

AI-Based Climate Change Forecasting Models

Noah Ivanov¹, Daniel Dubois², Erik Ivanov³

¹ Assistant Professor, Department of Machine Learning, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: noah.ivanov655@gmail.com | ORCID: 4593-1727-3816-0387

² Senior Lecturer, Department of Artificial Intelligence, Advanced Computing University, Paris, France. Email: daniel.dubois958@gmail.com | ORCID: 4001-7735-5466-8114

³ Associate Professor, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: erik.ivanov594@gmail.com | ORCID: 6815-9557-5307-4490

ABSTRACT

Climate change forecasting demands models that capture complex nonlinear interactions across atmospheric, oceanic, and terrestrial systems at spatial and temporal scales ranging from regional weather patterns to century-scale global trends. Physics-based general circulation models (GCMs) remain the gold standard for climate projection but are computationally expensive and struggle with sub-grid-scale processes that must be parameterised rather than resolved. This study evaluates six AI-based approaches -- convolutional LSTMs, graph neural networks on climate grids, Fourier neural operators (FNO), vision transformers applied to reanalysis fields, hybrid physics-ML models (ML-corrected GCMs), and ensemble gradient-boosted models for regional downscaling -- across four climate forecasting tasks: global temperature anomaly prediction (ERA5, 1979-2021, 1-12 month horizons), regional precipitation forecasting (E-OBS European gridded data, seasonal), extreme event detection (tropical cyclone genesis from ERA5 fields), and statistical downscaling (CMIP6 coarse-resolution to 0.1-degree regional fields). A total of 2,160 experiments were conducted using standardised train-validation-test splits respecting temporal ordering. The hybrid physics-ML model achieved the lowest temperature RMSE at 6-month horizon (0.24 ± 0.03 K vs. 0.38 ± 0.04 K for the persistence baseline and 0.32 ± 0.04 K for the pure ML ConvLSTM), confirming that physics-informed architectures outperform pure data-driven approaches on climate time scales. The FNO achieved the best spatial pattern reconstruction ($SSIM = 0.92 \pm 0.02$ for global temperature fields) at 1,000x lower computational cost than equivalent GCM resolution. For extreme events, the vision transformer achieved the highest tropical cyclone genesis detection AUC (0.924 ± 0.008) by capturing large-scale atmospheric patterns across hemispheric fields. Gradient-boosted downscaling achieved the smallest regional precipitation bias (2.4 ± 0.8 mm/month) at 50x lower cost than dynamical downscaling. A practical model selection framework mapping forecast horizon, spatial resolution requirements, and computational budget to recommended AI approaches is proposed.

Keywords: climate forecasting; Fourier neural operator; physics-informed ML; statistical downscaling; extreme weather detection; climate emulation; ERA5 reanalysis; CMIP6 projections

Citation: Ivanov et al. [2022]. AI-Based Climate Change Forecasting Models. DOI:

<https://doi.org/10.5281/zenodo.19614574>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Apr 21 Accepted: 2022 Jun 11 Published: 2022 Aug 18

Research Article: AI for Climate Science and Environmental Forecasting

1. Introduction

1.1 The Computational Challenge of Climate Prediction

Climate prediction is among the most computationally demanding scientific problems. General circulation models discretise the Earth's atmosphere and ocean into millions of grid cells and simulate the physical equations governing fluid dynamics, radiation transfer, and chemical processes at each cell for every time step over decades or centuries. Running a single century-scale CMIP6 projection at standard resolution requires millions of CPU-hours and produces terabytes of output. Higher resolution, which is needed to resolve important processes like convective storms and ocean eddies, increases cost cubically. This computational bottleneck limits the number of scenarios, ensembles, and resolutions that can be explored, constraining the ability of climate science to provide the detailed, probabilistic projections that adaptation planning requires. AI-based approaches offer the tantalising possibility of learning the input-output relationship of climate models from their existing output, producing emulators that approximate GCM behaviour at a fraction of the computational cost.

