

Short-Term Probabilistic Load Forecasting Based on Conformalized Quantile Regression

Lucija Matulin^{1,2*}, Tomislav Capuder¹, Tomislav Plavšić²

1 University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Energy and Power Systems

2 Croatian Transmission System Operator (HOPS), System Operation Department

(*Corresponding Author: lucija.matulin@fer.hr)

ABSTRACT

Short-term load forecasting is a fundamental task in reliable and secure power system operation, particularly in the current landscape marked by increased integration of renewable energy sources and electric vehicles, which introduces stochasticity and raises uncertainty. To express uncertainty in load predictions in the form of a probabilistic forecast, prediction intervals are generated. The variability in load values exhibits higher volatility during the day due to increased human activities, contrasting with lower variability at night. Classic methods for constructing prediction intervals cannot correctly model the variability in uncertainty leading to overly conservative prediction intervals. In this paper, we propose a novel approach – conformalized quantile regression – to create more informative, variable-length prediction intervals. Experimental results, based on a real load dataset from the Croatian Transmission System, showcase the method's superior performance in capturing adaptive-length prediction intervals. This translates to achieving higher coverage with shorter prediction intervals compared to conventional methods.

Keywords: probabilistic load forecasting, short term load forecasting, conformalized quantile regression, prediction interval, random forest

NONMENCLATURE

Abbreviations

STLF	Short-Term Load Forecasting
PI	Prediction Interval
CQR	Conformalized Quantile Regression
QR	Quantile Regression
CP	Conformal Prediction
TSO	Transmission System Operator

LF	Load Forecasting
AI	Artificial Intelligence
ML	Machine Learning
ANN	Artificial Neural Network
PLF	Probabilistic Load Forecasting
KDE	Kernel Density Estimation
RF	Random Forest
QRF	Quantile Random Forest
PICP	Prediction Interval Coverage Probability
MPIW	Mean Prediction Interval Width
<i>Symbols</i>	
L_{h-n}	Load in the n -th hour before the current hour, h
k	Indicator binary variable: is the test set observation or not covered by the constructed PI?
y_i	i -th observation in the test set
y_{li}	Lower bound of the i -th PI
y_{ui}	Upper bound of the i -th PI

1. INTRODUCTION

Recently, the impacts of the rapidly changing climate have become more prominent and manifest in extreme weather conditions, already affecting the entire world. Many countries, organizations, and initiatives are committed to reducing emissions, decreasing carbon footprint, and achieving carbon neutrality. Several significant initiatives address climate change, such as the European Green Deal [1], which aims to achieve a climate-neutral Europe by 2050, and the Paris Agreement [2], which many countries worldwide have signed. The Paris Agreement targets to limit global warming to 1.5 degrees Celsius and aims to achieve carbon neutrality by the mid-21st century. Carbon neutrality within power and energy systems can be

accomplished by adopting and increasing the integration of renewable energy sources, more energy-efficient technologies, and low-carbon technologies.

Introducing renewable energy sources presents unique challenges and uncertainties to complex electric power systems and their operation. The imbalance of generation and load affects the secure and reliable operation of the electric power system. Traditionally, the electric load or energy demand has always been stochastic and unpredictable due to various factors such as human activities and weather conditions, and generation was planned based on the forecasted load values. With the increase in intermittent renewable generation, new sources of stochastic and variable behavior are emerging on the generation side. Developing new forecasting methods to quantify emerging uncertainties in power systems is essential.

Accurate short-term load forecasting (STLF) is a crucial task for efficient and reliable power system operation and planning, traditionally accomplished by operators and planners [3]. With the development of market-oriented operations, it has become a fundamental task for all participants in the electricity market. Short-term load forecasting usually refers to forecasting horizons up to two weeks, according to some load forecasting (LF) classifications [4]. During the past three decades, researchers and practitioners developed numerous load-forecasting methods based on artificial intelligence (AI) and machine learning (ML) models. The power industry saw one of the earliest commercial applications of AI in the 1990s when it employed an artificial neural network (ANN) for STLF [3]. In the early stages of LF research and practice, the main focus was deterministic LF [4]. The objective was to forecast a single load value without quantifying the uncertainties. The most notable and frequently used ML-based techniques for deterministic LF include linear regression, artificial neural networks, support vector machines, gradient-boosting machines, decision trees, and ensembles [4], [5]. Probabilistic load forecasting (PLF) has become increasingly important as an emerging research topic [4]. PLF can quantify prediction uncertainty by assigning a probability to the expected value. Probabilistic forecasts can be given as quantiles, prediction intervals, or probability density functions [4].

