



Article

Quantifying the Uncertainty of Electric Vehicle Charging with Probabilistic Load Forecasting

Yvenn Amara-Ouali ^{1,2,*}, Bachir Hamrouche ¹, Guillaume Principato ^{1,2,3} and Yannig Goude ^{1,2}

¹ OSIRIS, EDF R&D, 91120 Palaiseau, France; bachir.hamrouche@edf.fr (B.H.)

² LMO, University Paris-Saclay, 91400 Orsay, France

³ CELESTE, INRIA, 91400 Orsay, France

* Correspondence: yvenn.amara-ouali@universite-paris-saclay.fr

Abstract: The transition to electric vehicles (EVs) presents challenges and opportunities for the management of electrical networks. This paper focuses on developing and evaluating probabilistic forecasting algorithms to understand and predict EV charging behaviours, crucial for optimising grid operations and ensuring a balance between electricity demand and generation. Several forecasting approaches tailored to different time horizons are proposed across diverse model classes, including direct, bottom-up, and adaptive approaches. In all approaches, the target variable can be the load curve quantiles from 0.1 to 0.9 with 0.1 increments or prediction sets with a target coverage of 80%. Direct approaches learn from past load curves using GAMLSS or QGAM methods. Bottom-up approaches predict individual charging session characteristics (arrival time, charging duration, and energy demand) with mixture models before reconstructing the load curve. Adaptive approaches correct in real-time the prediction sets issued by direct or bottom-up approaches with conformal predictions. The experiments, conducted on real-world charging session data from Palo Alto, demonstrate the effectiveness of the proposed methods with regard to different metrics, including pinball loss, empirical coverage, and RPS. Overall, the results highlight the importance of quantifying uncertainty in load forecasts and the potential of probabilistic forecasting for EV load management.



Academic Editors: Joeri Van Mierlo, Myoungho Sunwoo and Namwook Kim

Received: 22 December 2024

Revised: 27 January 2025

Accepted: 6 February 2025

Published: 9 February 2025

Citation: Amara-Ouali, Y.; Hamrouche, B.; Principato, G.; Goude, Y. Quantifying the Uncertainty of Electric Vehicle Charging with Probabilistic Load Forecasting. *World Electr. Veh. J.* **2025**, *16*, 88. <https://doi.org/10.3390/wevj16020088>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the World Electric Vehicle Association. 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

A key lever for reducing greenhouse gas emissions in the transport sector is the large-scale deployment of electrical vehicles (EVs). This has led to many governments implementing strong pro-EV policies, resulting in an increase in the number of EVs in global markets [1]. The arrival of these vehicles creates challenges in the management of the electrical network while also bringing opportunities in terms of grid flexibility. Indeed, these vehicles will become important assets for managing electricity demand whose charging can be automatically postponed when constraints are high or even used as batteries to reinject power when demand is high (vehicle-to-grid) [2]. All these operations of load optimisation are referred to as smart charging. One of the key elements of an efficient smart charging solution is a strong understanding of charging behaviours, which requires the development of efficient forecasting algorithms to predict them.

Quantifying the uncertainty in load forecasts enhances our understanding of when the grid is most likely to face constraints due to a surge in EV demand (see Section 2 for the related work). Two probabilistic outputs can be given to stakeholders to better

anticipate the EV demand: *quantile forecasts* or *prediction sets*. Quantiles provide an estimated representation of the distribution of the forecasted load curve while the prediction sets capture data in a specific range. Please note that quantiles can be directly used to produce a prediction set, but other methods introduced later in the paper can also be used to produce prediction sets with stronger theoretical guarantees (see Section 3).

In this paper, the focus is given to quantiles from 0.1 to 0.9 to have a holistic representation of the distribution while also looking at an 80% prediction set, which enables accommodation for a majority of scenarios. These two types of forecasts (quantiles and prediction sets) have to be provided at various horizons depending on industrial needs:

- Mid-term: typically spanning days to weeks ahead, they help grid operators and electricity market participants plan for infrastructure investment to meet future demand.
- Day-Ahead: critical for scheduling generation and market trading as well as managing grid constraints for EV charging, ensuring the balance between electricity demand and production, and minimising last-minute costs.
- Real-Time: offer immediate insights into electricity demand and supply balance, responses to sudden changes like spikes in EV charging demand.

As shown in Table 1, we propose different methods and classes of models depending on the horizon, taking into account the characteristics/capabilities of each model.

Table 1. Proposed approaches in this paper for each horizon and their applications.

Horizon	Model Class	Proposed Approach	Applications
Mid-Term	Direct (Section 3.2)	Generalised Additive Models (GAM, GAMLSS, QGAM) [3,4]	Infrastructure Sizing OTC Market Trading
Day-Ahead	Bottom-up (Section 3.3)	SARIMA & Gaussian Mixture Models [5]	SPOT Market Trading Smart Charging (and V2G)
Real-Time	Adaptive (Section 3.4)	Adaptive Conformal Inference (ACI) [6]	Demand/Supply balance, Smart Charging

The main contributions of this paper are: (a) an extension of the work presented in [5] to probabilistic forecasts and (b) the application of a post hoc uncertainty quantification method, namely adaptive conformal prediction, to EV charging data.

The rest of the paper is articulated as follows: Section 2 reviews the literature of work that preceded and inspired this paper. Section 3 details the direct, bottom-up, and adaptive approaches used in this paper for providing forecasted quantiles and prediction sets. Section 4 presents and discusses the results obtained on the Palo Alto charging session data [5].

2. Related Work

Forecasting charging station occupancy is a crucial problem for electricity market stakeholders when it comes to optimising their production units in line with charging demand [7]. With the development of EVs, research activities in that field have significantly increased, and many papers investigate data-driven methods, as presented in the review paper [8] and more recently in [9], where the authors highlight the superiority of performance of learning-based techniques but at the same time point out the difficulty of quantifying uncertainty in demand data. Forecasts can be used for different purposes, e.g., demand scheduling [10], trading [11], or grid management [12]. Forecasting models range from statistical models to the last generation of AI tools [9], where a major problem is the lack of data of sufficient quality for an efficient calibration of the parameters. For example, ref. [13] proposes a spatio-temporal distribution of demand with graph convolutional networks trained on a big dataset of 76,000 private EVs in Beijing. Ref. [14] uses a spatial-temporal attention transformer to model the charging demand of 321 electric buses across regions

and time in Shanghai. Ref. [15] proposes an attention-based spatio-temporal multi-graph convolutional Network fitted on two datasets of around 10 stations in Beijing and Shanghai.

The rapid development of data-driven predictive models, including complex and large-scale machine learning models, creates a need for benchmarking studies, including uncertainty quantification. This work follows the one in [5], focusing on benchmarking point estimates of day-ahead load and occupancy forecasting solutions covering eight open charging session datasets presented in the data and method review [8]. This benchmark examined two sets of methods: direct approaches that would predict the aggregate load curve for all sessions at once, and bottom-up approaches that would model individual charging sessions before aggregating them to obtain the overall load curve. The bottom-up approaches, although more complex to estimate, offer more flexibility for the deployment of smart charging solutions [16], where knowing individual arrivals and durations at the charging point is essential. We propose to extend these benchmarks to probabilistic forecasting algorithms by exploring probabilistic variations of these forecasting models. The use of probabilistic forecasts is becoming increasingly important for the efficient operation of electricity systems, as highlighted in the Global Energy Forecasting Competitions [17] and in [18].

The need for probabilistic forecasts is particularly important in the management of EVs, as the optimisation of charging loads often requires a good quantification of the uncertainty around the forecasts to manage the best- and worst-case scenarios for grid management [19] or, more generally, to maximise profit on energy markets [11]. In [20], quantile regression machine learning methods are applied to a set of charging stations in the Netherlands (the EVnetNL charging stations). The authors show the advantages of exploiting local data (load and meteorological) to forecast the global load in a hierarchical way through an ensemble methodology. Ref. [21] proposes an approach to quantify the uncertainty of parking duration forecasts in EV management. In [22], the authors compare confidence intervals derived from quantile random forest and neural networks to forecast individual charging stations in two commercial buildings. They study both the question of forecasting an aggregated load or individual charging points and conclude similarly to [5] that the aggregated EV load forecasting is an easier task.

