h.

Weather variables (7 predictors)	temp, swflux, mod, rh, cft, prec and snow.
Historical loads (30 predictors)	Actual loads with lags 1, 2, . . ., 24, 48, 72, 96, 120, 144, 168.
Calendar variables (42 predictors)	Eleven dummies for the month, six dummies for the weekday, twenty-three dummies for the day hour, one dummy for holidays, and one dummy for daylight saving time.

In the case of LQR, to avoid quantile crossing, the 99 values obtained for each hour were reordered accordingly. Table 5 shows the deterministic (forecast for quantile 0.5) and probabilistic forecasting results obtained for the entire testing dataset.

Table 5. Forecasting results of the LQR models for the testing dataset.

	LQR48 (Horizon 48 h)	**LQR1 (Horizon 1 h)**
MAE (kW)	231.46	133.16
RMSE (kW)	313.59	185.34
CRPS (kW)	164.44	95.11

It is worth highlighting the improvement achieved, as expected when reducing the prediction horizon, in both the deterministic and probabilistic forecasts of the LQR1 compared to the LQR48 model. However, the LQR1 model exhibits a poorer calibration compared to the LQR48 model, which can be observed in Figure 2, where *PICP* and VRH are plotted for the testing period, applying both models. For the VRH, classes of length 10 were considered—that is, the first bar of the histogram shows the relative counts of percentiles from 1 to 10, the second bar shows the relative counts of percentiles from 11 to 20, and so on. Therefore, a flat VRH is expected for a calibrated model, with 10% of relative frequency for each class. While the LQR48 model exhibits an almost flat VRH, the LQR1 model clearly shows under-dispersion. Sharpness is depicted in Figure 3.

In the case of the *QRF*, the parameter tunning was performed by 10-fold cross validation, with ntree = 1000 (large enough), and searching for the minimum *CRPS* in a grid varying mtry and min.node.size. Note that *CRPS* was selected as the goodness-of-fit measure to evaluate the three properties of probabilistic forecasts jointly (accuracy, calibration, and sharpness).

For the prediction horizon of 48 h and 1 h, the optimal hyperparameters after the cross-validation process as well as some goodness-of-fit measures for the test dataset can be seen in Table 6.

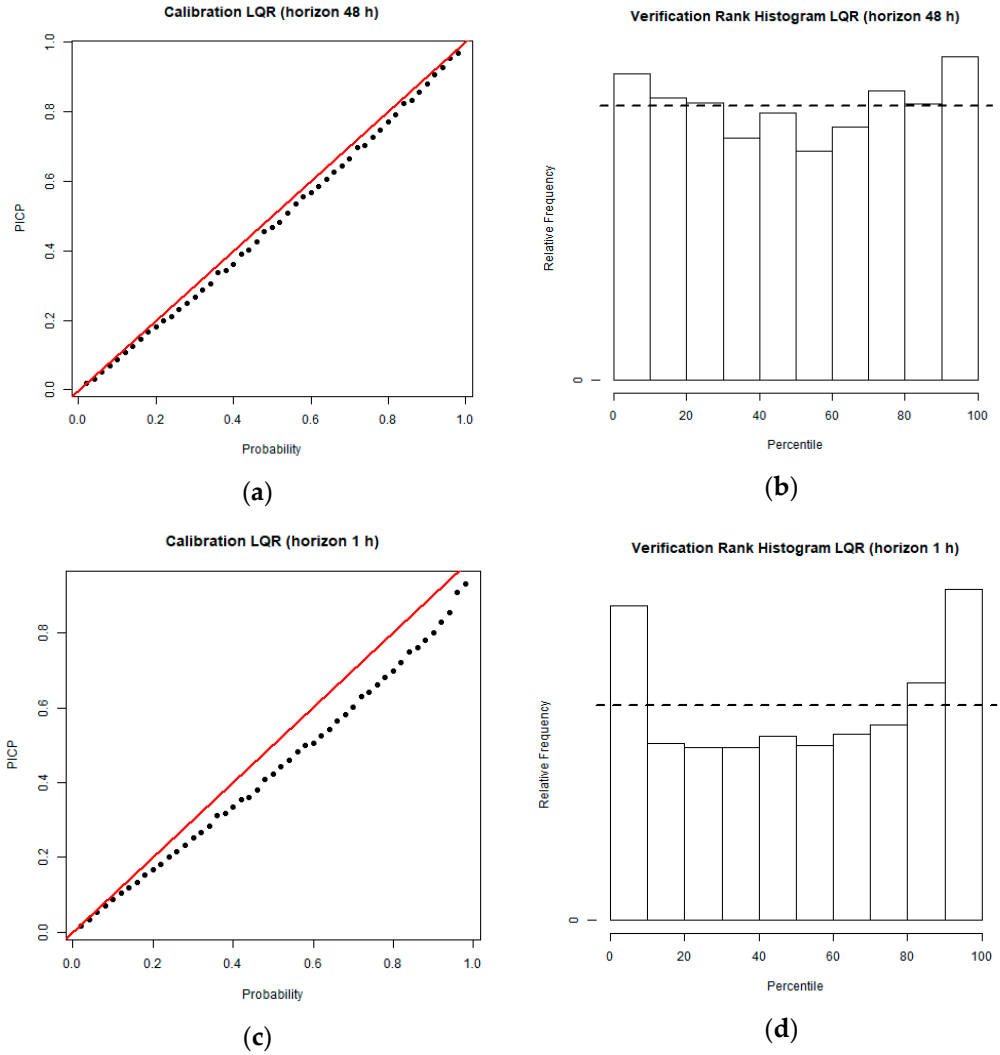


Figure 2. Calibration properties of the LQR models: (**a**) PCIP for the LQR48 model (black dots) and the bisector of the first quadrant (red line); (**b**) VRH for the LQR48 model; (**c**) PCIP for the LQR1 model (black dots) and the bisector of the first quadrant (red line); (**d**) VRH for the LQR1 model.

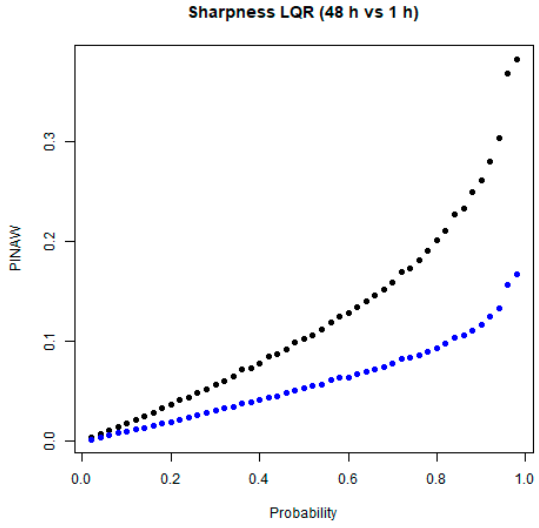


Figure 3. Sharpness property of the LQR models: *PINAW* for the LQR48 model (black dots) and for the LQR1 model (blue dots).

Table 6. Hyperparameters and forecasting results of *QRF* models for the testing dataset.

	QRF48 (Horizon 48 h)	QRF1 (Horizon 1 h)
Ntree	1000	1000
Mtry	27	61
min.node.size	5	5
MAE (kW)	154.8	99.8
RMSE (kW)	213.3	139.9
CRPS (kW)	112.3	72.3

Additionally, Figures 4 and 5 show the behavior of these probabilistic predictions as regards calibration and sharpness. The sharpness is higher for the prediction horizon of 1 h, as expected. However, in contrast to what happened with the LQR model, now the calibration property is better in the case of a prediction horizon of 1 h (QRF1) than for a prediction horizon of 48 h (QRF48).

Comparing the *QRF* and LQR methods, we can see that the calibration properties are similar for a horizon of 48 h, but much better for the *QRF* in the case of a horizon of 1 h. Regarding the sharpness properties, Figure 6 reveals a better behavior for the *QRF* method than for LQR, especially for the prediction horizon of 48 h.

