Conformal Multilayer Perceptron-Based Probabilistic Net-Load Forecasting for Low-Voltage Distribution Systems with Photovoltaic Generation

Anthony Faustine*† and Lucas Pereira†
*Center for Intelligent Power (CIP), Eaton Corporation, Dublin, Ireland; †ITI, LARSyS, Técnico Lisboa, Lisbon, Portugal
Email: anthonyafaustine@eaton.com, lucas.pereira@tecnico.ulisboa.pt

***Abstract*—Probabilistic net-load forecasting in Low-Voltage (LV) distribution networks is essential in light of the increased variability introduced by the widespread integration of renewable energy sources (RES). Various probabilistic approaches based on neural networks have been proposed to solve this challenge. This study introduces lightweight neural network-based conformal prediction (Conformal-MLPF) for net-load forecasting within an LV power distribution network. It uses Split Conformal prediction to transform a lightweight MLP-based point forecast into a probabilistic forecast. Our validation on two real-life LV substations datasets suggests that the proposed Conformal-MLPF achieves a better tradeoff between forecasting performance and model complexity without requiring restrictive assumptions about data distribution.**

***Index Terms*—Conformal prediction, Uncertainty Estimation, Low Voltage Distribution Network, Neural Network, Net-Load, Probabilistic Forecasting, Quantile Regression**

I. Introduction

The ongoing energy transition in Low Voltage (LV) networks, which are characterized by the massive integration of Renewable Energy Sources (RES) such as Photovoltaics (PV) and wind power resulting in growing complexity in operating and managing the grid [1], [2]. For instance, a prevalence of RESs in LV networks will result in bidirectional power flow, which can cause voltage fluctuations and uncertainty in LV network load profiles [2]–[4]. Additionally, with more households adopting PV systems [3], [5], it becomes more challenging for Distribution System Operator (DSO)s to manage the LV network effectively due to reduced visibility of the actual power demand [6], [7]. In this scenario, accurate forecasting of demand and generation is vital to optimizing the performance of the LV distribution network [8], [9].

Probabilistic forecasting methods have gained popularity in predicting net-load, which combines load demands and generation forecasts. These methods provide valuable insights into uncertainty and prediction intervals, enabling grid operators to develop risk-based strategies for effectively operating and managing the LV network. Among these methods, Deep Neural Networks (DNN)s have shown particular promise for net-load forecasting in LV power distribution networks. Researchers have explored several probabilistic DNN approaches, including Normalizing Flow [10], Gaussian distributions [6], and Quantile Regression (QR) [7], [11]. Other researchers use Bayesian Neural Networks (BNNs) [12], which produce probabilistic forecasts by treating the DNN model weights as random variables with a corresponding probability distribution.

Two main methodologies are employed in producing short-term net-load forecasting: disaggregated (additive) and aggregated (integrated). The disaggregated method involves the decomposition and prediction of the net-load in terms of its constituent parts, such as load demand and PV generation. Subsequently, the individual forecasts are merged to estimate the net-load [3], [8], [9], [13]. On the contrary, integrated methods exclusively rely on historical net-load data and other exogenous variables that impact RES and power demands to predict the future net-load. These forecasting methods assume that changes in the net-load are closely related to variations in RES generation and power demands [4]. Numerous studies in recent literature have proposed various integrated forecasting methods, as evidenced by research such as those mentioned in [6], [8], [9], [11], [12].

While several works on probabilistic net-load forecasting have been proposed, the effectiveness of Conformal Prediction (CP) in net-load forecasting remains unexplored. CP is a distribution-free uncertainty estimation method that constructs valid prediction intervals [14]. These intervals provide a range of potential values within which the forecasted data may fall, along with a specific confidence level. Importantly, CP achieves this without making any assumptions about the data distribution or the model [15], [16]. Recent applications of CP have been observed in both load forecasting and PV forecasting [15], [17]. However, there is limited research exploring the effectiveness of CP specifically for net-load forecasting, which combines both PV and load forecasting.

