Comparison of AI Weather Predictions and Traditional Numerical Weather Predictions (Or: How Good is AIFS, Really?)

Jared A. Lee[1], Stefano Alessandrini[1], Tom Brummet[1], Simone Sperati[2], and Sue Ellen Haupt[1]

[1] Research Applications Laboratory, NSF NCAR
[2] RSE (Ricerca sul Sistema Energetico), S.p.A. (Italy)

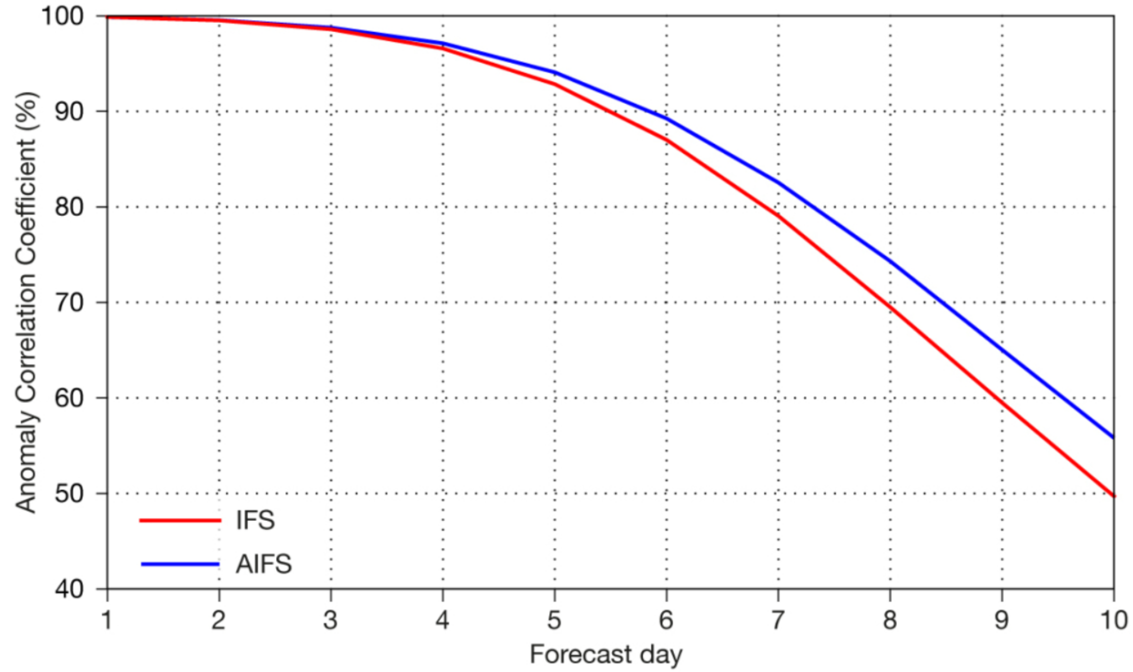
ESIG Forecasting & Markets Workshop ◆ Nashville, TN
24 June 2025

Photo by NASA on Unsplash

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Recent progress in data-driven weather forecasting

- Significant advancements have been made in data-driven weather forecasting
- Major tech companies (e.g., Google, Huawei, Nvidia) have developed purely data-driven models
- These models **outperform traditional physics-based NWP models** on key metrics, such as:
  - **Root mean square error (RMSE)**
  - **Anomaly correlation coefficient (ACC)** at 500 hPa geopotential height
- Training is based on historical datasets, typically subsets of **ECMWF's ERA5 reanalysis**
- Forecasts use **traditional NWP analyses** as initial conditions



AIFS forecast skill. We show the northern hemisphere Anomaly Correlation Coefficient (ACC) for geopotential height at 500 hPa of IFS forecasts (red, dashed) and AIFS forecasts (blue) for 2022. Higher values indicate better skill.

Adapted from: Lang et al. 2024

ECMWF AIFS overview

- The Artificial Intelligence Forecasting System (AIFS) employs a modern machine learning (ML) approach with an encoder–processor–decoder structure:

- **Encoder:** Utilizes attention-based graph neural networks (GNNs) to process input data
- **Processor:** Implements a sliding-window transformer to model temporal dependencies
- **Decoder:** It also utilizes attention-based GNNs to generate forecasts from processed data

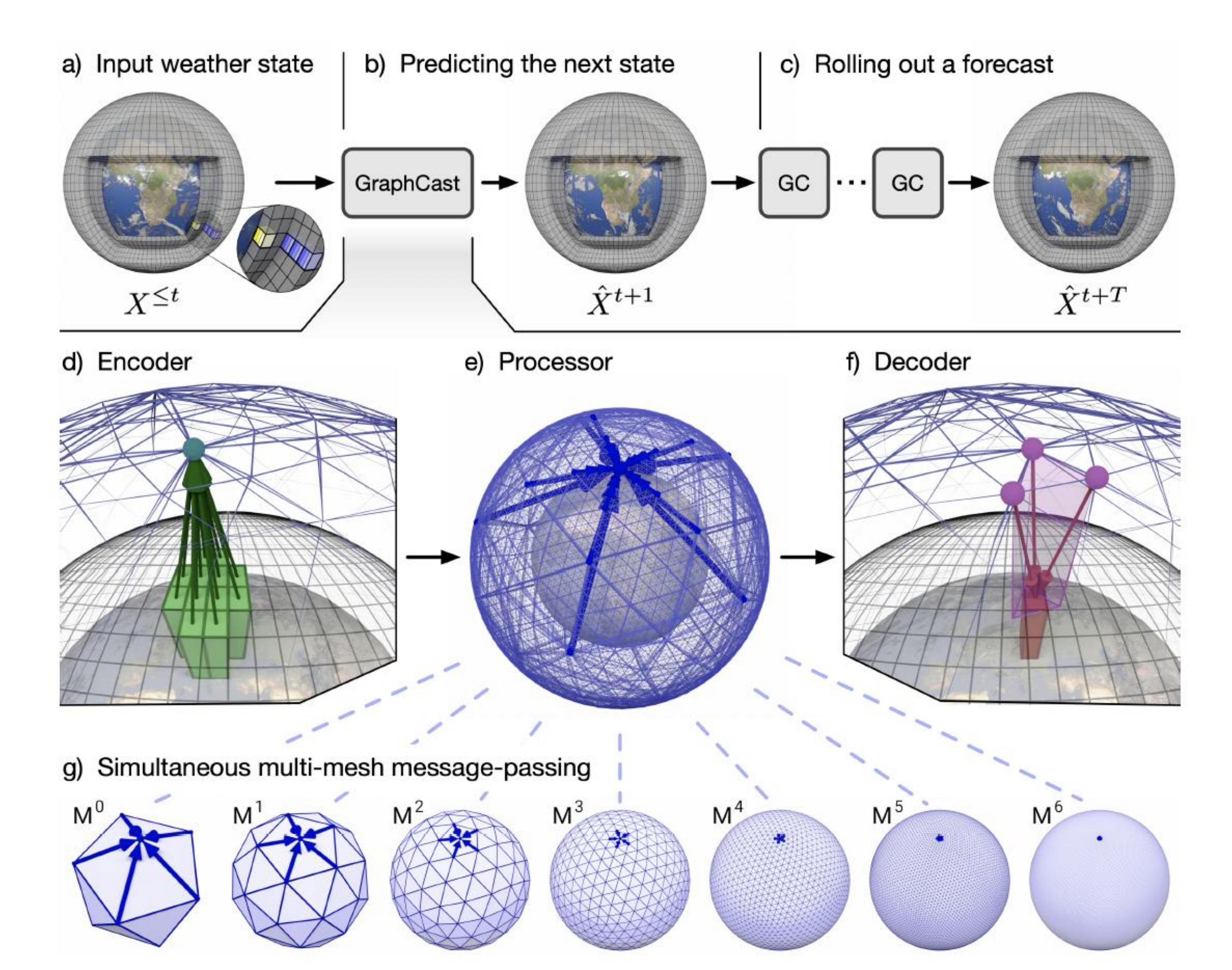
- This architecture is designed to handle high-resolution input data efficiently and supports parallelism for scalable training
- AIFS is trained to produce 6-hour forecasts. It receives as input a representation of the atmospheric states (ERA5 or ECMWF's operational analysis) at t–6h, t_0, and then forecasts the state at time t+6h.

