AirFuse

EPA's multi-pollutant fusion system

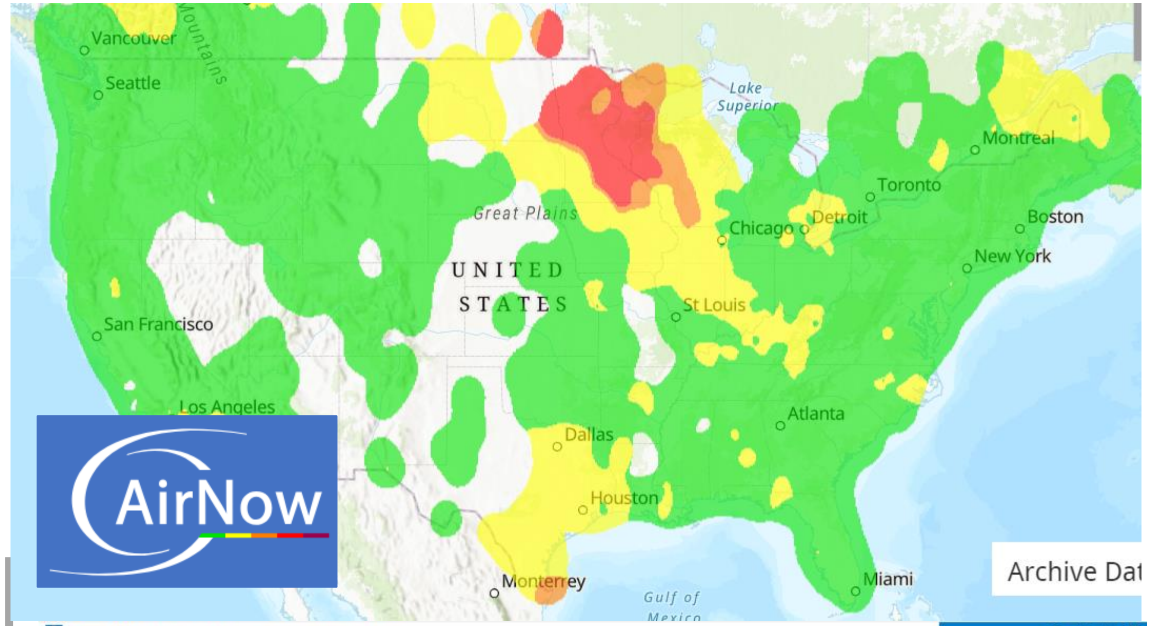
Barron H. Henderson and Phil Dickerson

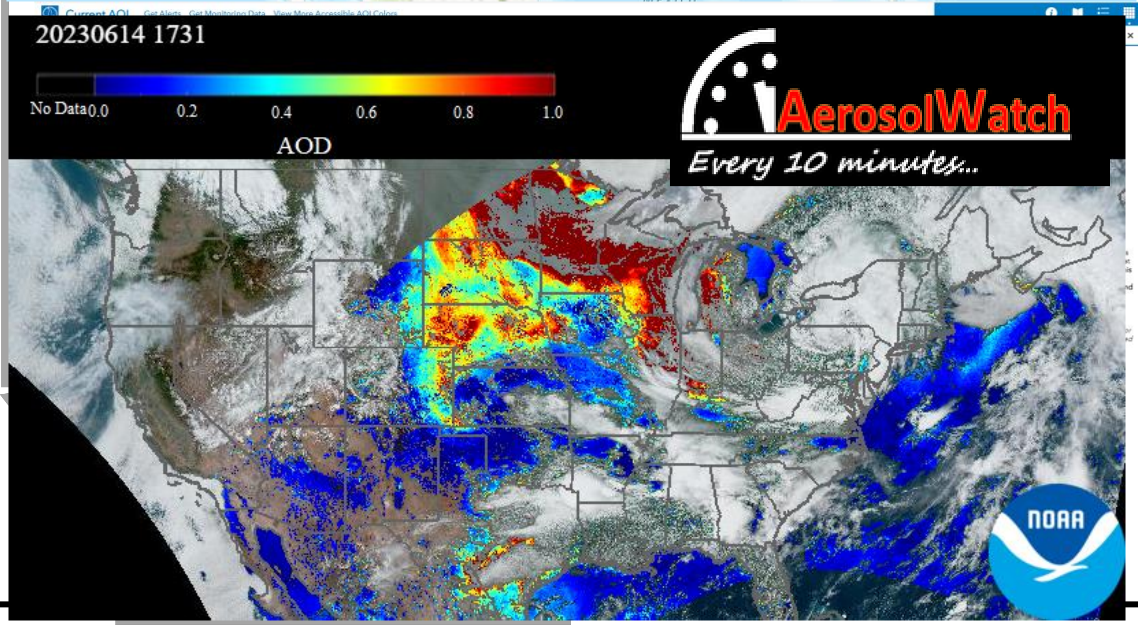
Co-authors: Pawan Gupta, Shobha Kondragunta, Alqamah Sayeed, Hai Zhang, Janica Gordon, Halil Cakir, Brett Gantt, Benjamin Wells, and HAQAST AirNow Team

Motivation

- AirNow communicates air quality in real time
  - Millions of visitors per day during fire seasons
  - Simple weighted (d^{-5}) contours monitors only
- 4x 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
- What about fusing AirNow, PurpleAir and satellites?
- Ozone is easier, because there is one data source.

Example Day in AirNow and Aerosol Watch



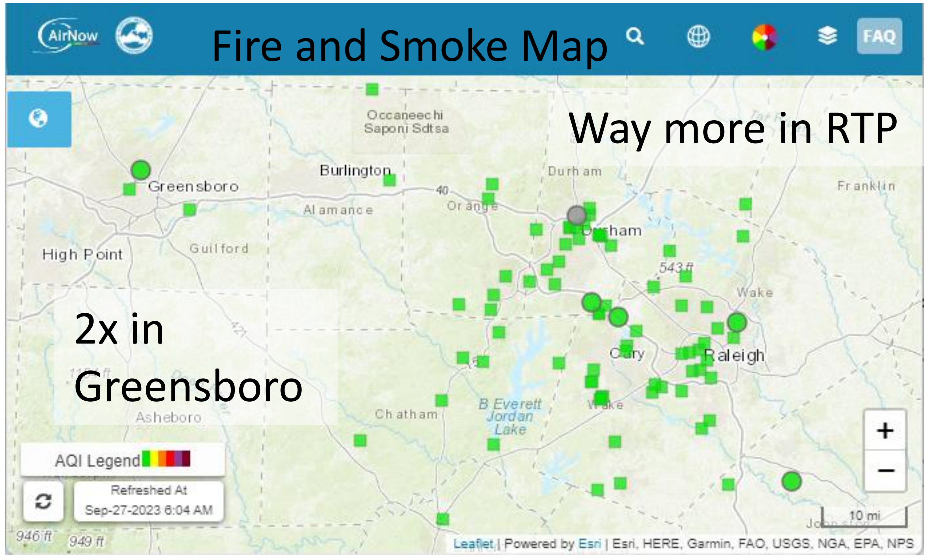


Why not simply use a numerical forecast?

- There are several forecast models available.
  - NOAA's National Air Quality Forecast Capability (NAQFC)
  - NASA's Goddard Earth Observing System Composition Forecast
  - European Copernicus Atmosphere Monitoring Service
- All models are wrong, some models are useful.
- NOAA provides a bias corrected model.
  - Forecasts model bias at monitors using a Kalman Filter Analog system
  - Interpolates bias to grid cells.
  - Removes bias from forecast.
- In AirNow, we don't have to forecast the bias – it already happened.

Monitors and PurpleAir sensors

- Many agencies report monitor data to AirNow
  - ~1000 reporting monitors per hour
  - Publicly available thru AirNowAPI
- Schulte et al (2020) using PurpleAir
  - Residual Kriging with both AirNow and PurpleAir
    - NOAA Forecast model
    - Model Correction : $Y = M_n - Krig(M_n - O_n)$
  - Improved performance of PM2.5 in leave-one-out validation and compared to Federal Reference Monitors
- We use corrected PurpleAir low-cost sensors
  - Barkjohn et al. 2021 developed a national correction
  - Extended correction via RSIG



Better RMSE in Los Angeles

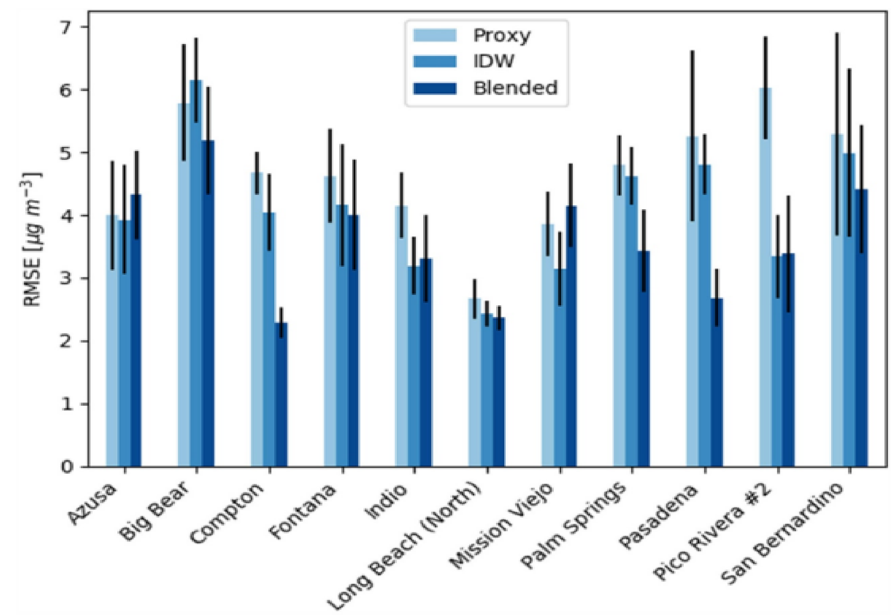


Figure 2. Root mean square error for $PM_{2.5}$ at each regulatory ambient monitoring station calculated using the independent gravimetric data. The RMSE is calculated from the 24-hour average concentrations. Vertical black solid lines are 95% confidence intervals.

