PIST-Mamba: A Physics-Informed Spatio-Temporal Mamba for Photovoltaic Power Forecasting

Hanyang Xu*, Yachen Tang†, Ying Wang§, Min Xia*, Liguo Weng*, and Shi Liang*‡

*Nanjing University of Information Science and Technology, Nanjing, 210044, China

†Univers, Santa Clara, CA, 95054, US

§Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education Southeast University, Nanjing, Jiangsu 210096, China

‡Corresponding author: Shi Liang (e-mail: liangshi@nuist.edu.cn)

***Abstract*—Accurate short-term photovoltaic (PV) power forecasting is critical for maintaining grid stability, yet data-driven approaches are fundamentally hindered by the stochastic nature of cloud cover. Spatio-Temporal Graph Networks (STGNs) model spatial correlations but lack physical inductive bias. This black-box characteristic limits their generalization capability. To bridge this gap, this paper presents PIST-Mamba, a Physics-Informed Spatio-Temporal Mamba framework designed as an interpretable gray-box solution. The proposed model introduces a dual-physics innovation: a Physics-Informed Kalman Filtering Graph Network (PI-KFGN) to model the horizontal physics of wind-driven cloud advection, and a Physics-Informed Prediction Head (PI-Head) to simulate the vertical physics of local energy conversion. By hierarchically modeling both the spatial transport and local conversion of solar energy, PIST-Mamba synergizes physical laws with data-driven patterns. Experiments on a large-scale, anonymized, real-world dataset from a province in Northwest China with 229 PV stations demonstrate state-of-the-art performance, reducing Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) by 50.4% and 41.2%, respectively, compared to a strong data-driven STG-Mamba baseline.**

***Index Terms*—Photovoltaic power forecasting, spatio-temporal graph networks, state space models, physics-informed neural networks, deep learning, renewable energy.**

I. Introduction

The escalating integration of photovoltaic (PV) power into modern energy grids is pivotal for decarbonization, yet it introduces significant volatility due to the inherent intermittency of solar energy [1]. The primary source of this volatility is cloud cover, whose rapid and localized movement induces sharp fluctuations, or ramp events, in PV generation, posing a serious threat to grid stability [2]. Consequently, accurate short-term PV power forecasting (typically spanning horizons from one hour to 48 hours) is of paramount importance for reliable grid management and economic operation [3].

The evolution of PV forecasting methodologies reflects a progressive effort to capture these complex dynamics. Early approaches based on physical models, such as Numerical Weather Prediction (NWP), simulate atmospheric physics but lack the resolution for localized, high-frequency events [4]. In parallel, statistical methods like ARIMA are constrained by linearity assumptions [5]. To overcome these limitations, the field has increasingly adopted deep learning, evolving from single-station Recurrent or Convolutional Neural Networks (RNNs, CNNs) [6] to Spatio-Temporal Graph Networks (STGNs) that explicitly model spatial correlations among distributed stations [7], [8].

Recently, a new class of sequence models based on State Space Models (SSMs), notably Mamba [9], has emerged as a compelling alternative to long-standing architectures like Transformers [10]. Mamba's linear-time complexity ($O(L)$) offers significant efficiency advantages over the Transformer's quadratic complexity ($O(L^2)$), while its continuous-time formulation provides a natural foundation for integrating physics. An initial exploration of this architecture for generic spatio-temporal tasks, STG-Mamba [11], forms the architectural basis of the proposed work.

However, a fundamental deficiency persists: these advanced models operate as black boxes, risking poor generalization and physically implausible predictions. This has catalyzed the rise of Physics-Informed Machine Learning (PIML), with recent reviews highlighting a variety of hybrid techniques [12]. Rather than merely using physics as a soft constraint in the loss function [13], a more potent approach is to construct "gray-box" models where physical knowledge is embedded into the network architecture, a paradigm known as Theory-Guided Machine Learning (TGML) [14].