Field	Level type	Input/Output
Geopotential, horizontal and vertical wind components, specific humidity, temperature	Pressure level: 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, 1000	Both
Surface pressure, mean sea-level pressure, sea-surface temperature, skin temperature, 2 m temperature, 2 m dewpoint temperature, 10 m horizontal wind components, total column water	Surface	Both
Total precipitation, convective precipitation	Surface	Output
Land-sea mask, orography, standard deviation of sub-grid orography, slope of sub-scale orography, insolation, latitude/longitude, time of day/day of year	Surface	Input

Table 1: Input and output variables of AIFS.

Adapted from: Lang et al. 2024

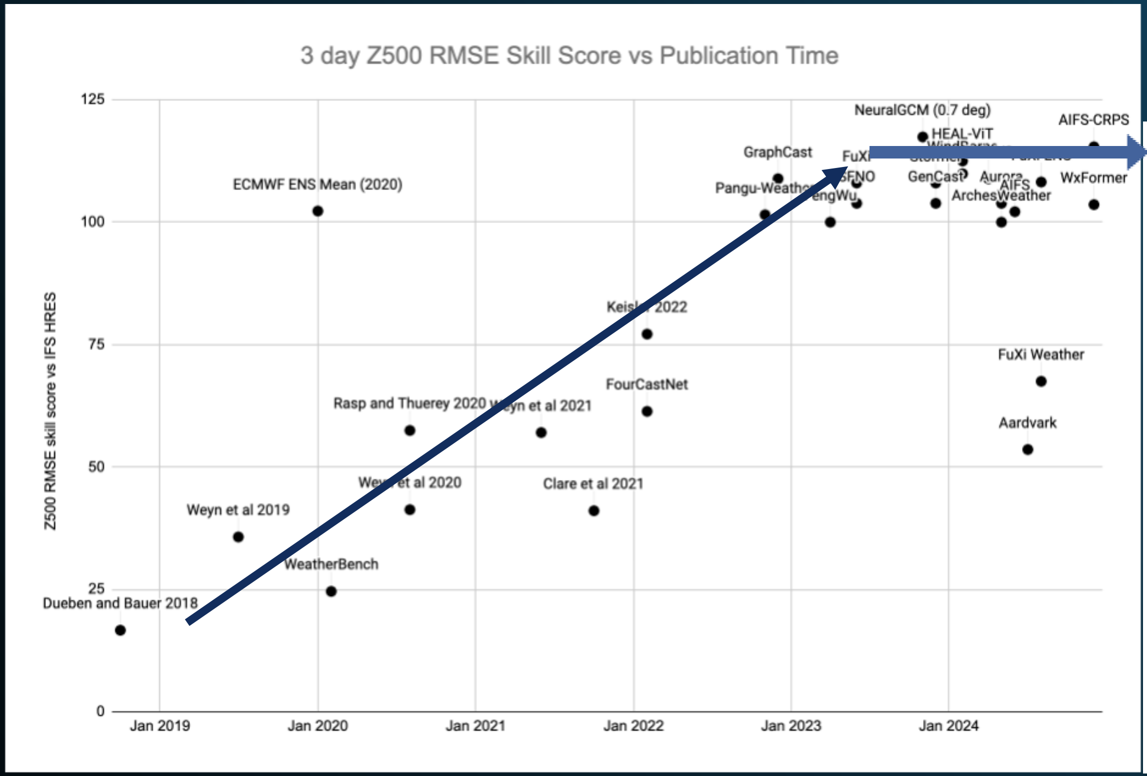
AIFS technology is similar to GraphCast

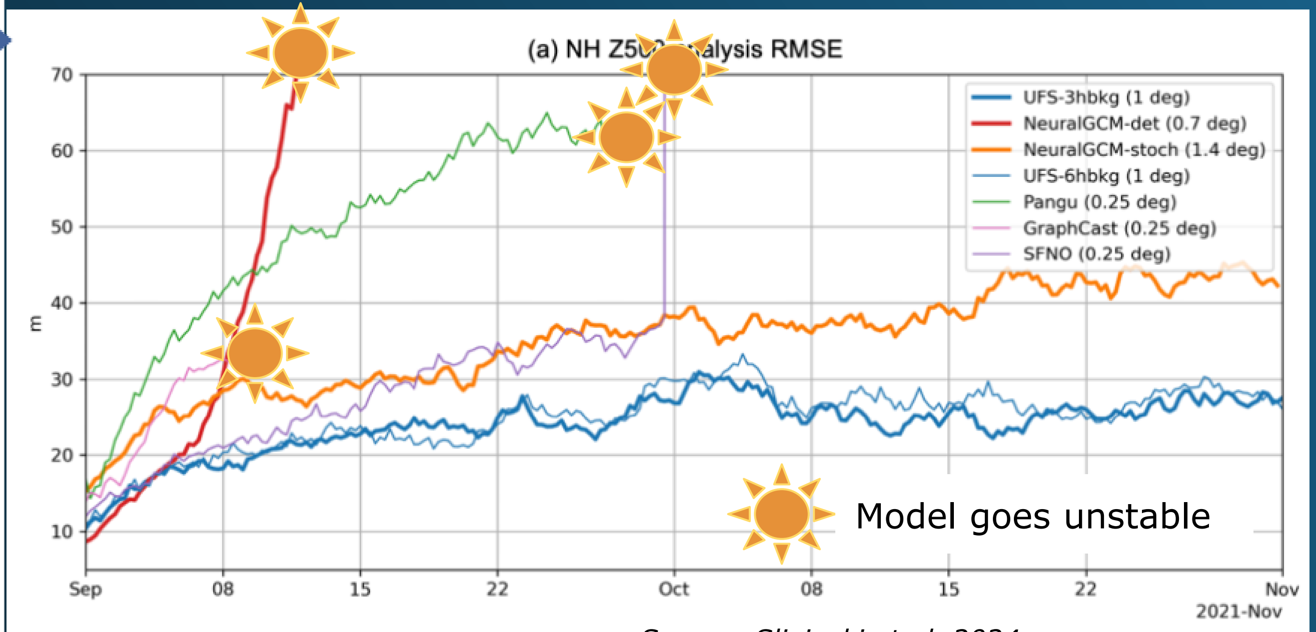


From: Lam et al., 2022. GraphCast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*.