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

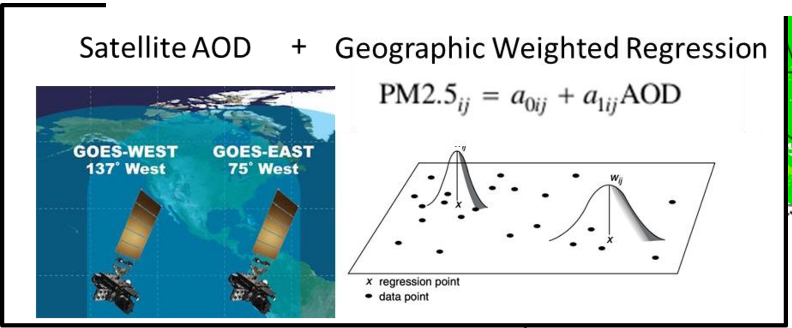
4. Enabling USEPA to ingest high-frequency satellite air quality data into the AirNow system

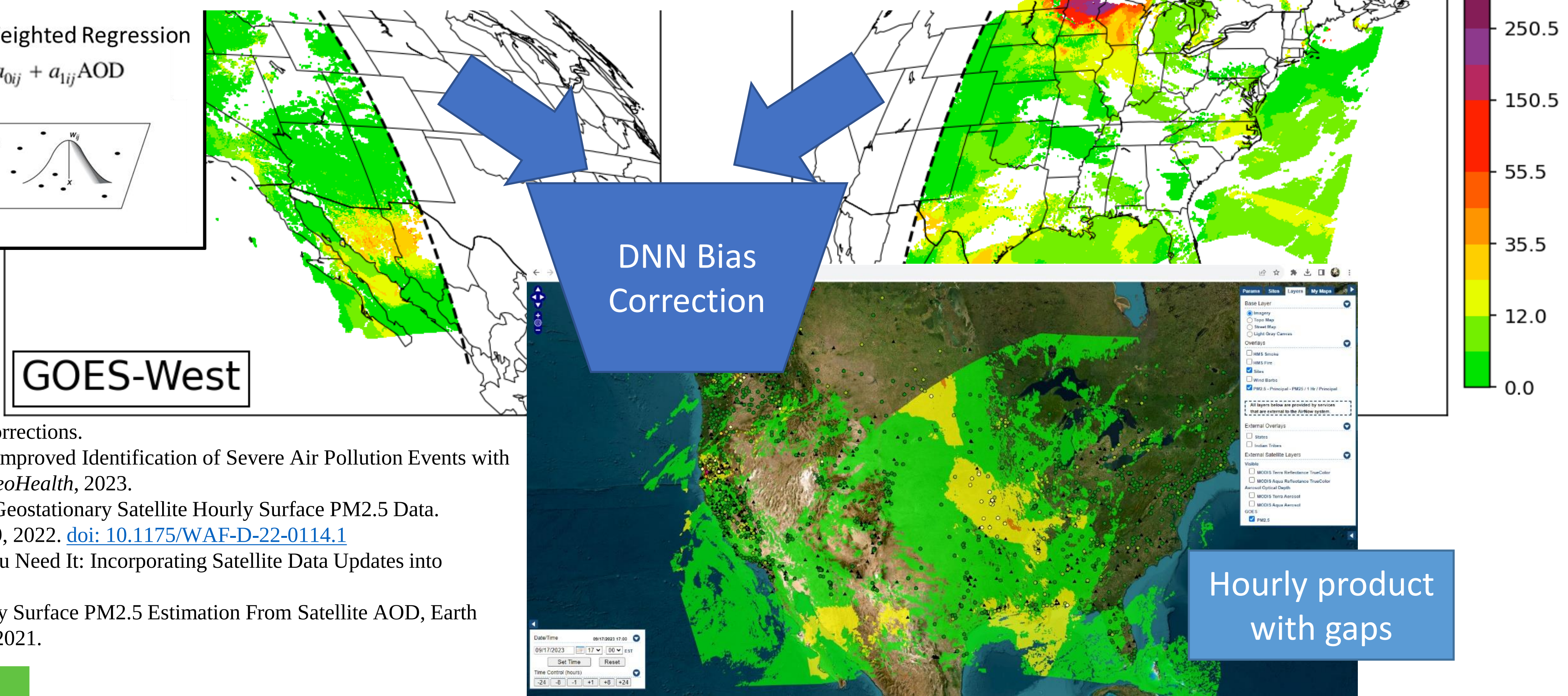
Team Lead: HAQAST investigator Pawan Gupta

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: Jianqiu 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.

https://haqast.org/tiger-teams/#2021-tiger-teams





1.Sayeed et al: Deep Neural Network bias corrections.
2.O'Dell et al.: Public Health Benefits from Improved Identification of Severe Air Pollution Events with Geostationary Satellite Data, *submitted to GeoHealth*, 2023.
3.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
4.Bratburd et al.: Air Quality Data When You Need It: Incorporating Satellite Data Updates into AirNow, EM Plus, 2022.
5.Zhang and Kondragunta.: Daily and Hourly Surface PM2.5 Estimation From Satellite AOD, Earth Space Sci, 8, doi: 10.1029/2020EA001599, 2021.

Hourly National-scale Fusion Ensemble

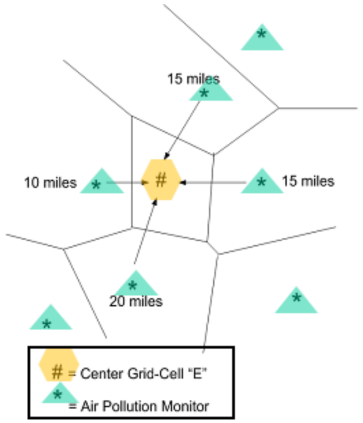
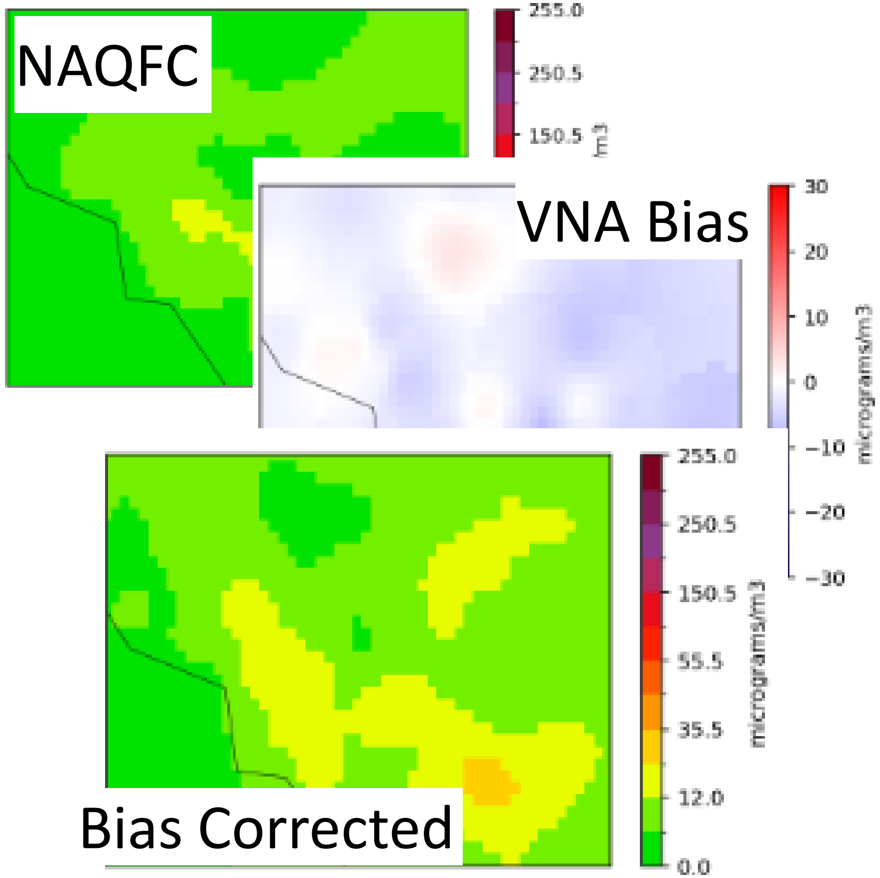


Figure courtesy of: Brian Timin

- Interpolating bias to "correct" the forecast model*
  - NOAA's Forecast Model (NAQFC) as mediating layer
  - VNA Bias = $\text{sum}(d_n^{-2}(m_n - o_n)) / \text{sum}(d_n^{-2})$ • n = Voronoi Neighbor
  - Y_i = NAQFC – VNA Bias_i
- One layer from AirNow (Y_{AN}) ***observations***:
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.



*A multiplicative corrector of this type is called extended VNA (eVNA)
**Piece-wise regression as in Fire and Smoke Map

Hourly National-scale Fusion Ensemble

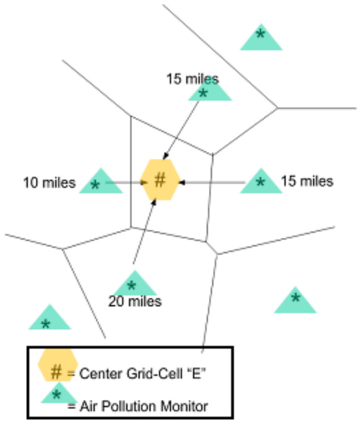
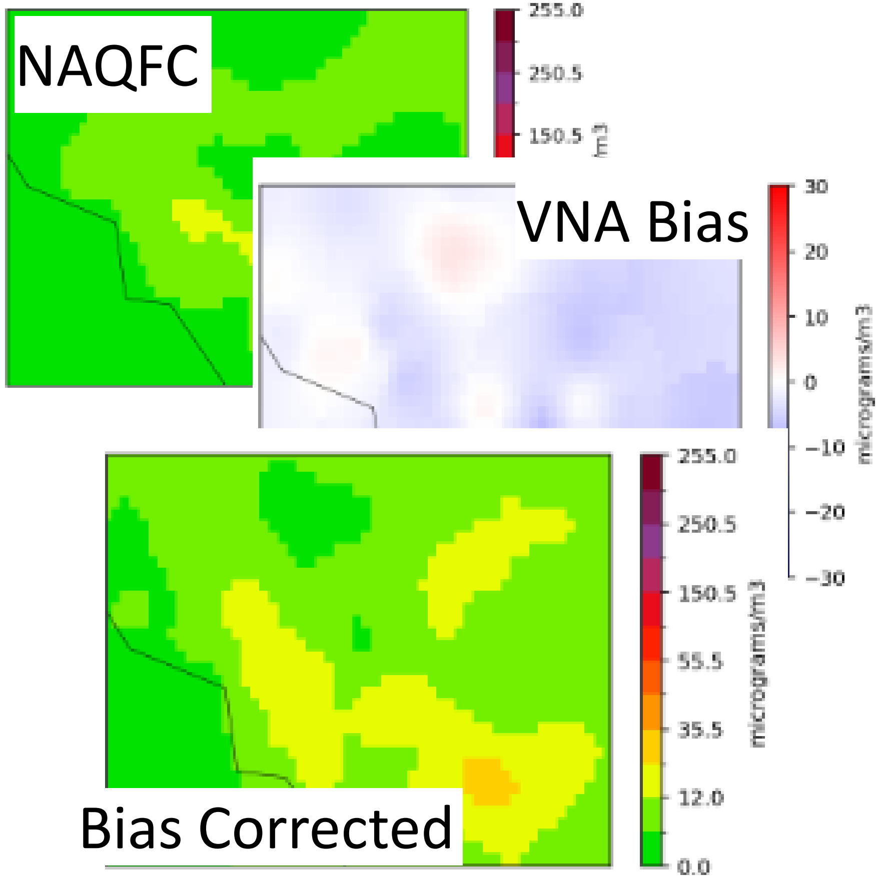


