JOURNAL OF ZHEJIAG UNIVERSITY-SCIENCE A

Physics-informed deep learning for data-efficient and robust photovoltaic power forecasting

Chang HUANG, Xuanbin HUANG, Jinmin GUO, Zhuo CAO, Ting HE, Wentao SHANG

Cite this as: Chang HUANG, Xuanbin HUANG, Jinmin GUO, Zhuo CAO, Ting HE, Wentao SHANG, 2026. Physics-informed deep learning for data-efficient and robust photovoltaic power forecasting. *Journal of Zhejiang University-SCIENCE A*, 27(8):885896. https://doi.org/10.1631/jzus.A2500582

PIDL–HM: A Synergistic Framework

White-box front-end

Leverages radiation transfer equations to transform raw meteorological data into stable, high-fidelity physical features.

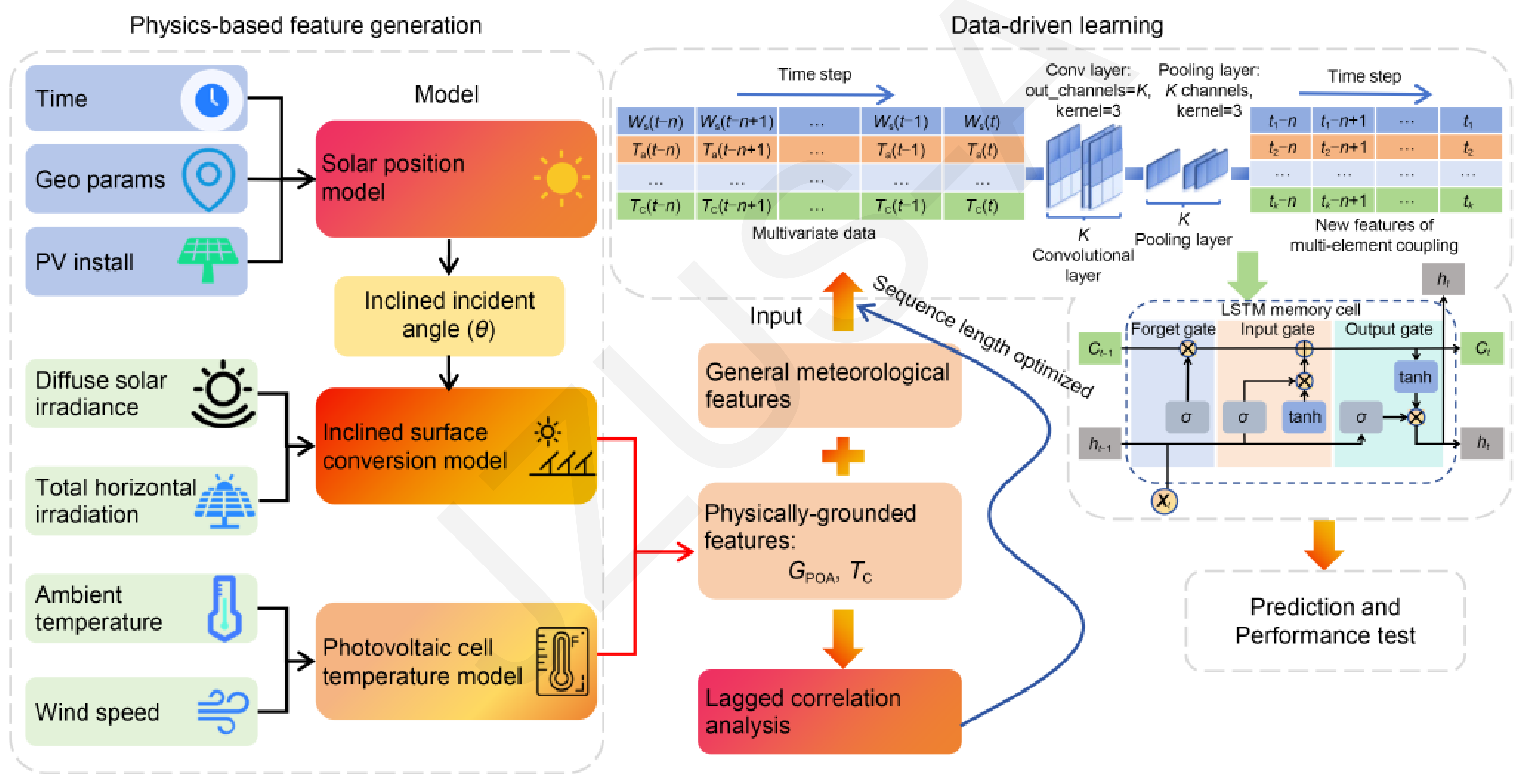
- Solar position modeling
- Plane-of-array irradiance
- Module temperature prediction

