Hierarchical Bayesian Tree-Weighted Light Gradient Boosting for Probabilistic Long-Term Conventional Load Forecasting

Quang-Manh Bui
ECE department
Clarkson University
Potsdam, NY, USA
buimq@clarkson.edu

Le Minh Vu
ECE department
Clarkson University
Potsdam, NY, USA
levu@clarkson.edu

Thomas H. Ortmeyer
ECE department
Clarkson University
Potsdam, NY, USA
tortmeye@clarkson.edu

Tony Vu
ECE department
Clarkson University
Potsdam, NY, USA
tvu@clarkson.edu

Jianhua Zhang
ECE department
Clarkson University
Potsdam, NY, USA
jzhang@clarkson.edu

***Abstract*—Accurate long-term load forecasting plays a vital role in strategic planning, resource adequacy assessment, and system reliability evaluation of modern power grids when electricity demand rises and uncertainties from renewable integration grow. However, existing deterministic forecasting techniques often fail to quantify predictive uncertainty, while many existing probabilistic methods are lack robust calibration over long horizons. This paper proposes a Hierarchical Bayesian Tree-Weighted LightGBM (BTW-LGBM) framework that preserves the accuracy and efficiency of LightGBM while adding a Bayesian layer on tree weights and a feature-dependent variance model. This combination yields closed-form prediction intervals and a decomposition of forecast uncertainty into model and data components. Using the GEFCom2014 dataset, BTW-LGBM significantly reduced the pinball loss and Winkler score compared to conventional neural network models. The model was also applied to all eleven load zones of the New York Independent System Operator (NYISO) to captures both zonal diversity and long-term uncertainty dynamics. The results highlight both the methodological superiority of BTW-LGBM and its practical value for utility-scale long-term load forecasting applications.**

***Index Terms*—Long-term load forecasting, Probabilistic load forecasting, Gradient boosting, Bayesian tree-weighting.**

I. Introduction

Electric power systems face rapid load growth driven by electrification and energy-intensive computing. NERC's latest Summer Reliability Assessment reports that peak demand across North America's 23 assessment areas increased by more than 10 GW between 2024 and 2025, more than twice the increase in the previous year [1]. This combination of rising demand, renewable integration, and market uncertainty increases the need for long term load forecasting methods that predict central trends and quantify uncertainty.

Long term load forecasting over horizons of months to years differs from short term operational forecasting and is better framed as a regression problem that uses explanatory variables such as calendar effects, weather, and socio-economic indicators rather than simple time series extrapolation. Compared with short- and mid-term horizons, long-term load forecasting is more challenging because structural changes in technology, policy, and the economy accumulate over time, so historical relationships become less stable and model uncertainty grows [2]. Over the last decade, deterministic machine learning models such as gradient boosting decision trees and LightGBM, together with sequence models such as LSTMs and attention based networks, have achieved very good point forecast accuracy for both short term and long term load tasks [3]. These models, however, produce only single trajectories and do not provide calibrated measures of uncertainty, which is problematic for planning under deep demand uncertainty.

To address this limitation, probabilistic load forecasting methods that generate prediction intervals or full predictive distributions have been widely studied. Quantile regression based approaches, including quantile LightGBM and advanced neural models, are now commonly used to construct prediction intervals and probabilistic forecasts [4]. However, quantile LightGBM trains separate models for each quantile and lacks a joint probabilistic structure, so the resulting intervals can be miscalibrated and mutually inconsistent, whereas a Bayesian Tree-Weighted formulation models the full conditional distribution within a single coherent layer. Recent reviews also highlight persistent issues with interval calibration, static coverage, and limited interpretability for quantile-based methods, particularly under distribution shifts and heteroscedastic conditions that are common in energy systems [4]. Bayesian Additive Regression Trees and related Bayesian ensemble methods offer a principled way to quantify uncertainty and provide posterior predictive distributions and credible intervals. Their reliance on computationally intensive Markov Chain Monte Carlo inference, however, restricts scalability and limits use in large scale or near real time applications [5]. Sample Reweighting methods can adapt models to nonstationary data by learning per sample weights, but they do not produce explicit probabilistic outputs or calibrated prediction intervals [6]. As a result, no existing approach provides a scalable, interpretable, and theoretically grounded Bayesian treatment of tree ensembles that delivers closed form, calibrated prediction intervals suitable for long term power system planning.

