Exploring Design Choices For Auto-regressive Deep Learning Climate Models

Florian Gallusser Simon Hentschel Anna Krause Andreas Hotho

Summary

We explore key design choices for autoregressive Deep Learning (DL) models to achieve stable 10-year rollouts that preserve the statistics of the reference climate. We quantitatively compare the long-term stability of three prominent model architectures — FourCastNet, SFNO, and ClimaX — trained on ERA5 reanalysis data at 5.625° resolution and systematically assess the impact of autoregressive training steps, model capacity and choice of prognostic variables.

Motivation

Earth system simulations on climate timescales require DL-based approaches to take thousands of autoregressive steps. DL models for medium-range weather prediction (MWP) often fail to maintain physically consistent trajectories far beyond 14 days [1, 2]. While a few atmospheric DL approaches with different model architectures and training strategies have demonstrated stability over decades [3, 4, 5], the key design choices enabling this remain unexplored.

Experiment

We train DL models that learn to predict the state of the atmosphere X_t from a previous state at time X_{t-6h}. We perform a grid search over the following hyperparameters and train each model configuration with 10 random seeds for 20 epochs:

- Models:
  - ClimaX [6] based on Vision Transformer (ViT)
  - FourCastNet (FCN) [7] combining the Fourier Neural Operator with a ViT backbone
  - SFNO [8] which generalizes Fourier Neural Operator approach to the sphere
- Number of autoregressive training steps: 1, 2, 4
- Number of prognostic variables: 8, 33
- Number of layers: 4, 6, 8
- Hidden dimension: 128, 256, 512

