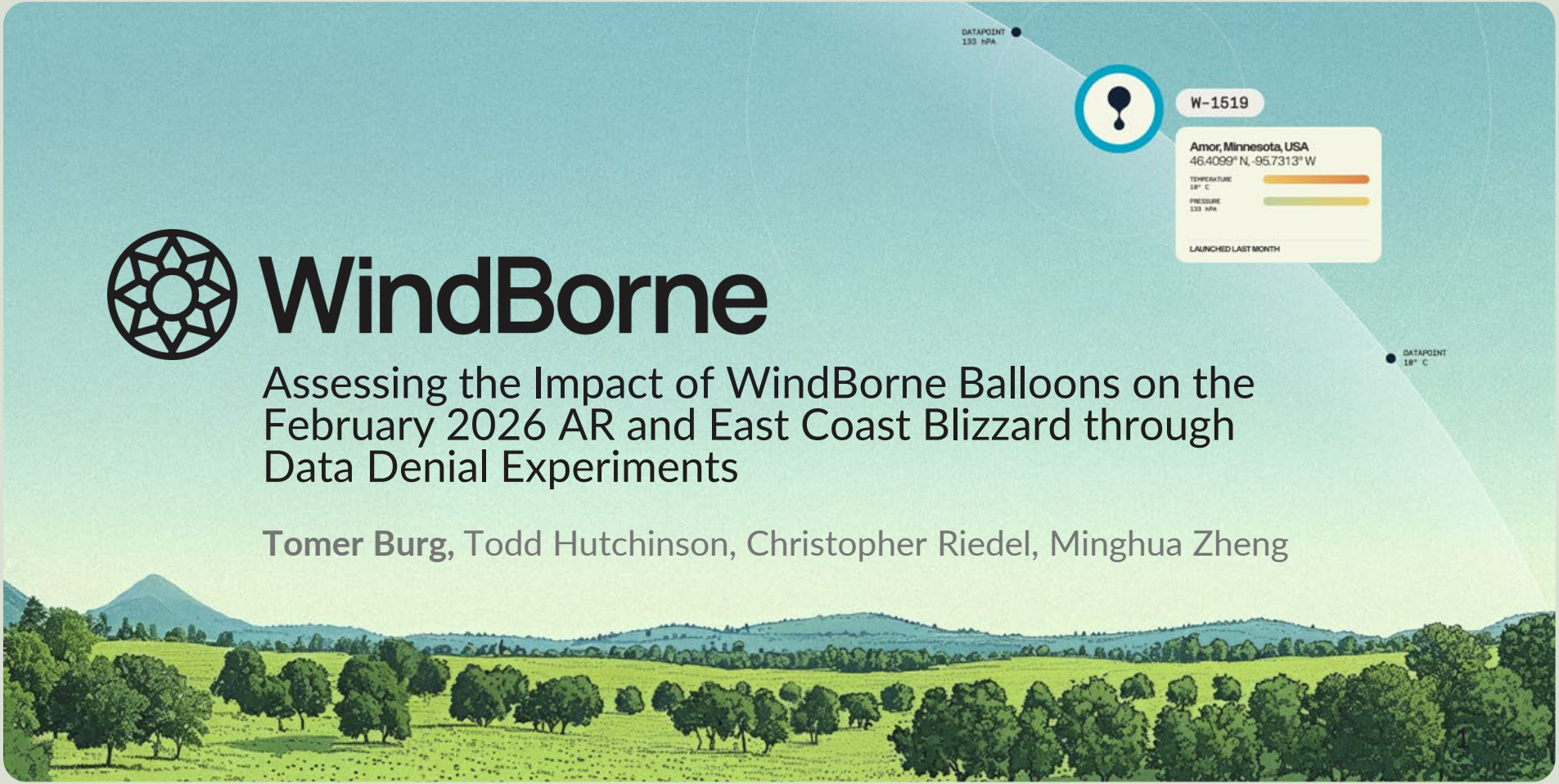
Special thanks to CW3E and NOAA for support



WindBorne

Assessing the Impact of WindBorne Balloons on the February 2026 AR and East Coast Blizzard through Data Denial Experiments

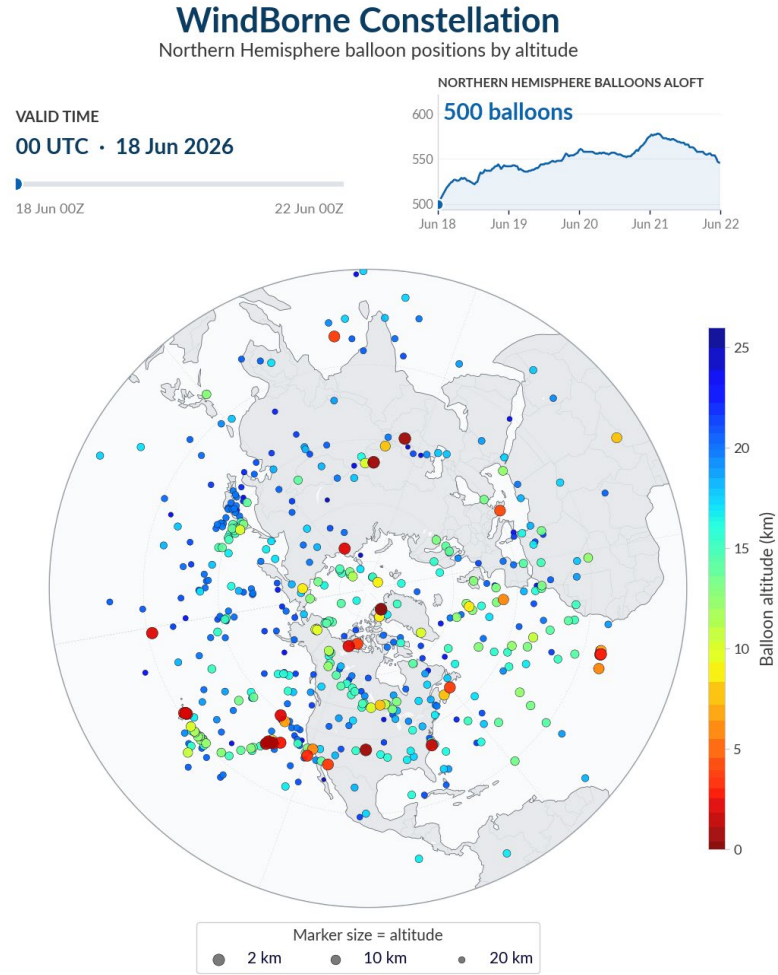
Tomer Burg, Todd Hutchinson, Christopher Riedel, Minghua Zheng

WindBorne Quick Facts

- WindBorne owns, operates, and manufactures the largest balloon constellation in the world
- **18 launch sites** in operation globally
- Rapid growth of observations collected globally, with over **600 simultaneous flights** as of June 2026

WindBorne in AR Recon

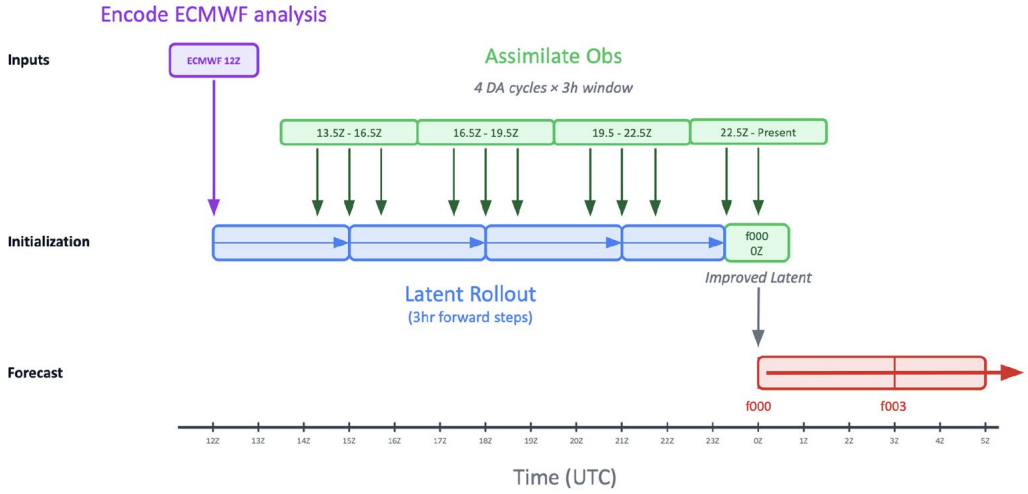
- WindBorne was an active participant in recent AR Recon campaigns
- During the 2025–26 campaign, WindBorne coordinated with AR Recon to collect vertical profiles in relevant areas outside of conventional flight ranges
- WindBorne had **1627 flights aloft in the North Pacific (2001 globally)** during the 2025–26 campaign
- WindBorne also collected **over 14,000 vertical profiles in the North Pacific** (over 36,000 globally)



WeatherMesh (WM) AI Model

- WeatherMesh (WM) is an **“all-in-one”** forecasting system – a transformer-based AI model combining DA and NWP
- Global 0.25° gridded output trained on ECMWF’s ERA5 reanalysis

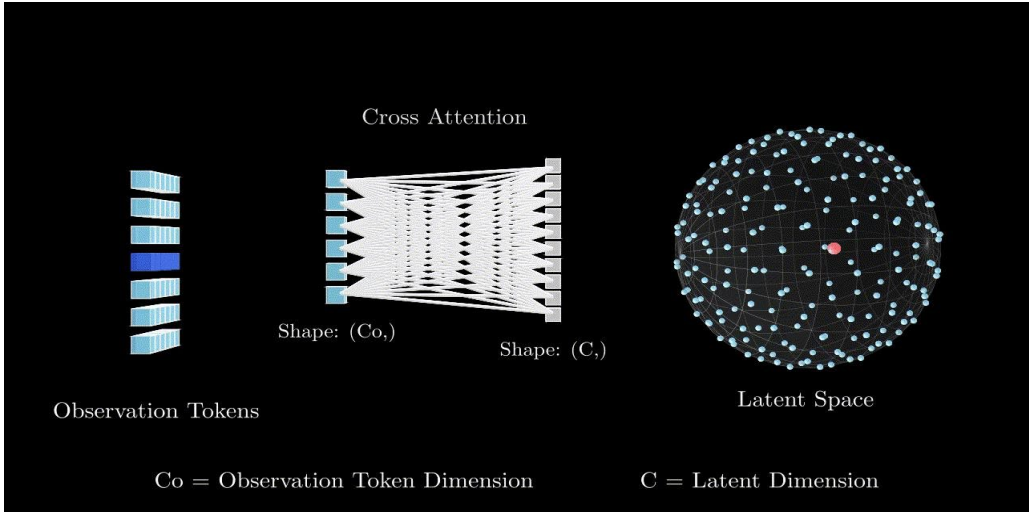
- Background initialized with IFS Ens analysis → iterative DA cycling assimilating observations into the analysis up to the present
- **WM performs DA and generates a 10-day forecast in as little as 5 minutes**