Figure courtesy of: Brian Timin

- Interpolating bias to "correct" the forecast model*
  - NOAA's Forecast Model (NAQFC) as mediating layer
  - VNA Bias = sum(d_n^{-2} (m_n - o_n)) / sum(d_n^{-2})
    - n = Voronoi Neighbor
  - Y_i = NAQFC – VNA $Bias_i$
- One layer from AirNow (Y_{AN}) ***observations***:
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.
- One layer from PurpleAir (Y_{PA}) ***observations***:
  - low-cost sensor hourly observations with calibration**
  - Aggregated within grid cells to create a pseudo-observation



*A multiplicative corrector of this type is called extended VNA (eVNA)
**Piece-wise regression as in Fire and Smoke Map

Hourly National-scale Fusion Ensemble

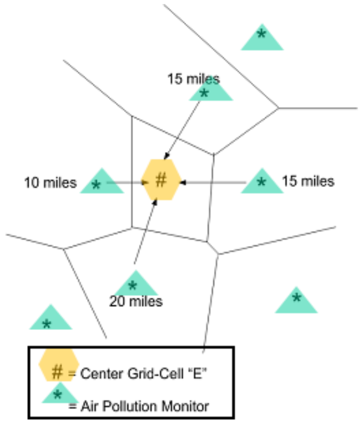
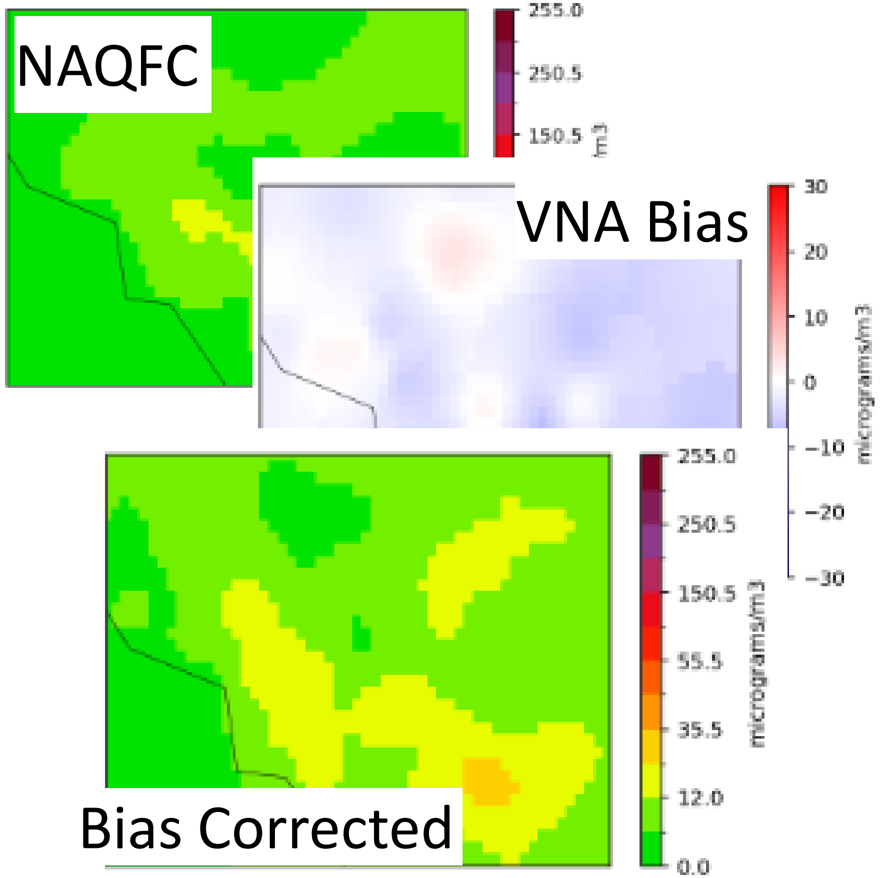


Figure courtesy of: Brian Timin

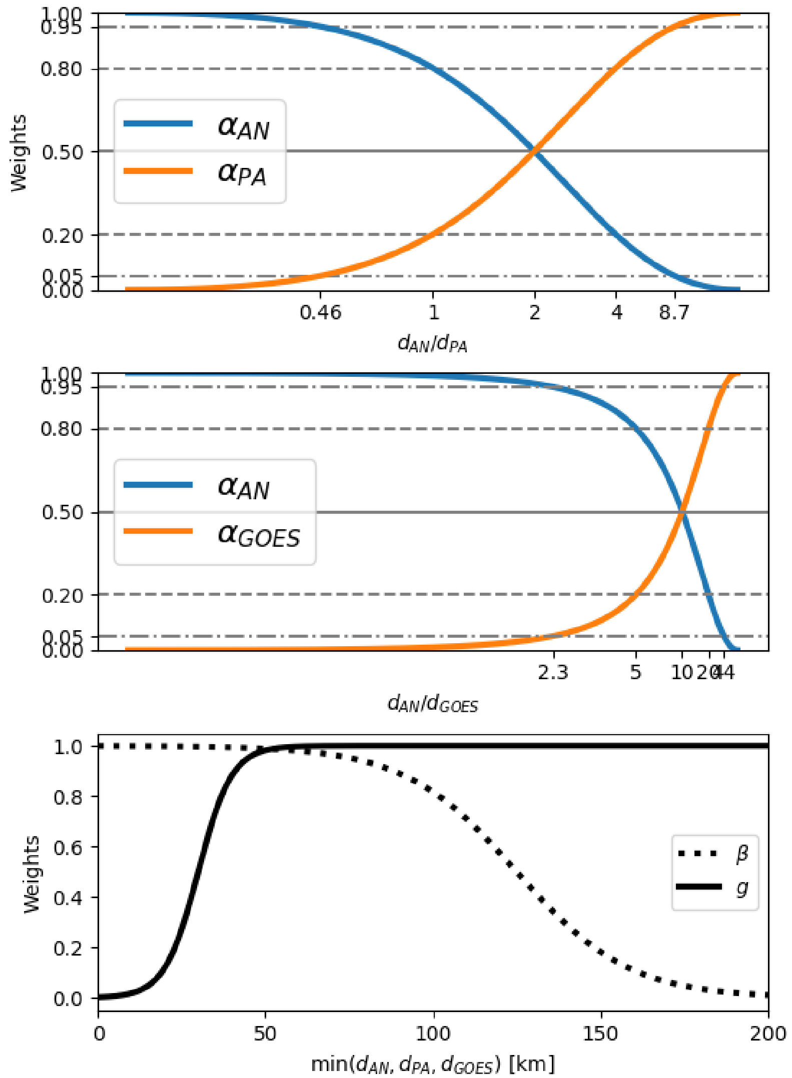
- Interpolating bias to "correct" the forecast model*
  - NOAA's Forecast Model (NAQFC) as mediating layer
  - VNA Bias = $\text{sum}(d_n^{-2}(m_n - o_n)) / \text{sum}(d_n^{-2})$ • n = Voronoi Neighbor
  - Y_i = NAQFC – VNA Bias_i
- One layer from AirNow (Y_{AN}) ***observations***:
  - mostly regulatory grade hourly observations
  - paired with collocated grid cell.
- One layer from PurpleAir (Y_{PA}) ***observations***:
  - low-cost sensor hourly observations with calibration**
  - Aggregated within grid cells to create a pseudo-observation
- One layer from GOES-PM25 (Y_{GOES}) "***observations***"
  - Geostationary Operational Environmental Satellite (GOES)
  - Aerosol Optical Depth from the GOES Advanced Baseline Imager
  - Geographic Weighted Regression (GWR) against AirNow
  - Deep Neural Network Corrected (Sayeed et al *in prep*)



*A multiplicative corrector of this type is called extended VNA (eVNA)
**Piece-wise regression as in Fire and Smoke Map