Black-box back-end

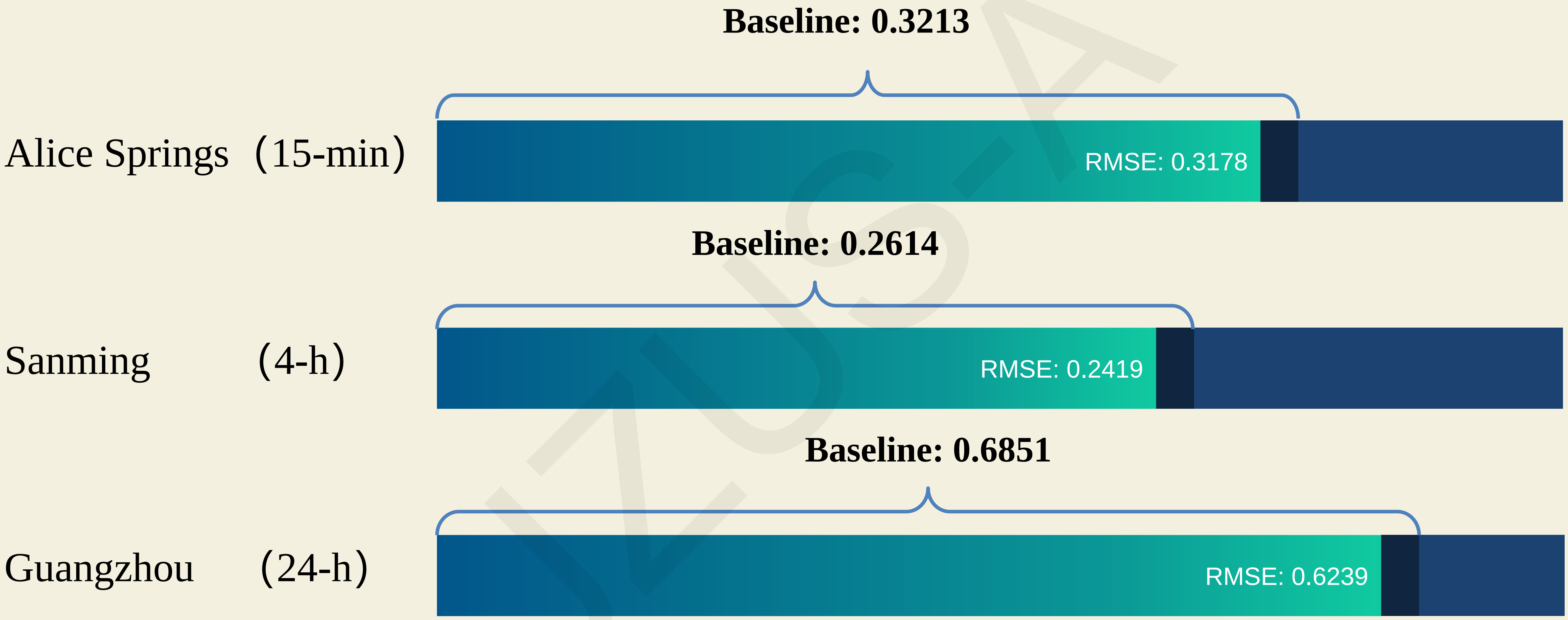
Uses a hybrid CNN–LSTM architecture to capture complex nonlinear dependencies and residual dynamics.