Transformer AI used in the DA component

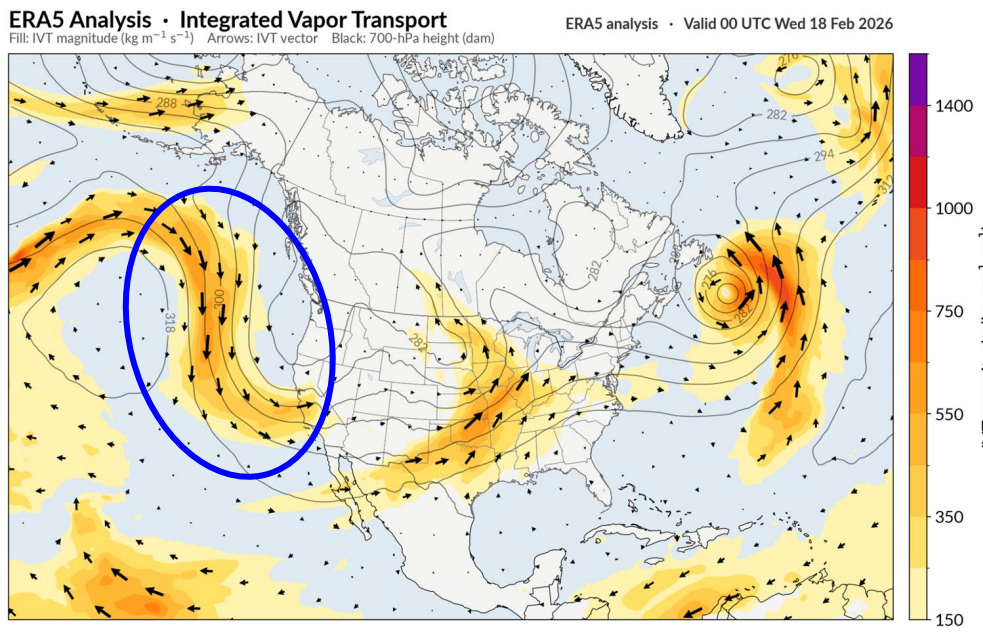
1. Encoding Phase

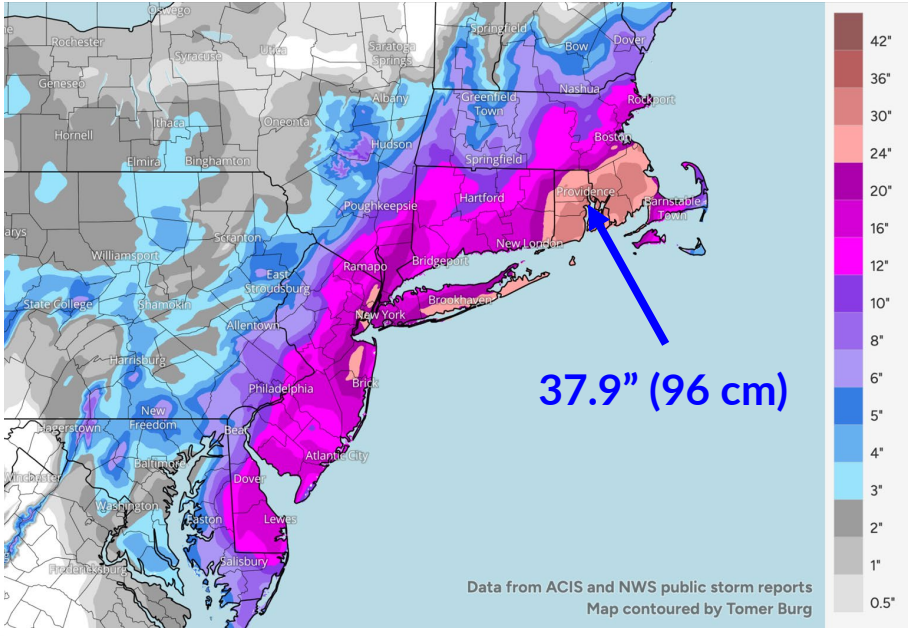
2. Assimilation Phase



February 2026 Blizzard: Background

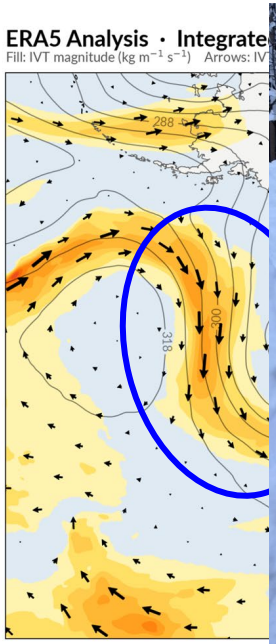
- A significant snowstorm impacted the northeastern United States on **February 22–23, 2026**, with as much as **96 cm (37.9 in)** of snow in Rhode Island
- This blizzard was driven by a complex interaction of numerous mid-upper tropospheric shortwave troughs, including a precursor Atmospheric River (AR) which was a target of AR Recon
- The blizzard was a significant forecast challenge, with many operational models simulating the cyclone too far east or failing to depict rapid cyclogenesis until short lead times

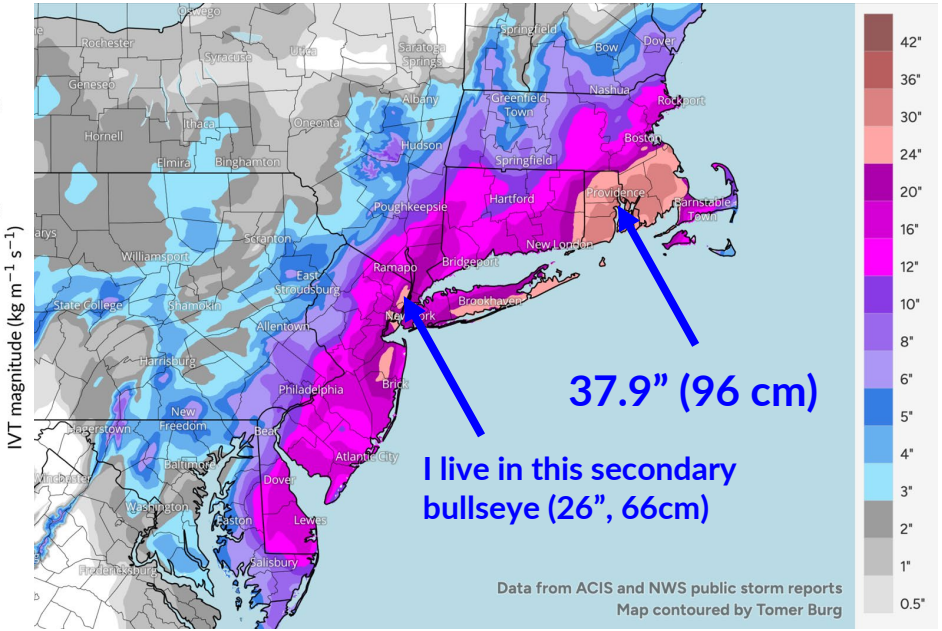




February 2026 Blizzard: Background

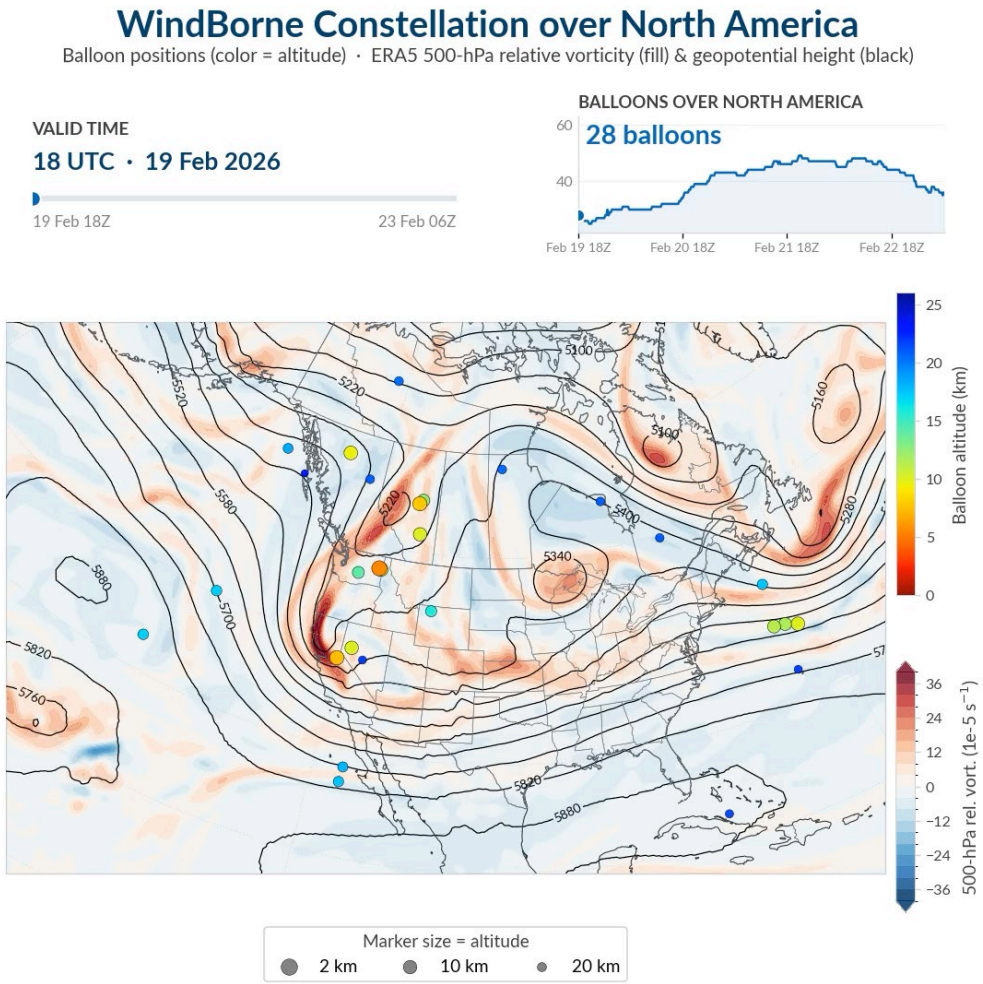
- A significant snowstorm impacted the northeastern United States on **February 22–23, 2026**, with as much as **96 cm (37.9 in)** of snow in Rhode Island
- This blizzard was driven by a complex interaction of numerous mid-upper tropospheric shortwave troughs, including a precursor Atmospheric River (AR) which was a target of AR Recon
- The blizzard was a significant forecast challenge, with many operational models simulating the cyclone too far east or failing to depict rapid cyclogenesis until short lead times





WindBorne collected profiles in dynamically relevant regions

- One of the key drivers of the rapid cyclogenesis was a **vorticity lobe emerging into the Upper Rockies** on 20–21 February
- In coordination with CW3E and NOAA as part of AR Recon, WindBorne collected numerous tropospheric **vertical profiles for balloons located directly within and surrounding this vorticity lobe**
- These observations were subsequently assimilated into the NOAA GFS model in real time