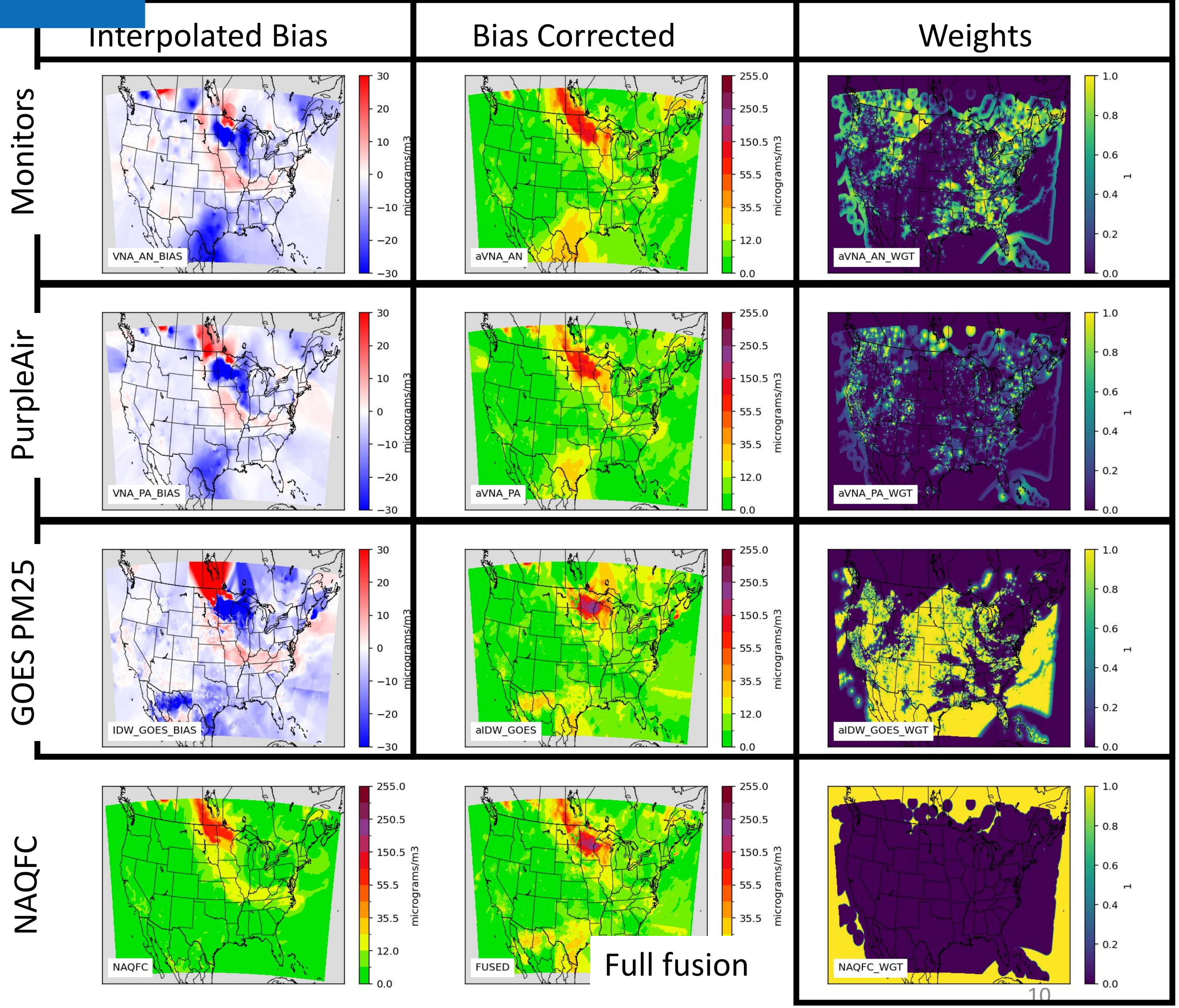
Ensemble Averaging Method

- Simple fusion of bias corrected surfaces
  - NAQFC, AirNow, PurpleAir, GOES-PM25
  - Fuse the surfaces based on distance
  - Apply different weights to ensembles
- $Y_{AN,PA,GOES} = \alpha_{AN} Y_{AN} + \alpha_{PA} Y_{PA} + \alpha_{GOES} Y_{GOES}$
  - $\alpha'_{AN} = (1 \times d_{AN})^{-2}$
  - $\alpha'_{PA} = (2 \times d_{PA})^{-2}$
  - $\alpha'_{GOES} = (10 \times d_{GOES})^{-2}$
  - $\alpha'_{tot} = \alpha'_{AN} + \alpha'_{PA} + \alpha'_{GOES}$
  - Normalize them all: $\alpha_i = \alpha'_i / \alpha'_{tot}$
- $Y_{AN,PA,GOES} = \beta \times Y_{AN,PA,GOES} + (1 - \beta) \times Y_{NAQFC}$

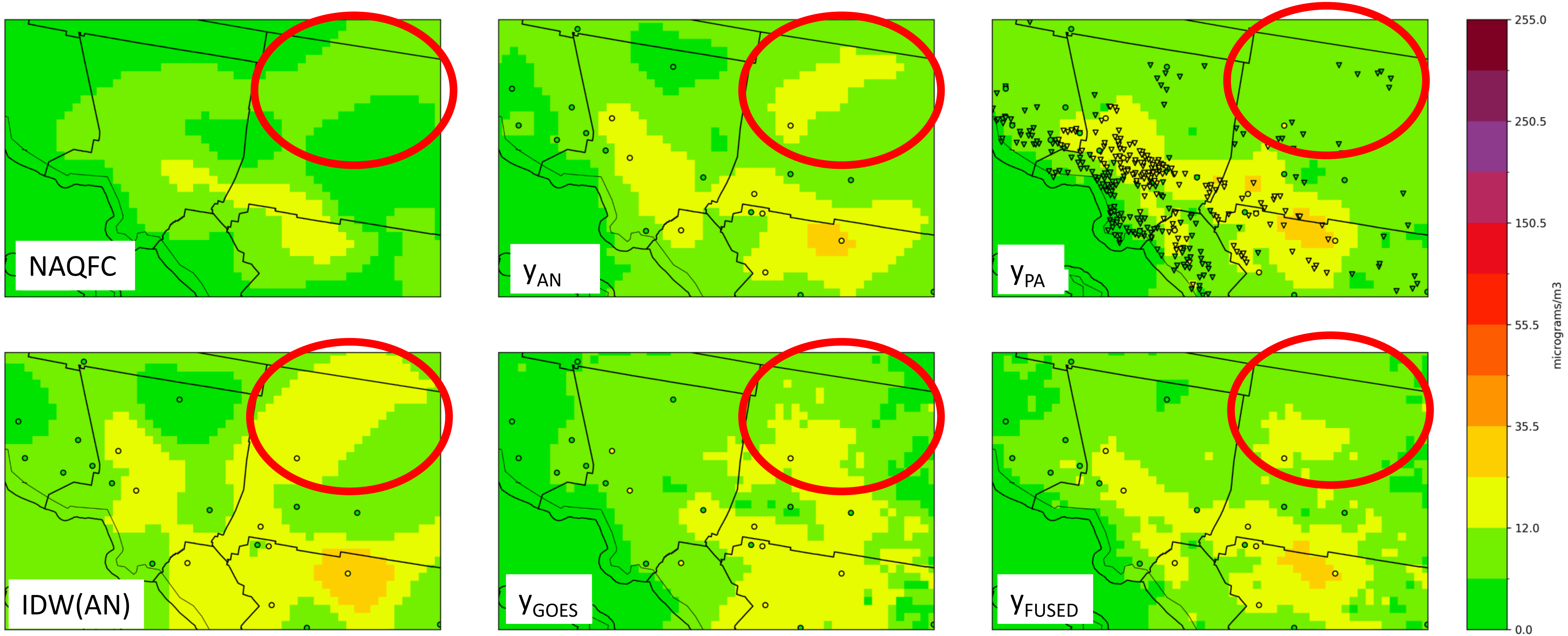


Case Study 2023-06-14T17Z

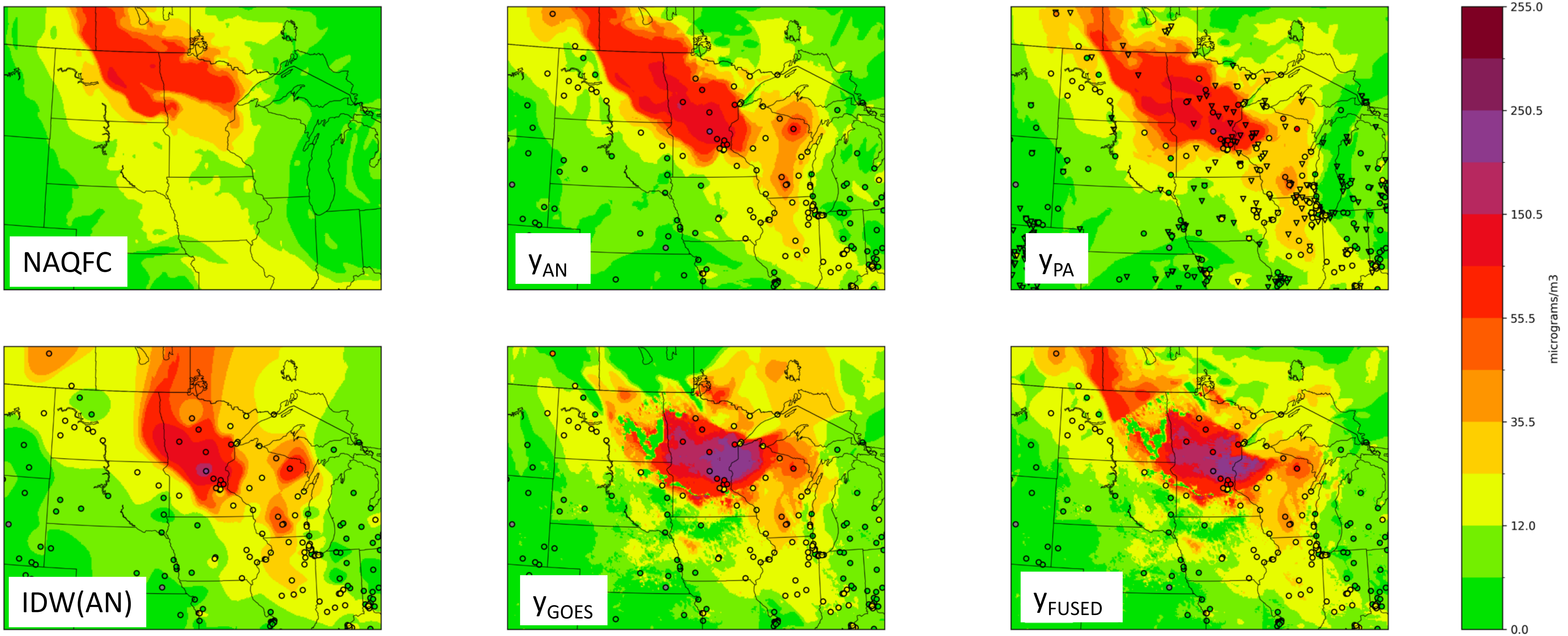
- Fairly typical day June day in the south western domain.
- Large fire contributions in Canada and sweeping down through Minnesota, Wisconsin and further
- 4 data sources
  - AirNow Monitors (top)
  - PurpleAir sensors
  - GOES PM25
  - NAQFC (bottom)
- Estimates
- Bias Corrections
- Full fusion