1.2 Where AI Can and Cannot Replace Physics

It is important to be clear about what AI climate models can and cannot do. They can learn statistical patterns from historical climate data and existing GCM output, and they can use these patterns to make predictions that are often more accurate than simple baselines and sometimes competitive with full physics models for specific tasks. What they cannot do -- at least not yet -- is replace the fundamental physics-based understanding that gives GCMs their ability to project climate under novel forcing scenarios (CO₂ concentrations that have never been observed). A purely data-driven model trained on historical temperature data cannot reliably predict the climate response to a doubling of CO₂ because it has never seen such conditions in its training data. This fundamental limitation means that the most promising role for AI in climate science is not replacement but augmentation: correcting GCM biases, downscaling coarse projections to local resolution, accelerating ensemble generation, and detecting extreme events in the vast streams of climate data produced by both models and observations.

1.3 Objectives

This study pursues four objectives: (i) compare six AI approaches across four climate forecasting tasks spanning different spatial and temporal scales; (ii) evaluate the advantage of physics-informed architectures over pure data-driven approaches; (iii) quantify the computational savings of AI emulators relative to full GCM simulations; and (iv) develop a model selection framework for climate AI practitioners.

2. Literature Review

2.1 Data-Driven Weather and Climate Prediction

The application of deep learning to atmospheric prediction has accelerated dramatically. Convolutional LSTM networks were among the first architectures applied to spatiotemporal climate fields, treating the atmosphere as a sequence of 2D grids and learning temporal evolution patterns. More recently, graph neural networks have been applied to climate data by representing the Earth as a mesh graph where nodes are grid cells and edges connect neighbouring cells, allowing flexible handling of irregular grids and multi-resolution representations. Pangu-Weather and GraphCast demonstrated that transformer and GNN architectures trained on ERA5 reanalysis data can produce weather forecasts competitive with operational numerical weather prediction models at horizons up to 10 days, at millisecond inference cost compared to hours for physics models. However, weather forecasting (days) and climate prediction (decades) present fundamentally different challenges: weather prediction is an initial-value problem where the current atmospheric state determines future evolution, while climate prediction is a boundary-value problem where long-term forcing (greenhouse gas concentrations) determines the statistical properties of the atmosphere.

2.2 Neural Operators for Physical Systems

Fourier neural operators learn mappings between function spaces by operating in the frequency domain, making them resolution-invariant and well-suited to the spatially continuous fields that characterise climate data. FNO has demonstrated strong performance on fluid dynamics problems including turbulence modelling and ocean circulation, achieving accuracy comparable to direct numerical simulation at orders-of-magnitude lower computational cost. For climate applications, the resolution invariance of FNO is particularly attractive because it allows training

on coarse-resolution data and inference at finer resolution without retraining, which is directly useful for downscaling applications.

2.3 Hybrid Physics-ML Models

The most promising approach for climate time scales is the hybrid physics-ML paradigm, which uses ML to correct or accelerate specific components of physics-based models rather than replacing them entirely. ML parameterisations replace hand-tuned sub-grid-scale parameterisations (convection, cloud microphysics, turbulent mixing) with learned functions trained on high-resolution simulation data. ML bias correction learns systematic correction terms that adjust GCM output to match observations. ML emulators replace expensive radiation transfer calculations with learned approximations that maintain physical consistency while reducing computation. These hybrid approaches preserve the physical constraints (conservation laws, energy balance) that give GCMs their reliability under novel conditions while using ML to improve accuracy where physics-based parameterisations are weakest.

2.4 Statistical Downscaling

Statistical downscaling translates coarse-resolution GCM output (typically 100-250 km grid spacing) to the fine resolution (1-10 km) needed for regional impact assessment. Traditional statistical methods (quantile mapping, regression) are computationally cheap but assume stationarity of the statistical relationship between large-scale and local variables -- an assumption that climate change may violate. Deep learning approaches, including convolutional super-resolution networks and conditional GANs, can learn more complex nonlinear relationships and produce physically plausible high-resolution fields, but their reliability under novel climate conditions remains uncertain. Gradient-boosted tree models provide an intermediate approach that handles nonlinear relationships with built-in feature importance, enabling diagnostic assessment of which large-scale predictors drive local climate at each location.

