

DATAPOINT
133 hPa

W-1519

Amor, Minnesota, USA
46.4099° N, -95.7313° WTEMPERATURE
18° CPRESSURE
133 hPa

LAUNCHED LAST MONTH

DATAPOINT
18° C

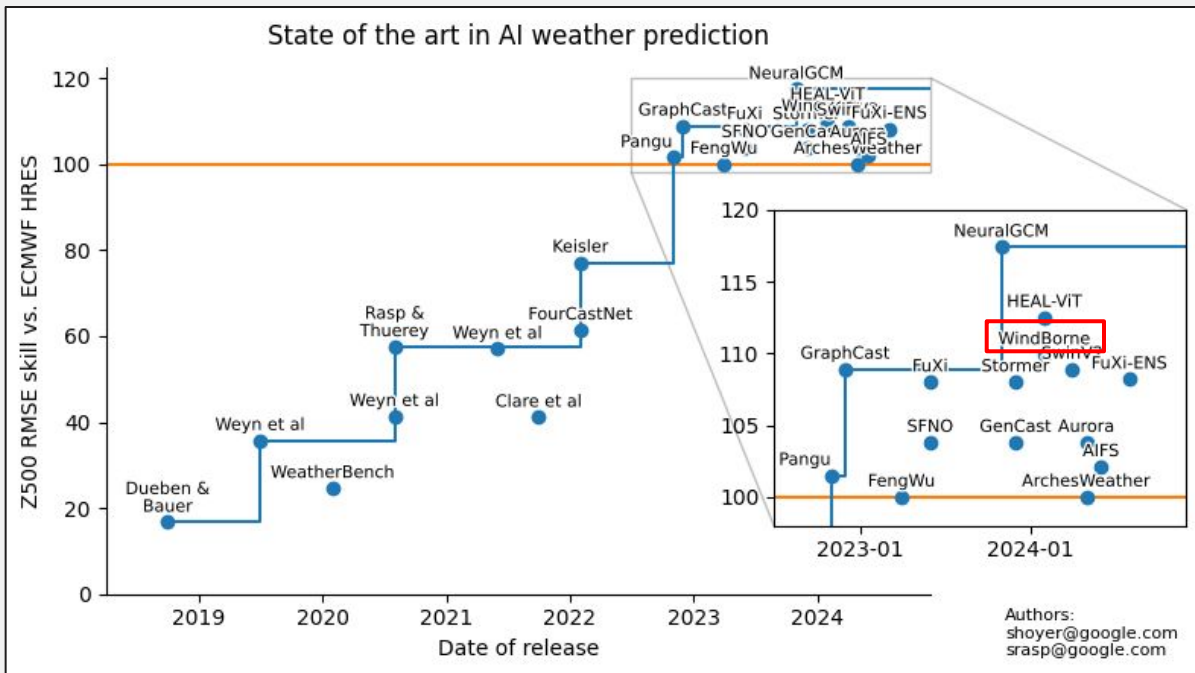
WeatherMesh-3: Fast and Accurate Operational Global Weather Forecasting

Haoxing Du*, Lyna Kim*, Joan Creus-Costa, Jack Michaels, Anuj Shetty, Todd Hutchinson, Christopher Riedel, John Dean

*WindBorne Systems
Palo Alto, CA*



AI for Numerical Weather Prediction is on the rise...

