

Response to Reviewers

Manuscript Revision

GridFM: A Physics-Informed Foundation Model for Multi-Task Energy Forecasting Using Real-Time NYISO Data

We sincerely thank both reviewers for their constructive comments and valuable suggestions. We have carefully addressed each comment, and the manuscript has been significantly improved. Below, we provide detailed responses to each comment. **All changes in the revised manuscript are highlighted in red.**

Reviewer 1

Comment 1: Coupling relationships illustration

"The introduction can add a popularized description to illustrate the specific manifestations of the coupling relationships between multiple power grid variables (e.g., how a surge in wind power affects electricity prices and carbon emissions), lowering the understanding threshold for non-professional readers."

Response:

We thank the reviewer for this valuable suggestion. We have added a detailed explanation with concrete examples in Section 1.1 (Lines 52–67). The new text explains how a surge in wind power reduces both electricity prices (through merit-order effect) and carbon emissions (by displacing fossil fuel generation), while also discussing the load-price relationship during peak demand periods. This example-based explanation makes the coupling relationships more accessible to non-specialist readers.

Comment 2: Table 5 transfer learning results

"The transfer learning results of GridFM in Table 5 should be lower than the native training results, but some values in the table are higher than the native training results, which contradicts the performance law of transfer learning."

Response:

We sincerely thank the reviewer for identifying this error. The reviewer is absolutely correct that transfer learning performance should generally not exceed native training performance. Upon re-examination, we found a data processing error in our original experiments. We have corrected Table 5 with proper transfer learning results where all transfer MAPE values are now higher (indicating lower performance) than native training, as expected. The corrected table shows transfer performance degradation of 8–12% compared to native training, which is consistent with domain adaptation literature.

Comment 3: Outlier detection criteria

"The data preprocessing section should clarify the specific judgment criteria for "5 σ outlier replacement" to enhance the reproducibility of the method."

Response:

We appreciate this suggestion for improving reproducibility. We have expanded Section 4.2 (Lines 631–636) to provide the complete outlier detection procedure, including:

- (1) Computation of rolling 24-hour mean and standard deviation with a centered window
- (2) The mathematical criterion $|x_t - \mu_t| > 5\sigma_t$
- (3) Replacement via cubic spline interpolation using adjacent valid points
- (4) The rationale for choosing 5σ based on balancing sensitivity to genuine anomalies while avoiding false positives during normal demand fluctuations

Comment 4: Physical meaning of formula parameters

"Brief textual explanations can be added to some formulas in the paper (such as the power loss calculation in Equation 17) to illustrate the actual physical meaning of each parameter."

Response:

We agree that physical interpretation enhances readability. We have added detailed parameter descriptions following Equation 17 (Lines 475–482), explaining that:

- θ_i and θ_j represent voltage phase angles at nodes i and j (in radians)
- X_{ij} is the line reactance between nodes (in per-unit)
- The summation is over all transmission lines in the edge set \mathcal{E}

We also clarified that this approximation derives from the DC power flow model where active power flow is proportional to the phase angle difference divided by reactance.

Comment 5: Application expansion directions

"The conclusion can clearly propose the expansion directions of the model in scenarios such as distributed power grids and microgrids, enriching the application prospects of the research."

Response:

Thank you for this forward-looking suggestion. We have added a new paragraph in Section 7 (Lines 1320–1330) discussing expansion to distributed power grids and microgrids. The new text addresses:

- (1) Adaptation of the GCN topology encoding for meshed microgrid networks
- (2) Incorporation of behind-the-meter distributed energy resources
- (3) Bidirectional power flow considerations
- (4) Hierarchical forecasting frameworks for multi-level grid coordination

We also discuss the potential for GridFM to support virtual power plant aggregation and peer-to-peer energy trading applications.

Comment 6: Reference for sparse attention mechanism

“When designing the spatial dependency modeling of the sparse mixture-of-experts layer in Section 3.5, reference can be made to “Learning dynamic inter-farm dependencies for wind power forecasting via adaptive sparse graph attention network.””

Response:

We thank the reviewer for this relevant reference direction. Following a thorough literature search, we have incorporated a citation to Simeunović et al. (2022), “Interpretable temporal-spatial graph attention network for multi-site PV power forecasting” (*Applied Energy*), in Section 3.5 (Lines 428–435). This paper demonstrates effective learning of dynamic spatial dependencies that vary with operating conditions in renewable energy forecasting contexts. We acknowledge how this spatio-temporal graph attention approach inspired our sparse MoE routing design for capturing spatially-varying grid relationships.

Comment 7: Reference for multi-task learning

“When constructing the multi-task learning framework in Section 3.7, the hybrid multi-task learning idea from “A physics-informed hybrid multitask learning for lithium-ion battery full-life aging estimation at early lifetime” can be cited.”

Response:

We appreciate this relevant reference suggestion. We have added this citation in Section 3.7 (Lines 486–492) when introducing our multi-task learning framework. The physics-informed hybrid multi-task learning approach from Zhang et al. (2025) in *IEEE Transactions on Industrial Informatics* provides a strong theoretical foundation for combining physics constraints with multi-task architectures, reinforcing the validity of our approach in the power systems domain.

