AirFuse

A multi-pollutant fusion system

Barron H. Henderson[1,4]

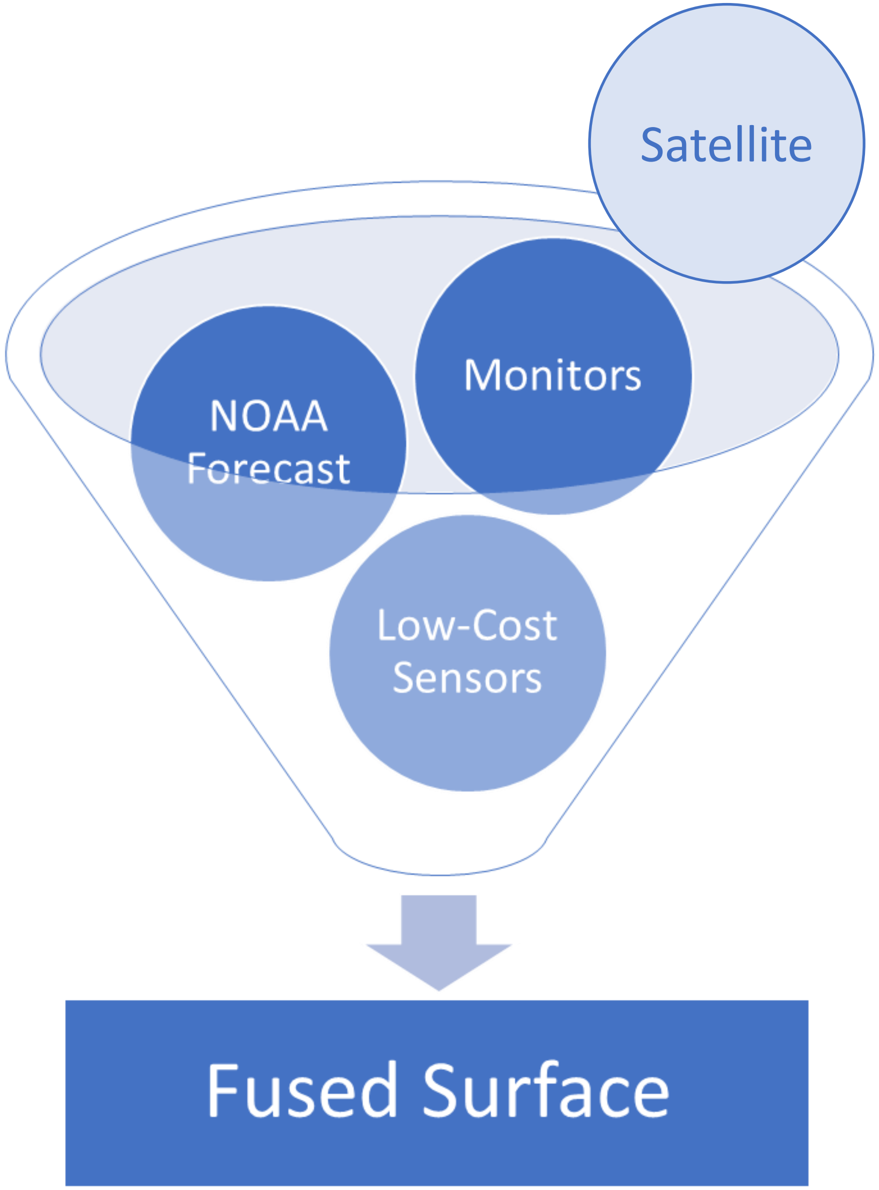
Contributors: Phil Dickerson[1,4] Pawan Gupta[2,4], Shobha Kondragunta[3,4], Yang Liu[4,6], Meng Qi[6], Alqamah Sayeed, Hai Zhang[3,4], Janica Gordon[5], Halil Cakir[1], Brett Gantt[1], Benjamin Wells[1], Marcus Hylton[7], Youngsun Jung[9], and the HAQAST AirNow Teams[4]

[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]Emory University; [7]AirNow Data Management Center; [9]National Oceanic and Atmospheric Administration / National Weather Service

AirFuse: hourly maps of $PM_{2.5}$ and ozone for AirNow

Fuses best available data sources

1. NOAA Forecast w/bias correction
2. AirNow monitors (~1000 per hour)
3. PurpleAir sensors (~9k per hour)
4. Near-real-time satellite observations
   - Recent development by NOAA/NESDIS/STAR
   - NASA HAQAST project connecting AirNow to NOAA geostationary satellite data



Ozone too, but no sensor or satellite data.

Available on airnowtech.org since March 2024

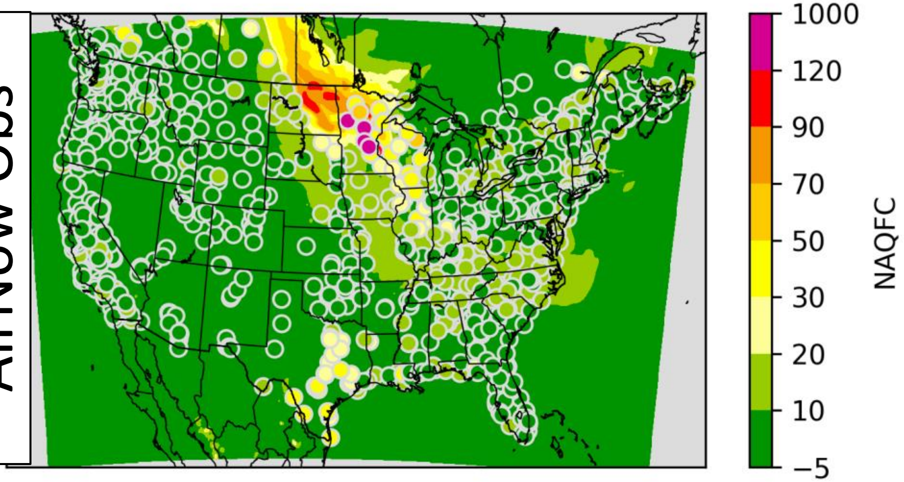
Calculate the bias of NOAA's forecast

Air Quality Forecast Guidance Viewer (*)