This paper focuses on probabilistic forecasting achieved by constructing prediction intervals (PIs) to quantify prediction uncertainty. With high probability, the response variable lies within an interval between lower and upper bounds known as PI [6]. Two common methods for constructing prediction intervals are

conformal prediction and quantile regression. The basic concept of conformal prediction is fitting a regression model on the training samples and using the residuals for a held-out validation attempting to construct locally variable-length PIs. On the other hand, quantile regression relies on estimating conditional quantile functions and forming the corresponding intervals based on specific levels, such as 5% and 95%, to achieve a 90% coverage. Y. Romano [6] first introduced the concept of combining conformal prediction and quantile regression in what is known as Conformalized Quantile Regression (CQR). This approach combines the strengths of both underlying methods, ensuring the validity of conformal prediction and the statistical efficiency of quantile regression [6]. Additionally, it is flexible enough to be used with any algorithm for quantile regression. Related work of adopting the CQR method in power system forecasting tasks include [7], [8]. In [7] authors proposed a deep neural network model, N-BEATS, for probabilistic STLF and demonstrated its effectiveness by constructing narrow PIs with expected nominal coverage. In [8] the authors presented a novel ensemble-based method, ensemble conformalized quantile regression, and tested the performance on several real-world datasets (i.e., electric load, solar and wind power production, air temperature).

This paper proposes a probabilistic STLF method based on CQR, as introduced in [6], using the random forest as a base algorithm. The proposed method is tested on a real-world electric load dataset provided by the Croatian Transmission System Operator, demonstrating the capability to produce valid and adaptive PIs, shorter in comparison to classic methods while maintaining satisfactory nominal coverage on two short-term forecasting horizons (one hour and one day ahead). The proposed probabilistic STLF method can be used with any other quantile regression algorithm (i.e., neural network or ensemble models), on other forecasting horizons, or with other time-series datasets. CQR method performance is analyzed based on conformity scores: PI coverage, PI length, and model-building time (including training and held-out validation).

2. METHODOLOGY

2.1 Dataset description

The research presented in this paper utilized a real-world load dataset provided by the Croatian Transmission System Operator, HOPS. The dataset consists of the following historical time-series covering the period from January 2018 to December 2022:

- historical load data,
- public holiday data,
- and air temperature data.

The temporal resolution of the provided datasets is hourly. The spatial granularity of the load dataset includes aggregate system-level load values or the mean active power absorbed by all loads on the transmission network level or transferred to the distribution network. The list of public holiday data for the Republic of Croatia was retrieved from the *holidays* [9] library, representing non-working days in Croatia. Air temperature for the four biggest Croatian cities (Zagreb, Split, Rijeka, and Osijek) was attained from the publicly available platform *Meteostat* [10].

Figure 1 represents the annual distribution of load values given in hourly resolution in the kernel density estimation (KDE) form. It illustrates the first (Q1) and the third (Q3) quartiles and median load values (Q2) by year, where dashed lines inside the violin represent quartiles. Violin width denotes the frequency of a particular load value in each year. The figure indicates that the violin widths and quartile values follow similar annual load patterns, except for 2020. During the COVID-19 lockdown periods in 2020, the load values were visibly lower, indicated in all quartile values lower than those in other years. Other than annual periodicity, the load dataset exhibits daily and weekly periodicities. The daily pattern is visible in multiple peak and valley values during the 24-hour period, which varies in occurrence and duration depending on the season. The weekly seasonality is observable in interchanging periods of higher load during the work week and lower load during the weekend.

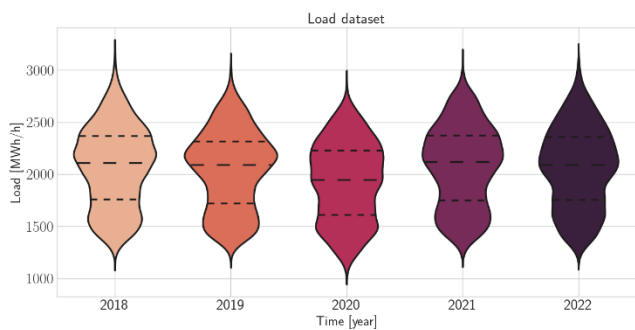


Figure 1: Hourly load distributions by year

The load values are impacted by several weather factors, with air temperature having the most significant influence [5]. The electricity load decreases when the air temperature is around 20°C and increases during higher heating and cooling usage.

