

FPSeq2Q: Fully Parameterized Sequence to Quantile Regression for Net-Load Forecasting with Uncertainty Estimates

Anthony Faustine and Lucas Pereira

Abstract—The increased penetration of Renewable Energy Sources (RES) as part of a decentralized and distributed power system makes net-load forecasting a critical component in the planning and operation of power systems. However, compared to the transmission level, producing accurate short-term net-load forecasts at the distribution level is complex due to the small number of consumers. Moreover, owing to the stochastic nature of RES, it is necessary to quantify the uncertainty of the forecasted net-load at any given time, which is critical for the real-world decision process. This work presents parameterized deep quantile regression for short-term probabilistic net-load forecasting at the distribution level. To be precise, we use a Deep Neural Network (DNN) to learn both the quantile fractions and quantile values of the quantile function. Furthermore, we propose a scoring metric that reflects the trade-off between predictive uncertainty performance and forecast accuracy. We evaluate the proposed techniques on historical real-world data from a low-voltage distribution substation and further assess its robustness when applied in real-time. The experiment's outcomes show that the resulting forecasts from our approach are well-calibrated and provide a desirable trade-off between forecasting accuracies and predictive uncertainty performance that are very robust even when applied in real-time.

Index Terms—Net-Load, Forecasting, Uncertainty, Deep Neural Network, Quantile Regression

I. INTRODUCTION

THE Paris Agreement has set forth an objective to reduce energy-related GHG emissions by more than 70% by 2050 through massive deployment of RES [1]. Consequently, distribution networks worldwide are experiencing massive adoption of RES and Electrical Energy Storage (EES) systems [2] leading to decentralized and distributed power systems.

The decentralized and distributed power systems with a high share of RES is likely to deliver myriad benefits such as reducing power outages and network losses, creating a competitive energy market, and increasing the overall efficiency of power systems [3]. However, managing the net-load, which is the difference in the amount of load and renewable output, is more challenging due to the intermittent nature of the RES generation and distributed [4]. Therefore the ability to provide adequate short-term forecasts from half an hour to a few days ahead of net-load is paramount for the planning and operation of distribution systems [5].

A. Faustine is with ITI, LARSyS, Técnico Lisboa, Portugal and Irish Manufacturing Research, Dublin, Ireland. e-mail: sambaiga@gmail.com

L. Pereira is with ITI, LARSyS, Técnico Lisboa, Portugal. e-mail: lucas.pereira@tecnico.ulisboa.pt

Accurate short-term forecasting of net-load at the distribution level is crucial not only to ensure the reliability of the system but also to increase the ability to accommodate additional RES [6], [7]. Net-load forecasts are a critical component of applications such as scheduling storage control for load-leveling and peak shaving [8], [9], resource allocation, and electricity market participation [10], [11].

Nevertheless, most of the existing forecasting approaches in power systems are specific to the transmission network use-cases and only consider the total load demand [12]. Still, when compared to the transmission system, forecasting load profiles at the distribution level is much more complex due to the impacts of end-user behavior [13]–[15] and the need to cope with the inherent time-varying nature of RES [16].

This work presents a short-term probabilistic net-load forecasting at the distribution level leveraging parameterized deep quantile regression. To be precise, we investigate how to leverage DNN and Quantile Regression (QR) to produce non-parametric probabilistic, day-ahead, net-load forecasts, and uncertainty estimates.

A. Net-load Forecasting

Net-load forecasting has emerged in the power systems research community as the fusion of load demand and RES generation forecasting problems. The net-load refers to the imbalance between the system load and the RES generation to be met by traditional forms of generation [7], [17]. For example, in a distribution system with distributed Photovoltaics (PV) and Wind generation, the net-load at a given instant, t , (NL_t) is given by Eq. (1):

$$NL_t = L_t^G - L_t^{PV} - L_t^{WND} \quad (1)$$

where L_t^G is the load demand, L_t^{PV} is the PV production, and L_t^{WND} is the wind production.

Overall, net-load forecasting approaches fall in two categories: 1) additive and 2) integrated [18]. Approaches under the former category work by decomposing the net-load into two parts: actual load profile and renewable generation, which are then forecasted individually. In contrast, in integrated approaches, renewable generation forecasts are used as inputs to the net-load forecasting model.

The base assumption behind additive methods is that the variation of each part is better explained, modeled, and forecasted individually [6]. Furthermore, by decomposing the net-load into total load and renewable generation, it is possible to

take advantage of the many forecasting alternatives available in the literature [11], [13], [19]–[21]. Approaches for additive net-load forecast have been presented in [6], [18], [22].

The working principle of integrated methods is based on the assumption that the variance in the net-load happens in response to variations on the renewable energy sources [18]. Therefore, by feeding historical net-load and RES forecasts, the learning algorithm would learn to quantify how variance in the RES affects the resulting net-load forecast. Approaches for integrated net-load forecasts have been proposed in [18], [23], [24].

The work in [18] presents a direct comparison of the two approaches in the context of a commercial micro-grid with high solar PV penetration (1 MW solar farm that can meet between 3-55% of the daily power demand). Ultimately, the integrated model outperforms the additive model by 10.69% in terms of Root Mean Squared Error (RMSE). Furthermore, the authors also show that following an integrated approach, it was possible to explain 98% of the integrated net-load forecasting errors using the solar forecast errors. In contrast, only 83% of the variance in the net-load forecasting could be predicted from the solar forecast errors in the additive models.

Ultimately, while no more direct comparisons are available in the literature, this strengthens the base assumption behind integrated models. It further suggests that more advanced models like DNNs may find a role in net-load forecasting. Nevertheless, despite DNNs, can significantly improve short-term (24h - 48h) load forecasting due to their ability to best capture the underlying non-linear relationships on the data [25], to the best of our knowledge, they are still not found in the net-load forecasting literature.

Another aspect to highlight from the literature is that most models are developed and tested using historical measurements for variables such as solar irradiation, wind speed, and temperature. However, full real-life scenarios will also require the inclusion of historical predictions of such variables to assess the robustness [23]. Furthermore, all the reviewed approaches assume that reliable forecasts for total solar or wind generation are readily available at the local level, which is most likely not the case in real-world scenarios [18].

As such, works that take into consideration real-world scenarios and that are more robust to the uncertainty of RES are critical to advance the State-of-the-Art (SOA) in net-load forecasting.

B. Uncertainty Estimation in Forecasting

As distributed RES begin to dominate the generation portfolios, it is also necessary to quantify the uncertainty of the developed forecasting models to assess their predictive confidence, which is critical for the real-world decision process [26]. Still, despite the success already demonstrated by DNNs methods for load forecasting problems to the detriment of more traditional approaches based on probabilistic models, they cannot often convey calibrated uncertainty estimates in their predictions [27].

Proper uncertainty estimates are thus vital for forecasting applications that interact with the physical world, where prediction errors may have serious repercussions [28]–[30]. For

instance, knowing the uncertainty of time series forecasting is critical for anomaly detection, resource allocation, and planning, and other related tasks [31], [32]. Additionally, uncertainty estimates help decide when to delegate high-risk predictions to human intervention or to proceed with a more conservative fall-back plan [32]. For example, high uncertainty in net-load predictions may trigger a set or more conservative actions when deciding to increase (in case of net-load over-estimation) or decrease (net-load under-estimation) the production from non-RES.

Predictive uncertainty can be categorized into two classes: 1) model (*epistemic*) and 2) data (*aleatory*) uncertainty [33]. Epistemic uncertainty indicates how uncertain the model is when explaining the observed data [33], [34]. Thus, providing enough representative training data can reduce this uncertainty type [35]. Aleatoric uncertainty, describes the variance of the conditional distribution of the target variable given input features [33], [34]. This uncertainty arises from the stochastic nature of the observed input features such as noise or measurement errors. In the particular case of net-load forecasting, aleatory uncertainty tends to dominate, especially due to the uncertainty of RES that greatly affects forecasts.

A variety of methods based on Bayesian approaches such as Bayesian Neural Networks (BNNs) or Monte Carlo Dropout (MCD) [36], [37] have been used to quantifying predictive uncertainty in DNNs [26]. These approaches use samples to represent the conditional distribution [27], [34]. Other methods are based on learning the parameters of parametric probability distributions such as Gaussian or Mixture of distributions with DNN [38], [39].

QR is another approach for estimating the conditional distribution that has been successfully applied in different time-series forecasting problems [40], including load forecasting [21], [41]. Compared to other uncertainty quantification approaches, QR can model complex distribution without making any apriori assumptions on the underlying distribution of the data [40], [42]–[45] and produce well-calibrated uncertainty estimates [46], [47]. However, unlike previous quantile forecasting approaches [19], [38], [48], we use a fully parametrized quantile function in which both the quantile fractions and quantile values are parametrized with a DNN.

C. Research Contributions and Paper Organization

This paper makes the following original research contributions:

- 1) We present the Full Parameterized Sequence to Quantile (FPSe2Q) forecasting model which estimates the entire conditional distribution of the target variable. Unlike most approaches for net-load forecasting applying quantile regression, the proposed FPSe2Q learns to best characterize the entire conditional distribution by learning both the quantile values and their associated probabilities.
- 2) We propose a scoring metric that quantifies the predictive uncertainty of the forecasted probabilistic net-load. The proposed metric exhibits the trade-off between sharpness, predictive calibration probability, and predictive forecasting error.