WM6 DA Configuration

- +4 hour observation cutoff for each cycle was selected to be consistent with ECMWF's operational DA window
- Cycling performed every 3 h for an 18 h period preceding the initialization time

All observations assimilated into WM6:

Satellite

- MWHS (MW)
- AMSU-A (MW)
- ATMS (MW)
- AMSR2 (MW)
- IAIS (IR)
- CrIS (IR)
- GPS-RO
- AMVs (cloud-derived winds)

Conventional

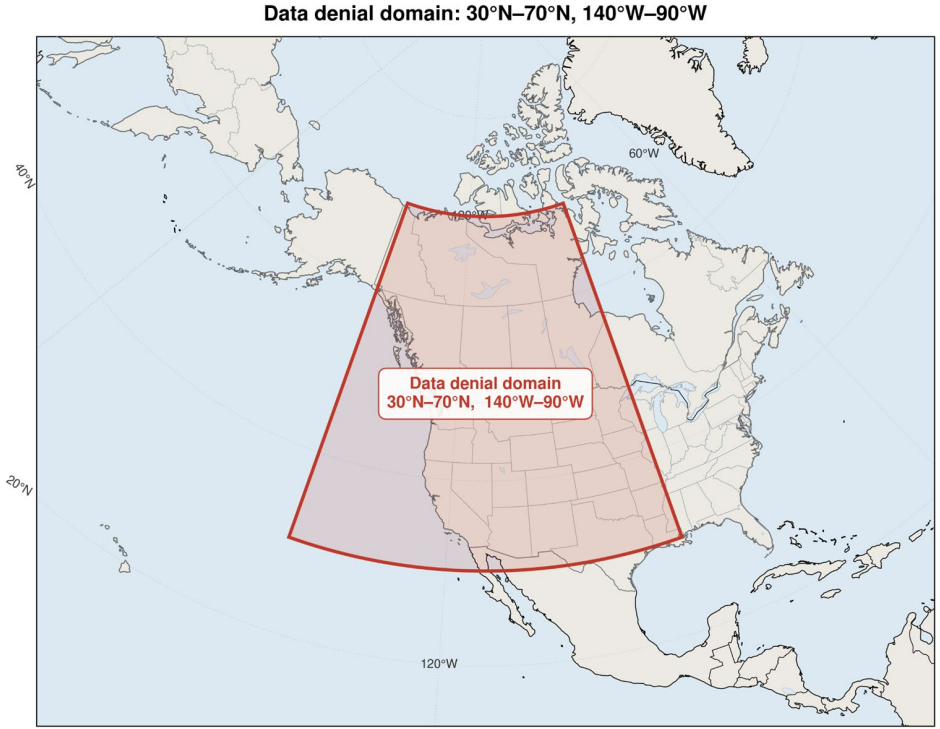
- Surface (METAR, buoys, etc)
- Radiosondes
- WindBorne GSBs

WM6 Run Configuration

- 4 simulations were performed per each cycle:
  - **Control (all data assimilated)**
  - No WindBorne
  - No ATMS (Microwave satellite)
  - No Radiosonde

- Simulations were performed for the following cycles *(lead time prior to peak cyclone deepening rate in parentheses)*:
  - 1800 UTC 20 February *(60 h)*
  - **0000 UTC 21 February *(54 h)***
  - 0600 UTC 21 February *(48 h)*

- Plots shown in these results are the ensemble mean of a 50-member ensemble



WM6 Run Configuration

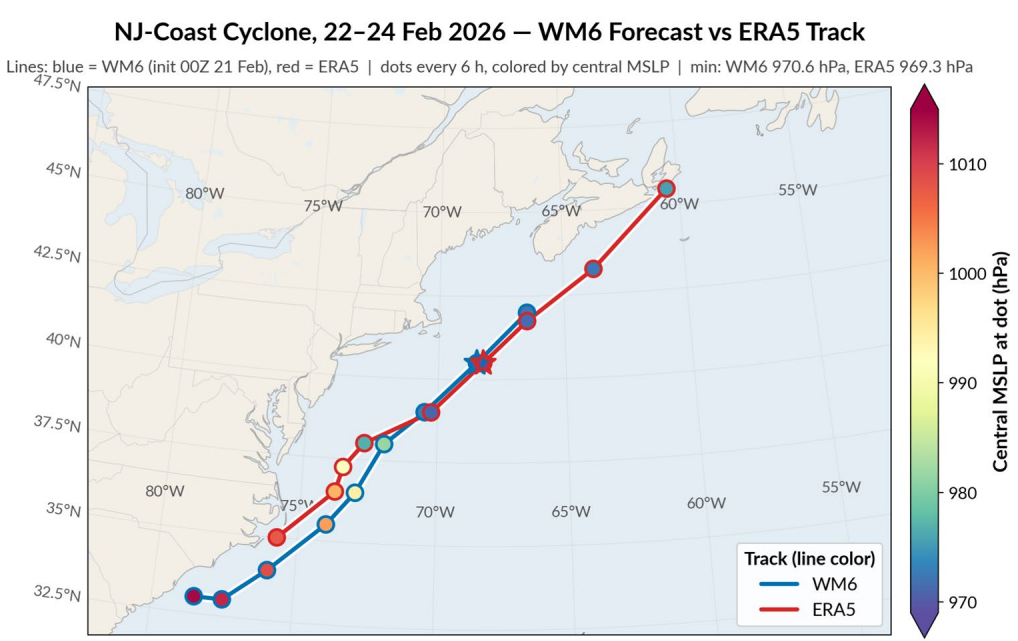
- 4 simulations were performed per each cycle:
  - **Control (all data assimilated)**
  - No WindBorne
  - No ATMS (Microwave satellite)
  - No Radiosonde

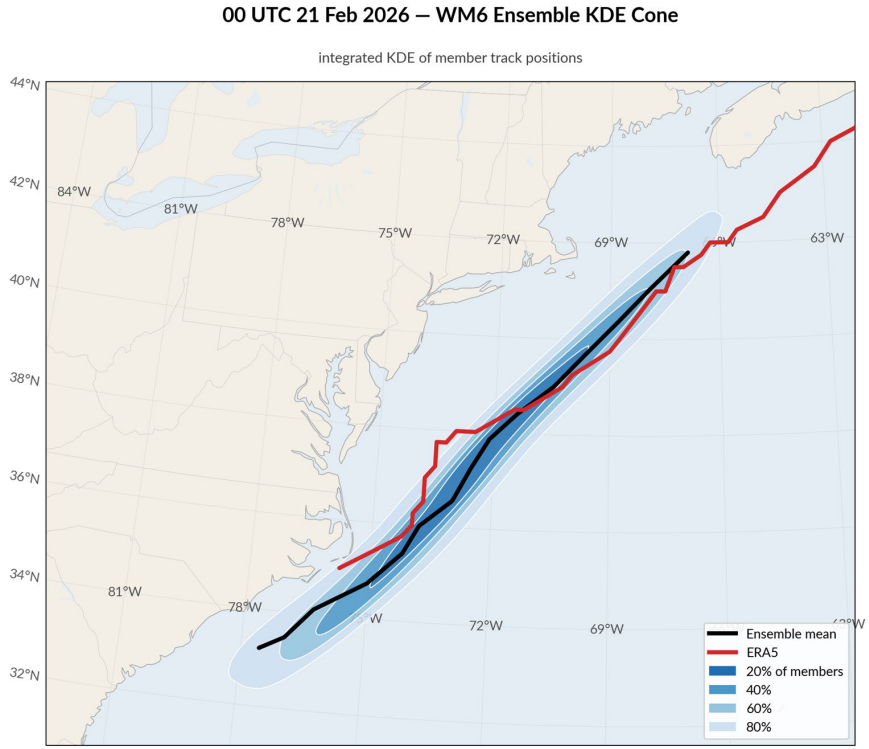
- Simulations were performed for the following cycles *(lead time prior to peak cyclone deepening rate in parentheses)*:
  - 1800 UTC 20 February *(60 h)*
  - **0000 UTC 21 February *(54 h)***
  - 0600 UTC 21 February *(48 h)*

- Plots shown in these results are the ensemble mean of a 50-member ensemble

Results : Control Simulation

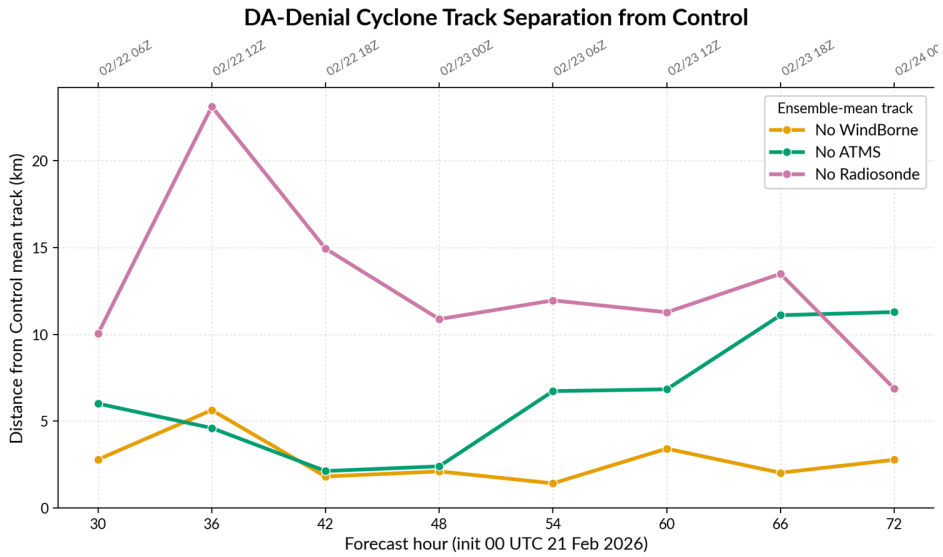
- For the initialization cycles of interest, the control WM6 simulation **represented the timing, peak intensity, and location of the cyclone well**, but was slightly too weak and southeast during peak deepening rates
- Ensemble spread was generally tightly clustered close to the ensemble mean

