Point Forecasts to Probability Clouds: Deep Learning Meets Electricity Market Uncertainty

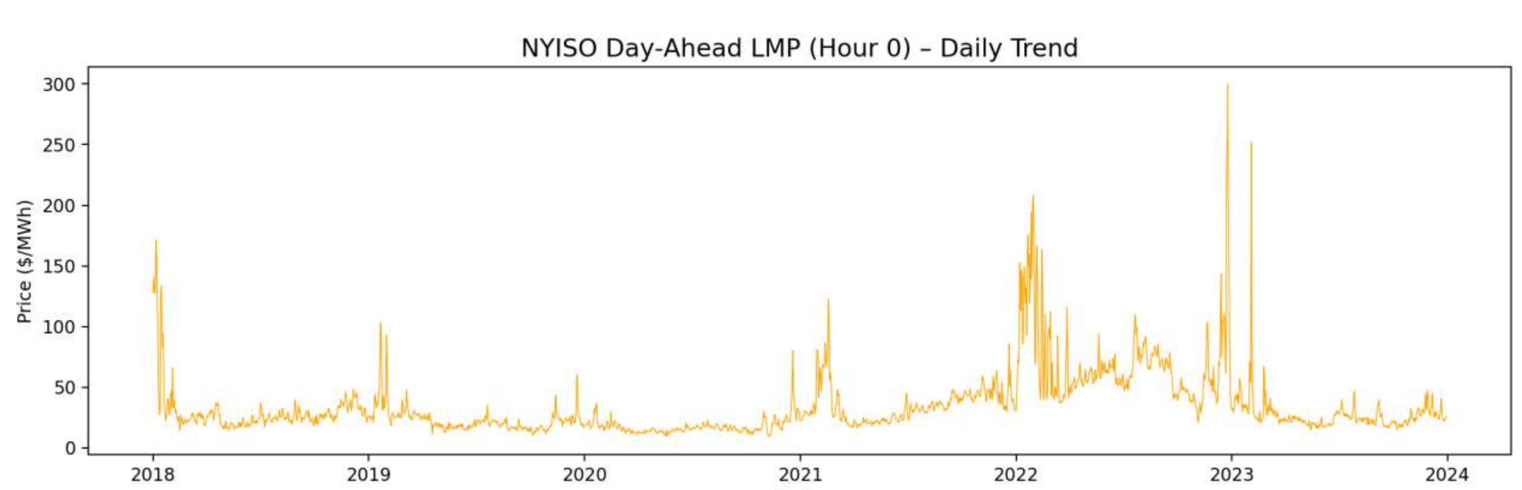
Author: Haixi Zhang(hz733@cornell.edu) Advisor: Lang Tong(lt35@cornell.edu)

