WeatherMesh: An All-in-One AI System for Data Assimilation and Weather Forecasting

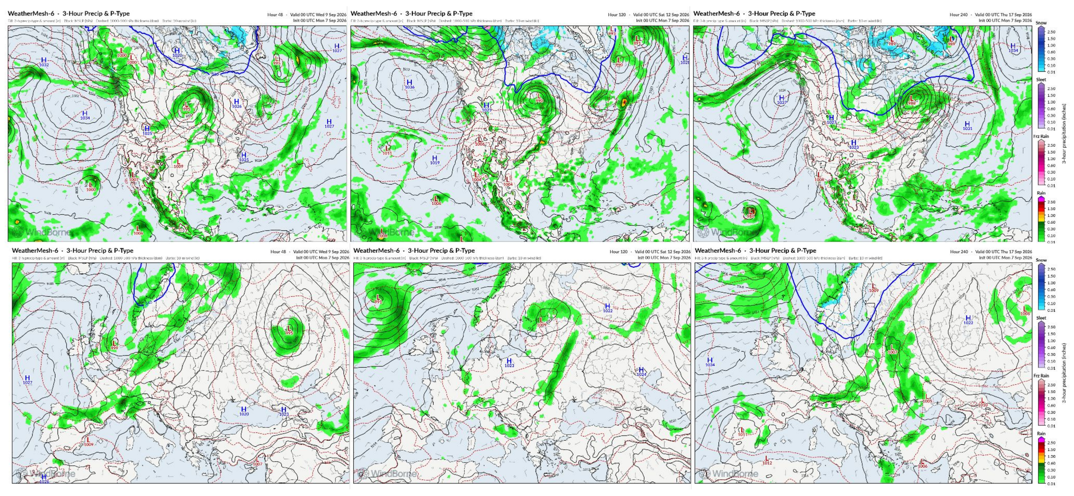
Todd Hutchinson, Joan Creus-Costa, Anuj Shetty, and Chris Riedel
WindBorne Systems, Redwood City, CA USA

INTRODUCTION

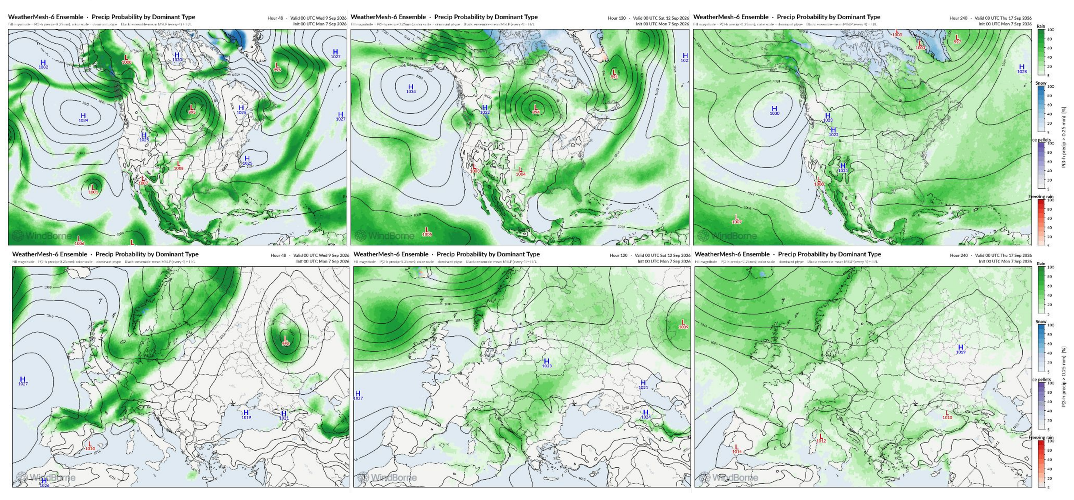
WeatherMesh, developed by WindBorne Systems, is a transformer-based AI weather model that **integrates data assimilation** directly into its forecast pipeline, assimilating **raw observations** to produce ensemble forecasts. WeatherMesh outperforms leading physics-based NWP systems and state-of-the-art AI weather models across nearly all parameters and lead times. WeatherMesh produces a **128-member ensemble** on pressure levels across a range of atmospheric variables.