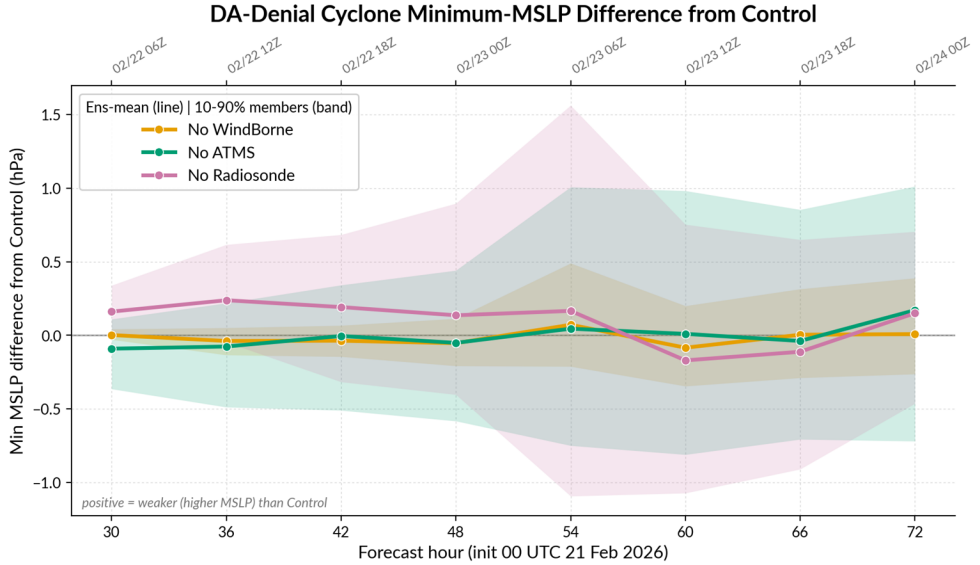




Results : Multi-Simulation Comparison

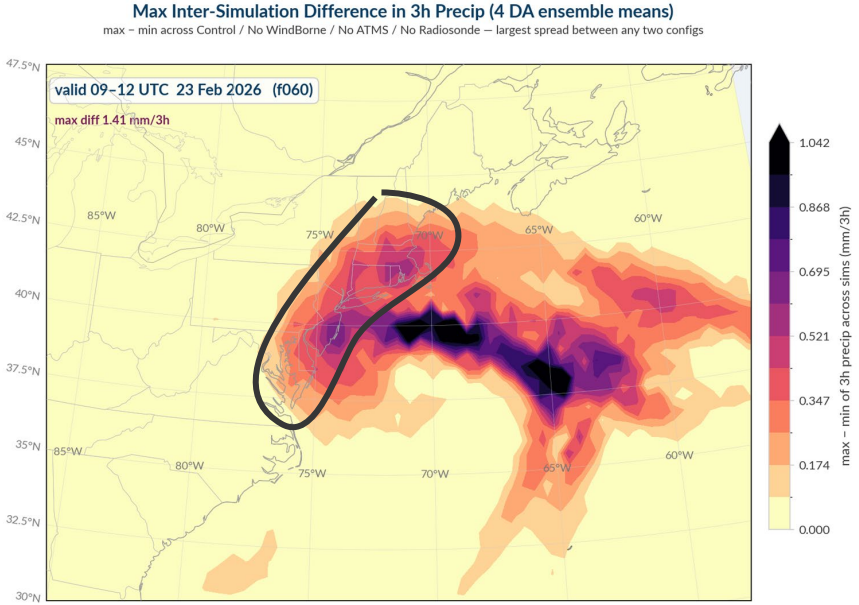
- Overall differences vs. control simulation from denying the three observation sources on track and intensity are **small (<=10–20 km, <=0.15 hPa)** for the ensemble means – up to 60–70 km for individual members
- These primarily result in sensible impacts to a narrow corridor of locations near sharp precipitation gradients at the north/western edge of heavy snowfall banding
- Denying radiosonde observations generally had the largest impact on cyclone track and intensity





Results : Multi-Simulation Comparison

- Overall differences vs. control simulation from denying the three observation sources on track and intensity are **small (<=10–20 km, <=0.15 hPa)** for the ensemble means – up to 60–70 km for individual members
- These primarily result in sensible impacts to a narrow corridor of locations near sharp precipitation gradients at the north/western edge of heavy snowfall banding
- Denying radiosonde observations generally had the largest impact on cyclone track and intensity



← Maximum difference between the 4 ensemble means along the major population corridor is on the order of 0.01–0.015” (0.25—0.4mm) per 3 h

Near negligible verbatim, but larger in practice given inherent failure of global AI models to resolve mesoscale precipitation structures and gradients

When comparing amongst individual members, this difference per gridpoint is as large as 0.15” (4mm) per 3 h

Results : DA Impact on Analysis

Impact of denying WB obs

What the plot shows:

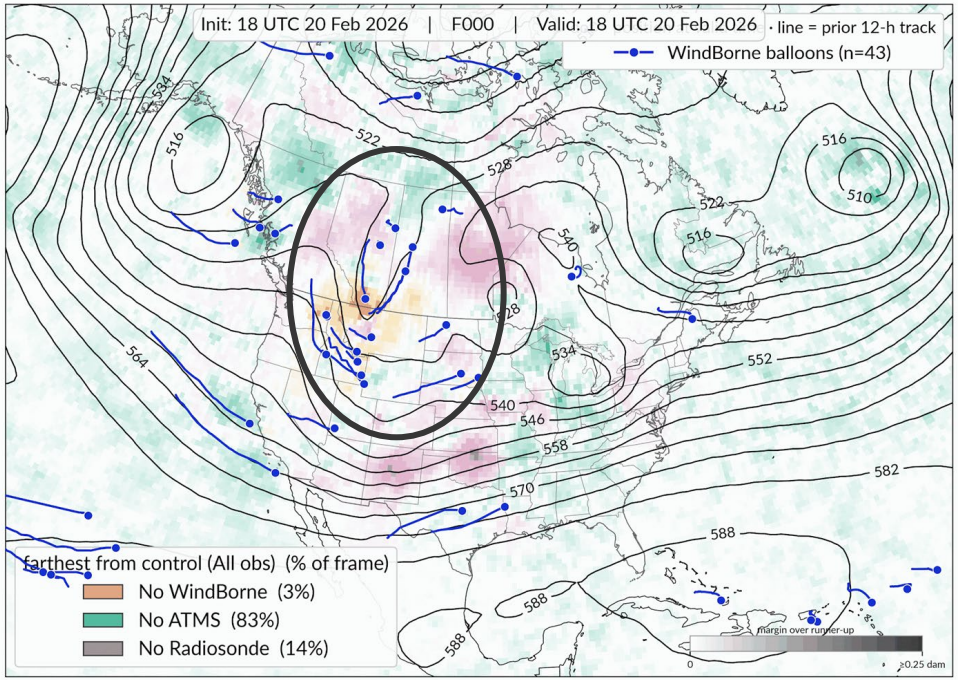
- **Variable**: 500-hPa geopotential height, WindBorne balloons in blue
- **Color**: which simulation 0h analysis was most different from the control 0h analysis
- **Intensity of color**: difference between farthest and 2nd farthest simulation

Analysis:

- Denying WB data had the biggest impact on the 500-hPa height analysis field within and surrounding the **vorticity lobe entering the Northern Rockies** and its **northern tail over Saskatchewan**
- This localized impact was confined to the greatest clustering of balloons surrounding both areas
- Clustering of WB balloons elsewhere had less impact than radiosondes & ATMS

500-hPa geopotential height: which run is farthest from control? (3 denial runs)

fill = farthest-from-control run; pale = narrow margin over 2nd, dark = wide margin | black contours = control (All obs)



Results : DA Impact on Analysis

WM6 analysis vs. ERA5 analysis

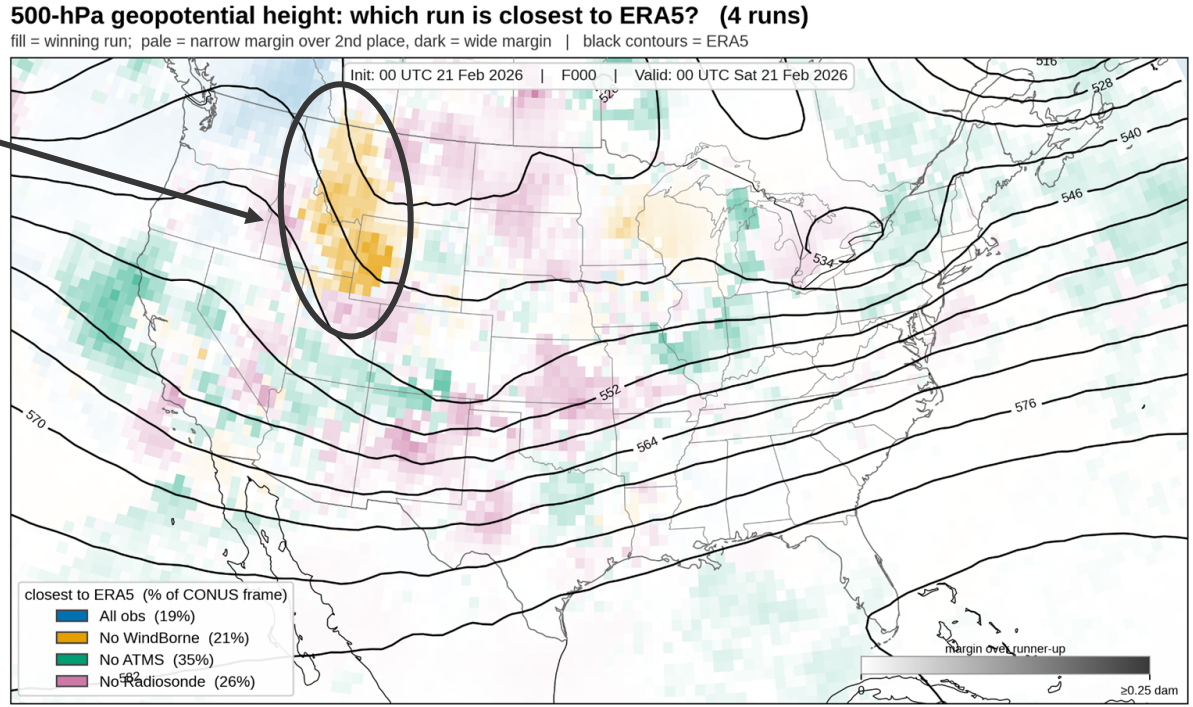
What the plot shows:

- **Variable**: 500-hPa geopotential height
- **Color**: which simulation 0h analysis was closest to ERA5 analysis
- **Intensity of color**: difference between closest and 2nd closest simulation

Analysis:

- **Denying** WindBorne data led to the WM6 0h analysis being **closest** to the ERA5 analysis, vs. denying other obs sources or even vs. control simulation

- **But… is this an expected result, given that ERA5 doesn't assimilate WindBorne observations?**



