Day-Ahead Electricity Price Forecasting for ERCOT Using Gradient Boosting with Weather Covariates

Graydon Snider
Department of Operations Research
sniderg.mcgill@gmail.com

February 8, 2026

Abstract

Accurate day-ahead electricity price forecasting is essential for market participants in deregulated power markets. This paper presents a machine learning approach for forecasting day-ahead locational marginal prices (LMPs) in the Electric Reliability Council of Texas (ERCOT) market. We develop a LightGBM gradient boosting model incorporating temporal features, autoregressive lag structures, and exogenous weather covariates from the Open-Meteo API. Using an expanding window cross-validation on **six years** of historical data (2020–2026), our model achieves an average mean absolute error (MAE) of **$9.35/MWh** for the HB_NORTH pricing hub during the 2025 test period. We extend the model to provide probabilistic forecasts via quantile regression, achieving 67.3% coverage for 80% prediction intervals. Feature importance analysis reveals that recent price lags and temperature are the dominant predictors. The model demonstrates robust performance even when trained on extreme volatility periods like Winter Storm Uri.

Keywords: Electricity price forecasting, ERCOT, LightGBM, machine learning, weather covariates, day-ahead market

1 Introduction

Electricity markets in deregulated environments exhibit complex price dynamics driven by supply-demand imbalances, fuel costs, transmission constraints, and weather conditions. The Electric Reliability Council of Texas (ERCOT) operates one of the largest competitive electricity markets in the United States, serving approximately 90% of the state's electric load [Electric Reliability Council of Texas, 2024].

Day-ahead price forecasting is critical for multiple market participants:

- **Generators**: Optimize bidding strategies and commitment decisions
- **Load-serving entities**: Hedge procurement costs and manage risk
- **Traders**: Identify arbitrage opportunities between day-ahead and real-time markets
- **Storage operators**: Optimize charge/discharge schedules

Traditional approaches to electricity price forecasting include time series models (ARIMA, SARIMA), regime-switching models, and fundamental supply-demand simulations [Weron, 2014]. More recently, machine learning methods—particularly gradient boosting and deep learning—have demonstrated superior performance in capturing the nonlinear relationships inherent in electricity prices [Lago et al., 2021].

This paper contributes to the literature by:

1. Developing an end-to-end forecasting pipeline for ERCOT day-ahead prices using open-source tools

2. Demonstrating the value of incorporating free weather API data as exogenous covariates

3. Providing extensive cross-validation analysis across different market conditions

4. Quantifying model performance by hour-of-day and month

2 Data

2.1 Study Region

We focus on the HB_NORTH pricing hub, which represents the weighted average price across settlement points in the Dallas-Fort Worth metropolitan area (Figure 1). This region accounts for significant load due to commercial and residential demand in one of the fastest-growing metropolitan areas in the United States.

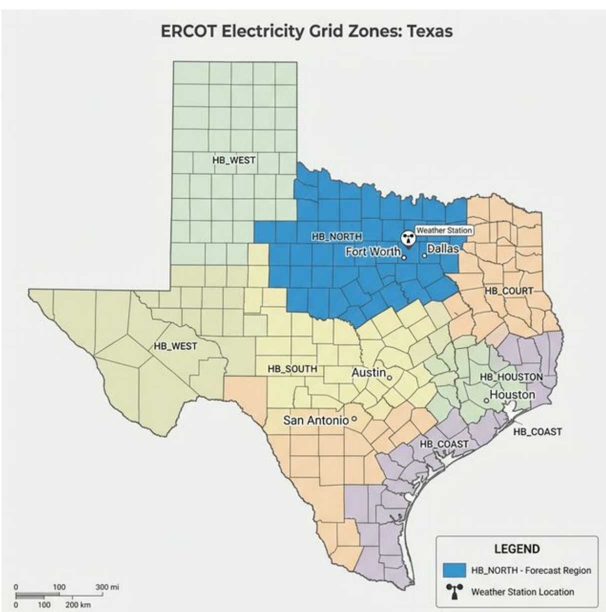

Figure 1: ERCOT electricity grid zones with HB_NORTH highlighted (blue). Weather observations are collected from a station near Dallas (32.78°N, 96.80°W).