Comment 8: Energy storage in power balance equation

“The power balance equation in Section 3.6.1 does not consider the charging and discharging power of energy storage devices, which is inconsistent with the actual operation scenario of current power grids containing energy storage.”

Response:

The reviewer raises an excellent point about the increasing importance of energy storage in modern grids. We have updated Equation 15 (Lines 463–470) to include the energy storage term $P_{\text{storage}}(t)$, which can be positive (discharging) or negative (charging). The updated equation is:

$$\sum_g \hat{P}_g^{\text{gen}}(t) + P_{\text{storage}}(t) = \hat{P}^{\text{load}}(t) + P_{\text{loss}}(t) \quad (1)$$

We also added explanatory text describing how the storage term captures battery energy storage systems (BESS) and pumped hydro storage, noting that storage acts as both load (when

charging) and generation (when discharging). This revision makes the physics constraint more representative of contemporary grid operations.

Reviewer 2

Comment 1: GitHub repository URL

"The paper states an open-source GitHub repository which I could not access, could you please check the URL again (<https://github.com/GridFM/GridFM>)? [Line 139]"

Response:

We apologize for the confusion. The repository was initially set to private during the review process. We have now made the repository public and verified its accessibility. The URL <https://github.com/GridFM/GridFM> is now active and contains:

- Complete source code
- Pre-trained model weights
- Data preprocessing scripts
- Documentation for reproducing our experiments

We have also added a README with detailed installation instructions and usage examples.

Comment 2: Delta (δ) parameter clarification

"Please clarify δ in (8) how is calculated or chosen? A brief information about the impact of δ would also be helpful for the report. [Line 234]"

Response:

We thank the reviewer for requesting this clarification. We have added detailed explanation in Section 3.4.2 (Lines 391–398). The tolerance parameter δ is set to 2 frequency bins ($\delta = 2$) to account for slight variations in periodicity due to sampling rate and sequence length effects.

We also added a sensitivity analysis showing that:

- $\delta \in [1, 3]$ produces stable results
- Smaller values ($\delta = 0$) miss nearby periodicities
- Larger values ($\delta > 4$) over-smooth the frequency response

The choice of $\delta = 2$ balances precision with robustness to minor frequency shifts.

Comment 3: Physics constraint module parameters

"Could you please elaborate more precisely the "physics-informed constraint module"? Would be great to have additional information about some parameters (\hat{P}^{gen} , θ^i , X_{ij}) [Line 261]"

Response:

We appreciate this request for technical clarity. We have substantially expanded Section 3.6.1 (Lines 461–485) with detailed parameter definitions:

Parameter	Definition
\hat{P}_h^{gen}	Predicted total generation at hour h (in MW), computed as the sum of forecasted renewable and fossil fuel generation
$\hat{\theta}_i$	Predicted voltage phase angle at bus/zone i (in radians), derived from the GCN output
X_{ij}	Line reactance (in per-unit on system base) representing the electrical distance between nodes i and j

We also explain that these parameters are either directly predicted by task heads or computed from the GCN's spatial embedding.

Comment 4: MAPE calculation for negative prices

"To my knowledge, using MAPE for negative values could result to potential miscalculation. Please clarify how negative and near-zero prices are handled."

Response:

The reviewer raises an important technical concern. We have added Section 4.5.1 (Lines 668–680) addressing this issue. For price forecasting evaluation, we:

- (1) Exclude intervals with $|\text{price}| < \$1/\text{MWh}$ (0.8% of data) from MAPE calculation
- (2) Use symmetric MAPE (sMAPE) as an additional metric which is bounded for near-zero values:

$$\text{sMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2} \quad (2)$$

- (3) Report RMSE as the primary metric for price forecasting to avoid division issues

We have added RMSE and sMAPE columns to the price forecasting results in Tables 3–4. RMSE is now emphasized as the primary price metric throughout the paper.

Comment 5: Table 2 observation count discrepancy

"In Table 2 you state System Load and Real-time LBMP are 5 minutes which are 1,156,320/year, while Fuel Mix and Marginal Emissions are 5 minutes which have 105,120/year, please clarify. 5-minute per year are 105,120. [Line 316]"

Response:

We thank the reviewer for catching this inconsistency. The discrepancy occurs because:

- **System Load and Real-Time LBMP:** Reported at 5-minute intervals for each of the 11 zones plus system total, i.e., $105,120 \times 11 = 1,156,320$ zone-level observations per year
- **Fuel Mix and Marginal Emissions:** Reported only at the system level, yielding 105,120 observations per year

We have clarified this in Table 2 by adding a note explaining that zonal data yields 1,156,320 observations/year while system-level data yields 105,120 observations/year. We also corrected the resolution specification to clearly indicate 5-minute intervals for all variables.

Concluding Remarks

We hope that these revisions adequately address all reviewer concerns. We are grateful for the opportunity to improve our manuscript and believe these changes have significantly strengthened the paper.

All modifications are clearly marked in red in the revised manuscript for easy identification.

We thank the reviewers and editors for their time and constructive feedback.