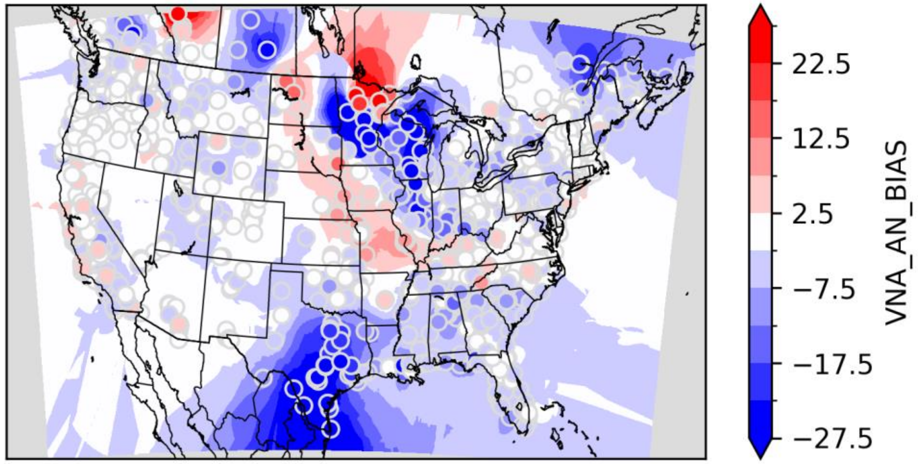
https://digital.mdl.nws.noaa.gov/

- NOAA's National Air Quality Forecast
  - CMAQ forecasts concentration
  - Kalman filter analog system forecasts bias
  - Bias interpolated to grid cells (Glahn et al.) to "correct" model
- Why not use this directly?
  - AirFuse corrects using multiple sources.
  - In AirNow bias has already happened, so the correction can be updated.
- Identify bias based on ***past*** observations
  - Bias using near-real-time observations.
  - Interpolation using Delaunay diagram

NAQFC+ AirNow Obs



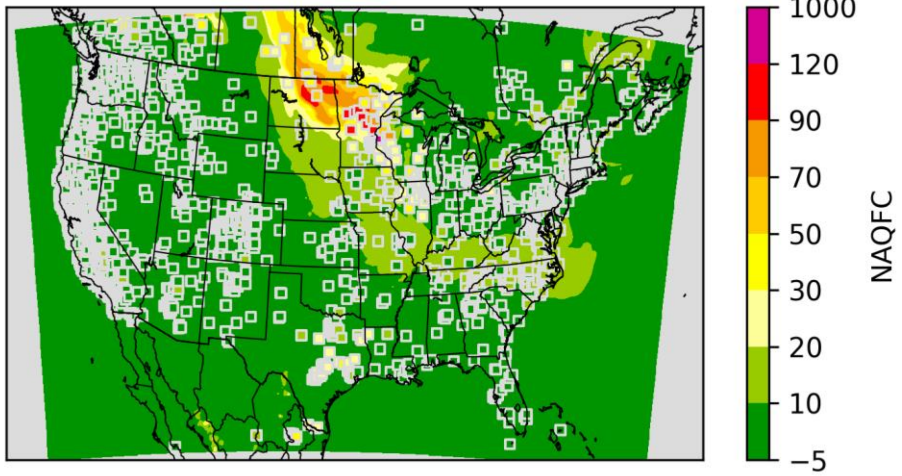
Real-time Bias



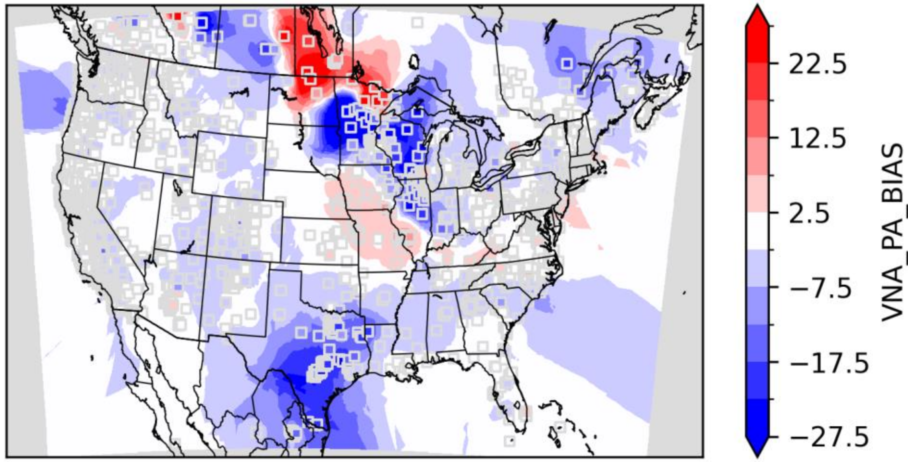
Calculate bias from sensors

- Schulte et al 2020 used PurpleAir
  - Model Correction : $Y = M_n - Krig(M_n – O_n)$
  - Observations (O) from both AirNow and PurpleAir
  - Improved validation statistics!
- Using the EPA national 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)
- Identify bias based on ***past*** observations
  - Bias using near-real-time observations.
  - Interpolation using Delaunay diagram