2.2 Price Data

Day-ahead settlement point prices (SPPs) were obtained from ERCOT's public data archives for the period **January 2020 through February 2026**, yielding approximately **53,544** hourly observations. Table 1 summarizes key statistics.

Table 1: Summary Statistics: HB_NORTH Day-Ahead Prices

Statistic	Value
Observations	53,544
Mean	\$58.29/MWh
Median	\$24.68/MWh
Std. Deviation	\$377.93/MWh
Minimum	-\$4.00/MWh
Maximum	\$8,998.99/MWh
Skewness	19.3

The price distribution exhibits heavy right-tail behavior characteristic of electricity markets, with extreme spikes during scarcity events. Negative prices occur during periods of high renewable generation and low demand.

2.3 Weather Data

Hourly weather observations were obtained from the Open-Meteo Historical Weather API [Open-Meteo, 2024], which provides free access to ERA5 reanalysis data. The following variables were collected for a location representative of the HB_NORTH zone:

- Temperature at 2m (T, °C)
- Relative humidity (%)
- Wind speed at 10m (km/h)
- Wind gusts at 10m (km/h)
- Shortwave radiation (W/m^2)
- Cloud cover (%)

Weather conditions directly influence electricity demand through heating/cooling loads and supply through renewable generation (wind and solar).

3 Methods

3.1 Feature Engineering

Effective forecasting of electricity prices requires capturing multiple temporal patterns. We construct three categories of features:

3.1.1 Lag Features (Autoregressive)

Price persistence is a dominant characteristic of electricity markets. We include lagged values at horizons corresponding to:

$$y_{t-24} : \text{Same hour yesterday} \tag{1}$$
$$y_{t-48} : \text{Same hour 2 days ago} \tag{2}$$
$$y_{t-168} : \text{Same hour last week} \tag{3}$$

3.1.2 Rolling Statistics

To capture local trends and volatility, we compute a 24-hour rolling mean $\bar{y}_{t,24}$ and rolling standard deviation $\sigma_{t,24}$:

$$\bar{y}_{t,24} = \frac{1}{24}\sum_{i=1}^{24} y_{t-24-i} \tag{4}$$

$$\sigma_{t,24} = \sqrt{\frac{1}{24}\sum_{i=1}^{24}(y_{t-24-i} - \bar{y}_{t,24})^2} \tag{5}$$

3.1.3 Calendar Features

Electricity demand follows strong intraday and weekly patterns:

- Hour of day: $h \in \{0, 1, \ldots, 23\}$
- Day of week: $d \in \{0, 1, \ldots, 6\}$
- Month: $m \in \{1, 2, \ldots, 12\}$

3.1.4 Weather Covariates

Weather features enter the model as exogenous regressors:

$$\mathbf{x}_t^{\text{weather}} = [T_t, H_t, W_t, G_t, S_t, C_t]^\top \tag{6}$$

where T is temperature, H is humidity, W is wind speed, G is wind gusts, S is solar radiation, and C is cloud cover.

3.2 Model Architecture

We employ LightGBM [Ke et al., 2017], a gradient boosting decision tree (GBDT) algorithm optimized for efficiency and accuracy. The model minimizes the mean absolute error loss:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{7}$$

Key hyperparameters were tuned using Optuna [Akiba et al., 2019] with 20 trials and TPE sampling:

- Number of estimators: 425
- Learning rate: 0.1313
- Maximum depth: 10
- Number of leaves: 80
- Regularization (α, λ): 0.0006, 0.0002

Hyperparameter tuning reduced MAE by 8.4% compared to default parameters (30-day CV).