This paper adopts a gray-box philosophy and proposes **PIST-Mamba**, a **P**hysics-**I**nformed **S**patio-**T**emporal Mamba framework. The proposed framework enhance the state-of-the-art STG-Mamba architecture by injecting physical principles into its core components. The main contributions can be summerized as:

- A **Physics-Informed Kalman Filtering Graph Network (PI-KFGN)** is designed to models the horizontal physics of cloud shadow advection via a wind-guided graph network.
- A **Physics-Informed Prediction Head (PI-Head)** is developed to capture the vertical physics of local energy conversion through a gated fusion mechanism.

The proposed PIST-Mamba is validated on a large-scale, anonymized dataset derived from a real-world regional power system in a province of Northwest China, demonstrating its superiority in both accuracy and interpretability.

979-8-3315-8138-1

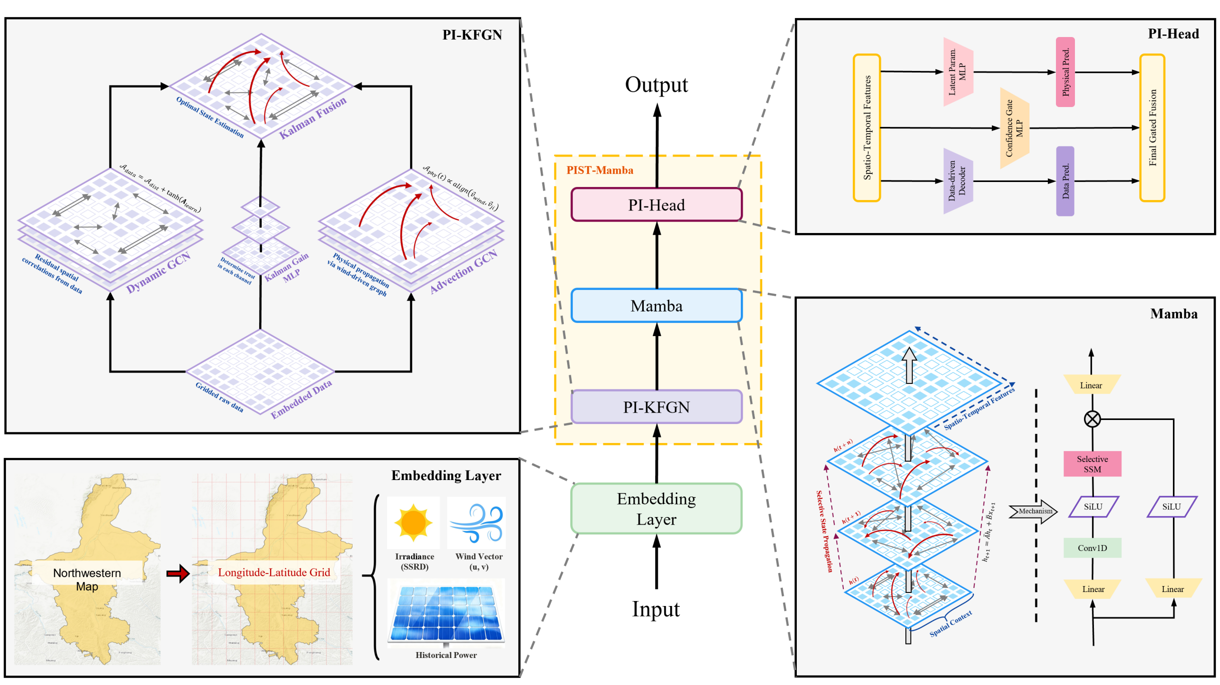

Fig. 1. Overall architecture of PIST-Mamba. The input is processed by stacked blocks containing a PI-KFGN for spatial physics and a Mamba block for temporal dynamics. The final prediction is generated by a physics-informed head.

II. Problem Formulation

The task of regional PV power forecasting is formalized as learning a predictive model on a dynamic spatio-temporal graph. A network of N PV stations is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ is the set of stations ($|\mathcal{V}| = N$) and $\mathcal{A} \in \mathbb{R}^{N \times N}$ is an adjacency matrix representing spatial proximity. At each time step t, a feature matrix $\mathbf{X}_t \in \mathbb{R}^{N \times F}$ contains the historical power output and exogenous meteorological variables for each station.