2.2 Feature extraction and feature selection

This section describes the process of obtaining features with data extraction and transformation techniques, which are inputs for regression from raw data described in the previous section. In this process, new variables that were not part of the original dataset are created. The complete feature set is included in Table 1.

There are three feature categories: historical, calendar, and meteorological, as indicated in Table 1. Historical features or lag features are created by shifting the actual load data for desired timesteps (i.e., L_{h-n} denotes load in the n -th hour before the current hour, h). Calendar features include the information of the current hour and day filtered from *DateTime* indices. Cyclical sine and cosine features are calculated using the sine and cosine functions for the current hour in the day or week. Meteorological features include the air temperature in the four biggest Croatian cities, Zagreb, Split, Rijeka, and Osijek, lagged for 24 hours, as these observed air temperature values are available at the time of forecast creation.

Table 1: Feature categorization and description

Category	Name	Description	Measurement Unit
Historical	L_{h-1}	Load value, 1h before	MWh/h
	L_{h-2}	Load value, 2h before	MWh/h
	L_{h-3}	Load value, 3h before	MWh/h
	L_{h-4}	Load value, 4h before	MWh/h
	L_{h-6}	Load value, 6h before	MWh/h
	L_{h-24}	Load value, 24h before	MWh/h
	L_{h-48}	Load value, 48h before	MWh/h
	L_{h-168}	Load value, 168h before	MWh/h
Calendar	holiday _h	Binary indicator variable, "is Holiday?"	-
	wday _h	Categorical variable, day of the week (0, 6)	day
	hour _h	Hour of the day	h
	sin_hour _h	Cyclical variable, hour of the day, sine function	h
	cos_hour _h	Cyclical variable, hour of the day, cosine function	h
	sin_week _h	Cyclical variable, hour of the week, sine function	h
	cos_week _h	Cyclical variable, hour of the week, cosine function	h
	Meteorological	temp_Zg _{h-24}	Air temperature in Zagreb, 24h before
temp_St _{h-24}		Air temperature in Split, 24h before	°C
temp_Ri _{h-24}		Air temperature in Rijeka, 24h before	°C
temp_Os _{h-24}		Air temperature in Osijek, 24h before	°C

Features were selected based on the Pearson correlation coefficient, measuring the linear correlation between each feature and the target load value. ML models used in the experimental part mainly share presented features, except the lag features which are not available at the time of forecast (i.e., lag features L_{h-1} are not available 24h in advance; as a result, models with a 24-hour forecasting horizon include L_{h-24} and higher).

2.3 Experimental setup

The presented research was conducted using *Python* and *Jupyter Notebook* [11] environment. *NumPy* [12] and *Pandas* [13] packages were used for time-series data analysis, while *Seaborn* [14] and *Matplotlib* [15] packages were used for generating all visualizations. The *scikit-learn* [16] package was used for implementing classic machine learning models and data preprocessing functions. Furthermore, the *holidays* [9] library was used to acquire public holiday data, and the *Meteostat Python API* [10] was utilized to acquire weather data. The *Nonconformist* [17] package was used for conformal predictions, and the *quantile-forest* [18] package was used for quantile forest regression implementation. Supplemental functions for implementing conformalized regression, quantile regression, and conformalized quantile regression were used from the original implementation in [19].

Experiments were performed as follows: firstly, the datasets described in section 2.1 were loaded, and the features were created as outlined in section 2.2. The dataset was split into a training set comprising of four years of historical data and a test set comprising of one year of data. This created the 80-20 ratio for the training-test split. Furthermore, the training set was further divided into two subsets: training and calibration set, in the 50-50 ratio. The data was standardized to have zero mean and unit variance using the *StandardScaler* preprocessing function, and then the output was scaled by its mean absolute value. This ensured the exchangeability of the original data (as per the original paper) [6]. Training was conducted, with the validation or calibration set utilized to optimize hyperparameters. The trained models were then tested on the test set with the defined evaluation metrics. In the next section, a detailed description of the models that have been used and compared in this research is provided.