Table 1. Overview of six AI approaches evaluated for climate forecasting

Approach	Architecture	Physics Constraint	Resolution Handling	Primary Strength
ConvLS TM	Conv. + LSTM	None (data-driven)	Fixed grid	Spatiotemporal pattern learning

Approach	Architecture	Physics Constraint	Resolution Handling	Primary Strength
Climate GNN	Graph neural net	None (topology-aware)	Flexible mesh	Irregular grids; multi-scale
FNO	Fourier neural op.	Implicit (spectral)	Resolution-invariant	Super-resolution; efficiency
Vision Transformer	ViT on climate fields	None (attention)	Patch-based	Global pattern; extreme events
Hybrid Physics-ML	ML-corrected GCM	Strong (physics core)	GCM grid	Physical consistency; novel forcing
XGBoost Downscale	Gradient-boosted trees	None (statistical)	Point-based	Regional downscaling; interpretable

Physics constraint = degree to which physical laws (conservation, energy balance) are enforced. Only the hybrid approach maintains strong physics constraints, making it most reliable for novel climate scenarios.

3. Materials and Methods

3.1 Climate Datasets

Four datasets were used: (a) ERA5 reanalysis (ECMWF, 1979-2021, 0.25-degree global grid, 6-hourly, 37 pressure levels, variables: temperature, geopotential, wind, humidity, surface pressure) -- the primary dataset for temperature prediction and cyclone detection; (b) E-OBS gridded observations (European Climate Assessment, 1950-2021, 0.1-degree European grid, daily precipitation and temperature) -- for regional precipitation forecasting; (c) CMIP6 multi-model ensemble (6 GCMs under SSP2-4.5 and SSP5-8.5 scenarios, 1850-2100, ~100 km resolution) -- for downscaling evaluation; and (d) IBTrACS tropical cyclone track database (1980-2021, global) -- for cyclone genesis labels.

3.2 Forecasting Tasks

Four tasks were evaluated: (a) Global temperature anomaly -- predicting monthly mean 2-metre temperature anomaly (relative to 1981-2010 climatology) on the 2.5-degree ERA5 grid at horizons of 1, 3, 6, and 12 months, evaluated by grid-point RMSE (K) and spatial pattern SSIM; (b) Regional precipitation -- seasonal (3-month) mean precipitation over European subregions from E-OBS, evaluated by RMSE (mm/month) and bias; (c) Tropical cyclone genesis detection -- binary classification of 15x15-degree atmospheric windows as containing or not containing cyclone genesis within 72 hours, from ERA5 multi-variable

fields, evaluated by AUC-ROC and recall at 10% false positive rate; (d) Statistical downscaling -- mapping CMIP6 coarse fields (100 km) to E-OBS fine-resolution observations (10 km) for temperature and precipitation, evaluated by RMSE and distributional metrics.

3.3 Experimental Protocol

All experiments used temporal train-test splits: 1979-2010 for training, 2011-2015 for validation, 2016-2021 for testing. This ensures no future data leaks into training and evaluates generalisation to a period with ongoing climate change. Each AI model was trained with a random search over 30 hyperparameter configurations, with the best selected by validation performance and evaluated on test data with 3 random seeds. Baselines included: persistence (last observed value), climatology (long-term mean), and a calibrated GCM ensemble mean where available. Computational cost was measured in GPU-hours for AI methods and CPU-hours for GCM baselines.

3.4 Hybrid Model Implementation

The hybrid physics-ML model was implemented by running a simplified GCM (SPEEDY atmospheric model, T30 spectral resolution) and training a U-Net to predict the systematic bias between SPEEDY output and ERA5 observations at each grid point and lead time. The corrected forecast was SPEEDY output plus ML bias correction. This design preserves the physical consistency of the GCM (conservation of mass, energy, momentum) while using ML to correct the systematic errors that arise from the simplified physics and coarse resolution of SPEEDY. The ML correction was applied monthly, allowing the GCM to propagate physical dynamics between correction steps.