	WM6-Global
Resolution / domain	0.25° global
Vertical	25 pressure levels up to 10 mb
Pressure level variables	Geopotential, temperature, winds, humidity
Surface/2D variables	2m T, 2m Td, min/max T, precipitation accum, snowfall, precipitation type, mslp, soil T, soil moisture, cloud cover, CAPE, PBL height, heat fluxes, solar and thermal radiation
Update / step	Hourly update; 3 h forecast output step
Forecast horizon	15 days
Forecast type	128 ensemble members

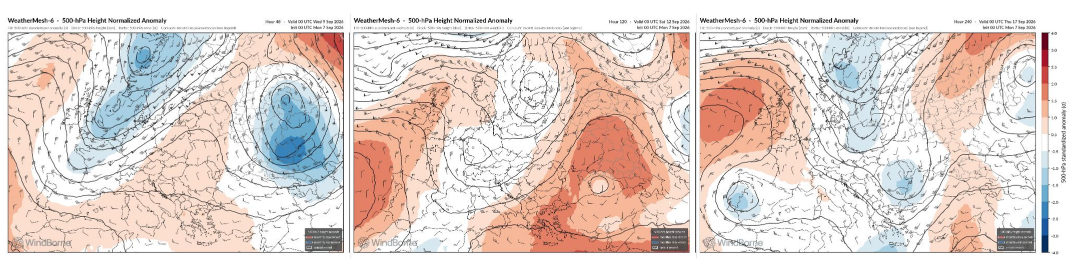
Precipitation type, accumulation, 500 hPa heights, and 1000 hPa heights illustrate the forecast evolution at Day 2, 5, and 10.



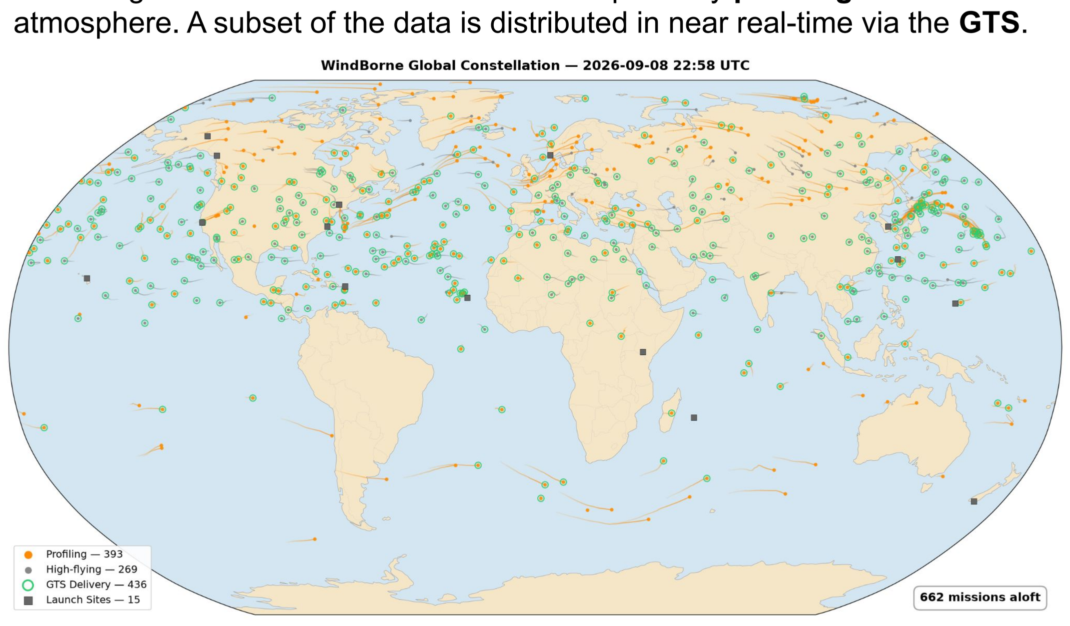
The 128-member ensemble enables probabilistic forecasts. Precipitation probability sharpens at short leads and broadens with increasing forecast hour, as expected.