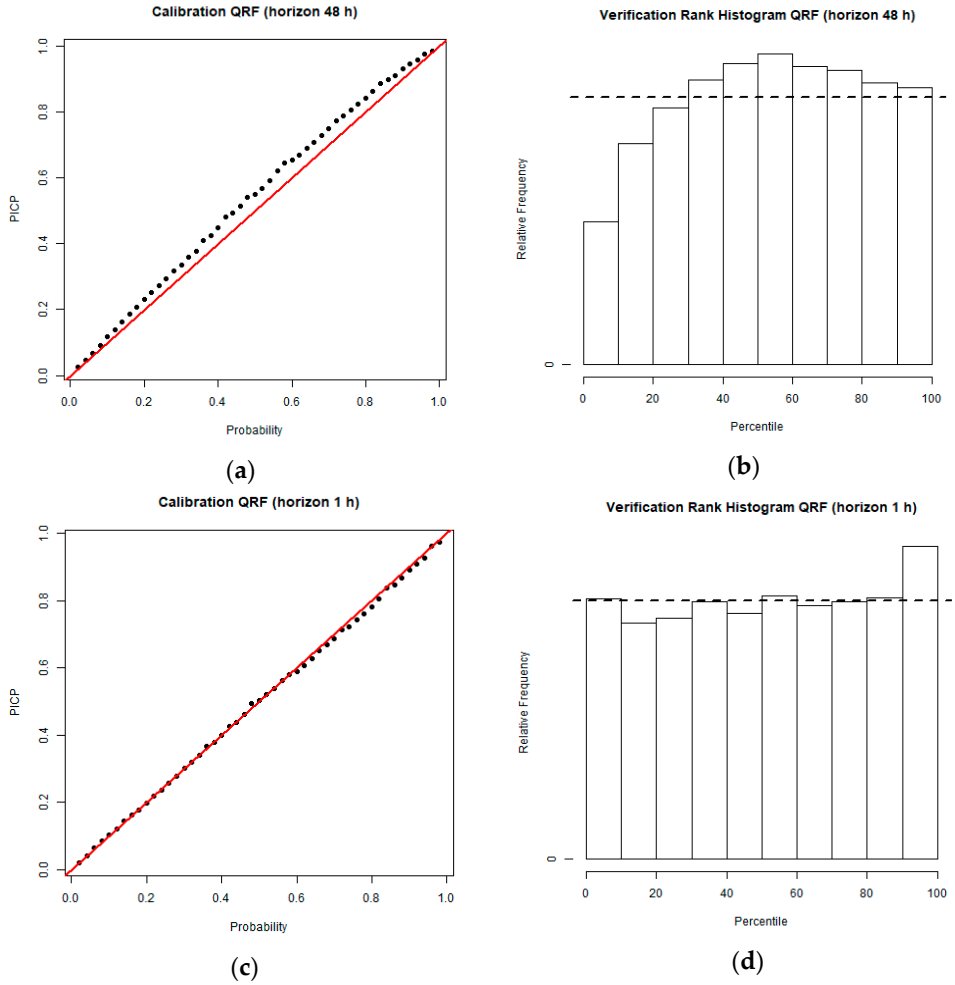


Figure 4. Calibration properties of the *QRF* models: (**a**) PCIP for the QRF48 model (black dots) and the bisector of the first quadrant (red line); (**b**) VRH for the QRF48 model; (**c**) PCIP for the QRF1 model (black dots) and the bisector of the first quadrant (red line); (**d**) VRH for the QRF1 model.

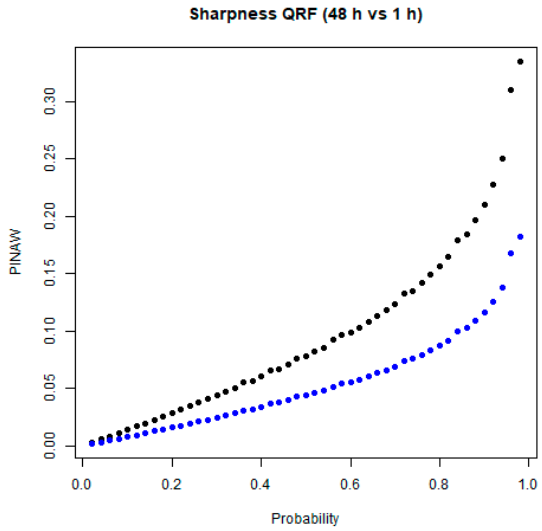
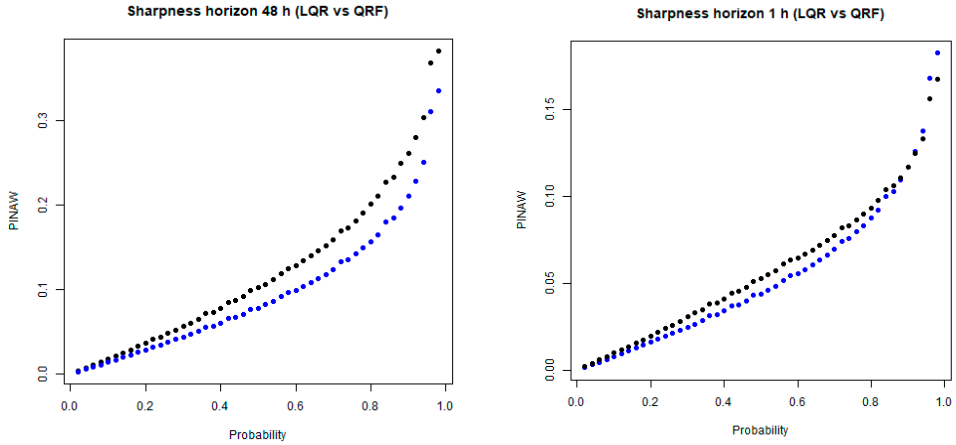


Figure 5. Sharpness properties of the *QRF* models: *PINAW* for the QRF48 model (black dots) and the QRF1 model (blue dots).



(**a**) (**b**)

Figure 6. Comparation of sharpness properties for the LQR (black dots) and *QRF* (blue dots) models: (**a**) *PINAW* for a prediction horizon of 48 h; (**b**) *PINAW* for a prediction horizon of 1 h.

3.2. Application 1: Quantifying the Coverage Probability for Daily DR Actions

In this section, we illustrate how to compute the daily DR capacity of the aggregation in terms of coverage probability. That is, for each day of the test period, we want to know which interval of uncertainty with respect to the predicted baseline can be covered by DR actions.

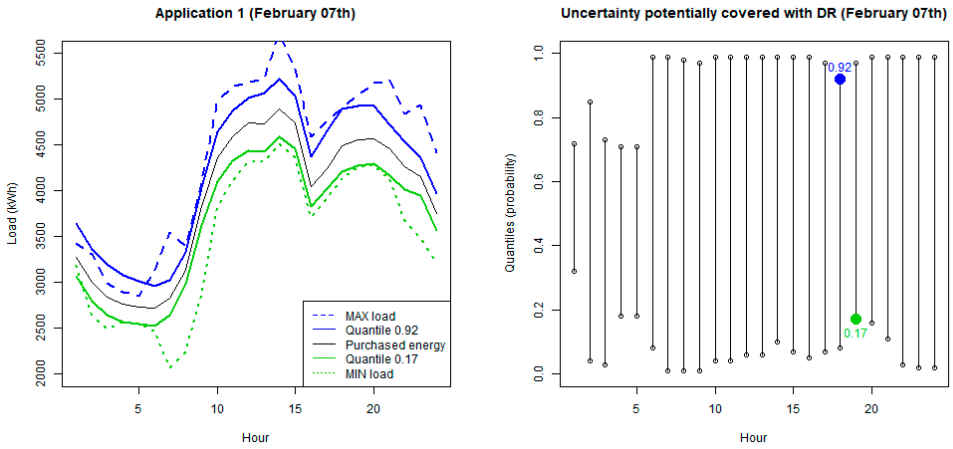
The process can be summarized as follows:

- Step 1: Select a day of the test period and obtain the predicted baseline of consumption for that day. In this step, the predicted baseline is given by the predicted alpha-quantile for each of the 24 h of the selected day, with alpha = 0.5 (the median) and a prediction horizon of 48 h.
- Step 2: Use the PBLM to simulate the maximum increase and maximum decrease in consumption by means of DR actions on the selected day. It is important to recall that comfort conditions must hold.

The 24 h curves of this step will be subtracted and added, respectively, to the baseline, providing the lower and upper bands of the energy covered by DR actions.

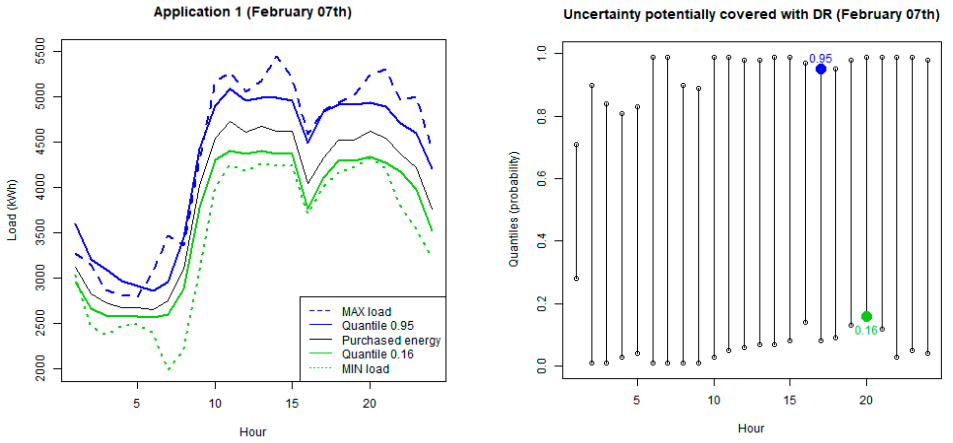
- Step 3: Determine the predicted quantile curves (horizon 24 to 48 h) that are closer to the bands obtained in the previous step.

For example, Figures 7 and 8 depict the three steps of the process for one day in the test period (7 February 2011) and the two predictive methods. In the case of the LQR method, we can ensure that at least 75% of the uncertainty can be potentially covered by DR actions during the main hours (from 10 a.m. to 8 p.m.), whereas in the case of the *QRF*, the percentage of coverage could reach 79%.



