Department Scientific Computing Center
Jr. Research Group Robust and Efficient AI

Studying Forecast Accuracy in Pangu-Weather through Ablation and Feature Enhancement

Deifilia To, Gholamali Hoshyaripour, Julian Quinting, Achim Streit, Charlotte Debus

Introduction

- Pangu-Weather (PGW) was the first data-driven model to perform better than Numerical Weather Prediction (NWP) [1]

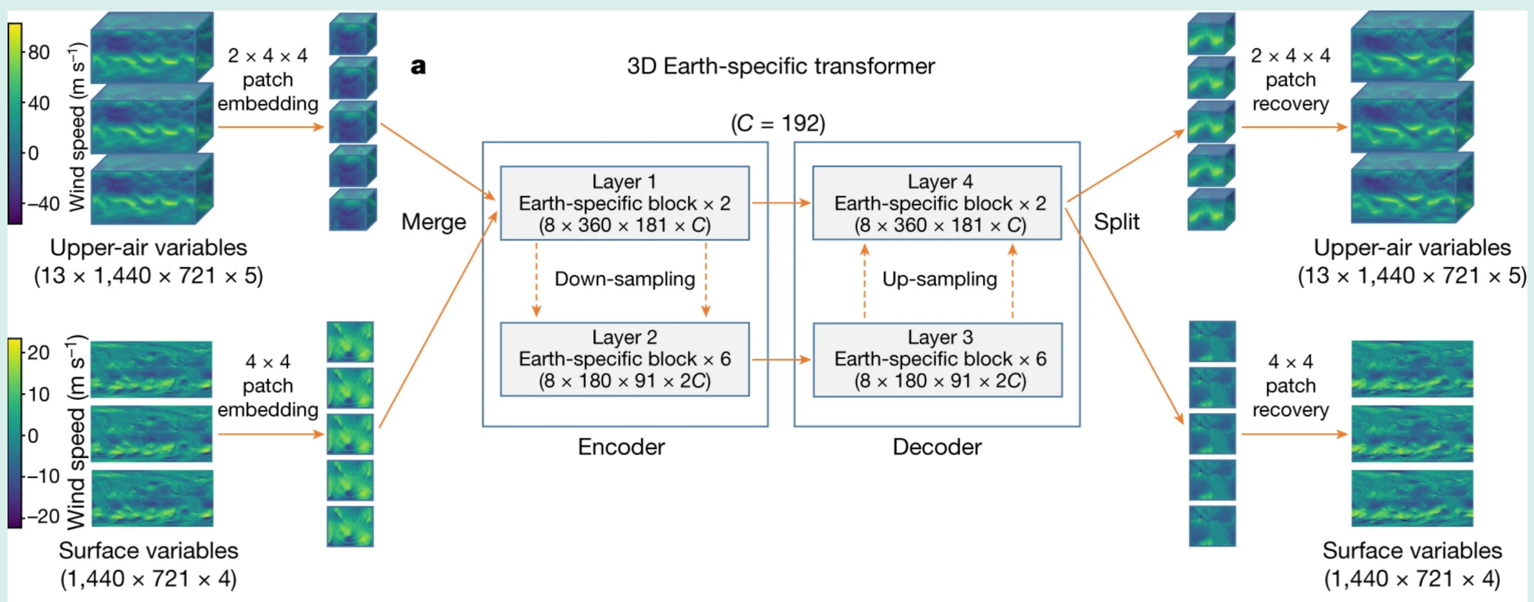


Figure: PGW architecture has 64 million parameters per lead time and features a 3D Transformer, earth-specific position bias, and one down- and up-sampling layer. Adapted from [1].

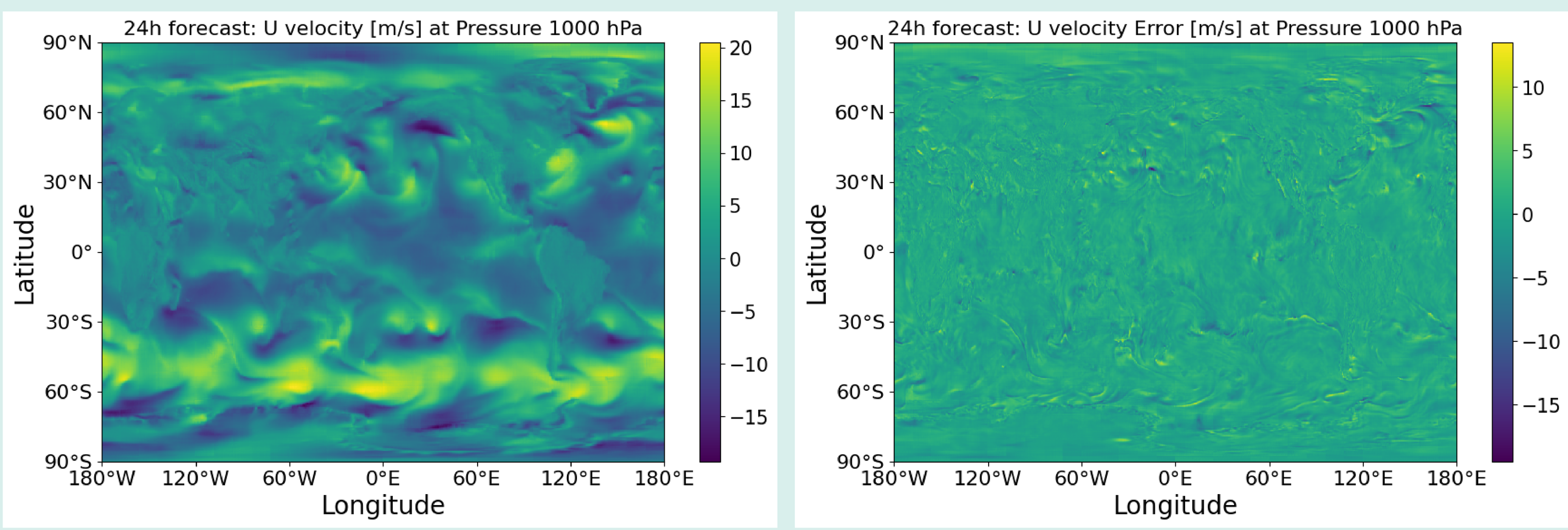
- **What components of PGW allow it to be so successful?**