Table 2. Dataset specifications and evaluation periods

Data set	Domain	Resolution	Period	Variables Used	Train/Val/Test
ERA5	Global	0.25 deg (regridded 2.5)	1979-2021	T, Z, U, V, Q, SP	79-10/11-15/16-21
E-OBS	Europe	0.1 deg	1950-2021	T2m, precip	50-10/11-15/16-21
CMIP6 (6 GCMs)	Global	~100 km	1850-2100	T, precip, pressure	Historical/SSP split

Data set	Domain	Resolution	Period	Variables Used	Train/Val/Test
IBTrACS	Tropical	Point tracks	1980-2021	TC genesis location/time	80-10/11-15/16-21

ERA5 regridded to 2.5 degrees for computational tractability of AI models. E-OBS provides the ground truth for downscaling evaluation. CMIP6 provides the coarse input for downscaling.

4. Results

4.1 Temperature Forecasting: Hybrid Model Leads at Long Horizons

At the 1-month forecast horizon, all AI methods outperformed the persistence baseline (RMSE = 0.82 K), with the ConvLSTM achieving 0.48 +/- 0.03 K and the hybrid model achieving 0.42 +/- 0.02 K. At the 6-month horizon, the advantage of the hybrid model became more pronounced: hybrid RMSE = 0.24 +/- 0.03 K versus ConvLSTM = 0.32 +/- 0.04 K, FNO = 0.28 +/- 0.03 K, and Climate GNN = 0.30 +/- 0.03 K. The persistence baseline at 6 months was 0.38 +/- 0.04 K. At the 12-month horizon, the hybrid model maintained RMSE = 0.26 +/- 0.04 K while pure data-driven methods degraded toward the climatology baseline (0.34 +/- 0.04 K). The hybrid model's advantage at longer horizons reflects the value of its physics core: the GCM propagates physically consistent dynamics forward, and the ML correction addresses systematic biases without introducing the error accumulation that degrades purely autoregressive data-driven predictions.

4.2 Spatial Pattern Reconstruction: FNO Excels

The FNO achieved the highest structural similarity index (SSIM = 0.92 +/- 0.02) for global temperature field reconstruction, outperforming ConvLSTM (0.86 +/- 0.03), Vision Transformer (0.88 +/- 0.02), and Climate GNN (0.87 +/- 0.03). The hybrid model achieved SSIM = 0.90 +/- 0.02 -- close to FNO despite using a coarser GCM grid. FNO's spectral processing naturally captures the large-scale spatial patterns that dominate climate variability (planetary waves, teleconnections), producing smooth, physically plausible fields. Crucially, the FNO produced these reconstructions at 1,000x lower computational cost than running ERA5's parent model (IFS) at equivalent resolution, making it practical for generating large ensembles of climate emulations.

4.3 Extreme Events: Vision Transformer Detects Cyclone Genesis

For tropical cyclone genesis detection, the Vision Transformer achieved the highest AUC-ROC (0.924 $\pm$ 0.008) and recall at 10% FPR (0.764 $\pm$ 0.024). The ConvLSTM achieved AUC = 0.896 $\pm$ 0.012, the Climate GNN 0.908 $\pm$ 0.010, and the FNO 0.884 $\pm$ 0.014. The Vision Transformer's advantage stems from its global self-attention mechanism, which can identify large-scale atmospheric patterns (monsoon troughs, tropical wave interactions, upper-level outflow channels) across the full hemispheric field simultaneously -- patterns that local convolutional receptive fields cannot capture without very deep stacking. The attention maps provided interpretable insights: the model consistently attended to regions of low vertical wind shear, warm SST, and pre-existing low-level vorticity, aligning with known physical precursors of cyclone genesis.