3. Materials and Methods

Two fundamental approaches have been retained to address the probabilistic forecasting task (see Figure 1). A direct approach that provides quantile forecasts with a GAMLSS or a QGAM (see Section 3.2) model trained directly on the load curve. A bottom-up approach (see Section 3.3), which predicts individual charging session characteristics to then reconstruct the load curve. With both approaches, 9 quantile forecasts are provided from 0.1 to 0.9 with 0.1 increments as well as a mean forecast. An additional approach based on Conformal Prediction (CP) is used to build prediction sets over both quantile and mean forecasts from direct or bottom-up approaches (see Section 3.4).

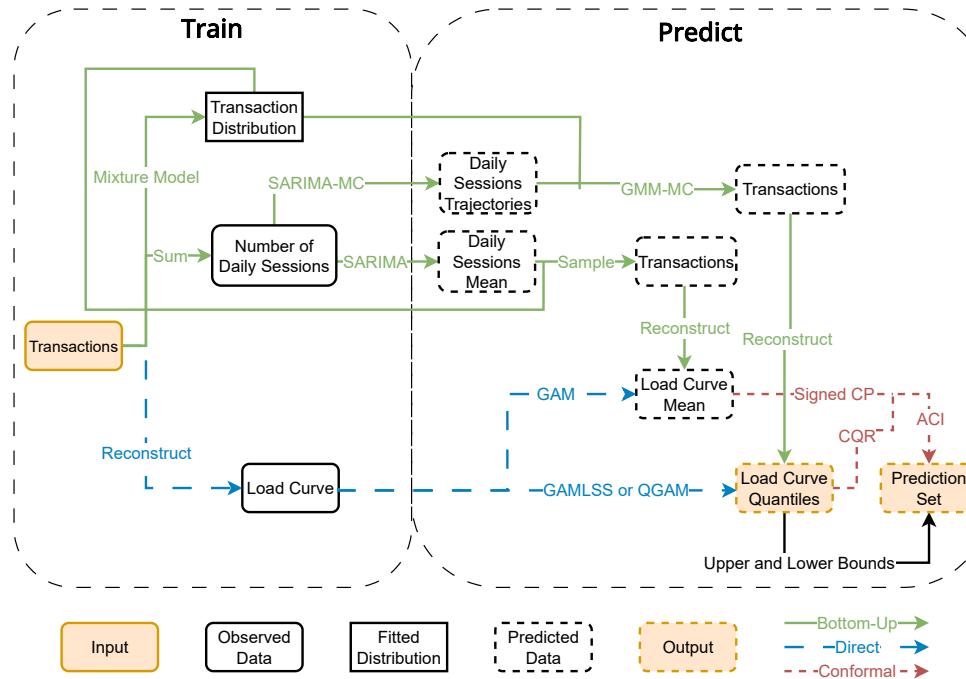


Figure 1. A summary of methods to produce quantile forecasts and/or prediction sets based on charging sessions.

3.1. Problem Statement

The load curve of an ensemble of charging stations at time $t \in \mathbb{R}^+$ can be written as a function $L(t)$. To reconstruct this load curve based on charging session data, three variables are required for each charging session i : the arrival time (a_i), the charge duration (d_i) and the energy demand (e_i). With these three variables, an elementary load curve can be reconstructed in the shape of a step function:

$$\ell_i(t) = \begin{cases} \frac{e_i}{d_i} & \text{if } t \in [a_i, a_i + d_i] \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The overall load curve is obtained by summing all elementary load curves:

$$L(t) = \sum_{i \in \mathcal{T}} \ell_i(t) \quad (2)$$

with \mathcal{T} the set of sessions observed over a given period of time. Thus, the assumption is that the electric vehicle supply equipment (EVSE) charges the EV at a constant power. It is an accurate approximation for EVSEs that are not subject to smart charging [5]. Please note that the overall load curve can also be reconstructed with a smart charging algorithm, which would yield a different formulation of Equation (1), out of the scope of this paper. In addition, to comply with the European market standard, the chosen time step for sampling the load curve is 15 min. Please note that this choice can easily be modified in the proposed implementation at this [GitHub](#) repository [23]. When necessary, switching from continuous to discrete time will be achieved with the notation L_t , $t \in \mathbb{N}$ instead of $L(t)$, $t \in \mathbb{R}^+$.

The final purpose of the following methods presented in Sections 3.2–3.4, is to provide an estimation of the predicted quantiles of L_t defined as $\hat{L}_t^{(\tau)}$, $\tau \in \{0.1, \dots, 0.9\}$ or a prediction set $\hat{C}_t^{(\alpha)}$ with $1-\alpha$ being the nominal coverage, set to 80% in the scope of this paper.

3.2. Direct Approach

Generalised Additive Models (GAMs) are statistical models that capture nonlinear relationships between a response variable and multiple predictors by representing them as the sum of smooth functions [3]. Let y be a continuous r.v. with conditional distribution $p(y|x)$, where x is a p-dimensional vector of covariates. Gaussian additive models assume that the conditional distribution $p(y|x)$ is Gaussian and such that : $y|x \sim \mathcal{N}(\mu(x), \sigma^2)$ with

$$\mu(x) = \sum_{j=1}^m f_j(x) \text{ and } \sigma^2 = \text{Var}(y|x) \quad (3)$$

where f_j are linear or nonlinear smooth functions.

Generalised Additive Models for location scale and shape (GAMLSS) [24] are an extension of GAMs that enable the fine modelling of multiple parameters of a single distribution. In this study, GAMLSS are used to model both the mean and variance of the load at charging points over time, assuming the conditional distribution of the load given the covariates is Gaussian. Gaussian location-scale models assume again that $y|x \sim \mathcal{N}(\mu(x), \sigma^2(x))$ where μ and σ are both modeled like previously as the sum of linear and nonlinear effects of the covariates.

Finally, let us also define the conditional quantiles of $p(y|x)$ by $\mu_\tau(x) = F^{-1}(\tau|x)$, where $\tau \in \{0,1\}$ is the quantile level and F^{-1} is the inverse conditional cumulative distribution function (c.d.f.) of y . QGAM models each conditional quantile as the sum of linear and non linear effects like previously. In all three cases : An effect f_j is expressed as a projection

$$f_j(x) = \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}(x), \quad (4)$$

where $(B_{j,k})_{k=1}^{m_j}$ is a spline basis of dimension m_j and $(\beta_{j,k})_{k=1}^{m_j}$ are the corresponding coefficients estimated by a ridge regression (GAM), a maximum likelihood (GAMLSS) or by minimizing the Extended log-F (ELF) loss for each quantile level (QGAM). The ELF loss is a smooth and convex generalisation of the pinball loss (see Section 4.2 for a definition of the pinball loss). It enables the model to capture more complex data that cannot be characterised by usual statistical distributions (e.g., Gaussian). The covariates used in all three models include exogenous weather variables such as the temperature, wind speed, and relative humidity, as well as calendar variables like the time of day, which are used as a non-linear effect. The weekday and the season are used as linear effects. Finally, as an additional baseline method, a persistence model is included in the study. This simple model uses the last 9 observations of the same weekday and time of day and calculates an empirical quantile for each level.

3.3. Bottom-Up Approach

Bottom-up approaches consist of predicting characteristics of individual charging sessions occurring over time. As explained in Section 3.1, three variables are required to reconstruct the load curve of an ensemble of charging stations in an uncontrolled charging environment: (a_i, d_i, e_i) the arrival time, charge duration, and energy demand of a charging session i . These three variables can be modelled using various statistical techniques. It was shown in [25] that mixture models and, in particular, Gaussian Mixture Models (GMM) [5] are an adequate choice of methods to represent individual charging sessions (see Equation (5)).

$$p(a_i, d_i, e_i) = \sum_{k=1}^K \pi_k \mathcal{N}(\mu_k, \Sigma_k) \quad (5)$$

with K the total number of components of the mixture, π_k , μ_k and Σ_k the weight, mean, and covariance of the k -th component. Assuming it is possible to predict the number of charging sessions X occurring each day with a time series model, X charging sessions from the mixture model distribution can be sampled to obtain a prediction for a particular day. In this work, SARIMA models are used as predictors for the number of daily charging sessions. A Monte-Carlo (MC) approach is used to obtain quantile forecasts $\hat{L}_t^{(\tau)}$ for quantile level $\tau \in [0, 1]$ as follows: first, $N_1 \in \mathbb{N}$ SARIMA trajectories are simulated for the given forecasting period, and for each of these N_1 SARIMA simulations, the associated number of sessions is generated N_2 times from the GMM. Therefore, $N = N_1 \times N_2$ sampled $\hat{L}_{n,t}$ of L_t are generated to then calculate empirical means and quantiles for every timestep in a given forecasting period. The index $n \in \{1, \dots, N\}$ is the simulation number.