- a) The input weather states are defined on a high-resolution latitude–longitude pressure-level grid
- b) GraphCast predicts the next state of the weather as the latitude–longitude–pressure levels grid
- c) A forecast is made by iteratively applying GraphCast to each previous predicted state, to produce a sequence of states which represent the weather at successive lead times

An example problem: AI weather prediction





Source: Slivinski et al. 2024

- After 4 years of rapid advancement in accuracy, further advancements in AI weather modeling have shown diminishing returns in improving global metrics
- Experiments with data assimilation and ensembles have revealed physical inconsistencies and instabilities that require more engagement with the data and physics to address

Models evaluated

- **ECMWF IFS (Integrated Forecasting System)**
  - **Type:** Global deterministic & ensemble NWP model
  - **Grid spacing:**
    - ~9 km (HRES deterministic)
    - ~18 km (ENS, ensemble system)
  - **Lead time:** Up to 10–15 days (HRES), 15 days (ENS)
  - **Updates:** 2x/day
  - **Strengths:** High skill in medium-range forecasts, advanced physics

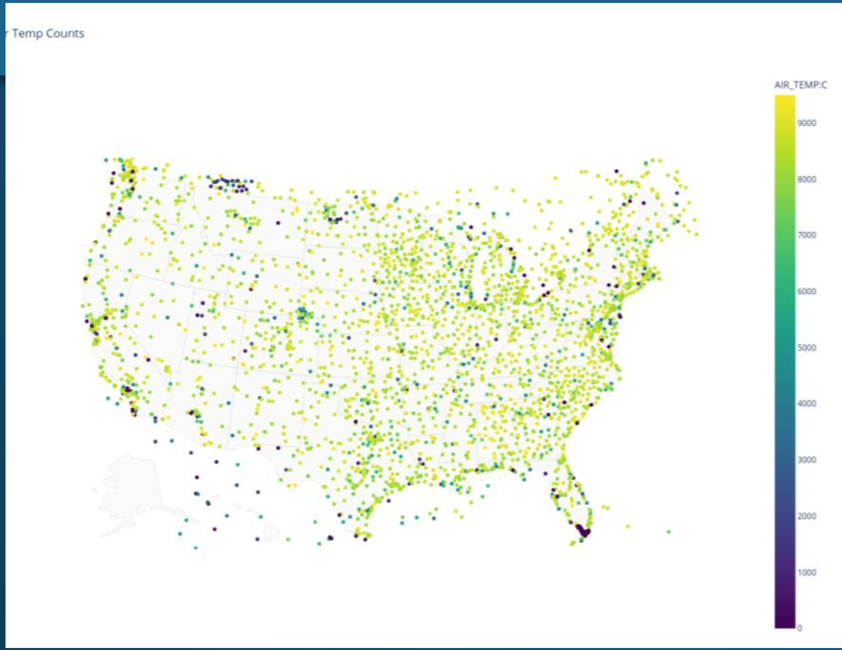
- **NOAA GFS (Global Forecast System)**
  - **Type:** Global NWP model (FV3 dycore)
  - **Grid spacing:** ~31 km (0.25°, publicly available)
  - **Lead time:** Up to 16 days
  - **Updates:** 4x/day
  - **Strengths:** Long-range forecasting, used for global guidance

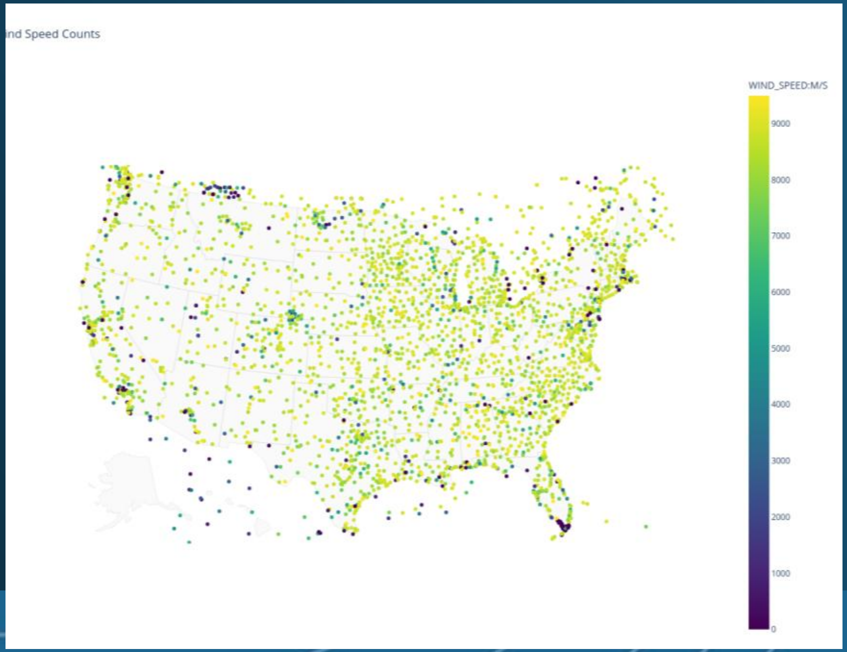
- **ECMWF AIFS (AI Forecasting System)**
  - **Type:** AI-based emulation of IFS forecasts
  - **Grid spacing:** ~31 km (0.25°) currently
  - **Lead time:** Up to 15 days
  - **Updates:** 2x/day
  - **Strengths:** Ultra-fast, energy-efficient forecasting using ML surrogates

- **NOAA HRRR (High-Resolution Rapid Refresh)**
  - **Type:** Regional, convection-allowing model (CONUS, AK, HI, PR), WRF-based
  - **Grid spacing:** 3 km
  - **Lead time:** 18–48 hours
  - **Updates:** Hourly
  - **Strengths:** High-resolution short-range forecasts, quite useful for severe weather, renewable energy, and other applications