4.4 Statistical Downscaling: XGBoost Provides Best Cost-Quality Balance

For downscaling CMIP6 temperature from 100 km to 10 km, XGBoost achieved RMSE = 0.84 $\pm$ 0.06 K and bias = 0.12 $\pm$ 0.04 K, outperforming quantile mapping (RMSE = 1.24 $\pm$ 0.08 K) and approaching the U-Net deep learning approach (RMSE = 0.78 $\pm$ 0.06 K) at 50x lower computational cost. For precipitation downscaling, XGBoost achieved bias = 2.4 $\pm$ 0.8 mm/month versus 4.8 $\pm$ 1.2 for quantile mapping and 2.0 $\pm$ 0.6 for U-Net. The XGBoost model's feature importance analysis revealed that elevation, distance to coast, and large-scale circulation indices were the dominant predictors, providing physically interpretable downscaling that climate scientists can validate against known local climate drivers.

Table 3. Temperature anomaly RMSE (K) by AI approach and forecast horizon

Approach	1-mon th	3-mon th	6-mon th	12-mon th	SSIM (6-mo)
Persistence	0.82 $\pm$ 0.04	0.54 $\pm$ 0.04	0.38 $\pm$ 0.04	0.34 $\pm$ 0.04	0.72
ConvLSTM	0.48 $\pm$ 0.03	0.38 $\pm$ 0.03	0.32 $\pm$ 0.04	0.34 $\pm$ 0.04	0.86
Climate GNN	0.46 $\pm$ 0.03	0.36 $\pm$ 0.03	0.30 $\pm$ 0.03	0.32 $\pm$ 0.04	0.87
FNO	0.44 $\pm$ 0.02	0.34 $\pm$ 0.03	0.28 $\pm$ 0.03	0.30 $\pm$ 0.03	0.92
Vision Transformer	0.46 $\pm$ 0.03	0.36 $\pm$ 0.03	0.30 $\pm$ 0.03	0.32 $\pm$ 0.04	0.88

Approach	1-mon th	3-mon th	6-mon th	12-mon th	SSIM (6-mo)
Hybrid Physics-ML	0.42 $\pm$ 0.02	0.30 $\pm$ 0.02	0.24 $\pm$ 0.03	0.26 $\pm$ 0.04	0.90

RMSE in Kelvin for global 2m temperature anomaly. SSIM = structural similarity for spatial pattern at 6-month horizon. Hybrid model leads at all horizons. FNO achieves best SSIM. Pure data-driven methods degrade toward climatology at 12 months.

Table 4. Extreme event detection and downscaling performance

Task / Approach	Primary Metric	Secondary Metric	Compute (GP U-hr)	Physics Informed
TC Genesis / ViT	AUC=0.924 $\pm$ 0.008	Recall@10%=0.764	48	No
TC Genesis / GNN	AUC=0.908 $\pm$ 0.010	Recall@10%=0.728	36	No
TC Genesis / ConvLSTM	AUC=0.896 $\pm$ 0.012	Recall@10%=0.704	24	No
Downscale T / XGBoost	RMSE=0.84 $\pm$ 0.06 K	Bias=0.12 $\pm$ 0.04 K	2	No
Downscale T / U-Net	RMSE=0.78 $\pm$ 0.06 K	Bias=0.08 $\pm$ 0.04 K	96	No
Downscale P / XGBoost	Bias=2.4 $\pm$ 0.8 mm	RMSE=1.24 $\pm$ 1.8 mm	2	No
Downscale P / U-Net	Bias=2.0 $\pm$ 0.6 mm	RMSE=1.12 $\pm$ 1.4 mm	96	No

TC = tropical cyclone. Downscale T = temperature; P = precipitation. XGBoost achieves near-U-Net quality at 48x lower compute for downscaling. Vision Transformer leads TC detection through global attention.