(**a**) (**b**)

Figure 7. Example of Application 1 for the date 7 February 2011 and the LQR method: (**a**) results of steps 1 to 3 using the LQR48 model; (**b**) quantiles potentially covered by DR using the LQR48 model.



(**a**) (**b**)

Figure 8. Example of Application 1 for the date 7 February 2011 and the *QRF* method: (**a**) results of steps 1 to 3 using the QRF48 model; (**b**) quantiles potentially covered by DR using the QRF48 model.

Let us focus on what happened from 10 a.m. to 8 p.m. on 7 February, using the *QRF* method (Figure 8). It can be seen that the quasi-homogeneous control group considered for the two end-uses was able to cover nearly 100% of the uncertainty at 10 a.m., a little less at 11 a.m., and so on. In particular, the lowest value of the upper coverage interval (95%, filled dot in blue color) was reached at 5 p.m., whereas the greatest value of the lower coverage interval (16%, filled dot in green color) was reached at 8 p.m. This means that during all the intervals of interest (from 10 a.m. to 8 p.m.), the flexibility provided by the QHCG would cover the interval determined by quantiles 16% and 95%, which implies that 79% (i.e., 95–16%) of the uncertainty due to predictions would have been covered.

3.3. Application 2: Determining the Number of Loads Needed for DR Actions

Another interesting application consists of determining the number of loads needed for DR actions for a given coverage probability. That is, we want to calculate how many

loads (of each type) we would need to cover with DR actions, with a given percentage of uncertainty with respect to the predicted baseline.

The process can be summarized as follows:

- Step 1: Select a day of the test period and obtain the predicted baseline of consumption for that day. In this step, the predicted baseline is given by the predicted α-quantile for each of the 24 h of the selected day, with $\alpha = 0.5$ (the median) and a prediction horizon of 48 h. Also, select a coverage probability $\beta \in (0, 1)$.
- Step 2: For each of the 24 h of the selected day, obtain the centered predicted quantiles corresponding to the given coverage probability β. Denote with $\hat{Q}_{h,\,\frac{1-\beta}{2}}$ and $\hat{Q}_{h,\,\frac{1+\beta}{2}}$ those predicted quantiles for each hour h.

  Compute the energy to rise by $\hat{Q}_{h,\,0.5} - \hat{Q}_{h,\,\frac{1-\beta}{2}}$ and the energy to fall by $\hat{Q}_{h,\,\frac{1+\beta}{2}} - \hat{Q}_{h,\,0.5}$ for each hour h.
- Step 3: Use the PBLM to simulate the number of loads needed to increase or decrease, with DR actions, the energy obtained in the previous step.

For example, Figures 9 and 10 depict the results of Application 2 for three days in the test period, a given coverage probability of 70%, and the two forecasting methods LQR and *QRF*. For the LQR method, the round number of loads/appliances needed during the testing period was 650 electric heaters and 900 Water Heaters. For the *QRF* method, the round number of appliances needed was 600 electric heaters and 750 Water Heaters. Obviously, each day has a different level of demand, and this evaluation makes it possible to consider 48 h ahead an estimated number of customers and loads that should be involved in DR. As an example, three different days have been selected to obtain the number of EHs and WHs involved. Table 7 shows the results.

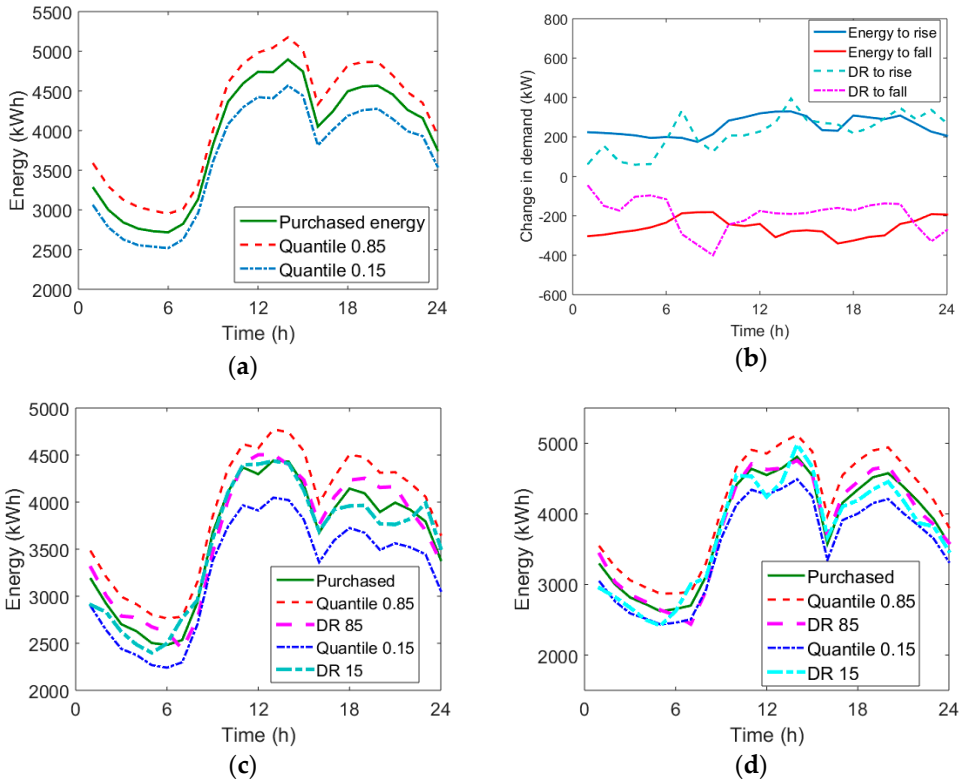


Figure 9. Example of Application 2 for three days in the test period. Coverage probability of 70% and method LQR: (**a**) purchased energy, and predicted quantiles 0.15 and 0.85 using the LQR48 model on 7 February 2011; (**b**) energy to rise and fall according to step 2 on 7 February 2011; (**c**) purchased energy, predicted quantiles 0.15 and 0.85 and DR efforts on 26 October 2010; (**d**) purchased energy, predicted quantiles 0.15 and 0.85 and DR efforts on 2 December 2010.

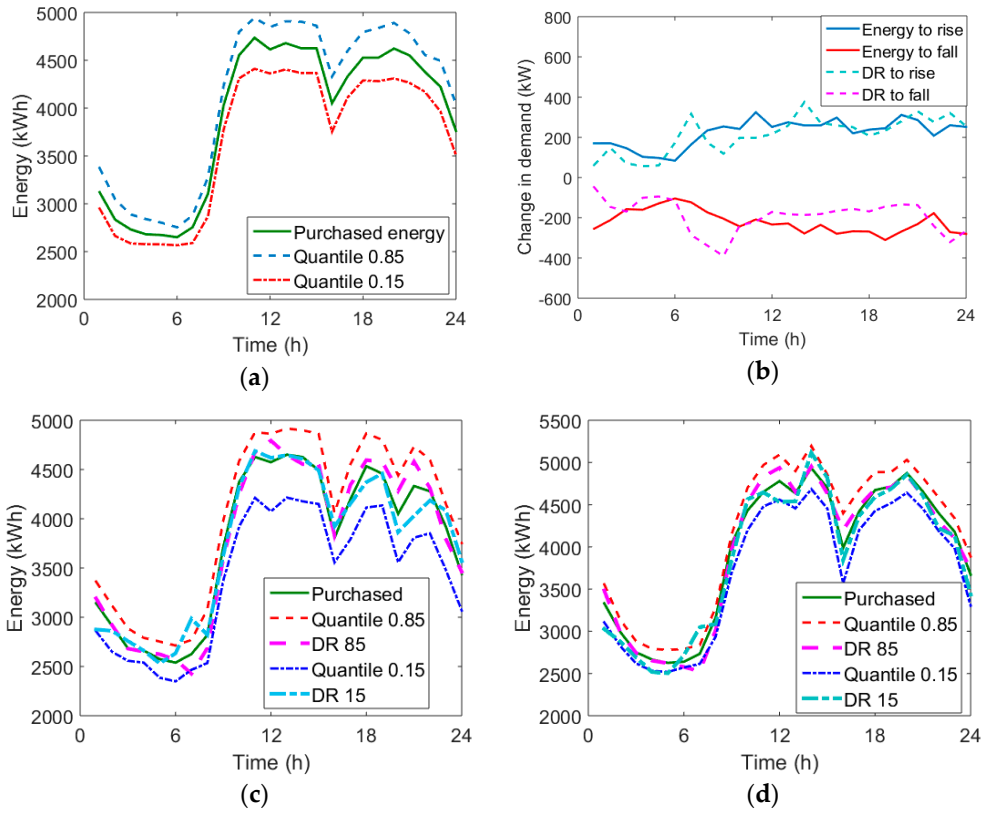


Figure 10. Example of Application 2 for three days in the test period. Coverage probability of 70% and method *QRF*: (**a**) purchased energy, and predicted quantiles 0.15 and 0.85 using the QRF48 model on 7 February 2011; (**b**) energy to rise and fall according to step 2 on 7 February 2011; (**c**) purchased energy, predicted quantiles 0.15 and 0.85 and DR efforts on 26 October 2010; (**d**) purchased energy, predicted quantiles 0.15 and 0.85 and DR efforts on 2 December 2010.