Given a historical sequence of features over a look-back window T_h, $\mathcal{X} = [\mathbf{X}_{t-T_h+1}, \ldots, \mathbf{X}_t]$, the objective is to learn a mapping function $\mathcal{F}$ that predicts the future power outputs $\mathcal{P} = [\mathbf{P}_{t+1}, \ldots, \mathbf{P}_{t+T_p}]$ over a prediction horizon T_p:

$$\mathcal{P} = \mathcal{F}(\mathcal{X}; \mathcal{G}), \tag{1}$$

where $\mathbf{P}_{t+i} \in \mathbb{R}^N$ is the power output vector at future time step $t+i$.

III. Methodology

The proposed PIST-Mamba is a gray-box architecture designed to synergize data-driven pattern recognition with physical principles. Starting from a foundational data-driven architecture, the paper then introduces the two key physics-informed innovations that transform it.

A. Foundational Architecture: Spatio-Temporal Graph Mamba

The foundation of our model is the Spatio-Temporal Graph Mamba (STG-Mamba) architecture [11], a powerful data-driven framework for generic spatio-temporal forecasting. Its structure consists of stacked residual blocks, where each block processes information spatially and then temporally.

1) Spatial Module: The spatial module employs a Dynamic Graph Convolutional Network. Its key feature is a trainable filter, $\mathbf{A}_{\text{learn}} \in \mathbb{R}^{N \times N}$, which is added to the static distance-based graph, $\mathcal{A}_{\text{dist}}$, to form an adaptive adjacency matrix. The graph convolution operation on the input features $\mathbf{X}' \in \mathbb{R}^{N \times D}$ can be summarized as:

$$\mathcal{A}_{\text{data}} = \mathcal{A}_{\text{dist}} + \tanh(\mathbf{A}_{\text{learn}}), \tag{2}$$

$$\mathbf{H}_{\text{spatial}} = \sigma(\mathcal{A}_{\text{data}} \mathbf{X}' \mathbf{W}_s). \tag{3}$$

2) Temporal Module: The temporal module utilizes a Mamba block [9] to model sequential dependencies. It implements a selective State Space Model (SSM) governed by the continuous-time linear ODE:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}u(t). \tag{4}$$

While effective, this foundational architecture is fundamentally a black-box model, which our work aims to enhance with explicit physical knowledge.

B. The Proposed PIST-Mamba Framework

The framework re-architects the STG-Mamba's components to create the PIST-Mamba, injecting physical principles into both the spatial module and the prediction head, as illustrated in Fig. 1.

1) PI-KFGN for Spatial Advection Physics: Inspired by the dual-channel structure of the Kalman Filter, our PI-KFGN is an interpretable data assimilation unit with three key components.

979-8-3315-8138-1

- **Physics Prediction Channel:** This channel models the transport of irradiance patterns via wind-driven advection. A time-varying, directed physical adjacency matrix is constructed as $\mathcal{A}_{\text{phy}}(t) \in \mathbb{R}^{N\times N}$. Its weight from an upstream node j to a downstream node i is determined by an alignment factor, $\alpha_{ij}(t)$, between the wind vector, $\vec{v}_{\text{wind}}(t)$, and the geographical vector $\vec{v}_{ji}$,

$$\alpha_{ij}(t) = \max\left(0, \frac{\vec{v}_{\text{wind}}(t)\cdot\vec{v}_{ji}}{||\vec{v}_{\text{wind}}(t)||\cdot||\vec{v}_{ji}||}\right). \quad (5)$$

The full matrix combines this directional information with geographical proximity, $\mathcal{A}_{\text{dist}}$:

$$\mathcal{A}_{\text{phy}}(t)_{ij} = \mathcal{A}_{\text{dist},ij}\cdot\alpha_{ij}(t). \quad (6)$$