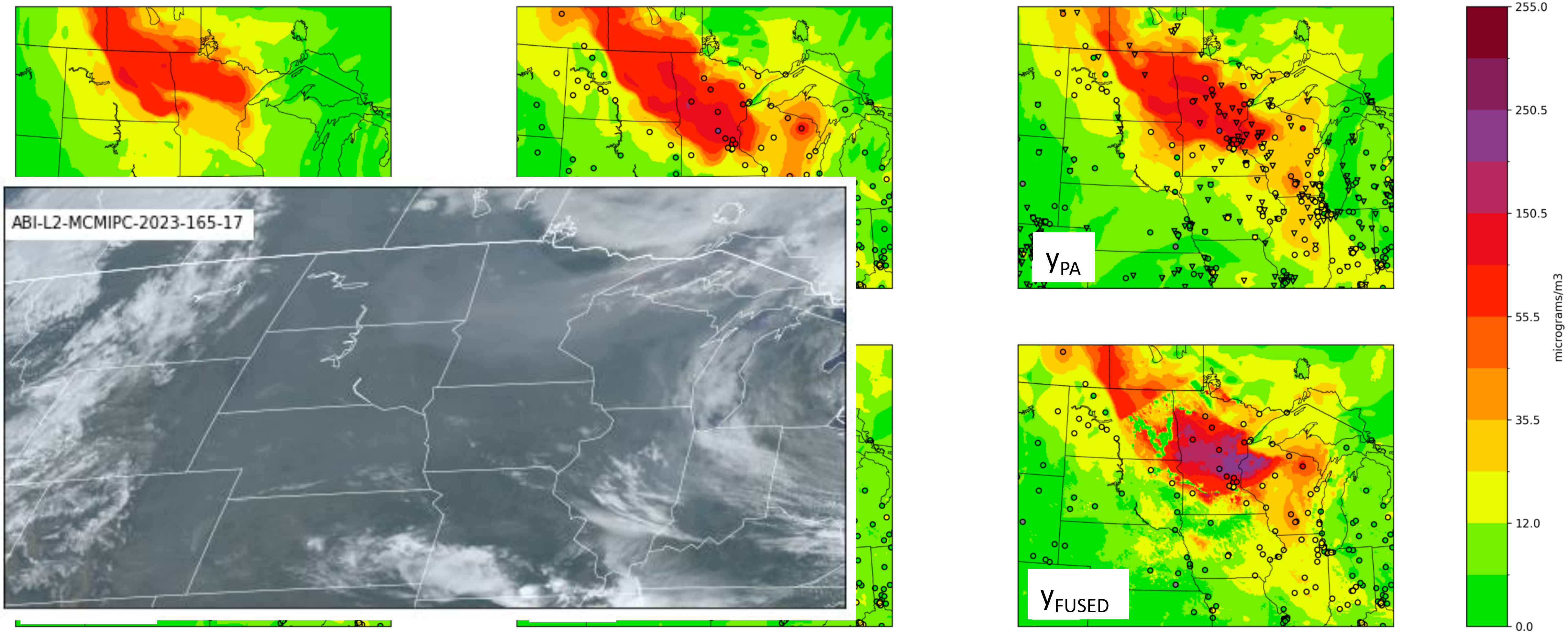
Los Angeles: 2023-06-14T17Z



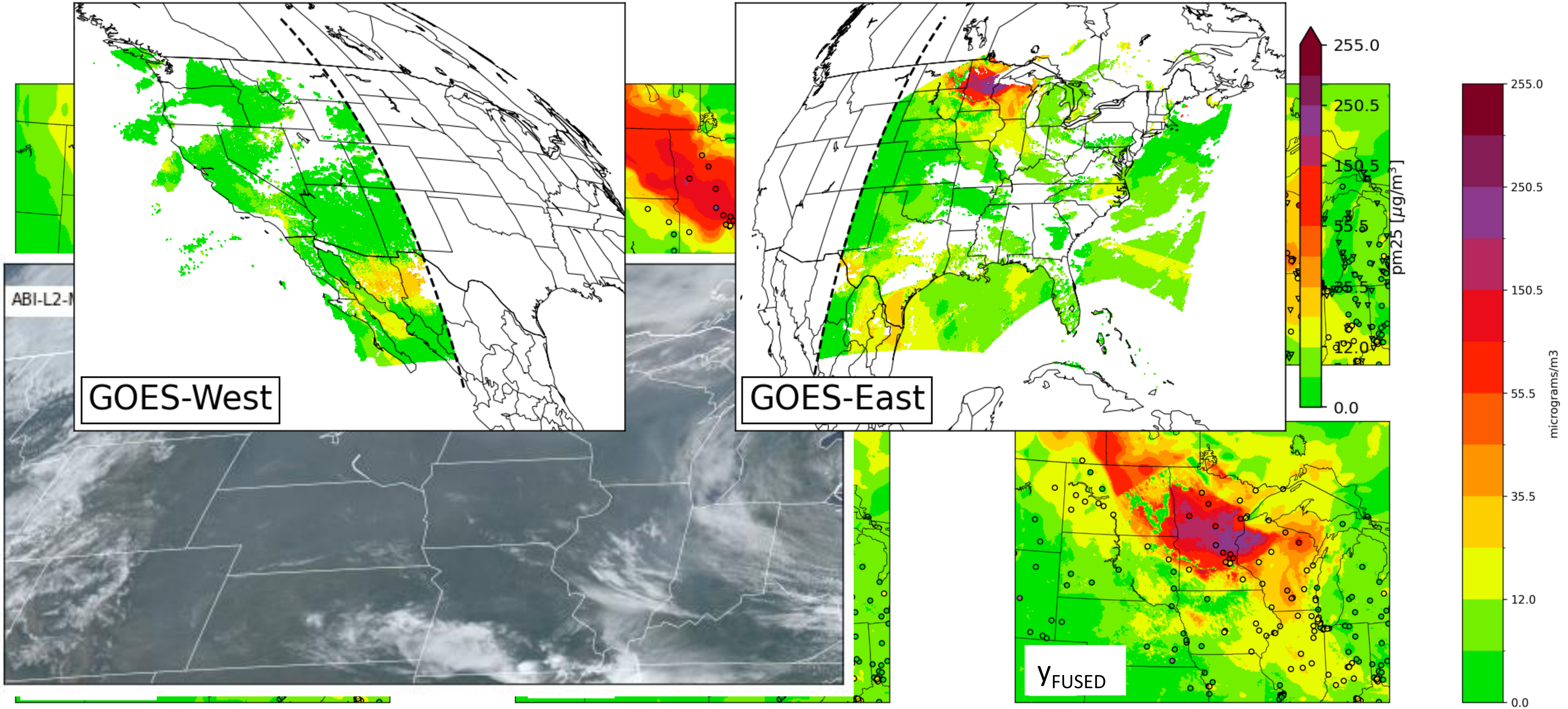
Canadian Wildfires: 2023-06-14T17Z



Canadian Wildfires: 2023-06-14T17Z

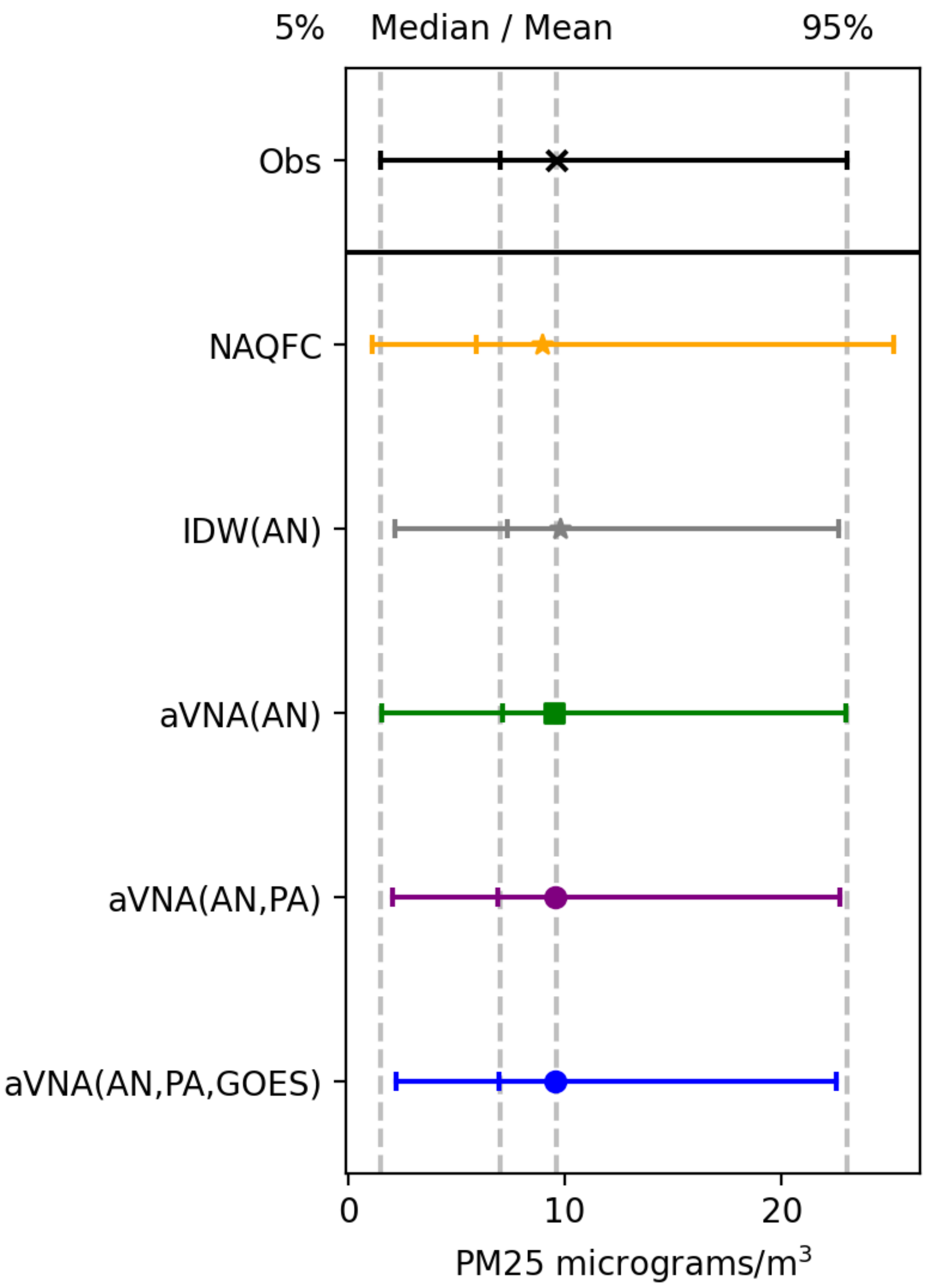


Canadian Wildfires: 2023-06-14T17Z



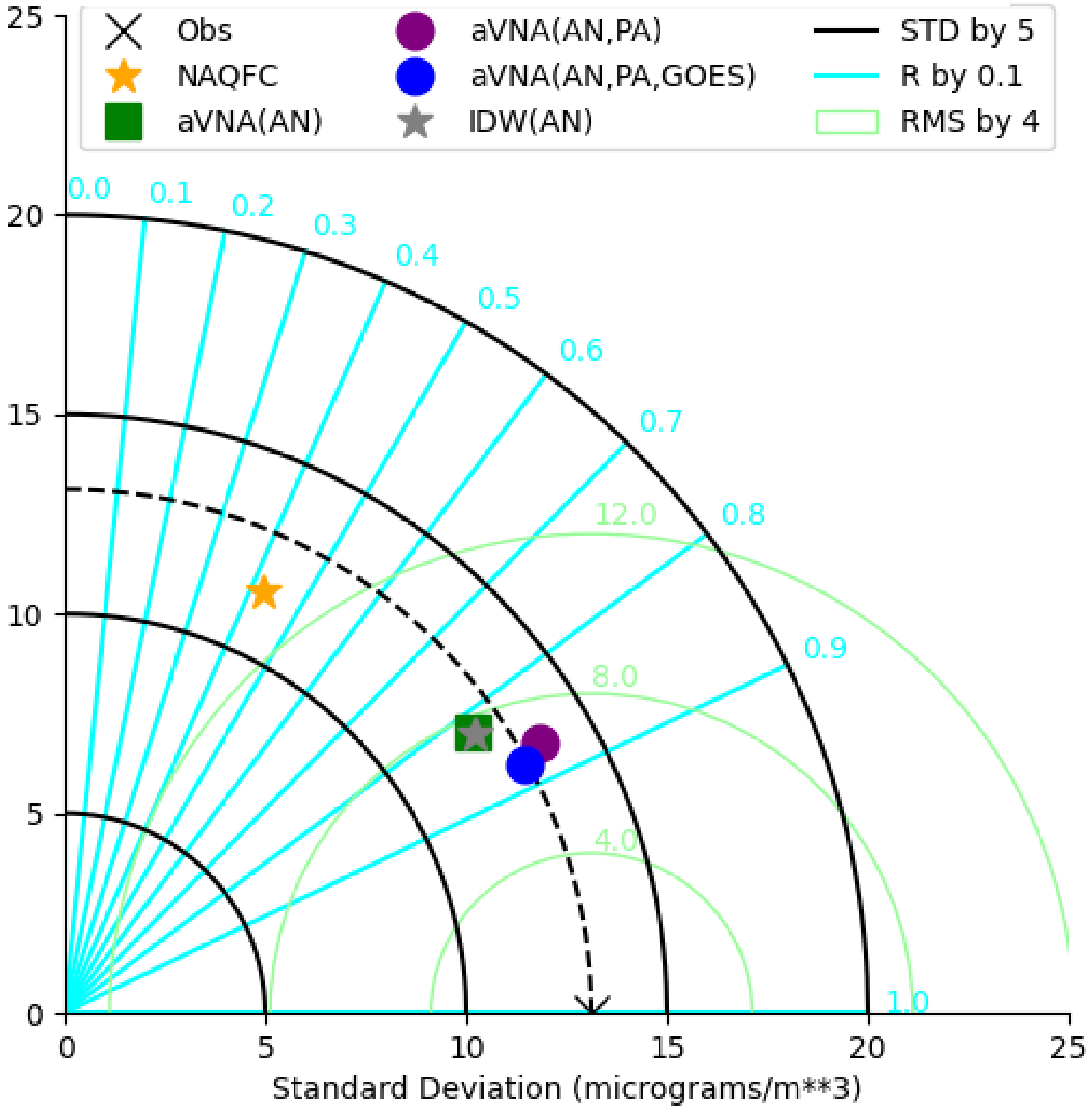
Evaluating the approach

- That was just one hour…
- Applied daylight from Jun 2023 to Sept 2023
  - IDW as in AirNow (*)
  - NAQFC from NOAA (*)
  - Corrected w/ AirNow: AN
  - Correction w/ AN and PurpleAir: AN+PA
  - Correction w/ AN, PA and GOES: AN+PA+GOES
- Predicted each AirNow monitor without that monitor in the fusion
  - n=1.3M = 12 h/d * 30 d/m * 3.75m * 1000 /h
- Statistics: Normalized Mean Bias, Normalized Mean Error, RMSE, Correlation.



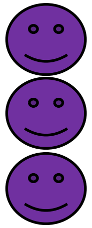
Performance Summary: June-Sept 2023 (daylight hours; n=1.3M)

- Multiple statistics matter
  - Pearson correlation (y-axis)