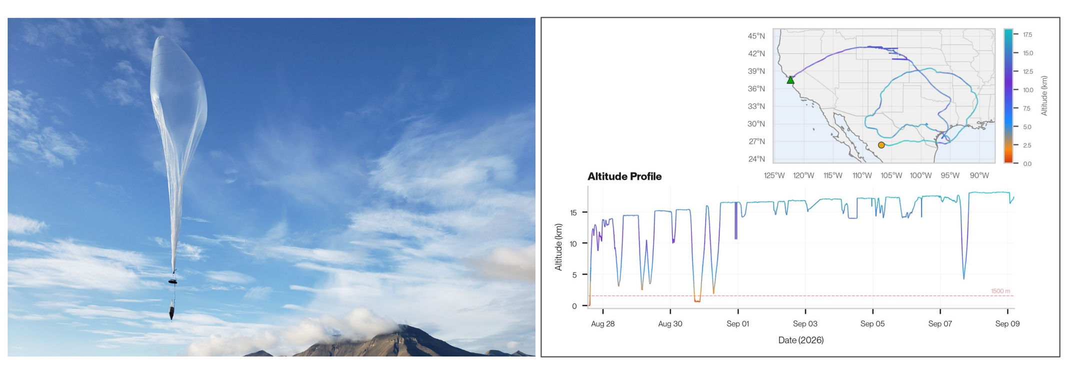
Anomaly forecasts of 500 hPa height highlight synoptic-scale pattern evolution, particularly valuable at extended range.



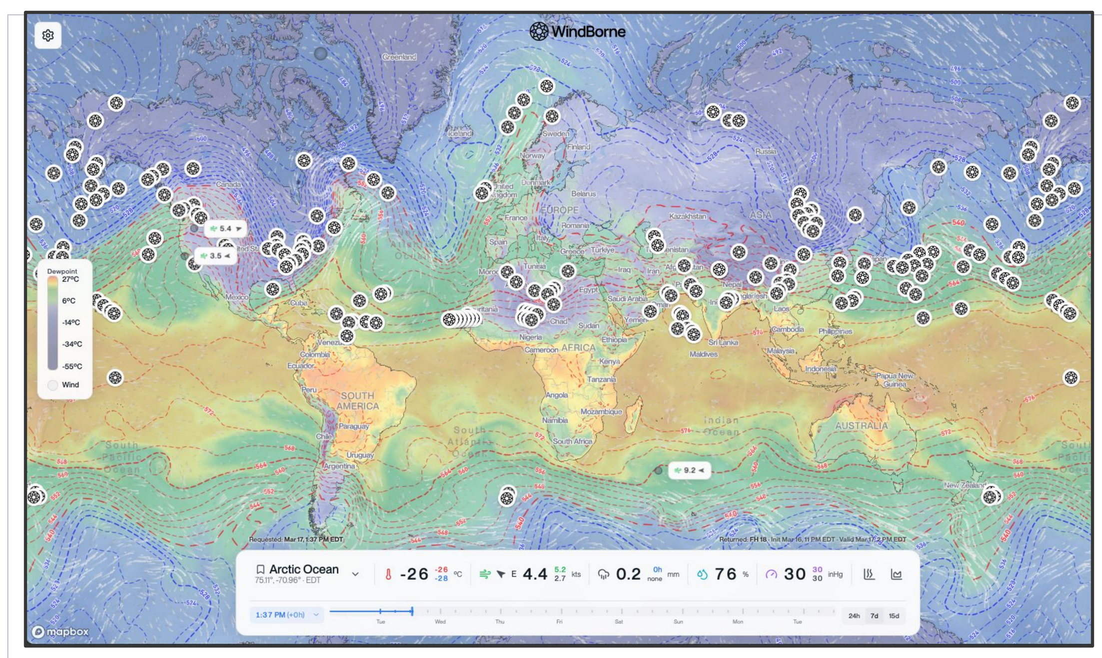
WeatherMesh assimilates a wide range of observational data sources; one unique source is **Atlas**, WindBorne's persistent global constellation of over **670 free-flying balloons**, currently collecting more than 1,500 vertical profiles per day across all ocean basins and continents. Observations in **data-sparse regions** — over oceans, the tropics, and the Southern Hemisphere — are expected to improve forecast skill globally. The constellation has grown several-fold over the past year and is on track to double again within the next year. Each element of the constellation is a **Global Sounding Balloon (GSB)** — a controllable, long-duration platform that remains aloft for **weeks**, traversing **thousands of kilometers** while repeatedly **profiling** the atmosphere. A subset of the data is distributed in near real-time via the **GTS**.