This paper addresses this gap by investigating CP 's effectiveness in net-load probabilistic forecasting, comparing it with established methods. Adopting an integrated approach, we introduce Conformalised Multilayer Perceptron based Forecast (Conformal-MLPF), a lightweight neural network model for accurate point forecasts with CP-based uncertainty estimation. Specifically, we employ Split Conformal Prediction (SCP), also called Inductive Conformal Prediction (ICP), which leverages calibration data to generate predictive intervals for any trained model. We evaluate this approach against QR and BNN, utilizing actual net-load data from LV substations in Portugal and the UK. This comparative analysis sheds light on this method's relative effectiveness and applicability in net-

load forecasting.

The structure of this paper is as follows: Section II discusses the background and introduces the Conformal-MLPF. Section III details the case study, including the dataset, input features, performance metrics, experiment descriptions, and evaluation methodology. Section IV presents and discusses the experimental results. Finally, Section V summarizes key findings, limitations, and future research directions.

II. Methods

A. Problem Formulation and Split Conformal Prediction

The main focus is to address short-term net-load forecasting problems, which involve predicting the near-future net-load $\mathbf{y}_{t+1:t+H}$ over a forecast horizon H ranging from several hours up to one week. To make these predictions, we use a base predictor $f_\theta(\mathbf{x}_L, \mathbf{c}_H)$ that utilizes historical features $\mathbf{x}_{t-L:t}$ with a time lag of L, as well as future covariates $\mathbf{c}_{t+1:H}$. Thus given set of observed data $\mathcal{D}_{train}$ our goal is to contsruct a predictor $f_\theta(\mathbf{y}|\mathbf{x}, \mathbf{c})$ that allows us to predict H steps confidence interval $\mathcal{C}$ into the future such that:

$$\hat{\mathbf{y}}_H \sim f_\theta(\mathbf{x}_L, \mathbf{c}_T) \tag{1}$$

CP methods can be broadly divided into two categories: those that involve retraining the model multiple times and those that use sample splitting, known as split conformal methods [18]. Compared to full conformal, SCP formulates conformal prediction using a dataset to train the predictive model and a calibration dataset on which to compute a critical non-conformity score, used to define the width of the prediction interval [14], [15], [19]. The SCP involves establishing a conformity score function $\gamma(.)$ to capture how well a sample conforms to the proper training set.

$$\gamma_k = |y_k - \mu_\theta(x_k)| \tag{2}$$

The non-conformity score, γ, measures the disagreement between the model's output and the ground truth in the calibration data $\mathcal{D}_c$ independent of the training set [19]. It assesses the conformity between the calibration's response values y and the predicted values μ_θ [15]. Thus, a larger γ implies a poor calibrated model. Therefore the goal of SCP is to obtain a prediction interval C_{k+1} for a new data point x_{k+1} given the calibration data score $\{\gamma_0, \dots \gamma_k\}$ such that

$$\mathrm{C}(x_{k+1}) = \{y \in \mathbb{R} : \gamma(x_{k+1}, y) \leq \mathcal{Q}_{1-\alpha}(\hat{F}_{k+1})\} \tag{3}$$

where $\hat{F}_{k+1}(.)$ is the empirical c.d.f of the samples $\gamma_{1:k} \cup \infty$. $\mathcal{Q}_{1-\alpha}$ denotes the $1-\alpha$-quantile [18]. This is equivalent of finding the p^{th} smallest non-conformity score ε such that

$$\varepsilon = \mathcal{Q}_p\left(\{\gamma_0, \dots \gamma_k\}\right) \tag{4}$$

where $p = \left\lceil \frac{(k+1)(1-\alpha)}{k} \right\rceil$ and k is the size of the calibration set. The ε threshold defines the size of the prediction set, which, in its simplest form, is centered on the predicted value such that

$$\mathcal{C}(x_{k+1}) = [\mu_\theta(x_{k+1}) \pm \varepsilon] \tag{5}$$

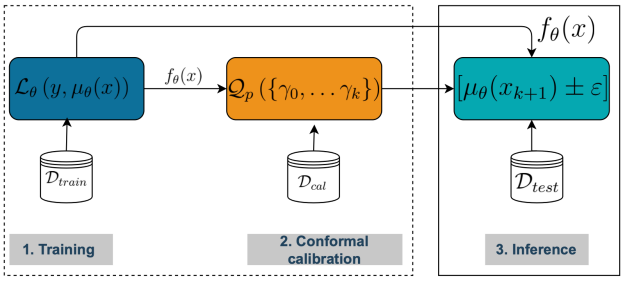


Fig. 1: Overview of conformalised-MLPF.