Table 7. Approximate number of loads to be involved in DR policies by day.

	QRF		LQR	
Day	**WH**	**EH**	**WH**	**EH**
26 October	620	575	640	595
2 December	870	625	735	530
7 February	455	435	445	420

3.4. Application 3: Establishing an Hourly Strategy for the Target Day

The third application involves probabilistic forecasting with a 1 h horizon to establish an effective DR strategy. The goal is to correct deviations between the electricity purchased (hourly baseline for a specific day) and actual consumption. To achieve this, we use 1 h horizon probability forecasts to estimate the energy to be saved or increased each hour, thereby setting a DR schedule for aggregation.

The process is as follows:

- Step 1: Select a day from the test period and obtain the hourly predicted baseline (purchased energy) for that day with a 48 h horizon, as in other applications.
- Step 2: Obtain the probabilistic forecasts (predicted quantiles) with a 1 h horizon for all hours of the selected day.
- Step 3: For each hour of the selected day, estimate the probability that actual consumption will be lower or higher than the purchased energy. This involves determining the position of the purchased energy among the predicted quantiles with a 1 h horizon.

For instance, if the purchased energy at hour h, E_h, is between the $\hat{Q}_{h,0.05}$ and $\hat{Q}_{h,0.06}$ predicted quantiles, we estimate $p_h = P(Y_h < E_h) = 0.05$, where Y_h denotes the actual consumption at hour h.

- Step 4: Establish criteria to determine DR actions, applying them only when there is a high certainty that saving or increasing energy is needed:

If p_h is in the range [0.25, 0.75], no DR actions are applied.

If $p_h < 0.25$, there is a high likelihood that actual demand will exceed purchased energy. In this case, DR actions are taken at hour h to save the energy difference between $\hat{Q}_{h,0.5}$ (the predicted median with a 1 h horizon) and the purchased energy E_h.

If $p_h > 0.75$, there is a high likelihood that actual demand will be less than purchased energy. In this case, DR actions (DR > 0, see Figure 11) are taken at hour h to increase the energy difference between E_h and $\hat{Q}_{h,0.5}$ (the predicted median with a 1 h horizon).

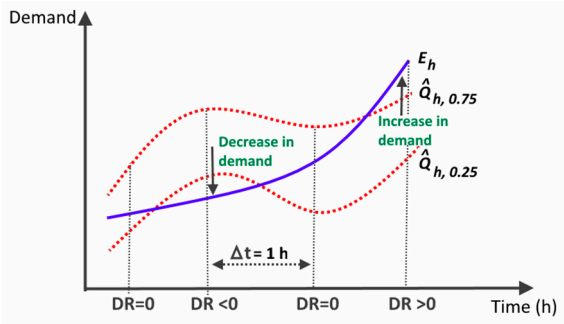


Figure 11. Example of application of step 4. The purchased energy E_h in markets is compared with 1 h predictions (quantiles are selected by the energy aggregator). If E_h is greater than energy defined by bands (in red), DR actions should increase demand—for example, through energy storage (DR > 0). If the blue curve is under the demand band, DR should be applied to clip demand (DR < 0).

To better illustrate the procedure from a practical point of view, and to present some different possibilities for the aggregator, three different days during the test period have been considered: 26 October and 2 December 2010, and 7 February 2011. The choice of these days is justified by the forecasted demand (and the considered representative quantiles) and the weather conditions that drive flexible end-use (WH and EH) demand. Due to these boundary conditions, the necessary flexibility requires different values and trends in different periods.

Figure 12a,b shows the demand quantile 0.50 for 48 h and 1 h forecasts for two of the three selected days with the *QRF* and LQR methods, labelled as "Demand (time-horizon, quantile)". These figures correspond to steps 1 and 2 of the proposed method/process. Note that flexibility requirements are different. As has been discussed before, in the short term, a 1 h forecast is more accurate (Table 4) and will drive the target for demand flexibility. In our case, quantile 0.50 with a horizon of 48 h has been chosen as the baseline for demand without DR. Obviously, the "real" load can increase and decrease in different magnitudes throughout the selected days with respect to the baseline in different periods. This represents interesting performance tests for demand flexibility (DR) and PBLM performance and usefulness.

According to step 3, the aggregator (and considering the number of loads defined in the previous subsection, i.e., Application 2) defines a target hour ahead (Target2 in Figure 13) to reduce or increase demand in each group of responsive loads (QHCG) with respect to the QHCG baseline demand. An example for EH and WH aggregated control groups on 26 October is depicted in Figure 13.

It is also interesting to analyze the stress that each load group suffers due to the compliance of DR targets. This can be considered through an effort curve (the ratio between the increase/decrease in each end-use demand over the baseline level for this end-use). Figure 14 shows this daily curve on 26 October again. In Figure 14a, each group develops a similar flexible pattern of demand, and it is seen that the effort does not surpass 0.3 p.u.

(the reader should note that higher effort values in the first hours of the day, i.e., from 1 to 5 a.m., are not significant because demand is very low in those periods and there is not a problem from the point of view of penalties due to failure in forecasts, e.g., energy imbalance concerns). A flexibility value around 30–40% is a credible level of response for WHs and EH.

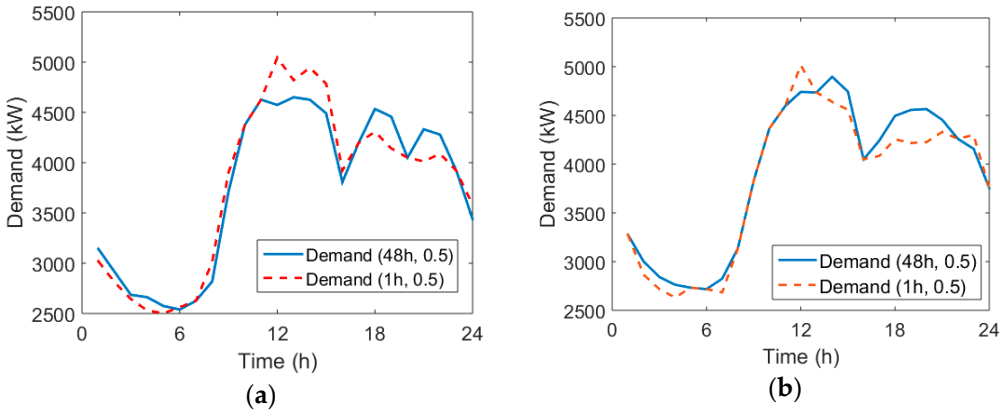


Figure 12. Demand forecasts at 48 h and 1 h (quantile 0.5) for two days and methods: (**a**) 26 October, method *QRF*; (**b**) 7 February, LQR method.

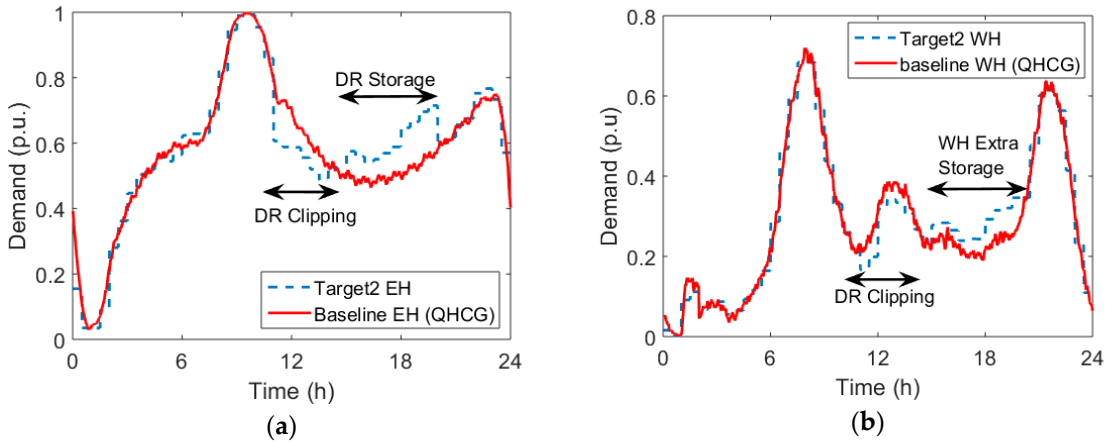


Figure 13. Targets for responsive demand in each QHCG on 26 October according to *QRF* forecast: (**a**) Electric Heating; (**b**) water heating.