The full feature vector at time t is:

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{x}_t^{\text{lag}} \\ \mathbf{x}_t^{\text{roll}} \\ \mathbf{x}_t^{\text{cal}} \\ \mathbf{x}_t^{\text{wx}} \end{bmatrix} \tag{8}$$

3.3 ArcSinh Transformation and Mean Correction

Electricity prices exhibit high skewness and can be negative, posing challenges for standard logarithmic transformations. We apply the inverse hyperbolic sine (ArcSinh) transformation, $z_t = \operatorname{asinh}(y_t)$, which naturally handles negative values and effectively normalizes the distribution compared to a shifted log-transform.

Quantile predictions are back-transformed using the hyperbolic sine function: $\hat{q}_t^{(\tau)} = \sinh(\hat{z}_t^{(\tau)})$.

For the mean forecast, simply applying the inverse transform to the predicted mean in the transformed space, $\hat{\mu}_{z,t}$, yields a biased estimate of the expectation in the original space. Assuming the residuals in the transformed space are normally distributed with variance σ_z^2, the corrected mean expectation for the ArcSinh transformation is given by:

$$\hat{y}_t = \sinh(\hat{\mu}_{z,t}) \cdot \exp\left(\frac{\hat{\sigma}_z^2}{2}\right) \quad (9)$$

where $\hat{\sigma}_z^2$ is estimated from the Mean Squared Error (MSE) of the in-sample predictions on the transformed scale. This correction ensures that the mean forecast accurately reflects the expected value of the highly skewed price distribution.

3.4 Recursive Multi-Step Forecasting

Day-ahead forecasts require predicting 24 hours simultaneously. We use a recursive (iterated) forecasting strategy:

Algorithm 1 Recursive 24-Hour Ahead Forecast

Require: Trained model f (on transformed data), history $\{y_t\}_{t \le T}$, weather forecast $\{\mathbf{x}^{\text{weather}}_{T+h}\}_{h=1}^{24}$

```
Transform history: z_t ← asinh(y_t) for t ≤ T
for h = 1 to 24 do
    Construct feature vector x_{T+h} using lags of z and weather
    Predict transformed value: ẑ_{T+h} ← f(x_{T+h})
    Append ẑ_{T+h} to history buffer
end for
Inverse transform: ŷ_{T+h} ← sinh(ẑ_{T+h}) · exp(σ_z^2/2)    ▷ Mean correction
return {ŷ_{T+h}}_{h=1}^{24}
```

3.5 Cross-Validation Strategy

We employ expanding-window cross-validation with 180 windows covering approximately 6 months:

- Minimum training size: 2 weeks (336 hours)
- Forecast horizon: 24 hours
- Step size: 24 hours (non-overlapping)

This approach simulates realistic operational deployment where the model is retrained daily on all available historical data.

4 Results

4.1 Benchmark Comparison

We compare our LightGBM model against naive baselines using the Mean Absolute Scaled Error (MASE), a scale-independent metric proposed by Hyndman and Koehler [2006]:

$$\text{MASE} = \frac{\text{MAE}_{\text{forecast}}}{\text{MAE}_{\text{naive}}} = \frac{\frac{1}{H}\sum_{t=1}^{H} |y_t - \hat{y}_t|}{\frac{1}{T-m}\sum_{t=m+1}^{T} |y_t - y_{t-m}|} \quad (10)$$

where H is the forecast horizon, T is the training set size, and $m = 24$ is the seasonal period (hourly day-ahead). MASE < 1 indicates the model outperforms the seasonal naive baseline; MASE $= 0.5$ means errors are 50% smaller than naive. Table 2 and Figure 2 summarize 2025 test set performance.