3.4. Conformal Prediction

Conformal Prediction (CP, [26]) is a general framework for uncertainty quantification that produces a valid prediction set from a black-box forecast procedure in a finite sample under mild assumptions. Recently, CP has received increasing attention since [27] presented Split Conformal Prediction (SCP), a procedure based on data splitting. More precisely, SCP consists of splitting observations between $\mathcal{D}_{\text{train}}$, $\mathcal{D}_{\text{calib}}$ and $\mathcal{D}_{\text{test}}$, respectively, a training, a calibration, and a test set. We detail the data splitting chosen for this application in Section 4.1. The objective is then to provide a prediction set $\hat{C}_t^{(\alpha)}$ such that the probability of having L_t inside the prediction set $\hat{C}_t^{(\alpha)}$ is at least $1 - \alpha$ for $t \in \mathcal{D}_{\text{test}}$. To do so, the training set is used to learn a forecast function with a given (black-box) algorithm \mathcal{A} . The forecasts can either be estimates of the mean or quantiles, so \mathcal{A} can be; for example, a GAM, a QGAM (see Section 3.2), or a SARIMA + GMM (see Section 3.3). Then, the calibration set is used to compute non-conformity scores. The choice of the score function is very important since it determines the usefulness of the prediction sets [28]. If \mathcal{A} provides an estimate of the mean \hat{L}_t , we decide to use the residuals as non-conformity scores [29]:

$$\hat{s}_t = L_t - \hat{L}_t \quad (6)$$

If \mathcal{A} provides estimates of the quantiles $\alpha/2$ and $1 - \alpha/2$, $\hat{L}_t^{(\alpha/2)}$ and $\hat{L}_t^{(1-\alpha/2)}$, we use Conformalised Quantile Regression (CQR, [30]), which corresponds to the non-conformity scores:

$$\hat{s}_t = \max \{ \hat{L}_t^{(\alpha/2)} - L_t, L_t - \hat{L}_t^{(1-\alpha/2)} \} \quad (7)$$

If the non-conformity scores are given by (6), then for all $t \in \mathcal{D}_{\text{test}}$, the prediction set is:

$$\hat{C}_t^{(\alpha)} := \left[\hat{L}_t + \hat{s}_{\lfloor (\# \mathcal{D}_{\text{calib}} + 1) \alpha / 2 \rfloor}; \hat{L}_t + \hat{s}_{\lceil (\# \mathcal{D}_{\text{calib}} + 1) (1 - \alpha / 2) \rceil} \right] \quad (8)$$

with $\hat{s}_{(k)}$ being the k -th order statistics of the calibration non-conformity scores and $\lfloor x \rfloor$ being the integer part of x . If the non-conformity scores are given by (7), then for all $t \in \mathcal{D}_{\text{test}}$, the prediction set is:

$$\hat{C}_t^{(\alpha)} := \left[\hat{L}_t^{(\alpha/2)} - \hat{s}_{\lceil (\# \mathcal{D}_{\text{calib}} + 1) (1 - \alpha) \rceil}; \hat{L}_t^{(1-\alpha/2)} + \hat{s}_{\lceil (\# \mathcal{D}_{\text{calib}} + 1) (1 - \alpha) \rceil} \right] \quad (9)$$

Under the exchangeability assumption, these procedures provide valid (i.e., in terms of coverage) and efficient (i.e., in terms of sharpness) prediction sets [29,30]. However, this assumption does not hold in the context of this article since we are dealing with time series. Therefore, we use a extension of CP introduced by [6], namely Adaptive

Conformal Inference (ACI). In this variation, a surrogate adaptive α_t is used instead of using α in (8) and (9). This adaptive framework enables us to handle potential shifts in the distribution (i.e., if the behaviour of the electrical vehicles changes) and provides calibrated and efficient prediction sets for time series. For the signed non-conformity scores (6), ACI needs to be extended with an adaptive α_t for both the upper and the lower bounds of the prediction set (see details in Appendix A). This extension provides more flexibility, especially if the forecasts are biased and the main properties of ACI's prediction sets can be recovered using a symmetry argument. In [6], this surrogate quantile level α_t is defined sequentially for all $t \in \mathcal{D}_{\text{test}}$ by:

$$\alpha_{t+1} = \alpha_t + \gamma(\alpha - \text{err}_t) \quad \text{with } \text{err}_t := \begin{cases} 0 & \text{if } L_t \in \widehat{C}_t^{(\alpha)} \\ 1 & \text{if } L_t \notin \widehat{C}_t^{(\alpha)} \end{cases} \quad (10)$$

In this work, inspired by the latest extensions of ACI [31,32], we use an adaptive γ_t instead of a fixed γ in (10). As suggested in [32], we choose a decaying γ_t inversely proportional to the square root of the sum of the past $(\alpha - \text{err}_t)^2$. Moreover, we add a threshold so that $\gamma_t \leq 0.1$ in order to avoid instability at the beginning of the test set. This method provides theoretical guarantees (see Appendix A) both in terms of validity and in terms of efficiency for the prediction set while achieving good practical performances.

4. Results and Discussion

In the following sections, the results of the methods applied to a real-world EV charging session dataset are described and discussed. The dataset as well as the implementation of the models are available at the following [GitHub](#) repository in csv format and R code [23].

4.1. Presentation of the Dataset

Experiments were led on a dataset from the city of Palo Alto (California, USA). This dataset has been explored thoroughly in [8] and includes information on each charging session that takes place on some of the city's public chargers. Features such as the arrival time, charge duration, and energy demand for each session are available for a period of time that spans from 2011 to 2020. In this study, the focus is given to the last year of data available before the pandemic (2019).

The weekly average load curve (see Figure 2) highlights two prominent peaks - one around midday and another in the late afternoon - present on all weekdays. Notably, Fridays exhibit a more pronounced second peak, while Sundays display only a single peak. Saturdays demonstrate two peaks of similar heights, contrasting with weekdays where the midday peak significantly exceeds the afternoon peak. This dataset is openly accessible and forms part of a collection of EV charging session datasets available on this [GitHub](#) repository.

In addition to the dataset itself, exogenous variables were sourced from the `riem` R package [33], which aggregates meteorological features from weather stations worldwide. In this study, relative humidity, wind speed, and temperature from the Palo Alto weather station were used.

Given the highly non-stationary nature of the data (EV adoption and technological breakthroughs), the models are refitted every two weeks (see Figure 3). This frequent refitting allows the models to adapt to changes in the data distribution, ensuring that the predictions remain accurate and relevant. The calibration set is specifically used for the conformal predictions (see Section 3.4).

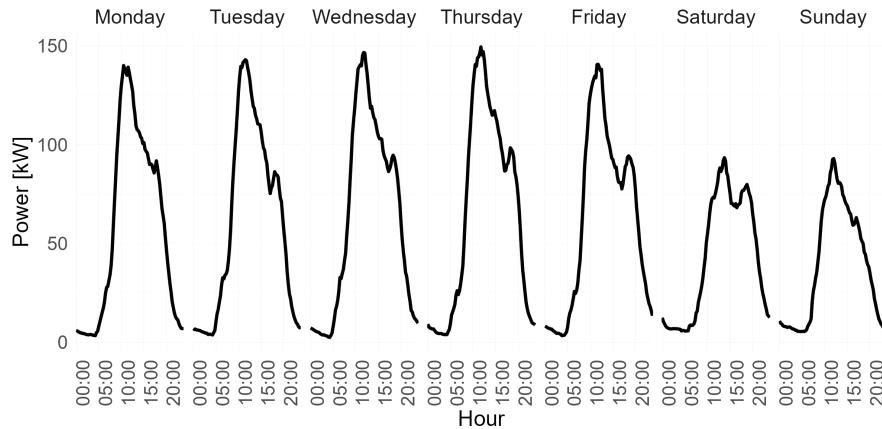


Figure 2. Average weekly load curve over the test period.