B. Conformal-MLPF

The Conformal-MLPF refines the Multilayer Perceptron based Forecast (MLPF) forecasting framework introduced in [7] to incorporate uncertainty quantification. It combines the MLPF with SCP to quantify the uncertainty of the point net-load forecast in a predictive interval. This integration enriches the forecasting process by generating predictive intervals that capture the expected variability of net-load values. The proposed framework is depicted Fig. 1 and consists of three stages: training, conformal calibration, and inference.

1) Training of deterministic MLPF: In this stage, we learn the parameters of the deterministic MLPF to obtain f_θ, which produces point forecast μ_θ. The MLPF is trained on the training dataset $\mathcal{D}_{train}$ using a supervised loss function $\mathcal{L}_\theta(y, \mu_\theta)$. This loss function combines the L_1 and L_2 norms such that:

$$\mathcal{L}_\theta(y, \mu_\theta) = \frac{1}{H}\sum_{t=1}^{H} \lambda(y_t - \hat{y}_t)^2 + (1-\lambda)|y_t - \hat{y}_t| \tag{6}$$

The training objective is to minimize the expected value of the loss function over the training data distribution: $\min_\theta \mathbb{E}_{(x,y)\sim\mathcal{D}_{train}}[\mathcal{L}_\theta(y, \mu_\theta)]$ where λ is the hyper-parameter that determines the weight of each error term in the loss function. The L_2 assigns greater weight to more significant errors, making the model more sensitive to outliers. This can be beneficial when specific extreme errors should be penalized more heavily, such as when predicting demand during extreme weather events. On the other hand, the L_1 is less sensitive to outliers, allowing the model to focus on capturing overall trends rather than being overly influenced by anomalous data points [20]. For a detailed description of the training procedure, refer to Section II-C.

2) Conformal calibration: The subsequent stage recalibrates the trained deterministic MLPF model f_θ using the calibration data $\mathcal{D}_{cal}$, following the SCP calibration procedure. Since the trained model f_θ is designed to produce multi-horizon forecasts, denoted as $\mathbf{y}_H = \mu_\theta(x)$, with H horizons ranging from $t+1$ to $t+H$, at each time step t and future time step h where $t \leq h \leq t+H$, we define $\mathcal{C}^k_{h|t}(x_k)$ as a prediction interval for the forecasted value $\hat{y}^k_h$ produced by the model.

For each data point k in the calibration data $\mathcal{D}_{cal}$, we group the obtained H forecasts to generate a vector of non-

conformity scores:

$$\gamma_H^k = \{\gamma_{t+1}, \ldots, \gamma_{t+H}\} \tag{7}$$

where γ_h is associated with the h^{th} forecast. Therefore, for new data point $k+1$, the calibrated forecast interval with a finite sample guarantee for each h-step is defined as:

$$\mathcal{C}_{h|t}(x_{k+1}) = [\mu_{\theta h}(x_{k+1}) \pm \varepsilon_h] \ \forall h \in [t, t+H] \tag{8}$$

Utilizing the non-conformity score defined in Eq. (2) for calibration is a common practice, as it aligns with the SCP methodology. However, as pointed out in [17], this score, the absolute error between the actual and forecasted values, can lead to asymmetric intervals where the probability mass below the lower and upper bounds may not be equal. As such, the effectiveness of signed score (Eq. (9)) was introduced in [17].

$$\gamma_{sgn}(x_k) = y_h^k - \mu_\theta(x_k)_h \tag{9}$$

Unlike the absolute non-conformity score, the sign score captures the direction of the forecast error.

3) Inference: Finally, during the inference stage, we leverage the trained deterministic model, denoted by f_θ, and the pth quantile non-conformity score, ε_h, to generate prediction intervals.

C. MLP Architecture and Training Procedures

We utilize a scalable Multilayer Perceptron (MLP) framework proposed in [11], which employs two encoders: the past encoder $g_\phi(\mathbf{x}_L, \mathbf{c}_L)$ and the future encoder $h\phi(\mathbf{c}_H)$. The two encoders comprised a set of Feed-forward neural network (FFN), BatchNorm, and a non-linear activation function as illustrated in Fig. 2b. These encoders aim to generate representations $\phi(gh)$ for historical variables and future covariates.

