Probabilistic Day-Ahead Power Forecasting in the Medium-Voltage Grid Using Structured State Space Models

Michael Mesarcik[1*], Jessica Loke[1], Jochem Wildeboer[1], Bob Lucassen[1]

[1]Alliander N.V., The Netherlands

*Email: misha.mesarcik@alliander.com

Keywords: DAY-AHEAD POWER FORECAST, MEDIUM-VOLTAGE GRID, PROBABILISTIC FORECASTING, STRUCTURED STATE-SPACE SEQUENCE MODELS, DEEP LEARNING

Abstract

Short-term power forecasting is becoming more challenging due to the growing integration of renewable energy sources. This paper presents a novel application of structured state space models for probabilistic day-ahead power forecasting in the medium-voltage grid. These probabilistic outputs enable grid operators to make informed-decisions with knowledge of prediction (un)certainties. The model receives both historical power measurements, historical weather measurements and weather forecasts to generate a power forecast. We validated our approach on a dataset comprising of 10 years of measurements from 312 substations across the Dutch electrical grid. Results showed that our approach achieves superior performance compared to traditional machine learning methods. Our findings suggest that structured state space models offer a promising direction for improving the accuracy of power forecast and providing grid operators with reliable uncertainty quantification.

1 Introduction

Short-term forecasting of both power demand and generation is essential for grid operators to guarantee grid safety and optimise flexible load management strategies. The integration of renewable energy sources, coupled with higher loads and increasing power flexibility, presents significant challenges for grid operators. Maintaining grid stability and ensuring reliable power supply under these conditions has become increasingly complex (1). Due to the complexity and amount of data required, machine learning-based load forecasting has become commonplace in the toolbox of grid operators.

Traditional machine learning-based regression techniques like decision tree ensembles and linear models are commonly used for power forecasting. Although these models have been valued for their simplicity, they face significant limitations, such as extensive feature engineering requirements, poor generalisability across different substations, and limited ability to capture complex temporal dependencies. These shortcomings have motivated the time series forecasting domain to explore more advanced deep learning-based methods. Deep learning architectures such as Long Short Term Memory networks (LSTMs) (2), Recurrent Neural Networks (RNNs) and transformers (3) have shown positive results in the time series domain but they face fundamental limitations. RNNs and LSTMs struggle with vanishing gradients and computational inefficiency over long sequences (4); whereas transformers face quadratic memory complexity and context-loss problems (3).

To remedy the limitations of the aforementioned architectures, state-space models (5) have emerged as a popular alternative in the time series forecasting domain due to their computational efficiency and state-of-the-art performance. State-space models are well suited to day-ahead power forecasting in the medium voltage grid because they can effectively capture both long-term seasonal trends and short-term hourly variations while readily incorporating additional features like weather forecasts and grid topology without extensive feature engineering efforts. We extend the state-space model architecture to be probabilistic as this approach is beneficial for grid operators in managing the electrical grid's complexity, handling its system dynamics, and accounting for the noise present in the data (6).

With this work, we make the following contributions: 1) a novel application of state-space models for power forecasting in the medium-voltage grid, achieving state-of-the-art performance; laying the foundations for larger models (7) 2) a simple extension to state-space models that enables probabilistic forecasting.

2 State space models for probabilistic power forecasting

State space models are a type of sequence model used to capture long-term dependencies in sequential data. State space models effectively map an input to a hidden state, and then using that hidden state the input is mapped back to some output, as given by

