

# Efficiency through Simplicity: MLP-based Approach for Net-Load Forecasting with Uncertainty Estimates in Low-Voltage Distribution Networks

Anthony Faustine, Nuno Jardim Nunes, Lucas Pereira,

**Abstract**—Power demand forecasting is becoming a crucial tool for the planning and operation of Low Voltage (LV) distribution systems. Most importantly, the high penetration of Photovoltaics (PV) power generation as part of Distributed Energy Resource (DER)s has transformed the power demand forecasting problem at the distribution level into net-load forecasting. This paper introduces a novel and scalable approach to probabilistic forecasting at LV substation with PV generation. It presents a multi-variate probabilistic forecasting approach, leveraging Quantile Regression (QR). The proposed architecture uses a computationally efficient feed-forward neural net to capture the complex interaction between the historical load demands and covariate variables such as solar irradiance. It is empirically demonstrated that the proposed method can efficiently produce well-calibrated forecasts, both auto-regressively or in a single forward pass. Furthermore, a benchmark against four state-of-the-art forecasting approaches shows that the proposed approach offers a desirable trade-off between forecasting accuracies, calibrated uncertainty, and computation complexity.

**Index Terms**—Deep Neural Networks (DNN) Feed-forward Neural Network (FFN) LV distribution Substation, Multilayer Perceptron (MLP), Net-Load, Probabilistic Forecasting, PV Generation, Quantile Regression

## I. INTRODUCTION

THE power grid is experiencing a transition from centralized power grids to decentralized ones, driven by the integration of Renewable Energy Sources (RES), Energy Storage System (ESS) and electrification of sectors such as Electric Vehicle (EV) and Electric Heating Systems (EHS). This shift is causing significant changes in power generation patterns and demands and poses new challenges for power flow management in LV distribution network [1]. Recent research emphasizes the importance of precise generation and demand forecasts at the LV distribution network level to ensure smooth network operation [1], [2]. Accurate short-term forecasts are indeed crucial for maintaining a stable and reliable LV distribution network. These forecasts, covering the next 30 minutes to a few days, play a significant role in ensuring that the LV grid operates within key grid constraints, such as voltage levels, thermal limits, and phase balance [3], [4]. The ability of Distribution System Operator (DSO)s to forecast power

demand and supply at LV distribution network is particularly critical due to the volatile demands and stochastic power generation from RES. For example, accurate forecasting of LV power demand and supply is essential in optimizing the operation of ESS. This allows for efficient charging during periods of abundant green electricity and discharge during times of high demand and low availability from renewables.

As a result of the growing uncertainty in load demands and generation, there has been a noticeable shift towards using probabilistic methods to forecast these variables in LV distribution, as seen in recent studies [5], [6]. Unlike deterministic forecasts, probabilistic forecasts provide uncertainty estimates, which are essential in accurately predicting LV power demands, particularly in situations where the use of DER exacerbates the level of uncertainty. [4]. However, many machine learning-based forecasting techniques, particularly those utilizing DNN, often overlook the importance of computational efficiency and scalability. These factors are crucial in reducing computation costs when deploying models in large LV distribution networks with numerous LV substations [7]. Furthermore, for effective decision-making, models with fast inference times are essential. This ensures that forecasting models can support real-time or near-real-time decision processes in the dynamic context of LV distribution networks, which often require quick responses to changing conditions and load demands [7]. An emerging technique in LV distribution forecasting is net-load forecasting, which combines power demand and generation forecasting at this level. Net-load forecasting at the LV level is increasingly critical due to the rising penetration of DER such as PV and wind turbines [8]–[10]. This method provides a more accurate estimate of the power required from the grid, particularly in regions with a high proportion of DER. For example, in a scenario with PV generation, the net-load  $P_{NL,t}$  at a given time instant  $t$  can be calculated using  $P_{NL,t} = P_{L,t}^D - P_{L,t}^{PV}$ , as illustrated in Fig. 1. where  $P_{L,t}^D$  is the load demand,  $P_{L,t}^{PV}$  is the PV production, and  $P_{L,t}^{W\&D}$  is the wind production.

Several researchers have emphasized the importance of accurate net-load forecasting for efficient grid management and planning, given the difficulties associated with directly measuring the generation from distributed generation at the LV distribution level [11]–[13].

It is worth noting that net-load forecasting at LV level is more challenging than at High Voltage (HV) transmission level due to the low number of aggregated demand profiles and volatile RES generation, resulting in less predictable

A. Faustine is with Center for Intelligent Power (CIP), Eaton corporation, Dublin, Ireland & ITI, LARSyS, Técnico Lisboa, Portugal. e-mail: sam-baiga@gmail.com

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

N. J. Nunes is with ITI, LARSyS, Técnico Lisboa, Portugal. e-mail: nunojnunes@tecnico.ulisboa.pt

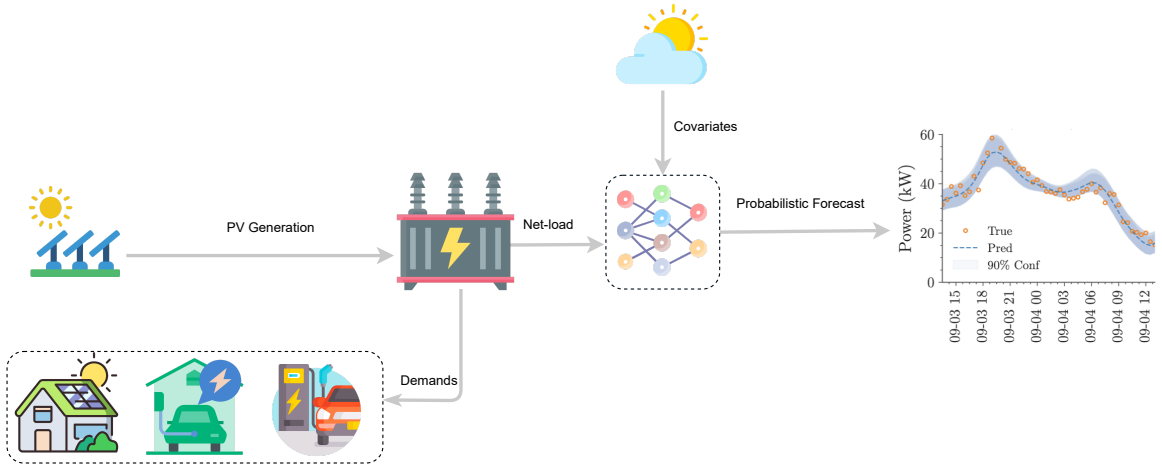


Fig. 1: Schematic diagram of an overview of the net-load forecasting at LV substation.

group behaviors. In addition, forecasting net-load at the LV distribution level remains challenging due to several factors, including the integration of behind-the-meter PV systems and the intermittent and varying nature of weather variables such as solar irradiation, which significantly impact the net-load, as highlighted in [14].

Furthermore, although different models have been proposed to tackle both problems, meteorological variables are the most common explanatory input, particularly temperature, solar irradiation, and wind speed [15]. The base assumption behind this design decision is that the strong correlation between load and temperature can help explain the variations in demand (e.g., an increase in demand in the winter due to colder temperatures). In contrast, solar irradiation and wind speed can help explain the variation in the yield of RES. However, the interactions between these variables or their effects on the forecasting results have been seldom explored, and the reported results are not unanimous. For example, in [9], the authors report that temperature can be detrimental to forecasting accuracy. In contrast, in [8], the authors report that the temperature accounted for half of the variation in the load. Further research is thus necessary to understand the complex relationships between these variables and their impact on forecasting accuracy.

In light of these considerations, our research aligns with the perspective that accurate net-load forecasting is crucial, and we adopt an integrated approach to capture the practical challenges encountered by DSOs in dealing with unobserved behind-the-meter PV installations [16]. The proposed methodology exclusively relies on historical net-load data and other exogenous variables that impact PV and power demands to predict the future net-load. Consequently, our methodology provides a more realistic and applicable framework for net-load forecasting in the presence of distributed energy resources since it does not assume previous knowledge of the installed PV capacity.