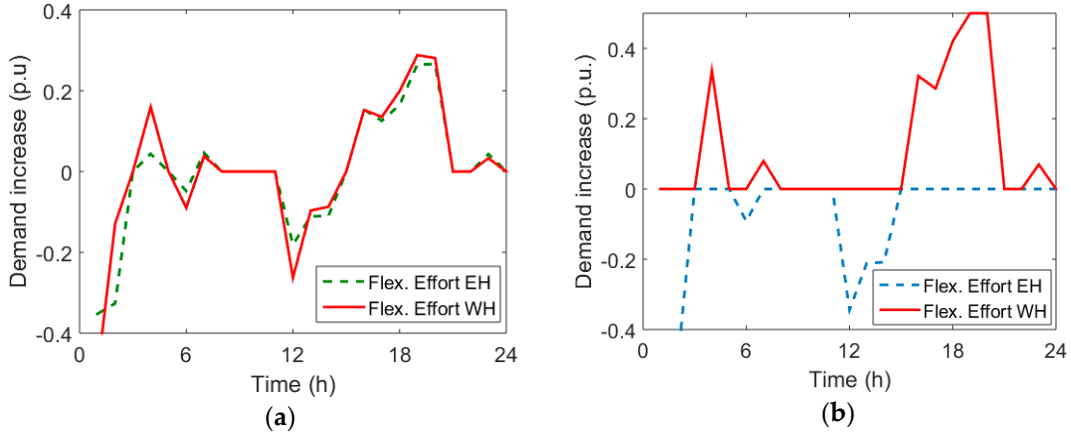


Figure 14. Flexibility effort for responsive demand in each QHCG on 26 October according to *QRF* forecast: (**a**) EH and WH QHCGs accomplish rise and decrease in demand; (**b**) each QHCG engages in a specific task (demand clipping or energy storage).

According to the experience of each aggregator with its customers' response, these efforts can be changed and QHCG targets can be different. For example, another option to be considered is to "specialize" each QHCG in a flexible task, i.e., WHs only increase demand and HVAC shaves demand. This is presented in Figure 14b. In this paper, this scenario is not considered because the main interest is to show the ability of flexible loads to perform in complex DR contexts. Nevertheless, draft results show that this specialization does not have a significant impact on the overall results, but from the point of view of energy efficiency this option could be of interest. For these reasons, a more detailed study will be presented in a future paper.

According to the DR targets that have been presented in Figure 13a,b, the QHCGs' response has been simulated, hour by hour, during a day. Figure 15a presents the performance of EH QHCGs with respect to the DR targets presented in Figure 13a. The figure shows that the DR targets are not reached perfectly, mainly because PBLM limits fast changes in demand (e.g., from 19 h to 20 h) to avoid short peaks and demand rebounds after clipping periods. As stated before, DR requires a surveillance of flexibility effects in load service (i.e., assuming some minimum thresholds for customer comfort—in our case, temperatures inside the dwelling or hot water temperatures, Section 2). The effects are depicted in Figure 15b through PBLM state variables (indoor temperature in the case of EH, or temperatures in the water tank in the case of WHs).

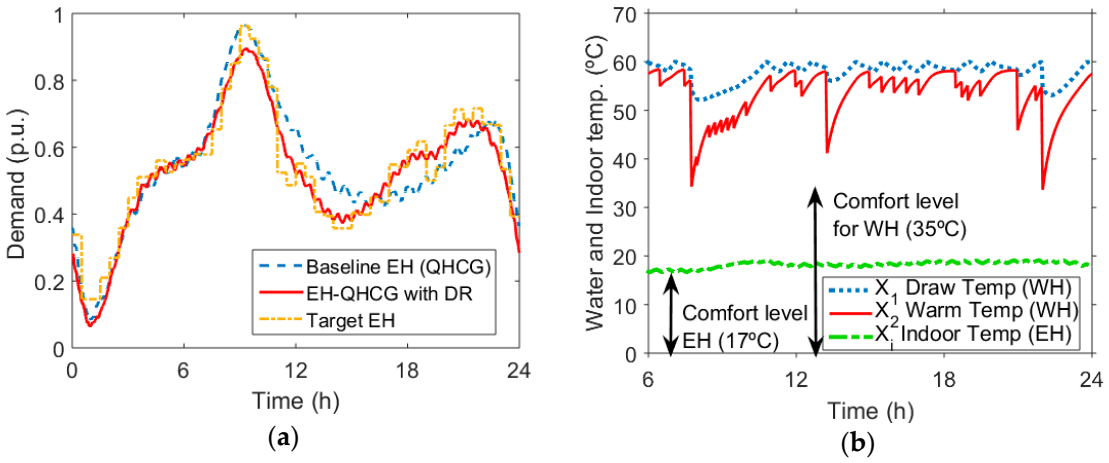


Figure 15. Response of QHCGs (26 October): (**a**) performance of EH QHCGs, (**b**) example of comfort surveillance during DR performance on EH and WH groups.

Another example is presented in Figure 16. On 7 February, the aggregator required (Figure 12b) an almost rising profile of demand to account for an excess of energy forecasted in the 48 h horizon (LQR method). This involved a non-typical DR action (i.e., a rise in demand). Figure 16a represents the target for WH loads, and Figure 16b presents simulation results though PBLMs. As can be seen, WH QHCGs can also follow changes in demand specified by a target (Target2 WH in Figure 13b) with good accuracy but filtering fast changes in the demand reference/target. To simplify the simulation and avoid overdamping, the control reference signal applies every 15 min, and any change is implemented in PBLM scripts if the foreseen change in demand (Target 2) is less than 0.03 p.u., i.e., a deadband is applied in PBLM control procedures.

From a practical point of view, and to implement a demand increase, the setpoint of the WH thermostat is changed from 58–60 °C to 63–65 °C during the "WH extra-storage" period (Figure 16a). On the other side, EH and WH QHCGs are controlled through pulse-wave-modulated "ON/OFF" signals to reduce demand ("DR clipping" label in Figure 13a). The change in the thermostat setting (a surplus around 1–2 °C) is applied to increase demand ("DR storage" label, Figure 13a).

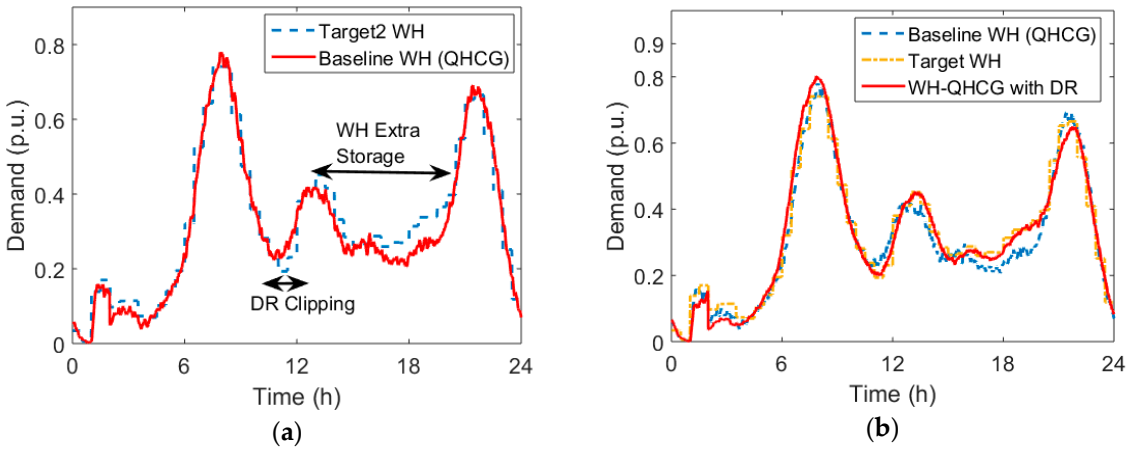


Figure 16. Response of QHCGs (7 February): (**a**) DR target and baseline for WH QHCG, (**b**) DR performance of WH groups in the daily period.

Once the targets, the performance of load groups, and the surveillance of customer comfort have been stated through some examples, it seems necessary to avoid an excessive repetition of the basic steps of the methodology for each end-use and day. For this reason, the main results showing the compliance with the objectives of this section are summarized in Figures 17 and 18 for the three selected days.