- 3) We demonstrate the efficacy of the proposed approach on real-world distribution substation data in which the proposed algorithm is tested using historical measurements of net-load and solar irradiation data to produce day-ahead forecasts. We also benchmark the proposed approach with existing point-forecast and probabilistic forecast models.
- 4) We demonstrate the real-world applicability of the proposed solution by conducting a real-time evaluation, in a real-world setup that is able to produce a day-ahead net-load forecasts every hour.

The source-code used in this work is available in an online code repository¹.

The remainder of this paper is organized as follows. Section II provides the theoretical background on net-load forecasting and quantile regression. Section III introduces the different methods utilized in this work, including the proposed algorithm, the implementation procedure, and the performance metrics employed. Section IV details the case study, including the data sources, input features, and the evaluation procedures. The outcomes of the evaluation are presented and discussed in Section V. Finally, Section VI summarises the contributions of this paper and provides an outlook on future work.

II. BACKGROUND

A. Net-load Forecasting Problem Formulation

We consider the net-load forecasting problem in discrete-time such that given input features $\mathbf{x}_{i,t:t-L}$ and $\mathbf{z}_{i,t+1:T}$, our goal is to learn day-head forecasting $\mathbf{y}_{i,t+1:T}$ by modeling a conditional distribution $p(\mathbf{y}_{i,t+1:T}|\mathbf{x}_{i,t:t-L})$. The input feature $\mathbf{x}_{i,t:t-L}$ is the historical features that can only be measured at each step and are unknown beforehand, and $\mathbf{z}_{i,t+1:T}$ is the features that are assumed to be known for all time points T such forecasted solar radiation. To this end we adopt multi-target quantile regression which estimate the entire conditional distribution of the target variable such that:

$$p(\mathbf{y}_{i,t0:T}|\mathbf{y}_{i,1:t-1}, \mathbf{z}_{i,1:T}) = \prod_t^T p(y_{i,t}, \tau_{1:N}|\mathbf{x}_{i,1:t-1}, \mathbf{z}_{i,1:T})$$

B. Quantile Regression

DNNs are widely used in supervised learning problems that involve learning to approximate the mean of the conditional distribution $\hat{y} = f_\theta(x)$ where $f_\theta(\cdot)$ is a differentiable non-linear function (a neural network) parameterized by θ . Still, for most situations, conditional averages are not usually good descriptions of data. Instead, QR provides a flexible framework for modelling the entire conditional probability distribution of the data such that:

$$P(y|X = x) = Q(\tau) : (0, 1) \rightarrow \mathbb{R} := \inf\{x \in \mathbb{R}, \tau \leq F(x)\} \quad (2)$$

where $Q(\tau)$ is the quantile function characterised by $p(x \leq Q(\tau))$ and $F_\theta(y|X = x) : \mathbb{R} \rightarrow [0, 1]$ is the cumulative

distribution. For a continuous and strictly monotonically increasing cumulative distribution $F_\theta(y|X = x)$ the quantile function $q^{(\tau)}$ of the random variable y is simply the inverse of cumulative distribution: $q^{(\tau)} = F_\theta^{-1}(\tau)$ for all quantile levels τ [45], [49]. Therefore a non-parametric probabilistic density estimate \hat{f}_y can be obtained by gathering a set of N_τ quantile estimates such that:

$$\hat{f}_y = \{F_\theta^{-1}(\tau_n|\mathbf{x} = x), n = 1, \dots, N_\tau \mid \tau \in (0, 1)\} \quad (3)$$

where $F_\theta^{-1}(\tau|\mathbf{x} = x) = \hat{y}^{(\tau)}$ [45].

To this end DNN could be used to learn $F_\theta^{-1}(\tau_n|\mathbf{x} = x)$ for fixed values of τ optimising pin-ball loss

$$\mathcal{L}(\rho_\tau(\xi)) = \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \rho_\tau(\xi^{\tau_{nt}}) \quad (4)$$

where: $\rho_\tau(\xi) = |\tau - \mathbb{I}\{\xi < 0\}|$ and $\xi_\tau = (y - \hat{y}^{(\tau)})$

This allows the quantification of predictive uncertainty using a prediction interval $\alpha = [\hat{y}^{(\tau_l)}, \hat{y}^{(\tau_u)}]$, which gives a lower and an upper bound between which the predictions lie with a certain probability $p_\tau \in [0, 1]$ where τ_l and τ_u is the lower and upper quantile [29]. Furthermore, it is shown in [32], [33], that it is also possible to quantify both *epistemic* and *aleatory* uncertainty.

III. METHODS

A. Fully Parameterized Sequence to Quantile (FPSe2Q) Forecasting

Existing DNNs quantile regression assume that the quantile probabilities $\tau \in [0, 1]$ are fixed and thus only the quantile value $F_\theta^{-1}(\tau|\mathbf{x} = x)$ needs to be parameterized. As a result, the quantile probabilities (τ) are selected using heuristic approaches, which in most cases do not provide the best estimate of a quantile value [43], [44].

In this work we are interested in parametrizing both the quantile probabilities and the corresponding quantile values using DNNs. To achieve this, we follow the work in [44], and parameterize a non-parametric probabilistic density estimate $\hat{f}(y) = \{F_\theta^{-1}(\tau_n|\mathbf{x} = x), n = 1, \dots, N_\tau\}$ such that both the quantile fractions τ_n and the quantile value $F_\theta^{-1}(\tau_n|\mathbf{x} = x)$ are parametrized by two neural networks. Namely, the Fraction Proposal Network (FPN), and the Quantile Value Network (QVN) such that: $q^\tau = \text{QVN}(\phi(\mathbf{x}))$ and $\tau = \text{FPN}(\phi(\mathbf{x}))$ where $\phi(\mathbf{x}) = f_\theta(\mathbf{x}; \theta)$.

Thus, given the feature space $\phi(\mathbf{x})$, we first generate the set of fractions τ using the FPN and then obtain the quantile values corresponding to τ using the QVN.

The FPN learns to generate set of $N \times T$ adjustable quantile fractions $\tau = \tau_0^t \dots \tau_{N-1}^t \in [0, 1]$ such that $\tau_{i-1}^t < \tau_i^t$ with $\tau_0 = 0$ and $\tau_N^t = 1$. To this end we extend the fully parametrized quantile function proposed in [44] to multivariate distribution by estimating $N \times T$ quantile proposals.

The quantile distribution is approximated by a weighted mixture of $N \times T$ diracs δ given by

$$Q_{\tau,p}(x) := \sum_{i=0}^{N-1} (\tau_{i+1}^t - \tau_i^t) \delta_{p_\theta^i}(x) \quad (5)$$

¹<https://github.com/sambaiga/FPSeq2Q>

where $\delta_\theta(x)$ is the dirac function and $p_\theta^i(x)$ is an estimation of $F_q^{-1}(\hat{\tau}_i^t)$ at quantile level $\hat{\tau}_i^t = \frac{\tau_i^t + \tau_{i+1}^t}{2}$ with $0 \leq i \leq N-1$.

Following [44], the projection-based quantile estimate for any quantile level τ and $p_{\theta i}$ is given by

$$F_q^{-1, \tau, p_\theta^t}(\omega) = p_\theta^0 + \sum_{i=0}^{N-1} (p_\theta^{i+1} - p_\theta^i) H(\omega - \tau^i) \quad (6)$$

where H is the Heaviside step function.

Then the QVN maps the quantiles probabilities to a quantile values $F_\theta^{-1}(\tau_n^t | \mathbf{x} = x)$ $n = 1, \dots, N_{\tau^t} | \tau^t \in (0, 1)$.

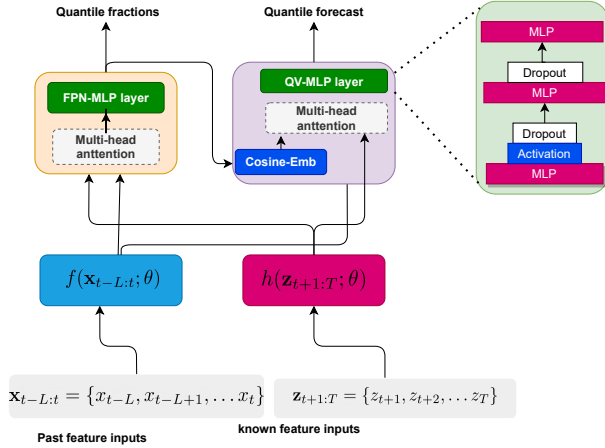


Fig. 1. The proposed FPSe2Q forecast architecture for multi-horizon forecasting.

B. FPSe2Q Model implementation

The implemented FPSe2Q architecture for net-load forecasting is illustrated in Fig. 1. It consists of two encoders: $f(\mathbf{x}; \theta)$ and $h(\mathbf{z}; \theta)$. The $f(\mathbf{x}; \theta)$ encoder takes in the time-varying known and unknown past inputs features $[\mathbf{x}_{i,t:t-L/2}; \mathbf{z}_{i,t:t-L}]$. The $h(\mathbf{z}; \theta)$ encoder, takes in the time-varying a priori known future inputs $\mathbf{z}_{i,t:t+T}$. Both $f(\mathbf{x}; \theta)$ and $h(\mathbf{z}; \theta)$ can be parameterized with any neural network architecture. More specifically, we experiment with the three commonly used neural network architectures for the $f(\mathbf{x}; \theta)$ encoder: Multilayer Perceptron (MLP), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), as described in Section IV. The $h(\mathbf{z})$ encoder is implemented using the recently proposed MLP-Mixer architecture [50].