Our study proposes a scalable architecture that utilizes MLPs to effectively capture the intricate relationship between historical net-load and meteorological variables. We specifically employ Quantile Regression (QR), which is fully

parameterized to make predictions with some level of uncertainty, and thus we name our architecture Parameterized Quantile Forecast using Multilayer Perceptron (MLPQF). The proposed approach is evaluated on two datasets: Madeira Low-Voltage distribution substation dataset in Portugal (MLVS-PT) and The Stentaway substation dataset in Plymouth-UK (SPS-UK), which represent real-world LV distribution networks in Europe. To state more precisely, this paper makes the following original research contributions:

- 1) We introduce an efficient MLP-based architecture for probabilistic net-load forecasting at the LV substation level. Our model utilizes QR parameterized by MLP to estimate the conditional distribution of net-load forecasts. It extends a previous model described in [17], streamlining and optimizing the architecture to make it more feasible for industrial applications.
- 2) We perform comprehensive experiments on two real-world datasets to evaluate the stability of our proposed approach and the accuracy of uncertainty estimates for the Auto-Regressive Inference (ARI) net-load forecasting task. The experiments also show the model's effectiveness in providing forecasts for various forecasting horizons and the effectiveness and accuracy of the uncertainty estimates for the ARI task.
- 3) We comprehensively analyze the influence of meteorological variables in specific temperature and solar irradiations on net-load forecasting performance. The empirical results suggest that solar radiation is a strong predictor of net-load demand.
- 4) Lastly, we benchmark the proposed approach against statistical, conventional machine-learning, and DNN state-of-the-art forecasting approaches. The empirical results suggest that MLPQF provide a desirable trade-off between forecasting accuracies, calibrated uncertainty, and computation complexity.

The remainder of this paper is structured as follows. The background and related research work are presented in Section II. Section III details the proposed forecasting approach's components and implementation. The case-study specification,

including the dataset, input features, performance metrics, description of the four conducted experiments, and the performance evaluation methodology, is described in Section IV. The results obtained from each experiment are presented and discussed in Section V. Finally, Section VI concludes the paper by providing an overview of the main implications of this work for real-world smart-grid applications, discussing its limitations, and suggesting future directions for further research.

## II. RELATED WORKS

### A. Short-term Net-load Forecasting

Short-term net-load forecasting can be performed through either additive or aggregated methods, as noted in works such as [11], [18], [19]. The additive approach involves decomposing the net-load into its constituent parts and separately forecasting each component, as described in studies such as [20]–[22]. Conversely, aggregated methods rely solely on historical net-load data and other variables (e.g., meteorological data) as the input to the forecasting model. However, the practical implementation of the additive approach is often limited, as it necessitates the availability of real PV power data, which may be challenging to obtain or prohibitively expensive, as noted in [11].

The study in [18], [19] compared the efficacy of additive and integrated net-load forecasting methods and found that the integrated model outperforms the additive model. Still, the availability of limited public and free datasets for LV distribution network presents a challenge for researchers in the field. As a result, most of the existing research has been benchmarked on hypothetical LV networks generated from aggregated smart-meter data rather than actual datasets [7]

### B. Probabilistic DNN Forecasting Approaches

DNNs have become popular for power forecasting due to their ability to capture complex nonlinear trends and dependencies between multivariate time series [19], [23]. As a result, there has been a surge in the use of DNNs for power forecasting, with various DNN architectures being employed. These include Recurrent Neural Networks (RNNs) like Long-Short Term Memorys (LSTMs) and Gated Recurrent Neural Networks (GRNNs) [24], as well as transformer-based approaches [25]. Transformer-based architectures have been developed to improve the accuracy and robustness of RNNs based forecasting models and have shown promising results in various applications, including power forecasting. Yet similar to RNNs, these approaches are computationally expensive [26].

As a result, there has been a shift towards DNN forecasting architectures based on MLP like NBEATS, [27] and NHiTS [26], which are more efficient in terms of computation but still offer comparable performance to RNNs and transformer-based approaches. Computational complexity becomes essential when considering the real-world application of demand forecasting at the LV distribution substations where DSOs have to manage hundreds of substations [28]. It is thus vital to have

forecasting models that are fast to train and infer and require few computational resources.

However, despite the superiority of DNN-based approaches to forecasting problems, as reported in recent literature [24], [26], they often do not convey calibrated uncertainty estimates in their predictions compared to the traditional probabilistic model. To address this issue, various techniques, including probabilistic DNN, Bayesian Neural Network (BNN), and model ensembles, have been proposed.

The probabilistic DNN learn the aleatoric uncertainty by approximating the forecasting distribution through the use of parametric or non-parametric distributions, such as the Gaussian distribution, Gaussian Mixture density networks [6], Normalizing Flow [5], [6] and QR [17], [29]. On the other hand, BNN, such as variational inference [30], and dropout-based inference [31], learn a posterior distribution of DNN parameters that quantify epistemic uncertainty [12]. Finally, the model ensembles learn DNN uncertainty by training multiple DNN networks through bootstrapping or ensembling methods.

In this respect, the non-parametric QR based approach has recently received attention for probabilistic load forecasting as it models complex distributions without making any *a priori* assumptions on the data. Compared to other techniques, in most cases, QR produces well-calibrated uncertainty estimates in short-time horizons such as day-ahead [17], [29]. However, the capability of these models in delivering well-calibrated uncertainty estimates for multi-variate load and net-load forecasting in LV substation over extended time horizons remains an area that requires further investigation. This is particularly crucial in practical scenarios, as DSOs are required to make simultaneous predictions for multiple targets, encompassing different time horizons ranging from a few minutes to multiple days.

## III. PROPOSED METHODS

This section describes in detail the components of the proposed forecasting approach.

### A. Problem Definition

Short-term load forecasting aims to predict the near-future load  $\mathbf{y}_{t+1:H}$  over the forecast horizon  $H$  (ranging from a few hours to one week). The prediction is made based on input variables, which may include past observed net-load demands  $\mathbf{y}_{t-L:t}$  with a time lag of  $L$  and covariates features  $\mathbf{c}$  such as weather forecasts, temperature, and solar irradiance. Time-index features like holidays can also be included as covariates.

The input variables in this work are divided into two categories: historical or past features  $\mathbf{x}_{t-L:t}$ , and future covariates  $\mathbf{c}_{t+1:t+H}$ . Historical features refer to past data (including historical load demands  $\mathbf{y}_{t-L:t}$ , historical covariates  $\mathbf{c}_{t-L:t}$  and any other variables, which may be useful for understanding the underlying trends and patterns in the data. In contrast, future covariates  $\mathbf{c}_{t+1:t+H}$  are known in advance and may help make predictions of power demand  $\mathbf{y}_{t+1:t+H}$ . Thus our goal is to provide probabilistic forecast  $\mathbf{y}_H = \{y_{t+1}, \dots, y_{t+H}\}$  given the historical features (including the past targets)  $\mathbf{x}_L = \{x_{t-L}, x_{t-L+1} \dots x_t\}$  and future

covariates  $\mathbf{c}_H = \{c_{t+1}, c_{t+2} \dots c_{t+H}\}$  such that;  $\hat{\mathbf{y}}_H \sim p(\mathbf{y}_H | \mathbf{x}_L, \mathbf{c}_H)$

### B. Uncertainty Modelling with QR

We use a non-parametric QR to model the conditional distribution  $p(\mathbf{y}_H | \mathbf{x}_L, \mathbf{c}_T)$ . Specifically, we express it as:

$$p(\mathbf{y}_H | \mathbf{x}_L, \mathbf{c}_T) = Q_\theta(\hat{\tau}_\theta) \quad (1)$$

where  $\hat{\tau}_\theta \in [0, 1]$  is a set of  $N \times H$  quantile probabilities satisfying:

$$\hat{\tau}_{\theta t}^1 < \hat{\tau}_{\theta t}^2 < \dots < \hat{\tau}_{\theta t}^{N-1} < \hat{\tau}_{\theta t}^N \quad (2)$$

$Q_\theta(\hat{\tau}_\theta)$  is  $N \times H$  quantile functions.