The resulting physics-based state estimate, $\mathbf{H}_{\text{phy}}(t)$, is then computed via graph convolution:

$$\mathbf{H}_{\text{phy}}(t) = \sigma(\mathcal{A}_{\text{phy}}(t)\mathbf{X}'(t)\mathbf{W}_{\text{phy}}). \quad (7)$$

- **Data Observation Channel:** In parallel, a standard dynamic GCN, as defined in Eq. 3, learns residual spatial correlations to yield a data-based observation, $\mathbf{H}_{\text{data}}(t)$.
- **Fusion Layer:** A learnable Kalman Gain, $\mathbf{K}(t) \in \mathbb{R}^{N\times D}$, adaptively weighs the two channels. The final fused spatial feature, $\mathbf{H}_{\text{final}}(t)$, is their dynamic combination:

$$\mathbf{H}_{\text{final}}(t) = (1-\mathbf{K}(t))\odot\mathbf{H}_{\text{phy}}(t)+\mathbf{K}(t)\odot\mathbf{H}_{\text{data}}(t). \quad (8)$$

2) Physics-Informed Prediction Head (PI-Head): After a Mamba backbone processes the sequence of spatially-aware features to produce a final latent representation $\mathbf{h}(t)$, the PI-Head simulates the vertical physics of local energy conversion based on the standard engineering model for PV power:

$$P_{\text{phy}} \propto \eta_{\text{eff}}\cdot G(t)\cdot[1+k_T(T_c(t)-T_{\text{ref}})], \quad (9)$$

where η_{eff} (effective efficiency) and T_c (cell temperature) are key unobserved variables. The proposed gray-box head consists of the following components.

- **Parameter Network:** An MLP takes $\mathbf{h}(t)$ to estimate unobserved physical parameters like effective efficiency $\hat{\eta}_{\text{eff}}(t)$ and cell temperature $\hat{T}_c(t)$:

$$[\hat{\eta}_{\text{eff}}(t), \hat{T}_c(t)] = \text{MLP}_{\text{param}}(\mathbf{h}(t)). \quad (10)$$

- **Parallel Decoders:** A *Physics Decoder* computes a physics-based power estimate, $\hat{\mathbf{P}}_{\text{phy}}(t)$. In parallel, a *Data Decoder* computes a purely data-driven prediction, $\hat{\mathbf{P}}_{\text{data}}(t)$:

$$\hat{\mathbf{P}}_{\text{data}}(t) = \text{MLP}_{\text{data}}(\mathbf{h}(t)). \quad (11)$$

- **Gated Fusion:** A final confidence gate, $g(t) \in [0,1]$, is also predicted from $\mathbf{h}(t)$ to weigh the trustworthiness of the two paths. The ultimate model output is their learned interpolation:

$$\hat{\mathbf{P}}_{\text{final}}(t) = g(t)\odot\hat{\mathbf{P}}_{\text{phy}}(t) + (1-g(t))\odot\hat{\mathbf{P}}_{\text{data}}(t). \quad (12)$$

This architecture anchors the final prediction in physical plausibility while retaining the flexibility to fit complex real-world data.

IV. Case Study

A. Case Scenario

The proposed PIST-Mamba is validated on a anonymised, real-world case study from a regional power system in Northwest China, featuring 229 utility-scale solar PV stations. We utilize a four-month dataset with 15-minute resolution power records and hourly meteorological data (SSRD, wind velocity) from the ERA5 reanalysis dataset [15]. The dataset is chronologically partitioned into training (70%), validation (10%), and testing (20%) sets. All models are trained using the Adam optimizer to minimize MSE loss. The experiments are conducted on a workstation with an NVIDIA GeForce RTX 4090 GPU and an Intel Core i9-13900K CPU, utilizing the PyTorch framework.

B. Performance of PIST-Mamba

1) Comparative Analysis: To benchmark the performance of PIST-Mamba, we compare it against several representative models: a classical spatio-temporal method (**STGCN** [7]), a Transformer-based STGN (**PDFormer** [10]), a non-graph sequential model (**Mamba** [9]), and our direct data-driven counterpart without physics infusion (**STG-Mamba** [11]).