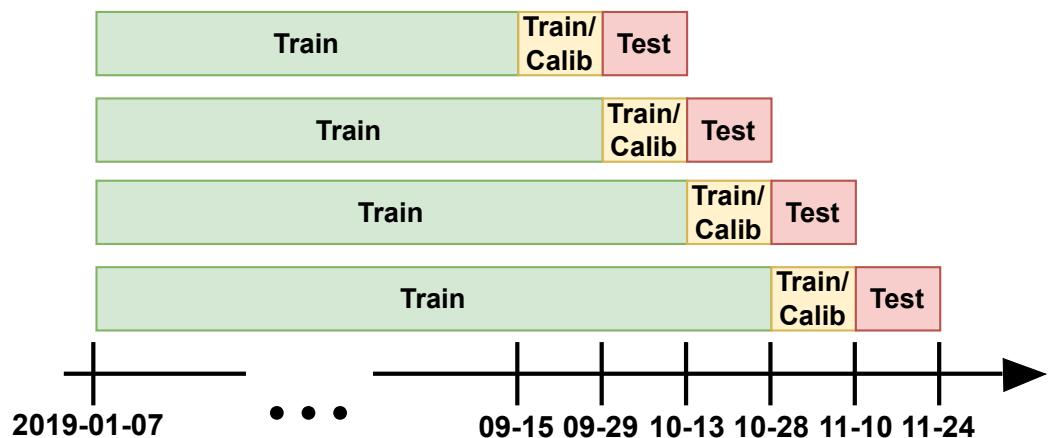


Figure 3. Data splitting for the four train, calibration, and test periods.

4.2. Quantile Forecasts

Forecasting quantiles 0.1 to 0.9 gives a complete representation of the distribution, which can be used in practice to better understand the uncertainty around the load curve forecast. Please note that in the case of the bottom-up approach, to calculate the empirical quantile and mean forecasts, 100 SARIMA trajectories with 100 GMM simulations for each of these trajectories were generated. In other words, a total of 10,000 simulated load curves were used to then calculate the quantile forecasts. In addition, one GMM is fitted per day of the week to better capture the variation in load profiles, as shown in Figure 2.

Three types of metrics have been used to evaluate quantile forecast performances. First, we consider a classical loss used in quantile regression, namely the pinball loss defined as follows:

$$Pin_{\tau} = \frac{1}{\#\mathcal{D}_{\text{test}}} \sum_{t \in \mathcal{D}_{\text{test}}} \rho_{\tau}(L_t - \hat{L}_t^{(\tau)}) \quad (11)$$

$$\rho_{\tau}(u) = \begin{cases} -(1 - \tau)u, & \text{if } u \leq 0 \\ \tau u & \text{if } u \geq 0 \end{cases} \quad (12)$$

It takes into account the desired imbalance between the true target values L_t and the predicted quantiles $\hat{L}_t^{(\tau)}$ with a degree of penalty depending on the chosen quantile level τ . Another classical metric that can be used for assessing the accuracy of quantile forecasts is the empirical coverage defined as:

$$Cov_{\tau} = \frac{1}{\#\mathcal{D}_{\text{test}}} \sum_{t \in \mathcal{D}_{\text{test}}} \mathbb{1}(L_t \leq \hat{L}_t^{(\tau)}) \quad (13)$$

Equation (13) is the empirical coverage of the predicted quantile, i.e., the proportion of observed values falling below the predicted quantile. Essentially, if $L_t^{\text{coverage}} = \tau$ the quantile forecast is said to be valid on the testing sample.

While Equations (11) and (13) evaluate the forecasts at a given quantile level, it is possible to evaluate them over all forecasted quantiles simultaneously. To do so, we define the continuous ranked probability score (CRPS) and its discrete approximation, namely the RPS:

$$\text{CRPS}(\hat{F}, L_t) = \int_{-\infty}^{+\infty} (\hat{F}(z) - \mathbb{1}_{L_t \leq z})^2 dz \quad (14)$$

where F is the predicted cumulative distribution function with $\tau_0 = 0$ and $\tau_{l+1} = 1$.

$$\text{RPS}(\{\hat{L}_t^{(\tau_1)}, \dots, \hat{L}_t^{(\tau_l)}\}, L_t) = \sum_{i \in [1,l]} \rho_i(L_t, \hat{L}_t^{(\tau_i)}) (\tau_{i+1} - \tau_{i-1}) \quad (15)$$

Figure 4 shows the average hourly quantile forecasts for a direct approach using QGAM and the bottom-up approach during each day of the week. It seems that both approaches capture the general shape of the observed curve in black with a peak demand in the middle of the day and another one in the late afternoon/evening for weekdays and Saturdays, and a single peak at midday on Sundays. However, the direct approach seems to better fit the finer details of the shape around the peaks, especially the second one. This is due to the fact that fitting the daily shape of the load, by time of day, is specifically expected of the QGAM or GAMLSS in the formulation of the splines optimisation problem, giving them an advantage. Tables 2 and 3 confirm these results, as the QGAM has better overall coverage, better pinball losses on all quantiles, and the best RPS.

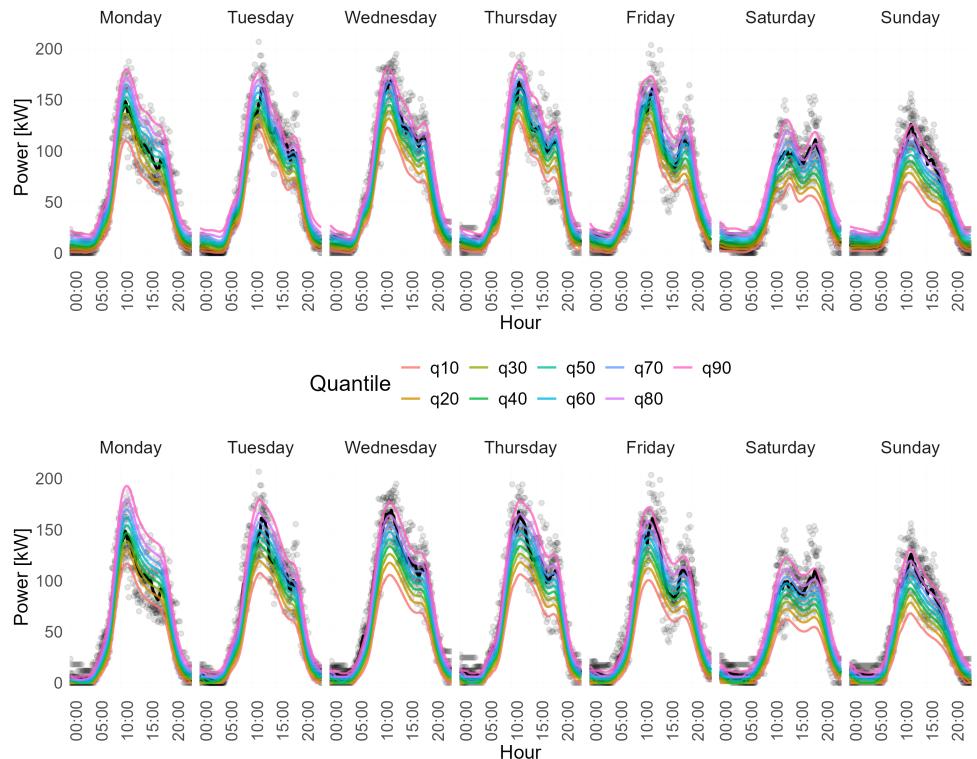


Figure 4. At the top are the QGAM averaged quantile forecasts for each quantile levels and at the bottom the same chart for the bottom-up approach. In black is the average daily load curve and the scatter plot of the daily observations over the test period. The plots are separated by day of the week.

This can be explained by the fact that QGAM minimizes the pinball loss without making any assumptions on the data distribution. On the other hand, GAMLSS assumes

that the data are Gaussian, which seems to deteriorate its performance, especially in higher quantiles. This could indicate that the real distribution of the data is slightly skewed to the right, in the higher values, compared to a Gaussian.

Finally, when analysing the performances of the bottom-up approach, it can be concluded that it is performing worse than direct approaches on most quantiles. However, this conclusion can be nuanced considering that:

- The bottom-up approach does not directly model the global load but the distribution of individual sessions. Which is, by nature, a more complex task.
- Unlike direct approaches, it does not use exogenous variables such as weather data.
- The pinball losses and empirical coverages for the 0.1 and 0.9 quantile forecasts are relatively close to their direct counterparts.

Considering all these shortcomings, the bottom-up approach giving results close to the direct approaches is promising, especially considering its key importance in smart charging applications.

Table 2. Pinball Losses (Pin_{τ}) and the empirical coverage (Cov_{τ}) for each quantile level. The best score is in bold.