Unlike most DNN for QR, which assume fixed quantile fractions, our approach parameterizes both quantile fractions and quantile functions with DNN. To this end, the quantile fractions and quantile function are represented by two FFNs: the Fraction Proposal Network (FPN) and Quantile Value Network (QVN) as illustrated in Fig. 8. The FPN learns to generate quantile fractions while the QVN maps these quantile fractions to the quantile function. This allows for end-to-end learning of the quantile fractions and the quantile function. Thus a non-parametric probabilistic density estimate  $p(\mathbf{y}_H | \mathbf{x}_L, \mathbf{c}_H)$  can be obtained by gathering a set of  $N$  quantile estimates such that:

$$p(y_t | \mathbf{x}_L, c_t) = \{Q_\theta(\hat{\tau}_{\theta t}^1), Q_\theta(\hat{\tau}_{\theta t}^2) \dots Q_\theta(\hat{\tau}_{\theta t}^N)\} \quad (3)$$

This allows the quantification of the uncertainty of a forecast using confidence region  $\Gamma^{(1-\alpha)} = [Q_\theta(\hat{\tau}_{\theta t}^L), Q_\theta(\hat{\tau}_{\theta t}^U)]$ , which gives a lower  $Q_\theta(\hat{\tau}_{\theta t}^L)$  and an upper bound  $Q_\theta(\hat{\tau}_{\theta t}^U)$  between which the predictions lie with a certain probability  $p_\tau = 1 - \alpha \in [0, 1]$  where  $\alpha$  is a small value that represents the level of uncertainty that is acceptable [32].

### C. MLPQF Model Architecture

The proposed MLPQF architecture is based on the work presented in [17]. In contrast to [17], the MLPQF is a lightweight architecture implemented solely using Multilayer Perceptron (MLP), with the capability to capture the sequence order of the time-series signal, which is vital in time-series forecasting [33], [34].

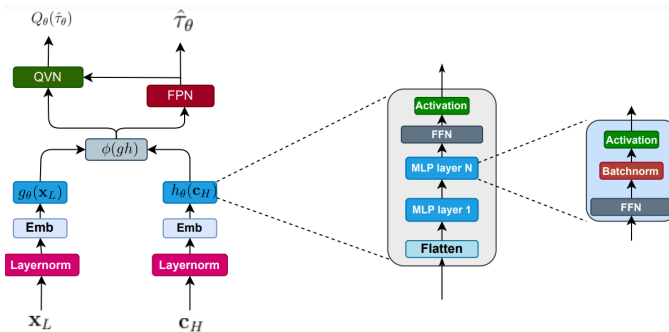


Fig. 2: The overall hierarchical architecture of MLPQF comprises two encoders, denoted as  $g$  and  $h$ . Each encoder consists of a flattening layer and lightweight MLP blocks.

As shown in Fig. 2, the proposed MLPQF is composed of various blocks constructed using FFN. The first block is the LayerNorm, which is a normalization step used to normalize the input features. The LayerNorm helps to reduce the internal covariate shift, which is the change in the distribution of the inputs to a layer that occurs during training. This helps to stabilize the training process and improve the performance of the MLPQF. The LayerNorm is followed by the input embedding block, which is used to enhance the FFN based encoders in capturing the sequence order of the time-series signal, which is vital in time-series forecasting [33], [34].

In our experimentation, we explored various strategies for positional embedding. Among these strategies, we found the Rotary Positional Embedding (RoPE) technique to be particularly effective, as detailed in Section V-A. The RoPE technique leverages a rotation matrix to convey absolute positional information, offering flexibility in accommodating different sequence lengths. This flexibility is crucial because it allows our MLPQF to learn context-dependent feature representations, ultimately enhancing its capability for long-term forecasting while effectively capturing underlying trends in the time-series data.

Following the input normalization and embedding stage, our architectural design includes two vital elements: the encoder for historical features, denoted as  $g(\mathbf{x}_L; \theta)$ , and the encoder for future covariates, represented as  $h(\mathbf{c}_H; \theta)$ . Our design choice was inspired by the research conducted by Chu et al. [35], which highlights the significant impact of covariate features, such as solar radiation, on predicting future net-load.

The two encoders comprised a flatten layer, a set of MLP block, Batchnorm, and a non-linear activation function as illustrated in Fig. 2. The sigmoid linear unit (silu) activation function defined as  $\text{silu}(x) = x \cdot \frac{1}{1+e^{-x}}$  is used as an activation function. The flattening layer reshapes multi-dimensional input data  $\mathbf{x}_L$  and  $\mathbf{c}_H$  into a one-dimensional feature representation, facilitating its subsequent processing by the MLP block. The  $g_\theta(\mathbf{x}_L)$  takes in the lagged inputs features  $\mathbf{x}_L$  to produce a feature representation  $\phi(g)$  while  $h(\mathbf{c}; \theta)$  takes in the time-varying a priori known future covariates  $\mathbf{c}_H$  and produces a feature representation  $\phi(h)$ . such that

$$\phi(g) = g_\theta(\mathbf{E}(\mathbf{x}_L)) \quad (4)$$

$$\phi(h) = h_\theta(\mathbf{E}(\mathbf{c}_H)) \quad (5)$$

$$\phi(gh) = \phi(g) + \phi(h) \quad (6)$$

where  $\mathbf{E}(\mathbf{x}; \theta) = \text{Dropout}[\text{Emb}(\text{LayerNorm}(\mathbf{x}))]$ .

For optimization, we employed the Adam optimizer with an initial learning rate of 0.001 for all methods. The learning rate was reduced by 0.1 when the number of iterations reached 75% and 90%, respectively.

Rather than utilizing the multi-head attention technique suggested in [17], we opted for merging the two feature representations to create a cohesive feature representation  $\phi(fh)$ . Unexpectedly, this uncomplicated method outperformed the application of multi-head attention as detailed in Section V-A. This result aligns with the findings in [25], which showed that intricate modules such as attention could have a detrimental effect on model performance in time-series forecasting.

Finally, the two encoders are succeeded by FPN and QVN block, which are responsible for converting the feature representation  $\phi(gh)$  into probabilistic forecasts. As explained in section Section III-B, the FPN's primary objective is to learn the quantile fractions  $\hat{\tau}_\theta$  that meet the requirements outlined in Eq. (4). To accomplish this, it employs an FPN a Dropout layer, and a Softmax activation function, as depicted in Fig. 3a. Thus the FPN receives the  $\phi(gh)$  feature as input and produces  $N + 1$  adjustable fractions  $\tau_t^{1:N+1}$  with  $\tau_t^1 = 0$  and  $\tau_t^{N+1} = 1$  such that:

$$\begin{aligned} \phi(\tau) &= \text{Dropout}[\text{FPN}(\phi(gh))] \\ \tau_{\theta t}^{1:N+1} &= [\tau_0, \text{csumSoftmax}(\phi(\tau))] \end{aligned}$$

where  $\tau_0$  is a vector of zeros. The  $\text{csumsoftmax}(x_k) = \sum_{i=1}^k \text{Softmax}(x_i)$  is cumulative sum of the Softmax output. This ensures that  $\tau_{\theta t}^{N+1} = 1$ . Finally the  $\hat{\tau}_{\theta t}^n$  estimate is obtained as:

$$\hat{\tau}_{\theta t}^n = \frac{\tau_{\theta t}^n + \tau_{\theta t}^{n+1}}{2} \quad (7)$$

The FPN layer's weight is initialized in such a way that the initial quantile probabilities follow a uniform distribution.