TABLE I
OVERALL FORECASTING PERFORMANCE COMPARISON ON THE TEST SET

Model	**MAE (MW) ↓**	**RMSE (MW) ↓**
Mamba [9]	6.2487	14.3891
STGCN [7]	4.4182	10.3221
PDFormer [10]	3.0623	9.8674
STG-Mamba [11]	2.9639	9.5907
PIST-Mamba (Ours)	**1.4711**	**5.6349**

As presented in Table I, PIST-Mamba achieves state-of-the-art performance, substantially outperforming all baselines. Notably, it reduces the MAE and RMSE by 50.4% and 41.2%, respectively, compared to its direct data-driven counterpart, STG-Mamba. This significant improvement underscores the profound impact of integrating physical principles, transforming the black-box learner into a more robust and accurate gray-box forecaster.

2) Scenario-Based Performance: Beyond aggregate metrics, we evaluate the model's performance in four distinct operational scenarios to assess its robustness. Fig. 2 visualizes the forecasting behavior during nighttime, sunrise (ramp-up), intermittent cloud cover (volatile), and sunset (ramp-down) periods.

The visualizations confirm the superiority of PIST-Mamba across diverse conditions. In low-generation and ramp scenarios (Fig. 2a, b, d), our model provides physically consistent and highly accurate forecasts. Most critically, during the highly volatile scenario (Fig. 2c), PIST-Mamba successfully captures the sharp, high-frequency fluctuations characteristic of fast-moving clouds, whereas purely data-driven models typically

979-8-3315-8138-1

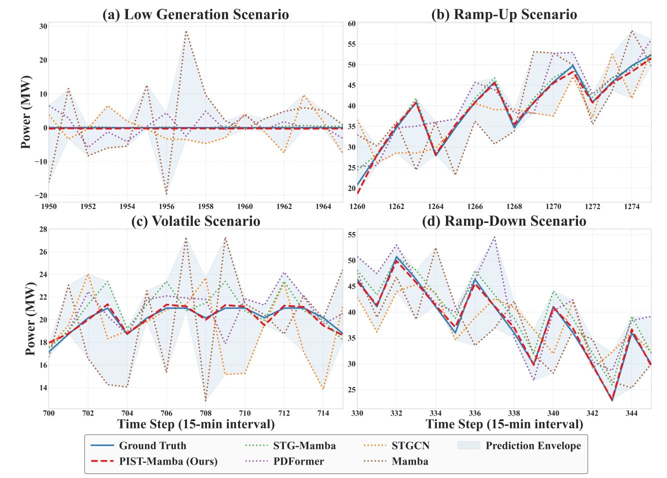

Fig. 2. Scenario-based performance analysis. PIST-Mamba (red, dashed) consistently demonstrates superior tracking of the Ground Truth (blue, solid) across all four typical scenarios compared to the baseline models (represented by the shaded prediction envelope).

yield overly smoothed and delayed responses. This demonstrates our model's enhanced capability to predict critical ramp events, a key requirement for reliable grid operation.

C. Component and Interpretability Analysis

To validate our dual-physics design and understand its inner workings, we conduct both an ablation study and a deep dive into the model's learned physical representations.

1) Ablation Study of Physics Components: We analyze the contribution of each physics-informed component by evaluating two degraded model variants: one without the spatial advection physics (*w/o PI-KFGN*) and another without the local conversion physics (*w/o PI-Head*). Table II reveals that removing either component leads to a significant performance drop, confirming that both innovations are crucial. The synergistic effect of combining the horizontal (spatial) and vertical (local) physics allows the full PIST-Mamba to achieve the best performance.