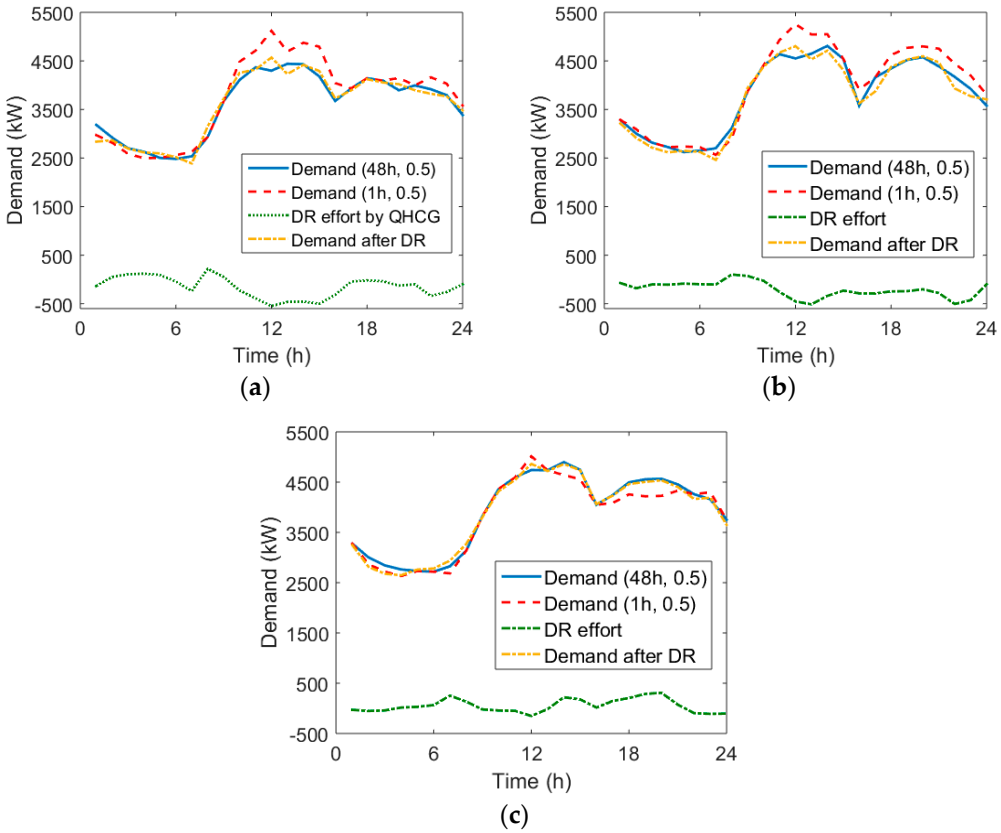


Figure 17. Results for Application 3: fitting of short-term 1 h demand forecast to purchased energy (48 h forecast through LQR) and overall demand flexibility of EH and WH QHCGs: (**a**) 26 October; (**b**) 2 December; (**c**) 7 February.

Figures 17 and 18 show the results for the LQR and *QRF* methods, respectively. These Figures show the energy purchased in markets (forecasted 48 h ahead through quantile 0.5 and labelled "Demand 48 h, 0.5") and the revisited demand, 1 h ahead ("Demand 1 h, 0.5"),

both forecasted through *QRF* and LQR. As stated before, a 1 h forecast is considered as the estimator for the actual demand. Moreover, the overall flexibility of both QHCG groups ("DR effort") and the adjustment of forecasted "actual demand" through DR (labelled "Demand after DR") are also shown in Figures 17 and 18. At first sight, the performance of QHCGs can be considered good enough in most of the time periods both to decrease and also to increase demand (a non-usual DR policy for loads).

Figure 18 shows the performance of DR but for *QRF* forecasts. The labels and days are the same, but logically forecast levels change (errors and patterns) and consequently flexibility requirements.

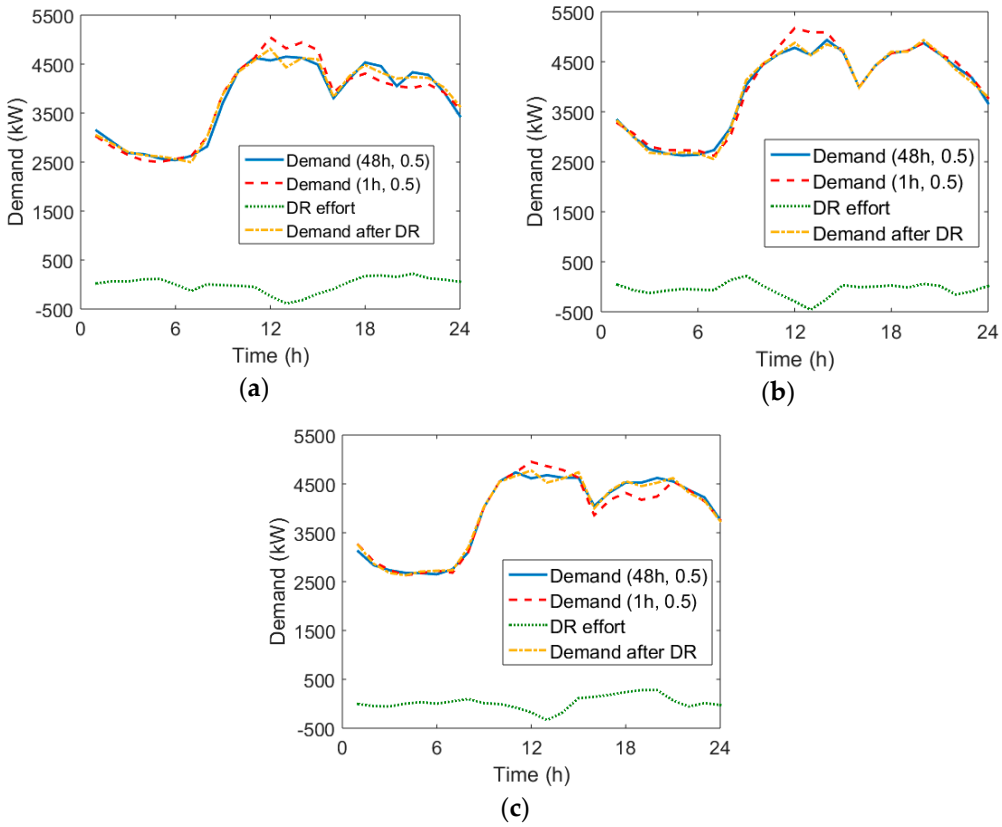


Figure 18. Results for Application 3: fitting of short-term 1 h demand forecast to purchased energy (48 h forecast through *QRF*) and overall demand flexibility of EH and WH QHCGs: (**a**) 26 October; (**b**) 2 December; (**c**) 7 February.

Some external factors can explain the patterns of demand from 1 h to 48 h forecasts. For instance, in Figures 17a and 18a, demand rises from 10 h to 14 h, perhaps due to a change in the external temperature that was not well forecasted by meteorological models 48 ahead. In this scenario, the aggregator requests flexibility to QHCGs to avoid penalties due to demand unbalance. In addition, demand decreases in the afternoon in Figure 18a (Demand (1 h, 0.5)) with respect to 48 h forecasts (Demand (48 h, 0.5)) but this pattern changes for the same day in Figure 17a when a different forecast method is applied. In this way, the aggregator needs to increase or decrease their aggregated QHCG demand according to energy bid levels sent to markets 48 h ago, irrespective of the previous flexibility trends required some hours before for the same day. In both cases, demand flexibility covers to a considerable extent and with good accuracy those needs.

It is also interesting to consider some insights from previous figures. First, the reader should note that soft changes in demand levels (from 48 h to 1 h forecasts) give a better fitting of flexibility to required levels (e.g., Figure 17c). Second, the required flexibility levels for the LQR method are higher than for the *QRF* method (around 100 kW up or down

in most cases). This can be partially explained through performance indicators for each forecast method (Tables 5 and 6).

Table 8 summarizes the improvement provided after running Application 3 to the three selected days in the test period and for the two probabilistic forecasting methods (LQR and *QRF*).

Table 8. Reduction in *RMSE* after applying Application 3 to the three selected days.

Day	Method	*RMSE* (Before Applying App3)	*RMSE* (After Applying App3)	Percentage of Error Reduction
26 October	LQR	301.37	178.61	40.73%
	QRF	210.02	14.45	29.79%
2 December	LQR	262.57	167.78	36.10%
	QRF	174.50	152.44	12.64%
7 February	LQR	173.73	149.91	13.71%
	QRF	156.52	119.44	23.69%

Additionally, Table 9 shows the error reduction in terms of the *RMSE* when the strategy proposed in Application 3 is applied to all days in the test period. In this case, perfect DR actions are assumed in the following sense: the energy to save or to increase at each hour, determined by the steps described in Application 3, is supposed to be achieved. As it can be seen, the *QRF* method is more accurate than the LQR, in the scenario of no DR actions as well as in the case of applying DR according to Application 3. However, the percentage of error reduction is greater for the LQR method, which can be considered an added value: the differences of accuracy among the predictive methods can be cushioned after applying the strategy proposed in Application 3.

Table 9. Reduction in *RMSE*s after applying Application 3 to the whole test dataset.

Method	*RMSE* (Before Applying App3)	*RMSE* (After Applying App3)	Percentage of Error Reduction
LQR	313.61	185.20	40.95%
QRF	213.33	141.61	33.62%

The evaluation of the error reduction is also depicted in Figure 19, which includes several comparing graphs with the evolution of the *RMSE* for all days in the test period. The significant reduction in error after applying Application 3 is latent in Figure 19a,b, corresponding to the LQR and *QRF* methods, respectively. On the other hand, Figure 19a,b reveals a better performance of the *QRF* method against LQR, which is cushioned after applying Application 3.

The lack of accuracy in predictions comes from different sources (weather forecast inaccuracies or the forecast method selected, among others). As we have confirmed, there is a significative difference in 48 h-ahead predictions' accuracy depending on the method employed (LQR versus *QRF*), and the impact of those differences was reduced after applying the DR strategy proposed in this paper. That is, the capacity for improvement is greater when the initial 48 h-ahead baseline is worse.

Although weather forecast inaccuracies affect demand predictions, it can also be proved that if we had used the real values of climatic factors (instead of their forecasts), the results would not have changed significatively. The reason is similar to before: the 48 h-ahead baseline of demand would be significatively affected by the lack of accuracy in the climatic factors, but not the 1 h-ahead baseline. Therefore, a suitable DR strategy like the one proposed in the paper will be the most important factor to obtain similar results (low deviations between the purchased energy and final consumption).