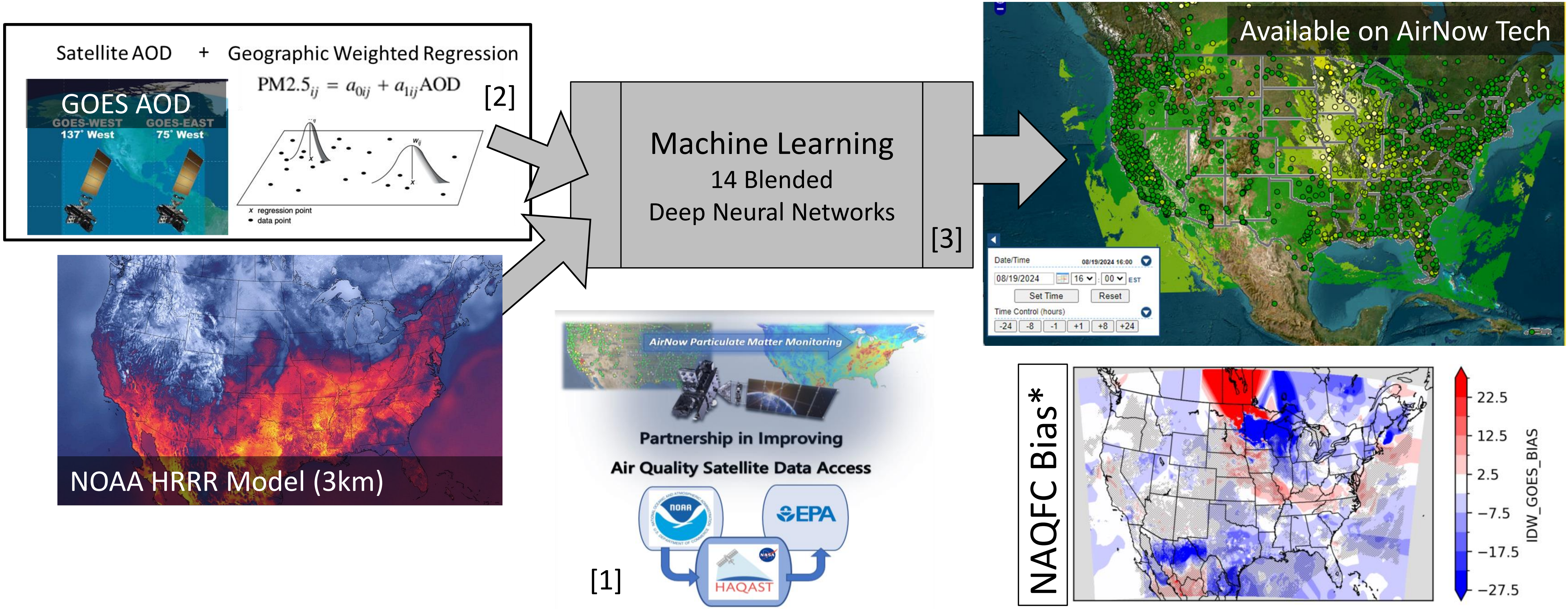
NAQFC+ with PurpleAir



NAQFC+ Bias According to PurpleAir



Calculate bias from geostationary satellite



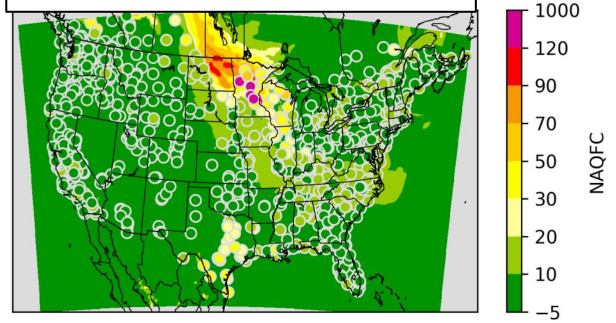
HAQAST Tiger Team Leads: Pawan Gupta and Yang Liu
Partners: Phil Dickerson and Barron Henderson (EPA), and Shobha Kondragunta (NOAA)

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.

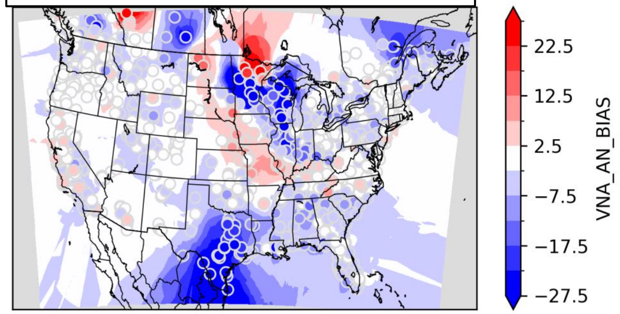
Hourly National-scale Fusion Ensemble

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

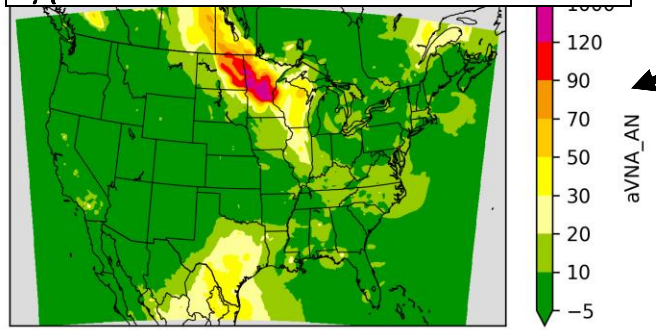
Model



Monitor Bias



Y_A = f(model, monitor)



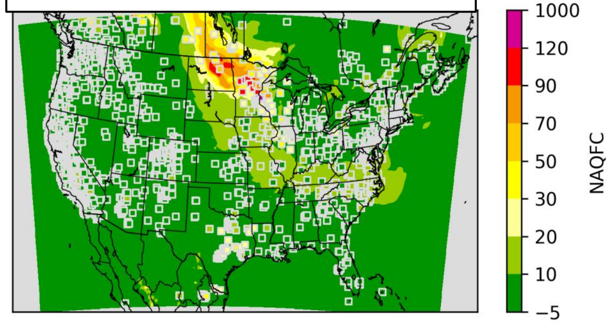
plicative corrector of this type is called extended VNA (eVNA)