The two encoders are followed by multi-head attention, and the FPN and QVN layers. Using a multi-head attention mechanism permits attending to parts of the input sequence differently, thereby encoding multiple relationships between the learned feature representation. First, we use multi-head attention to attend information from past-feature inputs $\phi(f)$ and future covariate inputs $\phi(h)$, where $\phi(h)$ is the query, and $\phi(f)$ is the key and value. We also, use the multi-head attention to attends information from the cosine embedding $\Phi(\tau)$, $\Phi(h)$ and $\Phi(f)$ in which $\Phi(\tau)$ is the query, and $\Phi(hf)$; the concatenation of $\Phi(h)$ and $\Phi(f)$ is the key and value respectively.

The FPN consists of one fully connected MLP_{fpn} whose weights are initialised such that the initial quantile probabilities are uniform. It take in the attention feature representation $\phi(A_{hf})$ between $\phi(f)$ and $\phi(h)$ from the two encoders and the previous day net-load demand $\mathbf{x}_{1,t:t-T}$ to produce the quantile probabilities $\tau \in [0, 1]$ such that

$$\phi(A_{hf}) = \text{MultiHead}(\phi(h), \phi(f), \phi(f)) \quad (7)$$

$$\text{Gating}(x) = \text{sigmoid}(\mathbf{x}_{1,t:t-T}) \odot \phi(A_{hf}) + \mathbf{x}_{1,t:t-T} \quad (8)$$

$$\tau_{\{n:N\}} = \text{csumsoftmax}[\text{MLP}_{fpn}(\text{Gating}(x))] \quad (9)$$

where the $\text{csumsoftmax}(\cdot)$ layer ensure that the produces quantile probabilities satisfy the following conditions: $\tau_{i-1} < \tau_i$, τ_i and $\sum_i \tau_i = 1$.

The QVN consists of cosine-embedding (CosineEmb) and MLP_{qvn} layers.

The cosine-embedding computes the embedding of τ denoted by $\Phi(\tau)$. It enables capturing the interaction of the learned τ in predicting the quantile values. The design of the cosine embedding was motivated by the result obtained in [43] which found simple linear embedding of tau to be insufficient to achieve good performance. We also did an ablation study and found that removing the cosine embedding results in unsatisfactory performance.

To produce the quantile values, the MLP_{qvn} layer takes in $\Phi(A_\tau)$, $\Phi(A_\tau)$ and $\phi(A_{hf})$ feature representation such that:

$$\Phi(\tau) := \left(\sum_{i=0}^{N-1} \cos(\pi \tau i) w_{ij} + b_j \right) \quad (10)$$

$$\Phi(A_\tau) = \text{MultiHead}(\Phi(\tau), \phi(hf), \phi(hf)) \quad (11)$$

$$F_q^{-1}(\tau_{n:N}) = \text{MLP}_{qvn}([\Phi(A_\tau) \odot \Phi(\tau)] + \phi(A_{hf})) \quad (12)$$

where $\phi(hf) = [\phi(h); \phi(f)]$.

The mean of the quantile forecast is then obtained as:

$$\bar{F}_q^{-1}(\tau) = \sum_{i=1}^{N-1} (\tau_{i+1} - \tau_i) F_q^{-1} \left(\frac{\tau_{i+1} + \tau_i}{2} \right) \quad (13)$$

To avoid large or small activation values which might result into exploding or vanishing gradient problem, the weights of all the layers (with exception to the FPN) are initialized using Normal Distribution such that $\mathbf{w} \sim \mathcal{N}(0, \sigma^2)$ [51].

C. Loss Function and Training Procedure

The FPN is trained to minimize the 1-Wasserstein metric given by:

$$W_1(q, \tau) = \sum_{m=1}^M \sum_{i=0}^{N-1} \int_{\tau_i}^{\tau_{i+1}} |F_q^{-1m}(\omega) - F_q^{-1}(\hat{\tau}_i^m)| d\omega \quad (14)$$

such that $\frac{\partial W_1}{\partial \tau} = \sum_{m=1}^M 2F_q^1(\tau_i^m) - F_q^1(\hat{\tau}_i^m) - F_q^1(\hat{\tau}_{i-1}^m)$

Traditionally the QVN is trained by minimising the pinball loss (\mathcal{L}) given by Eq. (4) which is an asymmetric loss function that penalises under-estimation by τ and over-estimation by $1 - \tau$.

One of the characteristics of the the pinball loss is that its gradients do not scale with the magnitude of the error,

but with the sign of the error and the quantile value τ [52]. This may lead to an increase in variances of the gradient, which in turn affect the model performance. To address this challenge, we follow the works in [42], [53], [54], that propose the application of the Huber quantile loss to train the QVN network. The Huber quantile loss allows the errors under some threshold κ to be scaled with their magnitude. The Huber quantile loss is given by Eq. (15):

$$\mathcal{L}(\rho_\tau^\kappa(\xi)) = \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \rho_\tau^\kappa(\xi^{\tau_{nt}}) \quad (15)$$

where: $\rho_\tau^\kappa(\xi) = |\tau - \mathbb{I}\{\xi < 0\}| \frac{\mathcal{L}_\kappa(\xi)}{\kappa}$ and $\mathcal{L}_\kappa(\xi)$ is the huber loss defined as

$$\mathcal{L}_\kappa(\xi) = \begin{cases} \frac{1}{2}\xi^2 & \text{if } |\xi| \geq \kappa \\ \kappa|\xi| - \frac{1}{2}\kappa & \text{if } \xi < 0, \end{cases}$$

We also introduce the penalty term to the Huber loss, in order, calibrate the sharpness of the predicted quantiles values such that;

$$\mathcal{L}(\gamma, \hat{\mathbf{y}}_\tau) = \beta \sum_{i=1}^N \max[\gamma, (\hat{y}_{\tau(i-1)} - \hat{y}_{\tau i})] \quad (16)$$

Thus, the parameters θ of the proposed FPSe2Q are learned by jointly minimizing the Huber quantile loss $\mathcal{L}(\rho_\tau^\kappa(\xi))$, the 1-Wasserstein metric with entropy regulation $W_1(q, \tau) + \lambda H(\tau)$ and the sharpness calibration loss $\mathcal{L}(\gamma, \hat{\mathbf{y}}_\tau)$

$$J_\theta(\mathbf{y}, F_q^{-1}(\tau), \tau) = \mathcal{L}(\rho_\tau^\kappa(\xi)) + W_1(q, \tau) + \lambda H(\tau) + \mathcal{L}(\gamma, \hat{\mathbf{y}}_\tau)$$

We use the mini-batch Adam optimizer, a gradient-based optimization of stochastic objective functions [55], with a batch size of 128 to optimise the parameter θ . Particularly, the parameters of the QVN, multi-head attention layer MultiHead, and the $f(\mathbf{x}; \theta)$ and $h(\mathbf{z}; \theta)$ encoders are optimized with Adam with initial learning rate of $1e^{-3}$. The learning rate is decayed by 0.1 when performance stops improving for at least 25 iterations. The parameters of the FPV layer are optimised using another Adam optimiser with a fixed learning rate of $1e^{-4}$. We set the total number of epochs to 200 with proper early stopping to avoid overfitting.

D. Performance Metrics

We adopt point-forecast metrics, predictive interval metrics and probabilistic forecast metrics to assess the performance of FPSe2Q. The forecast accuracy is assessed using Normalised Root Mean Squared Error (NRMSE) point-forecast metric [4], [56], [57]. NRMSE Eq. (17) provide normalised forecast accuracy that may be more interpretable, enabling the comparison between models with different scales.

$$\text{NRMSE} = \frac{1}{R} \sqrt{\sum_{t=1}^{t=T} \frac{(\hat{y}_t - y_t)^2}{T}} \quad (17)$$

In practice, the NRMSE can use different normalisation constants such as mean, standard deviation, interquartile range

and the difference between the maximum and minimum value (R). This work chooses the difference between maximum and minimum (R) as the normalising constant since it reflects the metrics used in net-load forecasting problems where the installed capacity is used as the normalising constant. We calculate the R as the difference between the maximum and minimum value of the test set.

We use the Continuous Ranked Probability Score (CRPS) defined in Eq. (18) to benchmark the performance of probabilistic forecasting and measure the compatibility of the estimated conditional distribution with the ground truth [58]–[60].

$$\text{CRPS} = \frac{1}{T} \sum_{t=1}^T \int_{-\infty}^{\infty} (F(\hat{y}_t) - \mathbb{I}\{\hat{y}_t \leq y_t\})^2 \quad (18)$$

where $\mathbb{I}\{\hat{y}_t \leq y_t\}$ is the indicator function which is 1 if $\hat{y}_t \leq y_t$ and 0 otherwise. $F(\hat{y}_t)$ is cumulative distribution function (CDF) defined as $F(\hat{y}_t) = \int_{-\infty}^{\hat{y}_t} p(y) dy$ and $p(y_t)$ is the PDF of y . This metric attains its minimum when the estimated predictive distribution and the data distribution are agreeable. Furthermore, we use the Probability Integral Transform (PIT) value [60] to test the reliability of the proposed FPSe2Q and assess the statistical consistency of the produced net-load forecasts.

$$\text{PIT} = F(y_t) = \int_{-\infty}^{y_t} p(x) dx \quad (19)$$

The probability predictions are considered reliable when the PIT values are subject to a uniform distribution between 0 and 1.