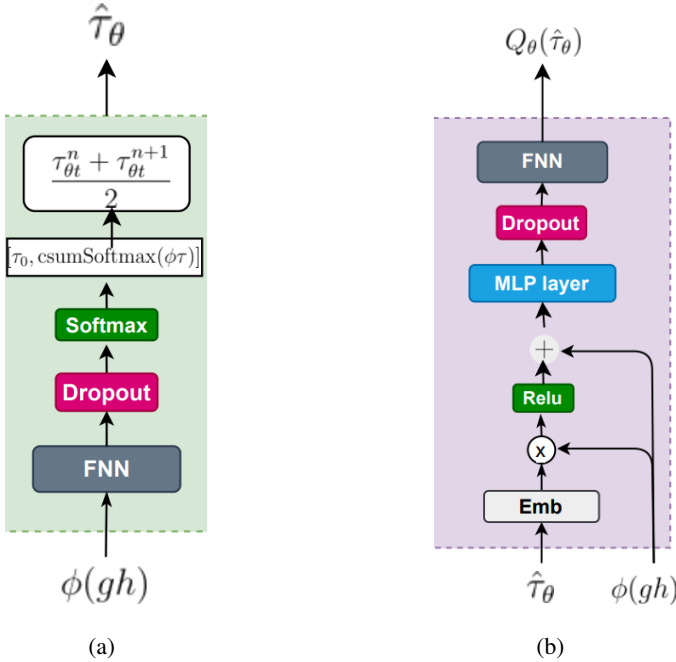


Fig. 3: FPN and QVN architectures (a) FPN (b) QVN.

After learning the quantile fractions, the estimate of the forecasted distribution is estimated with QVN (ref to Fig. 3b). The QVN takes in  $\hat{\tau}_\theta$  and  $\phi(fh)$  to produce  $Q_\theta(\hat{\tau}_\theta)$

$$\begin{aligned} \Phi(\hat{\tau}_\theta) &= \text{Relu}(\text{Emb}(\hat{\tau}_\theta)) \\ Q_\theta(\hat{\tau}_\theta) &= \text{Decoder}([\Phi(\hat{\tau}_\theta) \odot \phi(gh)] + \phi(gh)) \end{aligned}$$

The purpose of  $\text{RoPE}(\hat{\tau}_\theta)$  is to capture the interaction of the learned  $\hat{\tau}_\theta$ , which is then utilized to predict the quantile value  $Q_\theta(\hat{\tau}_\theta)$ . It's worth noting that the method used in this study for computing the embedding  $\Phi(\hat{\tau}_\theta)$  is different from the one employed in [17], as it doesn't require the use of additional hyper-parameters.

#### D. Loss Function and Training Procedures

The parameters of the MLPQF can be optimized by minimizing the pinball loss, as defined in Eq. (8):

$$\mathcal{L}_\tau(\epsilon_\tau) = \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \max[\epsilon_{\tau_t^n} \cdot \tau, (1 - \tau) \cdot \epsilon_{\tau_t^n}] \quad (8)$$

where  $\epsilon_\tau = y_t - Q_\theta(\hat{\tau}_{\theta t})$ . However, the pinball loss has a non-differentiable point at the origin, as reported in [36]. To address this issue, we adopt a smooth approximation of the pinball loss, using the Huber function, as defined in Section III-D:

$$\rho_\kappa(\epsilon_{\tau_t^n}) = \begin{cases} \frac{1}{2} \epsilon_{\tau_t^n}^2 & \text{if } |\epsilon_{\tau_t^n}| \geq \kappa \\ \kappa |\epsilon_{\tau_t^n}| - \frac{1}{2} \kappa & \text{if } \epsilon_{\tau_t^n} < 0, \end{cases}$$

This results in the definition of the Huber quantile loss, which is expressed in Eq. (9):

$$\mathcal{L}_\tau(\rho_\kappa(\epsilon)) = \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N |\tau_t^n - \mathbb{I}\{\epsilon_{\tau_t^n} < 0\}| \frac{\rho_\kappa(\epsilon_{\tau_t^n})}{\kappa} \quad (9)$$

This loss function enables the scaling of errors  $\epsilon_{\tau_t^n}$  that fall below a certain threshold  $\kappa$ , based on their magnitude. To further refine the sharpness of the predicted quantile values, we add a penalty term to the Huber quantile loss, as follows:

$$\mathcal{L}(\gamma, Q_\theta(\hat{\tau}_{\theta t})) = \frac{\beta}{T} \sum_{t=1}^T \sum_{n=1}^{N-1} \max[\gamma, Q_\theta(\hat{\tau}_{\theta t}^{n+1}) - Q_\theta(\hat{\tau}_{\theta t}^n)] \quad (10)$$

where  $\beta$  is a scalar value controlling the magnitude of the penalty term and  $\gamma$  regulates the sharpness of the quantiles.

While the Huber quantile loss optimizes the quantile function as specified in Eq. (9), we also incorporate the 1-Wasserstein metric in our learning process for the quantile fractions. The 1-Wasserstein metric is defined as:

$$W_1(Q_\theta(\hat{\tau}_{\theta t}), Q_\theta(\tau)) = \sum_{t=1}^T \sum_{n=0}^{N-1} \int_{\tau_t^n}^{\tau_t^{n+1}} |Q_\theta(\tau_t^n) - Q_\theta(\hat{\tau}_{\theta t}^n)| d\tau \quad (11)$$

and its derivative with respect to  $\tau$  is given by:

$$\frac{\partial W_1}{\partial \tau} = \sum_{t=1}^T 2Q_\theta(\tau_t^n) - Q_\theta(\hat{\tau}_{\theta t}^n) - Q_\theta(\hat{\tau}_{\theta t}^{n-1}) \quad (12)$$

As a result, the parameters  $\theta$  of the proposed MLPQF model are learned by jointly minimizing the following loss function using Adam optimizer with a batch size of 64,

$$\begin{aligned} \mathcal{J}_\theta(\mathbf{y}, Q_\theta(\hat{\tau}_{\theta t})) &= \frac{1}{K} \sum_{i=1}^K \mathcal{L}_\tau(\rho_\kappa(\epsilon))^i + \mathcal{L}(\gamma, Q_\theta(\hat{\tau}_{\theta t}))^i \\ &\quad + W_1(Q_\theta(\hat{\tau}_{\theta t}), Q_\theta(\tau))^i \end{aligned}$$

where  $K$  is the number of data samples in the training set. The initial learning rate was set to  $1e^{-3}$  and was later decreased by 0.1 when the number of iterations reached 75% and 90%, respectively. The values of  $\beta$ ,  $\gamma$ , and  $\kappa$  were set to  $3.6e^{-4}$ ,  $3e^{-6}$ , and 0.5, respectively.

#### IV. CASE STUDY SPECIFICATION

The proposed approach is evaluated on two real-life substation datasets: 1) the MLVS-PT [17], and 2) the SPS-UK [37].

The MLVS-PT data is sourced from a LV distribution substation located on a geographically isolated island in Southern Europe serving around 100 consumers. The substation has a 250 kVA transformer and 36 kWp of PV generation. The net-load demand is measured at one-minute intervals and covers the period from March 2019 to June 2022. MLVS-PT also includes meteorological data at different sample rates from 5 to 60-minute intervals, obtained from SolCast [38]. For this study, the net-load and meteorological data were resampled to a resolution of 30 minutes.

The SPS-UK encompasses distribution network demand, PV generation, and weather information for a primary substation located in Plymouth, UK. The load demand and PV power generation, recorded in MW, were measured at 30-minute intervals between November 2017 and July 2020. The net-load was calculated by subtracting the PV generation from the load demand. The weather data were interpolated linearly to match the 30-minute resolution of the PV and load demand.

##### A. Input Features

The input to the model consists of historical demand (net-load) and covariates, both past and future. The covariates features include temperature (Temp), global horizontal irradiance (Ghi), and time-based features that account for factors such as weather conditions, natural cycles, and calendar days, as described in [17], [39]. In line with the approach of [17], the date-time related features are transformed using sinusoidal and cosine transformations to capture daily and yearly patterns such that  $\mathbf{ET}_{i,t} = \left[ \sin\left(\frac{2\pi t_i}{T_{si}}\right), \cos\left(\frac{2\pi t_i}{T_{si}}\right) \right]$  where  $T_{si}$  is the period for each of the exogenous features. For instance, the day of the week has a periodicity of 7, whereas the hour of a day has a periodicity of 24. The following time-based features were selected based on a correlation analysis: hour of the day, session of the day (night or day), day of the week, and day of the month. The original time-based features and their corresponding sine and cosine transformations were analyzed to inform the selection.