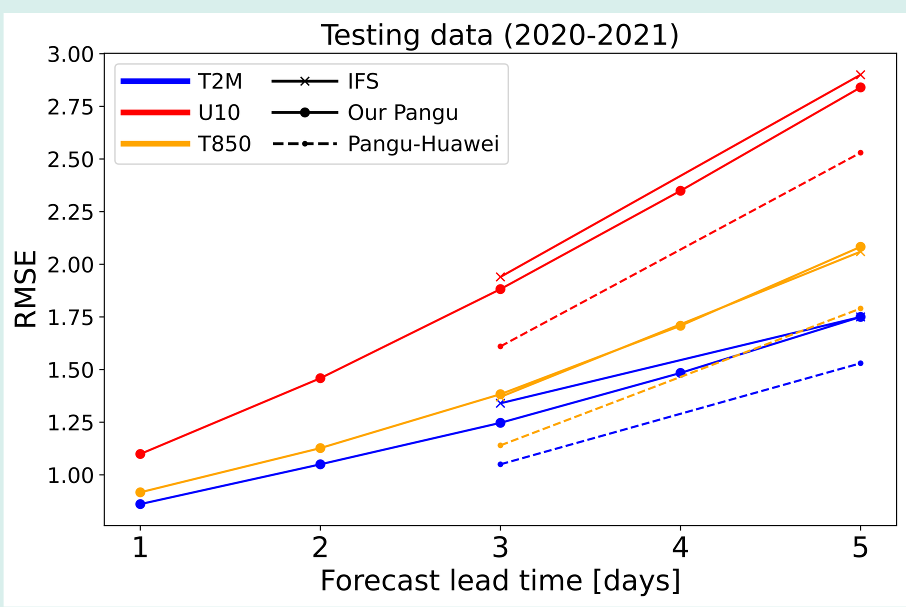
Method

- Replicate PGW in a modular manner that allows for substitution of different architectural components
- Validate PGW's 24 lead time model by verifying that the RMSE of key prognosis variables (T850, Z500, T2M, U10, V10) are better than NWP
- We observe that more data leads to better results—we are data limited
  - Train 24-hour Lite models on restricted data—1979–2018, 6-h data (15% of total dataset)

Validation

- Full model performs better than IFS for most variables
  - Results of PGW-Lite could not be validated





Models

- Trained with Adam optimizer, initial learning rate = $5 \cdot 10^{-4}$, scheduler: ReduceLROnPlateau

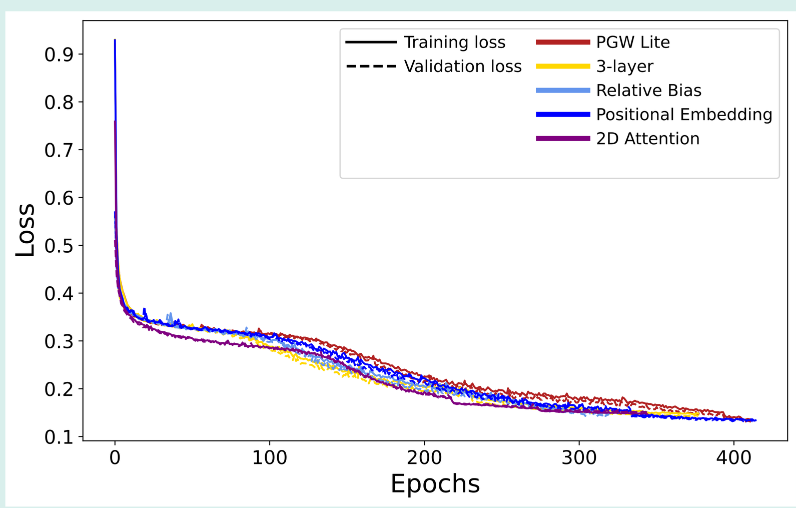
PGW Lite trained with...