In addition, we use two competing predictive-interval based metrics, the Predictive Interval Coverage Probability (PICP) and the Normalised Mean Prediction Interval Width (NMPI) to evaluate and quantify the quality of the prediction intervals. The PICP measures the fraction of the ground-truth falling within the predictive interval, whereas the NMPI measures the average width (sharpness) of the prediction interval [29], [46].

$$\text{PICP} = \frac{1}{T} \sum_{t=1}^T I(y_t \in [\hat{y}^{\tau_U}, \hat{y}^{\tau_L}]) \quad (20)$$

$$\text{NMPI} = \frac{1}{TR} \sum_{i=1}^T \hat{y}^{\tau_U} - \hat{y}^{\tau_L} \quad (21)$$

where \hat{y}^{τ_U} and \hat{y}^{τ_L} are upper and lower predictive interval. A well-calibrated model should give predictive intervals that are both correct with higher PICP and a narrow NMPI.

These two metrics are inversely proportional such that, reducing NMPI will lower PICP, and vice-versa. For instance, if the interval is wide enough, it is easy to have higher PICP close to 100%. Still, such a wide interval does not provide enough information on the model uncertainty. To address this, a coverage width based criterion (CWC) that combines PICP and NMPI is proposed in [61], [62] such that:

$$\text{CWC} = \text{NMPI} + \lambda_{\text{PICP}} e^{(-\gamma|\text{PICP} - \alpha|)} \quad (22)$$

where α is the predetermined confidence level, γ is the penalty paramete:

$$\lambda_{\text{PICP}} = \begin{cases} 0 & \text{if PICP} \geq \alpha \\ 1 & \text{otherwise} \end{cases} \quad (23)$$

However, this metric only considers the predictive intervals scores and ignores the forecasting accuracy metrics. Instead, we need a metric that provides a better trade-off between forecasting accuracy and predictive uncertainty. Motivated by this, we define a scoring metric (CWE) that combines both predictive uncertainty and point forecasting metrics, and avoid the need for the penalty parameter γ .

To achieve this, we first define $\frac{2\Sigma}{R}$ as the theoretical upper bound of NMPI where Σ is the ground truth standard deviation. We then define a PIC-based score $\gamma_{\text{nmipi}} \in [0, 1]$ and NMPI based score $\gamma_{\text{picp}} \in [0, 1]$ such that:

$$\gamma_{\text{picp}} = \frac{(1 - \Gamma^E)}{1 + \Delta_{\text{picp}}} \cdot \exp(-\Delta_{\text{picp}}) \quad (24)$$

$$\gamma_{\text{nmipi}} = \frac{(1 - \Gamma^E)}{1 + \Delta_{\text{nmipi}}} \cdot \exp(-\Delta_{\text{nmipi}}) \quad (25)$$

where is $\Gamma^E \in [0, 1]$ is the point-wise forecasting metric defined as $\Gamma^E = \frac{1}{T} \sum_{i=1}^N \text{NMRSE}(\hat{y}_t^{\tau_i}, y_t)$, $\Delta_{\text{picp}} = 1 - \text{PICP}$ and $\Delta_{\text{nmipi}} = |2\Sigma/R - \text{NMPI}|$

Finally, the CWE is set to be the harmonic mean of γ_{picp} and γ_{nmipi} :

$$\text{CWE} = 2 \cdot \frac{\gamma_{\text{nmipi}} \cdot \gamma_{\text{picp}}}{\gamma_{\text{picp}} + \gamma_{\text{nmipi}}} \quad (26)$$

IV. CASE STUDY SPECIFICATION

The proposed methodology is tested using real-world data from a Low-Voltage (LV) distribution substation in a Southern Europe geographically isolated island. This substation has a transformer with an apparent power of 250kVA, which transforms voltage from the transmission grid (6000 V) to the distribution grid (400 V). This substation feeds a LV distribution grid supplying around 100 consumers consisting mostly of domestic, small businesses, and agricultural facilities. There is also a total PV generation capacity of 36 kWp distributed over nine micro-producers for which no historical production profiles are available. The substation has a daily average net-load of 30.80 kW, off-peak net-load of 10 kW, and a peak net-load of 74.08 kW. An overview of the net-load forecasting problem considered in this paper is provided in Fig. 2

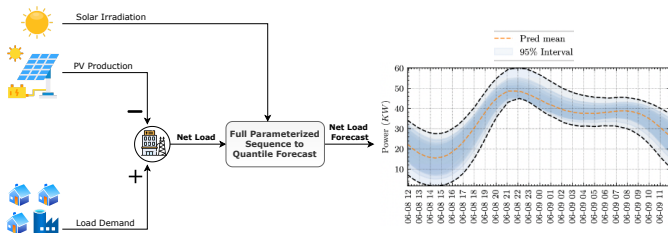


Fig. 2. Schematic diagram of an overview of the net-load forecasting considered in this work

A. Net-load and Solar Irradiation Data

The net-load (NL_t) demand is metered at 1-minute intervals in each phase and our historical dataset spans from March 2019 to September 2020. Over this period, the installed PV generation capacity remained the same. Still, the demand showed a slight increase due to more extended periods at home due to COVID-19 contingency measures in place during 2020 as depicted in Fig. 3. The real-time net-load measurements are also available at 1-minute resolutions through a web-service that was created specifically for this paper to enable real-time evaluation.

The solar irradiation historical data (Ghi_t) was downloaded from Solcast [63] using the geographical coordinates of the substation. The downloaded data spans from March 2019 and September 2020 and are available at 5, 10, 15, 30 and 60-minute resolutions. Real-time solar radiation forecasts are also available for the next seven days at 30-minute resolution through a web service that Solcast provides.

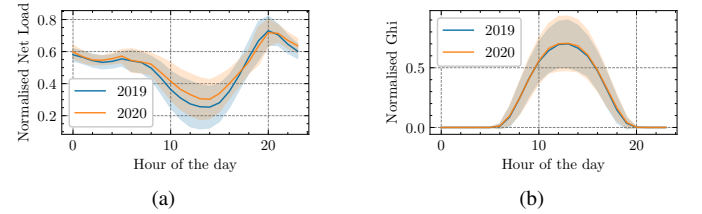


Fig. 3. Seasonality in the trends for the net-load (a) and solar irradiation (b).

B. Input Features

For this work, we use the 30 minutes historical and forecasted net-load and solar irradiation data as input features. Moreover, we also created additional features, ($LGhi_t$) and exogenous time features (\mathbf{ET}_t), which are derived from net-load and solar irradiation signals. The ($LGhi_t$) feature combines net-load and irradiation in a single feature to stress the effects of radiation on the load profile and is obtained as follows $LGhi_{t:t+T} = NL_{t-T:t} - Ghi_{t:t+T}$

After analysing the data we found the net-load demands varies between weekdays and the time of the day. Hence, calendar information was also used as features. More precisely, the following exogenous features were derived (\mathbf{ET}_t): DAY-OFWEEK, HOUR, WEEKDAY, WEEKEND, SATURDAY and SUNDAY, which are obtained as follows:

$$\mathbf{ET}_{i,t} = \left[\sin\left(\frac{2\pi t_i}{T_{si}}\right), \cos\left(\frac{2\pi t_i}{T_{si}}\right) \right] \quad (27)$$

where T_{si} is the period for each of the exogenous features. For instance DAYOFWEEK has a periodicity of 7 whereas HOUR a periodicity of 24. The domain and exogenous features are illustrated in Fig. 4.

The input features are normalised using min-max scaling to have a range between 0 and 1. The forecasted variable is normalised using the same normalisation strategy but with a range of -1 and 1. The model receives 96 data samples corresponding to sequence length (L) as input features to produce a one-day ahead forecast corresponding to 48 samples forecasting horizon (T).

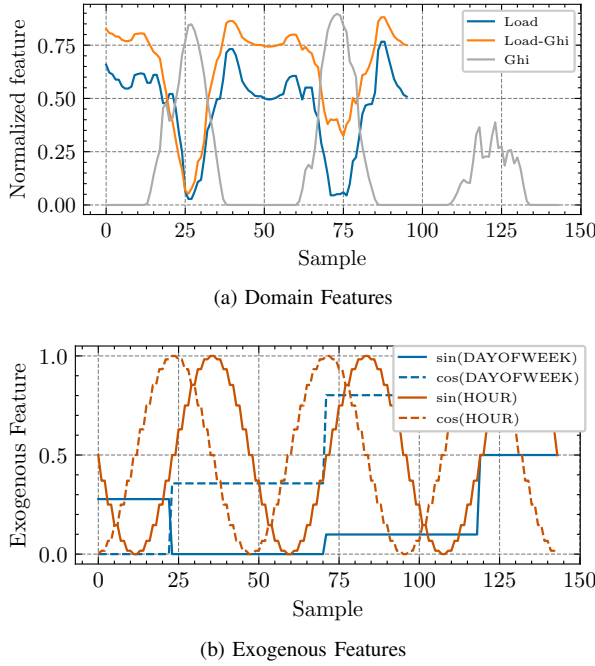


Fig. 4. Illustration of time-varying features used for net load forecasting.