This paper addresses these challenges by proposing the BTW-LGBM framework with the major contributions as follows:

1) Developing a novel probabilistic forecasting model that augments LightGBM with a hierarchical Gaussian prior

979-8-3315-8138-1

over tree leaf contributions. This transforms deterministic boosting into a fully Bayesian predictive model.

2) Conducting extensive validation on benchmark via GEFCom2014 [7] dataset and utility-scale NYISO case.

II. Proposed Methodology

This section introduces the proposed BTW-LGBM framework in Fig. 1 and the tailored Bayesian tree-weighting layer in Algorithm. 1. The method starts from a deterministic LightGBM base model and augments it with a Bayesian tree-weighting layer, a heteroscedastic variance head, and heavy-tailed predictive modeling to produce prediction intervals with an explicit decomposition of model and noise uncertainty.

A. Problem Formulation

Let $\{(x_t, y_t)\}_{t=1}^{N}$ denote the hourly historical load data, where $y_t \in \mathbb{R}$ is the integrated load at time $t \in \{1, 2, \dots, N\}$ and $\boldsymbol{x}_t \in \mathbb{R}^d$ is the corresponding feature vector including calendar and weather variables. The goal is to learn a probabilistic forecasting model that, for each future time t, provides a predictive distribution $p(y_t \mid \boldsymbol{x}_t, \mathcal{D}_{\text{train}})$, from which point forecasts $\hat{y}_t$ and central prediction intervals can be derived. $\mathcal{D}_{\text{train}}$ denotes the subset of historical data used for model estimation. The proposed BTW-LGBM framework is designed to: (i) capture nonlinear load–weather–calendar interactions via tree-based boosting, (ii) regularize tree contributions through a hierarchical first-order autoregressive prior across trees, (iii) account for heteroscedastic residual variance via a separate learned variance head, and (iv) ensure marginal coverage of the prediction intervals through Bayesian posterior inference.

B. Deterministic LightGBM Base Model

The first stage of BTW-LGBM is a deterministic LightGBM regressor that maps the feature vector $\boldsymbol{x}_t$ to a point prediction of load. Let $F(\boldsymbol{x}_t)$ denote the fitted LightGBM model: $F(\boldsymbol{x}_t) = \sum_{m=1}^{M} f_m(\boldsymbol{x}_t)$, where $f_m(\cdot)$ is the m-th tree and M is the number of boosting iterations. The LightGBM model is trained with

$$\min_{F \in \mathcal{F}} \sum_{t \in \mathcal{D}_{\text{train}}} (y_t - F(\boldsymbol{x}_t))^2 + \Omega(F), \tag{1}$$

with standard regularization Ω on leaf values. After training, this deterministic model provides a strong nonlinear basis for load forecasting. However, it lacks a probabilistic interpretation and does not separate model uncertainty from residual noise. The proposed approach addresses this by introducing a hierarchical Bayesian layer on top of the tree ensemble.

C. Bayesian Tree-Weighting Layer

1) Tree basis representation: For each input $\boldsymbol{x}_t$, the trained LightGBM model produces a vector of per-tree contributions, $\boldsymbol{\phi}(\boldsymbol{x}_t) = \left[f_1(\boldsymbol{x}_t), f_2(\boldsymbol{x}_t), \dots, f_M(\boldsymbol{x}_t)\right]^\top \in \mathbb{R}^M$. Collecting these across all training samples yields the matrix

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{x}_1)^\top \\ \boldsymbol{\phi}(\boldsymbol{x}_2)^\top \\ \vdots \\ \boldsymbol{\phi}(\boldsymbol{x}_N)^\top \end{bmatrix} \in \mathbb{R}^{N \times M}. \tag{2}$$