Each flight can is autonomously directed to ascend and descend to collect profiles of data throughout the troposphere. Frequently, descents reach down to several hundred meters above the surface.



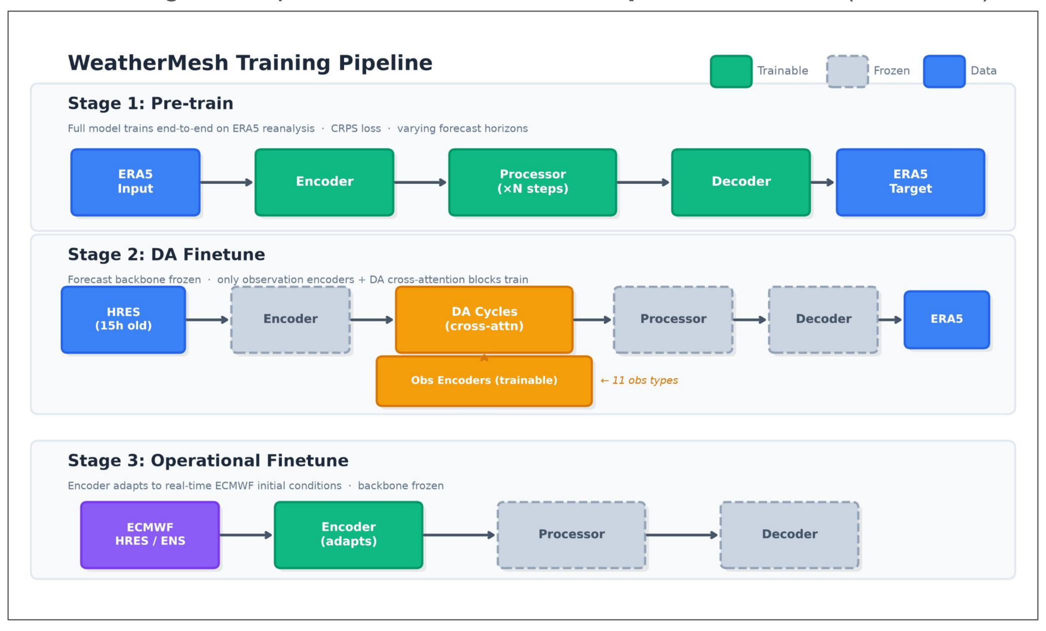
WindBorne Atlas By the Numbers	
670+	Balloons currently aloft
13,000+	Missions
~1,750	Atmospheric soundings per day
16	Permanent Launch sites