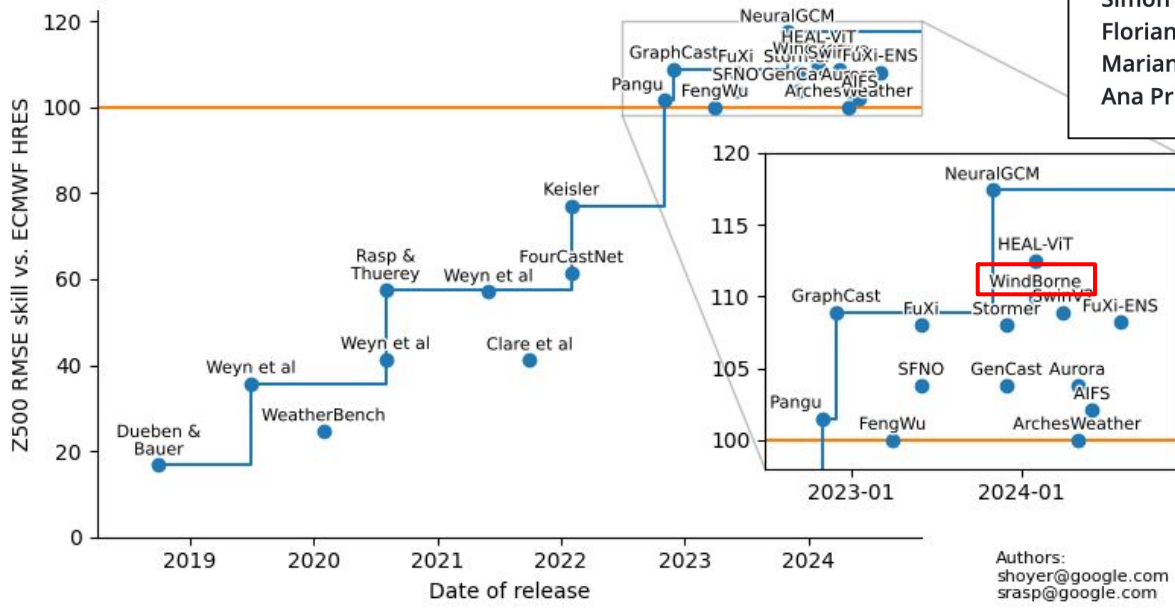


AI for Numerical Weather Prediction is on the rise...

NEWS AIFS: a new ECMWF forecasting system

Simon Lang, Mihai Alexe, Matthew Chantry, Jesper Dramsch, Florian Pinault, Baudouin Raoult, Zied Ben Bouallègue, Mariana Clare, Christian Lessig, Linus Magnusson, Ana Prieto Nemesio

State of the art in AI weather prediction



But future AI4NWP systems must provide differential value over NWP

1. **Speed/Compute**

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

2. **Model Skill**

Improve forecasting ability to on longer lead times with same inits

3. **Customizability**

Enable models tailored to maximize performance on local weather events

WeatherMesh-3 enables the unique contribution of AI4NWP models in an operational setting

