AirFuse

A multi-pollutant fusion system

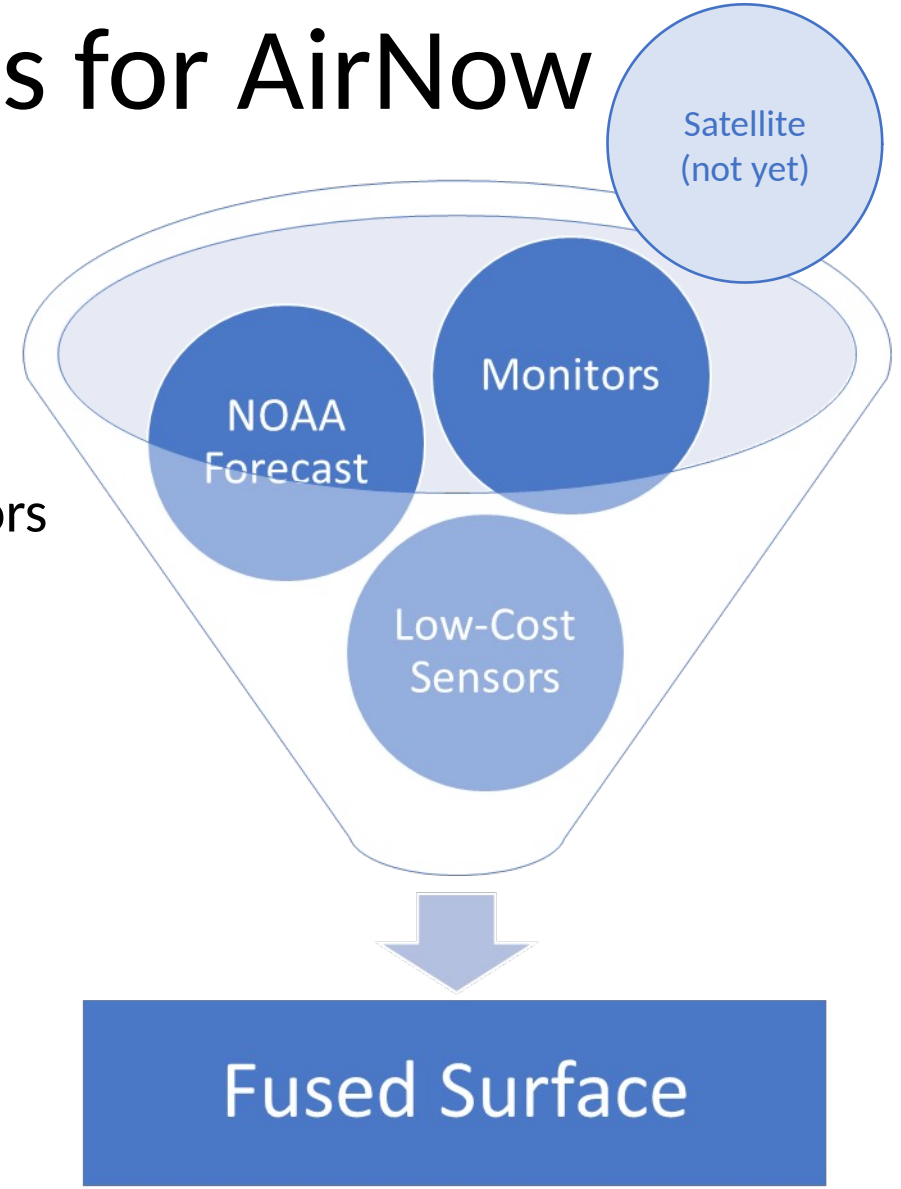
Barron H. Henderson[1,4] and Phil Dickerson[1,4]

Co-authors: Pawan Gupta[2,4], Shobha Kondragunta[3,4], Yang Liu[4,5], Alqamah Sayeed, Hai Zhang[3,4], Janica Gordon[5], Halil Cakir[1], Brett Gantt[1], Benjamin Wells[1], Marcus Proctor[6,7], Steven Proctor[6,8], Youngsun Jung[9], and the HAQAST AirNow Teams[4]

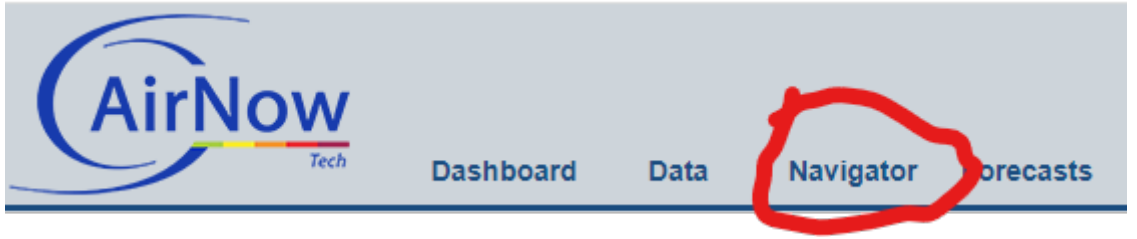
[1]US EPA Office of Air Quality Planning and Standards; [2]National Aeronautics and Space Administration; [3]National Oceanic and Atmospheric Administration / National Environmental Satellite, Data, and Information Service; [4]NASA Health and Air Quality Applied Sciences Team and Tiger Teams; [5]North Carolina Agricultural and Technical State University; [6]AirNow Data Management Center; [7]Tantus Technologies Inc.; [8]GoldSystems Inc.; [9]National Oceanic and Atmospheric Administration / National Weather Service

AirFuse: better air quality maps for AirNow

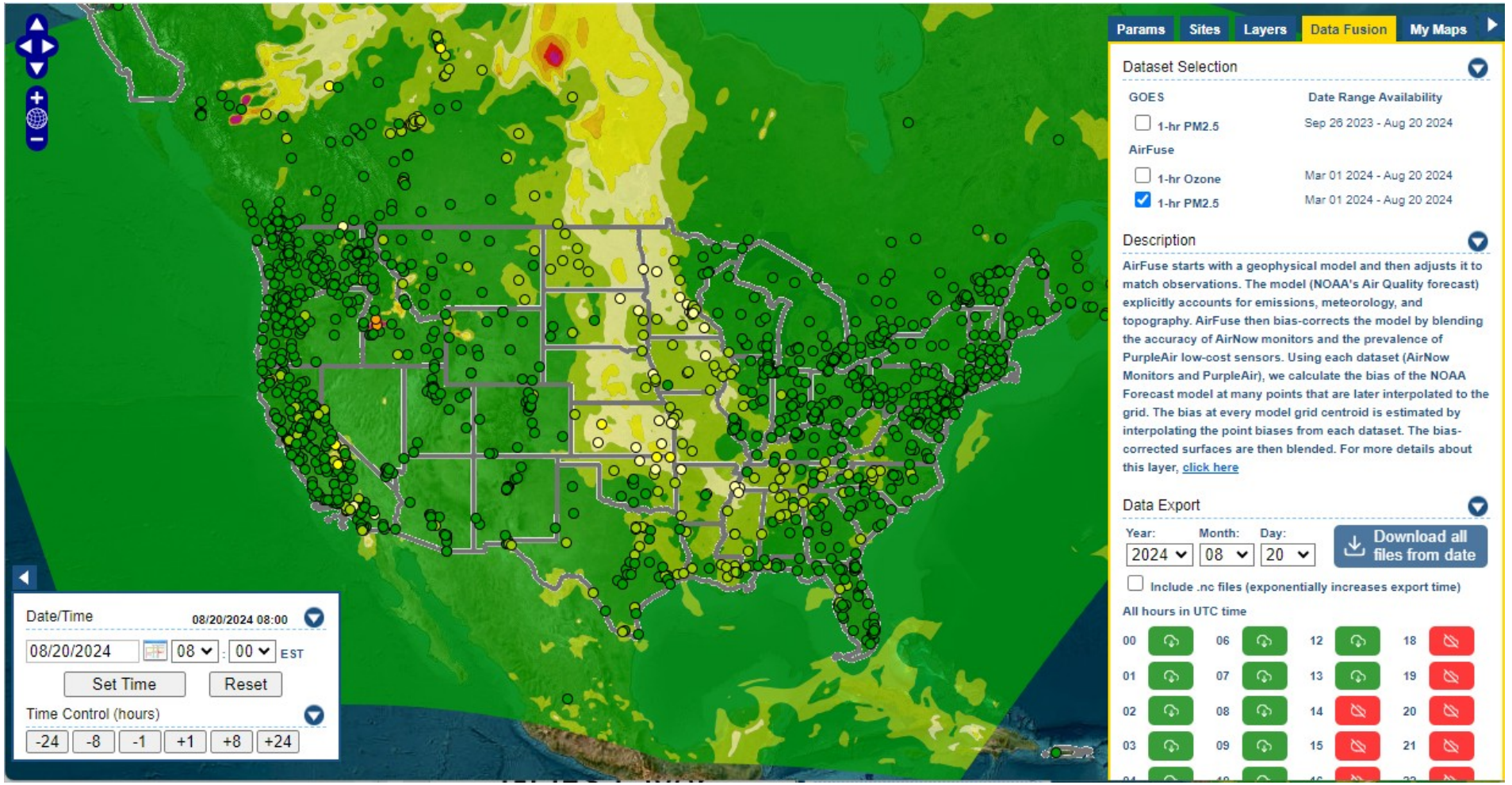
- AirFuse is a method of combining multiple observation sources
  - An OAQPS (AQAD & OID) collaborative product
  - Developed with input from NOAA and NASA collaborators
  - Tests show provides better skill than current AirNow mapping methods
- AirFuse brings together data from multiple sources to improve estimates over any single data source
  - Pilot will include NOAA forecast modeling, AQ monitors and PurpleAir $PM_{2.5}$ sensors
  - Future enhancements could include incorporation of satellite data to cover unmonitored areas



How-to: https://airnowtech.org/



- Login to ANT
- Choose Navigator
- On Navigator, choose the Data Fusion tab.
- Select an AirFuse Layer



Thanks AirNow team!