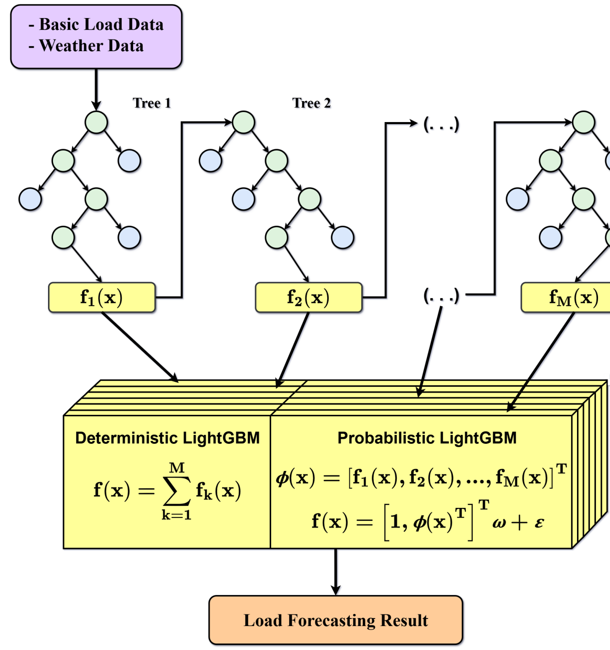

Fig. 1. Proposed BTW-LGBM Framework: Deterministic LightGBM Base Model and Bayesian Tree-Weights Layer.

An intercept term is appended, forming $\tilde{\boldsymbol{\Phi}} = \left[\mathbf{1}, \boldsymbol{\Phi}\right] \in \mathbb{R}^{N \times (M+1)}$, and the predictive model is a linear combination:

$$y_t = \tilde{\boldsymbol{\phi}}(\boldsymbol{x}_t)^\top \boldsymbol{\omega} + \varepsilon_t, \quad \tilde{\boldsymbol{\phi}}(\boldsymbol{x}_t) = [1, \boldsymbol{\phi}(\boldsymbol{x}_t)^\top]^\top, \tag{3}$$

where $\boldsymbol{\omega} \in \mathbb{R}^{M+1}$ is the tree weights vector and ε_t is a noise.

2) Hierarchical AR(1) prior across trees: To smooth contributions across successive trees and prevent overfitting, a hierarchical Gaussian prior is imposed on $\boldsymbol{\omega}$. The intercept $\boldsymbol{\omega}_0$ receives a weak ridge prior with small precision, while the remaining M tree weights $\boldsymbol{\omega}_{1:M}$ follow an AR(1) prior:

$$p(\boldsymbol{\omega}_{1:M}) \propto \exp\left(-\frac{\lambda}{2} \sum_{m=2}^{M} (\boldsymbol{\omega}_m - \rho \boldsymbol{\omega}_{m-1})^2\right), \tag{4}$$

where $\lambda > 0$ is a base penalty and $\rho \in (0, 1)$ controls the strength of autocorrelation between adjacent trees. The summation starts at $m = 2$ because an AR(1) prior penalizes the deviation of each weight $\boldsymbol{\omega}_m$ from its predecessor $\boldsymbol{\omega}_{m-1}$, and no such predecessor exists for $\boldsymbol{\omega}_1$. This prior introduces a Gaussian precision (inverse covariance) matrix $\boldsymbol{Q}_{\text{prior}} = \frac{1}{\tau^2} \begin{bmatrix} \lambda_0 & 0 \\ 0 & \boldsymbol{Q}_{\text{AR1}} \end{bmatrix}$, where $\tau > 0$ is a global scale parameter controlling the overall prior variance of the weights, λ_0 is a small ridge penalty for the intercept and $\boldsymbol{Q}_{\text{AR1}} \in \mathbb{R}^{M \times M}$ is a tridiagonal matrix. This hierarchical prior captures correlation among trees and stabilizes long-horizon extrapolation.