Quantile	Pinball Loss				Empirical Coverage			
	Persistence	GAMLSS	QGAM	SARIMA + GMM	Persistence	GAMLSS	QGAM	SARIMA + GMM
0.1	2.64	2.64	2.56	2.77	0.16	0.11	0.11	0.10
0.2	4.25	4.31	4.24	4.59	0.23	0.20	0.20	0.15
0.3	5.38	5.44	5.40	5.88	0.29	0.28	0.30	0.20
0.4	6.08	6.11	6.12	6.67	0.36	0.37	0.41	0.27
0.5	6.48	6.41	6.44	6.96	0.43	0.46	0.51	0.35
0.6	6.38	6.33	6.32	6.82	0.51	0.54	0.59	0.43
0.7	5.89	5.82	5.74	6.22	0.60	0.62	0.69	0.54
0.8	4.99	4.81	4.67	5.08	0.69	0.71	0.79	0.66
0.9	3.51	3.13	2.96	3.19	0.78	0.81	0.89	0.80

Table 3. Average RPS metric for each model. The best score is in bold.

Models	Persistence	GAMLSS	QGAM	SARIMA + GMM
RPS	9.12	9.00	8.89	9.64

4.3. Prediction Sets

As introduced in Section 3, an important tool for uncertainty quantification is the use of prediction sets. In this Section, we consider three different approaches to produce a prediction set with nominal coverage $1 - \alpha$. The first approach is to use the quantile forecast of level $\alpha/2$ for the lower bound and of level $1 - \alpha/2$ for the upper bound of the prediction set with any given procedure that outputs quantile forecasts (for example, QGAM, GAMLSS, or SARIMA + GMM). The two other approaches are based on conformal prediction, respectively, based on mean forecasts and on quantile forecasts as described in Section 3.4. In this study, we focus on prediction sets with a nominal coverage of 80%, which means that the quantile forecasts we will use are $\hat{L}^{(10\%)}$ and $\hat{L}^{(90\%)}$. This can be justified in practice as 80% usually represents a sweet spot between over-committing resources to rare events while still accounting for a reasonable level of uncertainty [34].

Figure 5 shows an example of prediction sets forecasted during the test period. Visually, it is hard to conclude which of Quantiles SARIMA + GMM (on the left) and CQR SARIMA + GMM (on the right) is more accurate, which makes it necessary to use specific metrics.

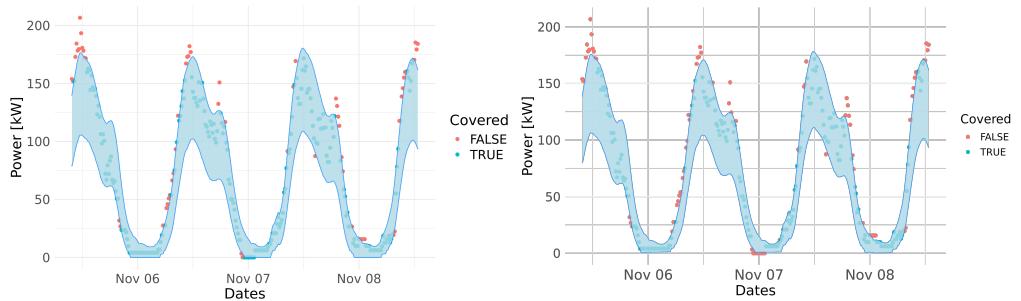


Figure 5. Coverage/Mean Width between conformal (right) and non-conformal (left) methods.

To do so, we extend the coverage metric introduced in Section 4.2 to prediction sets. Indeed, the first criteria to evaluate for a prediction set is validity, i.e., having an empirical coverage superior or equal to the target coverage $1 - \alpha$. However, this is not enough to ensure the quality of a prediction set. Indeed, one can achieve valid coverage with an infinite bound, which is completely uninformative. Therefore, to have a better understanding of the quality of a prediction set, we consider simultaneously a notion of coverage and one of efficiency by measuring its average length. More formally, denoting $\widehat{C}_t^{(\alpha)-}$ the lower bound and $\widehat{C}_t^{(\alpha)+}$ the upper bound of the prediction set $\widehat{C}_t^{(\alpha)}$, we have:

$$Len_\alpha = \frac{1}{\#\mathcal{D}_{\text{test}}} \sum_{t \in \mathcal{D}_{\text{test}}} (\widehat{C}_t^{(\alpha)+} - \widehat{C}_t^{(\alpha)-}) \quad (16)$$

The prediction sets produced directly from quantile forecasts are not valid, with a lower coverage than expected (see Figure 6). In Figures 6 and 7, the empirical coverage is computed sequentially, which explains the high degree of volatility at the beginning of the test set. The lack of coverage of the quantiles-based prediction sets in Figure 6 can be explained by a shift in the distribution that the quantile forecasts we used are unable to express.

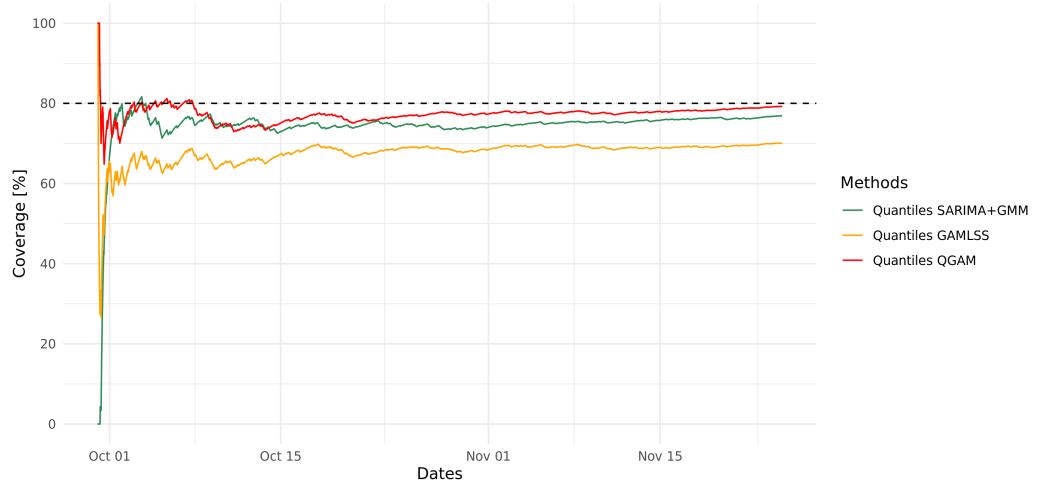


Figure 6. Empirical coverage for quantiles-based prediction sets.

Meanwhile, the conformal prediction based prediction sets are all valid, as shown in Figure 7.

We compare all the prediction sets by looking simultaneously at their coverages Cov_α and averaged lengths Len_α . Figure 8 emphasizes the trade-off between validity and efficiency since the prediction sets based on conformal prediction are valid but are less efficient than the Quantiles based prediction sets. However, Figure 8 shows that the best prediction sets appear to be Quantiles SARIMA + GMM, Quantiles QGAM and CQR

SARIMA + GMM because even if the Quantiles SARIMA + GMM and QGAM are not valid, their lack of coverage is small compared to the gain in efficiency. Among the conformal prediction based methods, the best one is based on the CQR non-conformity score which might highlight heteroscedasticity.

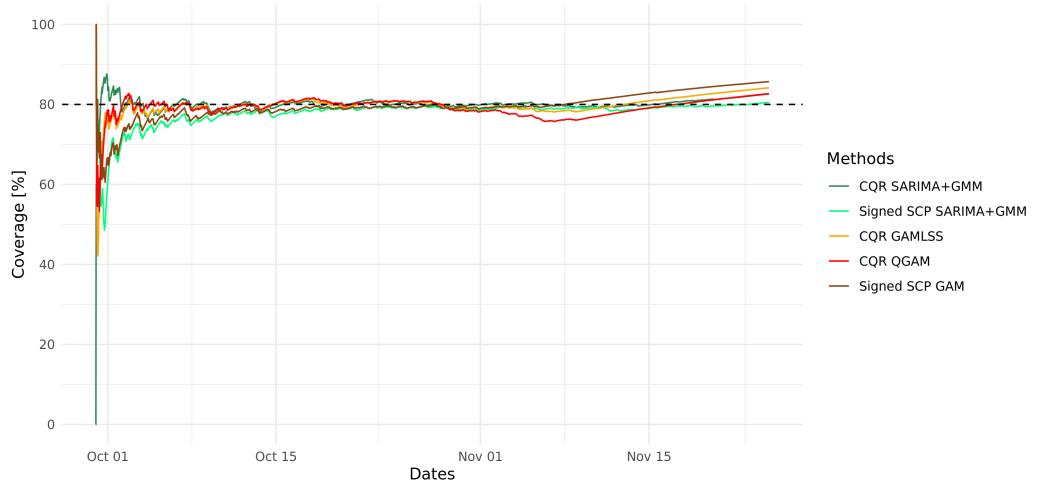


Figure 7. Empirical coverage for conformalised prediction sets.