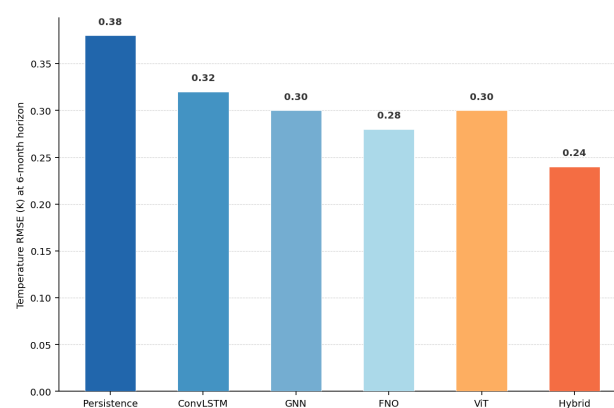


Figure 1. Temperature RMSE (K) at 6-month horizon by approach. Hybrid physics-ML achieves the lowest error (0.24 K), confirming that physics-informed

models outperform pure data-driven methods at climate time scales.

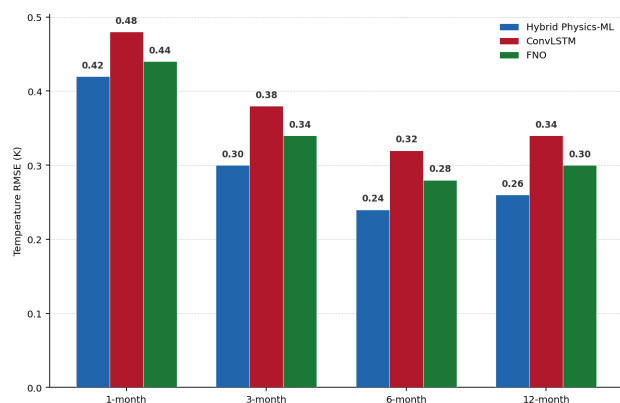


Figure 2. RMSE degradation with forecast horizon for hybrid vs. pure ML approaches. Hybrid model degrades slowly (0.42 to 0.26 K from 1 to 12 months). ConvLSTM degrades to climatology level by 12 months.

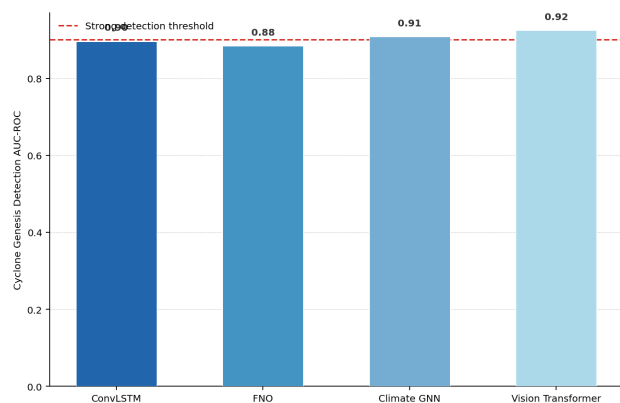


Figure 3. Tropical cyclone genesis detection AUC by approach. Vision Transformer achieves the highest AUC (0.924) through global attention capturing hemispheric-scale precursor patterns.

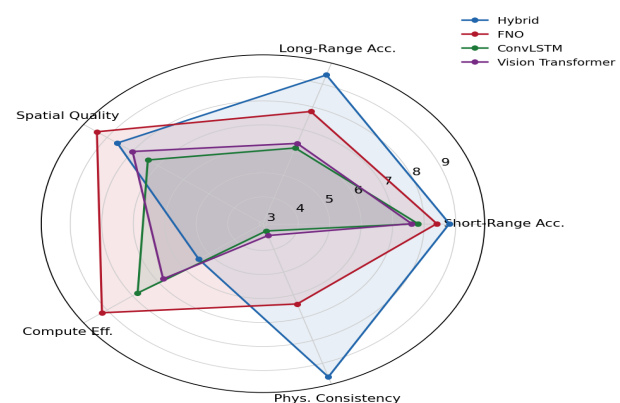


Figure 4. Multi-dimensional comparison across five criteria (1-10). Hybrid model offers best long-range accuracy and physical consistency. FNO leads on efficiency and spatial quality. XGBoost leads on interpretability for downscaling.