3) Heteroscedastic variance head: Residual analysis of the initial ridge fit reveals that the load forecast errors are heteroscedastic, with larger variance during extreme weather and peak hours. To capture this, BTW-LGBM employs a separate variance head that models the conditional noise variance $\sigma^2(\boldsymbol{x}_t)$ as a function of a subset of drivers $\boldsymbol{z}_t$

979-8-3315-8138-1

Algorithm 1 Bayesian Tree-Weighting Layer in BTW-LGBM

Require: Training data $\mathcal{D}_{\text{train}} = \{(\boldsymbol{x}_t, y_t)\}$, fitted LightGBM base model F, heteroscedastic variances $\hat{\sigma}^2(\boldsymbol{x}_t)$, AR(1) hyperparameters $(\lambda, \rho, \tau, \lambda_0)$.
Ensure: Posterior mean $\boldsymbol{\mu}_{\text{post}}$ and covariance $\boldsymbol{\Sigma}_{\text{post}}$.

1. Tree-basis extraction
for each training sample t **do**
  Evaluate $F(\boldsymbol{x}_t)$ and collect $f_m(\boldsymbol{x}_t)$ per-tree.
end for
Construct the matrix $\boldsymbol{\Phi}$ and augmented $\tilde{\boldsymbol{\Phi}}$.
2. Hierarchical AR(1) prior construction
Build the precision matrix $\boldsymbol{Q}_{\text{prior}}$ for the tree weights.
3. Heteroscedastic likelihood assembly
Form the diagonal noise covariance $\boldsymbol{\Sigma}_{\varepsilon}$.
Compute posterior precision $\boldsymbol{A}$ and vector $\boldsymbol{b}$.
4. Posterior inference
Solve $\boldsymbol{A}\boldsymbol{\mu}_{\text{post}} = \boldsymbol{b}$.
Compute posterior covariance $\boldsymbol{\Sigma}_{\text{post}}$.
return $(\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}})$.

comprising weather features and calendar indicators. Let $r_t = y_t - F(\boldsymbol{x}_t)$ denote residuals from an initial ridge fit without heteroscedasticity. The variance head is trained on $v_t = \log\left(r_t^2 + \epsilon\right)$, where ϵ is a small constant. The variance head is a separate LightGBM regressor that is trained only to model the residual variance. A LightGBM regressor $g(\cdot)$ is then fitted: $v_t \approx g(\boldsymbol{z}_t; \theta)$, and the conditional variance is recovered as $\hat{\sigma}^2(\boldsymbol{x}_t) = \exp\left(g(\boldsymbol{z}_t; \hat{\theta})\right)$. where θ denotes the internal parameters of the variance model and $\hat{\theta}$ denotes the parameter estimates after fitting the model on training data. By restricting the variance head to weather features only, the model avoids overfitting while still capturing key patterns.

4) Bayesian posterior inference: Given the design matrix $\tilde{\boldsymbol{\Phi}}$, the prior precision matrix $\boldsymbol{Q}_{\text{prior}}$ and the estimated noise variances $\hat{\sigma}^2(\boldsymbol{x}_t)$, the conditional likelihood is

$$y_t \mid \boldsymbol{\omega} \sim \mathcal{N}\left(\tilde{\boldsymbol{\phi}}(\boldsymbol{x}_t)^\top \boldsymbol{\omega},\ \hat{\sigma}^2(\boldsymbol{x}_t)\right), \tag{5}$$

which yields a Gaussian posterior over $\boldsymbol{\omega}$. Let $\boldsymbol{\Sigma}_{\varepsilon} = \operatorname{diag}\left(\hat{\sigma}^2(\boldsymbol{x}_t)\right)$ denote the diagonal noise covariance matrix. The posterior precision and mean are

$$\boldsymbol{A} = \tilde{\boldsymbol{\Phi}}^\top \boldsymbol{\Sigma}_{\varepsilon}^{-1} \tilde{\boldsymbol{\Phi}} + \boldsymbol{Q}_{\text{prior}}, \tag{6}$$

$$\boldsymbol{b} = \tilde{\boldsymbol{\Phi}}^\top \boldsymbol{\Sigma}_{\varepsilon}^{-1} \boldsymbol{y}, \tag{7}$$