Results : 0000 UTC 21 February Simulation

Impact of denying WB obs

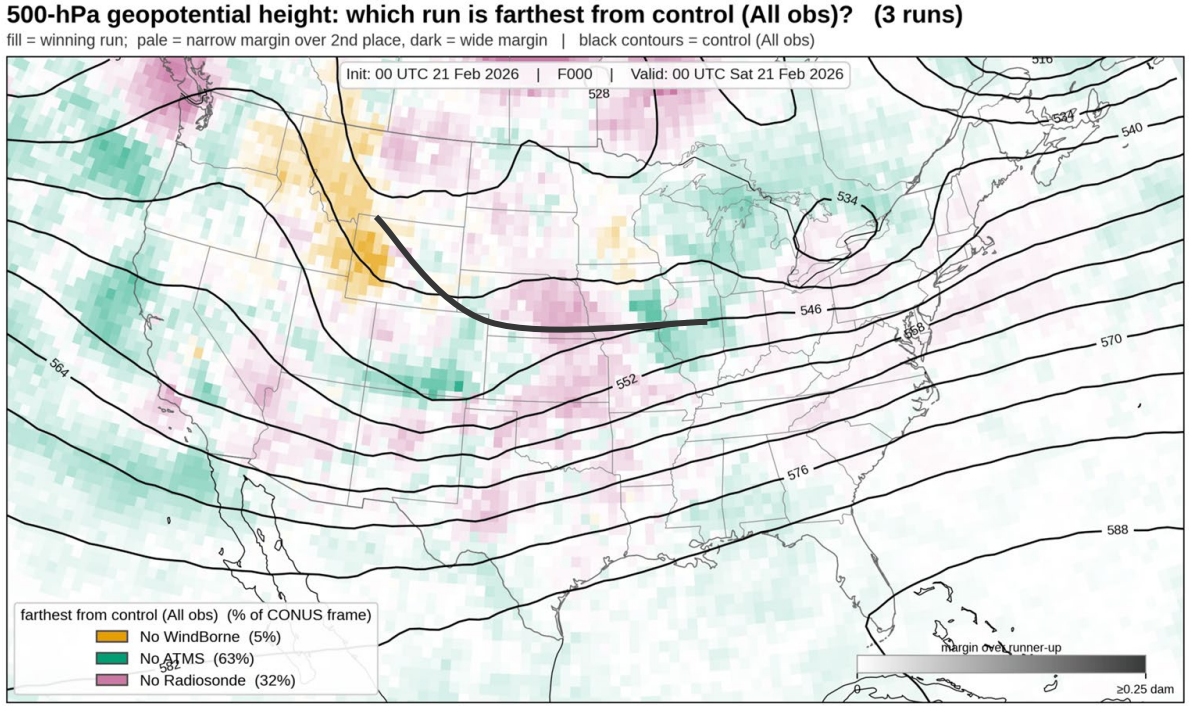
What the plot shows:

- **Variable**: 500-hPa geopotential height
- **Color**: which simulation was most different from the control simulation
- **Intensity of color**: difference between farthest and 2nd farthest simulation from control

Analysis:

- The area of greatest impact from denying WB obs, while spatially small, propagates downstream with the shortwave trough of interest

- By 30–36 h, denying WB obs leads to the largest differences vs. control **within the core of the upper-level trough** and the **amplitude of the downstream shortwave ridge**



Results : 0000 UTC 21 February Simulation

Best simulation vs. ERA5

What the plot shows:

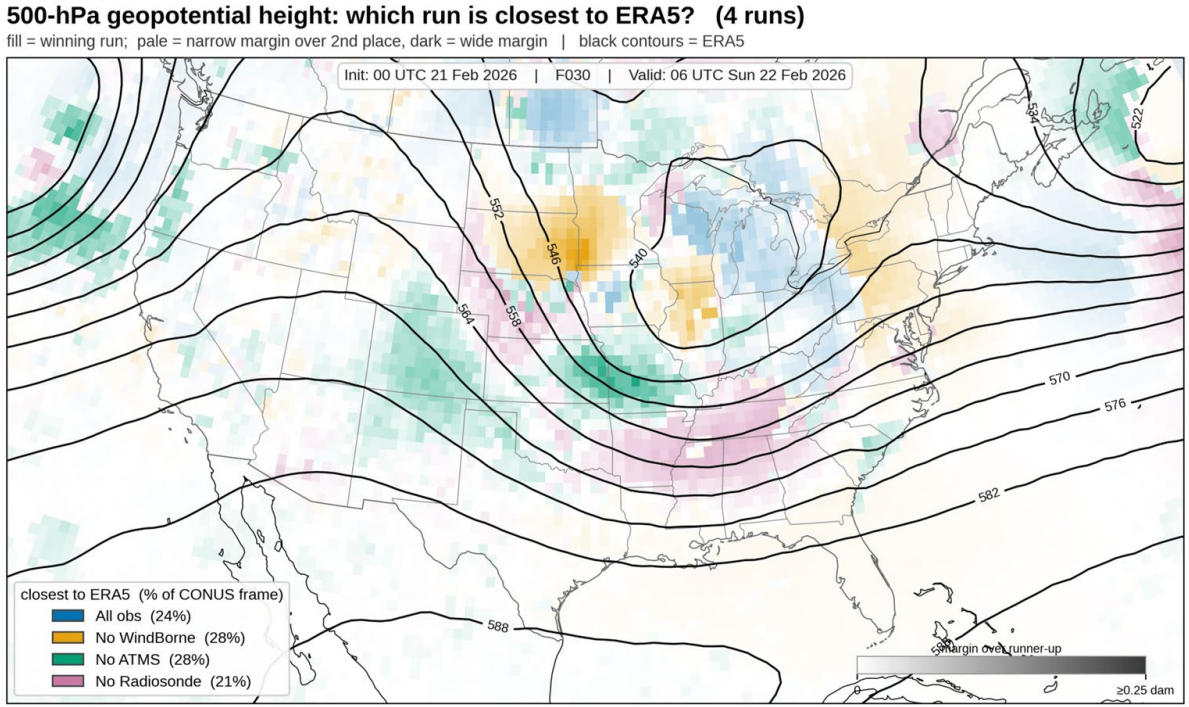
- **Variable**: 500-hPa geopotential height
- **Color**: which simulation was closest to ERA5 analysis
- **Intensity of color**: difference between closest and 2nd closest simulation from ERA5

Analysis:

- **Blue = desired** (all obs + WB leads to closest outcome to ERA5)
- **Gold = not desired** (denying WB leads to closer outcome to ERA5 than with WB)

WB obs have the most beneficial impact within:

- (1) the core of the upper-level trough as it becomes negatively tilted
- (2) timing of trailing shortwave trough



Results : 0000 UTC 21 February Simulation

Worst simulation vs. ERA5

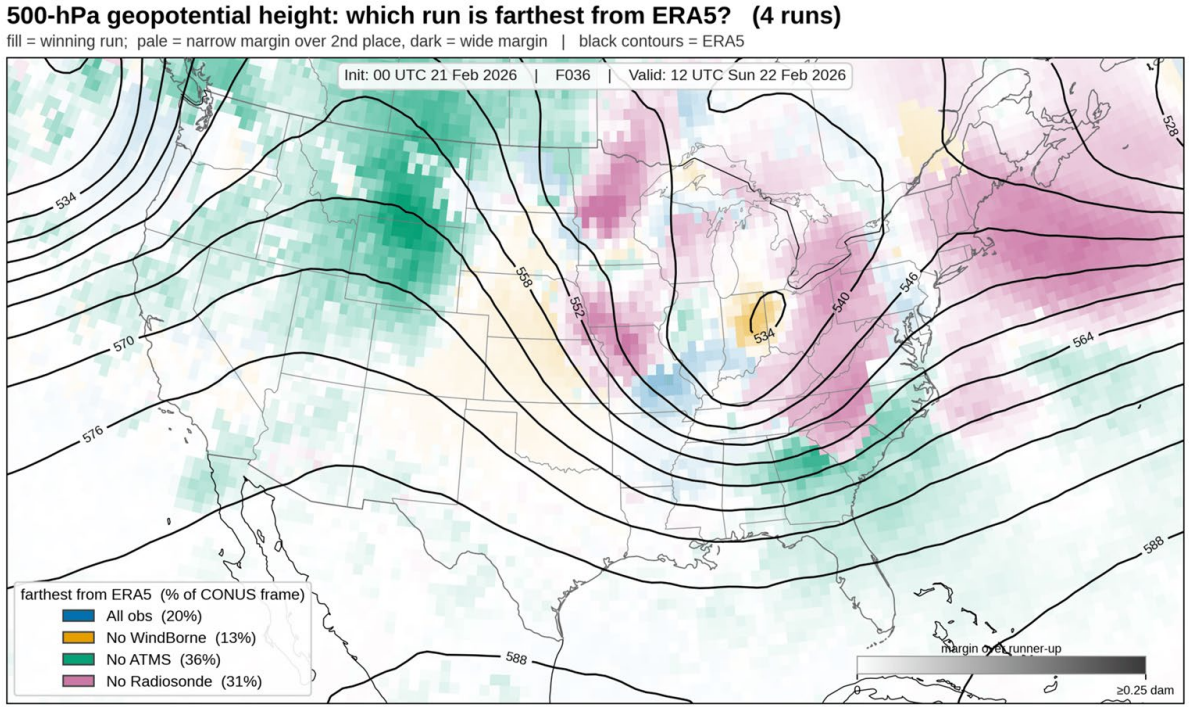
What the plot shows:

- **Variable**: 500-hPa geopotential height
- **Color**: which simulation was farthest to ERA5 analysis
- **Intensity of color**: difference between farthest and 2nd farthest simulation from ERA5

Analysis:

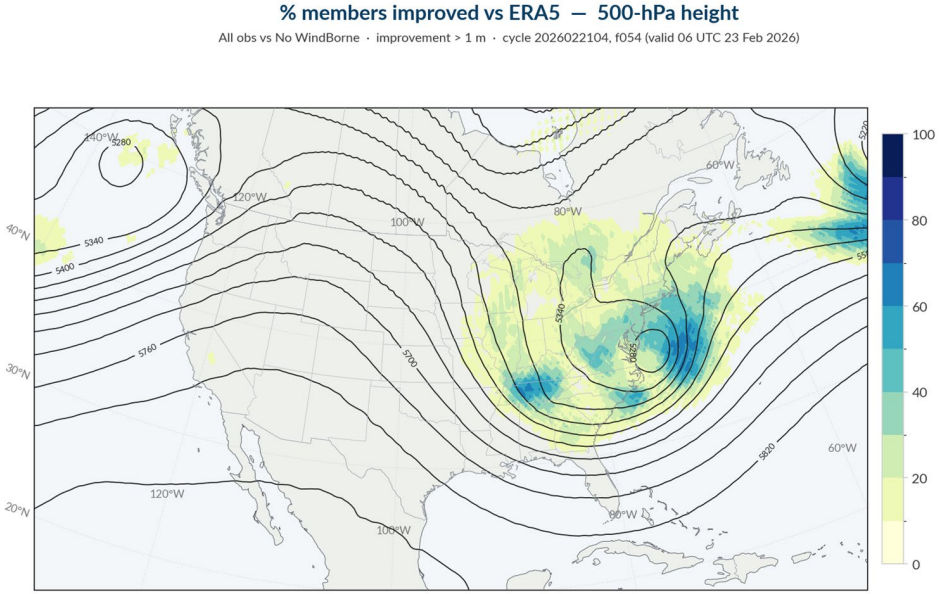
- **Gold = denying WB leads to greatest errors vs. ERA5 than assimilating WB or denying other observations**