TABLE II
ABLATION STUDY OF PHYSICS-INFORMED COMPONENTS

Model Variant	MAE (MW) ↓	RMSE (MW) ↓
PIST-Mamba (Full Model)	**1.4711**	**5.6349**
w/o PI-KFGN (No Spatial Physics)	2.6352	8.0181
w/o PI-Head (No Local Physics)	1.9255	7.1823
STG-Mamba (No Physics)	2.9639	9.5907

2) Interpretability of Innovation 1: Learned Advection Patterns: Our PI-KFGN is designed to interpret meteorological data and model the physical process of cloud shadow advection. To verify this, we visualize the theoretical information flow field generated by the physics-prediction channel under two distinct, hypothetical wind conditions. As depicted in Fig. 3, the model generates a physically coherent and dynamic advection field.

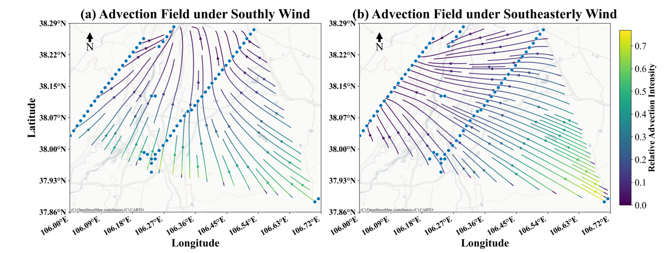

Fig. 3. Visualization of the theoretical advection field. The streamlines (colored lines) represent the learned direction of information flow. (a) Under a south wind, the entire flow field correctly moves northwards. (b) When the wind shifts to a southeasterly direction, the flow field rotates accordingly, moving towards the northwest.

The streamlines, representing the dominant paths of influence, clearly align with the prevailing wind direction. In Fig. 3(a), under a southerly wind, the information flow field shows a distinct northward propagation pattern. When the wind direction changes to a southeasterly wind in Fig. 3(b), the entire flow field rotates, correctly propagating towards the northwest. This provides direct, visual evidence that our PI-KFGN component successfully embeds the physical laws of advection.

3) Interpretability of Innovation 2: Learned Physical Parameter Correlations: A cornerstone of our gray-box design is its ability to learn physically meaningful internal representations. We validate this by analyzing the statistical correlations between key external variables and the internal parameters estimated by our physics-constrained head. Fig. 4 presents an enhanced correlation matrix based on the model's simulated behavior on the test set.

The matrix reveals several key relationships that confirm our model has learned physically consistent behavior:

- **Learned Thermodynamics:** A strong positive correlation ($r = +0.98$) is observed between the estimated effective cell temperature ($\hat{T}_{\text{eff}}$) and the input solar radiation (SSRD). This demonstrates that the model has successfully learned the fundamental thermodynamic principle of solar heating without any direct supervision on temperature. Furthermore, a clear negative correlation ($r = -0.62$) exists between the estimated efficiency ($\hat{\eta}_{\text{eff}}$) and temperature ($\hat{T}_{\text{eff}}$), correctly capturing the negative temperature coefficient effect inherent in PV cells.
- **Intelligent Gating Logic:** The most insightful finding is the strong negative correlation ($r = -0.78$) between the confidence gate ($g(t)$) and SSRD. This seemingly counter-intuitive result showcases the model's sophisticated reasoning. Since the NWP-derived SSRD is a coarse, regional average, it is often a poor predictor of the actual, highly variable irradiance at a specific site during cloudy conditions. The model learns this unreliability and adopts a strategy: when SSRD is high (i.e., during any daytime hour), the potential for unmodeled cloud effects