  - centered Root Mean Squared Error (x-axis)
  - Reproduction of standard deviation
- The NAQFC has the lowest correlation, the highest RMSE, and the worst standard deviation.
- The AirNow and IDW have similar correlation, AirNow has better standard deviation.
- The fusion with PurpleAir improves standard deviation, correlation, and root mean squared error.
- The fusion with GOES is even better.

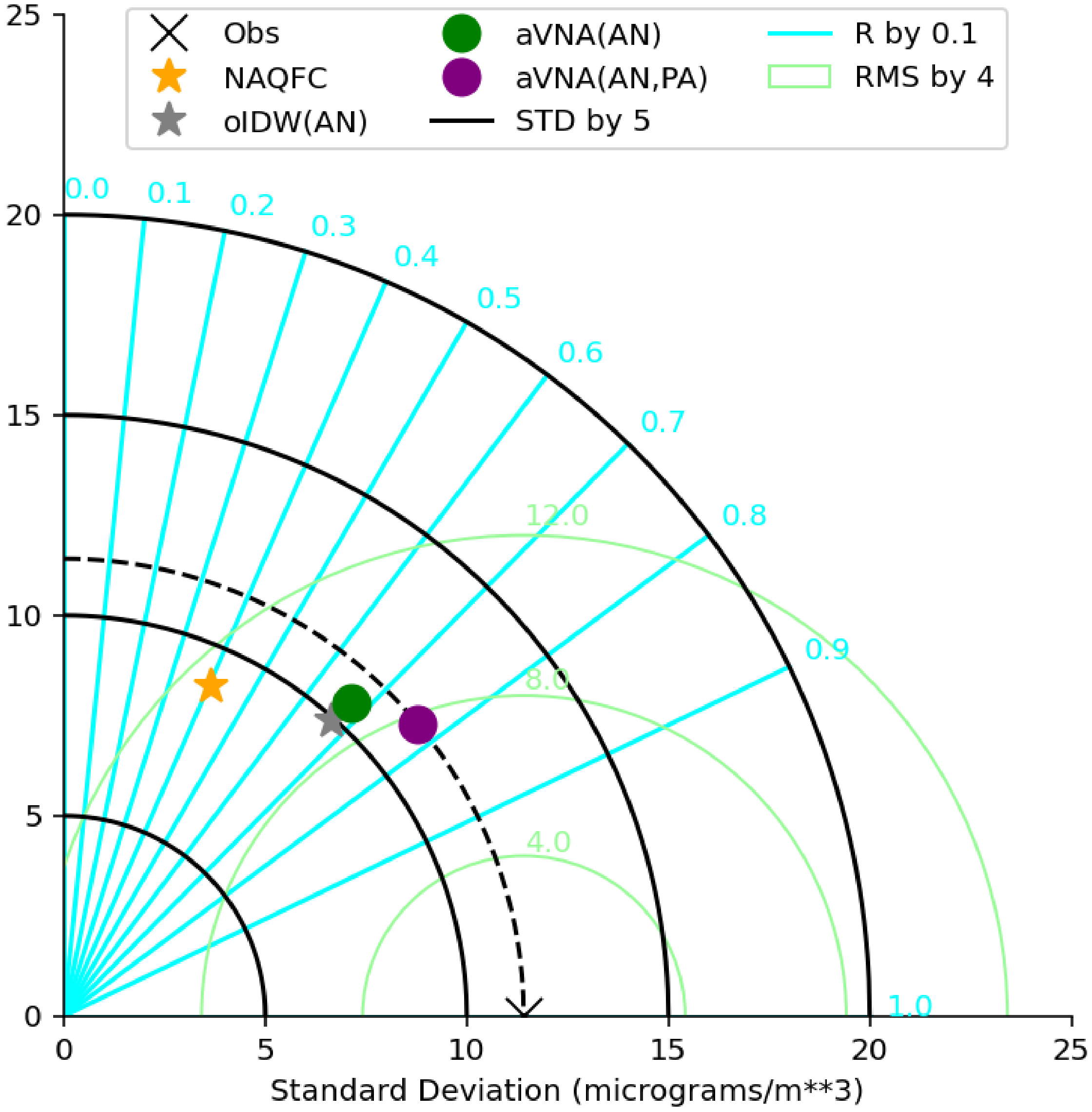


Performance Summary: June 2021-June 2022 (All hours; n=8M)

- Multiple statistics matter
  - Pearson correlation (y-axis)
  - centered Root Mean Squared Error (x-axis)
  - Reproduction of standard deviation
- The NAQFC has the lowest correlation, the highest RMSE, and the worst standard deviation.
- The AirNow and IDW have similar correlation, AirNow has better standard deviation.
- The fusion with PurpleAir improves standard deviation, correlation, and root mean squared error.