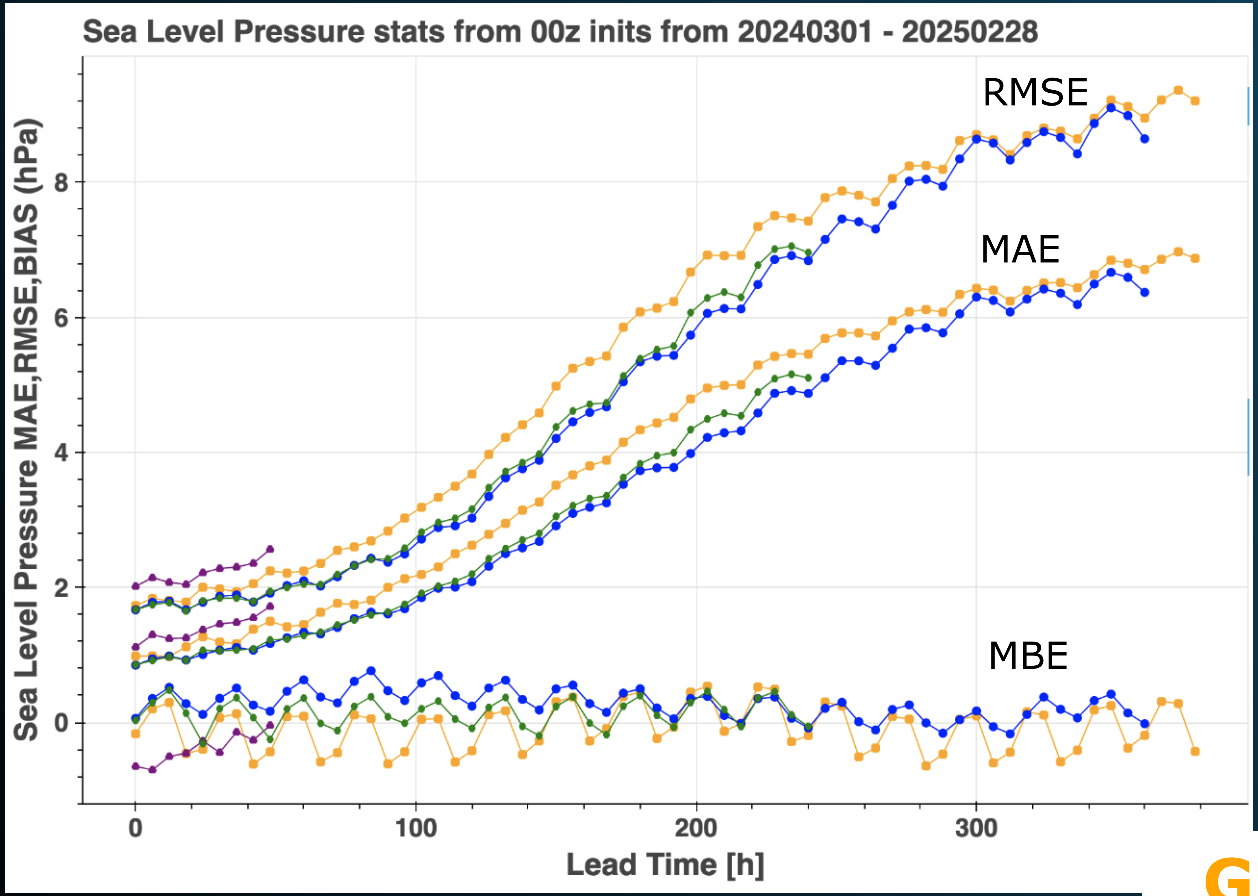
Observations

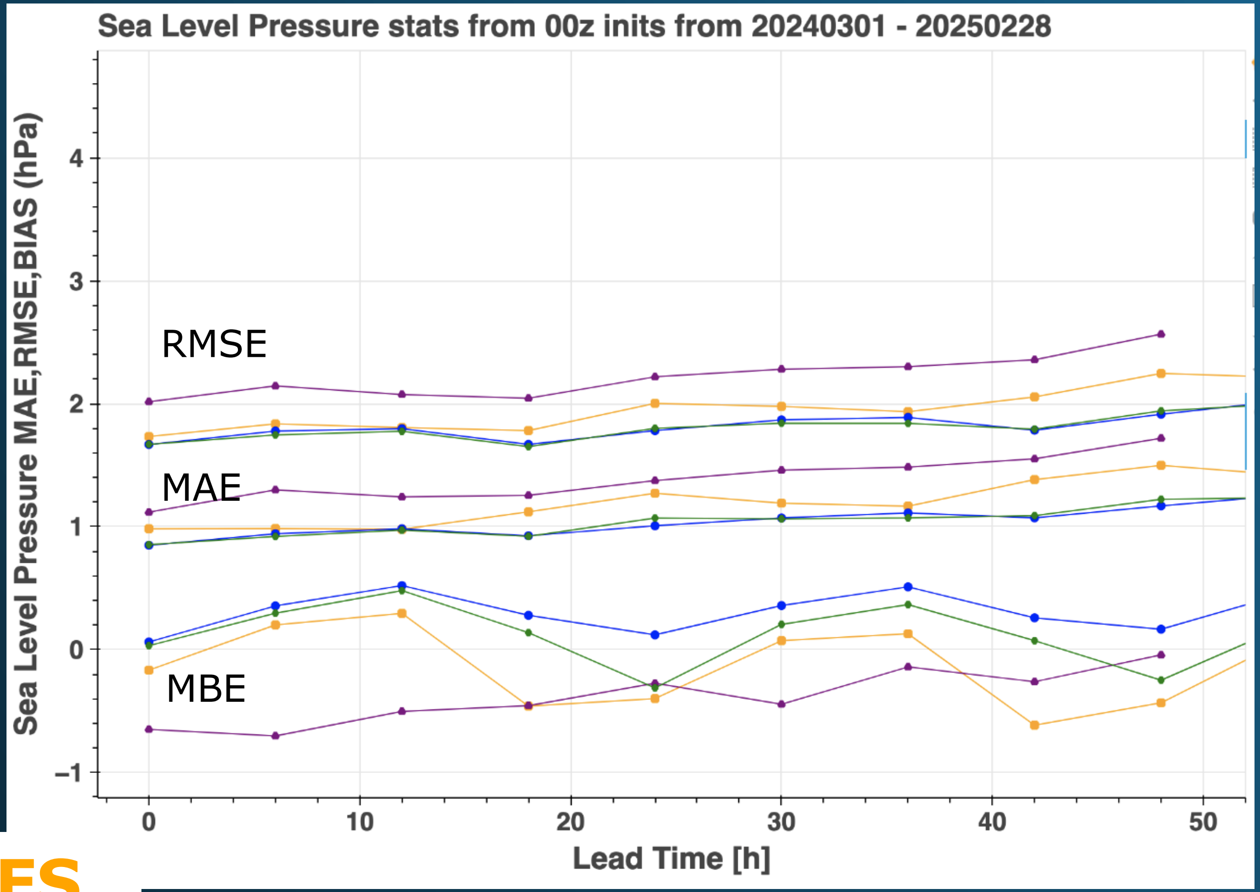
- **Automated Surface Observing System (ASOS) and Automated Weather Observing System (AWOS)** provide continuous, real-time surface weather observations across the U.S., primarily at airports
- **Key variables measured** include 2-m temperature, 2-m dewpoint, 10-m wind speed/direction, visibility, pressure, precipitation, and sky conditions (e.g., cloud height and coverage)
- **High temporal resolution**, with updates typically every 1 minute (AWOS) to 5–10 minutes (ASOS), supporting aviation, forecasting, and climatology
- **Maintained by NOAA, FAA, and state agencies**, offering a dense and standardized observational network essential for model validation and situational awareness