##### B. Performance Metrics

We adopt Normalized Root Mean Squared Error (NRMSE) to assess the forecasting performance. NRMSE defined as  $\frac{1}{P} \sqrt{\sum_{t=1}^{t=T} \frac{(\hat{y}_t - y_t)^2}{T}}$  is the forecast metric that provides a percentage of how high the mean squared error is compared to the installed capacity ( $P$ ) [17].

In addition, we follow [17] and use the combined coverage, width, and NRMSE scoring metric (CWE), which assess both the forecasting accuracy and predictive uncertainty of the probabilistic forecasting model. The CWE is the geometric harmonic mean of the  $\gamma_{pcip} \in [0, 1]$  a coverage-based score and  $\gamma_{nmpi} \in [0, 1]$  an interval-width score defined as per Eq. (13):

$$\text{CWE} = 2 \cdot (1 - \text{NRMSE}) \cdot \frac{\gamma_{nmpi} \cdot \gamma_{pcip}}{\gamma_{pcip} + \gamma_{nmpi}} \quad (13)$$

##### C. Evaluation Procedure

The proposed method is evaluated using back-testing cross-validation. This approach simulates making predictions for future time steps while considering the temporal dependencies present in the data, as reported in [40]–[42]. The process involves training the model on the complete historical data using a sliding window that progresses through the time series. At each step, a fixed segment of the future time series is designated as the testing set, immediately following the end of the training set thereby ensuring that the model is never tested on data that preceded the training data [41], [42]. More precisely, we adopt 10-fold back-testing cross-validation with an expanding window, in which the initial historical period is set to at least 12 months and the fixed future time window is set to 4 months. The sliding window is extended by two months as we move forward in the time series. Each model is trained on 90% of the training window for 100 iterations and validated on the remaining 10%.

TABLE I: Hyper-parameters used across the three experiments.

Parameter	Layers	Dropout	Emb size	N	Neurons
Value	2	0.25	16	90	256

##### D. Research Questions

To evaluate the proposed approach, we framed four research questions. Next, the different RQs and respective experimental procedures are described.

- RQ1 How do various model design choices impact the performance of the proposed MLPQF algorithm?
- RQ2 What is the influence of meteorological variables on net-load forecasting?
- RQ3 What is the effectiveness and accuracy of uncertainty estimates provided by the MLPQF for ARI net-load forecasting task?
- RQ4 Can a lightweight complex MLPQF algorithm rival state-of-the-art machine learning algorithms, such as DNN, in terms of forecasting performance and computational efficiency?

To address RQ1, we investigated the performance of the MLPQF model in various scenarios, including different forecasting horizons, lagged windows, the use of positional embedding, and the attention mechanism. Initially, we evaluated the performance of the proposed MLPQF model for forecasting horizons ranging from 1 hour to 7 days, representing very short to medium-term forecasting horizons as discussed in [43]. Additionally, we analyzed the impact of the length of the lagged (context) window denoted by  $L$  on the accuracy of future predictions, with window lengths varying from 1 day to 7 days. To conduct this analysis, we employed a backtesting cross-validation strategy, where each model underwent training on the SPS-UK dataset for approximately 50 iterations.

For RQ2, we examined the impact of covariate features on net-load forecasting performance. Specifically, we considered temperature (Temp), Global Horizontal Irradiance (Ghi), and

time-based features, as they were readily available in the SPS-UK and MLVS-PT datasets. As a result, four forecasting models were established, each incorporating a unique combination of covariate features, including [Ghi, Time], [Temp, Time], [Ghi, Temp, Time], and [Time]. These models were trained for a limited number of iterations, specifically 50 iterations, to assess their performance.

Regarding RQ3, we examined the robustness of the proposed proposed MLPQF in forecasting the ARI [44]. The ARI is a widely used forecasting method in real-time applications, where the model’s output is used to predict future time horizons based on the initial time horizon. An incorrect or improperly calibrated initial model can propagate errors to subsequent predictions [44]. To address this, the first step was to train a model on the three datasets using a back-testing cross-validation strategy with an expanding window, as described in Section IV-C. The resulting model was then evaluated using the ARI to produce  $k$ -day ahead forecasts, where  $k \in 2, 3, 4, 5, 6, 7$  days ahead [42].

Finally, to address RQ4, we conducted a comprehensive comparison of the performance of MLPQF against six well-established and state-of-the-art forecasting models using two datasets MLVS-PT and SPS-UK. To this end, two statistical models were selected: a naive seasonal (S-Naive) [40] and Multiple Seasonal-Trend decomposition using LOESS (MTL) [45] model. Along with these statistical models, two popular machine learning models, CatBoost (CAT) [46] and Random-Forest (RF) [47], were also included in the comparison.

Additionally, four state-of-the-art DNN-based forecasting models were considered, including N-BEATS [27], N-HITS [26], LSTM and Full Parameterized Sequence to Quantile (FPQ) [17]. FPQ-MLP is the genetic DNN architecture with full-parameterised QR forecasting model introduced in [17]. For a fair comparison with the proposed approach, we use an MLP encoder and name it FPQ-MLP. N-BEATS and N-HITS are simple architectures based on MLP and have shown competitive performance at lower computational costs. The LSTM has been defacto DNN architecture for power demand forecasting [48]. To mitigate any potential bias between the models, efforts were made to ensure their similarity. This was accomplished by constraining the number of parameters to comparable levels across all models. Further information on the model configurations can be found in Table I.

## V. RESULT AND DISCUSSION

### A. RQ1: MLPQF Model Analysis

The results presented in Fig. 4a show the forecasting performance of various forecast horizons with different context windows. The findings reveal that the MLPQF model is proficient in providing accurate short-term forecasts, ranging from 1-hour to 7-day horizons, with an average NRMSE score between 0.02 to 0.08. Additionally, the results show a slight decline in NRMSE score as the forecasting horizon increases from 1 day to 7 days. However, the MLPQF model remains competitive with a score below 0.1, which is equivalent to a forecasting accuracy of 90%. Remarkably, outliers exhibiting an NRMSE score falling between 0.2 and 0.5 were exclusively

detected in cases of extremely short forecasting horizons, spanning from just one hour to a single day. These findings suggest that forecasting horizons of at least one day ahead provide a favorable bias-variance trade-off, in contrast to very short forecasting horizons.

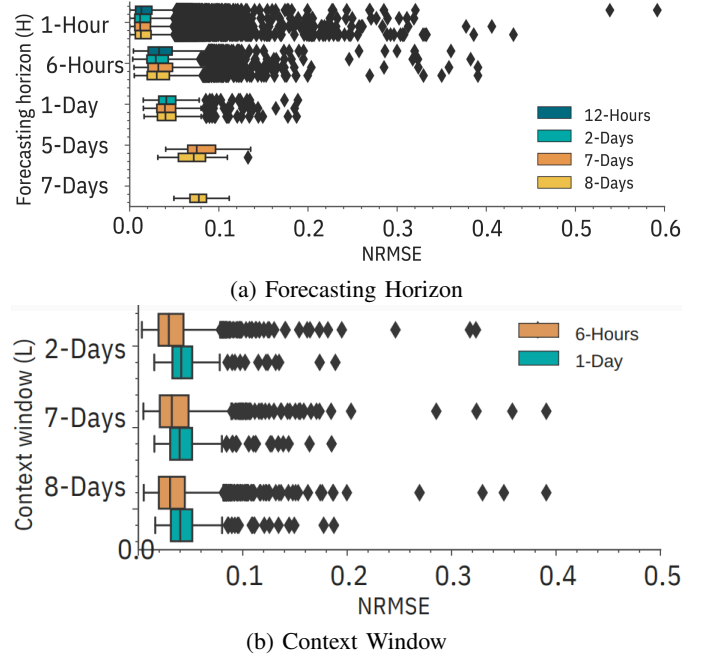


Fig. 4: Results of MLPQF for different forecasting horizons and lagging window (RQ1).