C. Experiment Description

To evaluate the proposed approach, we conducted two distinct experiments, as described next.

- 1) How does the proposed FPSe2Q compare with other existing time series forecasting approaches?
- 2) How robust is the proposed approach for real-time net-load forecasting?

1) *Experiment 1:* We first conduct an in-depth benchmark of the proposed approach against well-known forecasting approaches that exist in the literature. To achieve this, we implemented three different DNN encoders based on MLP (FPSe2Q-MLP), Gated Recurrent Unit (GRU) (FPSe2Q-GRU), and the FPSe2Q-UNET that uses one-dimensional CNN based on the U-Net architecture as proposed in [64].

These implementations are then benchmarked against six alternatives: four point-forecasting - Naive forecasting model based on SARIMA, Support Vector Regression (SVR) [18], Random Forest Regressor (RF) [65]–[67] and two probabilistic forecasting models: Multivariate Gaussian RNN (GRU-Gauss) and State-space model (SSM) based on Hidden Markov Model (HMM) [68], [69].

The parameters for the different models are summarised in Table I.

SARIMA is a traditional forecasting model based on a statistical model. The application of SVMs for additive net-load forecasting was previously proposed in [18], whereas RF represents a widely and high-performing conventional machine learning model for load forecasting [65]–[67]. AR-Net represents a family of forecasting models that combine traditional statistical and neural network models for time series modelling. The GRU-Gauss is equivalent to DeepAR model [38] in which we assume the net-load is distributed according to a multivariate Gaussian distribution with mean

TABLE I
SHARED PARAMETERS FOR EXPERIMENT 1.

Parameters	Values
N	82
Size of Cosine Embedding	64
Dropout	0.4
Latent size	64
Head size	4
Num layer	4
Activation	ReLU
SVR	kernel=rbf, C=1, gamma=auto, epsilon=0.01
RF	n_estimators=100

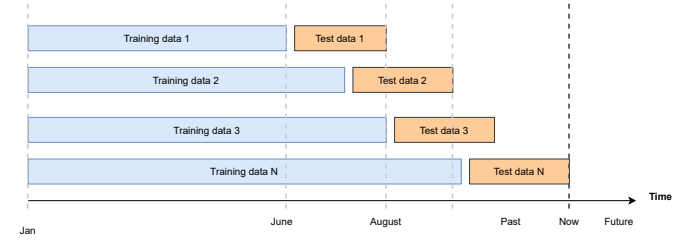


Fig. 5. Backtesting with expanding window

μ and covariance Σ . We then use Gated Recurrent Neural Networks (GRU) to model the dependence of μ and covariance Σ ; it takes the input features and outputs the parameter of the multivariate Gaussian distribution. Finally, the SSM is a linear dynamical system with hidden state [68], [69]. We use the SSM implementation in `pyro-forecasting`² that use LinearHMM with heavy-tailed distributions to model a heavy-tailed linear dynamical system. This model produces forecasts in the form of posterior joint samples at multiple future time steps.

We employ a backtesting approach to benchmark the performance of the proposed approach using the entire historical dataset, i.e., from March 2019 to September 2020. Backtesting is a standard out-of-sample strategy for evaluating time-series forecasting in which models are trained with different historical periods, and their forecast performance is evaluated on specific future time windows. More specifically, we applied the expanding window backtesting strategy, in which the historical periods are expanded by a specific window (in this case, 1 month) as we move forward in the time series. The initial historical period was set to 6 months while the testing set size is kept fixed (2 months), and it corresponds to the future time windows right after the end of the training set as illustrated in Fig. 5. Using expanding window backtesting strategy allowed us to simulate predictions that would have been historically obtained while incorporating all the historical information in the evaluation procedure.

2) *Experiment 2:* In this experiment, the best performing algorithm from experiment one is implemented and tested in real-time to assess how robust the models are to errors in solar irradiation forecasts. To this end, we follow a two-steps approach. First, between August 15th and November 30th of 2021, the experiment was conducted in real-time. More

²<https://docs.pyro.ai/en/latest/contrib.forecast.html>

precisely, every hour, net-load forecasts were produced for the next 24 hours. The input data for this experiments, net-load (historical) and solar irradiation (historical and forecast), were obtained in real-time from the dedicated APIs mentioned in Section IV-A. Then, the first step is replicated using only historical data. I.e., the solar irradiation forecasts are replaced with the actual measured values. For this experiment, we used the best-saved model trained for the two years data from 2019-March to 2020-August. The time window from March 2019 to August 2020 was used for training, while the data for September 2020 were used for validation.

V. RESULTS AND DISCUSSION

In this section we present and discuss the results of the two experiments.

A. Experiment 1: Benchmark

The results of experiment one are summarized in Table II, where each score is the average of the performances of the seven train-test splits.

TABLE II
EXPERIMENT 1 RESULTS

Model	NRMSE	CRPS.10 ³	PICP	CWE
SARIMA	0.49 ± 0.27	-	-	-
AR-Net	0.11 ± 0.02	-	-	-
SVR	0.12 ± 0.01	-	-	-
RF	0.11 ± 0.01	-	-	-
SSM	0.11 ± 0.01	4.09 ± 0.41	0.83 ± 0.05	0.67 ± 0.05
GRU-Gauss	0.08 ± 0.01	2.97 ± 0.16	0.70 ± 0.04	0.57 ± 0.02
FPSe2Q-MLP	0.07 ± 0.01	2.23 ± 0.10	0.91 ± 0.01	0.74 ± 0.02
FPSe2Q-GRU	0.08 ± 0.01	2.30 ± 0.15	0.87 ± 0.03	0.70 ± 0.04
FPSe2Q-UNET	0.08 ± 0.01	2.25 ± 0.16	0.88 ± 0.03	0.72 ± 0.04

It can be observed that the DNN based approaches achieve the best performances, with a NRMSE between 0.07 and 0.08. This shows that DNNs based approaches for net-load forecasting are capable of learning the data non-linear relationship to produce forecasts that are strongly correlated with the ground truth. Likewise, it is interesting to observe that the point-forecast and probabilistic forecasting algorithms also achieve very similar performances, around 0.11 and 0.12.

We also observe that the FPSe2Q improves the CRPS with a skill score between 25-45% when compared to the two probabilistic baselines. In this regard, the FPSe2Q-MLP achieves the lowest CRPS of 2.23×10^3 , meaning that the FPSe2Q is capable of learning the conditional distribution that is compatible with the target distribution. We further recognise that the FPSe2Q provide prediction intervals that contain the majority of the observations, which is reflected by an 87-91% PICP with narrow intervals (0.23-0.25 NMPI). This further suggests that the FPSe2Q is likely to produce more robust uncertainty estimates with a tighter interval and high coverage, which is often desired properties in practice.

Regarding the combined predictive metrics (CWE), the FPSe2Q achieve higher CWE (0.70-0.74), about a 10-21.25% increase compared to SSM and GRUGauss. This confirms that

FPSe2Q learn to estimate a good and well calibrated predictive uncertainty that provide a desirable trade-off between forecasting accuracies and predictive uncertainty performance.

To further understand the obtained results, Fig. 6 illustrates the net-load forecasts obtained by each model for a particular sequence of 48 hours, and the respective uncertainty intervals. This sequence was selected because it illustrates a day with no clouds (first 24 hours), and a day with clouds (second 24 hours), which considerably affect the net-profiles.

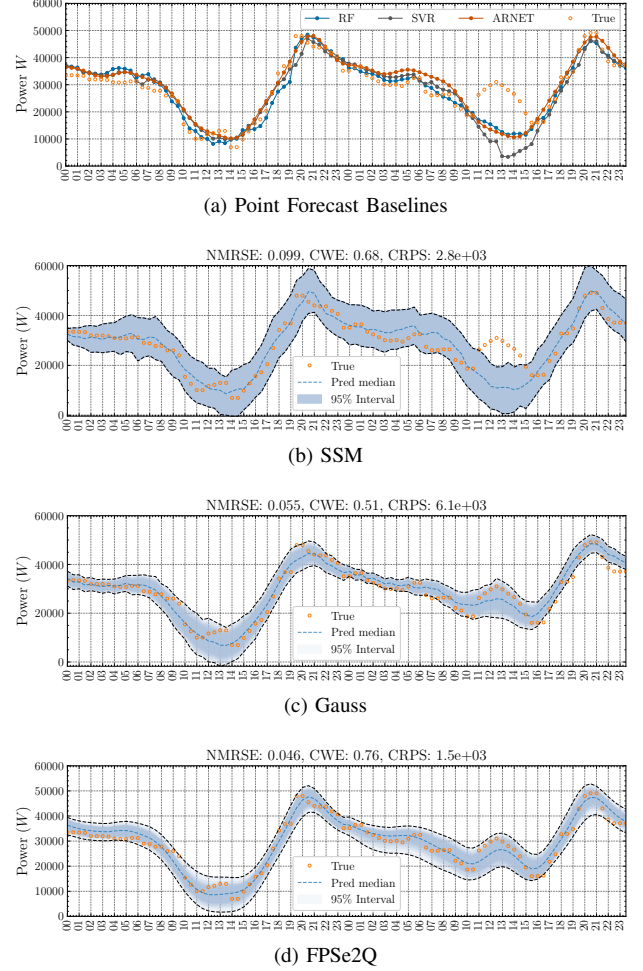


Fig. 6. Illustration of the net load forecast obtained by the different models for 48 hours

As it can be easily observed, the DNN based models are able to learn the two different patterns. In contrast, the remaining models are not able to correctly predict the pattern in the second 24 hours. This highlights the superior ability of DNNs when it comes to finding non-linearities on the data.