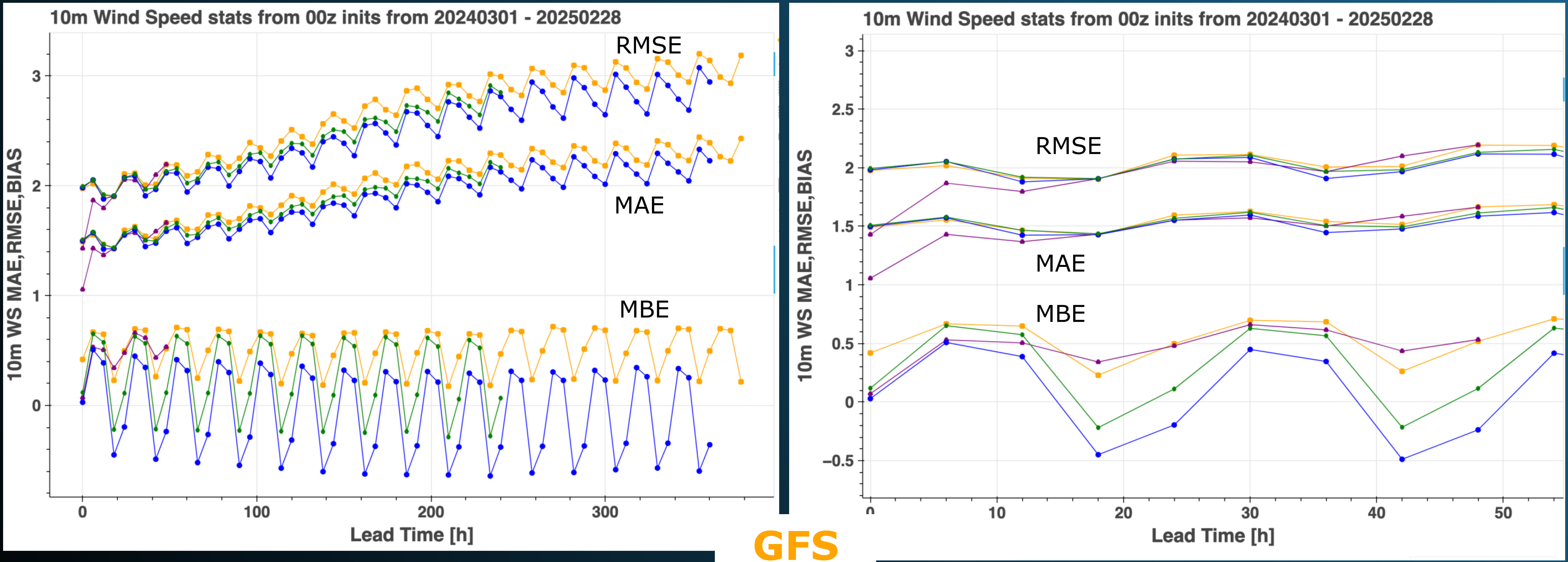
Sea-level pressure (RMSE, MAE, MBE)