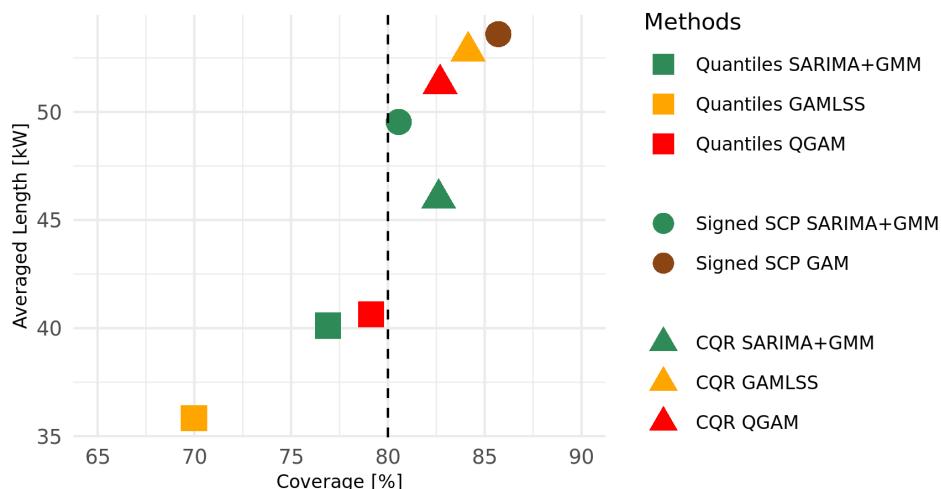


Figure 8. Coverage/Mean Width between conformal and non conformal methods.

To provide a better understanding of the improvement we obtained by using conformal prediction, we focus on two quantile forecast procedures, namely QGAM and SARIMA + GMM, along with their CQR conformalised version. In Figures 9 and 10, we see how conformal prediction improves the coverage when the quantile-based prediction sets are not large enough. Figures 9 and 10 also highlight that the prediction sets are not only valid marginally but also locally, i.e., conditionally valid with respect to a feature of interest, here the time of day. For example, a prediction set with a 100% coverage rate during the night and 70% throughout the day can be valid with respect to the 80% empirical coverage but will not be locally valid. This is very important since it is crucial to capture demand peaks for all applications detailed in Table 1.

Overall, the quantile methods produce prediction sets that capture the characteristics of the distribution but need to be recalibrated to be valid. We solved this issue online using conformal prediction, with the resulting prediction sets suffering a more or less significant loss in efficiency depending on the quantile forecasting method considered.

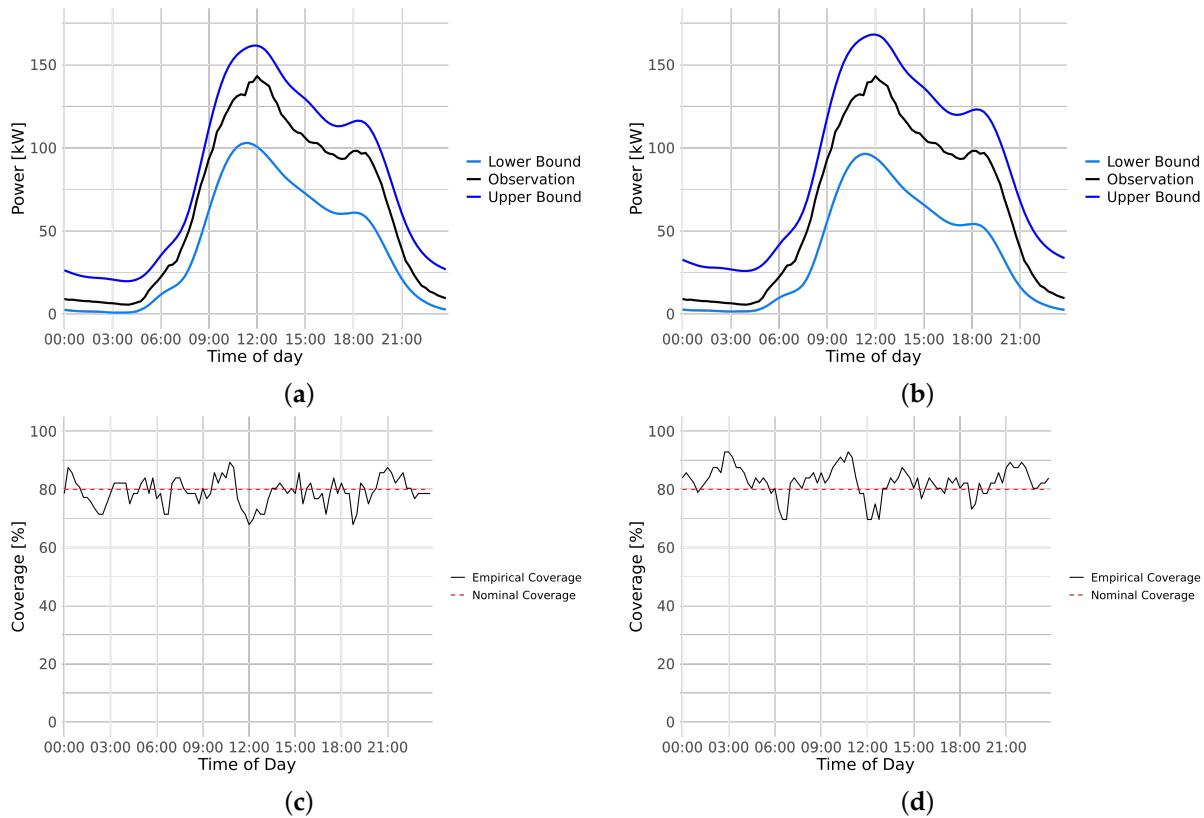


Figure 9. Daily profiles of QGAM (a) and CQR QGAM (b) and conditional coverages of QGAM (c), CQR QGAM (d).

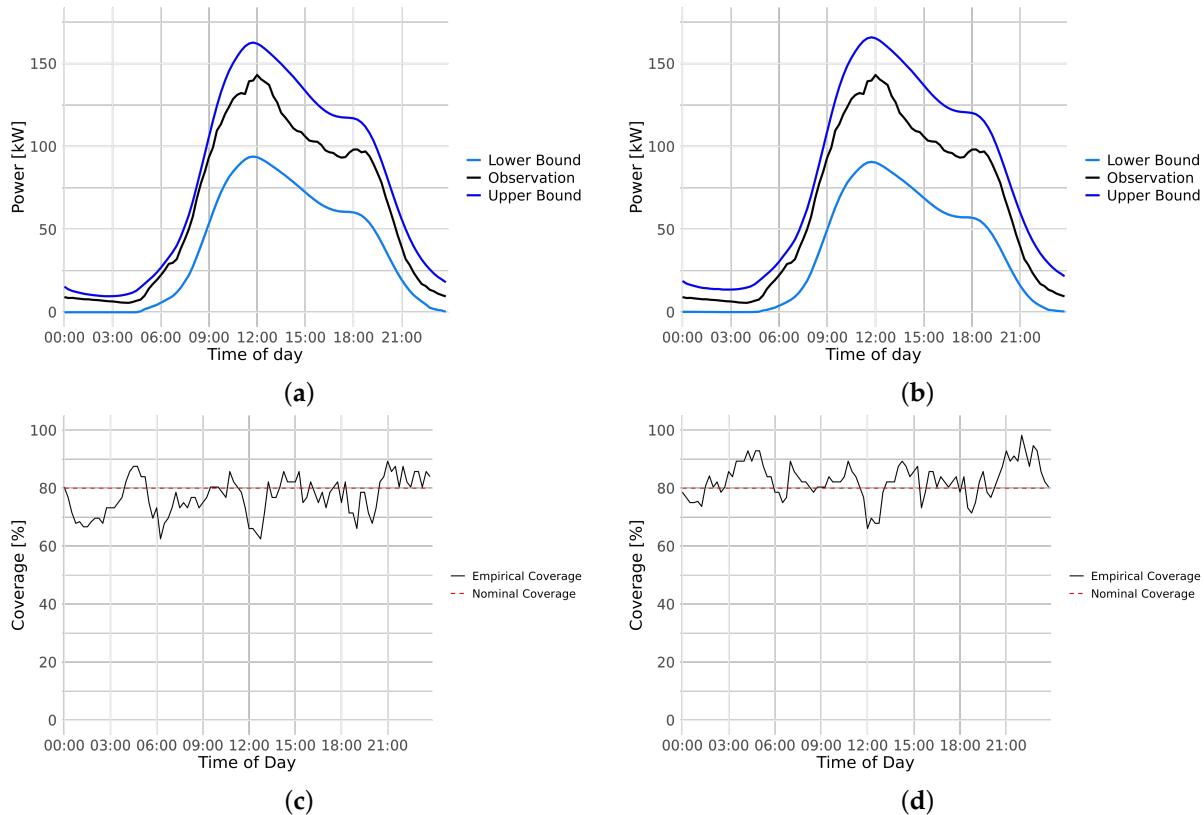


Figure 10. Daily profiles of SARIMA + GMM (a) and CQR SARIMA + GMM (b) and conditional coverages of SARIMA + GMM (c), CQR SARIMA + GMM (d).