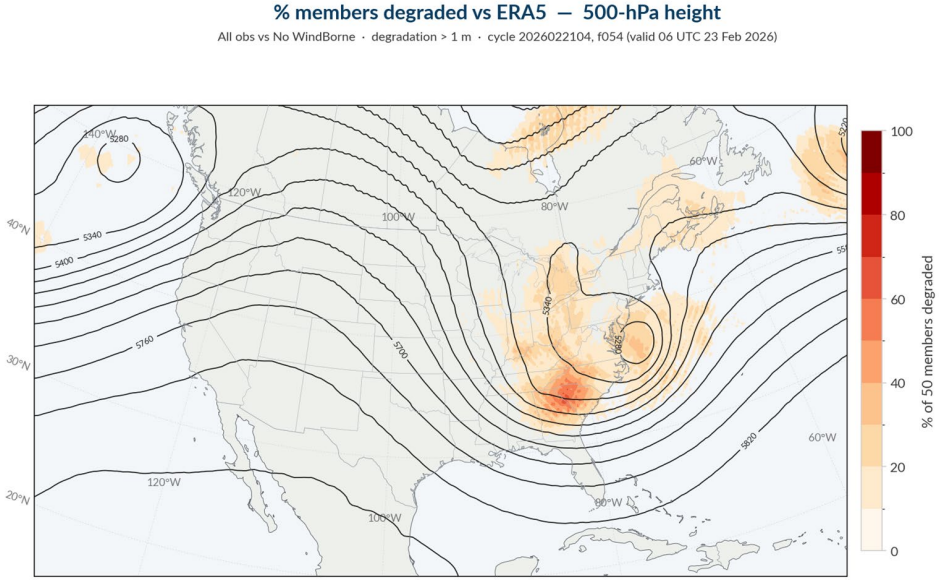
- Greatest negative impact from denying WB obs starts near the core of the upper low & back end of the longwave trough
- Impact grows in scale as the cyclone matures



DA Ensemble – How many members improved vs. degraded the forecast?

- On average, assimilating WB obs led to more members exhibiting increased vs. decreased errors compared against ERA5

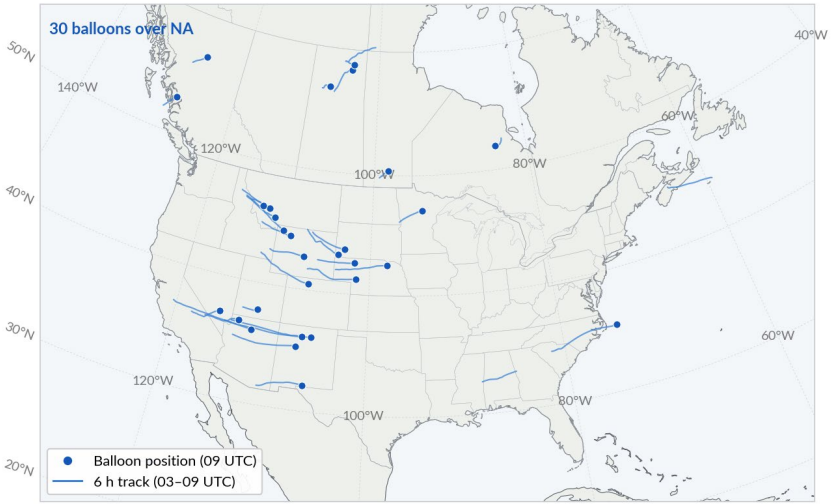




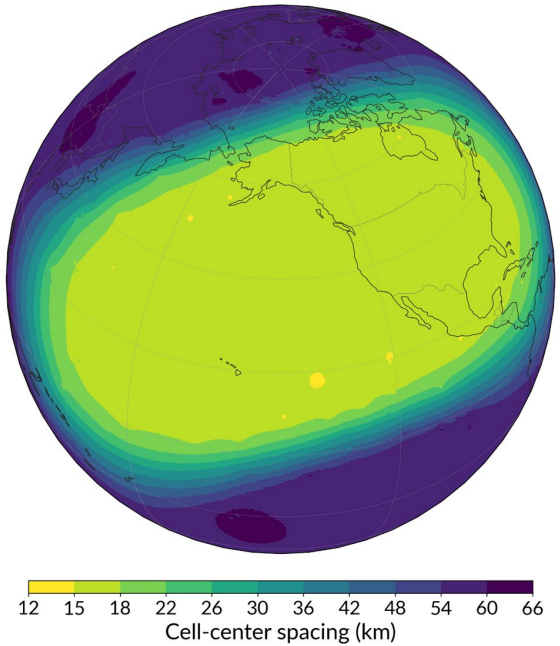
MPAS Run Configuration

- WindBorne observations were used in MPAS-JEDI DA
- 0600 UTC 21 February 2026 initialization
- +/- 3 h assimilation window → observations between 03 and 09 UTC 21 Feb assimilated

03-09 UTC 21 February 2026 WindBorne Balloon Tracks



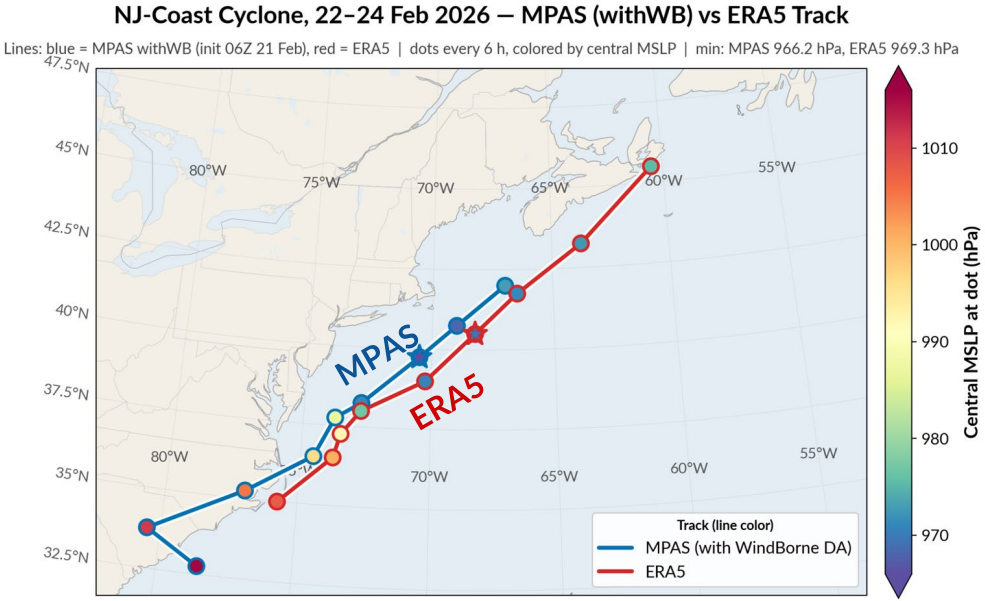
MPAS mesh domain & horizontal resolution
variable-resolution global mesh · ~16 km refined → ~62 km

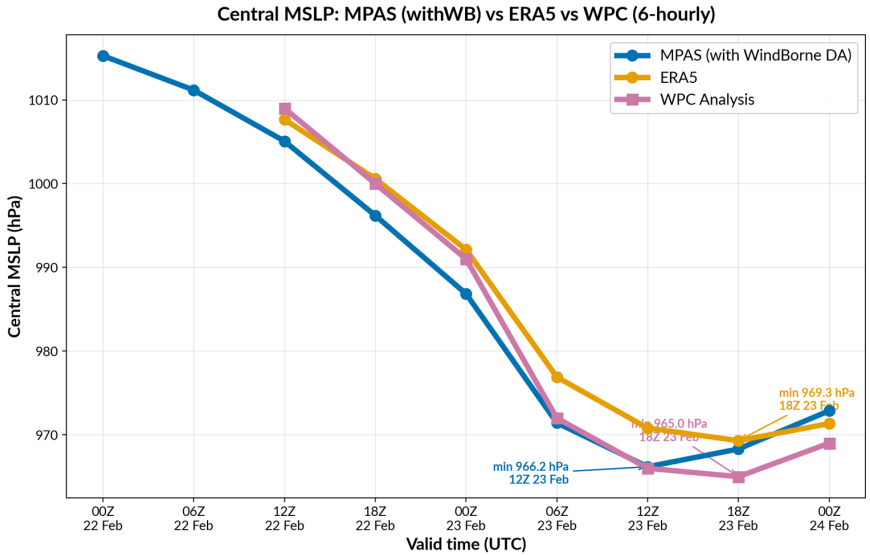


Special thanks to Minghua Zheng of CW3E for performing these simulations

Results : MPAS Control

- MPAS simulation with WindBorne data assimilated generally represented the synoptic-scale cyclone well, with a better representation of the surface cyclone deviating northwest during peak deepening than WM6
- MPAS exhibited lower minimum and slightly greater peak deepening rates MSLP than ERA5 analysis… **but ERA5 analysis was weaker and slower to deepen the cyclone than WPC operational surface analyses**
- Both analysis datasets agree MPAS was **too quick to decay the cyclone post-peak intensity**



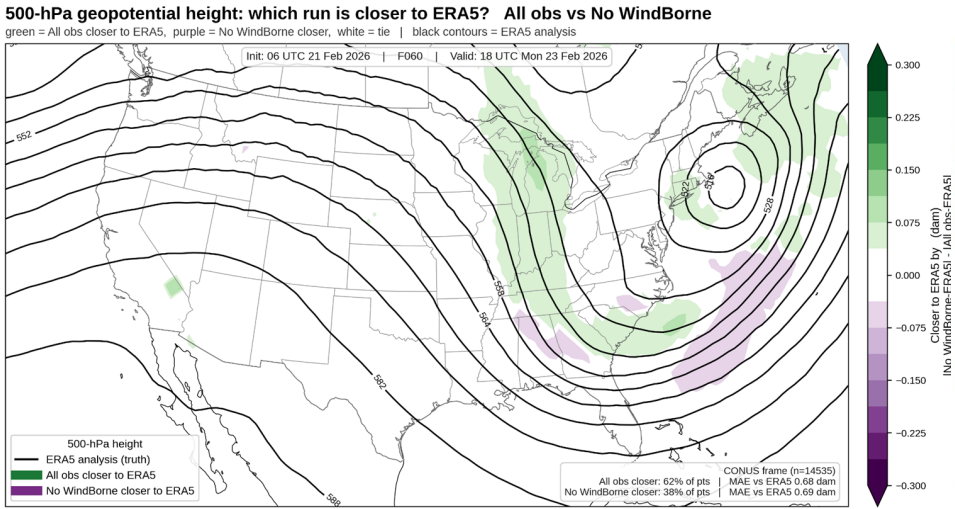


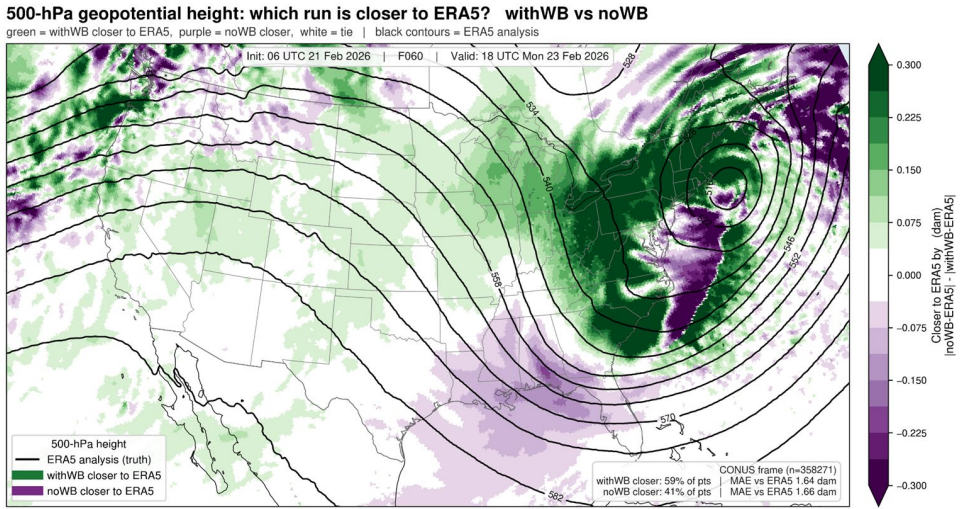
Results : MPAS vs. WM6 Experiments

- MPAS vs. WM6 DA experiments show a noteworthy but not unexpected finding: **Data denial for an AI ensemble mean has a smaller impact than data denial for deterministic higher-resolution NWP models at short lead times**
- Impacts of denying WB obs are generally spatially consistent between MPAS and WM6, but consistently exhibit a larger magnitude and larger spatial coverage in MPAS than WM6

