WeatherMesh-3: Fast and Accurate Operational Global Weather Forecasting

Haoxing Du*, Lyna Kim*, Joan Creus-Costa, Jack Michaels, Anuj Shetty,
Todd Hutchinson, Christopher Riedel, John Dean
{haoxing, lyna}@windbornesystems.com
ICLR 2025: Tackling Climate Change with Machine Learning

Takeaways

1. Lightweight operational transformer-based weather forecast
2. Achieves SOTA results with new methods: latent rollout + matepoint
3. Modular architecture for easy ingestion of additional custom data

1. Speed/Compute

Provide critical forecasts for time-sensitive events faster, more compute-efficiently

>100,000x speedup for generating global 14-day forecast over IFS

2. Model Skill

Improve forecasting ability to on longer lead times with same inits

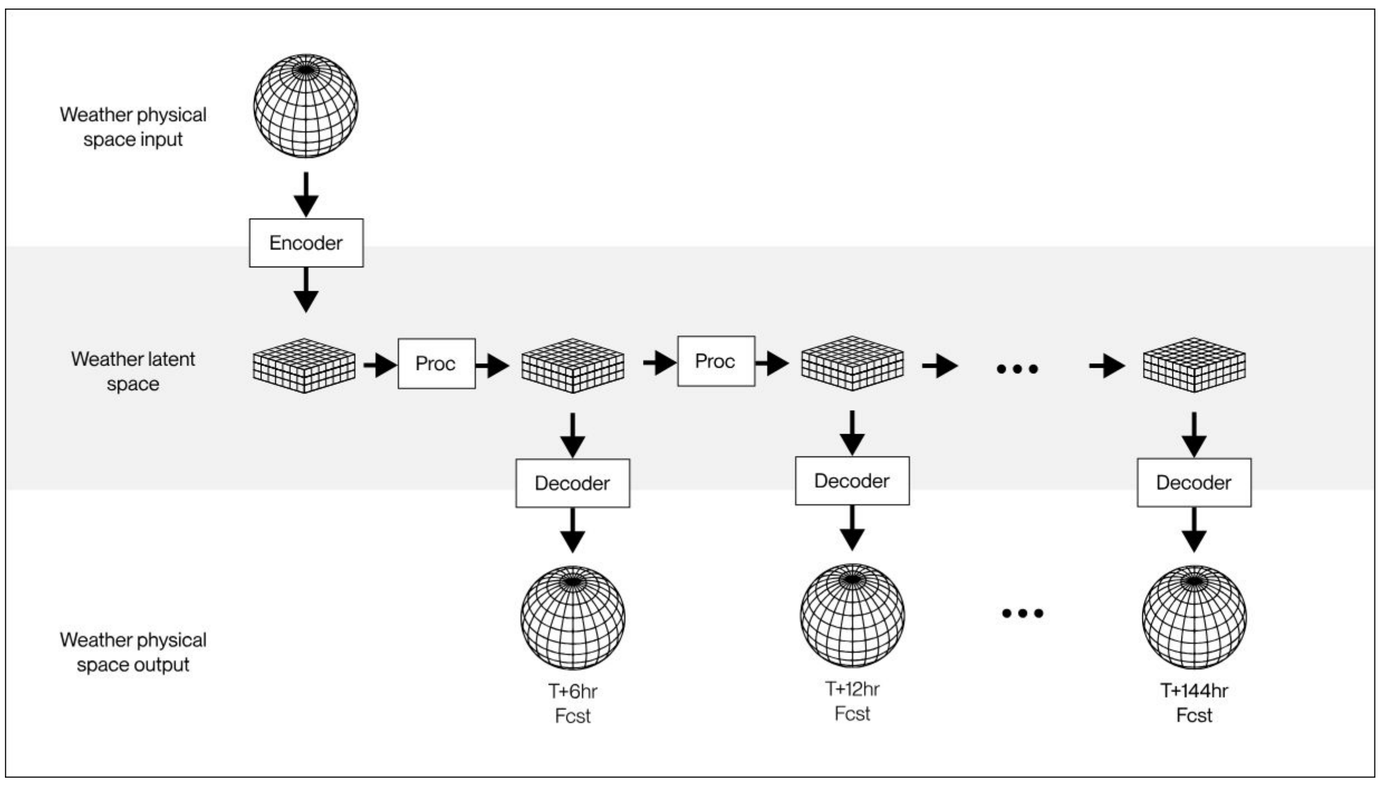
Outperforms IFS HRES by **up to 38%** across all but one of 690 targets