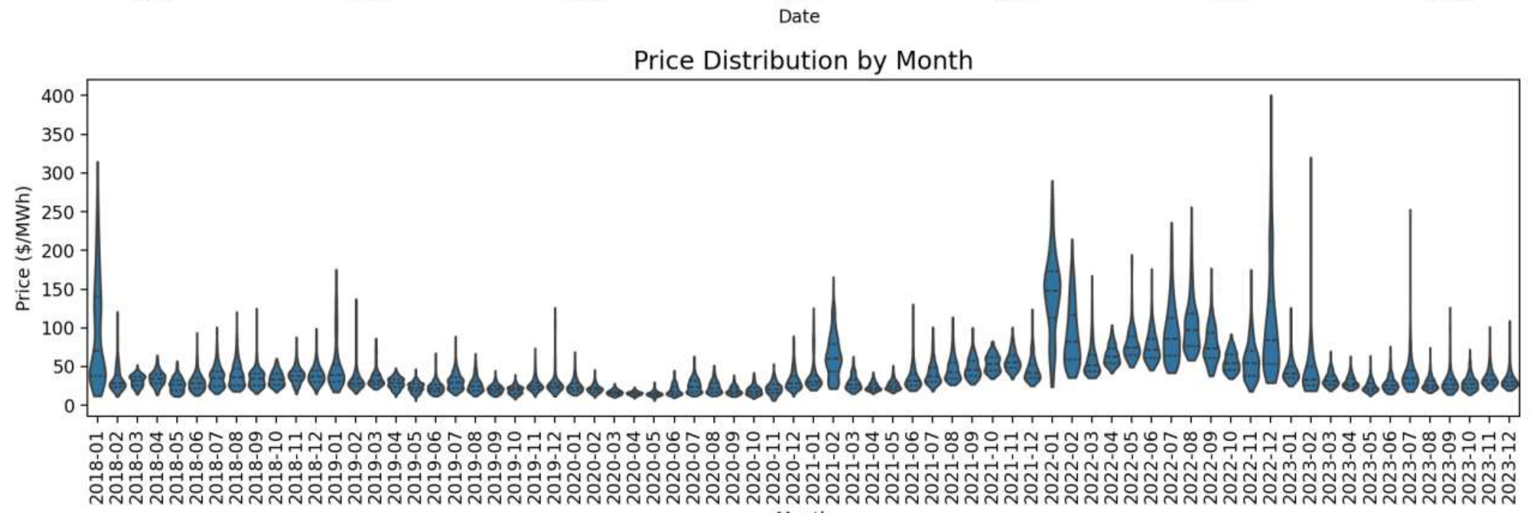
Overview

- **Electricity Market Forecasting:** Accurate forecasting of electricity prices, called Locational Marginal Prices (LMPs), is essential for grid operators and energy traders. Good forecasts support better scheduling, risk management, and overall market efficiency.
- **Challenges:** Day-ahead LMPs are highly volatile, affected by weather, demand changes, and unexpected outages. Traditional models struggle with the nonlinearity and sudden price spikes, especially as renewable energy makes the system even more unpredictable.
- **Deep Learning Opportunity:** Deep learning models like Transformers and RNNs can learn complex patterns in time-series data. These methods have been successful in other fields and offer a promising way to model electricity prices more accurately.
- **Need for Uncertainty Quantification:** Standard (deterministic) models only predict one price value, without showing how confident they are. In real-world markets, knowing the possible range of outcomes is crucial. Probabilistic forecasting solves this by predicting a full range of future prices, helping operators plan for best and worst cases.
- **Goal:** I compare several deep learning models on NYISO day-ahead prices and adapt a new probabilistic model — UWIAE-GPF (Univariate Weak Innovation Autoencoder for Generative Probabilistic Forecasting) — to produce not only accurate forecasts but also meaningful uncertainty estimates.

Data & Problem Setup

- **Dataset:** New York Independent System Operator (NYISO) Day-Ahead Market Locational Based Marginal Prices (LBMP) data for the N.Y.C. zone, covering January 1, 2018 to December 30, 2023 [1]. The dataset provides hourly clearing prices for electricity delivered the next day. Each day contains 24 hourly price points.
- **Forecasting Task:** Formulated as a time-series prediction problem. Given a window of past prices, the goal is to predict future prices.
- **Data Preparation:** The dataset is split chronologically into a training set (2018–2022) and a test set (2023). Prices are normalized based on historical means and variances during training and rescaled back for evaluation.
- **Characteristics:** The LMP time series shows regular daily cycles and weekly trends, but also sharp spikes during high demand or generator outages. These characteristics demand models that can capture periodic structure and react to anomalies. The problem setup thus evaluates whether advanced models can learn the base patterns while being flexible enough to predict rare events.





Model Families

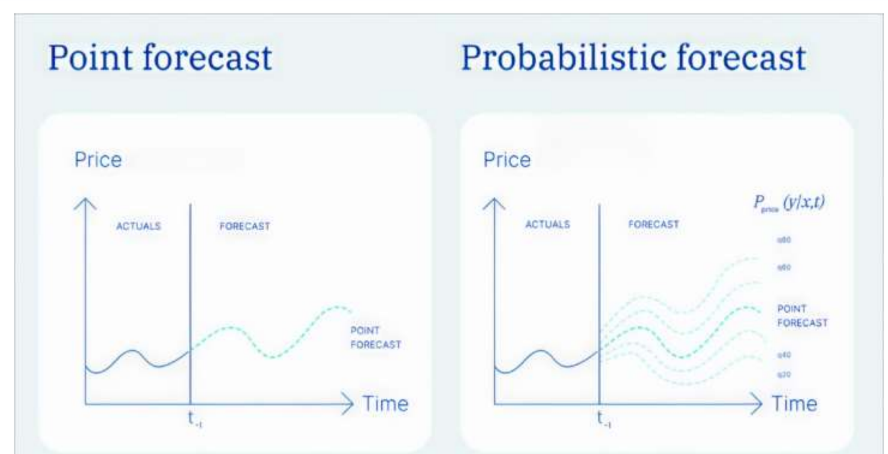
Deterministic · Classic Deep Learning Models
- GRU[2]– gated recurrent memory of past prices
- Nlinear[3] – season-trend linear residual model
- StemGNN[4] – spectral graph convolution over time