Furthermore, it is also possible to see that, although the FPSe2Q and the GRU-Gaus have very similar forecasting performances (0.046 vs. 0.055 in terms of NRMSE, the FPSe2Q shows a much higher CWE value and a much lower value in the Continuous Ranked Probability Score (CRPS) metric. Ultimately, this shows that the FPSe2Q provides a better trade-off between performance and uncertainty estimation.

We further use the Q-Q plots of the PIT values for the three FPSe2Q models shown in Fig. 7 to assess whether these

values are subject to a uniform distribution. We observe that the FPS_{2Q}' PIT values are evenly distributed around the diagonal, and the range evenly covers [0, 1]. Interestingly, all PIT points for the FPS_{2Q}-MLP are located in the Kolmogorov 5% significance band, indicating that the forecasted net-load distribution is reliable. This result is consistent with forecasting accuracy (NRMSE, MAE), probabilistic forecasting performance (CRPS, CWE), and interval prediction score (PICP, NMPI).

B. Experiment 2: Real-time Implementation

The results for the real-time experiment are summarized in Table III, where each score represents the average of all the 24-hours forecasting intervals between 2021-08-17 17:30 to 2021-11-30 09:30. We see that the real-time results are very close to those obtained with only historical data, suggesting that the real-time implementation is feasible. Still, we observe a 10% improvement for forecasting accuracies and predictive uncertainty metrics when historical radiation is used.

TABLE III
EXPERIMENT 2 RESULTS

Radiation	PERIOD	NRMSE	CRPS.10 ³	PICP	CWE
Forecasted	Aug-Sept	0.09 ± 0.03	3.61 ± 1.45	0.80 ± 0.15	0.71 ± 0.12
Forecasted	Oct-Nov	0.09 ± 0.03	3.53 ± 1.33	0.77 ± 0.16	0.69 ± 0.13
Historical	Aug-Sept	0.08 ± 0.03	3.53 ± 1.34	0.81 ± 0.13	0.71 ± 0.11
Historical	Oct-Nov	0.09 ± 0.03	3.53 ± 1.27	0.78 ± 0.14	0.69 ± 0.11

To further understand the effects of the solar radiation predictions in the net-load forecasts and how these effects are reflected in the uncertainty estimates, Fig. 8 depicts the results for two days with very different performances. As it can be observed, the performance of the real-time forecasts on the second day (Fig. 8d) are inferior (0.2 NRMSE), which is very likely to have happened due to the considerable deviation between the forecasted and the actual solar irradiation (Fig. 8b).

Nevertheless, it is also important to stress that both the CWE and CRPS metrics can capture this considerable drop in performance. For example, it is possible to see for the second day that when using forecasts (cresubfig:forecat-day2), the CWEE scores much lower when compared to the same period using the historical irradiation (Fig. 8f).

Finally, as expected, the real-time achieve a higher CWE score when the forecasted radiation is very close to historical radiation as depicted in Fig. 8c.

VI. CONCLUSION AND FUTURE WORK

This paper presents a day-ahead net-load probabilistic forecasting approach that uses historical net-load and solar irradiation as inputs. Specifically, we propose a deep quantile regression network that uses DNN to parameterize the quantile fractions and the associated quantile values to learn the net-load conditional distribution. We propose a scoring metric that combines the trade-off between sharpness, predictive calibration probability, and forecasting error. The proposed scoring metric provides the most suitable way to evaluate the

predictive uncertainty and forecast accuracy performance of the proposed approach.

Extensive evaluations were performed on real-world data from a low-voltage distribution substation in a geographically isolated Southern European island. We also evaluate the robustness of the models when applied in real-time using solar irradiation forecasts instead of historical forecasts. Despite being a very appropriate evaluation step, real-time evaluation of net-load forecast using solar irradiation forecasts is not found in the related literature. Ultimately, our results show that the proposed FPS_{2Q} architecture outperforms several state-of-the-art forecasting alternatives and produces well-calibrated forecasts that are very robust even when applied in real-time.

Similarly, by conducting a real-time evaluation, we found that the obtained forecasts are sensitive to errors in the irradiation forecasts suggesting that it is necessary to consider the uncertainty of the obtained radiation forecasts before making decisions. Unfortunately, this is not ordinarily available. Therefore, future work needs to address how uncertainty can be estimated beforehand by looking at historical data. One possible approach would be to explore modelling both the data distribution and conditional distribution leveraging advanced deep generative models.

The proposed scoring metric provides the best way to evaluate the quality of predictive interval coverage (PICP) with the assumption that higher PICP indicates that all points, including the peaks, will be within the predicated coverage. Nevertheless, it is essential for net-load forecasting at the distribution level to assess both the coverage and peaks. In the future, we will investigate the most suitable way to assess both the coverage and peaks effectively. Furthermore, future work should also seek to establish a scoring metric that combines the data and predictive uncertainty to assess the predictive uncertainty in the absence of ground truth.

While we evaluated our approach in real-world settings, this work's limitation is that we only considered one dataset. Future work should, therefore, consider assessing the performance of FPS_{2Q} in other datasets and other application domains. For instance, the considered sub-station is considerably small and does not have wind production. In the future, it would be relevant to study the applicability of the method in larger substations and its derivatives to situations where net-load is also affected by wind production.

Finally, another crucial aspect that the research community should address is how to combine forecasts and uncertainty to solve real-world downstream tasks such as load-levelling, peak-shaving, resource allocation and optimal control of the battery-integrated energy system. For example, in the context of battery storage control for load-levelling and peak shaving, accurate net-load forecasts, and the respective uncertainty estimates, are required in order to define the optimal storage set-points, as demonstrated in recent works (e.g., [8], [70]). Another example of how net-load forecasts and uncertainty estimates can play a role in the smart-grid is related to the coordination between Distribution System Operators (DSOs) and Transmission System Operators (TSOs), for the provision of active and reactive power from the former to the latter (e.g., [71], [72]). In this context, accurate uncertainty estimates of

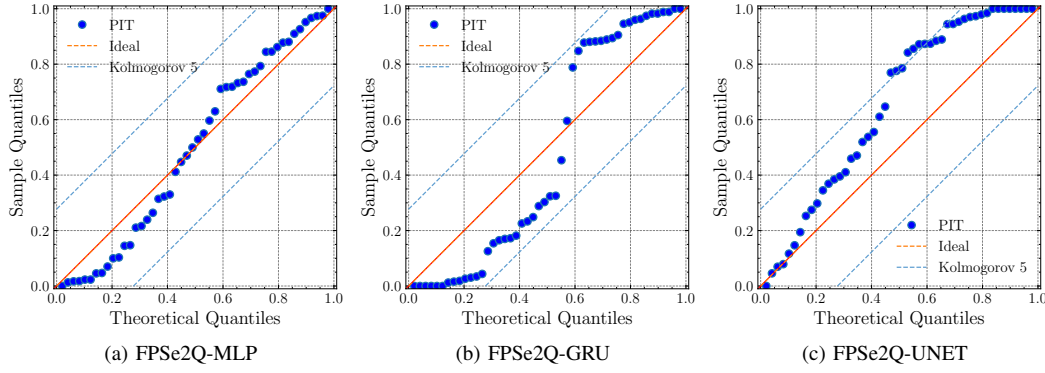


Fig. 7. Reliability diagrams (Q-Q plot) for FPSe2Q with three different encoders (MLP, GRU and UNET)