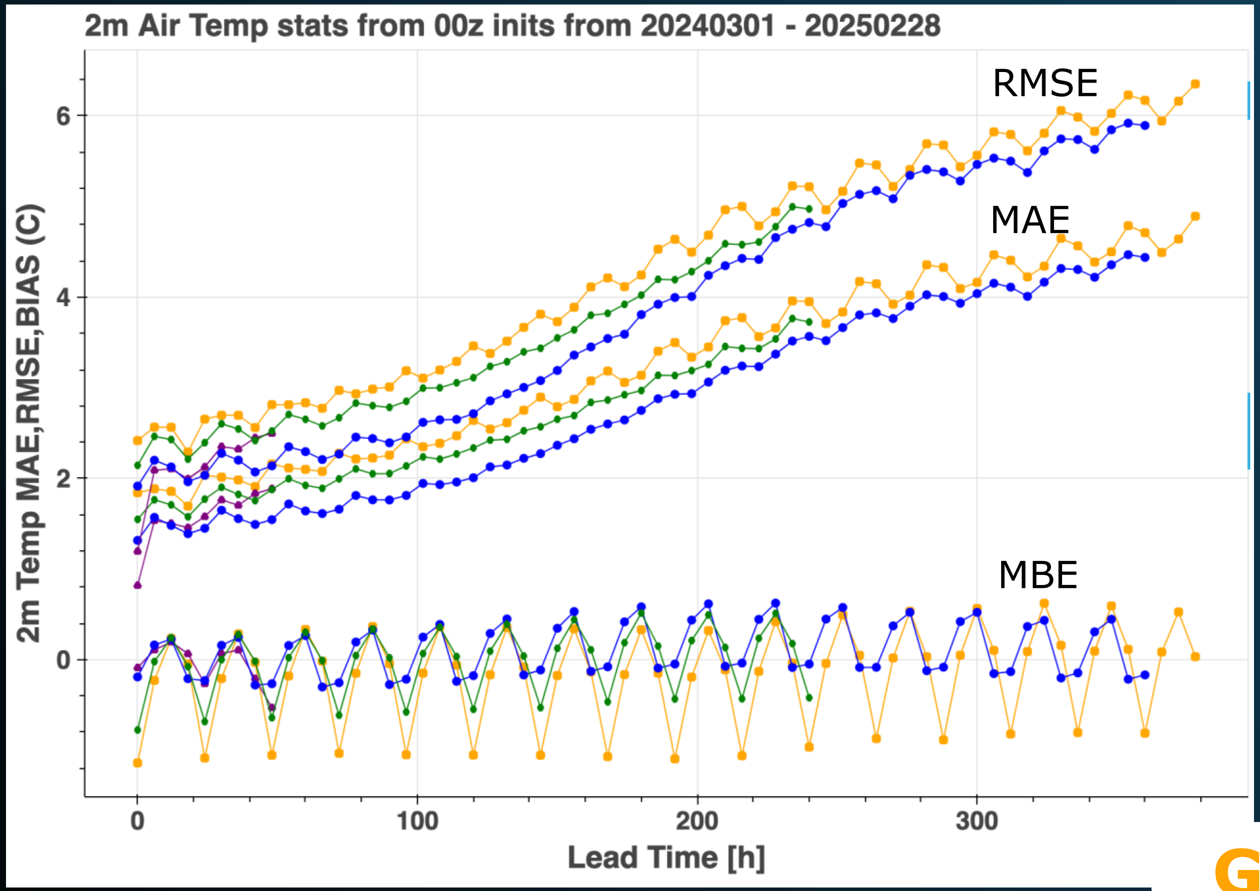
GFS
HRRR
IFS
AIFS

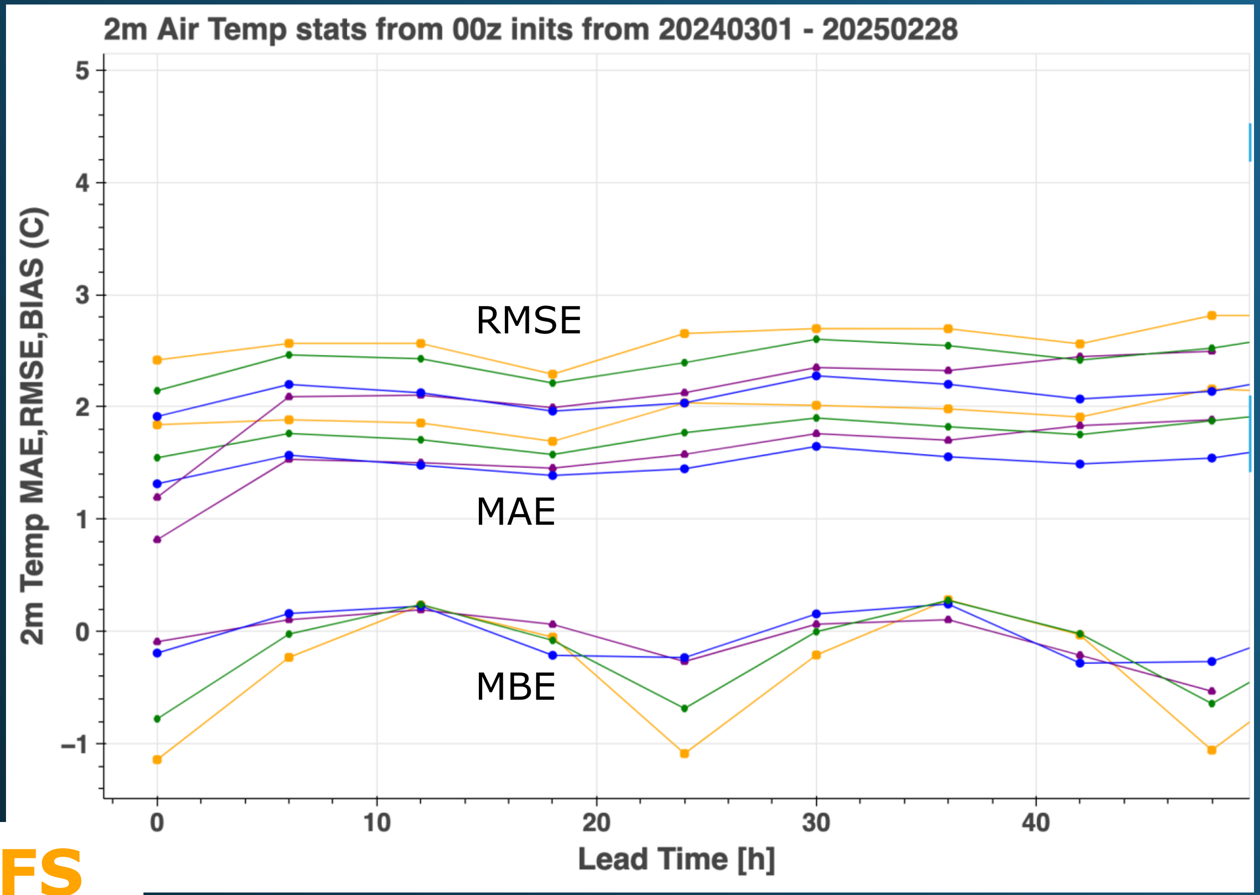
10-m wind speed (RMSE, MAE, MBE)



GFS
HRRR
IFS
AIFS

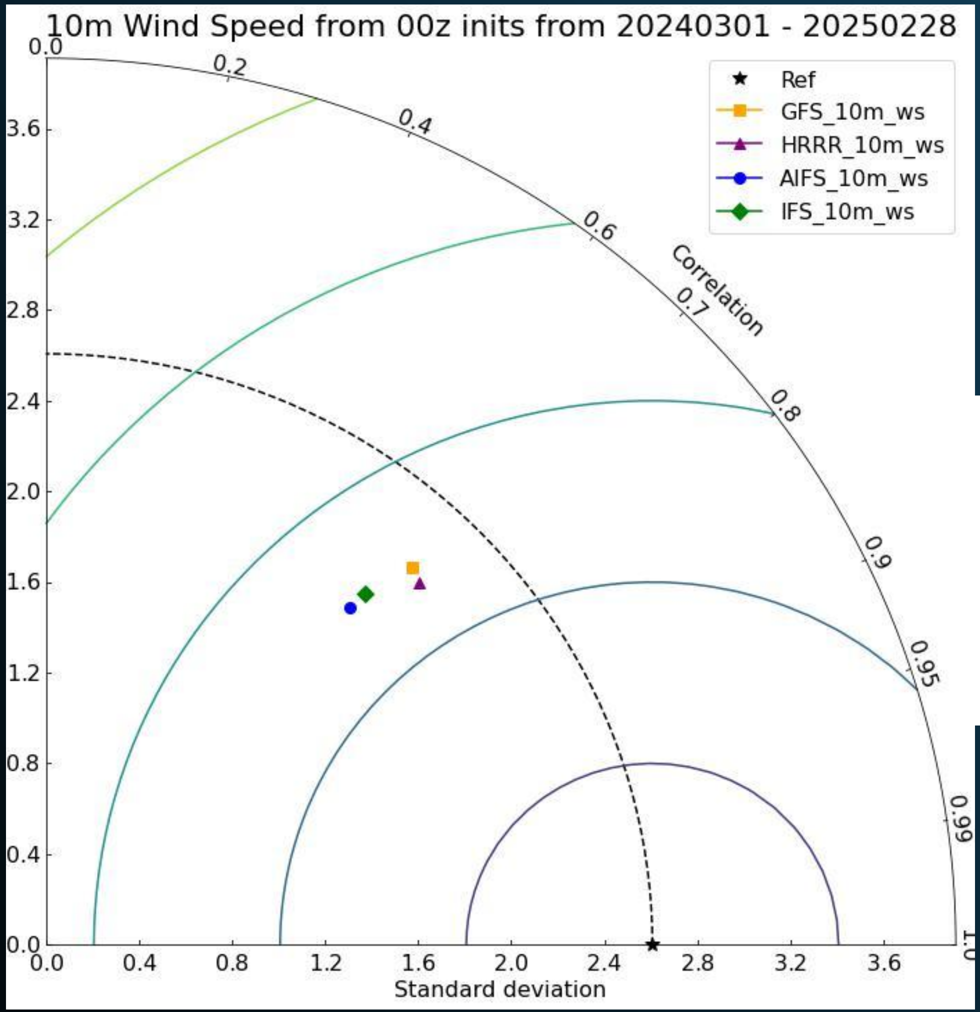
2-m air temperature (RMSE, MAE, MBE)