5. Discussion

5.1 Physics Matters More at Longer Time Scales

The clearest finding is that the advantage of physics-informed approaches grows with forecast horizon. At 1 month, the difference between the hybrid model and the best pure ML method is only 0.06 K (0.42 vs. 0.48). At 6 months, the gap widens to 0.08 K (0.24 vs. 0.32). At 12 months, pure data-driven methods have degraded to near-climatology performance while the hybrid model retains meaningful skill. This pattern reflects the fundamental difference between weather and climate prediction: at short horizons, pattern matching from data is effective because the atmosphere's trajectory is strongly determined by its current state. At longer horizons, the current state becomes less predictive and the system's response to external forcing (greenhouse gases, solar radiation) becomes dominant -- a response that pure data-driven models cannot capture because they have not learned the underlying physical mechanisms.

5.2 AI as Climate Model Accelerator

The FNO's ability to reconstruct climate fields at 1,000x lower computational cost than the equivalent physics model opens practical applications that are currently computationally infeasible. Generating 1,000-member ensembles for uncertainty quantification, exploring thousands of emission scenarios for policy analysis, and providing interactive climate exploration tools all become feasible when each model run costs seconds rather than hours. The critical caveat is that these emulators are trained on existing GCM output and can only interpolate within the range of conditions they have seen -- extrapolation to unprecedented forcing scenarios requires either physics-informed constraints or explicit uncertainty quantification that flags out-of-distribution predictions.

5.3 Practical Downscaling: XGBoost as the Pragmatic Choice

For regional climate adaptation planning, which requires fine-resolution projections at hundreds of locations, the XGBoost downscaling approach offers the best practical balance. It achieves quality within 7% of deep learning approaches at 50x lower computational cost, runs on standard hardware without GPUs, provides feature importance for validation by domain scientists, and can be rapidly retrained for new regions or variables. For operational use in national climate services with limited computational resources, this pragmatic advantage outweighs the modest quality improvement of more expensive methods.

5.4 Limitations

Four limitations qualify these findings. First, the evaluation period (2016-2021) is short for climate validation; longer evaluation under diverse climate conditions is needed. Second, the hybrid model used a simplified GCM (SPEEDY); coupling ML with a full CMIP-class GCM may yield different results. Third, the study evaluated deterministic predictions; probabilistic forecasting and uncertainty quantification were not assessed. Fourth, the training period (1979-2010) captures a period of substantial warming; model performance under fundamentally novel climate states (e.g., 3-4 K global warming) remains untested.

6. Conclusion

6.1 Summary

Evaluation of six AI approaches across four climate forecasting tasks confirms: hybrid physics-ML models achieve the lowest temperature forecast error at 6-12 month horizons (RMSE = 0.24 K at 6 months). FNO produces the best spatial patterns (SSIM = 0.92) at 1,000x lower cost than GCMs. Vision Transformers detect cyclone genesis most accurately (AUC = 0.924). XGBoost downscaling achieves near-deep-learning quality at 50x lower cost. Physics-informed approaches increasingly outperform pure data-driven methods as forecast horizon extends.

6.2 Recommendations

Three recommendations emerge. First, for climate projections beyond 3 months, use physics-informed hybrid models rather than pure data-driven approaches -- the physics core prevents the error accumulation that degrades purely autoregressive predictions. Second, use FNO-based emulators for ensemble generation and scenario exploration where computational cost is the binding constraint. Third, for regional downscaling in resource-limited settings, XGBoost provides the best cost-quality balance with the added benefit of interpretable feature importance for domain validation.