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

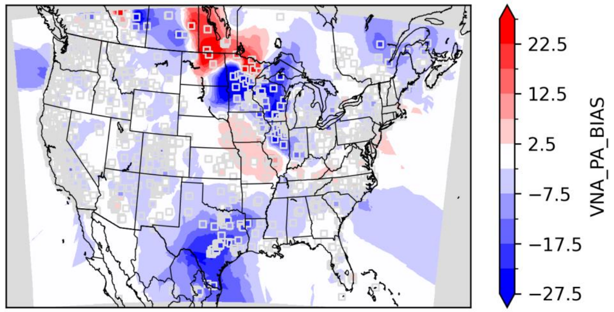
Hourly National-scale Fusion Ensemble

- One layer from AirNow monitor (Y_A) ***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

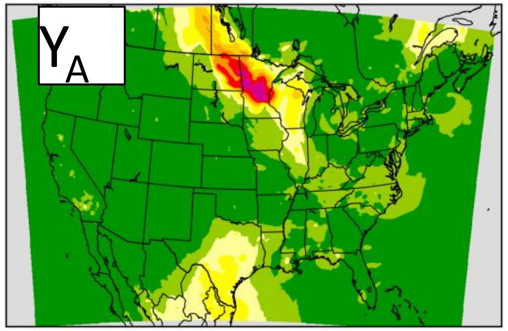
Model



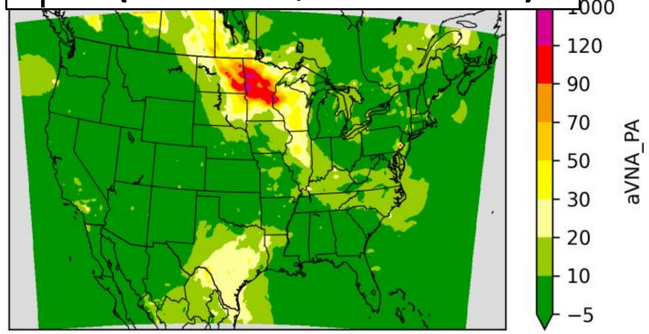
PurpleAir Bias



Y_A

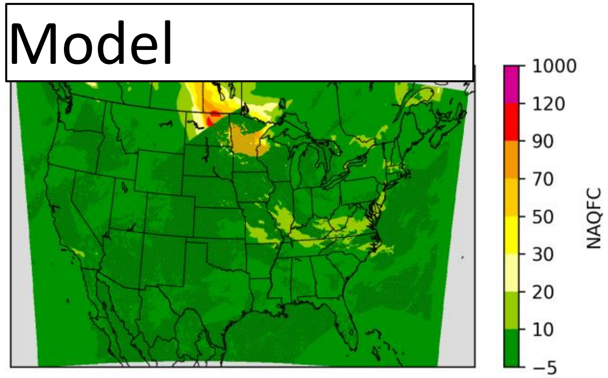


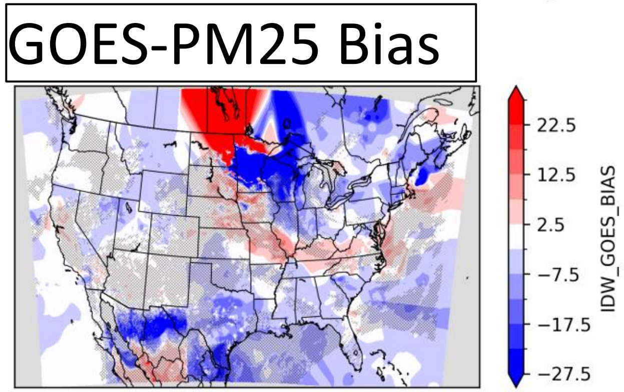
Y_P=f(model, sensors)



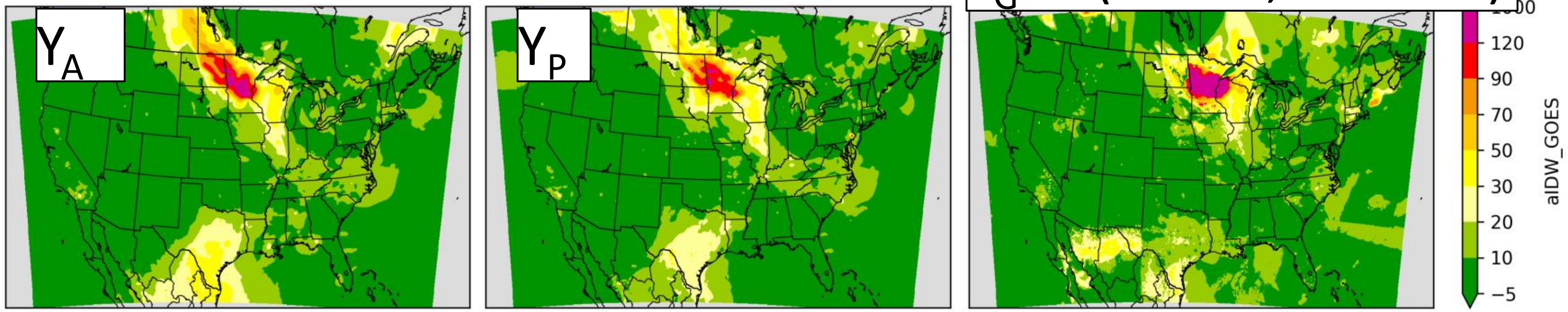
Hourly National-scale Fusion Ensemble

- One layer from AirNow monitor (Y_A) ***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.