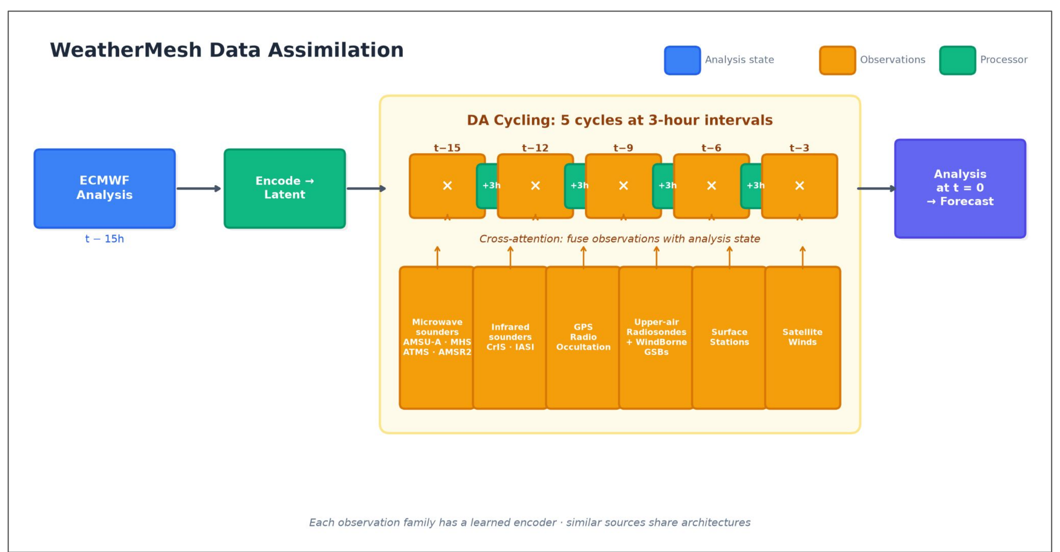
WeatherMesh DESIGN

WeatherMesh integrates data assimilation and weather forecasting into a **single end-to-end deep learning system**. Rather than treating DA as a separate preprocessing step, observations are assimilated directly within the model's learned **latent space** alongside the forecast pipeline. The system is trained in three stages, produces ensemble forecasts operationally, and runs every hour.

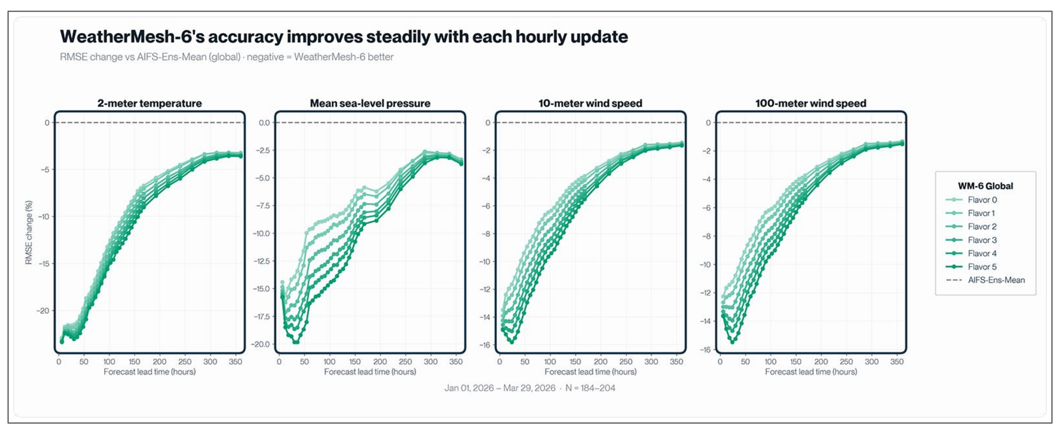
Training. WeatherMesh is a transformer-based model operating on a HEALPix icosahedral mesh, providing quasi-uniform global coverage without polar singularities. The model learns in a latent space: the processor advances the representation forward in 3-hour increments via neighborhood attention, applied repeatedly for longer forecasts. Pre-training uses **CRPS loss** across varying forecast horizons, producing calibrated ensemble forecasts naturally. All three stages complete in under **100,000 steps** on **16 GPUs** (~3 weeks).