3. Customizability

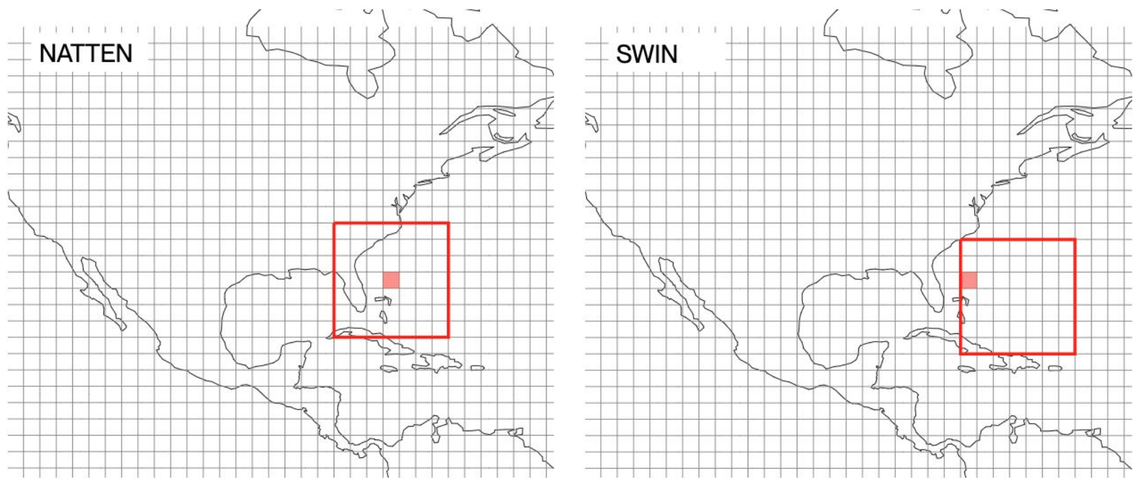
Enable models tailored to maximize performance on local weather events

Generates 0.25°, 14-day forecasts in **12 seconds** on a **single RTX 4090**

Methods

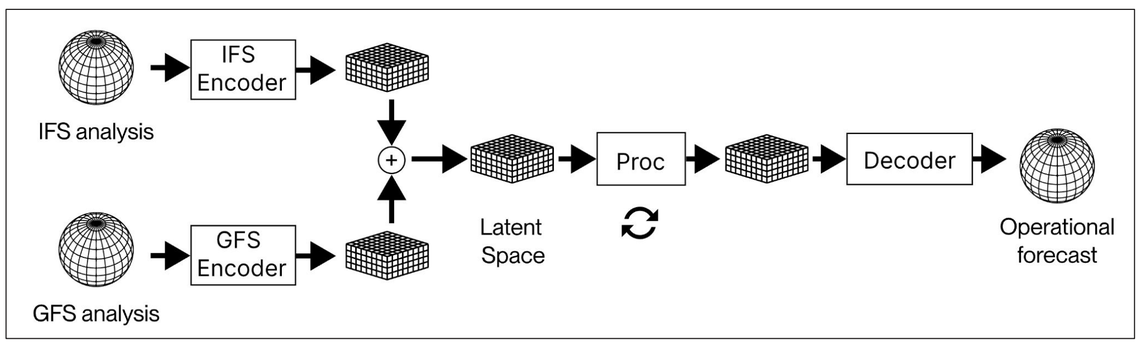


1. **Latent rollout**
   1hr/6hr processor "rolls out" the forecast in latent space to desired lead time, avoiding duplicative encode/decode steps
2. **Neighborhood Attention[1] vs Sliding Window[2]**
   Better inductive bias due to its consistent locality of attention for information transfer between patches