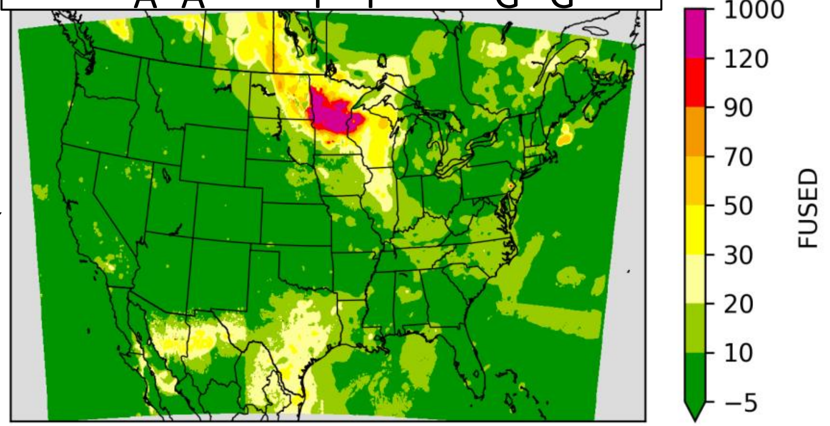
Y_G = f(model, satellite)

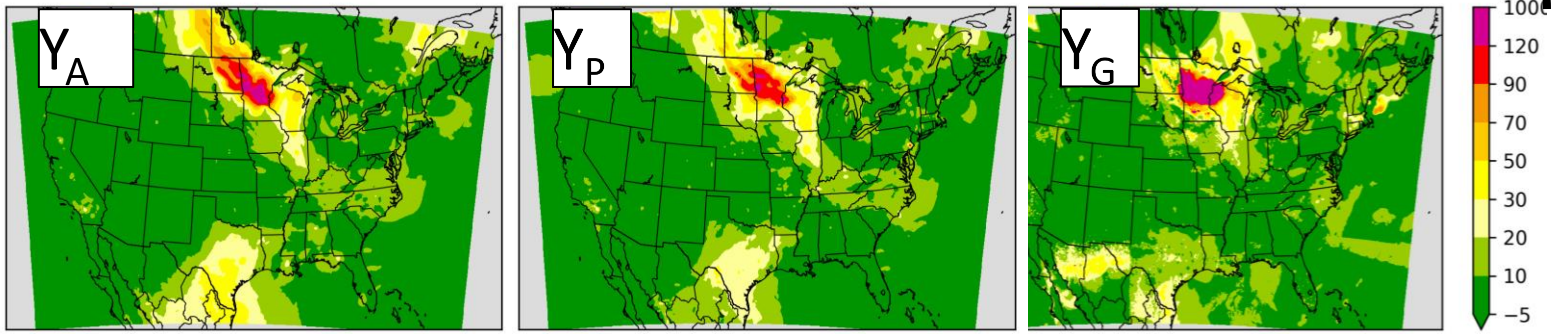


Hourly National-scale Fusion Ensemble

- One layer from AirNow monitor (Y_A) ***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 (w_A, w_p, w_g)