Green = assimilating WB closer to ERA5

Purple = denying WB closer to ERA5





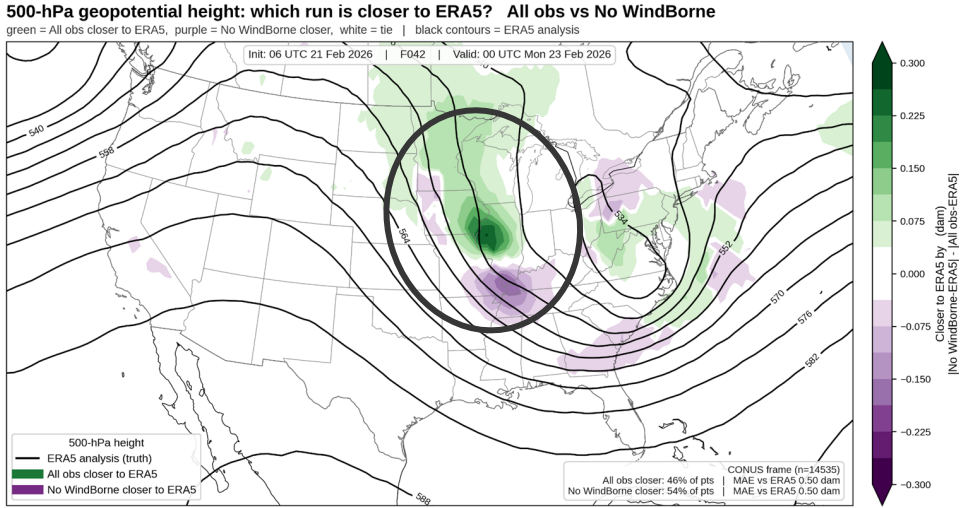
Note: Color scales are equal between both plots, -0.3 dam to +0.3 dam

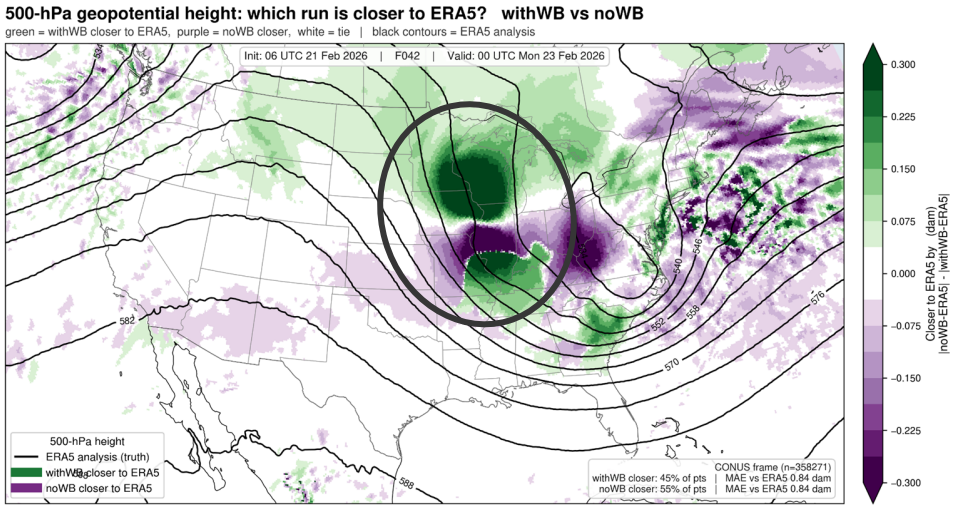
Results : MPAS vs. WM6 Experiments

- Recall that we earlier noted denying WB obs also had the largest impact on the analysis over the **northern tail of the vorticity lobe over Manitoba**
- This tail evolved into an embedded shortwave trough on the back side of the longwave trough – the impact of denying WB obs manifested as timing differences, **consistent between WM6 and MPAS denial simulations**

Green = assimilating WB closer to ERA5

Purple = denying WB closer to ERA5





Note: Color scales are equal between both plots, -0.3 dam to +0.3 dam

Impact of Results

- The magnitude of differences between WM6 data denial simulations was **generally very small – on the sub-synoptic scale** – with sensible impacts mostly limited to areas near sharp precipitation gradients on the north/west end of the heavy snowfall corridor

- Nonetheless, there are useful takeaways from these results:

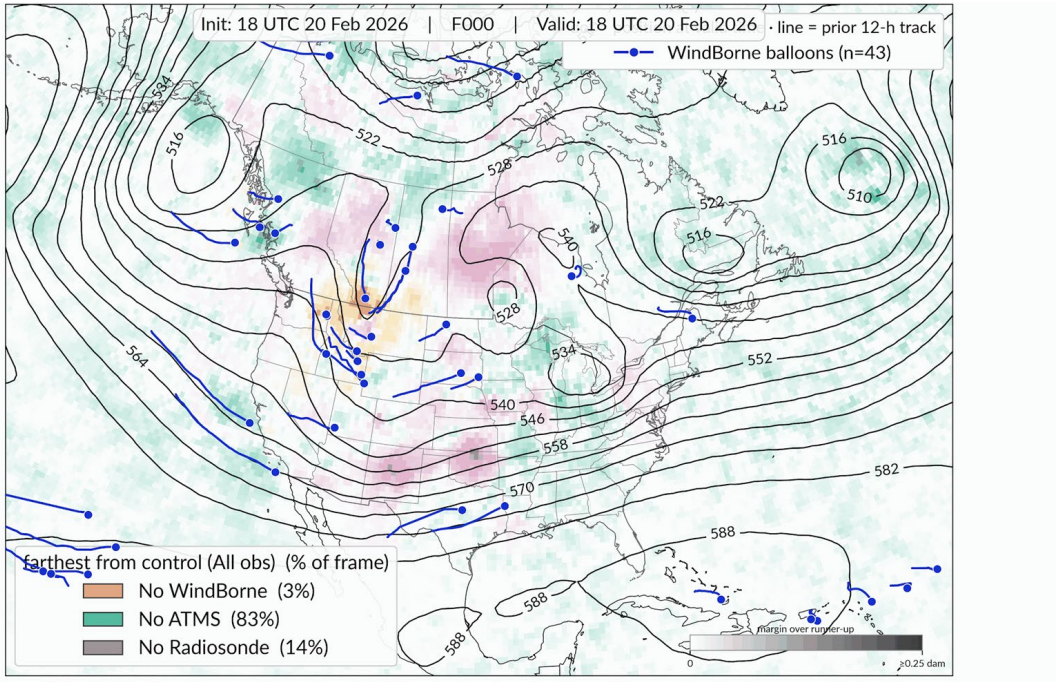
    - AI Model data denial experiments can be performed very quickly, much faster than NWP-only DA experiments

    - Despite a very limited number of WindBorne observations, clustered observations near dynamically relevant features can have a comparable impact on the development of associated synoptic-scale cyclones to other conventional observation sources such as satellites or radiosondes

    - Data denial experiments applied to AI models in this case have generally similar results to NWP data denial experiments, despite the impacts being smaller in magnitude in the short range

Key Takeaway Points

- High-impact blizzard impacted the Northeast U.S. in late February 2026, associated with a precursor AR
- Clustered WindBorne vertical profiles had a larger impact on the analysis fields near a relevant shortwave trough than radiosondes and ATMS satellite data
- This impact propagated downstream with slight forecast skill improvements vs. ERA5 with the longwave trough evolution through assimilating WB
- WindBorne obs in MPAS-JEDI had a greater impact than in AI ensemble mean



Questions?