- CNN: spatial correlation extraction
- LSTM: long-term temporal memory
- Sequential integration strategy

Workflow of PIDL-HM



Multi–Horizon Validation

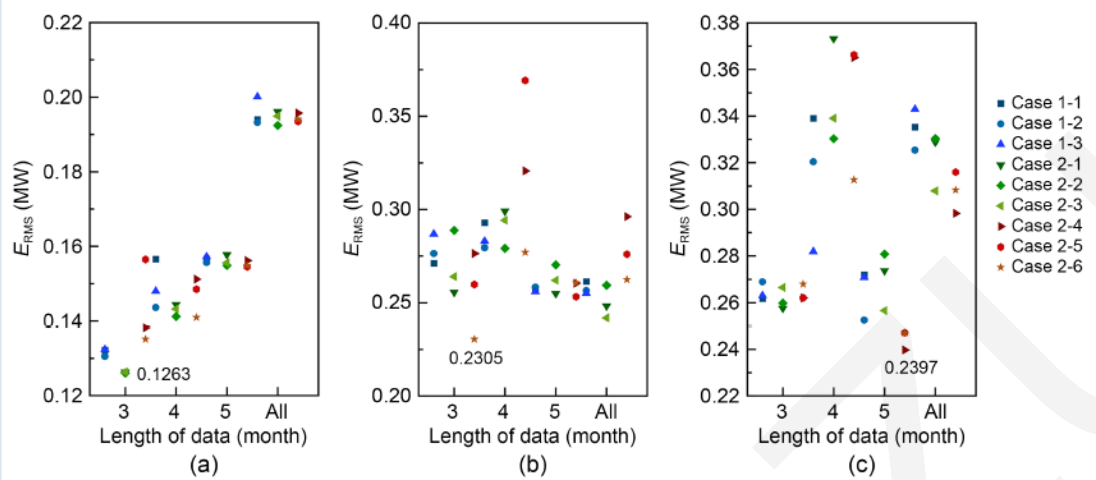


Validated across China and Australia, the PIDL-HM demonstrates superior performance:

- **Up to 8.93% reduction in RMSE compared to purely data-driven baselines**

Data Efficiency and Data Distribution Shift Experiment

Data efficiency: success with small-sample



- ➢ Taking the Sanming region as an example, it can be seen that with a small sample size, the experimental groups that embed prior physical knowledge outperform the baseline model driven purely by data.

Data distribution shift experiment

- ➢ The more pronounced the distribution shift is, the greater the benefit of embedded physical priors. Accuracy improves by 6.87% during strong seasonal transitions.

Synergy via Model Fusion

Forecasting scenario	XGBoost baseline	PIDL-HM (proposed)	Combined (fusion)
15-min (Alice Springs)	0.3138 MW	0.3178 MW	**0.3083 MW**
4-h (Sanming)	0.2426 MW	0.2419 MW	**0.2333 MW**
24-h (Guangzhou)	0.6059 MW	0.6239 MW	**0.5789 MW**

PIDL-HM establishes a physically constrained foundation for ensemble learning. The "PIDL-HM+XGBoost" ensemble consistently outperforms individual models, proving that physical constraints enhance the entire forecasting pipeline.

Conclusions

- Deep synergy: blends physical laws with AI, reducing day-ahead RMSE by up to 8.93%.
- Data efficient: outperforms baselines by 17.8% with just 3 months of data—ideal for new PV plants.
- High robustness: resists cross-seasonal distribution shifts, improving accuracy by 6.87%.
- Explainable AI: principle-guided feature engineering ensures transparent and practical deployment.
- Ensemble-ready: enhances overall pipeline performance when used as a base model (e.g., PIDL-HM+XGBoost).

Current limitations include the model's dependence on the quality of meteorological forecasts and the potential computational overhead for ultra-large-scale applications. Future research will focus on developing explicit NWP bias-correction techniques and exploring more efficient physics-informed architectures to enhance the framework's scalability and practical scope.