$$w_A Y_A + w_P Y_P + w_G Y_G$$

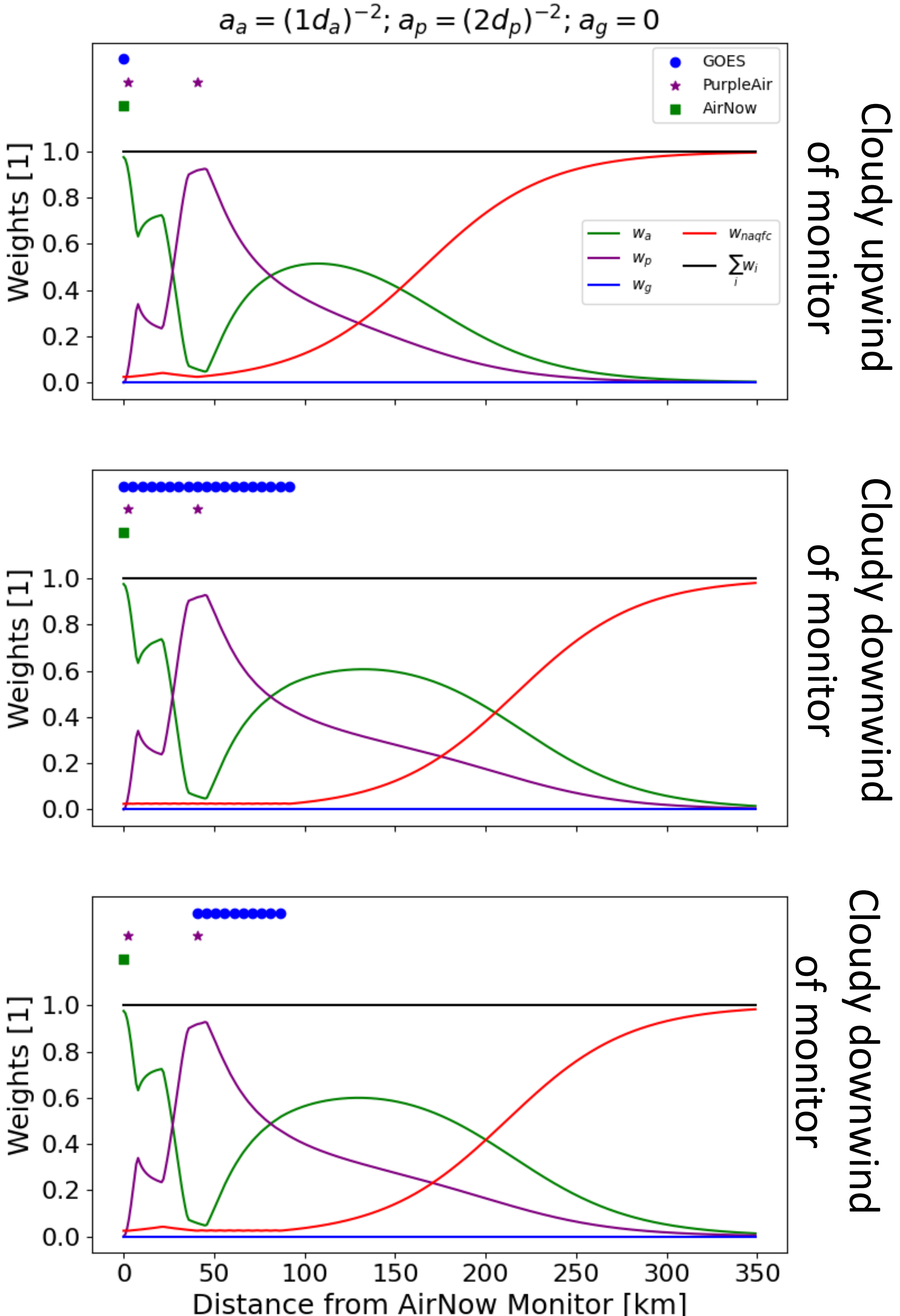




Weight the ensemble of surfaces on distance

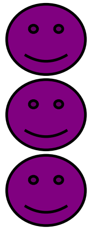
- Three scenarios to illustrate weights
  - AirNow on at the left.
  - PurpleAir near and a bit downwind.
  - GOES-PM25 coverage varies
- Pilot project began without satellite data
- $Y = w_A Y_A + w_P Y_P + w_G Y_G + w_N Y_N$
  - $a_a = (1 \times d_{AN})^{-2}$
  - $a_p = (2 \times d'_{PA})^{-2}$: $d'_{PA} = \max(d_{PA}, 3.6)$
  - $a_g = 0$
  - Normalize
    - $w_n = 1 / (1 + \exp(k * (d_{apg} - x_0)))$
    - $w_a = a_a (1 - w_n) / (a_a + a_p + a_g)$
- Performance
  - Adding PurpleAir improved performance.
  - Optimized weight of PurpleAir
  - Statistical performance is good even without satellite
  - But, the AirNow monitor is no the best data downwind.