979-8-3315-8138-1

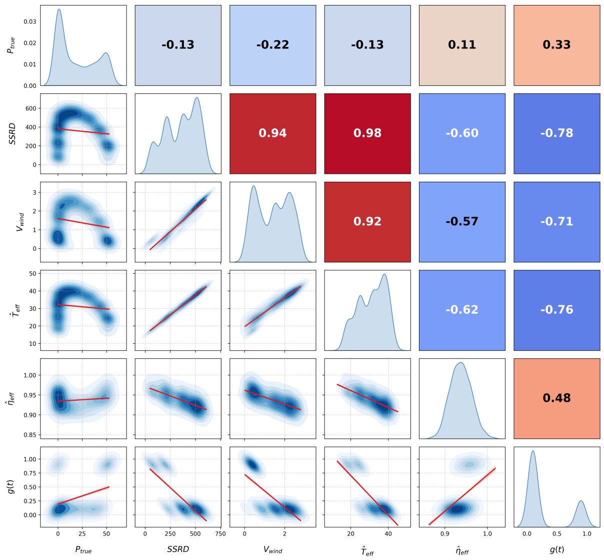

Fig. 4. Correlation matrix of external and learned internal variables during daytime conditions. The upper triangle displays the Pearson correlation coefficient, with the cell background color-coded from red (strong positive) to blue (strong negative). The diagonal shows the kernel density estimate (KDE) for each variable. The lower triangle presents 2D-KDE plots overlaid with a linear regression line.

is also high, thus it reduces its trust ($g(t)$ is low) in the simple physics formula. Conversely, the moderate positive correlation ($r = +0.33$) between the gate and the ground truth power (P_{true}) confirms that the model correctly learns to increase its trust in the physics path when the actual power output is high and stable, which is characteristic of clear-sky conditions.

This analysis confirms that the internal variables of PIST-Mamba are not arbitrary latent features but interpretable parameters that behave in accordance with physical laws and an intelligent, data-informed understanding of the physical model's limitations.

V. Conclusion

In this paper, the proposed PIST-Mamba, a novel gray-box spatio-temporal framework that injects physical principles into a state-of-the-art STG-Mamba architecture to address the challenges of short-term photovoltaic power forecasting. Our core contribution is a dual-physics innovation: a PI-KFGN that models the horizontal physics of cloud advection and a PI-Head that simulates the vertical physics of local energy conversion. Comprehensive experiments on a large-scale real-world dataset confirmed the superiority of our design. PIST-Mamba achieved state-of-the-art performance, reducing MAE by 50.4% and RMSE by 41.2% over a strong data-driven baseline, and demonstrated enhanced interpretability by learning to dynamically synergize physical laws with data-driven observations. While effective, our embedded physical models remain simplified. Future work will focus on integrating more sophisticated atmospheric models and on investigating the framework's adaptability to other spatio-temporal systems.

Acknowledgments

This work was supported by the Science and Technology Project of SGCC, named Research on AI Trustworthy Assessment Technology for Power System Dispatching in Typical Scenarios (5108-202417039A-1-1-ZN).

References

[1] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.

[2] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-term spatio-temporal forecasting of photovoltaic power production," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 538–546, 2018.

[3] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[4] R. Meenal, D. Binu, K. Ramya, P. A. Michael, K. Vinoth Kumar, E. Rajasekaran, and B. Sangeetha, "Weather forecasting for renewable energy system: a review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2875–2891, 2022.

[5] S. A. Haider, M. Sajid, H. Sajid, E. Uddin, and Y. Ayaz, "Deep learning and statistical methods for short-and long-term solar irradiance forecasting for islamabad," *Renewable Energy*, vol. 198, pp. 51–60, 2022.

[6] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, p. 113315, 2019.

[7] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[8] L. Cheng, H. Zang, T. Ding, Z. Wei, and G. Sun, "Multi-meteorological-factor-based graph modeling for photovoltaic power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 3, pp. 1593–1603, 2021.

[9] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.

[10] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, 2023, pp. 4365–4373.

[11] L. Li, H. Wang, W. Zhang, and A. Coster, "STG-Mamba: Spatial-temporal graph learning via selective state space model," *arXiv preprint arXiv:2403.12418*, 2024.

[12] L. de Oliveira Santos, T. AlSkaif, G. C. Barroso, and P. C. M. de Carvalho, "Photovoltaic power estimation and forecast models integrating physics and machine learning: A review on hybrid techniques," *Solar Energy*, vol. 284, p. 113044, 2024.

[13] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[14] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2318–2331, 2017.

[15] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

979-8-3315-8138-1