Pilot Goals

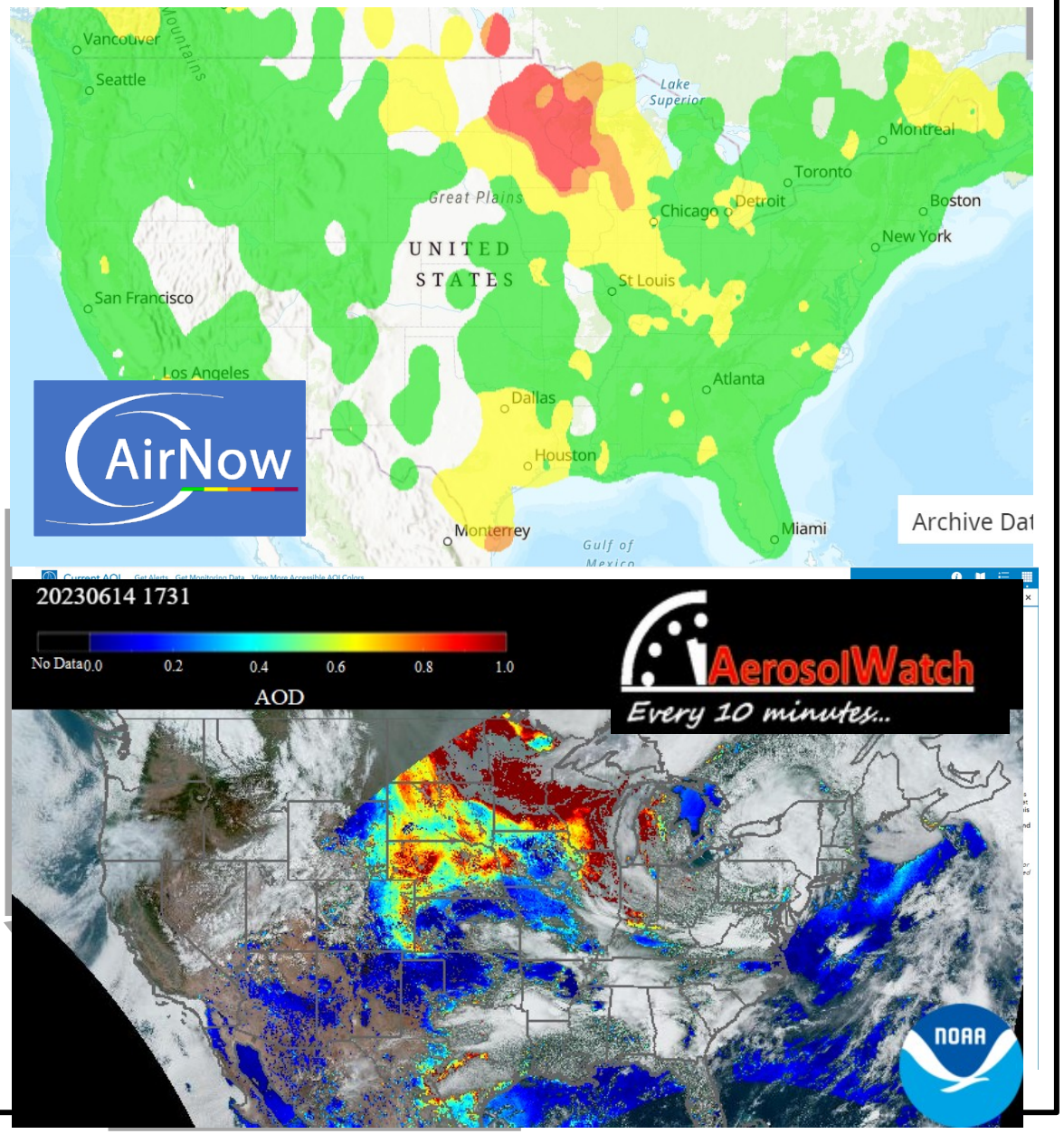
- Pilot AirFuse in AirNowTech from March 2023 to March 2024
  - State/Local/Tribal Agency input is critical
  - What is working well?
  - What needs improvement?
  - We want your feedback to ensure there are no surprises
- During the pilot
  - Investigate incorporation into other parts of AirNow
  - AirNow reports highest-monitor AQI in SLT defined groups*
  - Better AQI maps may allow for more local information
- *So, what is AirFuse and how does it work?*
- *Okay, but how has AirFuse been doing?*

*Groups are called reporting areas and their scope changes from day to day.

AirFuse

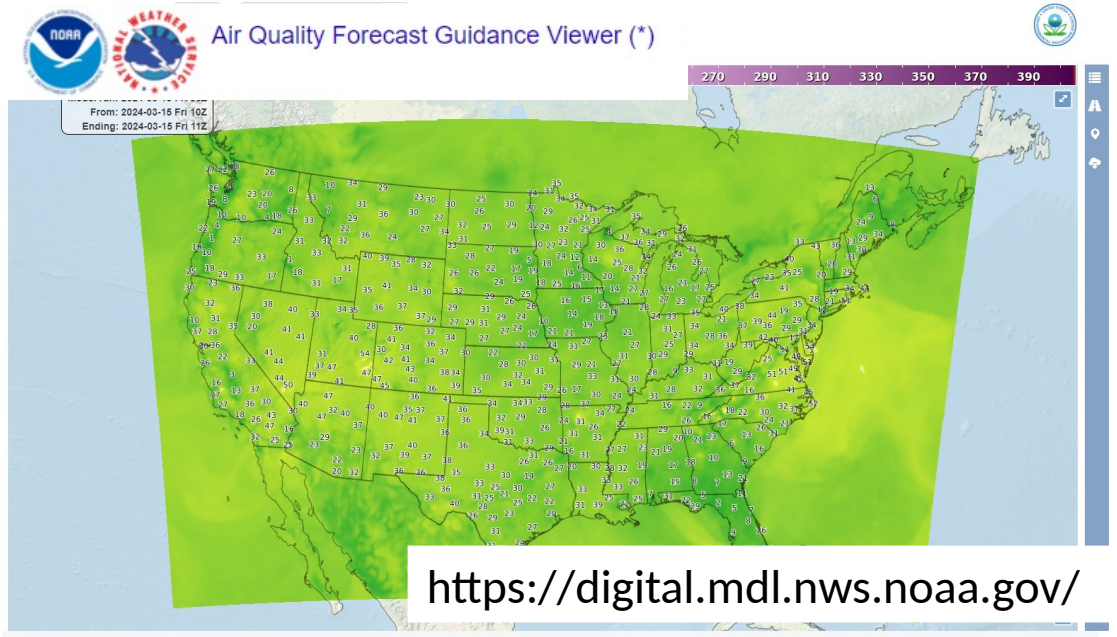
- AirFuse uses a NOAA's forecast and integrates multiple sources of observations.
- 8x+ more PurpleAir sensors than monitors
  - Increased the spatial coverage of monitored particulate matter.
  - Spoiler alert: sensor data improves predictions.
- Near-real-time satellite observations
  - Recent development by NOAA/NESDIS/STAR
  - NASA HAQAST project connecting AirNow to NOAA geostationary satellite data
- Ideally, use similar methods for ozone too.

Example Day in AirNow and Aerosol Watch



Starts with a numerical forecast.

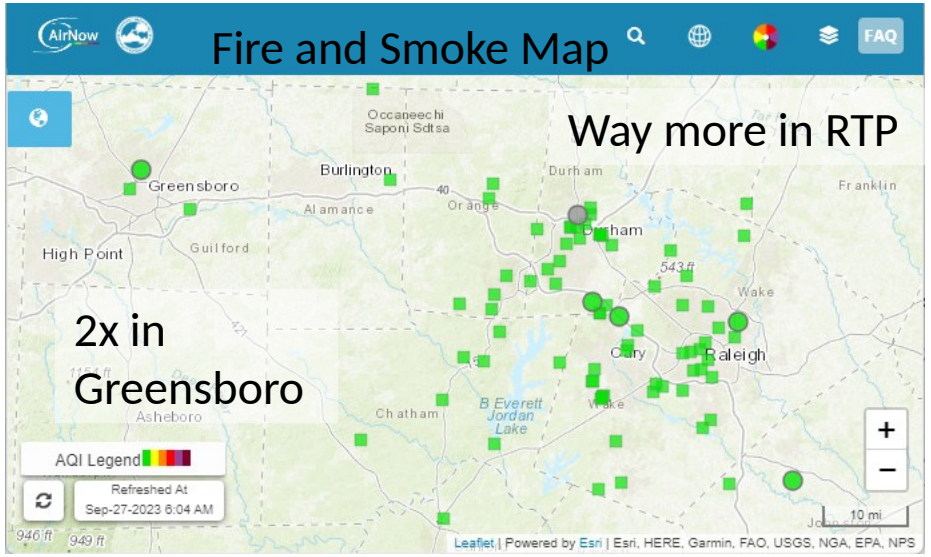
- All models are wrong, some are useful. – George Box
- NOAA reports a bias corrected forecast.
  - Forecasts concentration with CMAQ.
  - Forecast bias at monitors (Kalman Filter Analog)
  - Interpolates bias to grid cells and correct model.