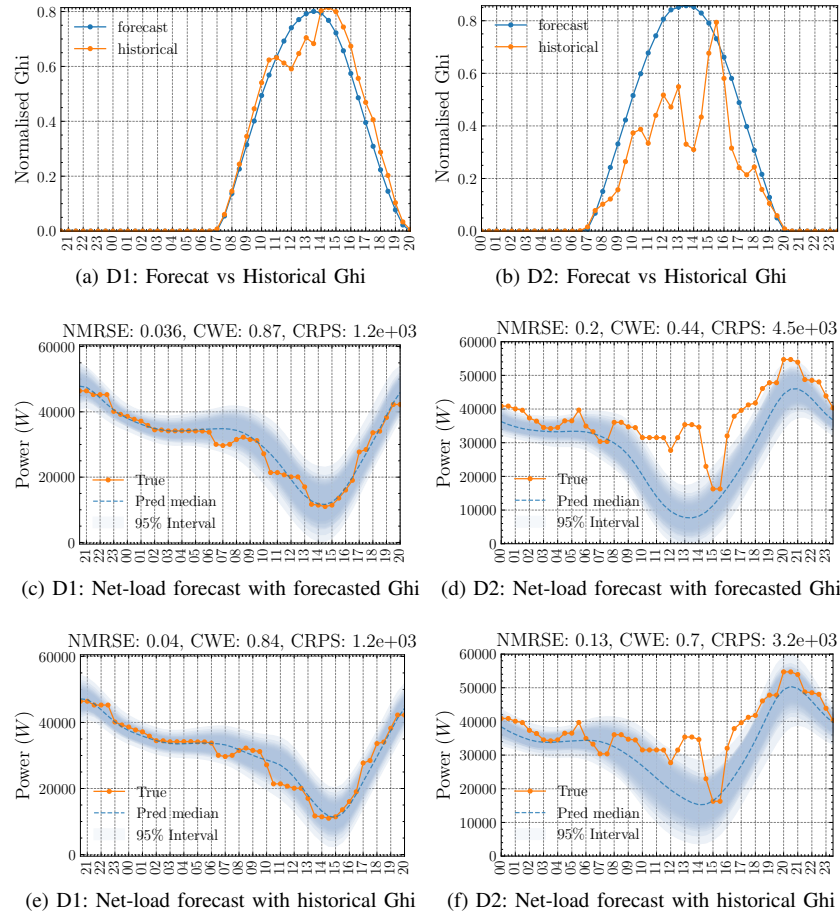


Fig. 8. Real-time experiment results obtained for one day with almost perfect radiation forecast (D1) and another day with poor radiation forecasts (D2).

RES forecasts are crucial to define the active and reactive power ranges that must be met by the DSO without incurring penalties for not being able to meet the established services provision contracts.

ACKNOWLEDGMENT

Lucas Pereira has received funding from the Portuguese Foundation for Science and Technology (FCT) under grants CEECIND/01179/2017 and UIDB/50009/2020.

REFERENCES

- [1] M. Lima, L. Mendes, G. Mothé, F. Linhares, M. de Castro, M. da Silva, and M. Stel, "Renewable energy in reducing greenhouse gas emissions: Reaching the goals of the paris agreement in brazil," *Environmental Development*, vol. 33, p. 100504, Mar. 2020.
- [2] A. Azizivahed, A. Arefi, S. Ghavideh, M. Shafie-khah, L. Li, J. Zhang, and J. Catalao, "Energy management strategy in dynamic distribution network reconfiguration considering renewable energy resources and storage," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, 2020, pp. 1–1.
- [3] D. N. Anthony Faustine, Lucas Pereira and L. Benabbou, "Leveraging machine learning for sustainable and self-sufficient energy communi-

- ties,” in *NeurIPS 2020 Workshop Tackling Climate Change with Machine Learning*, 2020.
- [4] M. Alipour, J. Aghaei, M. Norouzi, T. Niknam, S. Hashemi, and M. Lehtonen, “A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration,” *Energy*, vol. 205, p. 118106, Aug. 2020.
 - [5] B. Zhou, Y. Meng, W. Huang, H. Wang, L. Deng, S. Huang, and J. Wei, “Multi-energy net load forecasting for integrated local energy systems with heterogeneous prosumers,” *International Journal of Electrical Power & Energy Systems*, vol. 126, p. 106542, Mar. 2021. [Online]. Available: <https://doi.org/10.1016/j.ijepes.2020.106542>
 - [6] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, “Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV,” *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3255–3264, 2018.
 - [7] S. Sreekumar, K. C. Sharma, and R. Bhakar, “Grey System Theory Based Net Load Forecasting for High Renewable Penetrated Power Systems,” *Technology and Economics of Smart Grids and Sustainable Energy*, vol. 5, no. 1, p. 21, Oct. 2020.
 - [8] S. Chapaloglou, A. Nesiadis, P. Iliadis, K. Atsonios, N. Nikolopoulos, P. Grammelis, C. Yiakopoulos, I. Antoniadis, and E. Kakaras, “Smart energy management algorithm for load smoothing and peak shaving based on load forecasting of an island’s power system,” *Applied Energy*, vol. 238, pp. 627–642, Mar. 2019.
 - [9] M. Uddin, M. F. Romlie, M. F. Abdullah, C. Tan, G. Shafiullah, and A. H. A. Bakar, “A novel peak shaving algorithm for islanded microgrid using battery energy storage system,” *Energy*, vol. 196, p. 117084, Apr. 2020.
 - [10] A. Kaur, L. Nonnenmacher, and C. F. Coimbra, “Net load forecasting for high renewable energy penetration grids,” *Energy*, vol. 114, pp. 1073–1084, Nov. 2016. [Online]. Available: <https://doi.org/10.1016/j.energy.2016.08.067>
 - [11] S. E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D. B. Smith, and M. Minakshi, “From load to net energy forecasting: Short-term residential forecasting for the blend of load and pv behind the meter,” *IEEE Access*, vol. 8, pp. 224 343–224 353, 2020.
 - [12] T. Hong and S. Fan, “Probabilistic electric load forecasting: A tutorial review,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016.
 - [13] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, and M. T. Miller, “An Efficient Approach to Short-Term Load Forecasting at the Distribution Level,” *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2526–2537, 2016.
 - [14] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, “A Short-Term and High-Resolution Distribution System Load Forecasting Approach Using Support Vector Regression With Hybrid Parameters Optimization,” *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3341–3350, 2018.
 - [15] B. P. Hayes, J. K. Gruber, and M. Prodanovic, “Multi-nodal short-term energy forecasting using smart meter data,” *IET Generation, Transmission & Distribution*, vol. 12, no. 12, pp. 2988–2994, 2018.
 - [16] J. A. P. Lopes, A. G. Madureira, M. Matos, R. J. Bessa, V. Monteiro, J. L. Afonso, S. F. Santos, J. P. S. Catalão, C. H. Antunes, and P. Magalhães, “The future of power systems: Challenges, trends, and upcoming paradigms,” *WIREs Energy and Environment*, vol. 9, no. 3, p. e368, 2020.
 - [17] L. Jin, D. Cong, L. Guangyi, and Y. Jilai, “Short-term net feeder load forecasting of microgrid considering weather conditions,” in *2014 IEEE International Energy Conference (ENERGYCON)*, 2014, pp. 1205–1209.
 - [18] A. Kaur, L. Nonnenmacher, and C. F. M. Coimbra, “Net load forecasting for high renewable energy penetration grids,” *Energy*, vol. 114, pp. 1073–1084, 2016.
 - [19] X. Zhou, K. Gong, C. Zhu, J. Hua, and Z. Xu, “Optimal energy management strategy considering forecast uncertainty based on lstm-quantile regression,” in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, 2020, pp. 2753–2757.
 - [20] H. Aprillia, H.-T. Yang, and C.-M. Huang, “Statistical Load Forecasting Using Optimal Quantile Regression Random Forest and Risk Assessment Index,” *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1467–1480, Mar. 2021.
 - [21] W. Zhang, H. Quan, O. Gandhi, R. Rajagopal, C.-W. Tan, and D. Srinivasan, “Improving Probabilistic Load Forecasting Using Quantile Regression NN With Skip Connections,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5442–5450, Nov. 2020.
 - [22] P. Fonte, C. Monteiro, and F. M. Barbosa, “Net load forecasting in presence of renewable power curtailment,” in *2016 13th International Conference on the European Energy Market (EEM)*, 2016, pp. 1–5.
 - [23] P. Kobylinski, M. Wierzbowski, and K. Piotrowski, “High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources,” *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105635, 2020.
 - [24] J. Browell and M. Fasiolo, “Probabilistic Forecasting of Regional Net-Load With Conditional Extremes and Gridded NWP,” *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5011–5019, Nov. 2021.
 - [25] B. Völker, A. Reinhardt, A. Faustine, and L. Pereira, “Watt’s up at Home? Smart Meter Data Analytics from a Consumer-Centric Perspective,” *Energies*, vol. 14, no. 3, p. 719, Jan. 2021.
 - [26] Y. Ovidia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, “Can you trust your model’s uncertainty? Evaluating predictive uncertainty under dataset shift,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019.
 - [27] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning,” in *International Conference on Machine Learning*. PMLR, Jun. 2016, pp. 1050–1059.
 - [28] M. Teye, H. Azizpour, and K. Smith, “Bayesian uncertainty estimation for batch normalized deep networks,” in *Proceedings of Machine Learning Research*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. Stockholm: Stockholm Sweden: PMLR, 10–15 Jul 2018, pp. 4907–4916. [Online]. Available: <http://proceedings.mlr.press/v80/teye18a.html>
 - [29] N. Tagasovska and D. Lopez-Paz, “Single-Model Uncertainties for Deep Learning,” in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.
 - [30] M. Teye, H. Azizpour, and K. Smith, “Bayesian Uncertainty Estimation for Batch Normalized Deep Networks,” in *Proceedings of the 35th International Conference on Machine Learning*. PMLR, Jul. 2018, pp. 4907–4916.
 - [31] L. Zhu and N. Laptev, “Deep and confident prediction for time series at uber,” in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 103–110.
 - [32] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059. [Online]. Available: <http://proceedings.mlr.press/v48/gal16.html>
 - [33] A. Loquercio, M. Segu, and D. Scaramuzza, “A general framework for uncertainty estimation in deep learning,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3153–3160, 2020.
 - [34] M.-H. Laves, S. Ihler, J. F. Fast, L. A. Kahrs, and T. Ortmaier, “Well-calibrated regression uncertainty in medical imaging with deep learning,” in *Medical Imaging with Deep Learning*, 2020. [Online]. Available: <https://openreview.net/forum?id=AWfzf1d-IG2>
 - [35] M. Teye, H. Azizpour, and K. Smith, “Bayesian uncertainty estimation for batch normalized deep networks,” in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 4907–4916. [Online]. Available: <http://proceedings.mlr.press/v80/teye18a.html>
 - [36] K. O. Ellefsen, C. P. Martin, and J. Tørresen, “How do mixture density rnns predict the future?” *ArXiv*, vol. abs/1901.07859, 2019.
 - [37] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, and M. Khazaei, “Deep-based conditional probability density function forecasting of residential loads,” *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3646–3657, 2020.
 - [38] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, “Deepar: Probabilistic forecasting with autoregressive recurrent networks,” *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207019301888>
 - [39] S. Choi, S. Hong, S. Lim, and K. Lee, “Task agnostic robust learning on corrupt outputs by correlation-guided mixture density networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3871–3880, 2020.
 - [40] D. Kitahara, K. Leng, Y. Tezuka, and A. Hirabayashi, “Simultaneous spline quantile regression under shape constraints,” in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 2423–2427.
 - [41] W. Zhang, H. Quan, and D. Srinivasan, “An Improved Quantile Regression Neural Network for Probabilistic Load Forecasting,” *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4425–4434, Jul. 2019.
 - [42] A. Aravkin, A. Lozano, R. Luss, and P. Kambadur, “Orthogonal matching pursuit for sparse quantile regression,” in *2014 IEEE International Conference on Data Mining*, 2014, pp. 11–19.