In terms of the context window, the study revealed that a context window  $L$  that is equal to or slightly larger than the forecasting horizon  $H$  is adequate in achieving accurate forecasting performance. This was evidenced by the comparable performance of the historical window when forecasting for a specific horizon, as illustrated in Fig. 4b. This aligns with the results in Fig. 4a, which suggest that a historical window of less than a day is particularly effective for very short-term forecasting (one hour to six hours). Thus, small context windows are adequate for accurate short-term forecasting. As a result, the subsequent experiments were carried out using 96 data points as input and forecasting 48 steps. This configuration enabled the models to predict the upcoming day’s net-load based on the data from the two preceding days.

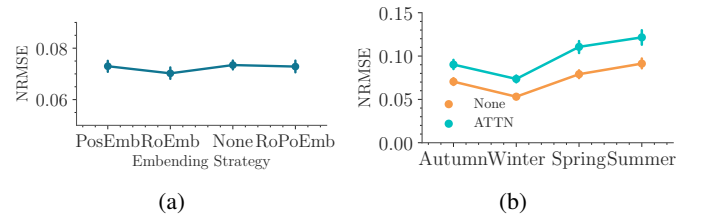


Fig. 5: Impact of positional encoding and attention mechanism in the MLPQF design (RQ1).

We also analyze the effectiveness of the positional embedding and multi-head attention building blocks. As shown in

Fig. 5a, embedding leads to improved forecasting compared to the absence of embedding (None), with Rotary Embedding achieving a higher score. However, we were surprised that combining different embedding strategies did not result in better performance. Finally, we evaluated the efficacy of two different techniques for combining feature representations: the multi-head attention method proposed in [17] and a simple addition operation (represented as "None" in the graph). As shown in Fig. 5b, the latter method outperformed the former, indicating that using more complex modules like attention may not always enhance model performance and can introduce unnecessary complexity.

### B. RQ2: Influence of Meteorological Variables on Net-load Forecasting

The obtained results for this experiment are presented in Figs. 6a and 6b. It is observed that Ghi exhibits a strong

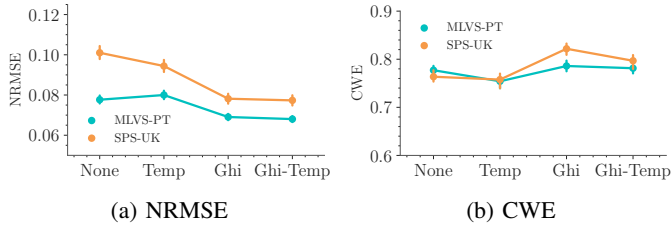


Fig. 6: Results of MLPQF for different covariates variable: (Experiment 1).

predictive capacity for net-load forecasting in both the MLVS-PT and SPS-UK datasets, as evidenced by average CWE scores of  $0.75\pm 0.15$  and  $0.77\pm 0.17$ , respectively. On the other hand, Temp accounts for an average CWE score of 0.72 on the two datasets.

Unexpectedly, the combination of Ghi and Temp does not enhance forecasting accuracy, with both datasets achieving an almost identical average CWE score of 0.76. These findings imply that Ghi could be a more effective predictor for net-load forecasting. We also observe that the forecasting performance is penalized when only time-derived features are used as inputs (None in the graph).

### C. RQ3: Robustness of ARI Forecasting

The obtained results for the second experiment are summarised in Fig. 7. We observe that the proposed MLPQF achieves competitive Normalised Root Mean Squared Error (NRMSE) ( $< 0.083$ ) and CWE ( $> 0.65$ ) on average for ARI with different days ahead up to one week. Still, a minimal score deterioration as the number of days increases can be observed from the results. The obtained result implies that the MLPQF is capable of learning a well-calibrated day-ahead forecast that could be leveraged for ARI as visualized in Figs. 7a and 7b.

### D. RQ4: Benchmark Against Baselines

The results of the benchmark experiment are shown in Fig. 8a. The performance of our MLPQF model was found

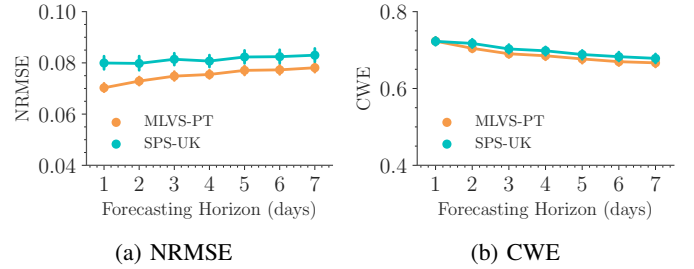


Fig. 7: ARI results for MLPQF (Experiment 2).

to be state-of-the-art, with an NRMSE score of  $0.07\pm 0.03$  and  $0.08\pm 0.04$  for the MLVS-PT and SPS-UK datasets, respectively. The FPQ-MLP model also exhibited similar performance, scoring  $0.07\pm 0.02$  and  $0.09\pm 0.04$  on MLVS-PT and SPS-UK datasets. These findings suggest that the modifications made to the MLPQF architecture did not compromise the forecasting accuracy. This represents a significant improvement of around 12% over the LSTM model for the MLVS-PT dataset and a gain of more than 30% over the other baselines. For the SPS-UK dataset, our MLPQF model showed an improvement of around 20% compared to the LSTM and CAT models and more than 30% compared to the other baselines.

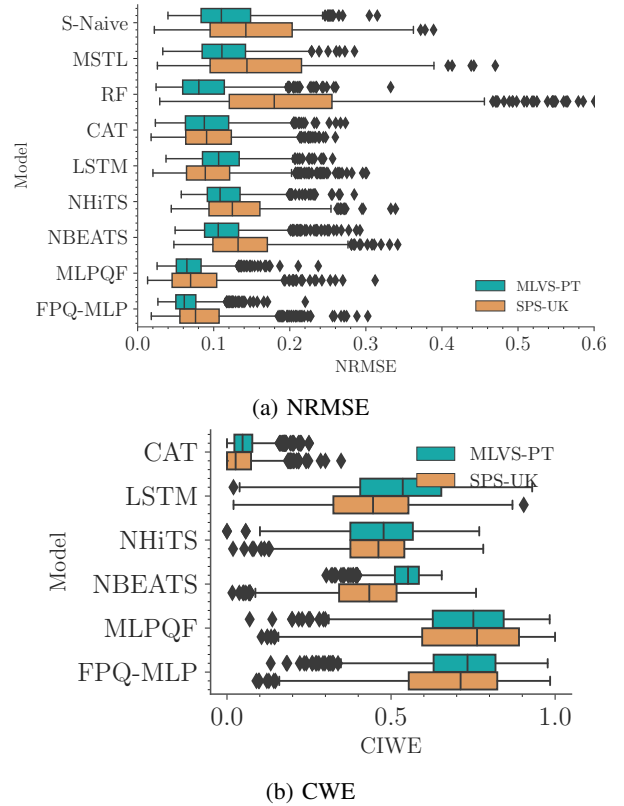


Fig. 8: Forecasting ability of the MLPQF against the alternatives (Experiment 4).

Upon analyzing the results presented in Fig. 8b, we found that our proposed model outperforms other probabilistic models by exhibiting a CWE score of  $0.73\pm 0.15$  for both datasets.