Deterministic · Transformer Variants
- Informer[5] (prob-sparse attention)
- iTransformer[6] (inverted FFT mixing)
- PatchTST[7] (patch embedding of subsequences)
- Autoformer[8] (auto-correlation decomposition)

Probabilistic · Diffusion-Based Model
- TimeGrad[9] – latent denoising-diffusion process

Probabilistic · Flow-Based Models
- GRU-RealNVP [10]
- GRU-MAF [10]
- Transformer-MAF [10]

Probabilistic · Innovation Autoencoder
- WIAE_GPF [11]– Original Weak-Innovation Autoencoder
- UWIAE_GPF – Univariate version with improved performance



Experiments

Training Setup

Models are trained on the historical NYISO data using PyTorch. I use a rolling window of past 1, 2, 7, 14, 28 days (24 – 672 h) of prices as input to predict the next-day hourly LMP (24 h) for N.Y.C. zone.

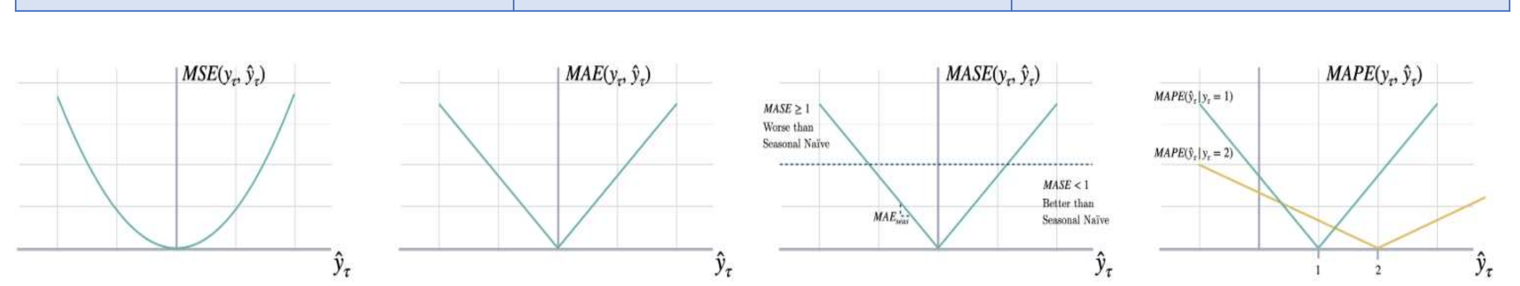
Probabilistic Forecast Generation

- After convergence, 1 000 latent resamplings per forecast window → ensemble of 24-h price paths.
- From the ensemble we extract:
  - Mean (μ) and Median (m) trajectories — used as point forecasts.
  - Symmetric central intervals (90 % / 50 % / 10 %) — used for risk bands.
- Transformer / LSTM outputs are single-valued ⇒ treated as μ only; no built-in interval.

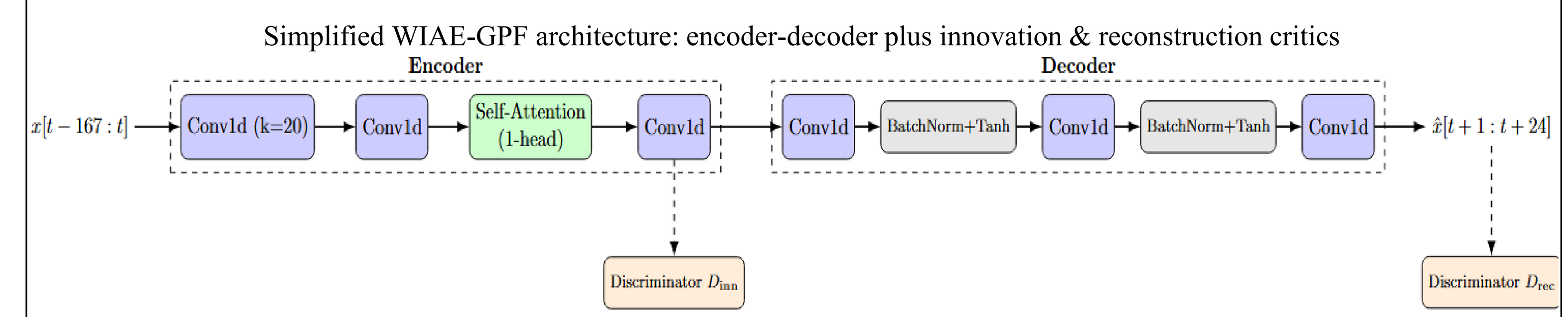
Evaluation Metrics

All errors are normalized (N-) by dividing by the empirical price range over the test set.