- Why not use this directly?
  - We don't have to forecast the bias, it already happened.
  - Correcting based on multiple observations.

Monitors and PurpleAir sensors

- Many agencies report monitor data to AirNow
  - ~1000 reporting monitors per hour
  - Publicly available thru AirNowAPI
  - Gold standard data!
- PurpleAir low-cost sensors with EPA correction
  - Barkjohn et al. 2021 developed a national correction
  - When PurpleAir is less than 210 micrograms/m3, PM is reduced by 0.0862 x Relative Humidity% (50%: -4.31 and 35%: -3.02)
- South Coast uses a similar system with PurpleAir
  - Schulte et al (2020) residual Kriging
    - NOAA Forecast model
    - Model Correction : $Y = M_n - Krig(M_n - O_n)$
  - Observations (O) from both AirNow and PurpleAir
  - Improved performance of PM2.5 in leave-one-out validation and compared to Federal Reference Monitors
- What about satellites?



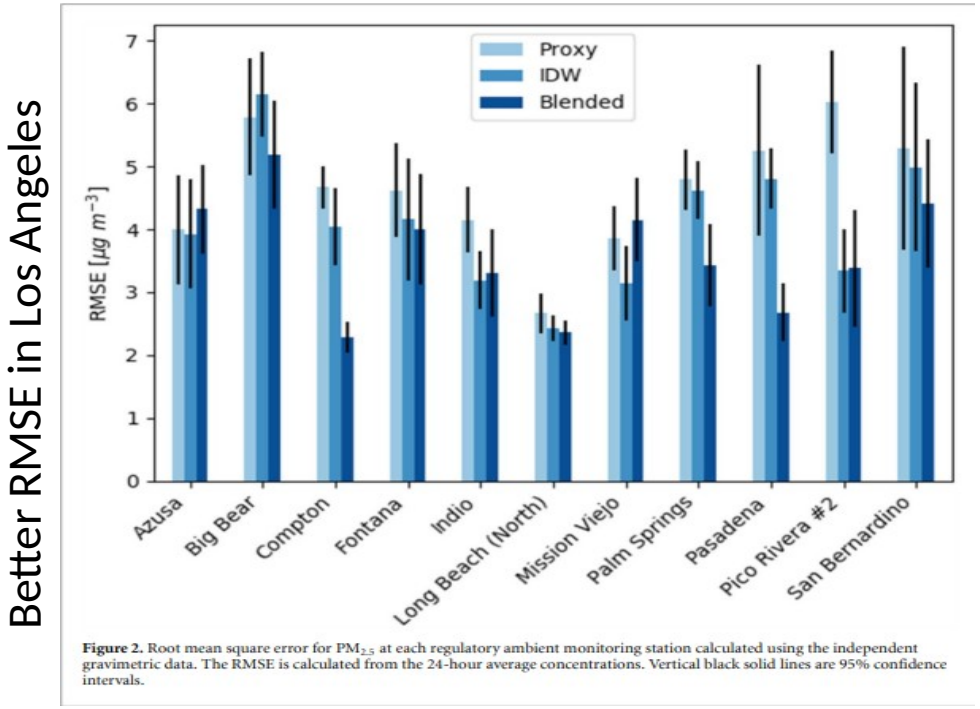


Figure 2. Root mean square error for $PM_{2.5}$ at each regulatory ambient monitoring station calculated using the independent gravimetric data. The RMSE is calculated from the 24-hour average concentrations. Vertical black solid lines are 95% confidence intervals.

Schulte et al 2020 (10.1088/1748-9326/abb62b)

HAQAST "AirNow" Tiger Team

Team Lead: HAQAST investigator Pawan Gupta

Partners: Phil Dickerson and Barron Henderson with the US Environmental Protection Agency (EPA), and Shobha Kondragunta with the National Oceanic and Atmospheric Administration (NOAA)

HAQAST Members and Collaborators: Jingqiu Mao, Yang Liu, Kel Markert, Robert Levy, Randall Martin, Amber J. Soja, Martin Stuefer, Jenny Bratburd, Emily Gargulinksi, Yanshun Li, and Daniel Tong also contribute to this team.

Pawan Gupta

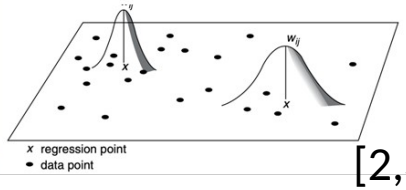
Yang Liu

lea by Yang Liu

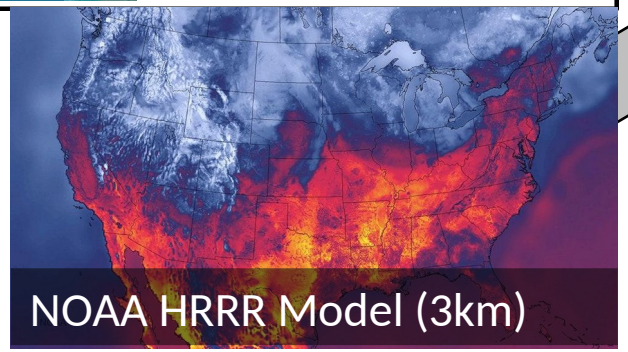
[1]

Satellite AOD + Geographic Weighted Regression

$$PM2.5_{ij} = a_{0ij} + a_{1ij}AOD$$

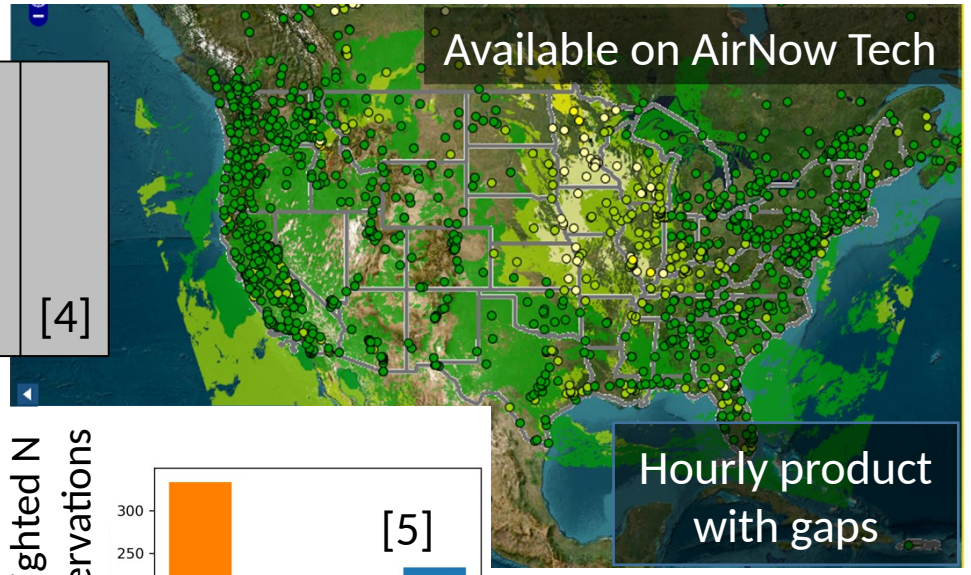


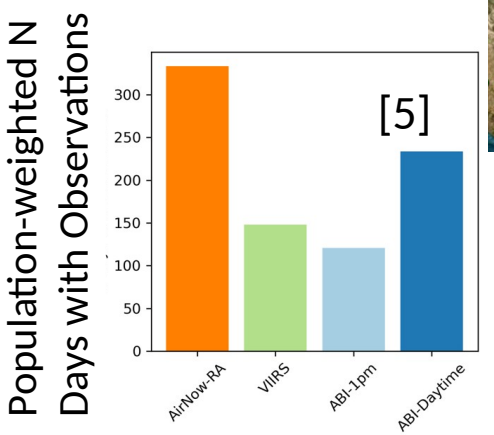
[2,3]



Machine Learning
14 Blended
Deep Neural Networks

[4]





[5]

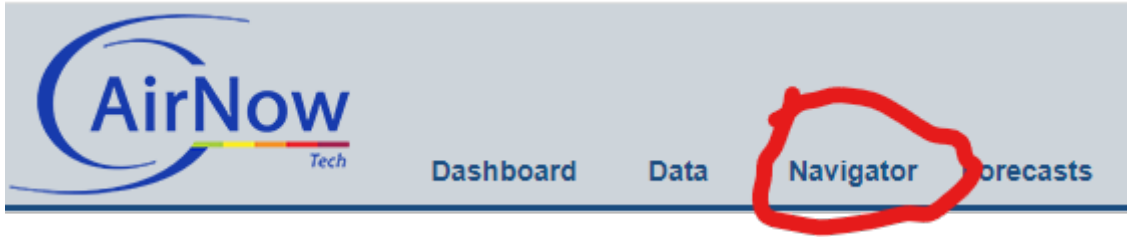
1.Bratburd et al.: Air Quality Data When You Need It: Incorporating Satellite Data Updates into AirNow, EM Plus, 2022.

