



Contribution ID: 69

Type: **not specified**

SHORT-TERM FORECASTING OF HETEROGENEOUS WIND POWER FLEETS: GLOBAL TEMPORAL FUSION TRANSFORMER LEVERAGING STATIC COVARIATES

Friday, March 27, 2026 12:10 PM (20 minutes)

Introduction and Motivation

With the rapid expansion of renewable energy sources, which accounted for 17% of global electricity production in 2024, accurate short-term wind power forecasting has become a prerequisite for grid stability and efficient electricity trading¹. While deep learning architectures have replaced statistical methods as the state-of-the-art, current industry practices predominantly rely on specialized local models trained for individual assets. These site-specific models often struggle to generalize when data is scarce and create significant operational complexity for large portfolios. Although global models offer a scalable alternative, they historically fail to outperform local baselines due to the high heterogeneity of wind farms, which differ significantly in local meteorology and technical turbine parameters. This study investigates whether a centralized, global Temporal Fusion Transformer (TFT) can outperform specialized local models in 48-hour-ahead forecasting by leveraging static context variables to learn shared physical representations across diverse sites.

Methodology

The study employs a Global Temporal Fusion Transformer (TFT) trained on a heterogeneous portfolio of 100 wind farms across Germany. The dataset combines synthetic power profiles with real-world Numerical Weather Prediction (NWP) data from the ICON-D2 model. To address the challenge of site heterogeneity, the model architecture explicitly incorporates static covariates—specifically farm age, surface altitude, turbine hub height, rotor diameter, and power curve parameters. These time-invariant features allow the neural network to condition its predictions on the specific physical identity of each asset. The forecasting framework utilizes a 48-hour prediction horizon with a corresponding 48-hour lookback window. A novel feature engineering pipeline is introduced to maximize physical consistency while minimizing data redundancy. This includes the derivation of a “rotor-equivalent air density,” which aggregates vertical density profiles using a linear weighted arithmetic mean across the rotor swept area. This approach reduces feature space dimensionality compared to using raw multi-level NWP data while maintaining physical fidelity. The model distinguishes between three input types: static covariates (context), past-observed inputs (historical power and weather), and known future inputs (NWP forecasts).

Experimental Setup

The performance of the global TFT was evaluated against distinct local models using a rigorous cross-validation scheme. The dataset includes 100 randomly distributed sites comprising six recurring turbine models (e.g., Vestas V112, Enercon E-82) with ages ranging from 2 to 24 years and elevations up to 900 meters. Hyperparameters were optimized using Optuna’s Tree-Structured Parzen Estimator (TPE) to minimize the Root Mean Squared Error (RMSE). To isolate feature-driven performance gains, the model was configured to generate deterministic point forecasts trained on Mean Squared Error (MSE), rather than probabilistic quantiles.

Results and Discussion

The empirical results demonstrate that the global TFT significantly outperforms the local baselines. The global model achieved a relative improvement of over 8% in the coefficient of determination R^2 and reduced the RMSE by 5% on the test set. Statistical analysis using the Wilcoxon signed-rank test confirmed the significance of these gains. A critical finding of this study is that this superiority is strictly conditional on the inclusion of static covariates. An ablation study revealed that a global model lacking asset-specific context variables (the “no-context” model) failed to generalize effectively, resulting in a performance decline of 1.28% compared

to local models. This confirms that explicit modeling of heterogeneity is a prerequisite for successful cross-learning between sites.

Interpretability

Leveraging the TFT's Variable Selection Network (VSN), the study provides novel physical insights into the forecasting process. Analysis of static feature importance identified farm age and altitude as the most critical determinants of site heterogeneity. This aligns with physical expectations, as turbine efficiency degrades with age and high-altitude terrain introduces complex turbulence. Furthermore, the model revealed a dual modeling strategy for dynamic inputs: the encoder prioritizes historical power output to reconstruct turbine efficiency, while the decoder minimizes forecast uncertainty by leveraging a combination of robust wind speed layers rather than relying on air density or single atmospheric levels.

Conclusion

This work supports a paradigm shift from single-site to scalable multi-site forecasting architectures. The results prove that a single global model, when correctly conditioned with static physical metadata, can not only reduce operational complexity but also enhance predictive accuracy for large-scale renewable energy portfolios. The proposed framework offers a pathway for operators to leverage fleet-wide data effectively, overcoming the limitations of isolated local models.

Author: WALTER, Viktor (Hochschule Karlsruhe - University of Applied Sciences)

Co-author: Prof. WAGNER, Andreas (Hochschule Karlsruhe - University of Applied Sciences)

Presenter: WALTER, Viktor (Hochschule Karlsruhe - University of Applied Sciences)

Session Classification: Renewable Expansion & Planning