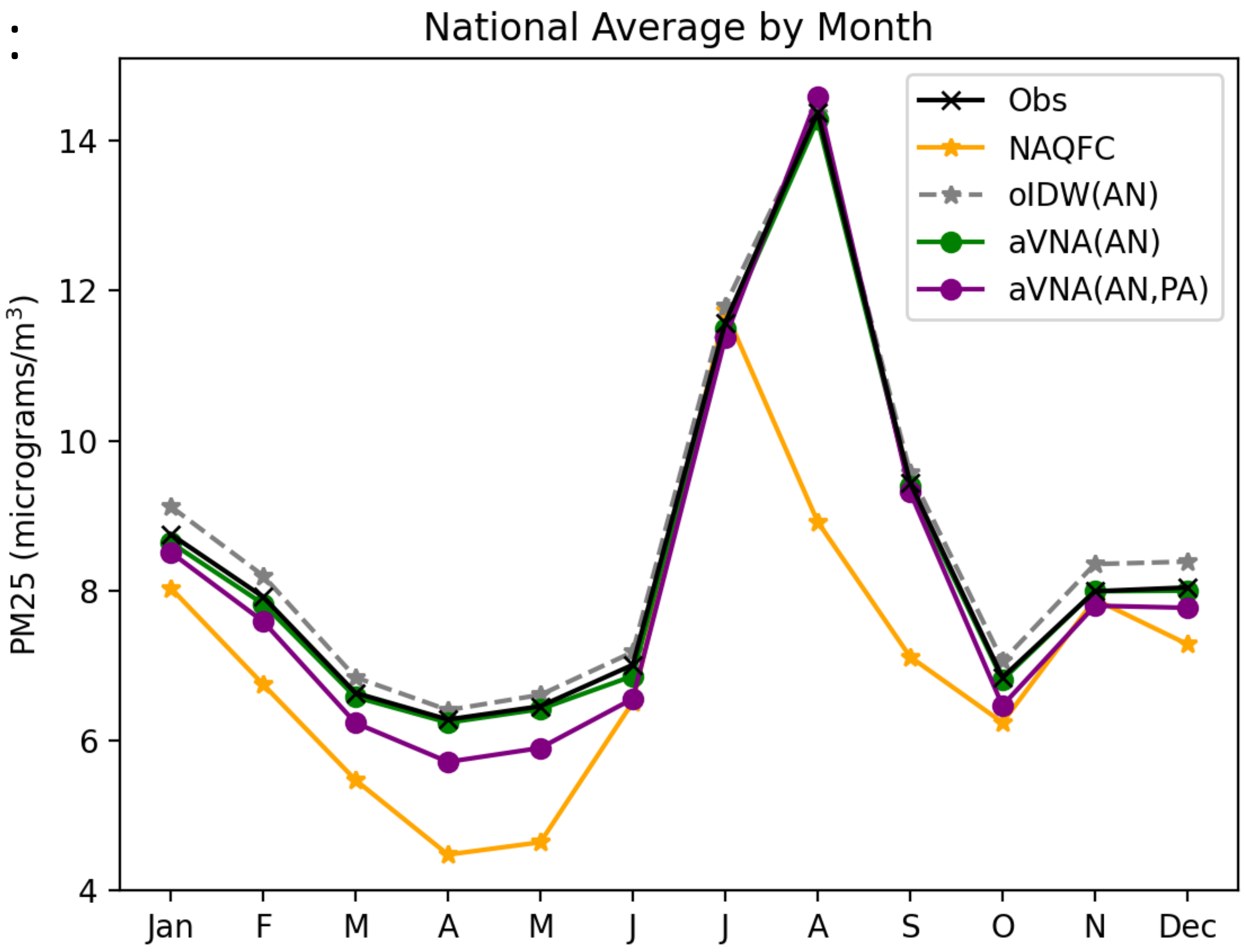


- Is the story more complex? When does one fail and the other succeeds?



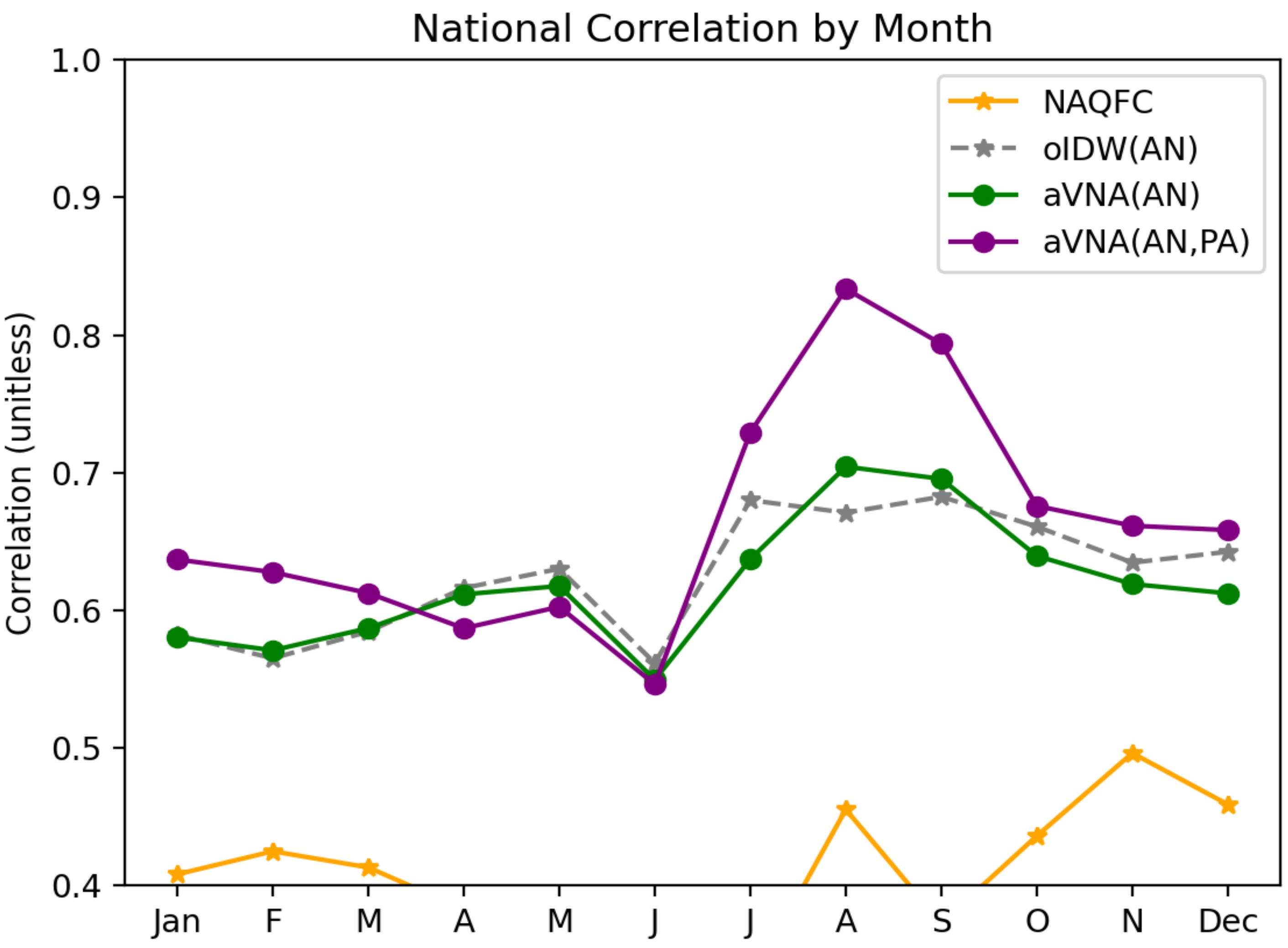
Leave-1-out Validation: National Average of Predictions

- This figure summarizes the concentration of PM2.5 over the months of the year by method.
- All methods peak during the fire season with the NAQFC peaking during July.
- Whereas the observations and other methods all peak during August.
- Remember, this is validation. In application, the prediction at the monitor is equal to the monitor.



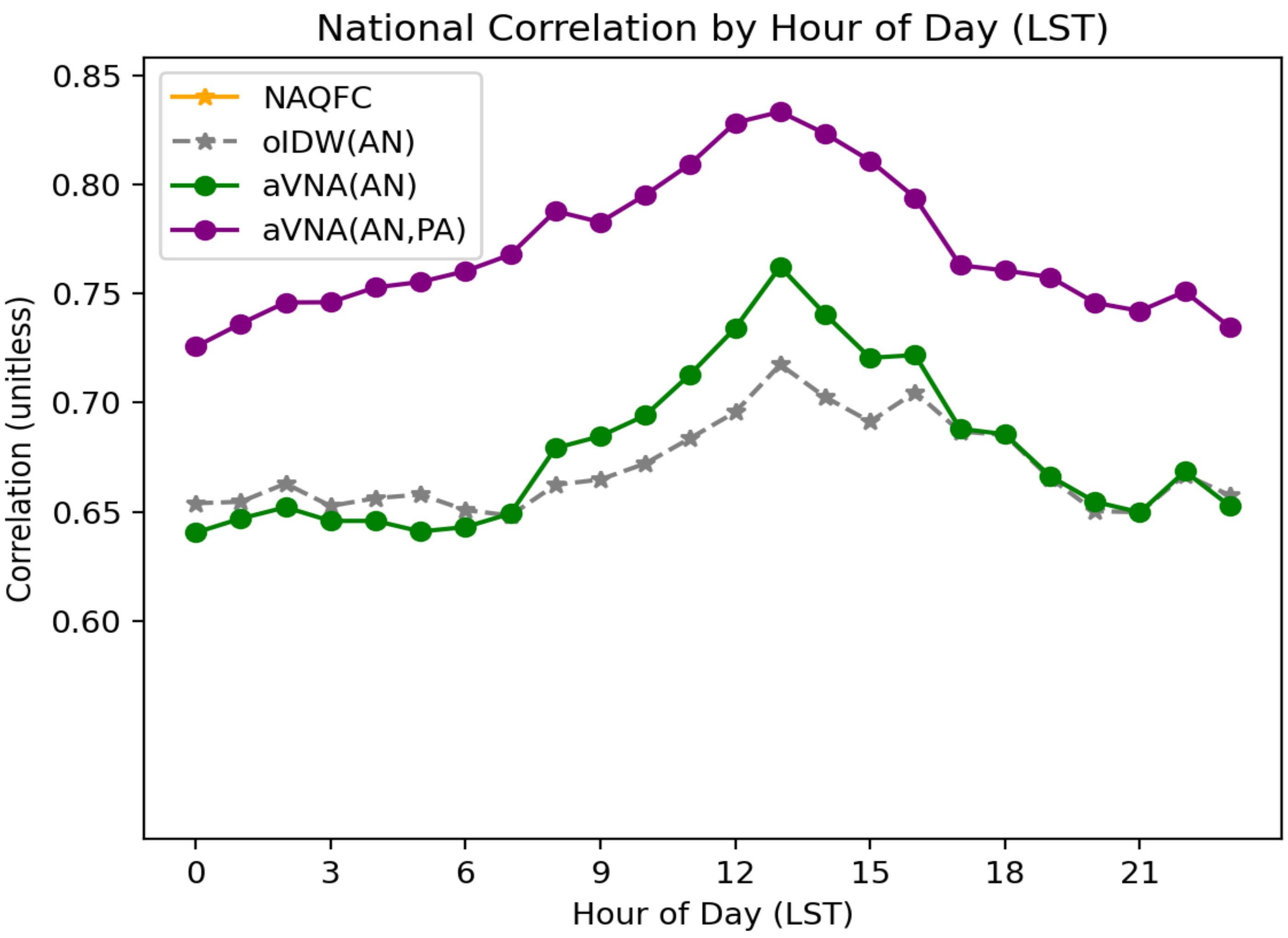
Leave-1-out Validation: National Correlation

- Incorporating PA improves the correlation especially during the fire season.
- aVNA(AN) has lowest correlation overall.
- aVNA(AN,PA) improves the correlation over the time of day.
- Remember, this is validation. In application, the prediction at the monitor is equal to the monitor.