2.4 Model description

As a baseline or persistence model, ridge regression, which is a linear regression model with l2-

norm regularization, was utilized. The regularization parameter is tuned via cross-validation. Ridge model implementation is used from the *scikit-learn* [16] package. Random forest regression is used as the underlying model for implementing several classic models:

- split conformal regression,
- local conformal regression,
- quantile regression,

and the proposed conformalized-quantile regression. All analyzed methods, except for classic quantile regression, require a calibration set for held-out validation. In this case, the training procedure for classic quantile regression is performed on the entire training set.

Random Forest (RF) is a bootstrapping-based ensemble method based on decision trees. It is a commonly used supervised machine learning technique for classification and regression tasks because of its predictive accuracy and ability to avoid and prevent overfitting. Neural networks are widely used models in literature. However, they require longer training time and larger datasets and often encounter the quantile crossing problem in quantile regression tasks. On the other hand, random forest algorithms are not affected by the quantile crossing problem. Quantile crossing occurs when two separate quantile regressions estimate a lower and an upper quantile, and there is no guarantee that the lower estimate will be smaller than the upper estimate.

Split conformal prediction is a regression technique that predicts the conditional mean by first dividing the training data into two subsets: training and calibration set. The calibration set is then used to compute the absolute residuals, which are used to calculate the quantile of the empirical distribution. However, a drawback of this method is that the prediction intervals are fixed in length and do not vary based on the query point.

A locally adaptive variant of conformal prediction uses scaled residuals instead of absolute residuals to calculate the quantile of the empirical distribution. Scaling is performed by dividing residuals with a measure of residual dispersion, and the problem is that it underestimates prediction errors since it uses residuals on training data. Prediction intervals are, as a result, less adaptive and longer.

Both split and local conformal regression implementations are from the *Nonconformist* [17] package, using the *scikit-learn* [16] Random Forest regression as the underlying model.

Quantile regression is a method that estimates the conditional quantile function of output values. By minimizing the pinball loss function, it provides adaptive prediction intervals. The prediction intervals are formed by fitting the conditional quantile function at the lower and upper percentage levels, creating corresponding intervals. Most ML algorithms can be integrated with QR. The Quantile Random Forest (QRF) model is implemented in the *quantile_forest* [18] package.

Conformalized Quantile Regression (CQR) inherits the best of both worlds: validity of prediction intervals from conformal prediction techniques and the statistical efficiency from quantile regression in terms of intervals adaptive to the local variability of provided samples. CQR is flexible and can be used with most algorithms, inheriting the strengths of QR. However, the key advantage that the CQR method offers is a rigorous control of the miscoverage rate, irrespective of the underlying regression algorithm, enhancing confidence in the validity of the results. This is a particularly valuable characteristic in applications where precision and accuracy are required, enhancing confidence in the validity of the results. CQR implementation used in this research is from the original paper [6] and CQR repository [19].

When referring to classic methods in this paper, it means split and local conformal prediction, as well as quantile regression methods. Classic methods are compared to Conformalized Quantile Regression method using the same underlying algorithm: Random Forest regression. The goal is to achieve 90% nominal coverage, and the corresponding levels to fit the conditional quantile functions are at 5% and 95%. To have a fair comparison, hyperparameters of CQR are identical to classic methods. Further details about hyperparameters and comparisons are disclosed in the following chapter.

2.5 Evaluation metrics

In this paper, the accuracy of the constructed prediction intervals (PIs) is evaluated using two conformal scores: the prediction interval coverage probability (PICP), defined with Equation (1) and the mean prediction interval width (MPIW), defined with Equation (2). PICP indicates the extent to which the PIs cover the target values in the test dataset, or in other words, the proportion of test data values that fall within the calculated PIs. On the other hand, MPIW provides information about the length of the PIs and measures the average absolute difference between the estimated lower and upper bounds of the target values.

PICP can be defined as:

$$PICP = \frac{\sum_{i=1}^n k_i}{n}, \quad (1)$$

MPIW is expressed as:

$$MPIW = \frac{1}{n} \sum_{i=1}^n (\widehat{y}_{li} - \widehat{y}_{ui}), \quad (2)$$

where n represents the test set length, vector k represents an indication of whether the test set observation is or is not covered by the constructed PI:

$$k = \begin{cases} 1, & \text{if } y_i \in [\widehat{y}_{li}, \widehat{y}_{ui}] \\ 0, & \text{if } y_i \notin [\widehat{y}_{li}, \widehat{y}_{ui}] \end{cases}, \quad (3)$$

y_i is i -th observation in the test set, and y_{li} and y_{ui} , respectively, represent lower and upper bounds of the i -th PI.