$$\begin{aligned} x'(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{1}$$

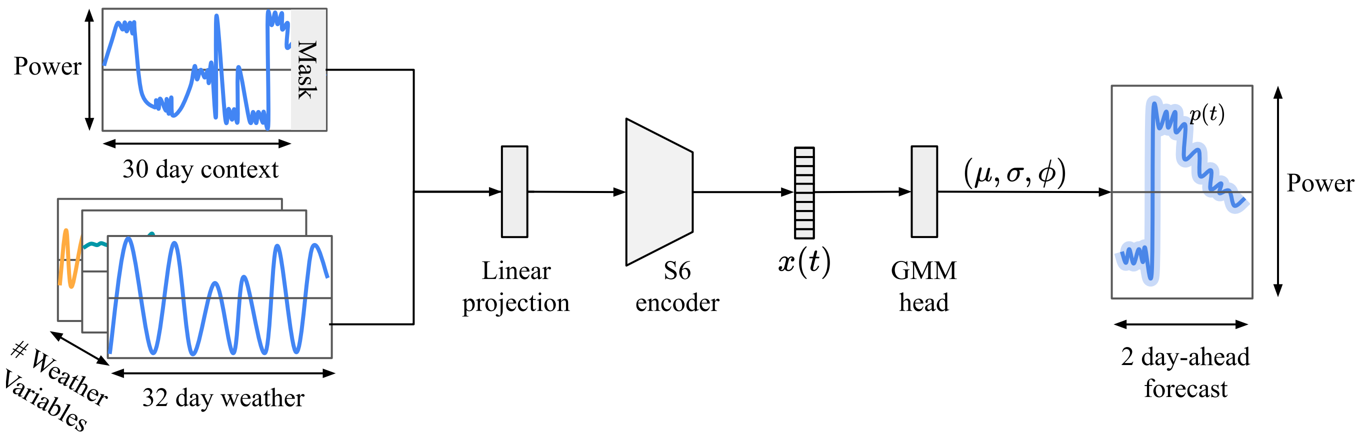

Figure 1: Diagram of the probabilistic state space model. At inference it ingests 30 days of context from a given substation as well as the weather measurements for the 30 day context and weather forecasts the 2 day prediction window. The model outputs a GMM per time sample given the encoding of the S6 model.

Where $u(t)$ is the input sequence, $y(t)$ is the output sequence, $x(t)$ is the current hidden state and matrices A, B, C and D are trainable parameters (5).

Recent developments have adapted these models for deep learning - *Structured State Space Sequence model* (S4) (5) and its successor *Mamba* (S6) (8). Building on this progress, we have utilised these models for probabilistic power forecasting in the electricity grid. To produce probabilistic forecasts, we project our input data using the S6 encoder to a latent representation $x(t)$, and then use a Gaussian Mixture Model (GMM) projection to produce the probabilistic forecast $p(t)$ as shown in Figure 1.

The probability distribution $p(t)$ for every time step t is described by a GMM with K components:

$$p(t) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\mu_i, \sigma_i), \tag{2}$$

such that the likelihood P of a power value ν at time step t is:

$$P(t, \nu) = p(t)|\nu. \tag{3}$$

Here, $\mathcal{N}$ represents a normal distribution with means μ_i, standard deviation σ_i, and ϕ_i as the normalised mixture component weights, such that $\sum_{i=1}^{K} \phi_i = 1$. This GMM projection provides a representation of uncertainty in our forecasts.

To train the state space model with a probabilistic output, we minimise the Negative Log-Likelihood (NLL) of the measured data under the predicted GMM distribution. The NLL objective trains the model to maximise the likelihood of the real observations, enabling it to produce forecasts that align with the true values while providing realistic uncertainty estimates. For each time step t, the loss is given by:

$$\mathcal{L}_{\text{NLL}} = -\log\left(p(t)|m(t)\right), \tag{4}$$

where $m(t)$ is the measured power value. Through minimising the NLL loss, the model learns to optimise the state-space parameters A, B, C, D and the GMM parameters, effectively capturing the temporal dynamics of power loads.

2.1 Hyperparameters and training

To validate our approach, we experimented with several architectural configurations. The optimal architectural configuration consisted of 6 layers using the S6 (Mamba) kernel (8), with an embedding dimension of 384 and 4 components in the GMM. We trained the model with a batch size of 30 with the Rectified Adam (RAdam) optimiser, a learning rate of 10^{-4}, a cosine learning rate schedule, and gradient-clipping of 0.1. During training, we simultaneously flipped the signs of both context and prediction window values with a 50% probability to improve the model's robustness to the bidirectional nature of power flow. We found that a context window of 30 days and a prediction window of 2 days provided optimal performance across our evaluation metrics MAE, NLL and F-β. Training took approximately 2 days on 4 NVIDIA A10G GPUs.

3 Data selection and preprocessing

We curated a dataset for medium voltage power forecasting consisting of measurements in the Dutch electrical grid. Due to privacy concerns, we are unable to release it publicly at this time.

Our dataset consists of 5-minute averaged measurements of active power (in MVA) collected across Alliander's electrical grid. It is split into training and testing parts, where the training data consist of 10 years of historical measurements (2013-2023) from 312 different substations. The test set comprises one year of measurements

(2023–2024) sampled at 15 minute intervals, from six substations known for their challenging forecasting behaviour. These substations exhibit complex dynamics, including consumption, solar generation, wind generation, responses to energy market pricing and congestion management actions.