$$\boldsymbol{\mu}_{\text{post}} = \boldsymbol{A}^{-1}\boldsymbol{b}, \tag{8}$$

$$\boldsymbol{\Sigma}_{\text{post}} = \boldsymbol{A}^{-1}, \tag{9}$$

where $\boldsymbol{y}$ is the vector of training loads $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^\top$. Let $\boldsymbol{x}_*$ denote a new test-set feature vector. The predictive mean and variance under the Gaussian model are

$$\hat{\mu}(\boldsymbol{x}_*) = \tilde{\boldsymbol{\phi}}(\boldsymbol{x}_*)^\top \boldsymbol{\mu}_{\text{post}}, \tag{10}$$

$$\hat{v}_{\text{model}}(\boldsymbol{x}_*) = \tilde{\boldsymbol{\phi}}(\boldsymbol{x}_*)^\top \boldsymbol{\Sigma}_{\text{post}} \tilde{\boldsymbol{\phi}}(\boldsymbol{x}_*), \tag{11}$$

$$\hat{v}_{\text{noise}}(\boldsymbol{x}_*) = \hat{\sigma}^2(\boldsymbol{x}_*), \tag{12}$$

$$\hat{v}_{\text{tot}}(\boldsymbol{x}_*) = \hat{v}_{\text{model}}(\boldsymbol{x}_*) + \hat{v}_{\text{noise}}(\boldsymbol{x}_*). \tag{13}$$

The total variance decomposes into a model-uncertainty component $\hat{v}_{\text{model}}$ and a noise component $\hat{v}_{\text{noise}}$, enabling an explicit uncertainty decomposition.

III. Case Studies

To comprehensively evaluate the performance and real-world applicability of the proposed model, two complementary sets of case studies are conducted. The first study benchmarks the proposed model on the widely used GEFCom2014 load forecasting dataset [7], enabling direct and fair comparison against state-of-the-art probabilistic load forecasting methods reported in recent literature, including the sample-reweighting approaches in [6]. While the GEFCom2014 tasks provide a controlled environment for cross-model comparison, such single-zone synthetic load dataset does not fully represent the operational complexity of large interconnected power systems. Therefore, a more comprehensive real-world case study is conducted using 11 load zones of NYISO. All experiments were executed on a workstation with Intel i7-13700 CPU model and 64 GB RAM memory.

A. Validation Study on the GEFCom2014 Load Dataset

The GEFCom2014 dataset consists of hourly load and 25 meteorological temperature features for the years 2011–2014, enabling reproducible benchmarking against state-of-the-art probabilistic forecasting methods. Following the paper [6], the dataset is divided similarly for training, validation, and testing with the data from 2011 to 2012, from 2013, and from 2014, respectively. To benchmark the proposed method, this study generates nine quantiles ($q = 0.1$–0.9) for probabilistic load forecasting and adopts the same probabilistic metrics including Average Pinball Loss (APL) and Winkler Score (WS). Lower values indicate better performance for both APL and WS.
Table I summarizes the probabilistic load forecasting performance of existing models and the proposed BTW-LGBM using the GEFCom2014 dataset. The proposed model achieves a substantial performance leap across all three metrics. The average pinball loss is reduced to 4.846, which corresponds to an improvement of approximately $\approx 86\%$ relative to the

TABLE I
Comparison of Probabilistic Forecasting Performance on the GEFCom2014 Dataset

Method	**Pinball Loss**	**WS (80% PI)**	**WS (60% PI)**
LR	50.280	609.763	462.990
LR+SR	50.229	607.124	462.920
ANN	38.658	467.163	354.953
ANN+SR	34.771	416.573	319.439
LSTM	36.072	439.336	329.696
LSTM+SR	35.129	413.291	319.917
GBRT	37.223	446.959	338.857
BTW-LGBM	**4.846**	**88.166**	**47.699**

PI: Prediction Interval, LR: Linear Regression, ANN: Artificial Neural Network, LSTM: Long Short Term Memory, GBRT: Gradient Boosting Regression Trees, and SR: variants enhanced by Sample Reweighting [6].