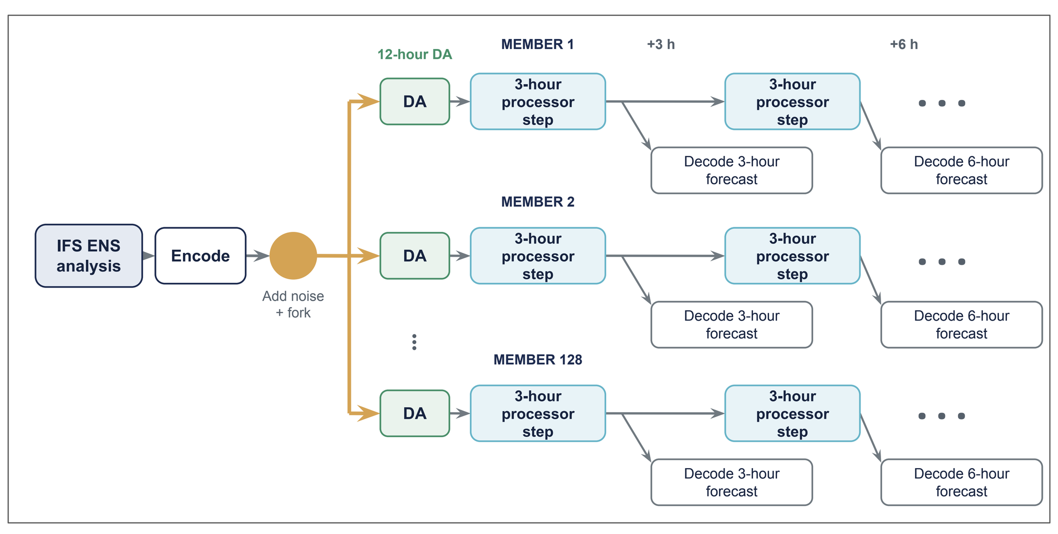
Data assimilation. DA occurs entirely within the forecast model's learned latent space — observations are not preprocessed into gridded fields, but assimilated directly through **learned cross-attention**. Multiple ensemble members pass through DA independently, producing **calibrated uncertainty** from analysis time onward. WindBorne balloon observations use **innovation-based encoding**: the model sees the difference between the measurement and the background prediction, enabling efficient learning from this rapidly growing data source.



Forecast accuracy improves steadily with each additional hour of data assimilation as more observations are incorporated through successive DA cycles. The benefit is largest at short lead times, where initial condition quality matters most.

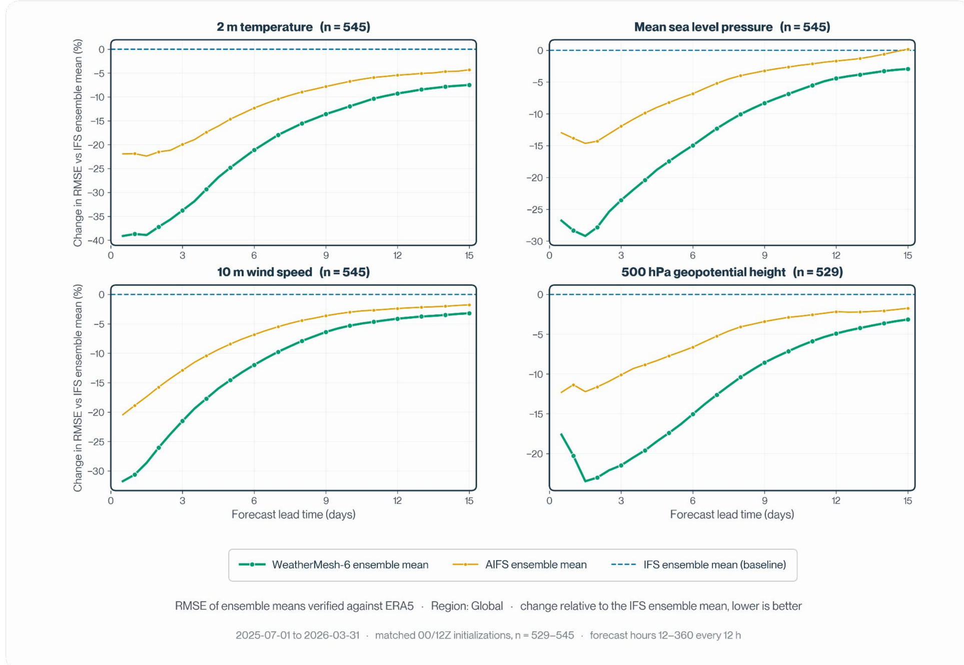


Inference. In operation, WeatherMesh produces a new 128-member ensemble **forecast every hour**. Each run ingests the two-cycle-old ECMWF analysis (to allow for more DA warm-up) along with all available observations from the preceding hours, and generates forecasts out to **15 days** at 3-hour intervals.