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

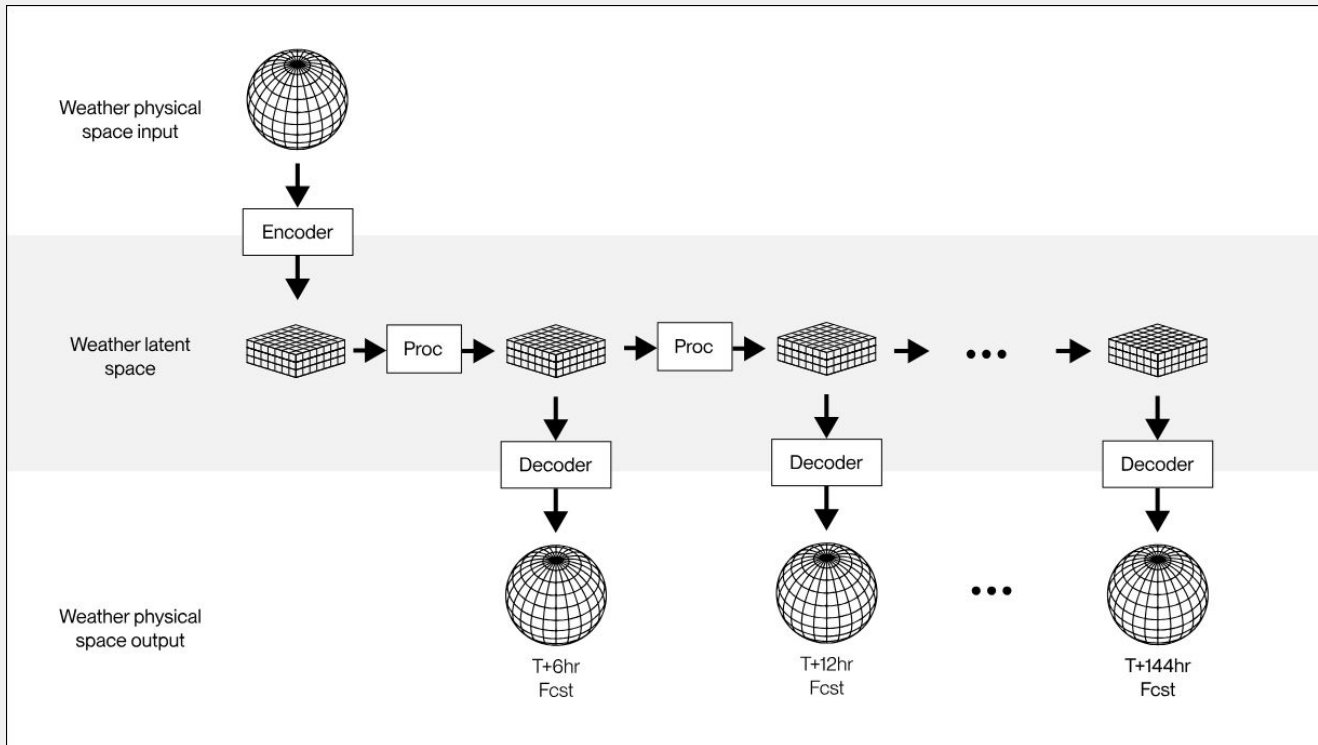
Generates 0.25°, 14-day forecasts in **12 seconds** on a **single RTX 4090**



Model + Training Details

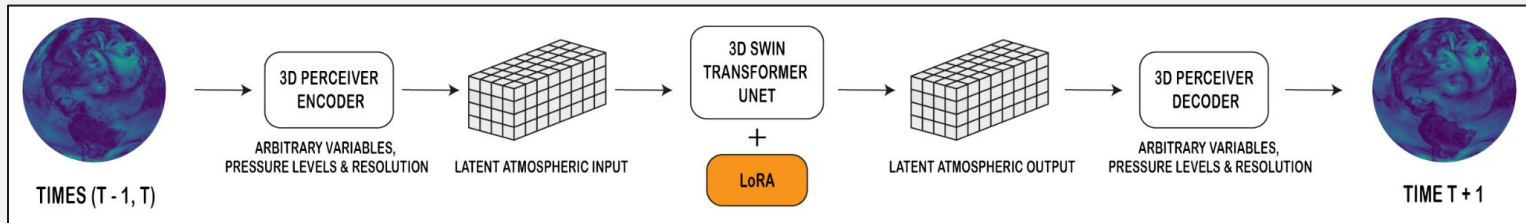


Transformer-based model with latent rollout



- 1hr/6hr processor “rolls out” the forecast in latent space to desired lead time
- Latent rollout avoids duplicative encode/decode steps