2.Zhang et al.: Nowcasting Applications of Geostationary Satellite Hourly Surface PM2.5 Data. Weather and Forecasting, 37(12), 2313-2329, 2022. doi: 10.1175/WAF-D-22-0114.1

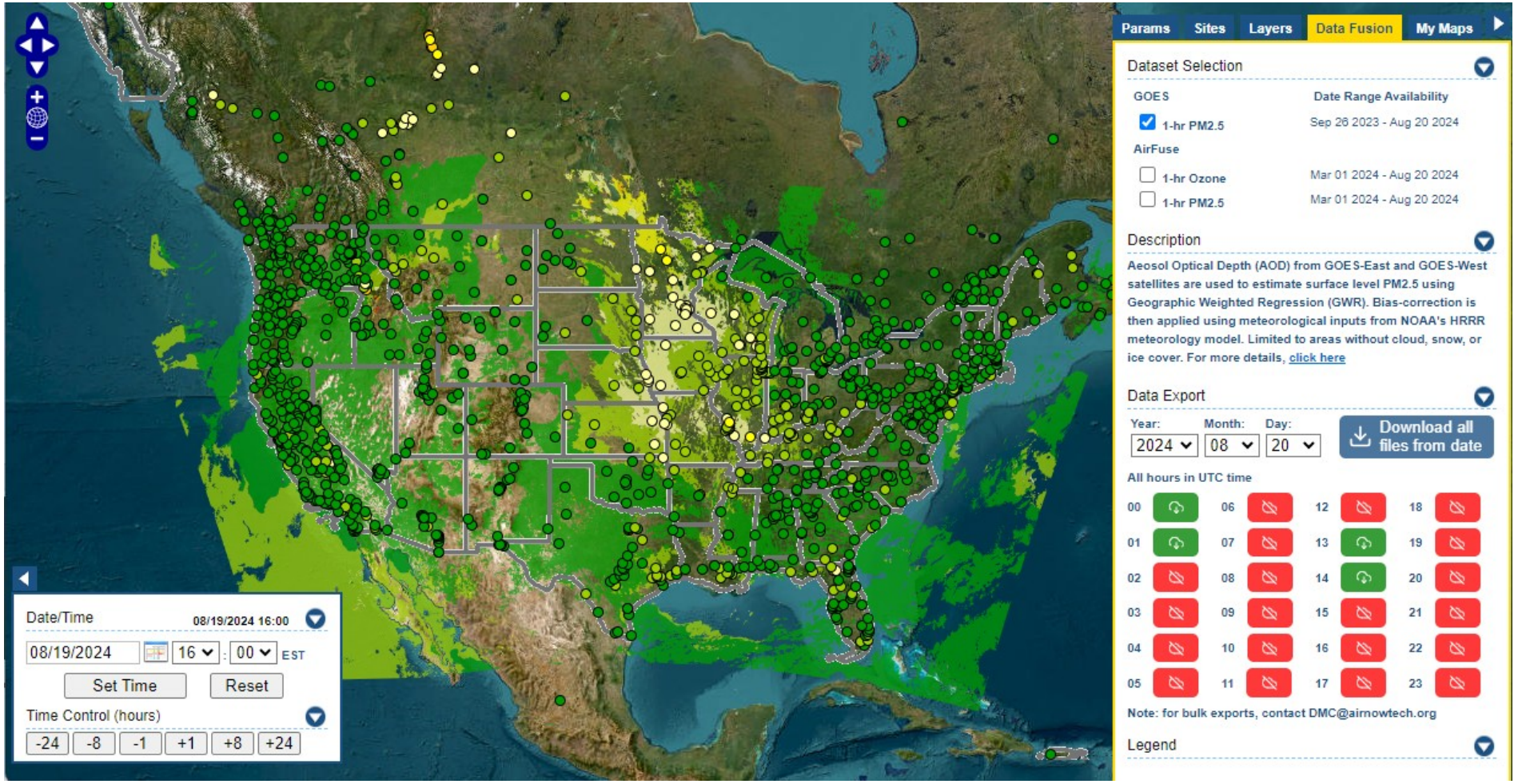
3.Sayeed et al: Deep Neural Network bias corrections (submitted);

4.O'Dell et al.: Public Health Benefits from Improved Identification of Severe Air Pollution Events with Geostationary Satellite Data, *GeoHealth*, 2023.

How-to: https://airnowtech.org/



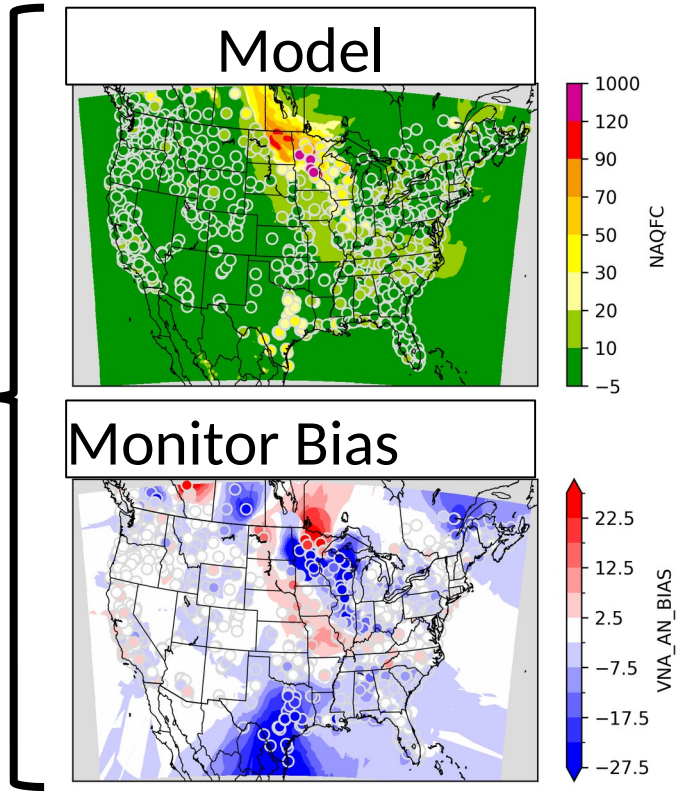
- Login to ANT
- Choose Navigator
- On Navigator, choose the Data Fusion tab.
- Select an AirFuse or GOES Layer



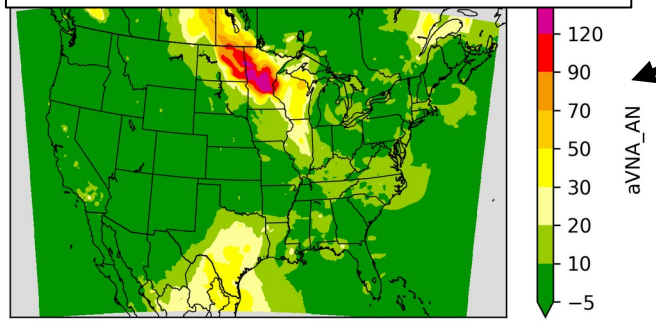
Thanks AirNow team!

Hourly National-scale Fusion Ensemble

- One layer from AirNow monitor (Y_M) ***observations:***
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.



Y_M = f(model, monitor)