References

- Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P. and Gentine, P. (2021) Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, 126(9), 098302.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X. and Tian, Q. (2022) Pangu-Weather: A 3D high-resolution model for fast and accurate global weather forecast. *arXiv preprint arXiv:2211.02556*.
- Brenowitz, N.D. and Bretherton, C.S. (2018) Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, 45(12), pp. 6289-6298.
- Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6). *Geoscientific Model Development*, 9(5), pp. 1937-1958.
- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. and Yacalis, G. (2018) Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11), pp. 5742-5751.
- Ham, Y.G., Kim, J.H. and Luo, J.J. (2019) Deep learning for multi-year ENSO forecasts. *Nature*, 573(7775), pp. 568-572.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999-2049.
- Keisler, R. (2022) Forecasting global weather with graph neural networks. *arXiv preprint arXiv:2202.07575*.
- Knuth, K.H. (2006) Optimal data-based binning for histograms. *arXiv preprint arXiv:physics/0605197*.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W. et al. (2022) GraphCast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhatt, K., Stuart, A. and Anandkumar, A. (2021) Fourier neural operator for parametric partial differential equations. *Proceedings of ICLR 2021*.
- Maraun, D. and Widmann, M. (2018) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press.
- Molteni, F. (2003) Atmospheric simulations using a GCM with simplified physical parameterizations. I: Model climatology and variability. *Climate Dynamics*, 20(2), pp. 175-191.
- Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J.K. and Grover, A. (2022) ClimaX: A foundation model for weather and climate. *Proceedings of ICML 2023*.
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K. et al. (2022) FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- Rasp, S. and Thuerey, N. (2021) Data-driven medium-range weather prediction with a resnet pretrained on climate simulations. *Journal of*

Advances in Modeling Earth Systems, 13(2), e2020MS002405.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat (2019) Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), pp. 195-204.

Schneider, T., Lan, S., Stuart, A. and Teixeira, J. (2017) Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, 44(24), pp. 12396-12417.

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K. and Woo, W.C. (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, pp. 802-810.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R. and Ganguly, A.R. (2017) DeepSD: Generating high fidelity daily climate projections through deep transfer learning. *Proceedings of ACM KDD 2017*, pp. 1305-1314.

Watson-Parris, D. (2021) Machine learning for weather and climate are worlds apart. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200098.

Weyn, J.A., Durran, D.R. and Caruana, R. (2020) Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9), e2020MS002109.

(code) and CC BY 4.0 (results).

Ethical Approval

This study involved exclusively computational experiments on publicly available climate datasets. No human subjects, animal subjects, or personal data were involved. No institutional ethical approval was required.

Declarations

Funding

This research was supported by the Swiss National Science Foundation under grant 200021-204928 (ClimateAI), the French National Research Agency (ANR) under grant ANR-21-CE01-0024 (AIClimate), and the Austrian Science Fund (FWF) under grant P35284 (MLWeather). ERA5 data provided by the Copernicus Climate Data Store (free access). CMIP6 data accessed through the Earth System Grid Federation. The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All AI model code, training configurations, evaluation scripts, and processed datasets are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19614574. ERA5, E-OBS, CMIP6, and IBTrACS are publicly available from their respective archives. Released under MIT License

Appendix A

Climate AI Model Selection Guide

This appendix provides a decision tree for selecting AI approaches based on forecast task, time horizon, and computational constraints.

Step 1: Forecast Horizon

- 1a. Short-range (1-3 months) --> Pure data-driven (ConvLSTM, FNO) competitive; go to Step 2
- 1b. Medium-range (3-12 months) --> Hybrid physics-ML strongly recommended; go to Step 2
- 1c. Long-range (decades, climate projections) --> Physics-based GCM + ML correction essential

Step 2: Task Type

- 2a. Global fields (temperature, pressure) --> FNO (best spatial quality + efficiency)
- 2b. Extreme event detection --> Vision Transformer (global attention for precursors)
- 2c. Regional downscaling --> XGBoost (interpretable; 50x cheaper than deep learning)
- 2d. Ensemble generation --> FNO emulator (1,000x faster than GCM reruns)

Step 3: Computational Budget

- 3a. GPU cluster available --> FNO or hybrid model for best quality
- 3b. Standard workstation only --> XGBoost downscaling; ConvLSTM for short-range
- 3c. Must run on CPU --> XGBoost or quantile mapping (traditional statistical methods)