- [43] W. Dabney, G. Ostrovski, D. Silver, and R. Munos, "Implicit quantile networks for distributional reinforcement learning," *CoRR*, vol. abs/1806.06923, 2018. [Online]. Available: <http://arxiv.org/abs/1806.06923>
- [44] D. Yang, L. Zhao, Z. Lin, T. Qin, J. Bian, and T.-Y. Liu, "Fully parameterized quantile function for distributional reinforcement learning," in *Advances in Neural Information Processing Systems*, 2019, pp. 6190–6199.
- [45] Y. Romano, E. Patterson, and E. Candes, "Conformalized quantile regression," in *Advances in Neural Information Processing Systems* 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 3543–3553. [Online]. Available: <http://papers.nips.cc/paper/8613-conformalized-quantile-regression.pdf>
- [46] F. Rodrigues and F. C. Pereira, "Beyond expectation: Deep joint mean and quantile regression for spatiotemporal problems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5377–5389, 2020.
- [47] Y. Chung, W. Neiswanger, I. Char, and J. Schneider, "Beyond pinball loss: Quantile methods for calibrated uncertainty quantification," 2020.
- [48] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," 2020.
- [49] K. Hatalis, A. Lamadrid, K. Scheinberg, and S. Kishore, "A novel smoothed loss and penalty function for noncrossing composite quantile estimation via deep neural networks," *ArXiv*, vol. abs/1909.12122, 2019.
- [50] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "Mlp-mixer: An all-mlp architecture for vision," 2021.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015. [Online]. Available: <http://arxiv.org/abs/1502.01852>
- [52] G. Ostrovski, W. Dabney, and R. Munos, "Autoregressive quantile networks for generative modeling," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. Stockholmsmässan, Stockholm Sweden: PMLR, 10–15 Jul 2018, pp. 3936–3945. [Online]. Available: <http://proceedings.mlr.press/v80/ostrovski18a.html>
- [53] S. Zheng, "Gradient descent algorithms for quantile regression with smooth approximation," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 3, pp. 191–207, Jul. 2011. [Online]. Available: <https://doi.org/10.1007/s13042-011-0031-2>
- [54] K. Hatalis, A. J. Lamadrid, K. Scheinberg, and S. Kishore, "A novel smoothed loss and penalty function for noncrossing composite quantile estimation via deep neural networks," 2019.
- [55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [56] K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi, "Methods for comparing uncertainty quantifications for material property predictions," *Machine Learning: Science and Technology*, vol. 1, no. 2, p. 025006, may 2020. [Online]. Available: <https://doi.org/10.1088/2632-2153/ab7e1a>
- [57] P. Koponen, J. Ikäheimo, J. Koskela, C. Brester, and H. Niska, "Assessing and comparing short term load forecasting performance," *Energies*, vol. 13, no. 8, 2020. [Online]. Available: <https://www.mdpi.com/1996-1073/13/8/2054>
- [58] K. Rasul, C. Seward, I. Schuster, and R. Vollgraf, "Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting," *CoRR*, vol. abs/2101.12072, 2021. [Online]. Available: <https://arxiv.org/abs/2101.12072>
- [59] M. Bieshaar, J. Schreiber, S. Vogt, A. Gensler, and B. Sick, "Quantile surfaces—generalizing quantile regression to multivariate targets," *arXiv preprint arXiv:2010.05898*, 2020.
- [60] Z. Zhang, H. Qin, Y. Liu, L. Yao, X. Yu, J. Lu, Z. Jiang, and Z. Feng, "Wind speed forecasting based on quantile regression minimal gated memory network and kernel density estimation," *Energy Conversion and Management*, vol. 196, pp. 1395–1409, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0196890419306958>
- [61] L. Zhang, L. Xie, Q. Han, Z. Wang, and C. Huang, "Probability density forecasting of wind speed based on quantile regression and kernel density estimation," *Energies*, vol. 13, no. 22, 2020. [Online]. Available: <https://www.mdpi.com/1996-1073/13/22/6125>
- [62] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 337–346, 2011.
- [63] Solcast, "Solcast - Solar Forecasting & Solar Irradiance Data," <https://solcast.com/>, 2021.
- [64] A. Faustine, L. Pereira, H. Bousbiat, and S. Kulkarni, "Unet-nilm: A deep neural network for multi-tasks appliances state detection and power estimation in nilm," in *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, ser. NILM'20. New York, NY, USA: Association for Computing Machinery, 2020, p. 84–88.
- [65] A. A. Muzumdar, C. N. Modi, M. G. M., and C. Vyjayanthi, "Designing a robust and accurate model for consumer centric short term load forecasting in microgrid environment," *IEEE Systems Journal*, pp. 1–12, 2021.
- [66] N. Huang, W. Wang, S. Wang, J. Wang, G. Cai, and L. Zhang, "Incorporating load fluctuation in feature importance profile clustering for day-ahead aggregated residential load forecasting," *IEEE Access*, vol. 8, pp. 25 198–25 209, 2020.
- [67] H. Aprillia, H.-T. Yang, and C.-M. Huang, "Statistical load forecasting using optimal quantile regression random forest and risk assessment index," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1467–1480, 2021.
- [68] G. Kitagawa, *State Space Modeling of Time Series*. Dordrecht: Springer Netherlands, 1994, pp. 43–62. [Online]. Available: https://doi.org/10.1007/978-94-011-0866-9_4
- [69] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. Oxford University Press, May 2012. [Online]. Available: <https://doi.org/10.1093/acprof:oso/9780199641178.001.0001>
- [70] P. Iliadis, S. Ntomalis, K. Atsonios, A. Nesiadis, N. Nikolopoulos, and P. Grammelis, "Energy management and techno-economic assessment of a predictive battery storage system applying a load levelling operational strategy in island systems," *International Journal of Energy Research*, vol. 45, no. 2, pp. 2709–2727, 2021.
- [71] J. Silva, J. Sumaili, R. J. Bessa, L. Seca, M. A. Matos, V. Miranda, M. Caujolle, B. Goncer, and M. Sebastian-Viana, "Estimating the Active and Reactive Power Flexibility Area at the TSO-DSO Interface," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4741–4750, Sep. 2018.
- [72] T. Soares, L. Carvalho, H. Morais, R. J. Bessa, T. Abreu, and E. Lambert, "Reactive power provision by the DSO to the TSO considering renewable energy sources uncertainty," *Sustainable Energy, Grids and Networks*, vol. 22, p. 100333, Jun. 2020.



Anthony Faustine is a machine learning researcher with experience applying data analytics and machine learning techniques to real-world problems. He is presently pursuing PhD research at Further Energy and Environment research Laboratory (FEE-Lab), at ITI/LARSyS, Técnico Lisboa in Portugal. His research interest is on robust machine learning algorithms for industrial applications, focusing on learning algorithms that allow uncertainty quantification and are robust to low data settings. Currently, Anthony works with Irish Manufacturing Research, where he develops data analytics and AI solutions to help industries and manufacturing design more effective solutions.



Lucas Pereira received his Ph.D. in Computer Science from the University of Madeira, Portugal, in 2016. Since then, he is at ITI/LARSyS, where he leads the Further Energy and Environment research Laboratory (FEE-Lab). Since 2019 he is a research fellow at Técnico Lisboa. Lucas's research applies data science, machine learning, and human-computer interaction techniques towards bridging the gap between laboratory and real-world applicability of ICT for the sustainable development goals (SDGs). His current research focuses on future energy systems and sustainable built environments and it typically involves the real-world deployment and evaluation of monitoring technologies and software systems.