979-8-3315-8138-1

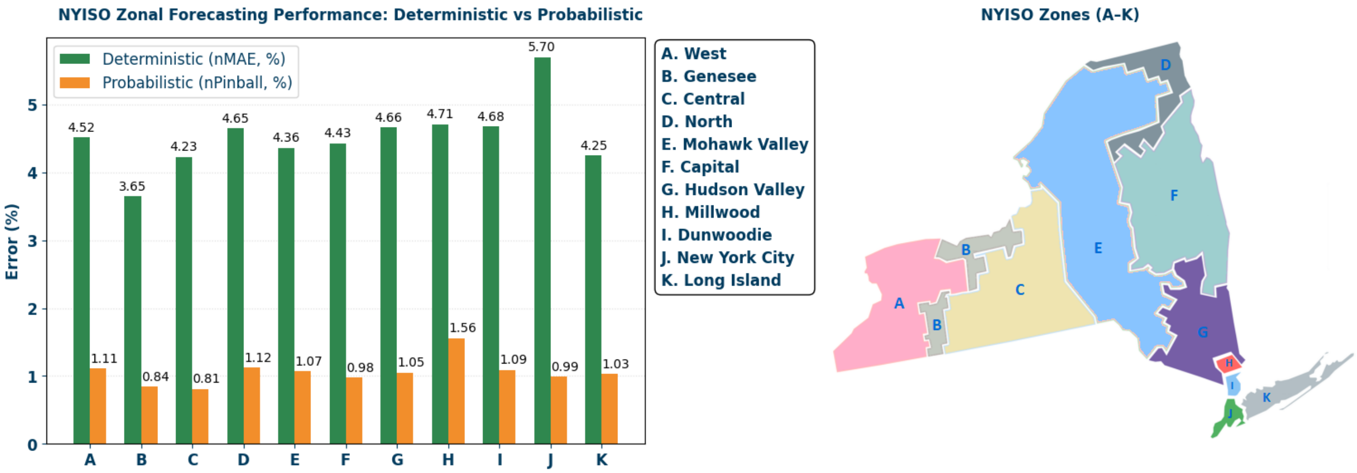

Fig. 2. NYISO zonal forecasting performance: deterministic mean prediction error (nMAE) and probabilistic normalized pinball loss (nPinball).

best baseline (ANN+SR). This drastic reduction indicates that the hierarchical Bayesian tree-weighting framework provides significantly sharper and better calibrated conditional distributions over the nine evaluation quantiles. A similar pattern is observed for the interval-based metrics. For the 80% prediction interval, the proposed method achieves a Winkler score of 88.166, representing about $\approx 78.7\%$ improvement over the best baseline (LSTM+SR) whereas the Winkler score drops from 319.439 (ANN+SR) to 47.699 with BTW-LGBM for the 60% interval. The simultaneous reduction in pinball loss and Winkler scores suggests that the proposed model is not merely producing narrow intervals, but is also improving reliable coverage. This validation results motivated the proposed model's application to more complex systems such as the NYISO case.

B. Application of BTW-LGBM to the NYISO Case Study

To assess the generalization of the proposed framework, a comprehensive zonal case study was performed across all eleven load zones of NYISO. Each zone represents a distinct control region that differs in generation mix, climatic conditions, and load composition, ranging from hydro-dominant upstate areas (Zones A–C) to the densely populated metropolitan zones (Zones J and K). Hourly conventional load data from 2001 to 2020, from 2020 to 2021, and from 2021 to 2024 were used for model training, validation, and testing, respectively. Weather data input is treated as observed during the test period to isolate the forecasting performance of the proposed model from uncertainty in upstream weather forecasts.

For the NYISO 11-zone study, models were trained independently per zone, and execution was timed as the average per-sample prediction time on the test set. The average training time per zone was five minutes, and the average execution time was thirty seconds per yearly horizon forecast. Key hyperparameters of the proposed model, including the AR(1) correlation coefficient ρ, the prior precision parameter λ, and the prior scale τ, were selected using a validation-based strategy within the training period. In particular, ρ controls the smoothness of tree-weight evolution across boosting iterations and was set to $\rho = 0.8$ to provide strong but not overly restrictive temporal regularization between consecutive trees. The prior precision λ was fixed at 10^{-4} to balance shrinkage and flexibility of the tree ensemble, while avoiding excessive damping of informative tree contributions. The prior scale τ was defined as a multiple of the empirical standard deviation of the training load, $\tau = 1.5\,\mathrm{std}(y)$, which adaptively normalizes the prior strength to the magnitude of the target variable. Two key evaluation metrics were employed including the normalized Mean Absolute Error (nMAE) for deterministic accuracy and the normalized Pinball Loss (nPinball) for probabilistic calibration. The testing results are shown in Fig. 2.