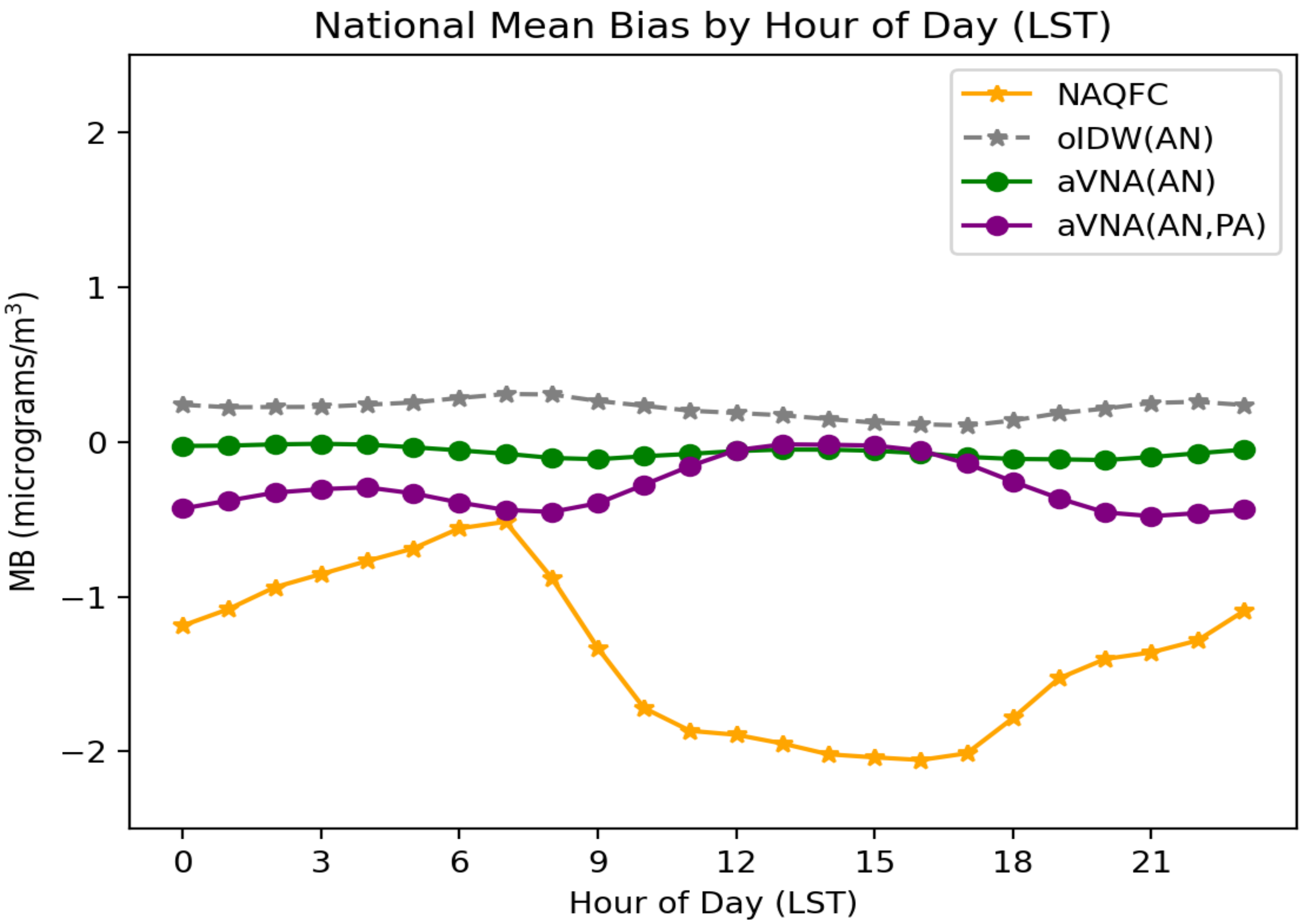
Leave-1-out Validation: National Correlation

- Incorporating PA improves the correlation especially during the fire season.
- aVNA(AN) has lowest correlation overall.
- aVNA(AN,PA) improves the correlation over the time of day.
- Remember, this is validation. In application, the prediction at the monitor is equal to the monitor.



Leave-1-out Validation: National Mean Bias

- oIDW and aVNA(AN) have the most consistent bias.
- aVNA(AN,PA) has highest bias at night but is still quite good.
    - Currently, we use a single bias correction for PurpleAir.
    - Humidity varies with time of day and may need more complex correction.
    - Also, FEM technologies are evaluated most strictly for daily average concentration.
- Remember, this is validation. In application, the prediction at the monitor is equal to the monitor.



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…

Questions?

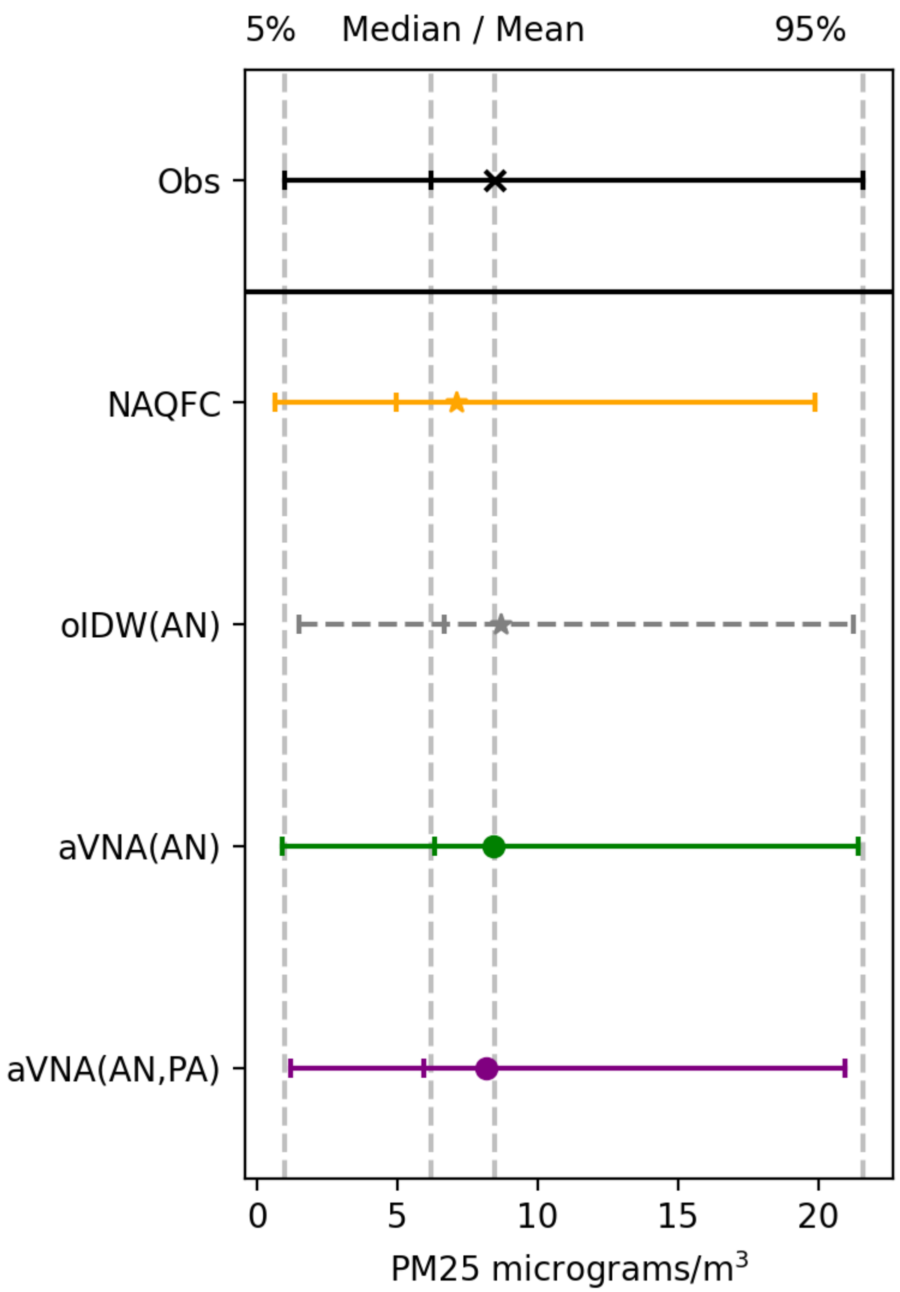
henderson.barron@epa.gov

Extra Slides

henderson.barron@epa.gov

Evaluating the approach

- That was just one hour…
- Applied hourly data from Jun 2021 to Jun 2022
  - IDW as in AirNow (*)
  - NAQFC from NOAA (*)
  - Corrected w/ AirNow: AN
  - Correction w/ AN and PurpleAir: AN+PA
  - ~~Correction w/ AN, PA and GOES: AN+PA+GOES~~
- Predicted each AirNow monitor without that monitor in the fusion
  - n=8M = 8760 h/y * 1000 /h
- Statistics: Normalized Mean Bias, Normalized Mean Error, RMSE, Correlation.



Alternative Ensemble Weighting Approaches

- We need a method to synthesize the products:
  - At this point, we have a potential of 4 fusion products
  - 2 bias correction methods (aVNA, eVNA), 2 data sources (AirNow, PurpleAir)
- Geographically Varying Weights (GVW)
  - Similar to Requia[1], but implemented like Skipper[2]
  - $Y_{fused} = \Sigma_i \alpha_i Y_i$
  - $\alpha_i = c_i + \beta_{i,0} x + \beta_{i,1} y + \beta_{i,2} d_{PA} + \beta_{i,2} d_{AN}$
  - c_i and all β are fit using least squares regression
- Random Forest Regression (RF)
  - Features: x, y, d_{PA}, d_{AN}, Y_i estimates from leave-one-out cross-validation
  - Configuration: Minimum 20 features for a split; 100 trees.
- **Not Shown**: Few day tests show
  - Both GVW and RF have better correlation than current approach; RF best.
  - But…. Current approach has better standard deviation than either
- This will likely need to be revisited when bringing in the GOES-PM25.

[1] Requia et al. https://pubs.acs.org/doi/10.1021/acs.est.0c01791
[2] Skipper et al. https://doi.org/10.1021/acs.est.0c08625

Leave-1-out Validation: National Mean Bias

- GOES shows structure in the bias that is associated with long-distance extrapolation…
- The fusion actually doesn't use those cells (too far away)
- Remember, this is validation. In application, the prediction at the monitor is equal to the monitor.