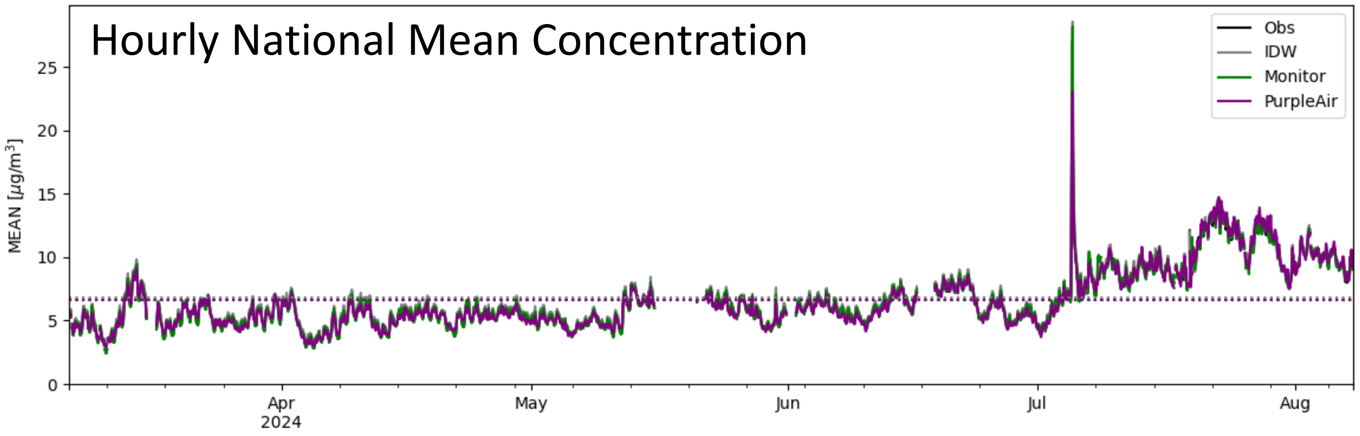
*Optimized parameters

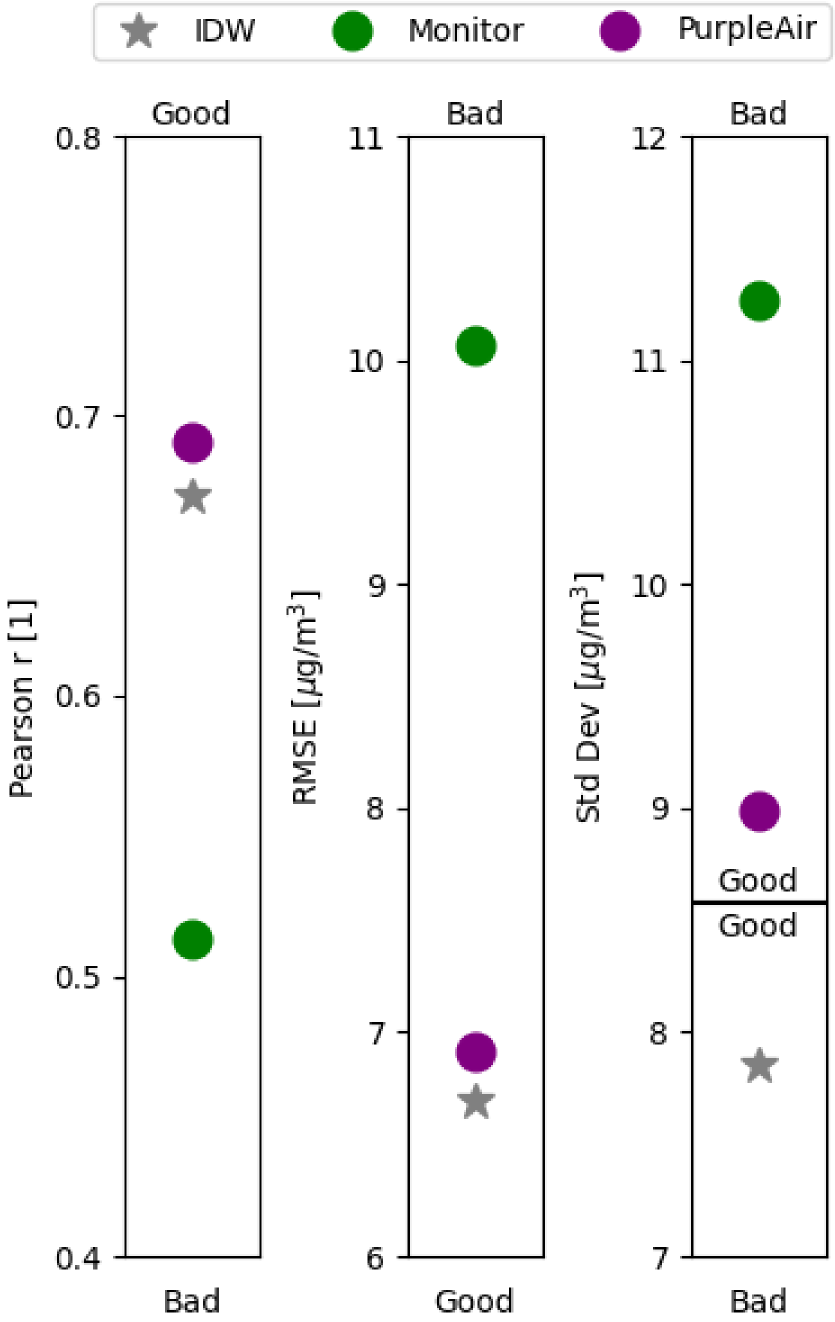


Pilot Validation Summary

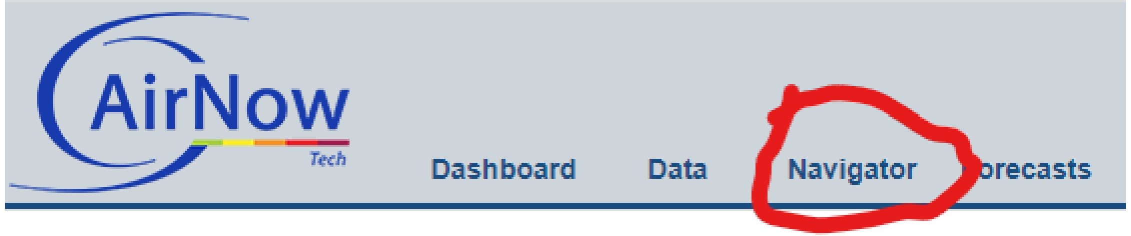
- Using only monitors aVNA performing worse than IDW
- Including PurpleAir improves:
  - Prediction standard deviation,
  - Prediction correlation, and
  - Root mean squared error.