The raw measurements contain a multitude of data errors, such as constant-valued readings, instantaneous offsets, sensor dead-banding and polarity switches. Several approaches to remove these errors were tested and an anomaly detection model was developed to detect and remove data errors from our training data. For brevity sake, we will discuss this anomaly detection model in more detail in an upcoming publication.

We trained our probabilistic state space forecasting model on both power measurements as well as weather measurements. Here, we experimented with several weather forecast providers, and reported our results with measurements from Open-Meteo[1]. We experimentally chose four weather variables: short-wave solar irradiance and direct normal irradiance, wind speed at 100m and air temperature at 2m.

We also needed to normalise both the historical power measurements and weather measurements. To prevent data leakage, we normalised these measurements per batch, with the mean and standard deviation calculated only from the context window - excluding data from the prediction window. Additionally, we clamped all measurements at 10 times their standard deviation.

4 Experiments

To evaluate the model's test set performance, we used two metrics which reflect key operational requirements - relative Mean Absolute Error (rMAE) and F-β score. The rMAE measures forecast accuracy of a forecasted quantile, while the F-β score assesses the model's ability to predict capacity limit exceedances. As rMAE and F-β score are deterministic metrics, we discretised the probabilistic state-space model's GMM output into its corresponding quantiles.

The F-β score is the harmonic mean of precision (ratio of correctly predicted exceedances to all predicted exceedances) and recall (ratio of predicted exceedances to all actual exceedances) where β is the weighting factor. We set $\beta = 2$ to weight recall more highly as this reflects our operational requirements; missing actual grid capacity exceedances is more critical than raising a false alarm.

The rMAE evaluates the difference between the forecasted values and actual measurements. Because this metric is sensitive to outliers, we scaled it using the 1st and 99th percentiles of the measured signal. This scaling ensures comparability across different signals.

[1]https://open-meteo.com/en/docs/historical-weather-api

Location #	**Linear**	**Gradient boosting machine**	**State space (ours)**
1	0.07	0.06	**0.05**
2	0.09	0.10	**0.08**
3	0.13	**0.10**	**0.10**
4	0.11	0.10	**0.09**
5	0.08	0.08	**0.06**
6	0.07	**0.06**	**0.06**
Mean	0.09	0.08	**0.07**

Table 1: Relative mean absolute error (rMAE) of 2-day ahead forecasts for the 6 test substation locations (2023-2024) for several different models. We report the rMAE of the 50$^{\text{th}}$ quantile for all test signals, with best in bold.

4.1 Baseline models

We compared our model against quantile regressors - linear and gradient boosting machines - implemented in openSTEF, an open-source forecasting framework from LF Energy. These baseline models, developed in part through Alliander's contributions to openSTEF[2], are currently used in production for grid operation decisions internally at Alliander. Each baseline model was trained independently per substation using three months of historical power measurements and historical weather forecasts, optimising a quantile regression task with pinball loss. For the linear model, an additional loss penalty was given for larger values to account for the importance of extreme values in daily operations.

A notable distinction in the training process was our state space model's use of actual weather measurements, while the gradient boosting machine and linear models were trained using historial 1-hour-ahead weather forecasts provided by KNMI [3].

4.2 Probabilistic forecasts

Table 1 presents the relative rMAE of 2-day-ahead active power (MVA) forecasts for the 50th quantile across six test substation locations and three models. The 50th quantile rMAE is a key metric as it assesses the accuracy of the median forecast relative to observed values.

Our state space model consistently outperforms or matches the baseline models across all locations, achieving rMAE improvements ranging from 0.01 to 0.03 points. On average, our state-space model achieves a mean rMAE of 0.07, outperforming the linear model (0.09) and gradient boosting machine (0.08). This demonstrates the superior accuracy of our approach for 2-day-ahead forecasts.

[2]https://github.com/OpenSTEF/openstef/releases/tag/v3.4.32
[3]https://dataplatform.knmi.nl/dataset/harmonie-arome-cy40-p3-0-2