The research of Chu et al. inspires the design of this framework [20], which emphasizes the significant influence of covariate features, such as solar radiation, on future net-load. The encoders aim to create representations for past and future variables, denoted as $\phi(gh)$, which are utilized as input to generate forecasts and their corresponding predictive uncertainty.

$$\phi(g) = g_\theta\left(\mathrm{E}(\mathbf{x}_L)\right) \tag{10}$$

$$\phi(h) = h_\theta\left(\mathrm{E}(\mathbf{c}_H)\right) \tag{11}$$

$$\phi(gh) = \phi(g) + \phi(h) \tag{12}$$

where $\mathrm{E}(\mathbf{x}; \theta) = \mathrm{Dropout}\left[\mathrm{RoPE}(\mathrm{LayerNorm}(\mathbf{x})\right]$.

We used two hidden layers to train all methods, each containing 256 nodes and the SiLU activation function in the hidden layers. 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. For implementation details, readers can refer to [7].

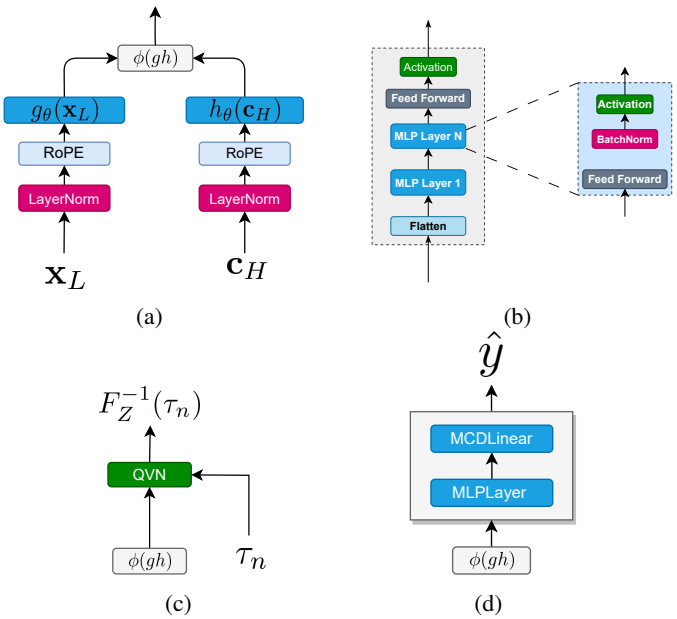
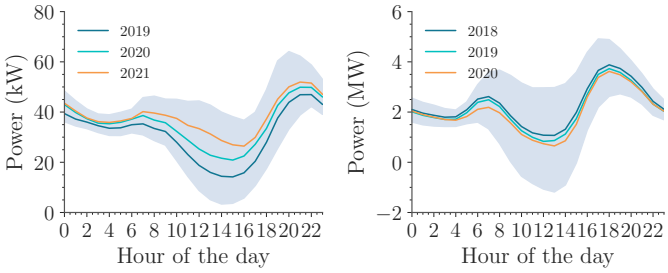


Fig. 2: (a) The hierarchical architecture of the scalable MLPF (b) The MLP encoder block (c) The final prediction layer for MLP Quantile Regression (MLP-QR) (d) The final prediction layer for Last-layer Monte Carlo dropout (MLP-MCD).

III. Experimental Design

A. LV Substation Datasets and Input Features

The proposed approach is evaluated on two real-life substation datasets: the Madeira Low-Voltage distribution substation dataset in Portugal (MLVS-PT) [11], and the The Stentaway substation dataset in Plymouth-UK (SPS-UK) [21] as visualized in Fig. 3.



(a) MLVS-PT (b) SPS-UK

Fig. 3: The net-load for the two datasets over time.

The MLVS-PT dataset includes net-load demand from a LV substation Madeira Island, serving 100 consumers. Data was recorded every minute from January 2019 to December 2021 and resampled to 30-minute intervals, with meteorological data from [22].

The SPS-UK dataset contains demand, PV generation, and weather data from a substation in Plymouth, UK[1]. Load demand and PV generation were recorded every 30 minutes from

[1]https://www.nationalgrid.co.uk/pod-data-science-challenge

November 2017 to July 2020. The net-load was calculated by subtracting PV generation from load demand, with weather data from MERRA-2[2] interpolated to 30-minute intervals.