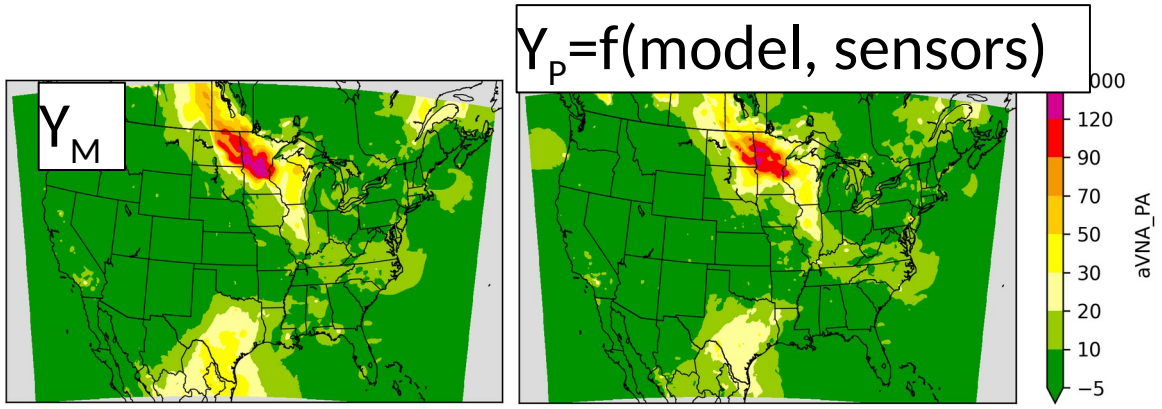


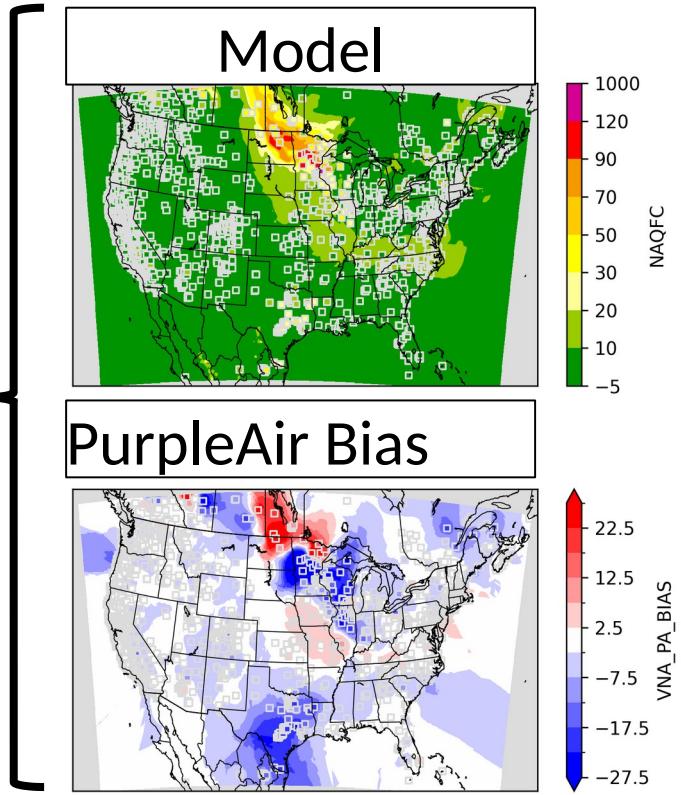
plicative corrector of this type is called extended VNA (eVNA)

**Piece-wise regression as in Fire and Smoke Map

Hourly National-scale Fusion Ensemble

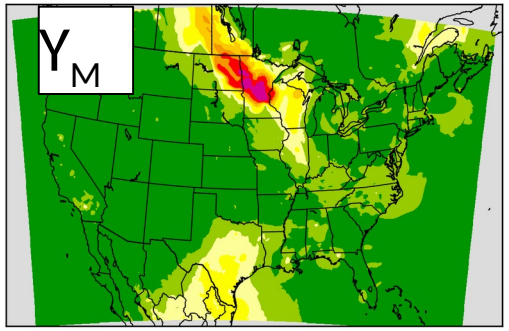
- One layer from AirNow monitor (Y_M) ***observations***:
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.
- One layer from PurpleAir (Y_P) ***observations***:
  - low-cost sensor hourly observations with calibration**
  - Aggregated within grid cells to create a pseudo-observation

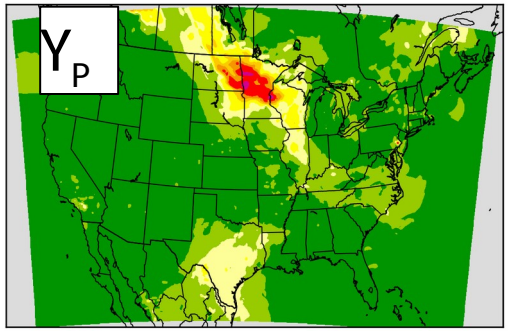


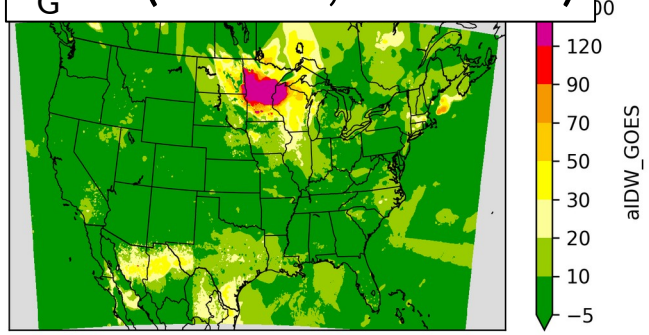


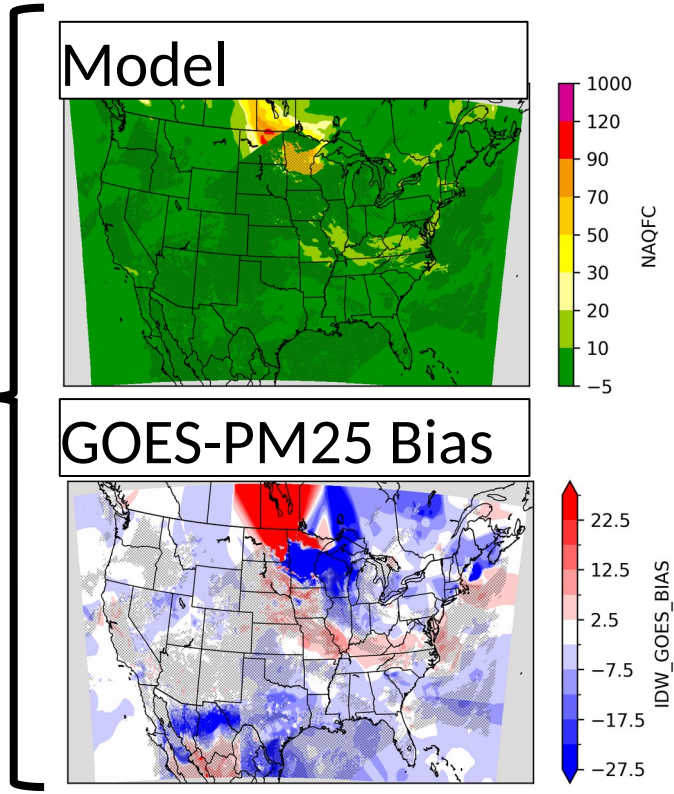
Hourly National-scale Fusion Ensemble

- One layer from AirNow monitor (Y_M) ***observations***:
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.
- One layer from PurpleAir (Y_P) ***observations***:
  - low-cost sensor hourly observations with calibration**
  - Aggregated within grid cells to create a pseudo-observation
- One layer from GOES-PM25 (Y_G) "***observations***"
  - Geostationary Operational Environmental Satellite (GOES)
  - Not clustered like monitors, so VNA interpolation is not necessary.



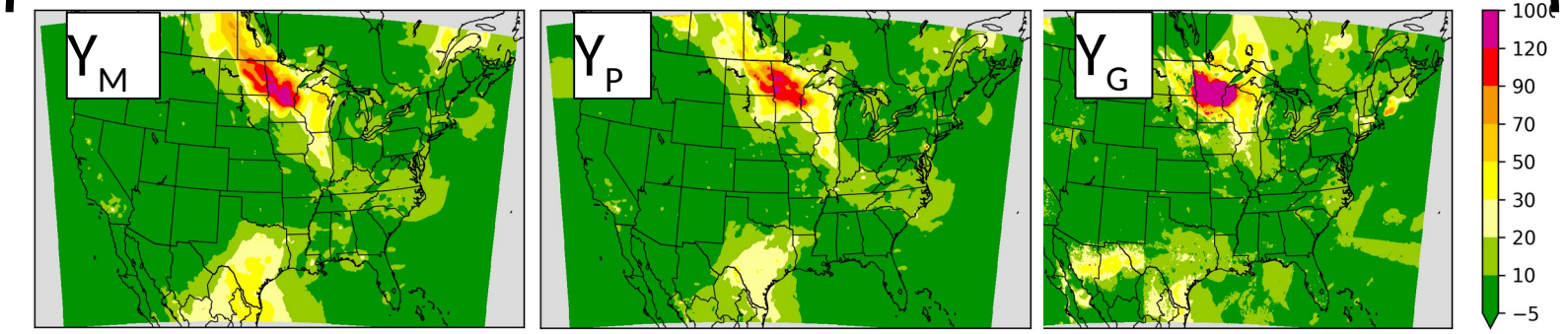


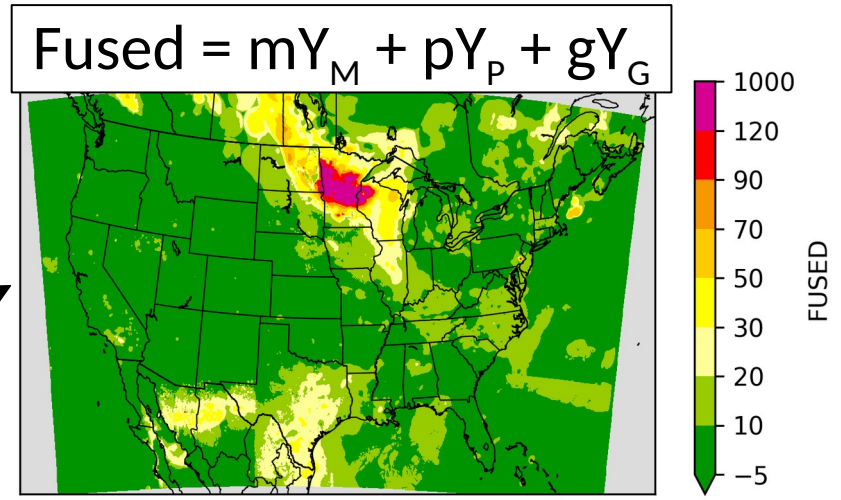




Hourly National-scale Fusion Ensemble

- One layer from AirNow monitor (Y_M) ***observations***:
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.
- One layer from PurpleAir (Y_P) ***observations***:
  - low-cost sensor hourly observations with calibration**
  - Aggregated within grid cells to create a pseudo-observation
- One layer from GOES-PM25 (Y_G) "***observations***"
  - Geostationary Operational Environmental Satellite (GOES)
  - Not clustered like monitors, so VNA interpolation is not necessary.
- Weight based on distance (m, p, g)