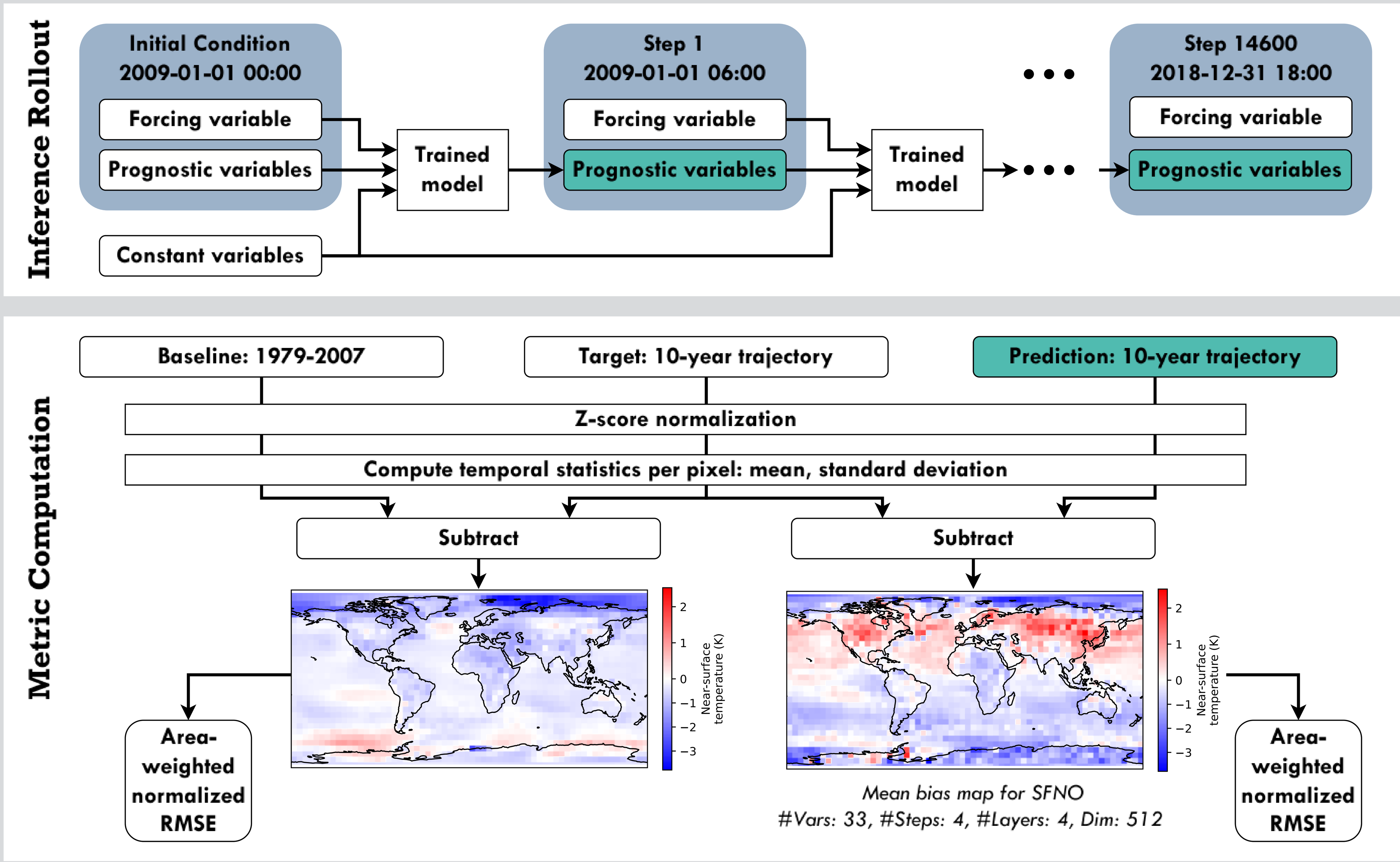
Dataset

We use the WeatherBench [8] dataset which was derived from ERA5 by bilinear interpolation:

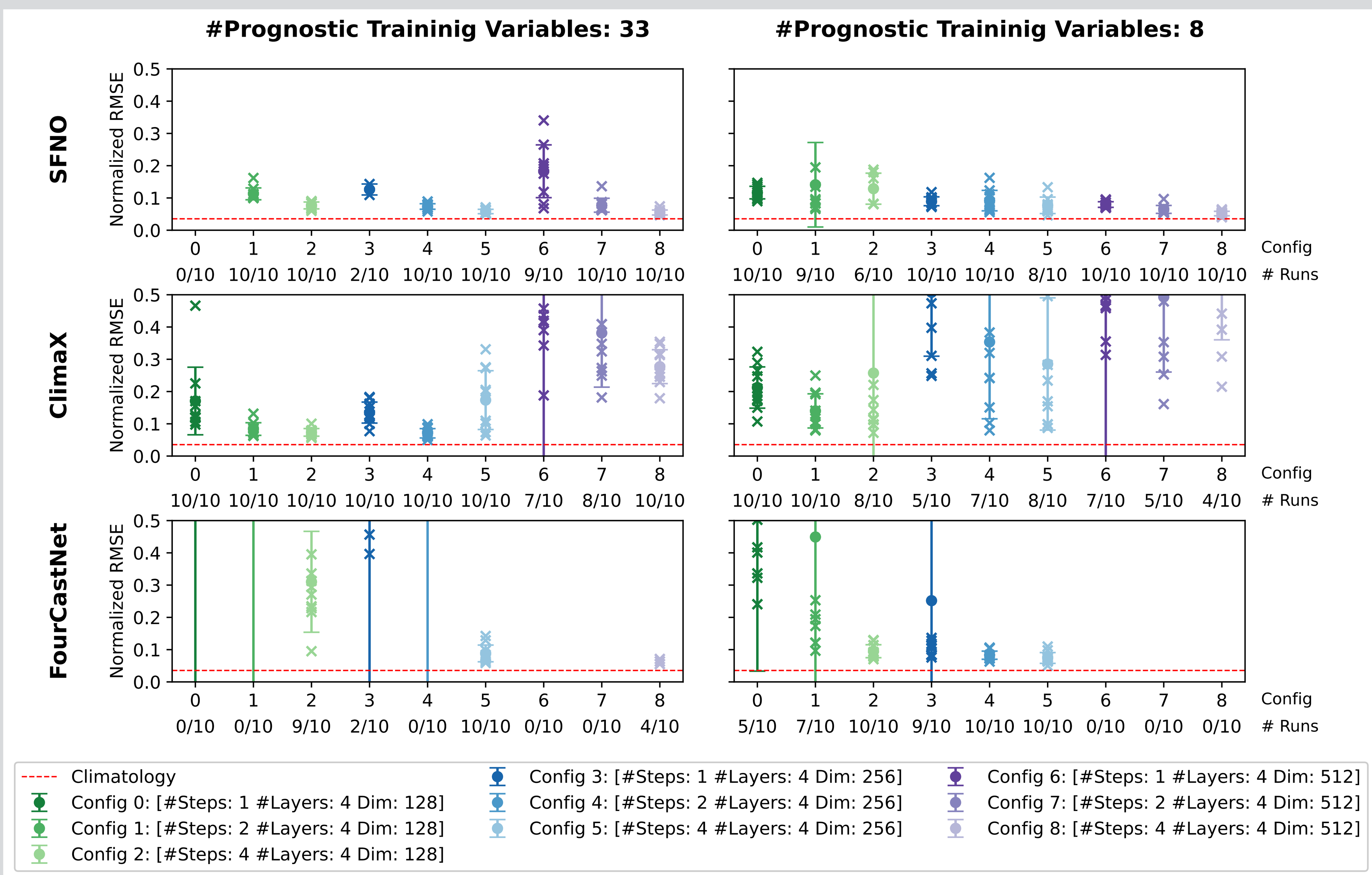
- Horizontal resolution: 5.625° (64x32 grid points)
- Temporal resolution: 6 hours
- Training years: 1979 to 2007
- Grid: Regular longitude-latitude grid
- Forcing variable: Top-of-the-atmosphere incident solar radiation
- Constants: Land-sea mask, orography, lon, lat
- Two sets of prognostic variables
  - 33 variables: Near-surface temperature (tas) and winds (uas/vas) as well as temperature (ta), geopotential height (zg), horizontal winds (va, ua) and specific humidity (q) at 925, 850, 700, 600, 500, 250 hPa
  - 8 variables: Near-surface temperature and winds as well as temperature at 850 hPa, geopotential height at 1000, 700, 500, 300 hPa