The input to the model consists of a combination of historical net-load with past and future covariates. The covariates include the global horizontal irradiance, temperature, and various time-based features that account for weather conditions, seasonal effects, and calendar days. Specifically, the model incorporates time-derived features: hour, day of the week, day of the month, and session. The session variable is a boolean variable that takes the value of 1 during the day and 0 at night. These features were selected based on their high correlation with net-load demands. In line with the approach of [11], the date-time-related features are transformed using sine and cosine transformations to capture daily and yearly patterns. The input features and forecasted variables were normalized using z-score normalization.

B. Benchmarks

We empirically evaluate the performance of the proposed conformal-MLPF compared to two commonly used probabilistic forecasting methods: QR (MLP-QR) and Monte Carlo dropout (MLP-MCD) [23]. The MLP-MCD, as illustrated in Fig. 2d, extends the MLPF by adding dropout decoder layers. Dropout is a regularization technique commonly used in DNN to prevent overfitting [23]. In the context of the MLP-MCD method, characterized by a dropout probability denoted as π, the q distribution takes the following form:

$$q(\theta_l) = \prod_{i=1}^{d_{out}} \pi q(\theta_l) + (1-\pi) q(\theta_l) \tag{13}$$

The MLP-MCD runs multiple forward passes through the network during test time with dropout enabled to obtain a probabilistic forecast. This stochastic sampling allows the model to capture the uncertainty in its predictions [24] such that;

$$\hat{\sigma} = \sqrt{\sigma_{\mathcal{D}}^2 + \sigma_{\epsilon}^2} \tag{14}$$

where $\sigma_{\mathcal{D}} = \sqrt{\sigma_s^2 + \exp(\mathbf{z}_\epsilon)^2}$, σ_s^2 is sample variance defined as $\sigma_s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{y}_i - \hat{\mu})$, and $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i)$. σ_ϵ^2 is the variance of the homogeneous noise ϵ such that $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$ which account for the fact that the MLP-MCD is only an approximation.

For QR, we use a common DNN-based QR approach with MLP-based architecture similar to the one used in conformalised-MLPF, as depicted in Fig. 2c. In this method, only the quantile value function is parameterized by MLP using fixed and predefined quantile fractions. The quantile fractions remain fixed throughout the training and inference process. The parameters of the MLP-QR are optimized by minimizing the pinball loss, as defined in Eq. (15):

$$\mathcal{L}_\tau(\epsilon_\tau) = \frac{1}{T}\sum_{t=1}^{T}\sum_{n=1}^{N} \max\left[\epsilon_{\tau_n^t} \cdot \tau, (1-\tau)\cdot \epsilon_{\tau_n^t}\right] \tag{15}$$

[2]https://github.com/emilylaiken/merradownload

where $\epsilon_\tau = y_t - Q_\theta(\hat{\tau}_{\theta t})$

C. Performance metrics and evaluation procedure

The following performance metrics are used to assess the quality of a probabilistic forecast: Predictive Interval Coverage Probability (PICP), Normalized Mean Prediction Interval width (NMPI), and Combined Coverage, Width and Forecasting Error (CWE). The PICP (Eq. (16)) measures the percentage of true values that fall within the prediction intervals, whereas the NMPI (Eq. (17)) is the normalized width of the prediction intervals, which gives the average width of the prediction interval [25].

$$\text{PICP} = \frac{1}{H}\sum_{t=1}^{H}\begin{cases}0, & \notin [\mathcal{C}_t^U, \mathcal{C}_t^L] \\ 1, & y_t \in [\mathcal{C}_t^U, \mathcal{C}_t^L]\end{cases} \tag{16}$$

$$\text{NMPI} = \frac{1}{R}\text{median}(\mathcal{C}_d) \tag{17}$$

where $\mathcal{C}_d = \{\mathcal{C}_1^U - \mathcal{C}_1^L, \dots \mathcal{C}_H^U - \mathcal{C}_H^L\}$, $\mathcal{C}_t^U$ and $\mathcal{C}_t^L$ are upper and lower predictive interval, H is the forecasting horizon and $R = \max\{y_1, \dots y_H\}$ is range of true value.