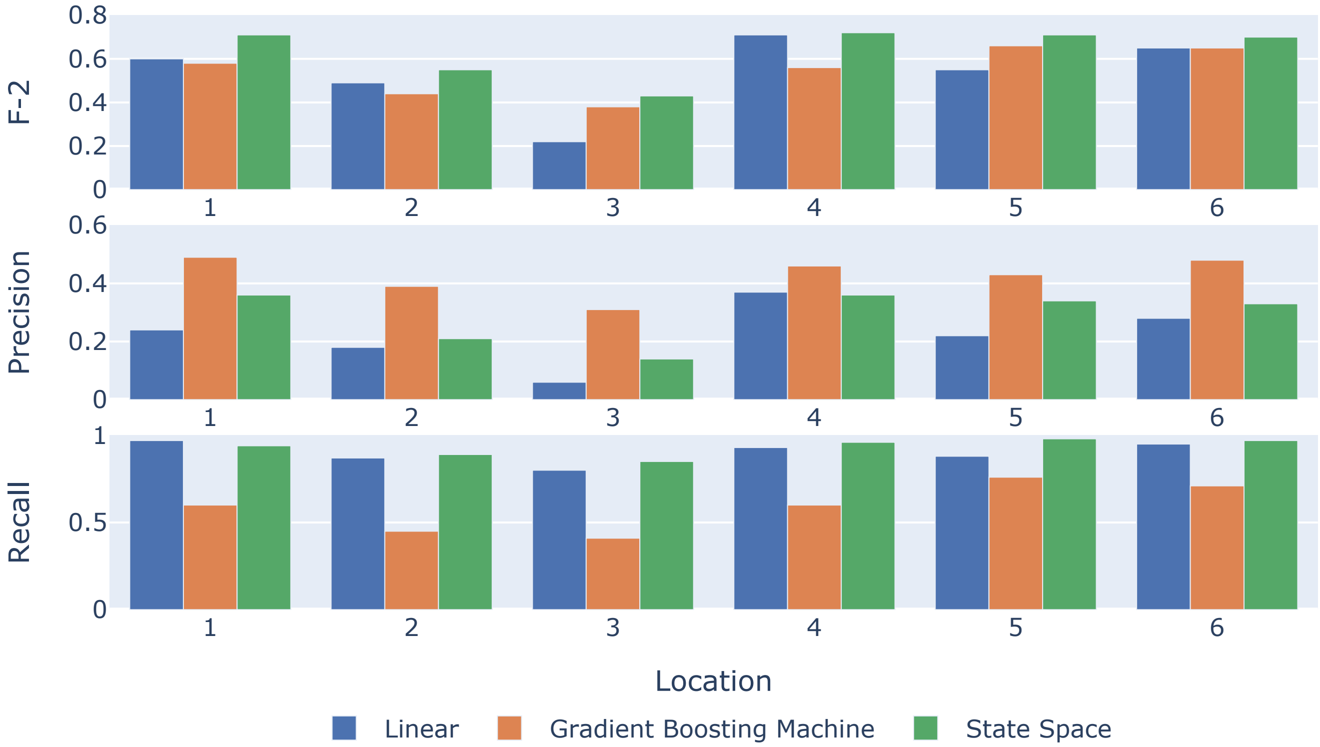

Figure 2: F-2, precision and recall scores of the 2-day ahead forecasts for the 6 test substations (2023-2024) for several different models on predetermined quantiles used for grid congestion management.

4.3 Forecasting for congestion management

We evaluated each model's accuracy in predicting grid capacity exceedances, as shown in Figure 2. Using the 5$^{\text{th}}$ quantile forecast and substation-specific transport capacity limits, we computed precision, recall and F_2 scores across all test signals.

It is clear that our proposed state space model obtains optimal F_2 score across all test signals. We find that the state space model offers improvements ranging from 0.01 to 0.21 relative to the linear and gradient boosting machine models. Notably, we find the gradient boosting machines model to give most balanced precision and recall forecasts, whereas the linear model favours recall over precision. We see a similar trend in the state-space model as the linear model, while maintaining consistently higher precision relative to the linear model. The congestion management results track closely with the rMAE-based results.

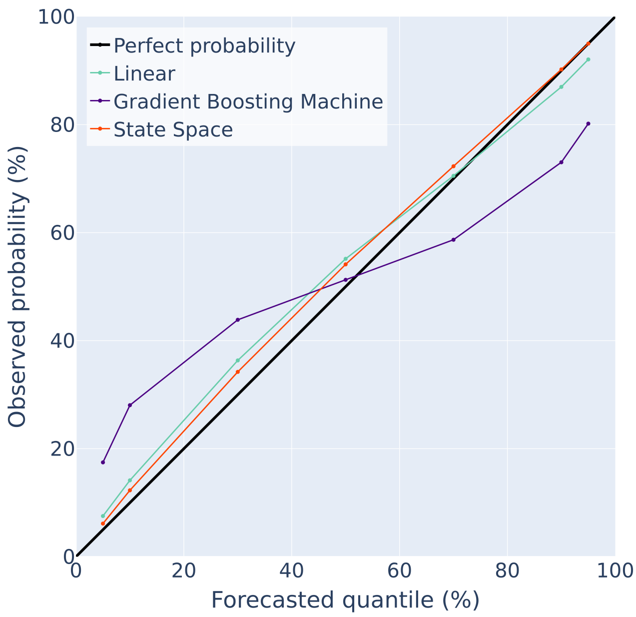

Figure 3: Example calibration plot comparing forecasted quantiles with observed probabilities for different models for the 6$^{\text{th}}$ location. The black line shows perfect ligament with observed probability and forecasted quantile.