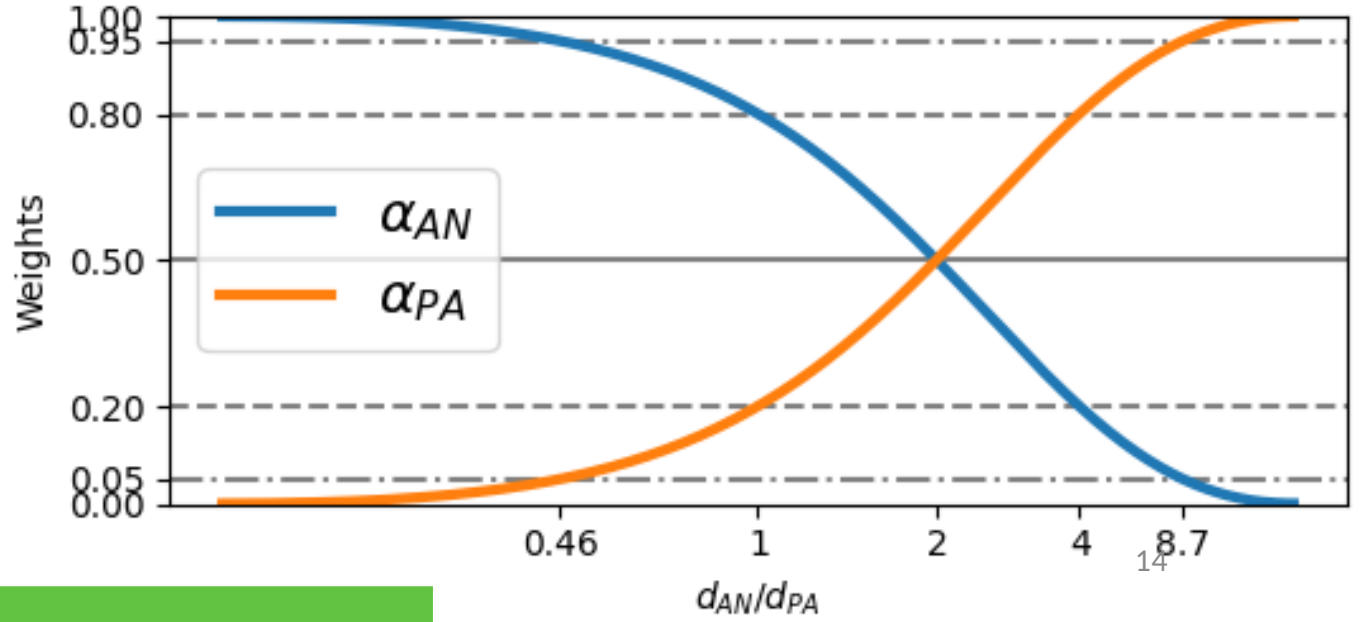




Weight the ensemble of surfaces on distance

$Y = mY_M + pY_P + gY_G$

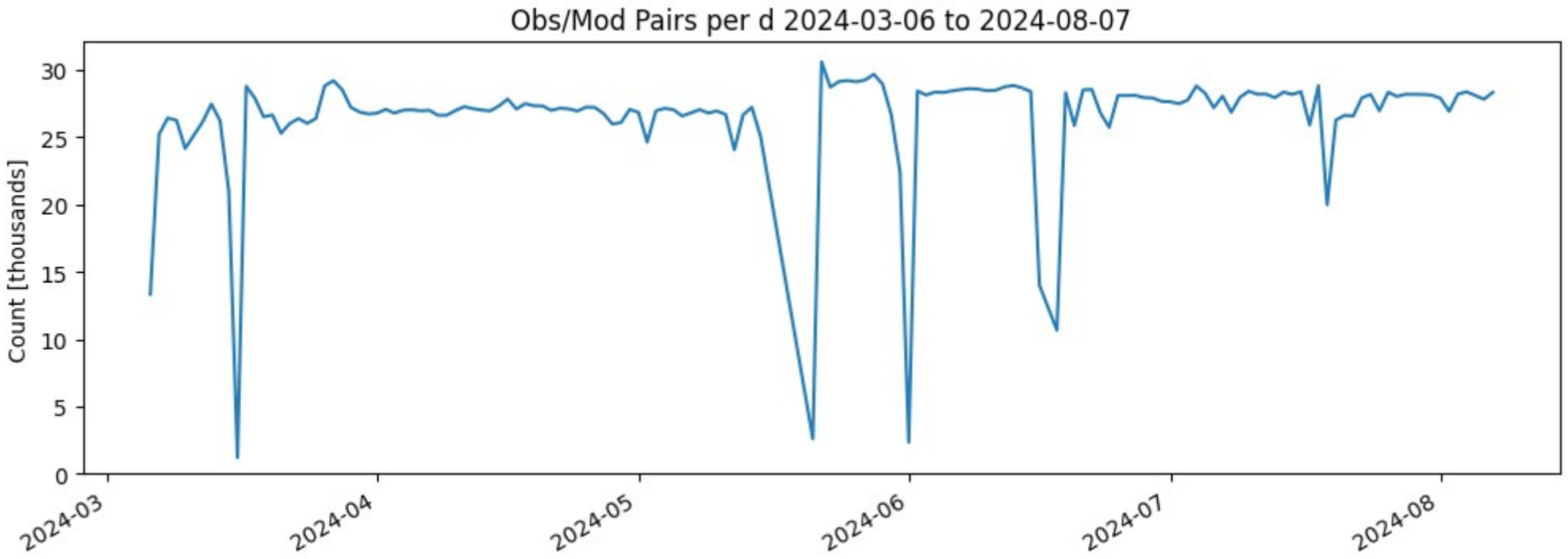
- $m' = (1 \times d_{AN})^{-2}$
- $p' = (2 \times d'_{PA})^{-2}$: $d'_{PA} = \max(d_{PA}, 3.6)$
- $g' = 0$ # **Not yet!**
- Normalize : e.g., $m = m' / (m' + p' + g')$



Validation Methodology

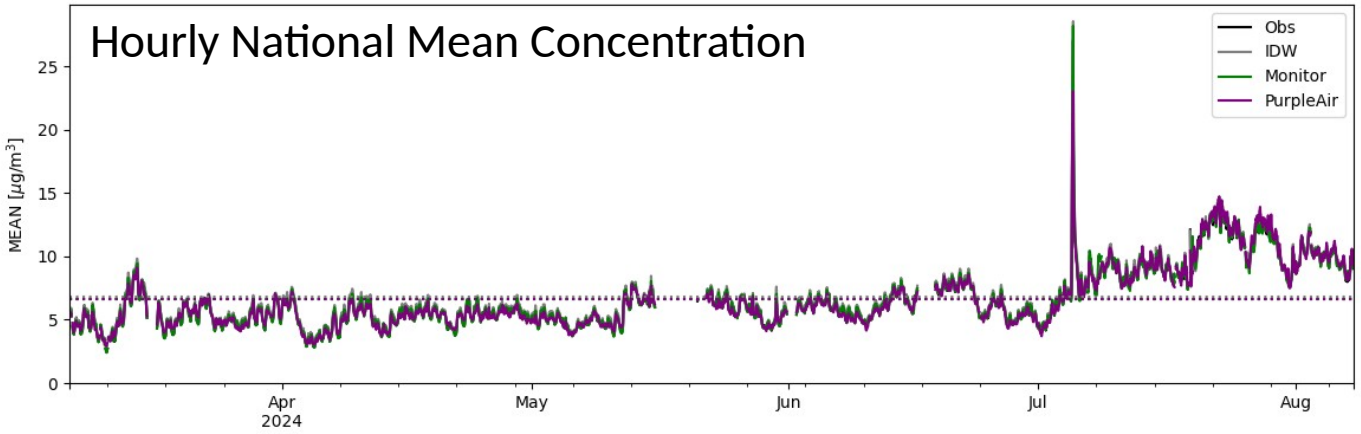
- Running hourly since March 6th
- ~30k obs/model pairs per day
- ~4M obs/model pairs so far