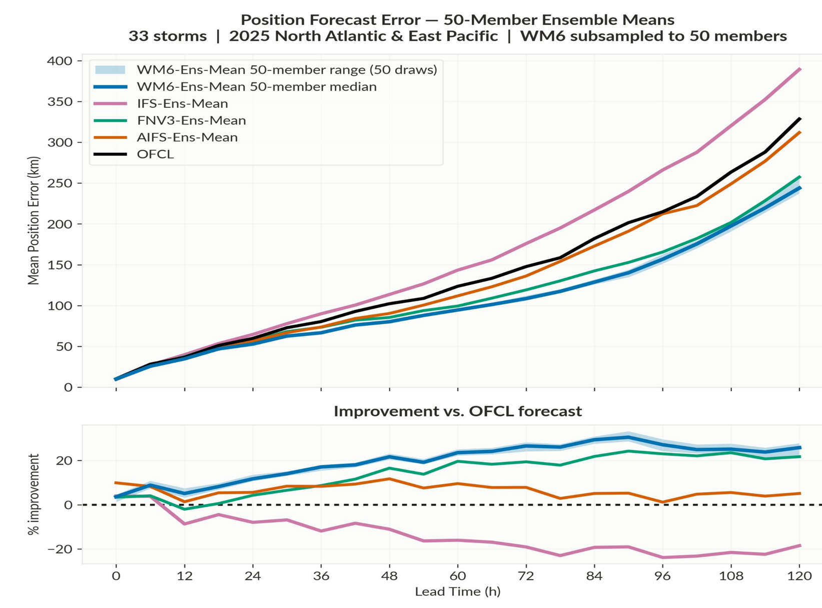


WeatherMesh RESULTS

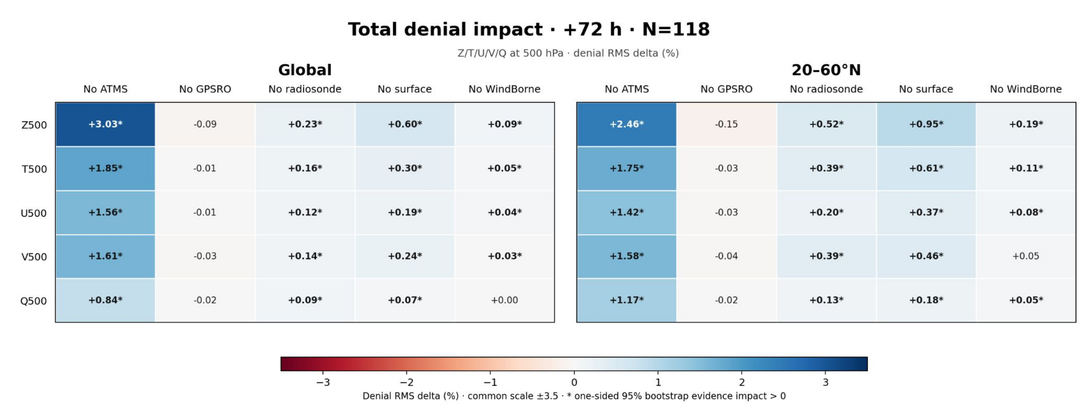
Global statistics. WeatherMesh-6 **outperforms** leading NWP and AIWP models, such as the IFS ensemble mean and the AIFS ensemble mean, for nearly all forecast parameters at all forecast leads. The advantage is particularly striking at short leads, where 2-meter temperature, MSLP, 10-meter wind speeds, and 500 hPa geopotential heights show RMSE reductions of **20-40%** globally.