This is slightly higher compared to the FPQ-MLP’s score of  $0.71 \pm 0.15$  (2%  $\uparrow$ ) on MLVS-PT and  $0.68 \pm 0.18$  (6%  $\uparrow$ ) on SPS-UK dataset. Compared to other models, we see an improvement of about 21%. These results suggest that the proposed lightweight architecture not only maintains forecasting accuracy but also improves probabilistic performance. From Figs. 9a and 9b, we see that the net-load forecasting

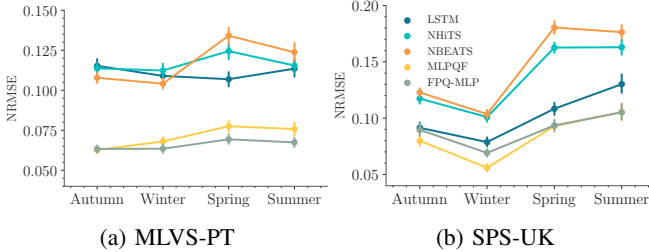


Fig. 9: Forecasting ability of the MLPQF against the alternatives: Seasonal variations.

performance varies across seasons, with the lowest NRMSE in winter and autumn, in contrast to summer and spring. This variation may be due to the seasonal variability in power demand and generation patterns. Specifically, summer and spring demand patterns may be influenced by unpredictable factors, such as weather fluctuations or changes in consumer behavior during holidays or vacations, making it more difficult to accurately predict demand compared to the more stable demand patterns observed in winter and autumn. This highlights the need to consider seasonality in the development and evaluation of electricity demand forecasting models.

To thoroughly assess the computational efficiency of our proposed approach, we conducted a thorough comparison against eight different baselines. Our evaluation considered both computational training and inference time, as illustrated in Fig. 10. The results (Fig. 10a) clearly show that the S-Naive and MSTL statistical baselines required the least amount of training time, followed by conventional machine learning models RF and CAT when compared to our MLPQF approach. However, MLPQF is an eager learner, meaning that a model is created during the training phase, which takes more time to train than statistical and conventional machine learning baselines. Nonetheless, our method is still faster to train compared to LSTM (82.4%  $\downarrow$ ) and FPQ-MLP (51.4%  $\downarrow$ ). Furthermore, its training speed is slightly higher than that of NHITS (13.4%  $\downarrow$ ) and NBEATS (18.8%  $\downarrow$ ).

On the other hand, MLPQF demonstrated a much shorter inference time comparable to S-Naive (refer to Fig. 10b). Specifically, it only required 0.064 seconds to forecast a 4-month data horizon, in contrast to FPQ-MLP, which necessitated 0.76s (91.6%  $\downarrow$ ). Additionally, we observed that the inference speed of MLPQF surpassed that of MSTL (74.0%  $\downarrow$ ) and RF (94.0%  $\downarrow$ ) models. Notably, it outperformed state-of-the-art DNN models such as NBEATS and NHITS by reducing inference times by a remarkable 98.5%. Furthermore, it exhibited over 300 times faster inference than LSTM, highlighting its superior efficiency.

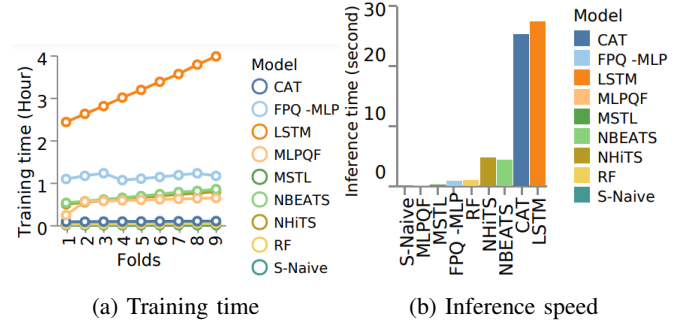


Fig. 10: Computational performance of the MLPQF against the four alternatives (Experiment 4).

## VI. CONCLUSION

This paper presented a MLPQF approach, which constitutes a novel and scalable net-load probabilistic forecast at the LV secondary substations with high PV generation. The proposed architecture leverages simple but effective MLP to effectively capture non-linear and complex relationships between historical variables and future covariates, thereby enabling accurate forecasting of the net-load of LV substations. Through empirical experimentation on two real-world datasets, we have demonstrated the efficacy of the proposed architecture for net-load forecasting at a LV substation distribution level.

By utilizing NRMSE as a measure of forecasting accuracy, it is shown that the proposed approach achieves a consistent and accurate predictive performance. Specifically, our model achieves a low NRMSE of approximately 0.07 and 0.08 on the two datasets, showcasing its superior forecasting capability in both day-ahead and ARI forecasting. This represents a substantial improvement, ranging between 12% to 30% when compared to baseline models. In addition to forecasting accuracy, the model’s complexity was evaluated by measuring the execution time for both training and inference. Notably, the proposed approach approximates the speed of a seasonal naive approach, being 300 times faster compared to state-of-the-art benchmarks such as LSTM.

Ultimately, the obtained results validate that uncomplicated architectures, such as those based on MLP, can yield competitive performance while keeping computation requirements low, unlike more intricate DNN models such as transformers. Additionally, the study illustrates that it is feasible to use ARI to generate forecasts for varying timeframes without requiring the training of supplementary models.

However, a limitation of this work is the geographical bias inherent in the datasets used for evaluation. Both datasets were collected from the same European region, which might hinder the generalizability of our proposed MLPQF architecture to other geographic locations characterized by distinct energy consumption patterns. To address these limitations, future research should focus on evaluating the performance of MLPQF using additional LV distribution substation datasets with diverse geographical locations and installed capacities. In this respect, it is important to stress the need for continued efforts in deploying monitoring technology at the LV level and

sharing other real-world datasets with the public in order to continually expand the research in this area.

This work has also highlighted the crucial role of covariate features in the performance of the power load forecasting model. Specifically, we have shown that solar radiation (Ghi) is an essential covariate feature that significantly enhances the accuracy of net-load forecasts. This result suggests that solar irradiance might be a crucial predictor for net-load as it explains the fluctuations in PV production. Consequently, to enhance net-load forecasting, utilities should consider integrating solar irradiation measurements into their forecasts instead of relying only on temperature. However, it should be stressed that this observation requires further investigation as the presented analysis was limited to the two datasets. A comprehensive study encompassing a wider range of datasets and diversifying covariate factors is essential to establish the generalizability of these observations.

#### ACKNOWLEDGMENT

This work was supported by Eaton Corporation’s Center for Intelligent Power (CIP), Dublin, Ireland (Anthony Faustine). The authors would also like to acknowledge the Portuguese Foundation for Science and Technology (FCT) for their support through projects 10.54499/LA/P/0083/2020; 10.54499/UIIDP/50009/2020 & 10.54499/UIIDB/50009/2020, and grant CEECIND/01179/2017 (Lucas Pereira).