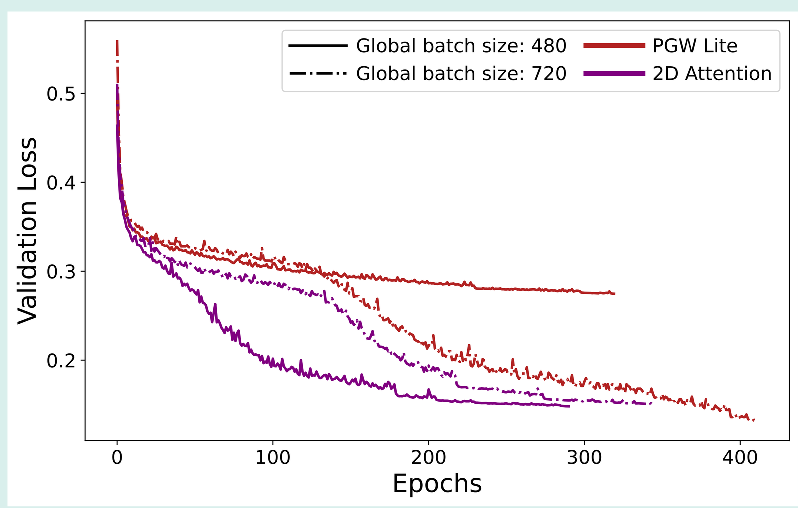
	Model	Number of parameters (millions)
1.	Absolute bias (orig.)	44.6
2.	Relative bias	24.3
3.	Positional embedding	24.3
4.	2D Attention	57.2
5.	Two up/down sampling layers	108.9

- Bias: term added to attention matrix
- Absolute bias: different bias based on latitude and height of patch
- Relative bias: based on relative position within attention matrix
- Positional embedding: learnable position term when patching; not in attention layers

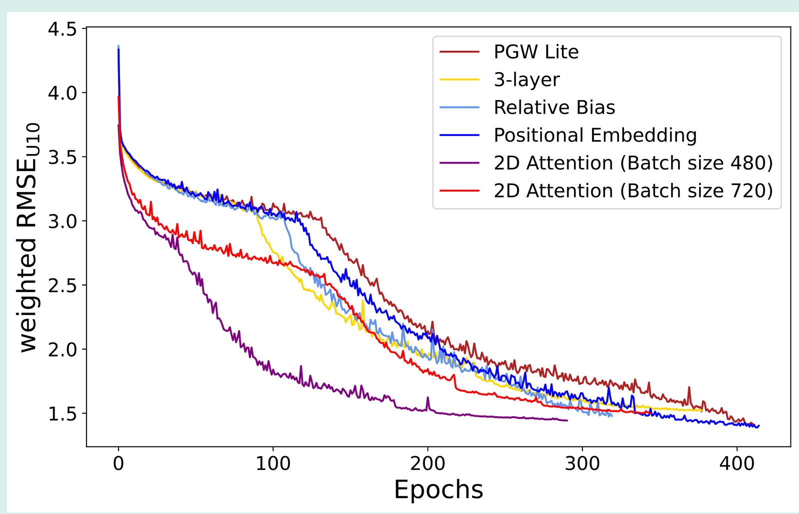
Results

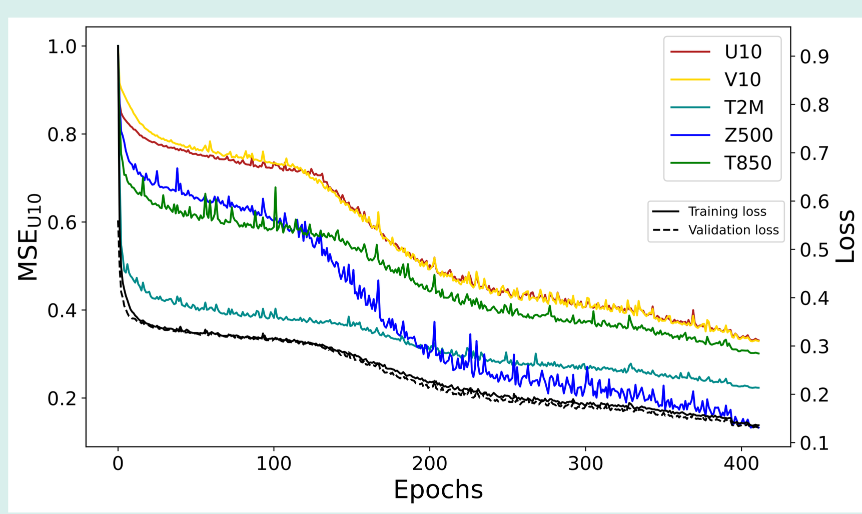
- Bias term can be replaced with patch embedding
- Two-dimensional transformer converges faster and performs competitively
  - 2D Transformer = 2× local batch size = ≈40% reduction in compute required
- Training is highly sensitive to batch size





- Atmospheric variables are minimized at different rates throughout training





References

1. K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian. Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619(7970):533–538, July 2023.

deifilia.to@kit.edu