The deterministic forecasting accuracy of the proposed model, evaluated through nMAE, remained consistently robust across all NYISO zones, with errors ranging between 3.65% and 5.70%. Zones B (3.65%) and C (4.23%) exhibited the lowest nMAE values, reflecting smoother load patterns. By contrast, Zone J (NYC) recorded the highest nMAE of 5.70%, attributed to the complex consumption behavior, and sensitivity to temperature extremes typical of metropolitan regions. The moderate nMAE observed in Zones E through G suggests that the model effectively preserved region-specific dynamics while maintaining model stability across varying grid topologies. This strong deterministic forecasting performance can be attributed to the hierarchical AR(1) prior which reduces overfitting to noise in the historical load data and produces a more reliable mean prediction. In terms of probabilistic accuracy, the nPinball loss ranged from 0.81% to 1.56%, confirming the strong calibration of the predictive intervals. The lowest pinball losses were observed in Zones C (0.81%) and B (0.84%), signifying precise uncertainty quantification in these zones. Even in zones with more dynamic consumption profiles, such as Zones J and K, the probabilistic scores remained below 1.1%, demonstrating the model's capacity to maintain calibration under higher volatility. Overall, the proposed model delivers well-calibrated uncertainty quantification (low nPinball) and a reliable median forecast (low nMAE), demonstrating that it can jointly enhance the representation of uncertainty and the stability of the mean prediction.

Fig. 3 presents the testing-period forecasting results for Zone J

979-8-3315-8138-1

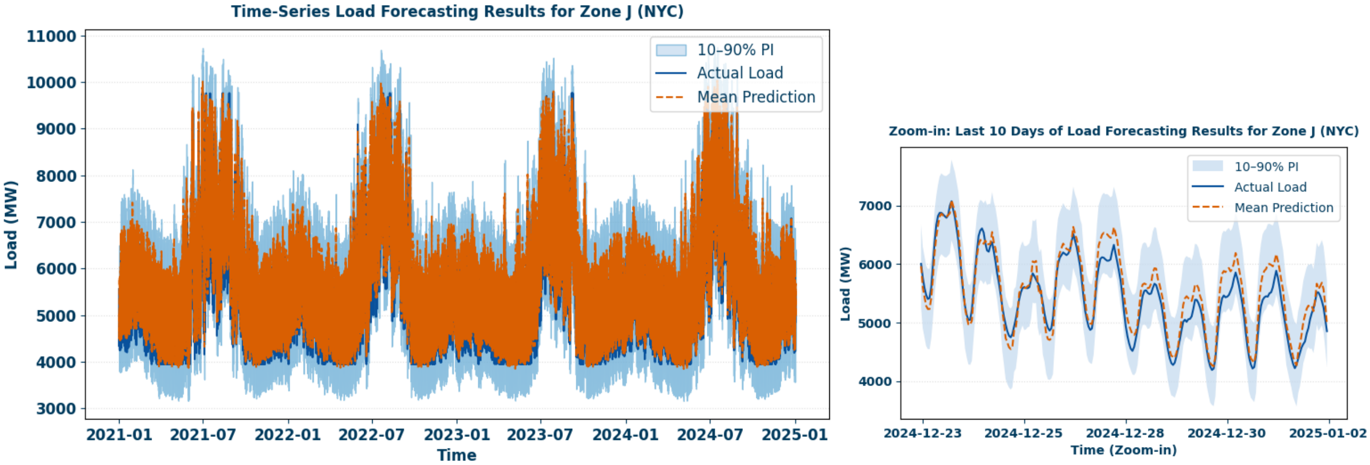

Fig. 3. Testing-period probabilistic load forecasting results for Zone J (New York City): the left plot shows the full multi-year forecast and the right plot presents the last 10-day zoomed-in view.