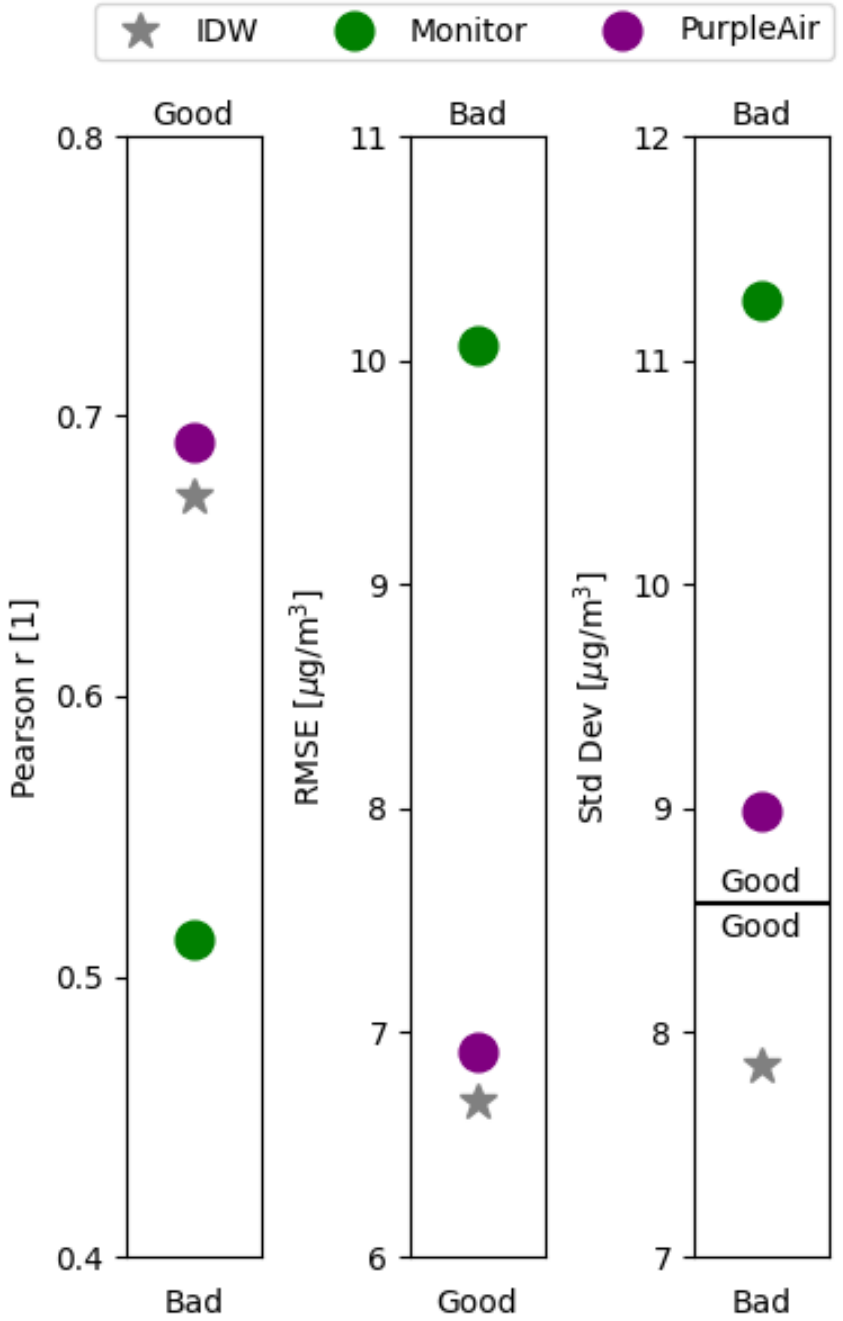
- Validation by 10-fold cross validation
- Validation by Leave-One-Out (LOO) validation



Pilot Validation Summary

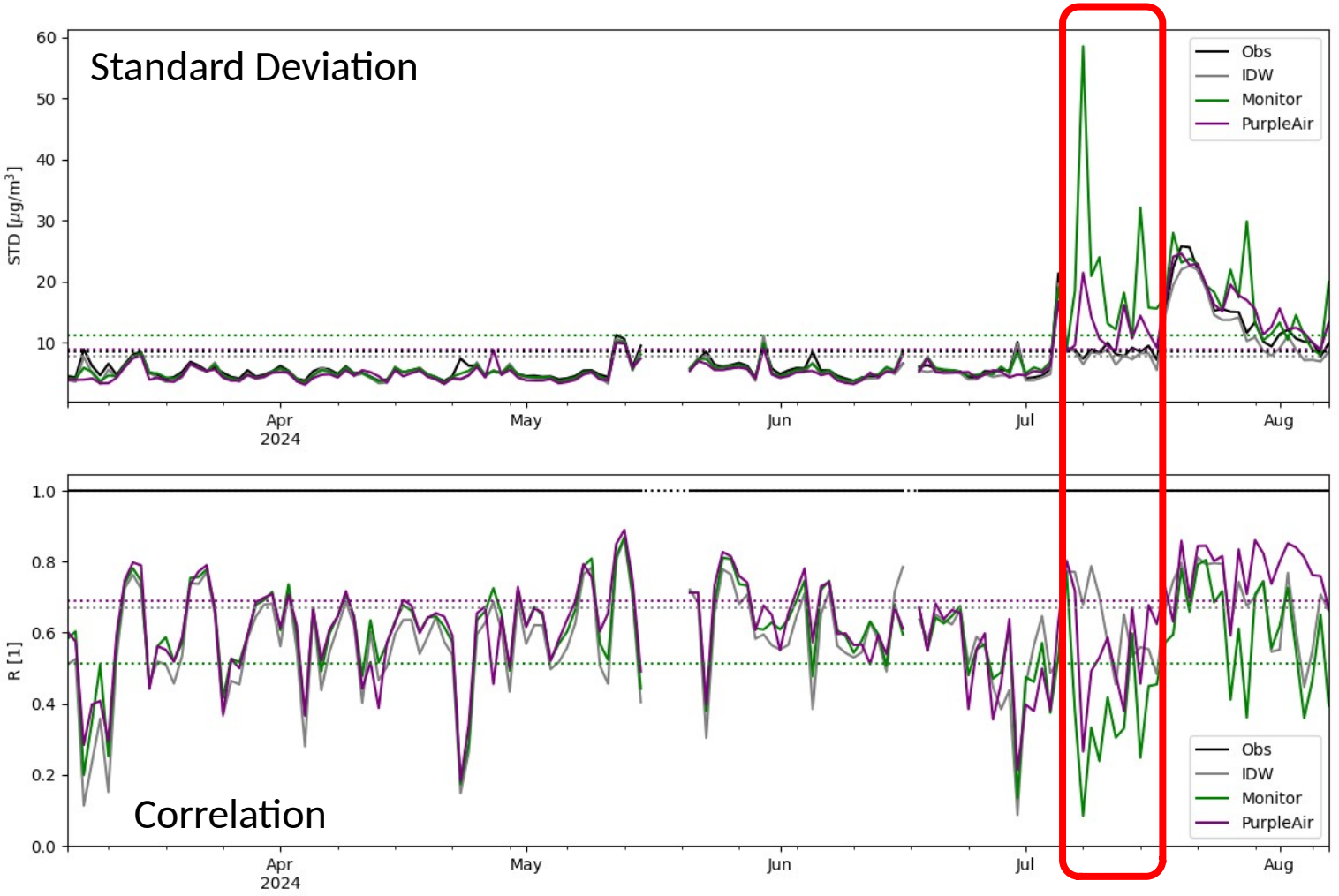
- Using only monitors aVNA performing worse than IDW
- Including PurpleAir improves:
  - Prediction standard deviation,
  - Prediction correlation, and
  - Root mean squared error.





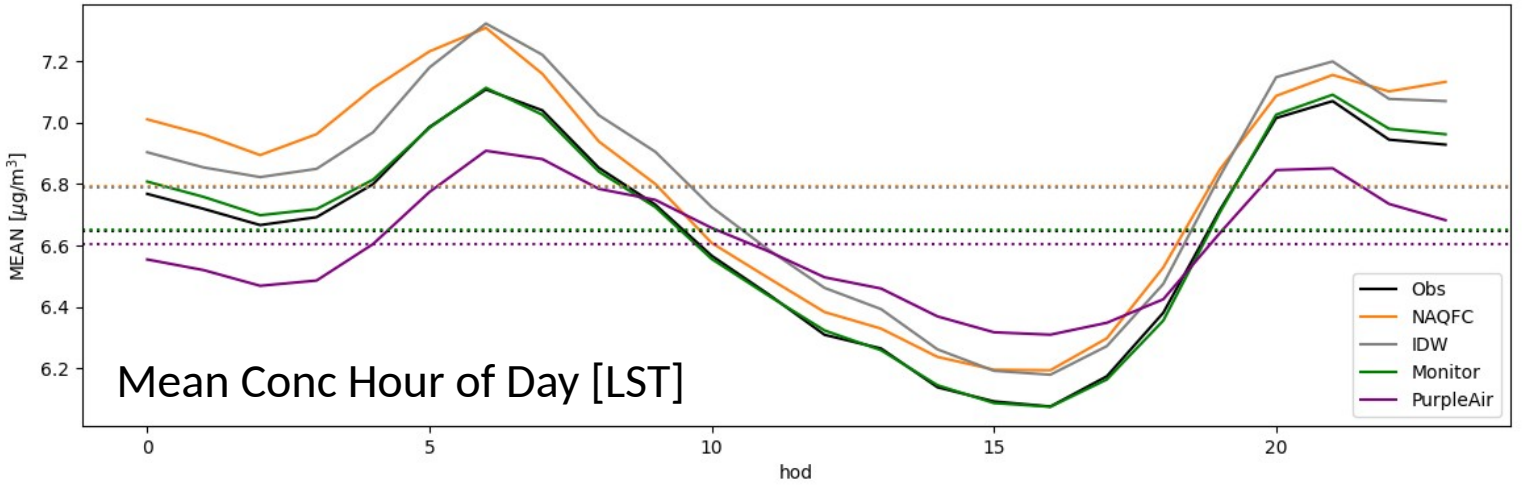
Leave-1-out Validation: National Correlation