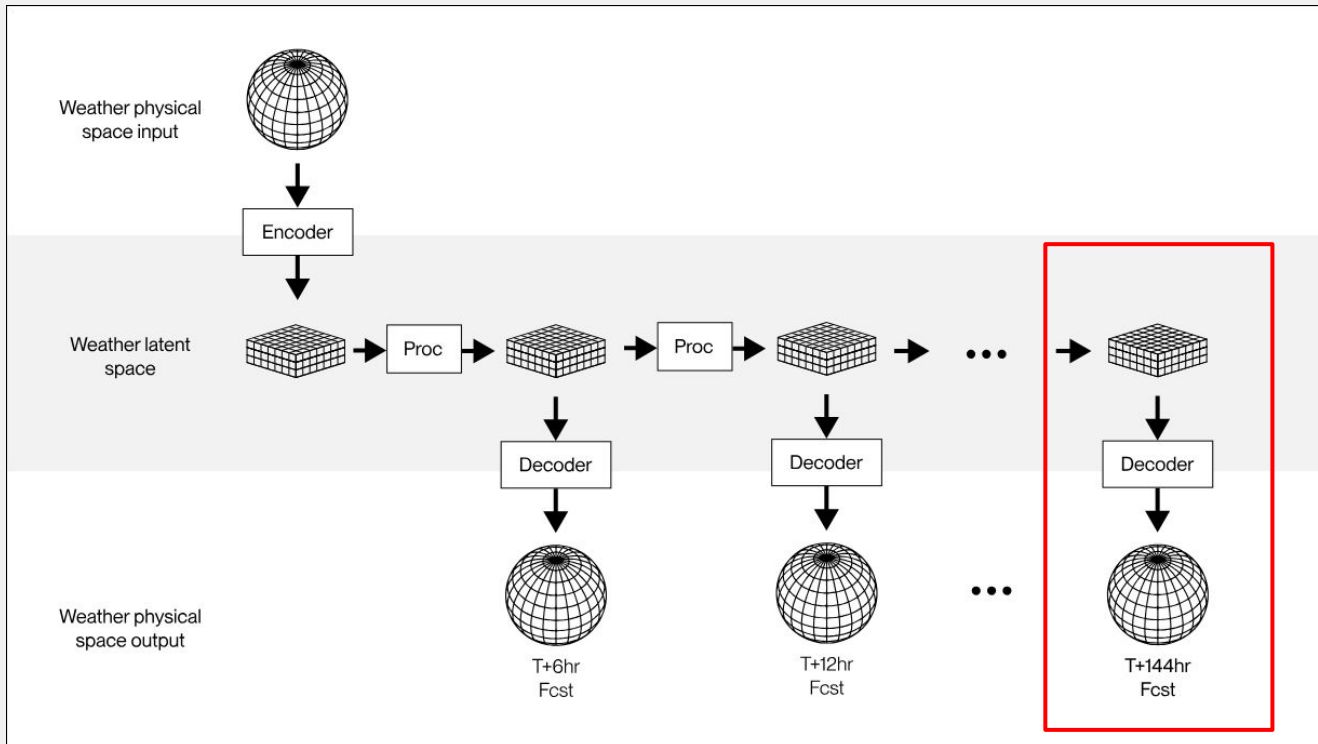
Transformer-based model with latent rollout



Typical autoregressive layout example, from Bodnar et al., 2024:

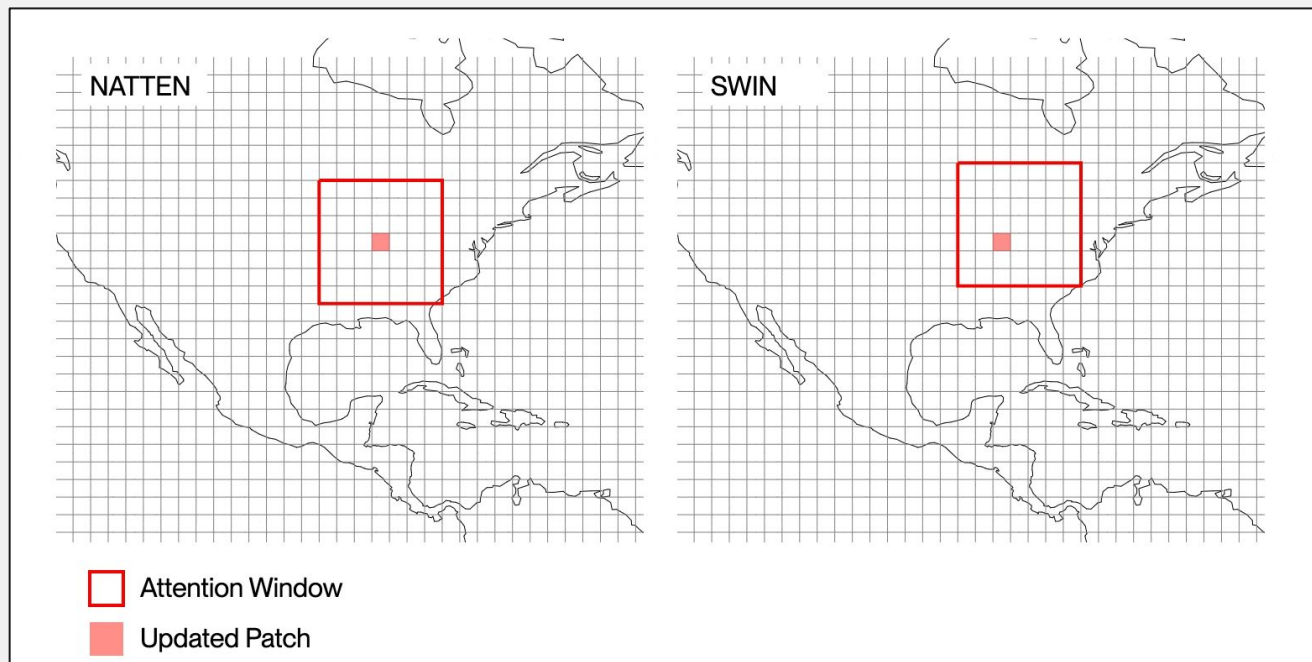
- Physical weather state (at time t) encoded
- Forecast “stepped forward” in latent space
- Decoded from latent space to physical space for time $t+1$
- Repeat from $t+1$ for $t+2$

Transformer-based model with latent rollout



- 1hr/6hr processor “rolls out” the forecast in latent space to desired lead time
- Latent rollout avoids duplicative encode/decode steps

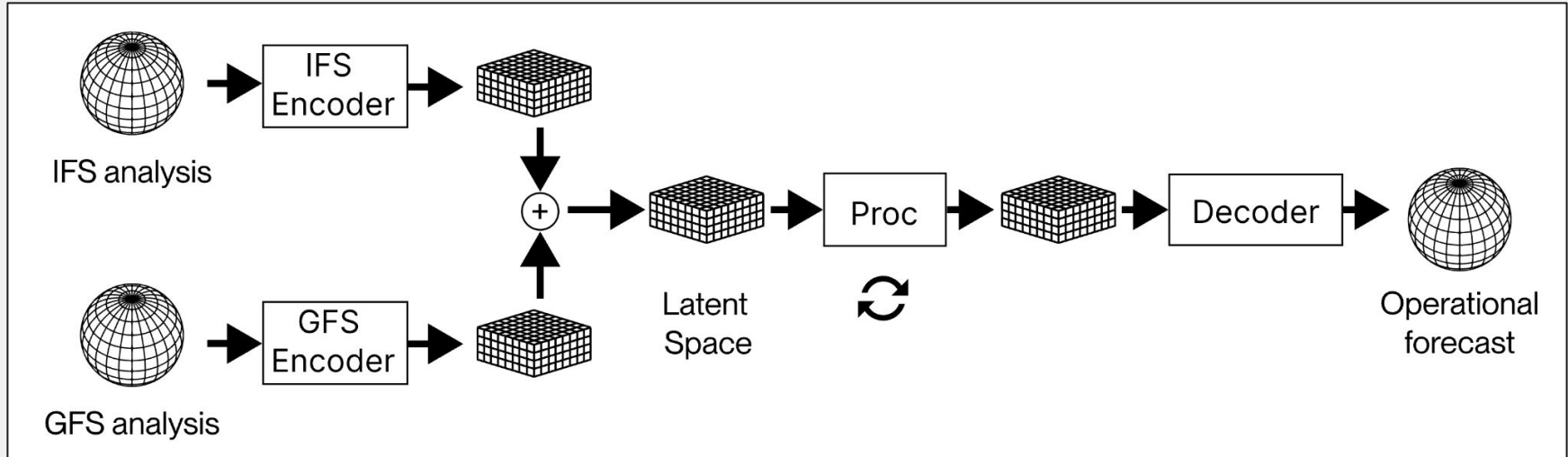
Neighborhood Attention + matepoint



- NATTEN (vs. SWIN) provides better inductive bias due to its consistent locality of attention for information transfer between patches
- Matepoint optimizes VRAM utilization (scheduled release)