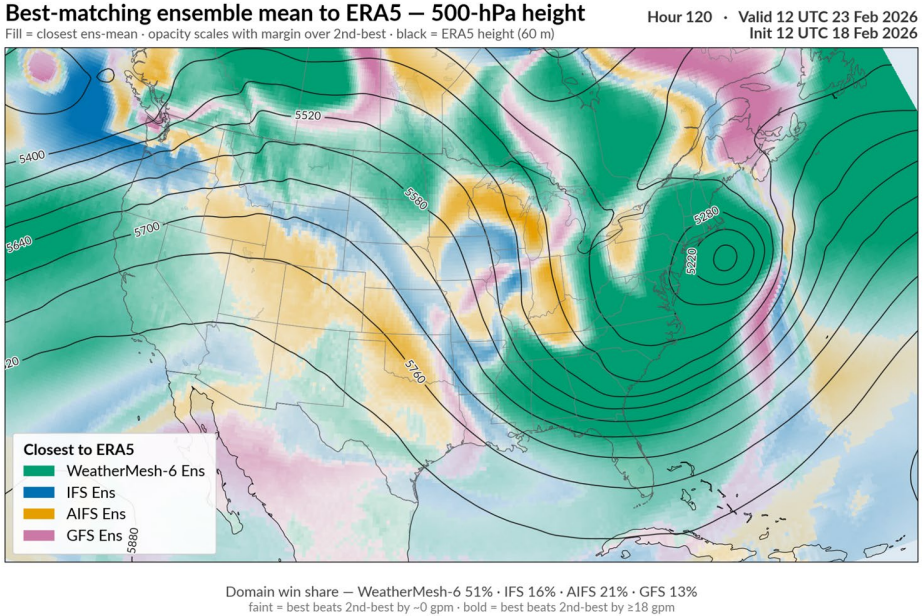
tomer@windbornesystems.com

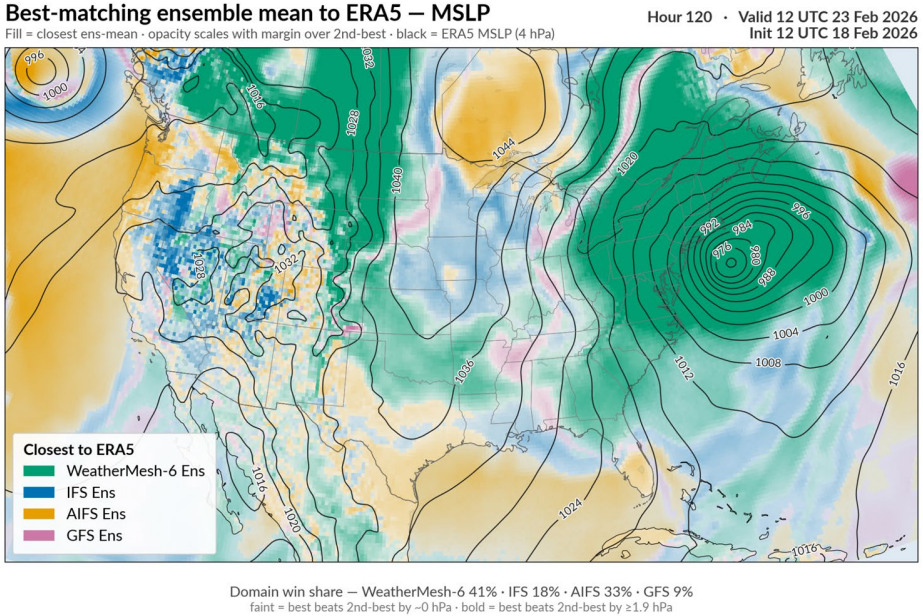
Extra Slides

WeatherMesh 120 h Forecast

- At medium-range lead times, the WM ensemble mean was the best-performing ensemble for the cyclone's structure, track and amplitude among the leading operational ensemble suites

- ...unfortunately, WindBorne had few balloons in areas of large ensemble sensitivity or significant synoptic-scale features of interest





Results : 0000 UTC 21 February Simulation

Worst simulation vs. ERA5

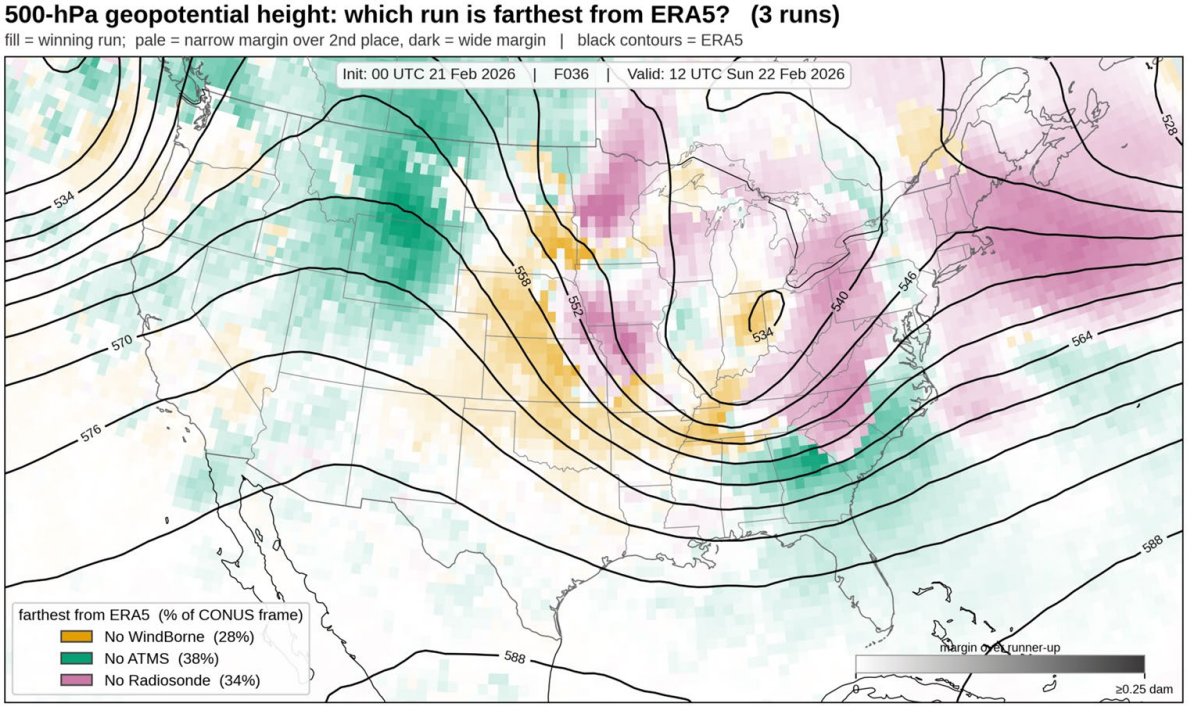
What the plot shows:

- **Variable**: 500-hPa geopotential height
- **Color**: which simulation was farthest to ERA5 analysis **(excluding control)**
- **Intensity of color**: difference between farthest and 2nd farthest simulation from ERA5

Analysis:

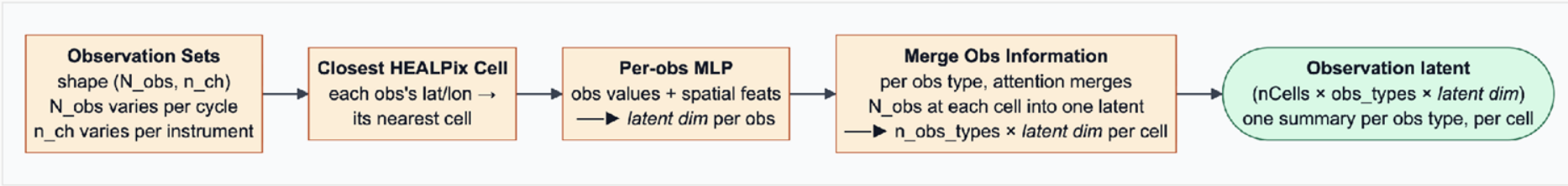
- **Gold = denying WB leads to greatest errors vs. ERA5 than denying other observations**

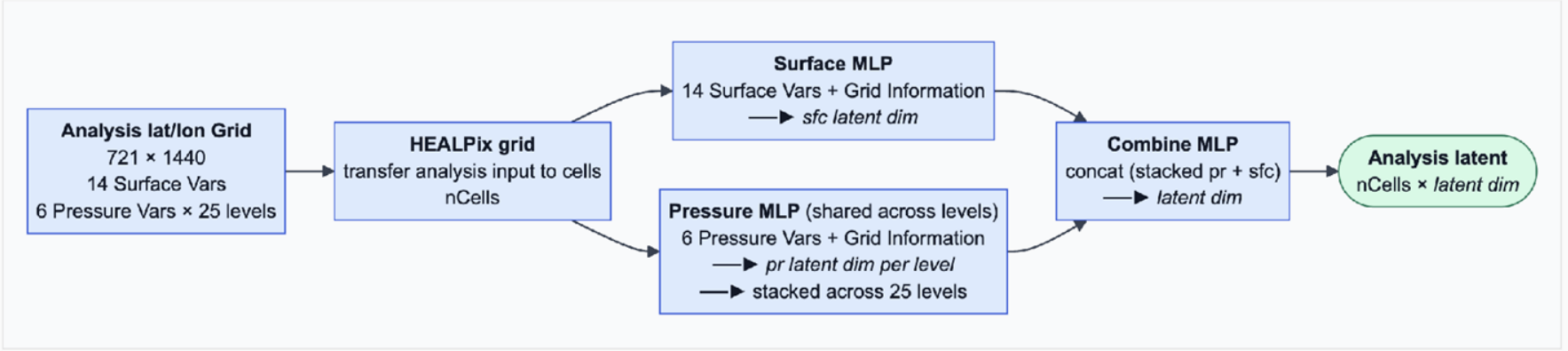
- Solely comparing the 3 data denial simulations, denying WB has the most negative impact on forecast skill in the backside of the longwave trough *(associated with timing and amplitude differences)*



WeatherMesh (WM) AI Model

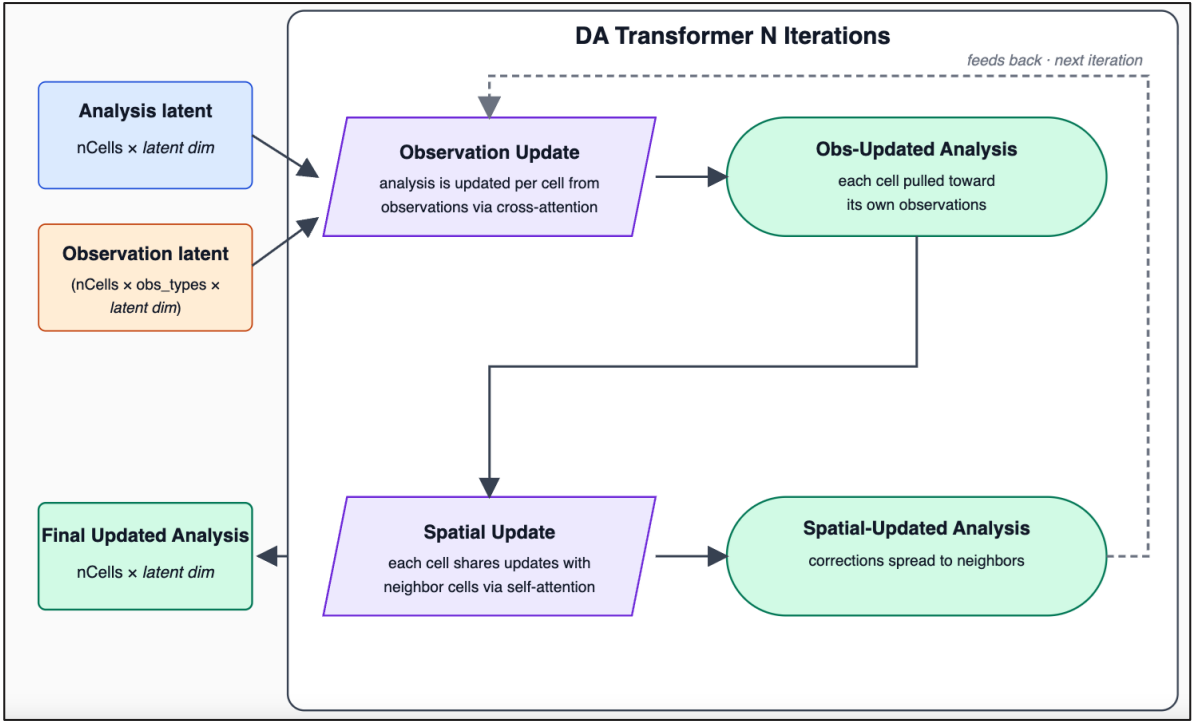
1. Encoding Phase: Both analysis and observations encoded into a shared latent space, so DA transformer can update analysis from observations





WeatherMesh (WM) AI Model

2. Assimilation Phase: Observations update analysis in the shared latent space through different types of attention in DA transformer's iterative loop.



Example Spatial Updates