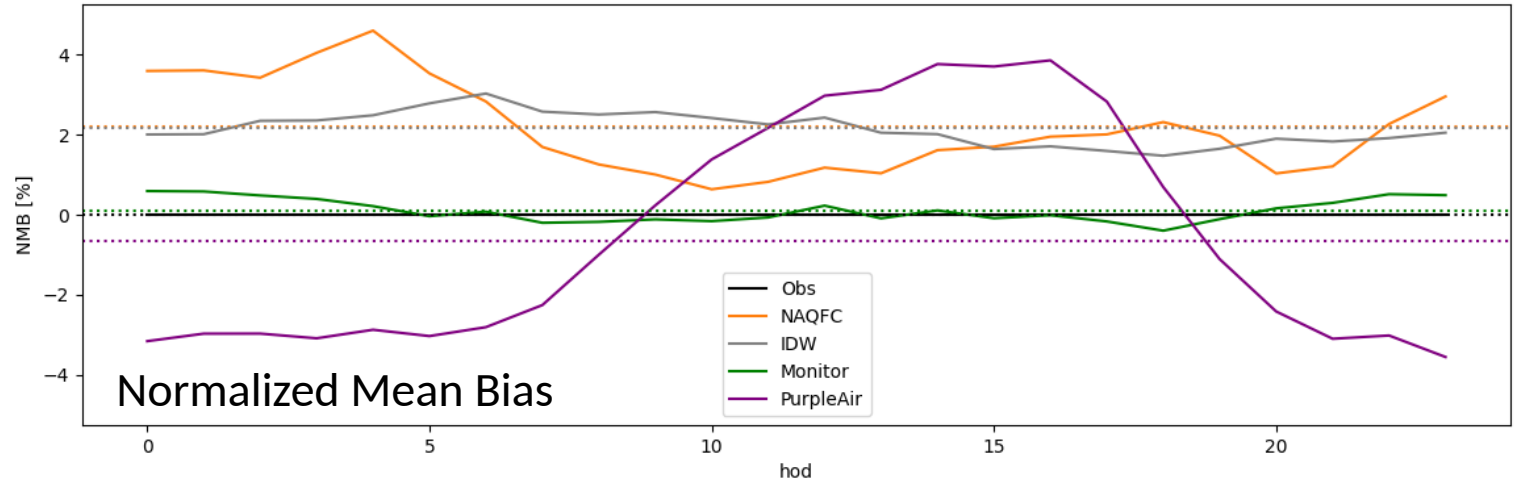
- Days with large spatial variability are hard for any technique.
  - Especially with mobile monitors in fires
- NAQFC has a few days with spotty huge concentrations (1000s).
- Adding PurpleAir “tacks” prevents large deviations.
- Longer-term NAQFC filtering is probably necessary and results like these would not reach the public.



Diurnal Variation of PM and AirFuse

- Hourly particulate matter is highest at night during high humidity.
- CMAQ forecast over does the variability
- IDW and AirFuse w/out PurpleAir capture that variability.
- Adding PurpleAir mutes the diurnal variability.*



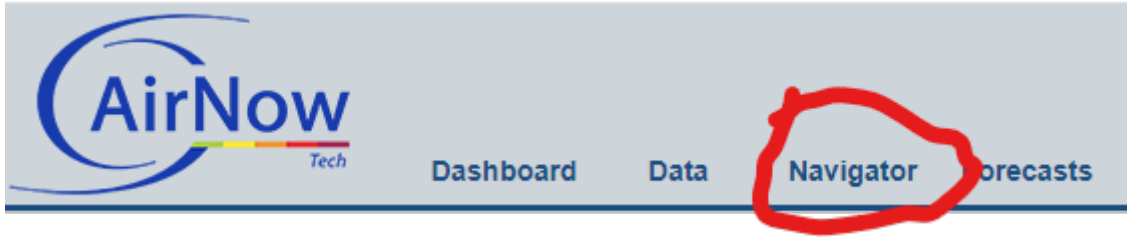


*Correction was developed from daily averages, so investigating update to correction.

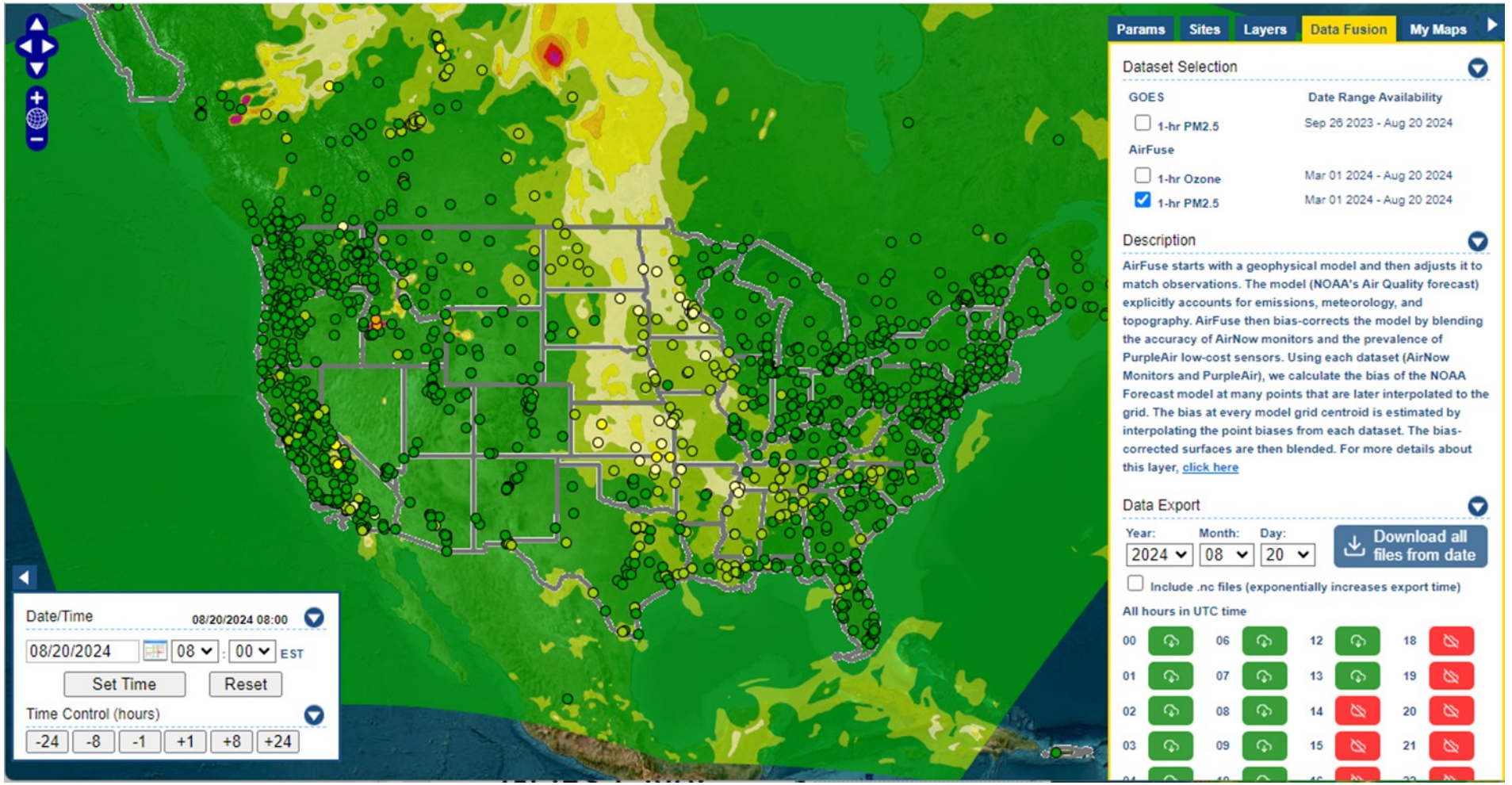
Summary

- AirNow needs an updated interpolation method.
  - EPA has long used models and statistical fusion to fill gaps with regulatory but has not incorporated these methods into AirNow.
  - Schulte et al. demonstrated including models and PurpleAir improved on simple interpolations and applied it in an AirNow-like system.
  - HAQAST Tiger Team evaluated GOES PM25 for real-time-applications.
- Fusion with PurpleAir is ready.
  - Discontinuities are less stark than GOES because datasets are more spatially consistent (ie sparse in the same places).
  - Value of PurpleAir is obvious because they are dense near monitors.
- Fusion with GOES PM25 ongoing work
  - HAQAST Tiger Team 2021 (Gupta) – now 2023 (Yang Liu)
  - Conceptually, the satellite value is highest away from monitors and sensors… making it hard to evaluate
  - ~5% of monitors are further than 30km from their nearest withheld monitor…
- Need your feedback!
  - Statistics will only tell us so much.
  - How does your area look?
  - When does AirFuse give weird answers?

How-to: https://airnowtech.org/



- Login to ANT
- Choose Navigator
- On Navigator, choose the Data Fusion tab.
- Select an AirFuse or GOES Layer



Thanks AirNow team!

Piloting AirFuse

- AirFuse maps are available in AirNowTech!
  - Running hourly from now until March 2025
  - Preloaded back to Mar 1, 2024
  - Download outputs to investigate further
- Get your feedback
  - Collect feedback from forecasters and SLTs
  - Like it? Great. Don’t like it? Even better.
  - Your input can help us make this better.
- Currently, updated once per hour – may not include lately submitted data.

Questions?

henderson.barron@epa.gov