The CWE (Eq. (18)) metric measures how well a model is calibrated. A calibrated model gives predictive intervals that are both correct and narrow.

$$\text{CWE} = 2 \cdot \frac{\gamma_{nmpi} \cdot \gamma_{pcip}}{\gamma_{picp} + \gamma_{nmpi}} \tag{18}$$

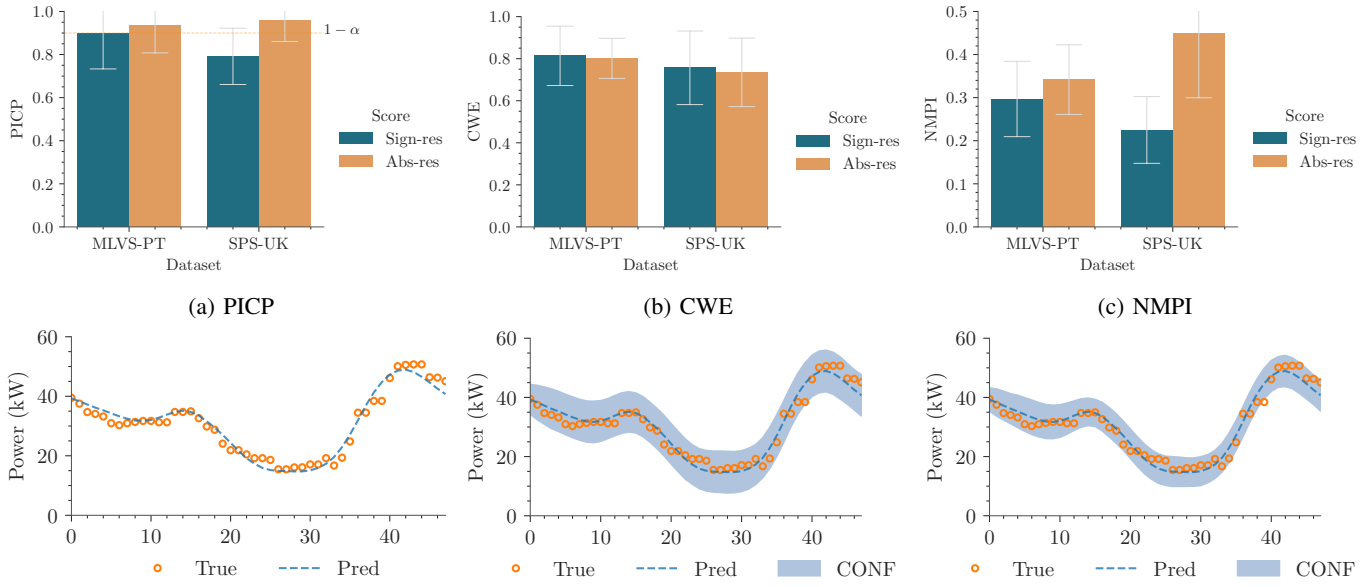
Finally, we use the Normalized Root Mean Squared Error (NRMSE), a point-forecast metric, to assess the quality of the point forecast.

The training was done using the backtesting cross-validation technique outlined in [11]. Specifically, we implemented 10-fold backtesting cross-validation with expanding window strategies, setting the initial historical period to a minimum of 12 months and the fixed future time window to 6 months. As we moved through the time series, the sliding window was extended by three months. In each fold, only 90% of the data were used as training set $\mathcal{D}_{train}$, and the remaining 10% were used as calibration set $\mathcal{D}_{cal}$.

D. Experiments Description

The evaluation focuses on two key aspects: (1) how the two non-conformity scores (absolute error and signed error) influence the quality of the prediction intervals, and (2) the overall performance of conformal-MLPF against the two baseline methods (MLP-MCD and MLP-QR).

The first experiment investigated how the two non-conformity scores (absolute and signed errors) impacted the quality of the prediction intervals generated by conformal-MLPF. The MLPF model was calibrated using each non-conformity score, and the resulting prediction intervals were compared. In the second experiment, we compared the overall performance of conformal-MLPF with two baseline methods: MLP-MCD and MLP-QR. The comparison focused on point forecasts and interval metrics.



(a) PICP (b) CWE (c) NMPI

(d) Deterministic (e) Abs-res (f) Sign-res

Fig. 4: Results comparison for four Conformal-MLPF with sign residual and absolute residual non-conformity score.