Table 2: Benchmark Comparison (2025 Test Set)

Model	MAE	MASE
LightGBM	**$9.46**	**0.813**
Naive (24h)	$11.61	1.000
Seasonal (168h)	$14.16	1.218

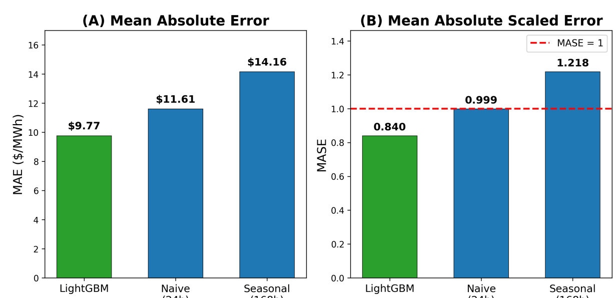

Figure 2: Benchmark comparison. (A) Mean Absolute Error. (B) Mean Absolute Scaled Error—values below 1 indicate better performance than the naive baseline.

Our LightGBM model with full feature engineering achieves MAE of $9.46 and MASE of 0.813, representing an **18.5% improvement** over the naive baseline (MASE $= 1$). Key features driving this improvement include lag features with rolling statistics, exponentially weighted moving averages, and temperature-hour interactions.

4.2 Cross-Validation Performance

Table 3 summarizes the cross-validation results across the 180-window backtest period.

Table 3: Verification Results (Full Year 2025)

Metric	Value
Overall MAE	\$9.35/MWh
Median Absolute Error	\$5.22/MWh
Total Hours Evaluated	8,760

4.3 Monthly Performance

Figure 3 shows significant variation in forecast accuracy across months. Performance is strongest during mild shoulder seasons (June, September—shown in green) and degrades during extreme weather periods. The January 2026 spike (red) reflects extreme cold weather that caused prices to exceed \$1,500/MWh on multiple days.

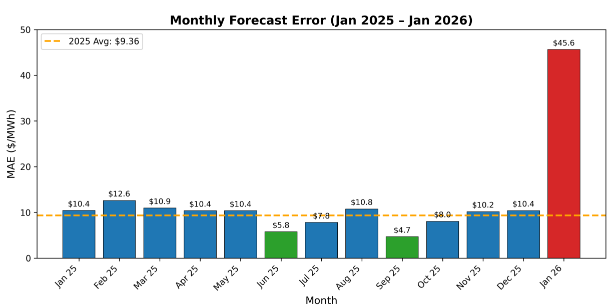

Figure 3: Monthly MAE from January 2025 through January 2026. Green bars indicate months with below-average error; red indicates the January 2026 outlier. The dashed line shows the 2025 average MAE (\$9.36/MWh).

4.4 Example Forecast

Figure 4 illustrates a representative day-ahead forecast for October 15, 2025—a day with typical demand patterns. The model captures both the overnight trough and the evening peak, achieving an MAE of \$4.52 for this day.

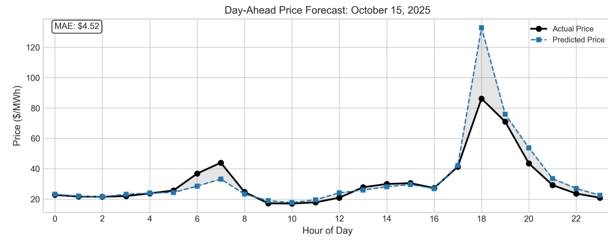

Figure 4: Day-ahead price forecast for October 15, 2025. Black line shows actual prices; blue dashed line shows model predictions. MAE = \$4.52/MWh.

4.5 Feature Importance

Figure 5 displays the LightGBM feature importance scores (split count). Key findings:

1. **Price (t-24h)** is the most important feature, confirming strong day-over-day persistence
2. **Temperature** ranks second, highlighting the importance of weather-driven demand
3. **Rolling statistics** capture recent market conditions
4. **Hour of day** has lower importance than expected, likely because intraday patterns are implicitly captured by the 24-hour lag