Evaluation

We expect our models to approx. reproduce the climate of the 10 years following the training period since we assume the effects of climate change within this short period to be neglectable.



Results



Area-weighted normalized RMSE scores averaged over the variables tas, uas, vas, ta850 and zg500 for model configurations 0 — 8 with layer size L = 4. The mean (dots) and standard deviation (error bars) were computed across 10 seeds (crosses) with finite RMSE. For better readability, the y-axis is cut-off at 0.5 and the number of displayed runs out of 10 is shown on the x-axis.

Conclusions

- ➢ With the right choice of hyperparameters all model architectures can yield 10-year rollouts that match important distributional characteristics of the target climate
- ➢ Stability of 10-year rollouts with SFNO is much less sensitive to the choice of hyperparameters
- ➢ Multi-step training enhances the stability of long rollouts
- ➢ For the non-geometry aware model architectures – ClimaX and FCN – increasing the hidden dimension leads to instabilities of long rollouts
- ➢ For all evaluated model architectures, some random seeds lead to instabilities of long rollouts

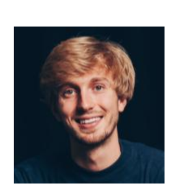
References

[1] Karlbauer et al. Advancing Parsimonious Deep Learning Weather Prediction Using the HEALPix Mesh. *Journal of Advances in Modeling Earth Systems*, 2024.
[2] McCabe et al. Towards Stability of Autoregressive Neural Operators. *Transactions on Machine Learning Research*, 2023.
[3] Watt-Meyer et al. ACE: A fast, skillful learned global atmospheric model for climate prediction. *arXiv:2310.02074*, 2023.
[4] Cresswell-Clay et al. A Deep Learning Earth System Model for Stable and Efficient Simulation of the Current Climate. *arXiv:2409.16247*, 2024.
[5] Karlbauer et al. Comparing and Contrasting Deep Learning Weather Prediction Backbones on Navier-Stokes and Atmospheric Dynamics. *arXiv:2407.14129*, 2024.
[6] Nguyen et al. . ClimaX: A foundation model for weather and climate. *In Proceedings of the 40th International Conference on Machine Learning*, 2023.
[7] Pathak et al. FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. *arXiv:2202.11214*, 2022.
[8] Bonev et al. . Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere. *In Proceedings of the 40th International Conference on Machine Learning*, 2023.
[9] Rasp et al. WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting. *Journal of Advances in Modeling Earth Systems*, 2020.

Funded by the European Union

Computational resources were provided by HPC@FAU under project ID b214cb.

Contact



Code

Chair of Data Science (Informatik X) | Am Hubland, 97074 Würzburg, Germany | https://informatik.uni-wuerzburg.de/datascience/