PICP is expressed as a coverage percentage, while MPIW is calculated on scaled load values without a measurement unit.

3. CASE STUDIES

This chapter provides descriptions of the analyzed cases for two forecasting horizons: one hour ahead and one day ahead (24 hours). Real-time system operation procedures typically apply one-hour-ahead (or lower, i.e., 15-min-ahead) load forecasts, while one-day-ahead load forecasts are utilized for system operational planning tasks.

For each forecasting horizon, all the previously mentioned methods are calculated. The results are then compared based on defined evaluation metrics and model-building time, which includes the time employed for model training and calibration (where applied). To ensure credible results, the experiment is conducted using ten random training-calibration splits for each forecasting horizon and method. In every experiment, a random seed is generated to shuffle the training set and set the random state parameter in the RF regression model. Table 2 provides a list of hyperparameter values used. Hyperparameters are the same for both horizons and all methods using the RF as an underlying algorithm.

For the one-hour-ahead models, historical and calendar features are exclusively used. This is because recent lag features, such as load value one hour prior, can better predict the output value. However, for the one-day-ahead models, meteorological features are also included, along with historical and calendar features. Additionally, these models only consider lag features available for more than 24 hours ahead.

Table 2: List of hyperparameter values

Hyperparameter	Value
<i>n_estimators</i>	500
<i>min_samples_leaf</i>	50
<i>max_features</i>	X_train.shape[1]
<i>significance</i>	0.9
<i>lower quantile level</i>	0.05
<i>upper quantile level</i>	0.95

4. RESULTS & DISCUSSION

This chapter presents the results obtained from experiments. Figure 2 and Figure 3 present the results of all analyzed methods for one-hour-ahead and one-day-ahead forecasting horizons, respectively, based on ten different training-calibration splits. The left plot illustrates the PICP metric, representing the average PI coverage achieved by each method across ten calculations. The desired expected coverage was set at 90% (defined with a black dashed line on the left, PICP plot). The right plot indicates the MPIW, the scaled average PI length. Ideally, PICP mean values (middle line on the boxplot) should be over the expected 90% value, and the mean MPIW values should be as low as possible, indicating shorter or more narrow PIs.

When observing the aggregate results from ten different training-calibration splits on models with one-hour-ahead forecasting horizons in Figure 2, it is observable that all analyzed algorithms achieved over the expected 90% coverage, which is visible in all mean values denoted on boxplots. The QRF method yields a mean PICP of 96.963% with the longest PIs, indicating an overly conservative approach to prediction. The Linear (Ridge) model serves as a baseline to compare with more complex RF-based models for improved results. This can be observed by examining the mean coverage over 90% (91.062) and relatively short PIs (mean MPIW is 0.089). The Ridge model performs even better than the QRF model for this forecasting period. It is noticeable that the locally adaptive version of CRF performs better than the standard or split version of CRF. The coverage is almost the same, but the PIs are shorter. The proposed CQR RF method displays the best results, covering the expected 90% significance level (with mean PICP value of 90.171) and the lowest MPIW value (0.084), representing the narrowest PIs.

Regarding Figure 3, which represents the combined outcomes of models with a one-day-ahead forecasting period, it should be noted that the CQR RF model has a mean PICP value of 89.821, which is slightly below the anticipated 90% coverage. Nevertheless, it has the shortest PIs among all the models. Compared to the other classic conformal methods, RF and the locally adaptive RF version deliver similar results, slightly below the expected 90% coverage. The locally adaptive version outperforms the standard CRF, while the QRF is too conservative, with a mean PICP of 93.971. The linear (Ridge) model exceeds the desired 90% coverage with too long PIs. Although CQR RF did not achieve the desired significance level of 90%, it outperformed other methods. However, further fine-tuning of the parameters and cross-validation may help achieve the desired significance level. It is important to note that the study's goal was not to find the absolute best CQR RF model but rather to compare different methods.