IV. Results and Discussion

A. Experiment 1: Influence of Non-Conformity Score

The results are depicted in Fig. 4. It is evident from Fig. 4a that the absolute residual non-conformity score (Abs-score) achieves higher coverage ($\geq 1-\alpha = 0.9$) on both datasets, while the sign-residual non-conformity score (Sign-score) attains a PCIP ($1-\alpha = 0.9$) solely on the MLVS-PT dataset. In both cases, the Abs-score's PCIP is higher than the sign-score's, about 4% higher on MLVS-PT and over 20% higher on the SPS-UK dataset. These results imply that prediction intervals constructed using Abs-score tend to encompass the actual values more frequently. However, this higher coverage for abs-score comes at the cost of broader coverage intervals, meaning that the absolute score provides a less precise estimate of the forecast error. We observe from Fig. 4c that the sign-score exhibits a lowerNMPI of 0.30 compared to 0.34 for the absolute score on the MLVS-PT dataset and 0.23 NMPI on the SPS-UK dataset, which is 48% lower compared to that of the Abs-score. This implies that the intervals constructed using the abs-score are broader, as illustrated in Fig. 4e, potentially leading to less precise predictions, unlike the sign-score, which offers a narrow predictive interval while covering most of the true value as shown in Fig. 4f.

The key difference between the abs-score)and the sign-score lies in how they treat forecast errors. The abs-score prioritizes magnitude, encompassing positive and negative errors, leading to wider prediction intervals with higher coverage (as seen in Fig. 4a). Conversely, the sign-score focuses solely on direction, resulting in narrower intervals but potentially sacrificing coverage. However, this tradeoff can be beneficial, as evidenced by the sign-score's slightly higher CWE score on the MLVS-PT dataset (refer to Fig. 4b). The sign-score achieves greater precision by focusing on direction, even if it misses some errors of the same magnitude but with the opposite sign.

B. Experiment 2: Comparison with Baselines

Our second experiment compared the proposed Conformal-MLPF against two baseline models. The results are summarized in Table I. The comparison shows that Conformal-MLPF and MLP-QR have similar forecasting accuracy, achieving NRMSE scores of 0.09 and 0.13 on the MLVS-PT and SPS-UK datasets, respectively. MLP-MCD exhibited slightly lower forecasting power, reflected by higher NRMSE scores of 0.13 and 0.19 on the MLVS-PT and SPS-UK datasets.

When considering probabilistic metrics, the proposed Conformal-MLPF performed competitively, achieving a coverage probability PICP of 0.84, on par with the well-established MLP-QR. However, Conformal-MLPF has a slightly higher NMPI (0.26) compared to MLP-QR (0.22), leading to a marginally lower CWE. Notably, both Conformal-MLPF and MLP-QR achieved a balance between coverage and predictive interval width, suggesting good calibration, unlike MLP-MCD. These results indicate that the proposed Conformal-MLPF can be competitive with established methods such as QR without imposing any restrictive assumptions about the underlying data distribution.

On the SPS-UK dataset, Conformal-MLPF appears to be slightly conservative, as indicated by its lower NMPI. This resulted in lower PICP and CWE scores compared to MLP-QR. However, its CWE score remains higher than that of MLP-MCD. The wider intervals produced by MLP-MCD led to a lower CWE score than the other non-parametric QR and MLP-LD models, indicating less precise predictive intervals.

TABLE I: Results of comparison with baselines.

Dataset	Model	NRMSE	PICP	NMPI	CWE
MLVS-PT	MLP-MCD	0.13	0.72	0.22	0.75
	MLP-QR	0.09	**0.84**	**0.26**	**0.82**
	Conformal-MLPF	0.09	**0.84**	0.28	0.79
SPS-UK	MLP-MCD	0.19	0.82	0.36	0.75
	MLP-QR	0.13	0.96	0.32	0.83
	Conformal-MLPF	0.13	0.78	**0.24**	0.74

V. Conclusion

This paper presented Conformal-MLPF, a lightweight and efficient CP-based net-load probabilistic forecast using neural networks. Empirical results on two real-world datasets show that Conformal-MLPF competes with established methods like QR without requiring restrictive data distribution assumptions. The study highlights the importance of the sign nonconformity score, which focuses on forecasting error direction and improving the balance between prediction interval coverage and width. This has significant implications for power demand forecasting and the feasibility of applying CP in real-world scenarios.