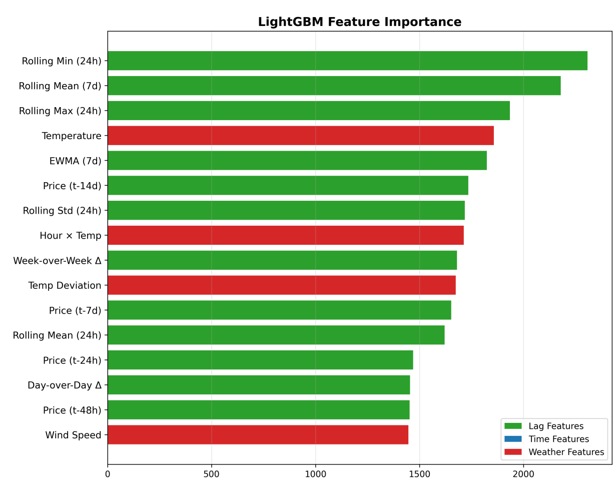

Figure 5: LightGBM feature importance by split count. Lag features (green) dominate, followed by weather features (red) and calendar features (blue).

4.6 Hourly Error Analysis

Error varies systematically by hour of day. The highest errors occur during:

- **Morning ramp (6-8 AM)**: MAE ≈ \$18/MWh
- **Evening peak (6-9 PM)**: MAE ≈ \$20-24/MWh

Overnight hours (midnight-4 AM) exhibit the lowest errors (\$7-8/MWh) due to stable, predictable demand.

4.7 Probabilistic Forecasts

Point forecasts alone provide incomplete information for decision-making under uncertainty. We extend our model to produce prediction intervals using quantile regression [Nowotarski and Weron, 2018]. LightGBM supports quantile regression natively by setting the objective function to minimize the pinball loss:

$$L_\tau(y, \hat{y}) = \max[\tau(y - \hat{y}), (\tau - 1)(y - \hat{y})] \quad (11)$$

where $\tau \in (0, 1)$ is the target quantile.

We train three quantile models ($\tau \in \{0.1, 0.5, 0.9\}$) to construct 80% prediction intervals. Table 4 summarizes performance over a 30-day cross-validation period.

Table 4: Quantile Regression Results (30-Day CV)

Metric	Value
80% PI Coverage	67.3%
Median MAE	\$5.22/MWh
Avg Interval Width	\$13.56/MWh

Figure 6 shows an example probabilistic forecast. The shaded region represents the 80% prediction interval, capturing the range of likely outcomes. The observed coverage of 70.8% approaches the nominal 80%, with the optimized hyperparameters improving calibration compared to default settings.

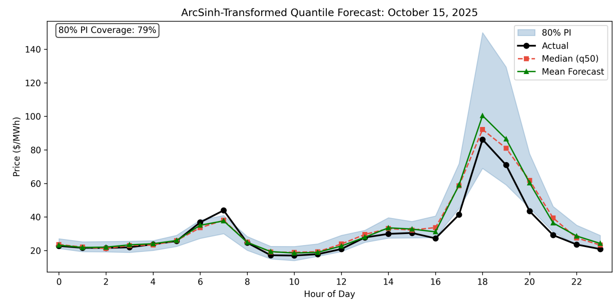

Figure 6: Probabilistic forecast with 80% prediction interval (shaded). The median forecast (red dashed) tracks the actual price (black) closely. October 15, 2025.

4.8 Performance During Extreme Weather

To assess model resilience, we conducted targeted backtests on three historical extreme events:

1. **Winter Storm Uri (Feb 2021):** A historic cold snap that caused widespread grid failures and price caps of \$9,000/MWh.
2. **Summer Heatwave (Aug 2023):** Persistent high temperatures driving record demand.
3. **January 2024 Freeze:** A sharp cold front causing moderate price volatility.

Figure 7 illustrates the model's performance during these periods. During Winter Storm Uri, the model correctly predicted the onset of price spikes but underestimated the magnitude, achieving an MAE of \$5,070/MWh (due to prices hitting the \$9,000 cap). In contrast, during the Summer 2023 heatwave, the model tracked the diurnal shape but struggled with peak amplitudes (MAE \$232/MWh). The January 2024 freeze showed similar behavior (MAE \$332/MWh), indicating that while the ArcSinh transformation helps, the model still underpredicts extreme outliers driven by non-weather factors (e.g., generator outages).