4.4 Calibration of forecasts

To verify our probabilistic forecast, we evaluated model calibration by comparing forecasted quantiles against observed probabilities. Specifically, we computed the observed probability for each quantile and showed the discrepancy between the observed and forecasted probabilities. Figure 3 shows the calibration results for location 6, where the state space and linear models closely follow the ideal calibration line with only minor deviations at extreme quantiles. In contrast, the gradient boosting model shows significant miscalibration, with its 95th percentile forecast corresponding to the 80th percentile of observed data. While such miscalibration could be addressed using conformalized quantile regression techniques, our state space model already achieves the desired quantile coverage.

This calibration performance directly impacts operational reliability: accurate probabilistic forecasts from our state space model enable grid operators to make informed decisions about risk assessment and capacity planning with properly calibrated confidence levels.

5 Conclusion

This paper demonstrates the effectiveness of state space models for probabilistic day-ahead power forecasting in the medium-voltage grid. Our experimental results show three key advantages of the proposed approach.

First, our state space model achieves superior accuracy, reducing the relative mean absolute error by 3 points compared to traditional approaches across all test locations.

Second, we find that the state-space model provides improvements ranging from 0.01 to 0.21 compared to the linear and gradient boosting machine models. This performance boost is paired with well-calibrated probabilistic forecasts, where predicted quantiles align closely with observed probabilities. Additionally, the model consistently achieves high F_2 scores in predicting capacity exceedances, making it especially beneficial for grid operators tasked with managing congestion risks.

Third, our comprehensive evaluation on 10 years of measurements from substations across the Dutch electrical grid validates the model's practical applicability. The test results on six chosen challenging substations demonstrates robust performance across diverse operational conditions including mixed consumption and generation profiles and market-responsive behaviours.

Future work will focus on extending model capabilities by incorporating additional grid topology information and deploying these models in operational settings. Furthermore, we are investigating the publication of our dataset, which features diverse load patterns, generation types and grid conditions. Through continued research and validation in real-world deployment, we aim to help grid operators manage the challenges of renewable energy integration in their networks.

References

[1] N. Phuangpornpitak and S. Tia, "Opportunities and Challenges of Integrating Renewable Energy in Smart Grid System," *Energy Procedia*, vol. 34, pp. 282–290, 2013.

[2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997.

[3] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, S. Kapoor, J. Zschiegner, D. C. Maddix, H. Wang, M. W. Mahoney, K. Torkkola, A. G. Wilson, M. Bohlke-Schneider, and Y. Wang, "Chronos: Learning the Language of Time Series," *Transactions on Machine Learning Research*, 3 2024.

[4] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training Recurrent Neural Networks," *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, 11 2012.

[5] A. Gu, K. Goel, and C. Ré, "Efficiently Modeling Long Sequences with Structured State Spaces," in *International Conference on Learning Representations*, 10 2022.

[6] O. Doelle, N. Klinkenberg, A. Amthor, and C. Ament, "Probabilistic Intraday PV Power Forecast Using Ensembles of Deep Gaussian Mixture Density Networks," *Energies*, vol. 16, no. 2, p. 646, 1 2023.

[7] H. F. Hamann, B. Gjorgiev, T. Brunschwiler, L. S. Martins, A. Puech, A. Varbella, J. Weiss, J. Bernabe-Moreno, A. B. Massé, S. L. Choi, I. Foster, B.-M. Hodge, R. Jain, K. Kim, V. Mai, F. Mirallès, M. De Montigny, O. Ramos-Leaños, H. Suprême, L. Xie, E.-N. S. Youssef, A. Zinflou, A. Belyi, R. J. Bessa, B. P. Bhattarai, J. Schmude, and S. Sobolevsky, "Foundation models for the electric power grid," *Joule*, vol. 8, no. 12, pp. 3245–3258, 12 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2542435124004707

[8] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," 12 2023.