Figure 4 shows the summarized results for the one-hour-ahead forecasting horizon. It illustrates the output values from the test dataset, which consists of one-year load data in the form of a box-whisker plot filtered by the hour of the day denoted on the x-axis. The box-whisker plot shows the mean values in each hour and the variability on the test set. The orange lines represent the original PIs aggregated on an hour basis, calculated with the classic split conformal RF method, and the blue lines represent the conformalized, locally adaptive PIs calculated with the proposed CQR RF method. The

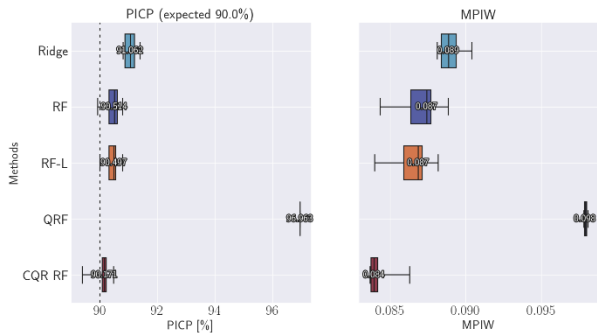


Figure 2: PICP and MPIW for one-hour ahead forecasting horizon, comparing different methods

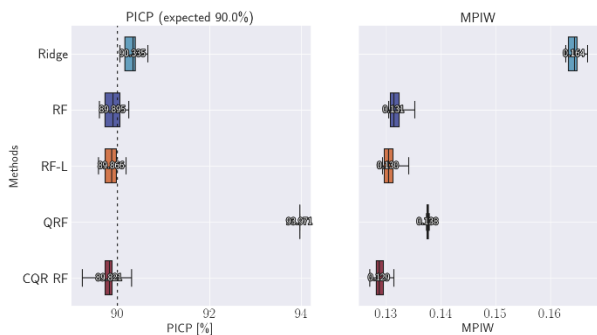


Figure 3: PICP and MPIW for one-day ahead forecasting horizon, comparing different methods

original PIs, shown in orange, are the same length for all the hours in the day, covering 90% PIs, and the conformalized PIs are adaptive to values in that particular hour. In the early hours (from h1 to h7), where the load is lower and has lower variability, the conformalized PIs are shorter in length than the original PIs. From h7, conformalized PIs start to be longer than the original PIs, with higher differences in length during the second valley and the second peak hours (from h18 to h23), which also exhibit high variability (observable in longer boxes). These differences in the length of the original and conformalized PIs not only justify using the CQR RF for the STLF task but also exhibit the strength of the proposed method in terms of recognizing the load uncertainty and variability, which is beneficial in both the planning and operation processes.

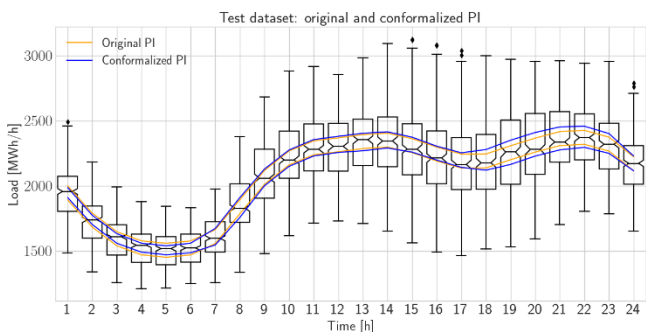


Figure 4: Original vs. conformalized PIs for test dataset, hourly distributions for one-hour ahead forecasting horizon

Table 3: Summarized numerical results

Forecasting horizon		1h			24h		
	Method	PICP [-]	MPIW [-]	Training time [s]	PICP [-]	MPIW [-]	Training time [s]
Classic methods	Ridge	91.0674	0.0890	0.0201	90.3083	0.1644	0.0292
	RF	90.4578	0.0870	5.7671	89.8961	0.1318	7.3429
	RF-L	90.4578	0.0864	37.3149	89.8710	0.1307	51.9612
	QRF	96.9635	0.0979	14.6583	93.9712	0.1376	17.8889
CQR	CQR RF	90.0811	0.0842	21.5642	89.8070	0.1289	21.7158

Table 3 serves as a numerical summary of the experimental results. All the classic and proposed CQR methods are compared over two forecasting horizons using the two conformal scores (PICP and MPIW) and the training time in seconds, which includes both training and the held-out validation where applicable. PICP and MPIW results are discussed in previous paragraphs.