5. Conclusions

The electrification of transport through the widespread adoption of electric vehicles (EVs) presents both challenges and opportunities in the management of electrical networks. This paper addresses the critical need for efficient forecasting algorithms to understand and anticipate EV charging behaviors, which are essential for optimising grid operations and ensuring stability. By quantifying the uncertainty in load forecasts, stakeholders can better anticipate surges in EV demand and plan accordingly. This study has evaluated probabilistic forecasting approaches at different horizons, including mid-term, day-ahead, and real-time forecasts, using various model classes tailored to each horizon's requirements. The research focused on providing forecasted quantiles and prediction sets, which offer insights into the distribution of forecasted load curves and provide a range of possible scenarios for grid management.

The experiments conducted on real-world charging session data from Palo Alto demonstrated the effectiveness of the proposed approaches. Results showed that both direct and bottom-up approaches captured the general shape of load curves, with the direct approach exhibiting better performance in capturing finer details around peak demand periods. Additionally, comparison metrics such as pinball loss and empirical coverage revealed the strengths and weaknesses of each model, with QGAMs proving to be the best quantile forecaster in terms of pinball loss. In addition, the study explored the generation of prediction sets using both direct quantile forecasts and Conformal Prediction techniques. The analysis highlighted the importance of validity and efficiency in prediction sets, with Conformal Prediction methods showing promise in correcting invalid forecasts and improving coverage. However, further research is needed to optimise the performance of prediction sets, particularly in balancing coverage and length of intervals.

Overall, the study contributes to the advancement of probabilistic forecasting algorithms for EV load management, providing valuable insights for grid operators, market participants, and policymakers in effectively integrating EVs into the electrical grid while ensuring reliability and stability. In particular, this work can help to mitigate the impact of EV charging on the grid by linking it to *smart charging* algorithms, especially robust optimisation and reinforcement learning models that can incorporate predictive inputs and thus be more realistic. Future research directions may include refining forecasting models, incorporating additional exogenous variables, and investigating advanced techniques for prediction set construction to further enhance grid management capabilities.

Author Contributions: The methodology, data preparation and codes is a joint work of Y.A.-O., B.H. and G.P. Writing, review and editing, is a joint work of Y.A.-O., B.H., G.P. and Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: All the authors were employed by the company EDF. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EV	Electric Vehicle
EVSE	Electric Vehicle Supply Equipment
GAM	Generalised Additive Model
GAMLSS	Generalised Additive Model for Location Scale and Shape

QGAM	Quantile Generalised Additive Model
SARIMA	Seasonal autoregressive integrated moving average
GMM	Gaussian Mixture Model
ACI	Adaptive Conformal Inference
V2G	Vehicle-to-Grid
CP	Conformal Prediction
ELF	Extended log-F
SCP	Split Conformal Prediction
MC	Monte-Carlo
CQR	Conformalised Quantile Regression
r.v.	Random Variable
c.d.f.	Cumulative Distribution Function

Appendix A. Details on Adaptive Conformal Inference

Appendix A.1. Extension of ACI for Signed Non-Conformity Scores

In Section 3.4, we introduced the signed non-conformity scores (6), which enables separate treatment of the lower and upper bounds of the prediction set. Consequently, this choice of non-conformity scores requires two adaptive quantile levels $\alpha_t^{(u)}$ and $\alpha_t^{(\ell)}$, defined recursively as:

$$\alpha_{t+1}^{(u)} = \alpha_t^{(u)} + \gamma_t \left(\frac{\alpha}{2} - \text{err}_t^{(u)} \right) \quad \text{with } \text{err}_t^{(u)} := \begin{cases} 0 & \text{if } L_t \leq \hat{L}_t + \hat{s}_{\lceil (\#D_{\text{calib}}+1)(1-\alpha_t^{(u)}) \rceil} \\ 1 & \text{if } L_t > \hat{L}_t + \hat{s}_{\lceil (\#D_{\text{calib}}+1)(1-\alpha_t^{(u)}) \rceil} \end{cases}$$

and

$$\alpha_{t+1}^{(\ell)} = \alpha_t^{(\ell)} + \gamma_t \left(\frac{\alpha}{2} - \text{err}_t^{(\ell)} \right) \quad \text{with } \text{err}_t^{(\ell)} := \begin{cases} 0 & \text{if } L_t \geq \hat{L}_t + \hat{s}_{\lfloor (\#D_{\text{calib}}+1)(\alpha_t^{(\ell)}) \rfloor} \\ 1 & \text{if } L_t < \hat{L}_t + \hat{s}_{\lfloor (\#D_{\text{calib}}+1)(\alpha_t^{(\ell)}) \rfloor} \end{cases}$$

This flexibility can be very useful if the estimate of the mean turns out to be biased because the resulting prediction set does not have to be centered on \hat{L}_t . Moreover, by construction of the recursive definition of $\alpha_t^{(u)}$ and $\alpha_t^{(\ell)}$, the theoretical guarantees in terms of coverage of the prediction set are the same than with the classical ACI procedures with adaptive γ_t .

Appendix A.2. Proof of Asymptotical Coverage

As stated in Section 3.4, the conformal prediction procedure used in this article based on the ACI version of CQR provides valid prediction sets in terms of asymptotic coverage, i.e., $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{\{L_t \notin \hat{C}_t^{(\alpha)}\}} \stackrel{a.s.}{=} \alpha$. This result can be obtained with one of the best theoretical convergence rate among the methods using an adaptive γ_t [32], namely $O\left(\frac{1}{\sqrt{T}}\right)$.

Proof. We recall, the recursive update scheme of α_t we use:

$$\alpha_{t+1} = \alpha_t + \gamma_t (\alpha - \text{err}_t) \quad \text{with } \text{err}_t := \begin{cases} 0 & \text{if } L_t \in \hat{C}_t^{(\alpha)} \\ 1 & \text{if } L_t \notin \hat{C}_t^{(\alpha)} \end{cases}$$

and $\gamma_t := \frac{\eta}{\sqrt{\sum_{s=1}^t (\alpha_s - \text{err}_s)^2}}$

With this recursive definition, we obtained (A3) using Abel transformation:

$$(\alpha - \text{err}_t) = \frac{\alpha_{t+1} - \alpha_t}{\gamma_t} \quad (\text{A1})$$

$$\sum_{t=1}^T (\alpha - \text{err}_t) = \sum_{t=1}^T \frac{\alpha_{t+1} - \alpha_t}{\gamma_t} \quad (\text{A2})$$

$$\sum_{t=1}^T (\alpha - \text{err}_t) = \frac{\alpha_{T+1}}{\gamma_T} - \frac{\alpha_1}{\gamma_0} + \sum_{t=1}^T \alpha_t \frac{\gamma_t - \gamma_{t-1}}{\gamma_t \gamma_{t-1}} \quad (\text{A3})$$