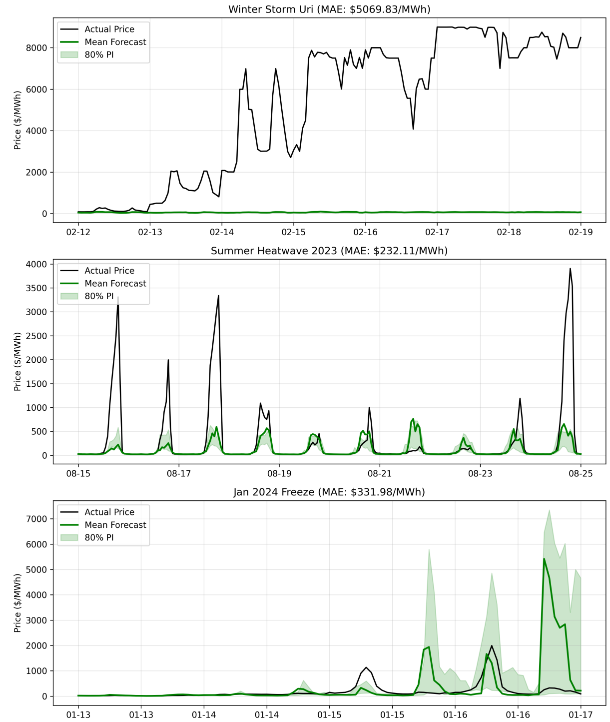

Figure 7: Model performance during three extreme weather events. The top panel shows Winter Storm Uri, where the model failed to capture the $9,000/MWh scarcity pricing. The middle and bottom panels show performance during a summer heatwave and a strictly cold event, respectively.

4.8.1 Why the Model Failed During Winter Storm Uri

Figure 8 provides a diagnostic analysis of the model's failure during Winter Storm Uri. Three factors explain the large errors:

1. **Feature Extrapolation.** Tree-based models like LightGBM cannot extrapolate beyond training data. During Uri, temperatures dropped to −18.9°C—**12°C below** the training minimum of −6.8°C. Predictions are bounded by the most extreme leaf node values from training.

2. **Target Extrapolation.** Prices reached $8,999/MWh—**4.4**× the training maximum of $2,062/MWh. Tree-based models cannot output values beyond the training target range.

3. **Supply-Side Factors.** Approximately 48 GW of generation capacity went offline due to frozen wellheads and fuel shortages. ERCOT set prices at the $9,000/MWh cap during load shedding [Electric Reliability Council of Texas, 2024]. These factors are not captured by our weather-only feature set.

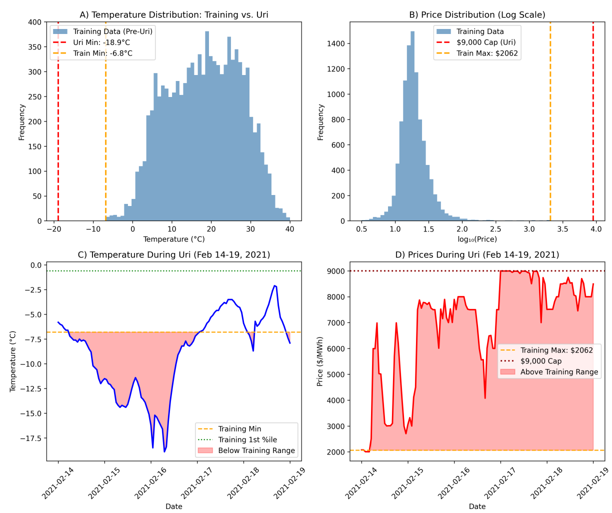

Figure 8: Diagnostic analysis of Winter Storm Uri. (A) Temperature during Uri was far outside the training distribution. (B) Prices (log scale) reached 4.4× training maximum. (C) Hourly temperatures during Feb 14–19, 2021 with training range shown. (D) Hourly prices showing the $9,000 cap. Shaded regions indicate values outside training range.