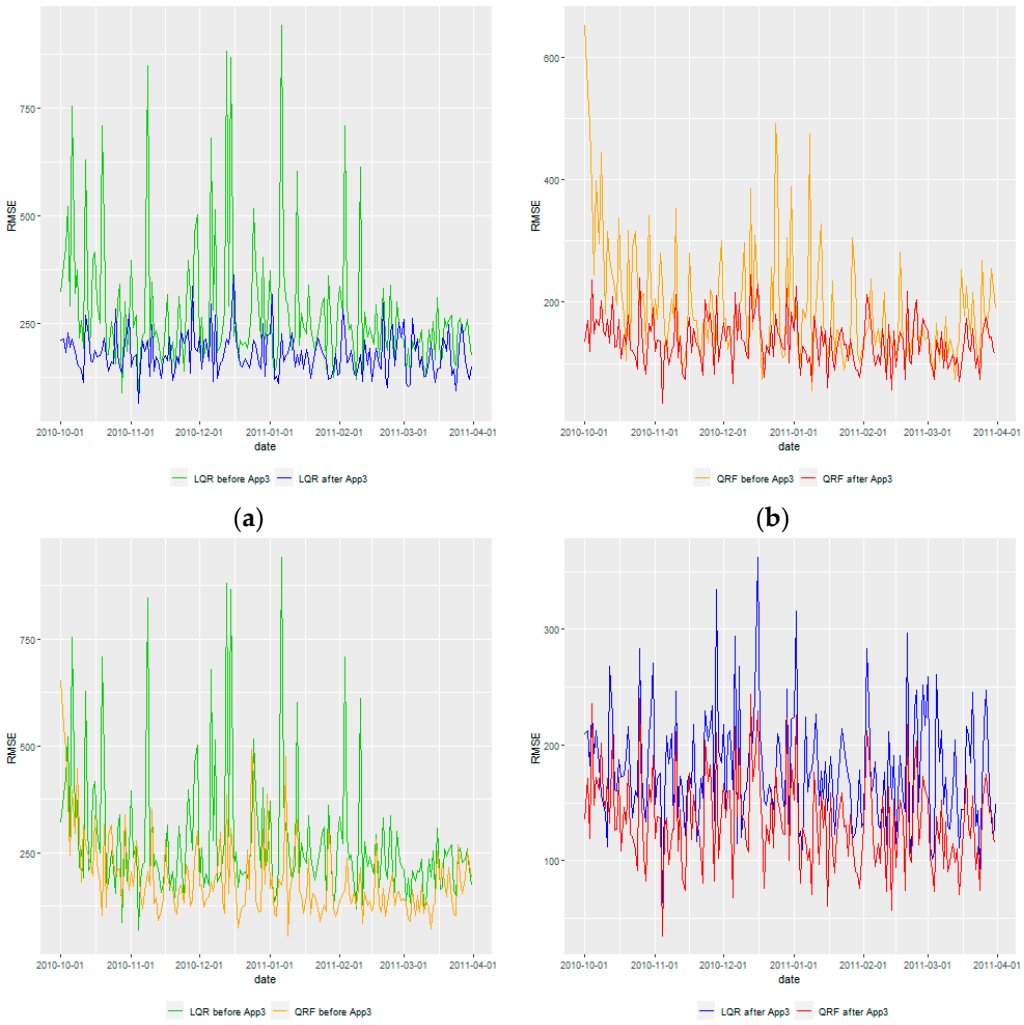


Figure 19. *RMSE* during the test period before and after applying Application 3 (assuming perfect DR actions): (**a**) comparative results for the LQR method; (**b**) comparative results for the *QRF* method; (**c**) comparative results between predictive methods before applying Application 3; (**d**) comparative results between predictive methods after applying Application 3.

4. Conclusions

Probabilistic forecasting is shown as a useful tool in the context of demand response, together with Physical-Based Load Models—an aspect that is reflected throughout the paper by means of three specific applications. On the one hand, it allows us to quantify the percentage of uncertainty that DR actions could potentially cover during a target day for predictions with a horizon of 48 h (the usual time horizon when purchasing the energy on the market). On the other hand, the forecasted quantiles can be used to compute the number of consumers needed for DR actions.

Two probabilistic forecasting methods have been used to illustrate the proposed applications: Lineal Quantile Regression and a Quantile Regression Forest. For the analyzed dataset (hourly data of a small city in Spain), the *QRF* method showed better performance properties than LQR regarding accuracy, calibration, and sharpness. In both cases, results reveal that the percentage of uncertainty covered by DR is lower in the day hours from 10 a.m. to 8 p.m., as expected.

Apart from these two applications, a third one is also developed in detail. Unlike the first application, in Application 3 we do not focus on the theoretical potential of DR but propose a specific strategy based on the quantiles resulting from the probabilistic forecast-

ing. The results reveal a significant reduction in errors (deviations of real consumption from purchased energy) for the two forecasting methods (LQR and *QRF*), with a greater reduction in the case of the one that is less accurate. Consequently, this work shows that demand response can provide high levels of flexibility in the residential sector. This is significant because flexibility needs to increase in the future to avoid security issues and high generation costs, thus achieving balance in the power system. Additionally, this flexibility substantially reduces short-term load forecasting uncertainty, enabling the creation of a new portfolio of more reliable resources on the demand side. Consequently, demand response alleviates some energy and balance concerns for customers, aggregators, distributors, and other stakeholders (e.g., balancing responsible parties) in future scenarios.

Moreover, the increase in flexibility (both upward and downward) is particularly valuable when generation sources, primarily renewables, can heighten the volatility of supply resources, thus increasing the probability of imbalances. This is critical from a security of supply perspective. These imbalances can occur at small scales (prosumers) or across the entire power system. The main challenge is to define flexibility targets using short-term load forecasting methods and to analyze the potential flexibility of demand-side resources through simulation models. Additionally, probabilistic methods allow for the consideration of different risk levels according to the profiles of customers and aggregators and, naturally, the potential benefits or revenues associated with those risks in markets or services.

Author Contributions: Conceptualization: L.A.F.-J., A.G. (Antonio Gabaldón), and M.C.R.-A.; methodology: L.A.F.-J., A.G. (Antonio Gabaldón), and M.C.R.-A.; software: L.A.F.-J., A.G. (Antonio Gabaldón), M.C.R.-A., and A.G. (Antonio Guillamón); validation: A.G. (Antonio Gabaldón) and A.G. (Antonio Guillamón); formal analysis: L.A.F.-J. and M.C.R.-A.; investigation: all authors; resources: M.C.R.-A. and A.G. (Antonio Guillamón); writing—original draft preparation: all authors; writing—review and editing: L.A.F.-J., A.G. (Antonio Gabaldón), and M.C.R.-A.; visualization: A.G. (Antonio Guillamón); supervision, L.A.F.-J. and M.C.R.-A.; project administration: A.G. (Antonio Gabaldón) and M.C.R.-A.; funding acquisition: L.A.F.-J., A.G. (Antonio Gabaldón), and M.C.R.-A. All authors have read and agreed to the published version of the manuscript.

Funding: Grants TED2021-129722B-C33 and TED2021-129722B-C32 funded by MICIU/AEI /10.13039/501100011033 and by the European Union Next Generation EU/PRTR.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy concerns of customers that have been monitored in some demand tests.

Acknowledgments: Authors are very grateful to the Agencia Estatal de Investigación (Spanish Government) for its financial support throughout the research projects TED2021-129722B-C32, TED2021-129722B-C33 and to the European Union for providing additional funds and research support through the NextGeneration/PRTR program.

Conflicts of Interest: The authors declare no conflicts of interest.

IGOs License: This is an open access article distributed under the terms of the Creative Commons Attribution IGO License (CC BY) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations

CBL	Customer Baseline Load
CPRS	Continuous Rank Probability Score
DERs	Distributed energy resources
EE	Energy efficiency

ERWH	Electric Resistance Water Heater
EH	Electric Heating
HPWH	Heat Pump Water Heater
HVAC	Heat and Ventilation Air Conditioning
LQR	Lineal Quantile Regression
MAPE	Mean Absolute Percent Error
MPE	Mean Percent Error
PBLM	Physical-Based Load Model
PINAW	Prediction Interval Normalized Average Width
PICP	Prediction Interval Coverage Probability
QHCG	Quasi-homogeneous control group (of loads)
QRF	Quantile Regression Forest
RMSE	Root Mean Square Error
WH	Water Heater
VRH	Verification Rank Histogram