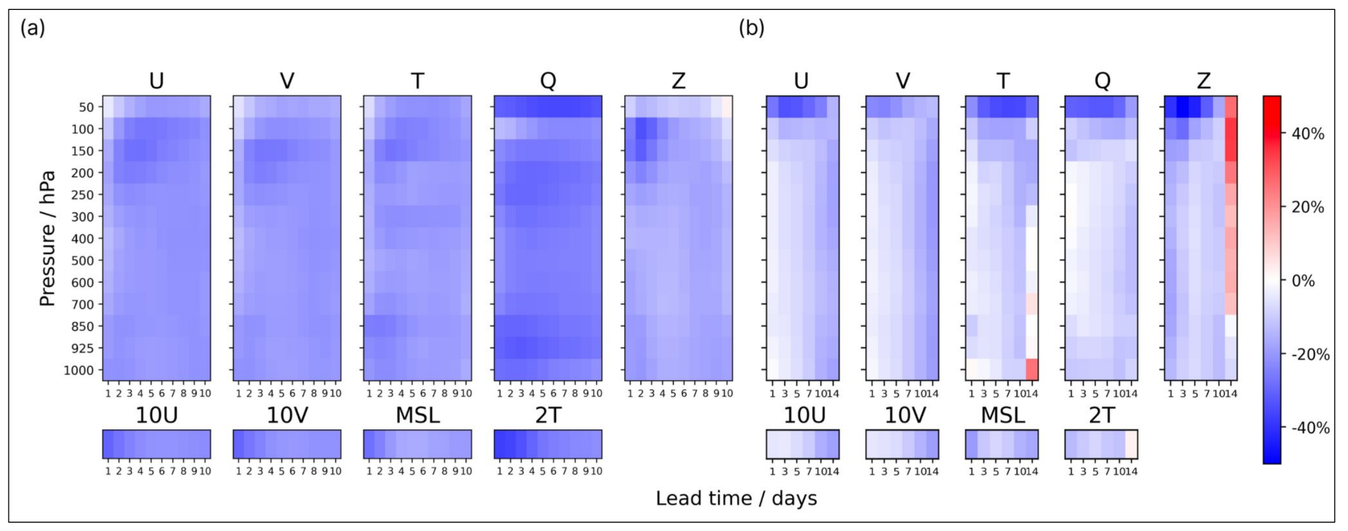


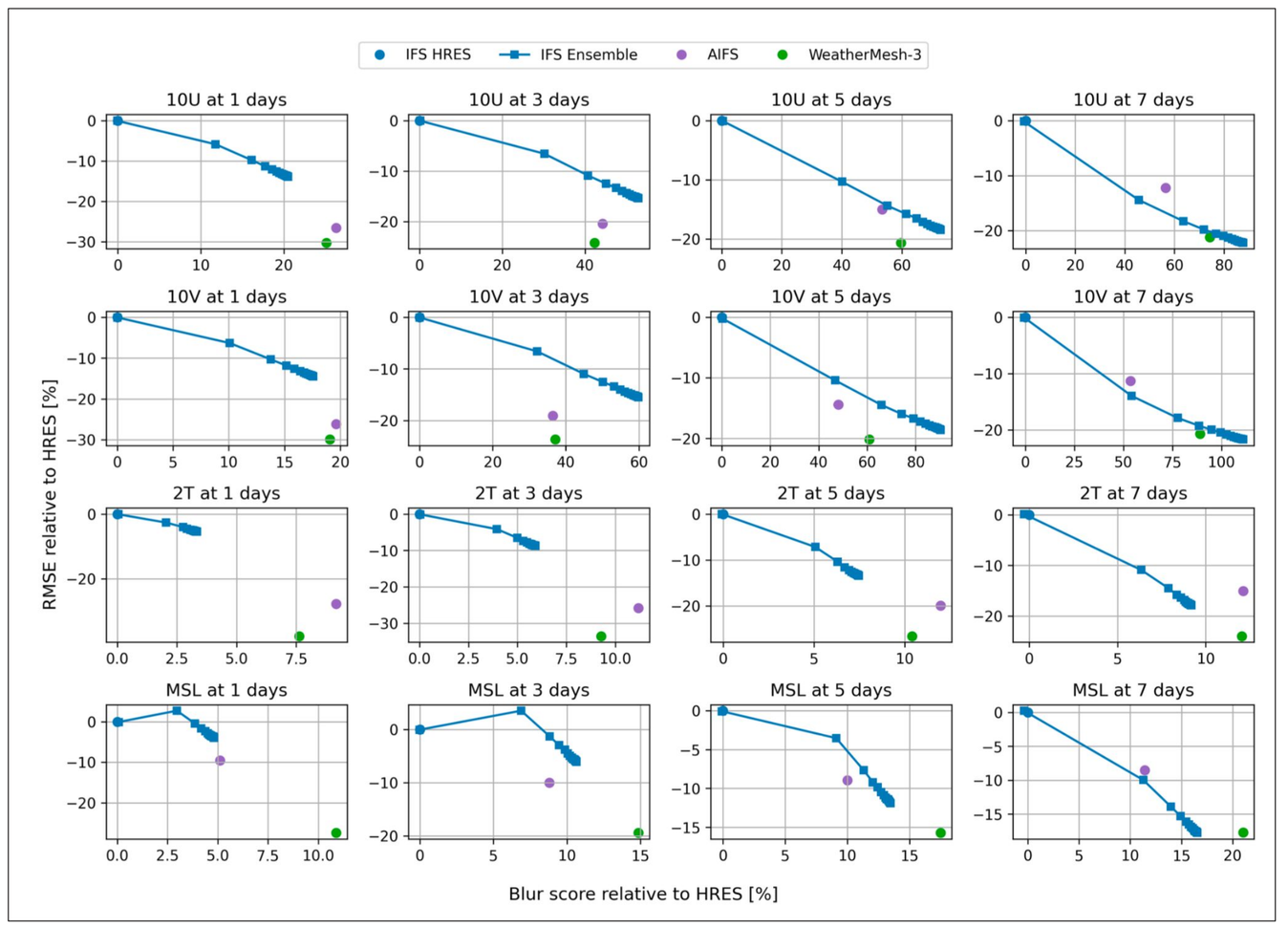
Training

- Pre-trained on ERA-5 from 1979-2022
- Mixed horizon prediction tasks introduced during training of multiple processors (1hr, 6hr)
- In preparation for operational settings, train additional encoders to use blended input of IFS + GFS initializations & tune processors to strengthen real-time performance



Results





Discussion

- **Compute efficiency democratizes access to high-quality weather forecasts.** WM-3 runs on hardware as light as a consumer grade laptop with 16GB VRAM & 32GB RAM, ~⅓ of Aurora's[3] VRAM usage.
- The modular architecture of **WM-3 enables users to ingest additional custom data sources that enhance forecast accuracy.**

Taken together, **organizations with limited computational resources can use WM-3** to run their own customized weather forecasting systems without compromising performance.

Future Work

1. **Large ensembles**
   For better uncertainty quantification on forecast reliability and probabilistic forecasting with representative spread
2. **Expand data assimilation sources**
   Integration of additional real-time data sources for improving quality of initial conditions via live data assimilation pipeline

References

1. Hassani, Ali, et al. 2023, "Neighborhood attention transformer."
2. Liu, Ze, et al. 2021, "Swin transformer: Hierarchical vision transformer using shifted windows."
3. Bodnar, Cristian, et al. 2024, "Aurora: A foundation model of the atmosphere."