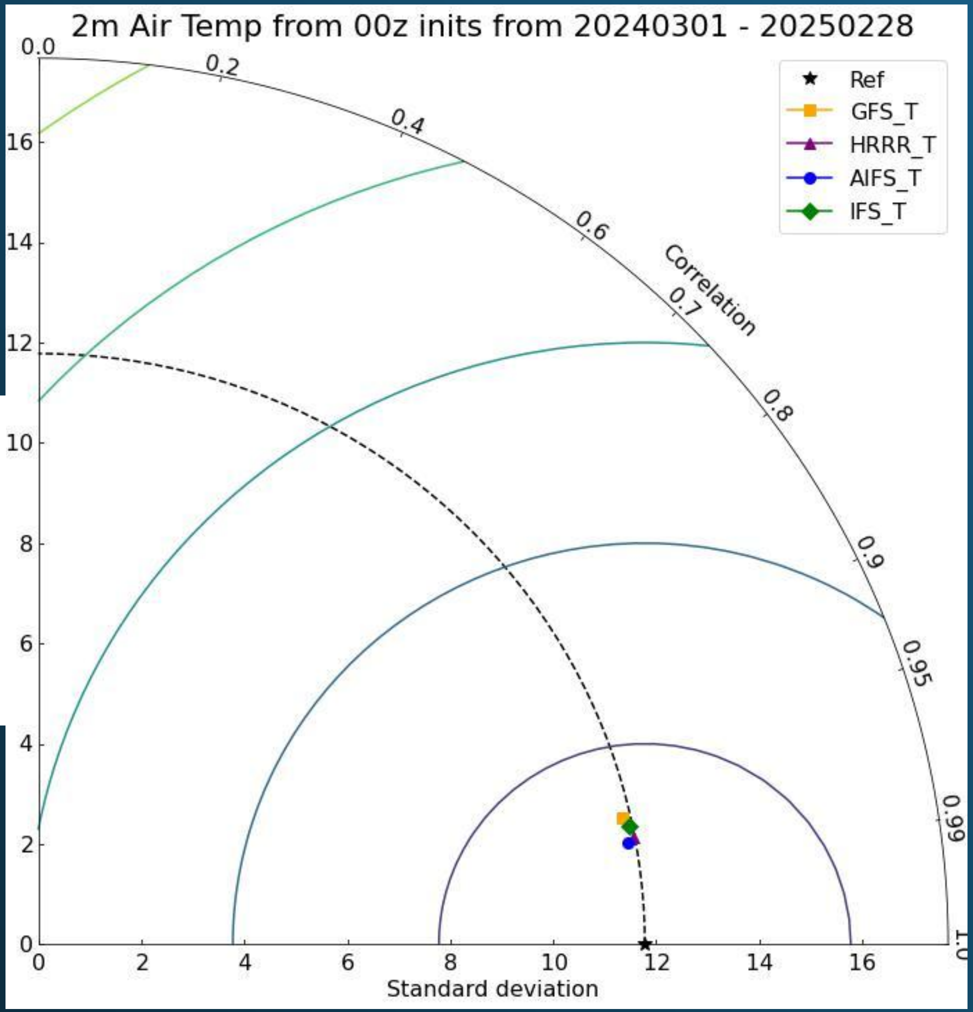


GFS
HRRR
IFS
AIFS

0–48-hour Taylor diagrams (std. dev., correlation, CRMSE)