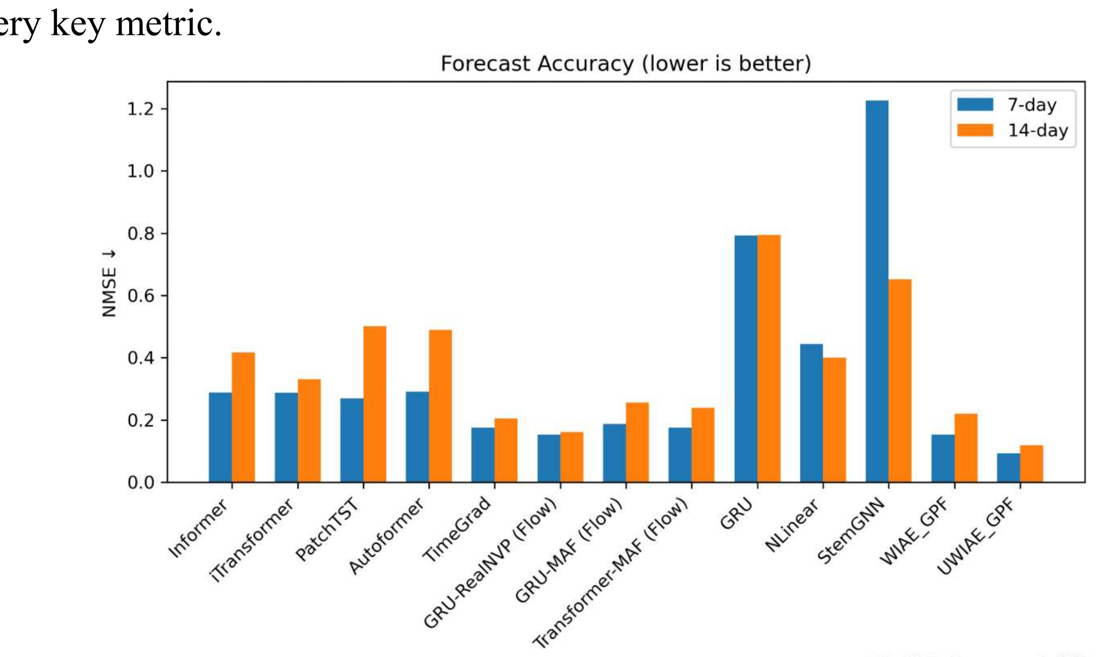
NMSE (Normalized Mean Squared Error)	Average of squared forecast errors, normalized by the variance (or price range) of the series	Penalizes large errors heavily
NMAE (Normalized Mean Absolute Error)	Average of absolute forecast errors, normalized by the price range	Easy to interpret; less sensitive to outliers than NMSE; normalized for fair comparison.
MAPE (Mean Absolute Percentage Error)	Mean of absolute errors divided by actual values, expressed as a percentage.	Intuitive percentage interpretation; useful for business/market communication.
MASE (Mean Absolute Scaled Error)	Mean absolute error scaled relative to a naïve benchmark (e.g., last observation).	Allows fair comparison to a simple baseline; values >1 indicate worse than naïve.