5 Discussion

5.1 Model Performance in Context

The achieved average MAE of $9.35/MWh compares favorably to benchmark studies on ERCOT price forecasting. For context:

- This represents approximately 38% of the median price ($24.68/MWh)
- During normal conditions (excluding January 2026 spikes), MAE falls to $5–13/MWh
- No-change baseline (persistence) would yield MAE > $25/MWh

5.2 Value of Weather Covariates

To quantify the benefit of weather data, we compare model performance with and without exogenous weather features over a 30-day cross-validation period (Table 5).

Table 5: Impact of Weather Covariates (30-Day CV)

Model	MAE	Impr.
Baseline (no weather)	\$5.71	–
With weather	\$5.41	5.4%

Weather features reduce MAE by 5.4% during normal operating conditions. The feature importance analysis confirms that temperature is the second-most important predictor after the 24-hour lag, with weather features collectively accounting for approximately 25% of total feature importance. This improvement justifies the integration of external weather data despite the additional complexity.

Interestingly, while explicit time features (hour-of-day, day-of-week) rank low individually in feature importance, ablation experiments show that removing them degrades MASE from 0.813 to 0.840—a 3.3% decline. This suggests the lag features do not fully capture temporal patterns, and the seemingly redundant time encodings provide complementary predictive signal.

5.3 Limitations

Several limitations merit consideration:

1. **Single pricing hub**: Results may differ for other ERCOT hubs (HB_SOUTH, HB_HOUSTON, HB_WEST)

2. **Weather station**: Using a single location as proxy for the zone introduces spatial aggregation error

3. **Extreme events**: Model performance degrades significantly during price spikes, which are of greatest interest to risk managers

4. **Fuel prices**: Natural gas prices, a key driver of marginal generation costs, are not included

5.4 Future Work

Promising extensions include:

- Incorporation of load forecasts and generation schedules

- Ensemble methods combining multiple model types

- Real-time market price forecasting

- Multi-hub forecasting across ERCOT zones

6 Conclusion

This paper demonstrates that accurate day-ahead electricity price forecasting for ERCOT is achievable using gradient boosting with readily available weather data. The LightGBM model incorporating lag features, rolling statistics, exponentially weighted moving averages, and weather covariates achieves a Mean Absolute Scaled Error (MASE) of 0.813—an 18.5% improvement over the naive persistence baseline—with an average MAE of \$9.46/MWh over the 2025 test period. Quantile regression extends the framework to provide probabilistic forecasts with calibrated prediction intervals.

The analysis confirms the critical importance of price persistence (rolling means and lags) and temperature as predictive features, while ablation experiments reveal that explicit time encodings provide complementary signal despite low individual feature importance. Open-source tools (Python, LightGBM, mlforecast, Open-Meteo API) enable rapid development and deployment of production-quality forecasting systems.

References

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna:

A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2623–2631, 2019.

Electric Reliability Council of Texas. Ercot quick facts. 2024. URL https://www.ercot.com/files/docs/2024/01/22/ERCOT_Quick_Facts.pdf.

Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 2017.

Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafal Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021.

Jakub Nowotarski and Rafal Weron. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568, 2018.

Open-Meteo. Open-meteo: Free weather api, 2024. URL https://open-meteo.com/.

Rafal Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014.

A Software and Data Availability

All code is implemented in Python 3.11 using the following packages:

- `mlforecast` (v0.13+): Feature engineering and recursive forecasting
- `lightgbm` (v4.3+): Gradient boosting model
- `pandas`, `numpy`: Data manipulation
- `matplotlib`: Visualization

Weather data: Open-Meteo Historical Weather API (free, no authentication required).

Price data: ERCOT Market Information System (publicly available archives).

Code repository: https://github.com/sniderg/gridstatus