(NYC), including both the long-horizon series (left) and the last 10-day zoomed-in view (right). The lower bound 10th-percentile and upper bound 90th-percentile enclose the range within which the true load value is expected to fall with approximately 80% probability. The width of this interval reflects forecast uncertainty: it widens during highly variable load periods and narrows when the system behavior is more stable. To assess the calibration quality of the proposed probabilistic model, an empirical coverage analysis was conducted for the prediction intervals. In particular, the 10th–90th percentile prediction interval (nominal 80% coverage) was evaluated using a held-out validation set under a split-conformal calibration framework. The resulting empirical coverage was approximately 80.02%, closely matching the nominal target and indicating good probabilistic calibration. Similar coverage behavior was observed on the test set across NYISO zones, supporting the reliability of the probabilistic outputs for long-term planning applications. From a high-level perspective, the model successfully captures the dominant seasonal and load-patterns characteristic of NYC. The prediction interval closely tracks these large-scale variations, widening during summer peaks when load becomes more sensitive to extreme temperatures, humidity, and urban heat-island effects. Conversely, the PI narrows during milder conditions, especially in summer/fall transitions, reflecting lower uncertainty in those periods. The zoomed-in plot provides further insight into short-term alignment between the model and actual load. Across the last 10-day horizon, the mean forecast does not simply track the center of the training distribution, but it dynamically adapts to day-to-day load variability. From a practical standpoint, the uncertainty bands resulting from the proposed BTW-LGBM can be directly used as a planning envelope for resource adequacy studies, while the upper bound can be used to stress-test capacity margins and operating reserve requirements.

IV. Conclusion

This research introduced a novel framework for probabilistic long-term conventional load forecasting. Two comprehensive case studies confirmed the effectiveness and generalization capability of the proposed approach. On the GEFCom2014 benchmark, BTW-LGBM achieved substantial improvements in probabilistic metrics, outperforming traditional regression and neural network baselines. When extended to the NYISO zonal forecasting study, the model consistently captured spatial and temporal dependencies under varying grid conditions.

Acknowledgment

This work was supported by the U.S. Department of Energy (DOE) under the OPTIMA program, Award No. DE-EE0011369, "ADMIRE-GridPlan: Advanced Methods for Integrating Renewables in Grid Planning." The Clarkson University team gratefully acknowledges the project leadership of Florida International University and the collaboration and support of Sandia National Laboratories, the University of North Carolina at Charlotte, GE Energy Consulting, the New York Power Authority, and INS Engineering.

References

[1] North American Electric Reliability Corporation, "2025 summer reliability assessment," North American Electric Reliability Corporation (NERC), Tech. Rep., May 2025.
[2] K. B. Walid and Y. iang, "Short-term load forecasting with advanced machine learning approaches: Spatial variability and model performance," in *2024 56th North American Power Symposium (NAPS)*, 2024, pp. 1–6.
[3] M. Hasan, Z. Mifta, S. J. Papiya, P. Roy, P. Dey, N. A. Salsabil, N.-U.-R. Chowdhury, and O. Farrok, "A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications," *Energy Conversion and Management: X*, Apr. 2025.
[4] D. Kaur, S. N. Islam, M. A. Mahmud, M. E. Haque, and Z. Y. Dong, "Energy forecasting in smart grid systems: recent advancements in probabilistic deep learning," *IET Generation, Transmission & Distribution*, Nov. 2022.
[5] R. Bai, M. R. Boland, and Y. Chen, "Scalable high-dimensional Bayesian varying coefficient models with unknown within-subject covariance," *Journal of Machine Learning Research*, no. 259, pp. 1–49, 2023.
[6] C. Wang, Y. Zhou, Q. Wen, and Y. Wang, "Improving Load Forecasting Performance via Sample Reweighting," *IEEE Transactions on Smart Grid*, vol. 14, no. 4, pp. 3317–3320, Jul. 2023.
[7] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," pp. 896–913, 2016.

979-8-3315-8138-1