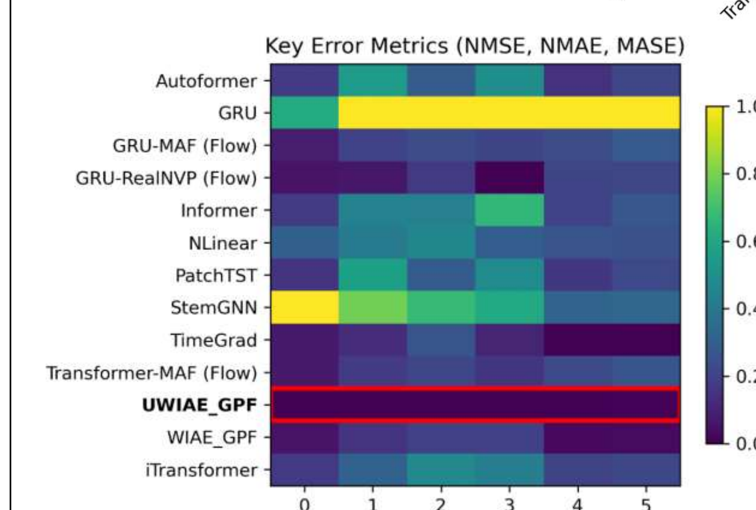


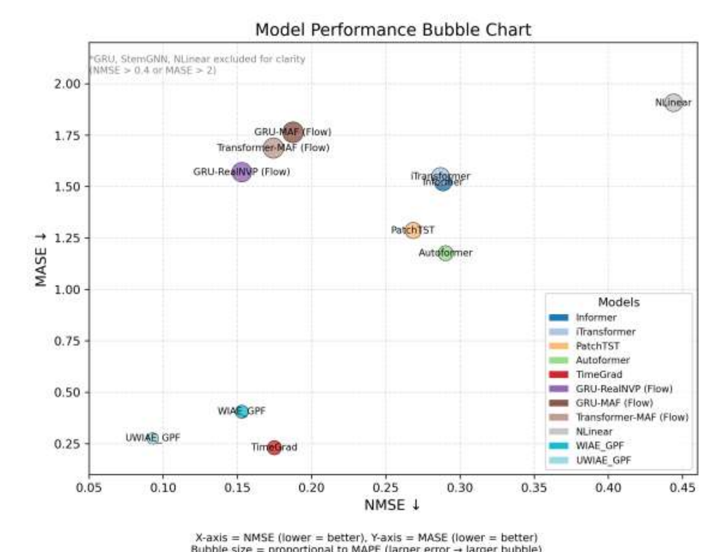
Results



- By fusing weak-innovation encoding with an upgraded adversarial generator, UWIAE_GPF sets a new benchmark for probabilistic electricity-price forecasting, outperforming both state-of-the-art Transformers and earlier generative baselines in every key metric.







Conclusions

- **Probabilistic Wins → Lower Risk**: My improved UWIAE-GPF model achieved the best accuracy with an NMSE of 0.093, which is **65.3%** lower than PatchTST and **39.2%** lower than the original WIAE, while maintaining a 4× faster per-epoch training time than PatchTST (the best-performing deterministic model).
- **Transformer vs. LSTM**: Attention nets (PatchTST > iTransformer > Informer) beat recurrent baselines, but gains flatten once the history window exceeds ~7 days.
- **Generative Edge**: Autoencoder + GAN critics learns spike-tail distributions; sample paths reveal multiple market “scenarios,” giving operators a quantitative feel for worst-case prices.
- I would like to thank Prof. Lang Tong for his guidance and for introducing the WIAE-GPF framework, originally proposed by Xinyi Wang, Qing Zhao, and Lang Tong. Special thanks to Ph.D. student Ze Hu for his valuable support on data processing and model choices.

References

[1] NYISO, “Custom Reports,” 2024. [Online]. Available: https://www.nyiso.com
[2] R. Dey and F. M. Salem, “Gate variants of GRU neural networks,” Proc. IEEE MWSCAS, 2017.
[3] A. Zeng et al., “Are Transformers effective for time-series forecasting?,” AAAI-23, pp. 11121-11128, 2023.
[4] D. Cao et al., “Spectral-temporal GNN for multivariate time-series forecasting,” NeurIPS-20, pp. 17766-17778.
[5] H. Zhou et al., “Informer: Efficient transformer for long-sequence forecasting,” arXiv:2012.07436, 2021.
[6] Y. Liu et al., “iTransformer: Inverted transformers for time-series,” arXiv:2310.06625, 2023.
[7] Y. Nie et al., “A time series is worth 64 words,” arXiv:2211.14730, 2022.
[8] H. Wu et al., “Autoformer: Decomposition transformers with auto-correlation,” NeurIPS-21.
[9] K. Rasul et al., “Autoregressive denoising diffusion models for probabilistic forecasting,” ICML-21.
[10] K. Rasul et al., “Multivariate probabilistic forecasting via conditioned normalizing flows,” arXiv:2002.06103, 2021.
[11] X. Wang, Q. Zhao, and L. Tong, “Probabilistic forecasting of real-time electricity market signals via interpretable generative AI,” arXiv:2403.05743, 2024.

Cornell Engineering
Electrical and Computer Engineering