#### REFERENCES

- [1] A. Azizivahed, A. Arefi, S. Ghavidel, M. Shafie-khah, L. Li, J. Zhang, and J. P. S. Catalao, “Energy management strategy in dynamic distribution network reconfiguration considering renewable energy resources and storage,” *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 662–673, Apr. 2020. [Online]. Available: <https://doi.org/10.1109/tste.2019.2901429>
- [2] 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.
- [3] S. Haben, M. Voss, and W. Holderbaum, *Core Concepts and Methods in Load Forecasting: With Applications in Distribution Networks*. Springer Nature, 2023.
- [4] Y. Wang, G. Hug, Z. Liu, and N. Zhang, “Modeling load forecast uncertainty using generative adversarial networks,” *Electric Power Systems Research*, vol. 189, p. 106732, Dec. 2020. [Online]. Available: <https://doi.org/10.1016/j.epr.2020.106732>
- [5] M. Arpogaus, M. Voss, B. Sick, M. Nigge-Uricher, and O. Dürr, “Short-term density forecasting of low-voltage load using bernstein-polynomial normalizing flows,” *arXiv preprint arXiv:2204.13939*, 2022.
- [6] A. Jamgochian, D. Wu, K. Menda, S. Jung, and M. J. Kochenderfer, “Conditional approximate normalizing flows for joint multi-step probabilistic electricity demand forecasting,” *arXiv preprint arXiv:2201.02753*, 2022.
- [7] C. Gilbert, J. Browell, and B. Stephen, “Probabilistic load forecasting for the low voltage network: forecast fusion and daily peaks,” *Sustainable Energy, Grids and Networks*, p. 100998, 2023.
- [8] C. J. Bennett, R. A. Stewart, and J. W. Lu, “Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system,” *Energy*, vol. 67, pp. 200–212, Apr. 2014.
- [9] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, “Short term load forecasting and the effect of temperature at the low voltage level,” *International Journal of Forecasting*, vol. 35, no. 4, pp. 1469–1484, Oct. 2019.
- [10] M. Dong and L. Grumbach, “A Hybrid Distribution Feeder Long-Term Load Forecasting Method Based on Sequence Prediction,” *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 470–482, Jan. 2020.
- [11] O. Rubasinghe, T. Zhang, X. Zhang, S. S. Choi, T. K. Chau, Y. Chow, T. Fernando, and H. H.-C. Iu, “Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high pv penetration,” *Applied Energy*, vol. 333, p. 120641, Mar. 2023. [Online]. Available: <http://dx.doi.org/10.1016/j.apenergy.2023.120641>
- [12] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, “Using bayesian deep learning to capture uncertainty for residential net load forecasting,” *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188–201, 2019.
- [13] C. O’Dwyer and D. Flynn, “Using energy storage to manage high net load variability at sub-hourly time-scales,” *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 2139–2148, 2015.
- [14] X. Sun and C. Jin, “Impacts of solar penetration on short-term net load forecasting at the distribution level,” in *2021 IEEE 4th International Electrical and Energy Conference (CIEEC)*, 2021, pp. 1–6.
- [15] S. Haben, S. Arora, G. Giasemidis, M. Voss, and D. Vukadinović Greetham, “Review of low voltage load forecasting: Methods, applications, and recommendations,” *Applied Energy*, vol. 304, p. 117798, 2021.
- [16] 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.
- [17] A. Faustine and L. Pereira, “Fpseq2q: Fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates,” *IEEE Transactions on Smart Grid*, pp. 1–1, 2022.
- [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] M. Beichter, K. Phipps, M. M. Frysztacki, R. Mikut, V. Hagenmeyer, and N. Ludwig, “Net load forecasting using different aggregation levels,” *Energy Informatics*, vol. 5, no. 1, pp. 1–21, 2022.
- [20] 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, 2017.
- [21] M. Toro-Cárdenas, I. Moreira, H. Morais, P. M. Carvalho, and L. A. Ferreira, “Net load disaggregation at secondary substation level,” *Renewable Energy*, 2022.
- [22] S. Sree Kumar, K. C. Sharma, and R. Bhakar, “Gumbel copula based aggregated net load forecasting for modern power systems,” *IET Generation, Transmission & Distribution*, vol. 12, no. 19, pp. 4348–4358, 2018.
- [23] P. C. Huy, N. Q. Minh, N. D. Tien, and T. T. Q. Anh, “Short-term electricity load forecasting based on temporal fusion transformer model,” *IEEE Access*, vol. 10, pp. 106 296–106 304, 2022.
- [24] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, “Probabilistic individual load forecasting using pinball loss guided lstm,” *Applied Energy*, vol. 235, pp. 10–20, 2019.
- [25] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?” *arXiv preprint arXiv:2205.13504*, 2022.
- [26] C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler, and A. Dubrawski, “N-hits: Neural hierarchical interpolation for time series forecasting,” *arXiv preprint arXiv:2201.12886*, 2022.
- [27] B. N. Oreshkin, D. Carпов, N. Chapados, and Y. Bengio, “N-beats: Neural basis expansion analysis for interpretable time series forecasting,” *arXiv preprint arXiv:1905.10437*, 2019.
- [28] N. Andreadou, M. G. Flammini, G. Fulli, M. Masera, G. Pretticco, and S. Vitiello, “Distribution system operators observatory 2018: Overview of the electricity distribution system in Europe.”
- [29] 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, 2021.
- [30] H. Wen, J. Ma, J. Gu, L. Yuan, and Z. Jin, “Sparse variational gaussian process based day-ahead probabilistic wind power forecasting,” *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 957–970, 2022.
- [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)*. IEEE, 2017, pp. 103–110.
- [32] 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.
- [33] J. Su, Y. Lu, S. Pan, B. Wen, and Y. Liu, “Roformer: Enhanced transformer with rotary position embedding,” *CoRR*, vol. abs/2104.09864, 2021. [Online]. Available: <https://arxiv.org/abs/2104.09864>
- [34] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting,”

*CoRR*, vol. abs/2106.13008, 2021. [Online]. Available: <https://arxiv.org/abs/2106.13008>

- [35] Y. Chu, H. T. Pedro, A. Kaur, J. Kleissl, and C. F. Coimbra, “Net load forecasts for solar-integrated operational grid feeders,” *Solar Energy*, vol. 158, pp. 236–246, 2017.
- [36] 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>
- [37] E. Borghini, C. Giannetti, J. Flynn, and G. Todeschini, “Data-driven energy storage scheduling to minimise peak demand on distribution systems with pv generation,” *Energies*, vol. 14, no. 12, 2021. [Online]. Available: <https://www.mdpi.com/1996-1073/14/12/3453>
- [38] J. M. Bright, “Solcast: Validation of a satellite-derived solar irradiance dataset,” vol. 189, pp. 435–449.
- [39] 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 and Energy Systems*, vol. 117, p. 105635, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0142061518335257>
- [40] H. Hewamalage, K. Ackermann, and C. Bergmeir, “Forecast evaluation for data scientists: common pitfalls and best practices,” *Data Min Knowl Disc*, Dec. 2022. [Online]. Available: <https://link.springer.com/10.1007/s10618-022-00894-5>
- [41] V. Cerqueira, L. Torgo, and I. Mozetič, “Evaluating time series forecasting models: an empirical study on performance estimation methods,” *Machine Learning*, vol. 109, no. 11, pp. 1997–2028, Oct. 2020. [Online]. Available: <https://doi.org/10.1007/s10994-020-05910-7>
- [42] H. Hewamalage, K. Ackermann, and C. Bergmeir, “Forecast evaluation for data scientists: Common pitfalls and best practices,” 2022. [Online]. Available: <https://arxiv.org/abs/2203.10716>
- [43] T. Hong and S. Fan, “Probabilistic electric load forecasting: A tutorial review,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207015001508>
- [44] H. Kamarthi, L. Kong, A. Rodriguez, C. Zhang, and B. A. Prakash, “When in doubt: Neural non-parametric uncertainty quantification for epidemic forecasting,” in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, K. Nguyen, P. S. Liang, J. W. Vaughan, and Y. Dauphin, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 19796–19807. [Online]. Available: <https://proceedings.neurips.cc/paper/2021/file/a4a1108bbcc329a70efa93d7bf060914-Paper.pdf>
- [45] K. Bandara, R. J. Hyndman, and C. Bergmeir, “Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns,” *arXiv preprint arXiv:2107.13462*, 2021.
- [46] A. V. Dorogush, V. Ershov, and A. Gulin, “Catboost: gradient boosting with categorical features support,” *arXiv preprint arXiv:1810.11363*, 2018.
- [47] 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.
- [48] 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.



**Nuno Jardim Nunes** is a Full professor at Técnico U. Lisbon and the President of the Interaction Technologies Institute (ITI) a research unit of the LARSYS Associated Laboratory. He’s also co-director of the [www.cmuportugal.org](http://www.cmuportugal.org) Carnegie Mellon International partnership and affiliated faculty at the HCII at CMU. Nuno’s research interests lie in applying models to software, systems, and service design for environmental sustainability and participatory culture. Nuno is a strong advocate of the role of design in engineering.



**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 (FEELab). Since 2019, he has been 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 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.

energy systems and sustainable built environments, and it typically involves the real-world deployment and evaluation of monitoring technologies and software systems.



**Anthony Faustine** is a lead data scientist at the Center of Intelligent Power (CIP) at Eaton’s global headquarters in Dublin, where he leads, designs and develops of cutting-edge AI and data analytics-driven solutions tailored to address emerging challenges within the industrial and energy sectors. His research interest is in robust machine learning algorithms for future energy systems, focusing on load forecasting, load disaggregation, and data-driven optimization. He is currently finalizing a Ph.D. Degree in Computer Science Engineering at Further Energy and

Environment Research Laboratory (FEELab), ITI/LARSYS, Técnico Lisboa in Portugal.