Operational fine-tuning for real-time use

- Pretrained on ERA-5
- In operational settings, uses mixed input of IFS + GFS initializations
- **Modular architecture** is a significant advantage in improving initial conditions

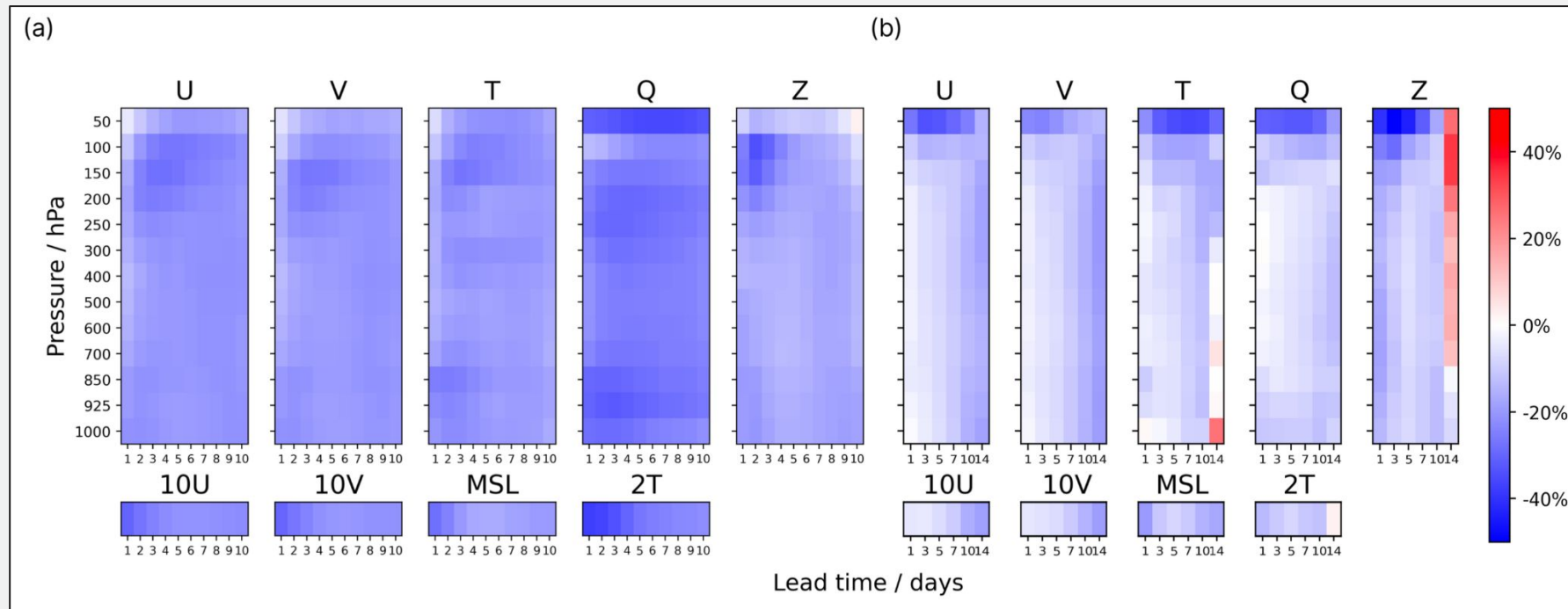




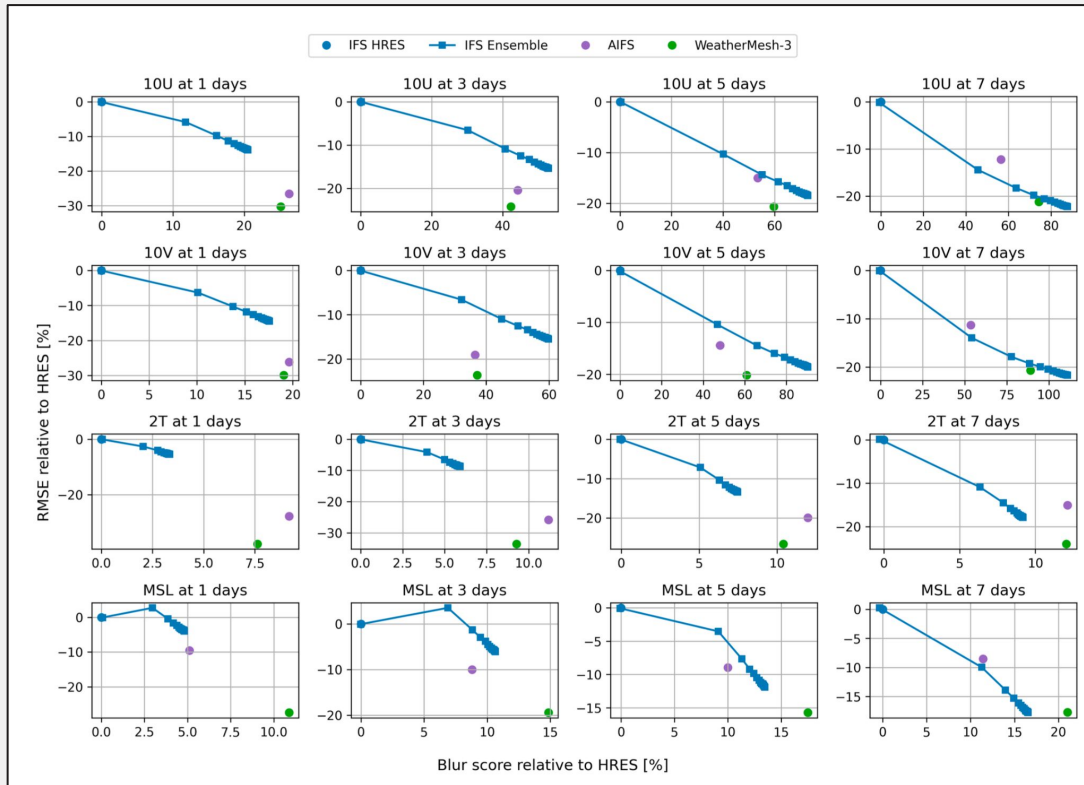
Results + Discussion



Significant RMSE improvements over both leading NWP and AI4NWP systems



Demonstrates Pareto-optimal blur-skill performance tradeoffs



- Blur calculated based on spectral power at 500km wavelength
- Note cases where WM-3 outputs are both more skillful and less blurry

WeatherMesh-3 is a state-of-the-art AI4NWP system that aims to deliver complementary benefits to NWP systems

1. Operational Performance

Methodological advancements allow WM-3 to achieve high performance, particularly along blur-skill boundary

2. Compute-Accessibility

Lightweight model makes weather forecasts accessible when compute-restricted, without compromising performance

3. Modular System for Customizability

Architecture with easily adjustable encoders / processors enables easy ingestion of custom data sources to enhance forecast accuracy

WeatherMesh-3 is a state-of-the-art AI4NWP system that aims to deliver complementary benefits to NWP systems

1. Operational Performance

Methodological advancements allow WM-3 to achieve high performance, particularly along blur-skill boundary

2. Compute-Accessibility

Lightweight model makes weather forecasts accessible when compute-restricted, without compromising performance

3. Modular System for Customizability

Architecture with easily adjustable encoders / processors enables easy ingestion of custom data sources to enhance forecast accuracy

AI4NWP systems are **complementary systems** to currently existing NWP efforts by agencies like NOAA or ECMWF.



Strong observation networks + support for NWP are critical to model performance.



Excited to share our work
with you at ICLR 2025!