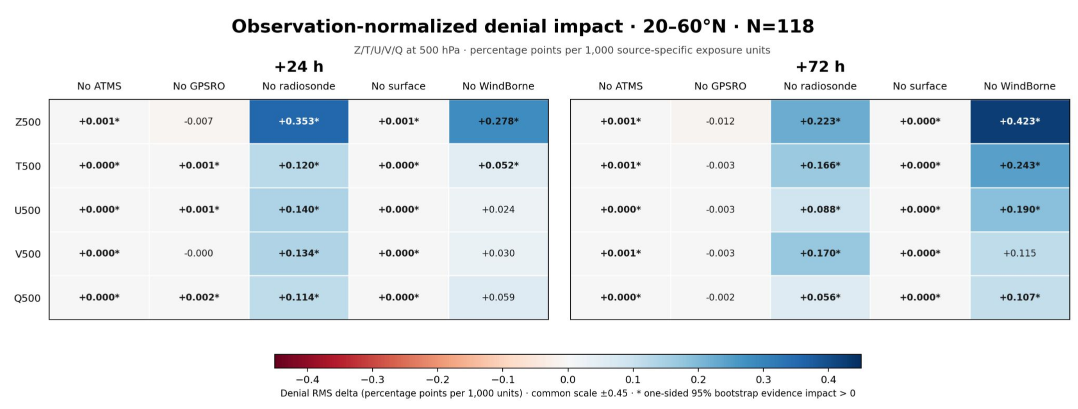
Tropical cyclones. Over the 2025 tropical cyclone season, WM-6 ensemble mean TC track forecasts had **lower errors** than the IFS, AIFS, WeatherNext-2, and the NHC's official forecast. WM-6 does still struggle to adequately capture TC magnitude (not shown), demonstrating low biases in wind speeds and MSLP.



Denial experiments. Fast deep-learning based models enable relatively efficient data denial and scaling experiments. We look at denials in a three month period for various observation sources. In total impact, **microwave sounders dominate**, matching existing FSOI physics-based analyses. WindBorne data has grown enough to show a **measurable positive impact**; however for the period of the analyses it was dwarfed by ~5x by traditional radiosondes in vertical distance covered. On a **per-profile basis** impact looks favorable vs radiosondes, with room to improve as we deploy better targeting.

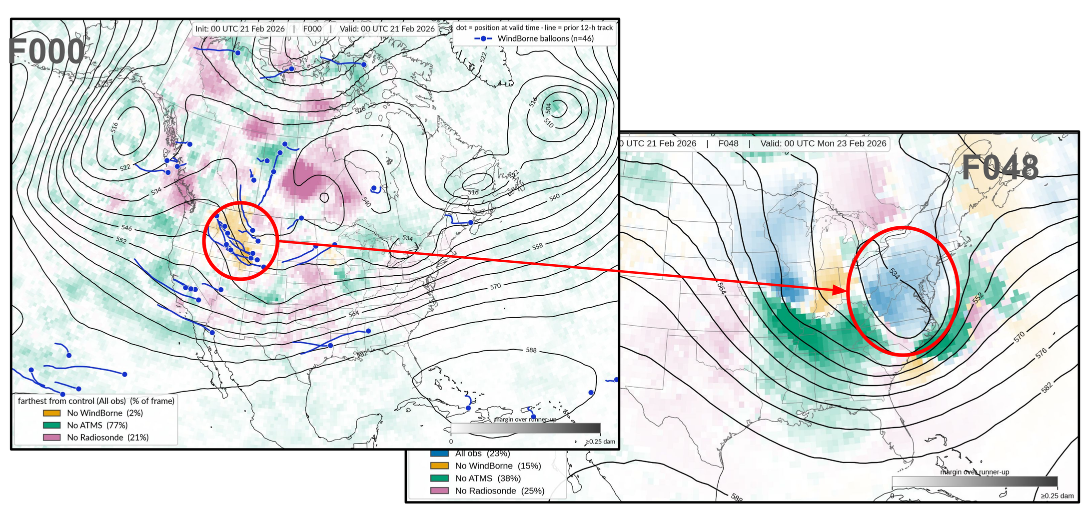


An interesting observation is the **outsized impact of surface observations** relative to our expectations, suggesting deep learning models might be more capable of extracting their full information and correlations. GPS-RO had **negligible-to-negative impact**; likely a limitation of our current encoders and network and should reasonably get closer to physics-based impacts with more work.



Looking at a denial case study rather than in aggregate allows us to hone in on the meteorological evolution of a high-impact snowstorm. A high concentration of WindBorne balloons were assimilated in the vorticity lobe entering the Rocky Mountains. Downstream, this led to better forecasts of the core of the upper-level trough.

Blue → Including WB obs helps (closer to ERA5)
Gold → Including WB obs hurts



WindBorne