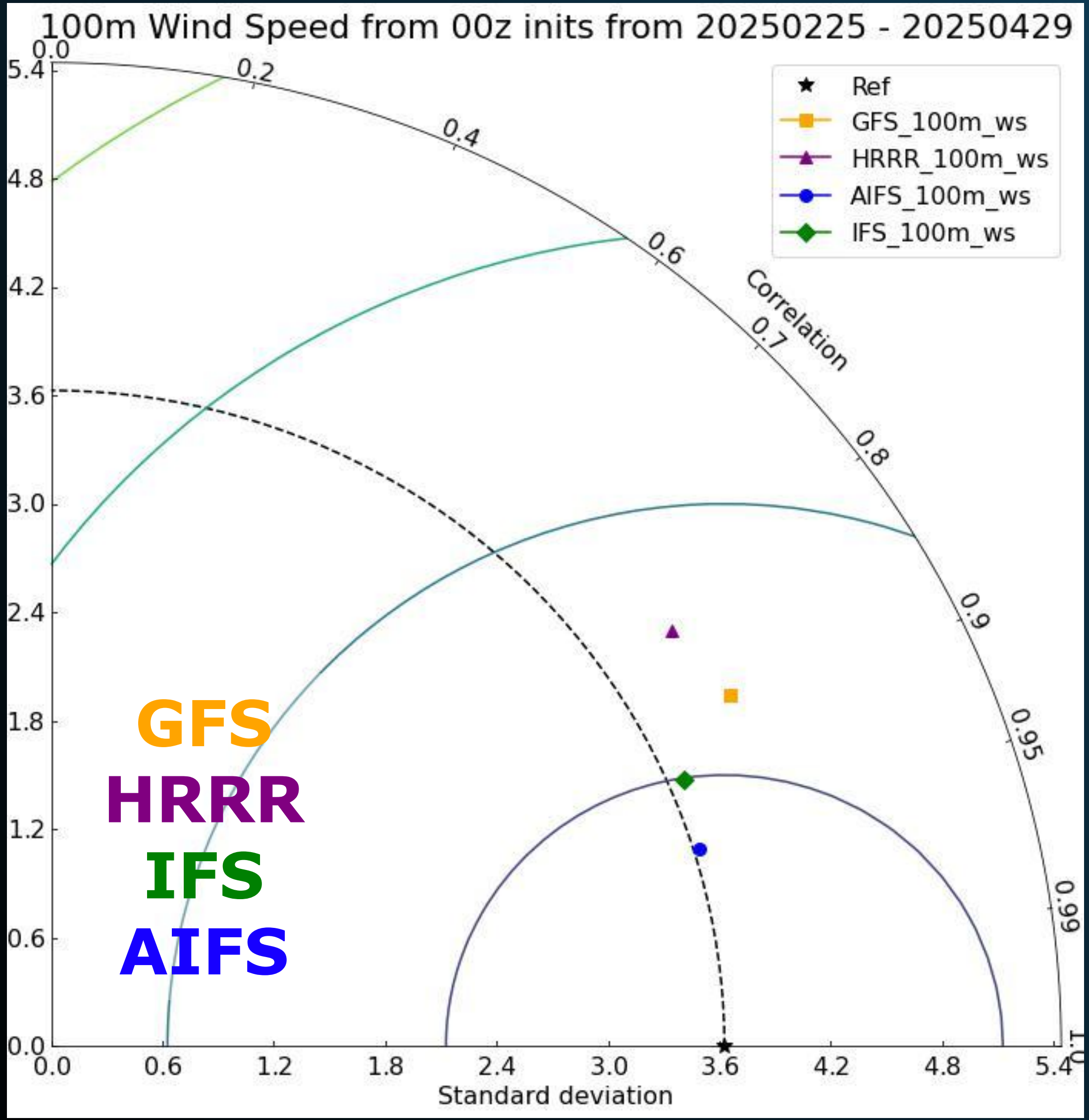


GFS
HRRR
IFS
AIFS



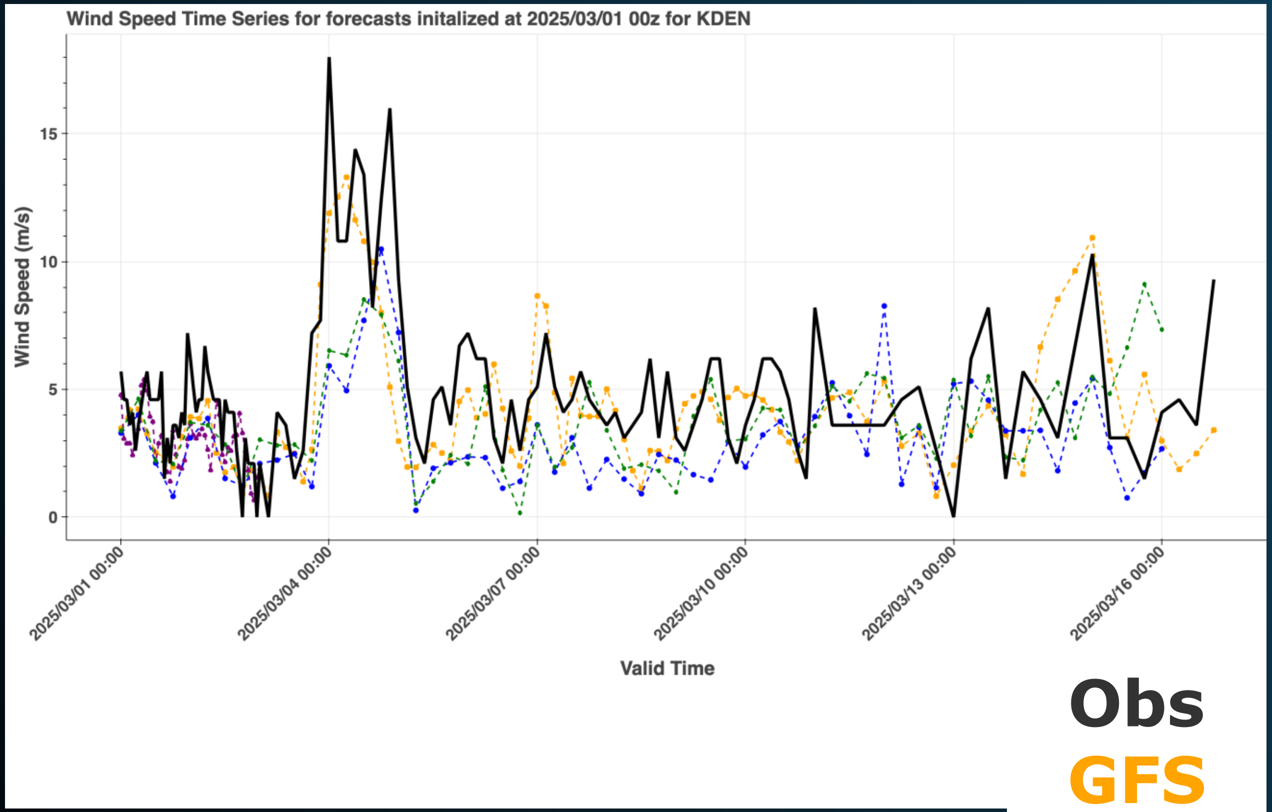
- AIFS has the lowest standard deviation (lowest variability) for wind speed
- HRRR has the best performance for wind speed

0–48-hour Taylor Diagrams (std. dev., correlation, CRMSE)
Validated against ERA5 100-m wind speed



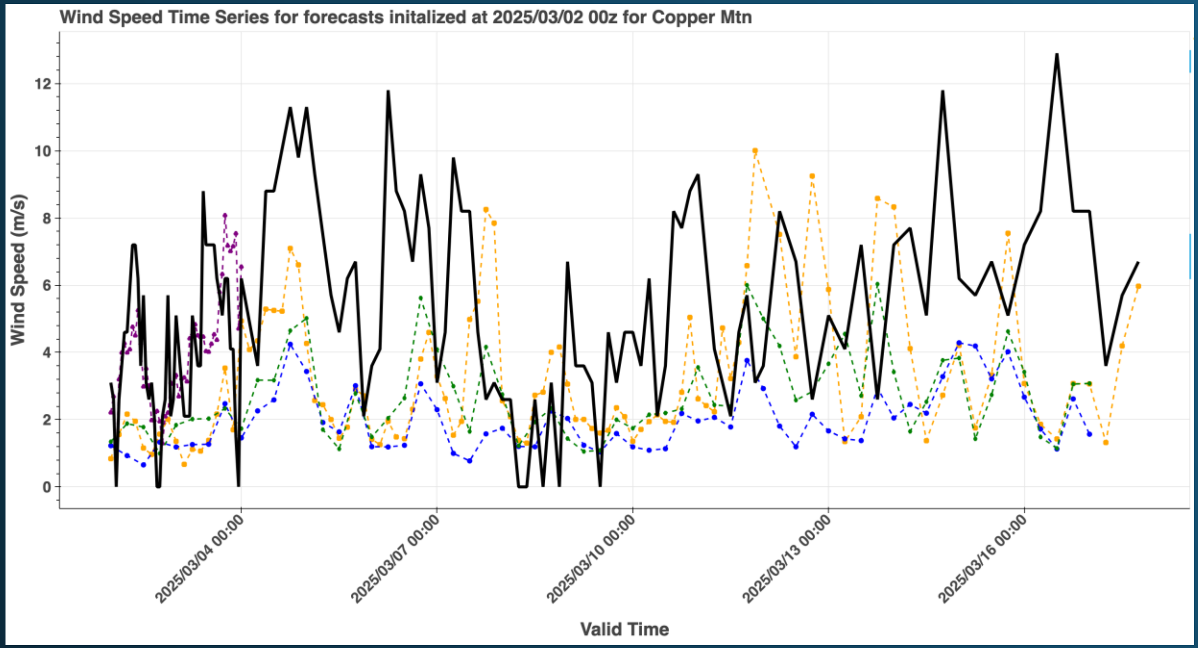
- AIFS has the best performance when validated against ERA5 100-m wind speed data
- ***However, this is not overly surprising, as AIFS was trained on ERA5 data up through 2018***
- Like any other model, ERA5 has its own biases and errors that should be accounted for
- It would be far preferable to validate against observations, but those are hard to find for 100-m wind speed
- Absent observations, want to validated against other analyses, too (HRRR analyses, GFS-FNL)
- It's also worth noting that HRRR does not output 100-m wind speed directly
  - Need to interpolate from 10-m & 80-m
- High-res HRRR likely penalized by low-res ERA5

10-m wind speed (time series for DEN and Copper Mtn)



- HRRR is the most realistic up to 48 hours
- AIFS seems to struggle the most in capturing the wind speed variability

Obs
GFS
HRRR
IFS
AIFS



NSF | NCAR Operated by UCAR

Conclusions

- **AI-based models (AIFS, GraphCast, etc.)** show strong potential and **can outperform traditional NWP models** in many global metrics (e.g., RMSE, ACC)
- **AIFS offers rapid and energy-efficient forecasts**, but currently with **lower resolution and limited lead times** compared to IFS HRES
- **HRRR remains superior for short-range and high-resolution forecasts**, particularly for surface variables like wind speed
- For wind speed, even after a simple bias correction (CRMSE), **AIFS does not outperform other models**
- **Further improvements are needed** in handling physical consistency, ensemble generation, and regional/local-scale performance
- **Data assimilation and hybrid approaches** may bridge the gap between physics-based and AI-based forecasting
- AIFS produces solar irradiance as **6-hourly *accumulated* GHI**, which is not yet suitable for solar energy forecasting applications

Plans & Outlook

- For our work…
  - Add more months to validation as 2025 progresses
  - Conduct regional validation, not just CONUS-wide
  - Conduct seasonal validation, not just the whole period
  - Conditional validation on different parts of variable distributions
  - Spatial validation maps, precipitation validation
  - Validate 100-m wind speed against other analyses
  - Also evaluate performance over Italy

- For AI weather forecasting in general…
  - Next steps: Develop **ensemble forecasts** using probabilistic training or **generative AI** (e.g., diffusion models)
  - Research underway to **couple AI models with observations** or train them directly on observational data
  - A new **ECMWF Member State pilot project** will support the development of AI models across global to local scales
  - Continued **rapid advances expected** in skillful, data-driven forecasting methods