Training time for both forecasting horizons is the lowest for the linear ridge model due to its simplicity. The locally conformalized RF method exhibits the highest training time. The CQR RF method shows significant improvement in terms of lower training time when compared to locally conformalized RF but slightly higher training time than QRF. The standard RF is relatively fast in the model-building phase.

5. CONCLUSION

This paper presents a novel approach to short-term probabilistic load forecasting. Specifically, we propose the use of conformalized quantile regression with random forest as the underlying algorithm. Through this method, we aim to achieve more informative, variable-length PIs to model the variability in load values. By leveraging the strengths of conformal prediction and quantile regression, we have shown that our proposed method provides superior results to classic methods, comparing it over two short-term forecasting horizons. Two standard conformal scoring methods were utilized in the defined case study: PICP, which measures PI coverage, and MPIW, which measures the PI length. The CQR method demonstrates the best results with the narrowest PIs, covering the expected significance levels of 90% and exhibiting improvement in terms of model-building time when compared to other locally-adaptive methods. The CQR method effectively addresses load uncertainty and variability, making it essential for real-time system operation and day-ahead planning. The significance of our findings lies in the potential for our method to improve both system operation and operational planning processes, leading to more efficient and sustainable power system operation.

REFERENCE

- [1] 'The European Green Deal'. Accessed: Sep. 13, 2023. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en
- [2] 'The Paris Agreement | UNFCCC'. Accessed: Sep. 13, 2023. [Online]. Available: <https://unfccc.int/process-and-meetings/the-paris-agreement>
- [3] T. Hong and P. Wang, 'Artificial Intelligence for Load Forecasting: History, Illusions, and Opportunities', *IEEE Power and Energy Magazine*, vol. 20, no. 3, pp. 14–23, May 2022, doi: 10.1109/MPE.2022.3150808.

- [4] T. Hong and S. Fan, 'Probabilistic electric load forecasting: A tutorial review', *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016, doi: 10.1016/j.ijforecast.2015.11.011.
- [5] T. Hong, 'Short Term Electric Load Forecasting', Dissertation, University of North Carolina, Charlotte, 2010.
- [6] Y. Romano, E. Patterson, and E. J. Candès, 'Conformalized Quantile Regression'. arXiv, May 08, 2019. Accessed: Sep. 13, 2023. [Online]. Available: <http://arxiv.org/abs/1905.03222>
- [7] H. Wen, J. Gu, J. Ma, L. Yuan, and Z. Jin, 'Probabilistic Load Forecasting via Neural Basis Expansion Model Based Prediction Intervals', *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3648–3660, Jul. 2021, doi: 10.1109/TSG.2021.3066567.
- [8] 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, doi: 10.1109/TNNLS.2022.3217694.
- [9] 'python-holidays — holidays documentation'. Accessed: Jun. 16, 2022. [Online]. Available: <https://python-holidays.readthedocs.io/en/latest/>
- [10] 'Meteostat Developers'. Accessed: Sep. 13, 2023. [Online]. Available: <https://dev.meteostat.net/>
- [11] 'Project Jupyter'. Accessed: Jun. 17, 2022. [Online]. Available: <https://jupyter.org>
- [12] 'NumPy'. Accessed: Jun. 17, 2022. [Online]. Available: <https://numpy.org/>
- [13] 'pandas - Python Data Analysis Library'. Accessed: Jun. 17, 2022. [Online]. Available: <https://pandas.pydata.org/>
- [14] M. Waskom, 'seaborn: statistical data visualization', *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/joss.03021.
- [15] 'Matplotlib — Visualization with Python'. Accessed: Jun. 17, 2022. [Online]. Available: <https://matplotlib.org/>
- [16] 'scikit-learn: machine learning in Python — scikit-learn 1.1.1 documentation'. Accessed: Jun. 16, 2022. [Online]. Available: <https://scikit-learn.org/stable/index.html>
- [17] H. Linusson, 'nonconformist'. Accessed: Sep. 13, 2023. [Online]. Available: <https://github.com/donlnz/nonconformist>
- [18] 'quantile-forest'. Zillow. Accessed: Sep. 13, 2023. [Online]. Available: <https://github.com/zillow/quantile-forest>
- [19] 'yromano - Repositories', GitHub. Accessed: Sep. 13, 2023. [Online]. Available: <https://github.com/yromano>