While this study explores the impact of conformal prediction, specifically SCP, for net-load forecasting, future research should incorporate advanced CP techniques. SCP provides only marginal coverage assurance with fixed predictive intervals, which may not guarantee conditional coverage for time series with strong seasonal and heteroscedastic behavior like net-load. Future work should investigate adaptive conformal prediction techniques, allowing predictive intervals to adjust to specific data points. Additionally, future research should evaluate Conformal-MLPF using distribution substation datasets from various geographical locations, installed capacities, and different RES such as wind power generation.

Acknowledgments

This paper is supported by the European Union's Horizon Europe R&I programme under grant agreement no. 101160665, project AHEAD (AI-informed Holistic EVs integration Approaches for Distribution grids). This research was also supported by FCT projects 10.54499/2022.15771.MIT; 10.54499/LA/P/0083/2020; 10.54499/UIDP/50009/2020; 10.54499/UIDB/50009/2020, and grant CEECIND/01179/2017 (LP).

References

[1] S. R. Deeba, R. Sharma, T. K. Saha, and D. Chakraborty, "A tool to estimate maximum arbitrage from battery energy storage by maintaining voltage limits in an LV network," in *2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2015, pp. 1–5.

[2] M. Saviozzi, S. Massucco, and F. Silvestro, "Implementation of advanced functionalities for distribution management systems: Load forecasting and modeling through artificial neural networks ensembles," *Electric Power Systems Research*, vol. 167, pp. 230–239, Feb. 2019.

[3] 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.

[4] S. Sreekumar, 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.

[5] J. Faraji, A. Ketabi, H. Hashemi-Dezaki, M. Shafie-Khah, and J. P. S. Catalão, "Optimal day-ahead self-scheduling and operation of prosumer microgrids using hybrid machine learning-based weather and load forecasting," *IEEE Access*, vol. 8, pp. 157 284–157 305, 2020.

[6] 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.

[7] A. Faustine, N. J. Nunes, and L. Pereira, "Efficiency through Simplicity: MLP-based Approach for Net-Load Forecasting with Uncertainty Estimates in Low-Voltage Distribution Networks," *IEEE Transactions on Power Systems*, pp. 1–11, 2024.

[8] 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.

[9] T. Zhang, X. Zhang, T. K. Chau, Y. Chow, T. Fernando, H. H.-C. Iu *et al.*, "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, 2023.

[10] J. de Vilmarest, J. Browell, M. Fasiolo, Y. Goude, and O. Wintenberger, "Adaptive probabilistic forecasting of electricity (net-)load," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4154–4163, 2024.

[11] 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.

[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] S. Lei, X. Liang, X. Wang, J. Ding, X. Ge, F. Wang, and J. Feng, "A short-term net load forecasting method based on two-stage feature selection and lightgbm with hyperparameter auto-tuning," in *2023 IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, 2023, pp. 1–6.

[14] V. Jensen, F. M. Bianchi, and S. N. Anfinsen, "Ensemble conformalized quantile regression for probabilistic time series forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2022.

[15] M. Zaffran, O. Féron, Y. Goude, J. Josse, and A. Dieuleveut, "Adaptive conformal predictions for time series," in *International Conference on Machine Learning*. PMLR, 2022, pp. 25 834–25 866.

[16] S. Sun and R. Yu, "Copula conformal prediction for multi-step time series forecasting," *arXiv preprint arXiv:2212.03281*, 2022.

[17] J. Jonkers, D. N. Avendano, G. Van Wallendael, and S. Van Hoecke, "A novel day-ahead regional and probabilistic wind power forecasting framework using deep cnns and conformalized regression forests," *Applied Energy*, vol. 361, p. 122900, 2024.

[18] S. I. Amoukou and N. J. Brunel, "Adaptive conformal prediction by reweighting nonconformity score," *arXiv preprint arXiv:2303.12695*, 2023.

[19] N. Seedat, A. Jeffares, F. Imrie, and M. van der Schaar, "Improving adaptive conformal prediction using self-supervised learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023, pp. 10 160–10 177.

[20] 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.

[21] 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.

[22] J. M. Bright, "Solcast: Validation of a satellite-derived solar irradiance dataset," *Solar Energy*, vol. 189, pp. 435–449, 2019.

[23] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[24] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021.

[25] 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.