References

1. ACER. European Environment Agency. 2024. Flexibility Solutions to Support a Decarbonised and Secure EU Electricity System. Available online: https://www.eea.europa.eu/publications/flexibility-solutions-to-support (accessed on 16 September 2024).
2. DSO Entity & ENTSO-E Public consultation on Network Code for Demand Response—European Network of Transmission System Operators for Electricity—Citizen Space. 2024. Available online: https://consultations.entsoe.eu/markets/public-consultation-networkcode-demand-response/ (accessed on 11 September 2024).
3. Williams, B.; Bishop, D.; Gallardo, P.; Chase, J.G. Demand Side Management in Industrial, Commercial, and Residential Sectors: A Review of Constraints and Considerations. *Energies* **2023**, *16*, 5155. [CrossRef]
4. Siano, P. Demand response and smart grids—A survey. *Renew. Sustain. Energy Rev.* **2014**, *30*, 461–478. [CrossRef]
5. Moghram, I.; Rahman, S. Analysis and Evaluation of Five Short-Term Load Forecasting Techniques. *IEEE Power Eng. Rev.* **1989**, *9*, 42–43. [CrossRef]
6. Mandal, P.; Senjyu, T.; Funabashi, T. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Convers. Manag.* **2006**, *47*, 2128–2142. [CrossRef]
7. Chen, Y.; Xu, P.; Chu, Y.; Li, W.; Wu, Y.; Ni, L.; Bao, Y.; Wang, K. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Appl. Energy* **2017**, *195*, 659–670. [CrossRef]
8. Dudek, G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies* **2022**, *15*, 7547. [CrossRef]
9. Çevik, H.H.; Çunkaş, M. Short-term load forecasting using fuzzy logic and ANFIS. *Neural Comput. Appl.* **2015**, *26*, 1355–1367. [CrossRef]
10. Waheed, W.; Xu, Q. Data-driven short term load forecasting with deep neural networks: Unlocking insights for sustainable energy management. *Electr. Power Syst. Res.* **2024**, *232*, 110376. [CrossRef]
11. Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1352–1372. [CrossRef]
12. Ali, S.; Bogarra, S.; Riaz, M.N.; Phyo, P.P.; Flynn, D.; Taha, A. From Time-Series to Hybrid Models: Advancements in Short-Term Load Forecasting Embracing Smart Grid Paradigm. *Appl. Sci.* **2024**, *14*, 4442. [CrossRef]
13. Xu, M.; Li, Q.; Zhao, Z.; Sun, C. Bilinear-DRTFT: Uncertainty prediction in electricity load considering multiple demand responses. *Energy* **2024**, *309*, 133067. [CrossRef]
14. Gneiting, T.; Katzfuss, M. Probabilistic Forecasting. *Annu. Rev. Stat. Its Appl.* **2014**, *1*, 125–151. [CrossRef]
15. Van der Meer, D.W.; Widén, J.; Munkhammar, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1484–1512. [CrossRef]
16. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
17. He, Y.; Xu, Q.; Wan, J.; Yang, S. Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. *Energy* **2016**, *114*, 498–512. [CrossRef]
18. Salinas, D.; Flunkert, V.; Gasthaus, J.; Januschowski, T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecast.* **2020**, *36*, 1181–1191. [CrossRef]
19. Lu, R.; Hong, S.H. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. *Appl. Energy* **2019**, *236*, 937–949. [CrossRef]
20. Sharma, A.K.; Saxena, A. A demand side management control strategy using Whale optimization algorithm. *SN Appl. Sci.* **2019**, *1*, 870. [CrossRef]
21. Taherian, H.; Aghaebrahimi, M.R.; Baringo, L.; Goldani, S.R. Optimal dynamic pricing for an electricity retailer in the price-responsive environment of smart grid. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 107004. [CrossRef]
22. Esnaola-Gonzalez, I.; Jelić, M.; Pujić, D.; Diez, F.J.; Tomašević, N. An AI-Powered System for Residential Demand Response. *Electronics* **2021**, *10*, 693. [CrossRef]

23. Ceccon, W.F.; Freire, R.Z.; Szejka, A.L.; Junior, O.C. Intelligent Electric Power Management System for Economic Maximization in a Residential Prosumer Unit. *IEEE Access* **2021**, *9*, 48713–48731. [CrossRef]
24. Giglio, E.; Luzzani, G.; Terranova, V.; Trivigno, G.; Niccolai, A.; Grimaccia, F. An Efficient Artificial Intelligence Energy Management System for Urban Building Integrating Photovoltaic and Storage. *IEEE Access* **2023**, *11*, 18673–18688. [CrossRef]
25. Mai, V.; Maisonneuve, P.; Zhang, T.; Nekoei, H.; Paull, L.; Lesage-Landry, A. Multi-agent reinforcement learning for fast-timescale demand response of residential loads. *Mach Learn* **2024**, *113*, 5203–5234. [CrossRef]
26. Bampoulas, A.; Pallonetto, F.; Mangina, E.; Finn, D.P. An ensemble learning-based framework for assessing the energy flexibility of residential buildings with multicomponent energy systems. *Appl. Energy* **2022**, *315*, 118947. [CrossRef]
27. MeteoGalicia. Numerical Weather Simulation Models. Available online: https://www.meteogalicia.gal/modelos/inicio.action (accessed on 16 September 2024).
28. Weather Research & Forecasting Model (WRF) | Mesoscale & Microscale Meteorology. Available online: https://www.mmm.ucar.edu/models/wrf (accessed on 16 September 2024).
29. IDAE (Institute for Energy Efficiency Spain). Energy Consumption in Households—Statistics Explained. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households#Energy_consumption_in_households_by_type_of_end-use (accessed on 19 September 2024).
30. Pinson, P.; Nielsen, H.A.; Møller, J.K.; Madsen, H.; Kariniotakis, G.N. Non-parametric probabilistic forecasts of wind power: Required properties and evaluation. *Wind. Energy* **2007**, *10*, 497–516. [CrossRef]
31. Wilks, D.S. Indices of Rank Histogram Flatness and Their Sampling Properties. 1 February 2019. Available online: https://journals.ametsoc.org/view/journals/mwre/147/2/mwr-d-18-0369.1.xml (accessed on 27 September 2024).
32. Khosravi, A.; Nahavandi, S.; Creighton, D. Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts. *IEEE Trans. Sustain. Energy* **2013**, *4*, 602–610. [CrossRef]
33. Verbois, H.; Rusydi, A.; Thiery, A. Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting. *Sol. Energy* **2018**, *173*, 313–327. [CrossRef]
34. Koenker, R.; Bassett, G. Regression Quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
35. Chernozhukov, V.; Fernández-Val, I.; Galichon, A. Quantile and Probability Curves without Crossing. *Econometrica* **2010**, *78*, 1093–1125.
36. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
37. Gabaldón, A.; García-Garre, A.; Ruiz-Abellón, M.C.; Guillamón, A.; Molina, R.; Medina, J. Management of Railway Power System Peaks with Demand-Side Resources: An Application to Periodic Timetables. *Sustainability* **2023**, *15*, 2746. [CrossRef]
38. Hofstädter, R.N.; Zero, T.; Dullinger, C.; Richter, G.; Kozek, M. Heat capacity and heat transfer coefficient estimation for a dynamic thermal model of rail vehicles. *Math. Comput. Model. Dyn. Syst.* **2017**, *23*, 439–452. [CrossRef]
39. Nejad, S.; Gladwin, D.T.; Stone, D.A. A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states. *J. Power Sources* **2016**, *316*, 183–196. [CrossRef]
40. Zhang, L.; Hu, X.; Wang, Z.; Sun, F.; Dorrell, D.G. A review of supercapacitor modeling, estimation, and applications: A control/management perspective. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1868–1878. [CrossRef]
41. Hu, M.; Xiao, F.; Wang, L. Investigation of demand response potentials of residential air conditioners in smart grids using grey-box room thermal model. *Appl. Energy* **2017**, *207*, 324–335. [CrossRef]
42. Nel, P.J.C.; Booysen, M.J.; Van der Merwe, B. A Computationally Inexpensive Energy Model for Horizontal Electric Water Heaters With Scheduling. *IEEE Trans. Smart Grid* **2018**, *9*, 48–56. [CrossRef]
43. Kiliccote, S.; Olsen, D.; Sohn, M.D.; Piette, M.A. Characterization of demand response in the commercial, industrial, and residential sectors in the United States. *WIREs Energy Environ.* **2016**, *5*, 288–304. [CrossRef]
44. Farooq, A.A.; Afram, A.; Schulz, N.; Janabi-Sharifi, F. Grey-box modeling of a low pressure electric boiler for domestic hot water system. *Appl. Therm. Eng.* **2015**, *84*, 257–267. [CrossRef]
45. Yu, Y.; Narayan, N.; Vega-Garita, V.; Popovic-Gerber, J.; Qin, Z.; Wagemaker, M.; Baue r, P.; Zeman, M. Constructing Accurate Equivalent Electrical Circuit Models of Lithium Iron Phosphate and Lead–Acid Battery Cells for Solar Home System Applications. *Energies* **2018**, *11*, 2305. [CrossRef]
46. Demand Response. Available online: http://www.demandresponse.eu/ (accessed on 17 October 2024).
47. Matlab. Simulink—Simulation and Model-Based Design—MATLAB. Available online: https://www.mathworks.com/products/simulink.html (accessed on 20 September 2024).
48. Capasso, A.; Grattieri, W.; Lamedica, R.; Prudenzi, A. A bottom-up approach to residential load modeling. *IEEE Trans. Power Syst.* **1994**, *9*, 957–964. [CrossRef]
49. Laurent, J.C.; Malhamé, R.P. A physically-based computer model of aggregate electric water heating loads. *IEEE Trans. Power Syst.* **1994**, *9*, 1209–1217. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.