If we use the convention that if $\alpha_t < 0$ then $\widehat{s}_{(\lceil (\#D_{\text{calib}}+1)(1-\alpha_t) \rceil)} = +\infty$ and if $\alpha_t > 1$ then $\widehat{s}_{(\lceil (\#D_{\text{calib}}+1)(1-\alpha_t) \rceil)} = -\infty$, we have that $\alpha_t = O(1)$, hence $\frac{\alpha_{T+1}}{\gamma_T} - \frac{\alpha_1}{\gamma_0} = O(\sqrt{T})$. Moreover, we can also bound the other term:

$$\left| \sum_{t=1}^T \alpha_t \frac{\gamma_t - \gamma_{t-1}}{\gamma_t \gamma_{t-1}} \right| \leq \left| \sum_{t=1}^T \alpha_t \left(\frac{1}{\gamma_{t-1}} - \frac{1}{\gamma_t} \right) \right| \quad (\text{A4})$$

$$\leq \max_t(|\alpha_t|) \cdot \left| \sum_{t=1}^T \left(\frac{1}{\gamma_{t-1}} - \frac{1}{\gamma_t} \right) \right| \quad (\text{A5})$$

$$\leq \max_t(|\alpha_t|) \cdot \left| \frac{1}{\gamma_0} - \frac{1}{\gamma_T} \right| = O(\sqrt{T}) \quad (\text{A6})$$

Which concludes that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{\{L_t \notin \widehat{C}_t^{(\alpha)}\}} \stackrel{a.s.}{=} \alpha$. \square

References

- IEA. Electric Vehicles, 2022. Available online: <https://www.iea.org/reports/electric-vehicles> (accessed on 2 March 2023).
- Hu, J.; Morais, H.; Sousa, T.; Lind, M. Electric vehicle fleet management in smart grids: A review of services, optimization and control aspects. *Renew. Sustain. Energy Rev.* **2016**, *56*, 1207–1226. [CrossRef]
- Hastie, T.; Tibshirani, R. Generalized additive models: Some applications. *J. Am. Stat. Assoc.* **1987**, *82*, 371–386. [CrossRef]
- Fasiolo, M.; Wood, S.N.; Zaffran, M.; Nedellec, R.; Goude, Y. Fast calibrated additive quantile regression. *J. Am. Stat. Assoc.* **2021**, *116*, 1402–1412. [CrossRef]
- Amara-Ouali, Y.; Goude, Y.; Hamrouche, B.; Bishara, M. A benchmark of electric vehicle load and occupancy models for day-ahead forecasting on open charging session data. In Proceedings of the Thirteenth ACM International Conference on Future Energy Systems, Virtual, 28 June–1 July 2022; pp. 193–207.
- Gibbs, I.; Candes, E. Adaptive conformal inference under distribution shift. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 1660–1672.
- Zhang, J.; Wang, Z.; Miller, E.J.; Cui, D.; Liu, P.; Zhang, Z. Charging demand prediction in Beijing based on real-world electric vehicle data. *J. Energy Storage* **2023**, *57*, 106294. [CrossRef]
- Amara-Ouali, Y.; Goude, Y.; Massart, P.; Poggi, J.M.; Yan, H. A Review of Electric Vehicle Load Open Data and Models. *Energies* **2021**, *14*, 2233. [CrossRef]
- Rashid, M.; Elfouly, T.; Chen, N. A comprehensive survey of electric vehicle charging demand forecasting techniques. *IEEE Open J. Veh. Technol.* **2024**, *5*, 1348–1373. [CrossRef]
- Hafeez, A.; Alammari, R.; Iqbal, A. Utilization of EV Charging Station in Demand Side Management Using Deep Learning Method. *IEEE Access* **2023**, *11*, 8747–8760. [CrossRef]
- Bampos, Z.N.; Laitos, V.M.; Afentoulis, K.D.; Vagropoulos, S.I.; Biskas, P.N. Electric vehicles load forecasting for day-ahead market participation using machine and deep learning methods. *Appl. Energy* **2024**, *360*, 122801. [CrossRef]
- Tappeta, V.S.R.; Appasani, B.; Patnaik, S.; Ustun, T.S. A Review on Emerging Communication and Computational Technologies for Increased Use of Plug-In Electric Vehicles. *Energies* **2022**, *15*, 6580. [CrossRef]
- Wang, S.; Chen, A.; Wang, P.; Zhuge, C. Predicting electric vehicle charging demand using a heterogeneous spatio-temporal graph convolutional network. *Transp. Res. Part C Emerg. Technol.* **2023**, *153*, 104205. [CrossRef]
- Zhang, Y.; Liu, C.; Rao, X.; Zhang, X.; Zhou, Y. Spatial-temporal load forecasting of electric vehicle charging stations based on graph neural network. *J. Intell. Fuzzy Syst.* **2024**, *46*, 821–836. [CrossRef]
- Shi, J.; Zhang, W.; Bao, Y.; Gao, D.W.; Wang, Z. Load Forecasting of Electric Vehicle Charging Stations: Attention Based Spatiotemporal Multi-Graph Convolutional Networks. *IEEE Trans. Smart Grid* **2023**, *15*, 3016–3027. [CrossRef]

16. Bogdanov, D.; Breyer, C. Role of smart charging of electric vehicles and vehicle-to-grid in integrated renewables-based energy systems on country level. *Energy* **2024**, *301*, 131635. [[CrossRef](#)]
17. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [[CrossRef](#)]
18. Hong, T.; Pinson, P.; Wang, Y.; Weron, R.; Yang, D.; Zareipour, H. Energy Forecasting: A Review and Outlook. *IEEE Open Access Power Energy* **2020**, *7*, 376–388. [[CrossRef](#)]
19. Nutkani, I.; Toole, H.; Fernando, N.; Andrew, L.P.C. Impact of EV charging on electrical distribution network and mitigating solutions—A review. *Iet Smart Grid* **2024**, *7*, 485–502. [[CrossRef](#)]
20. Buzna, L.; De Falco, P.; Ferruzzi, G.; Khormali, S.; Proto, D.; Refa, N.; Straka, M.; van der Poel, G. An ensemble methodology for hierarchical probabilistic electric vehicle load forecasting at regular charging stations. *Appl. Energy* **2021**, *283*, 116337. [[CrossRef](#)]
21. Phipps, K.; Schwenk, K.; Briegel, B.; Mikut, R.; Hagenmeyer, V. Customized uncertainty quantification of parking duration predictions for ev smart charging. *IEEE Internet Things J.* **2023**, *10*, 20649–20661. [[CrossRef](#)]
22. Khan, W.; Somers, W.; Walker, S.; de Bont, K.; Van der Velden, J.; Zeiler, W. Comparison of electric vehicle load forecasting across different spatial levels with incorporated uncertainty estimation. *Energy* **2023**, *283*, 129213. [[CrossRef](#)]
23. Yvenn, A.; Bachir H.G.P. ORFEVRE. 2024. Available online: <https://github.com/yvenn-amara/ORFEVRE> (accessed on 16 January 2025).
24. Stasinopoulos, D.M.; Rigby, R.A. Generalized additive models for location scale and shape (GAMLSS) in R. *J. Stat. Softw.* **2008**, *23*, 1–46.
25. Flammini, M.G.; Prettico, G.; Julea, A.; Fulli, G.; Mazza, A.; Chicco, G. Statistical characterisation of the real transaction data gathered from electric vehicle charging stations. *Electr. Power Syst. Res.* **2019**, *166*, 136–150. doi: 10.1016/j.epsr.2018.09.022. [[CrossRef](#)]
26. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: Berlin/Heidelberg, Germany, 2005.
27. Lei, J.; G'Sell, M.; Rinaldo, A.; Tibshirani, R.J.; Wasserman, L. Distribution-free predictive inference for regression. *J. Am. Stat. Assoc.* **2018**, *113*, 1094–1111. [[CrossRef](#)]
28. Angelopoulos, A.N.; Bates, S. Conformal prediction: A gentle introduction. *Found. Trends® Mach. Learn.* **2023**, *16*, 494–591. [[CrossRef](#)]
29. Linusson, H.; Johansson, U.; Löfström, T. Signed-error conformal regression. In Proceedings of the Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, 13–16 May 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 224–236.
30. Romano, Y.; Patterson, E.; Candes, E. Conformalized quantile regression. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
31. Bhatnagar, A.; Wang, H.; Xiong, C.; Bai, Y. Improved online conformal prediction via strongly adaptive online learning. In *International Conference on Machine Learning*; PMLR: Honolulu, HI, USA, 2023; pp. 2337–2363.
32. Angelopoulos, A.N.; Barber, R.; Bates, S. Online conformal prediction with decaying step sizes. In *International Conference on Machine Learning*; PMLR: Vienna, Austria, 2024; pp. 1616–1630.
33. Salmon, M.; Anderson, B. riem: Accesses Weather Data from the Iowa Environment Mesonet. 2022. Available online: <https://CRAN.R-project.org/package=riem> (accessed on 21 December 2024).
34. Van der Meer, D.W.; Munkhammar, J.; Widén, J. Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. *Sol. Energy* **2018**, *171*, 397–413. [[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.