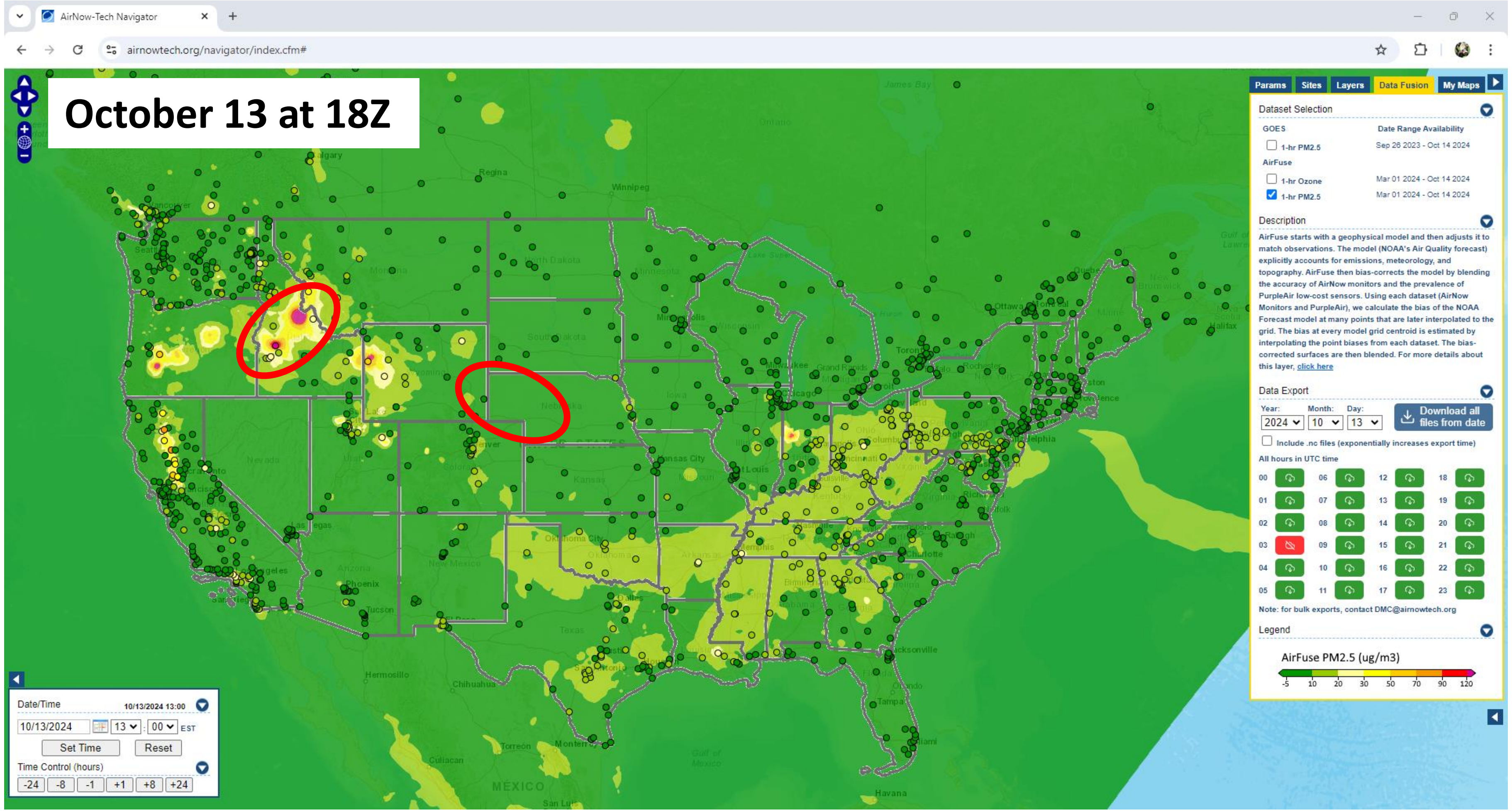




Data on: https://airnowtech.org/

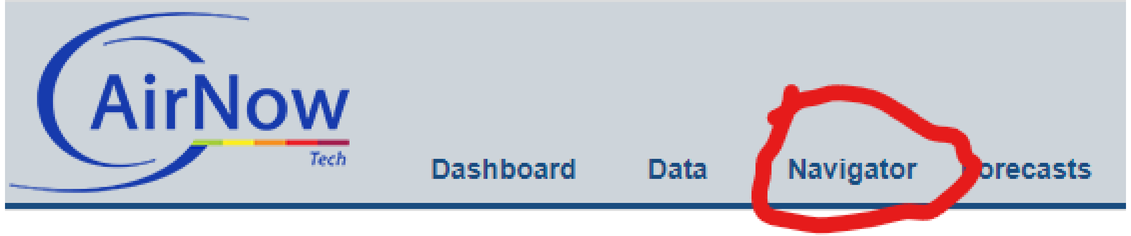


- Login to ANT
- Choose Navigator
- On Navigator, choose the Data Fusion tab.
- Select an AirFuse or GOES Layer

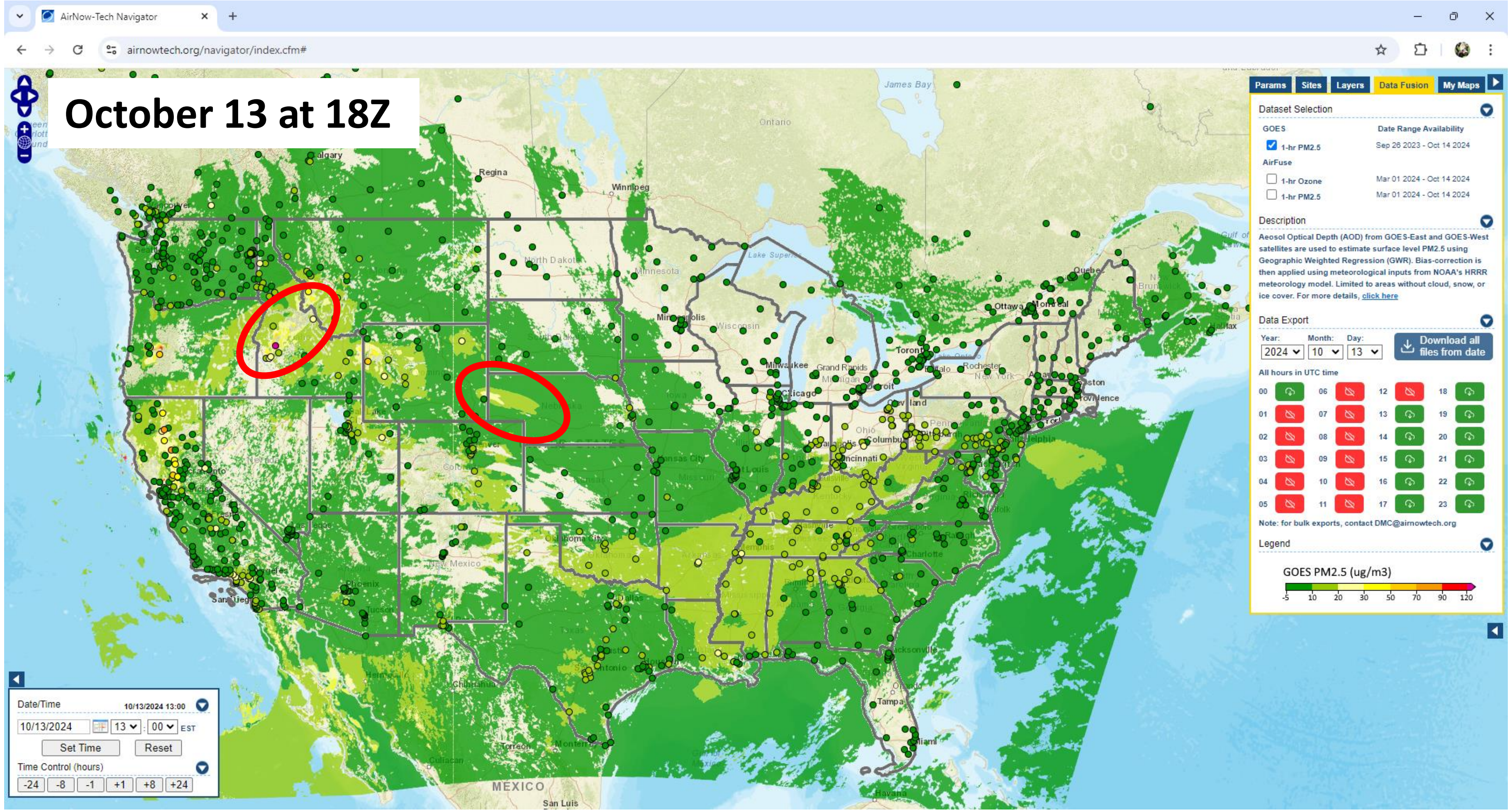


Thanks AirNow team!

Data on: https://airnowtech.org/



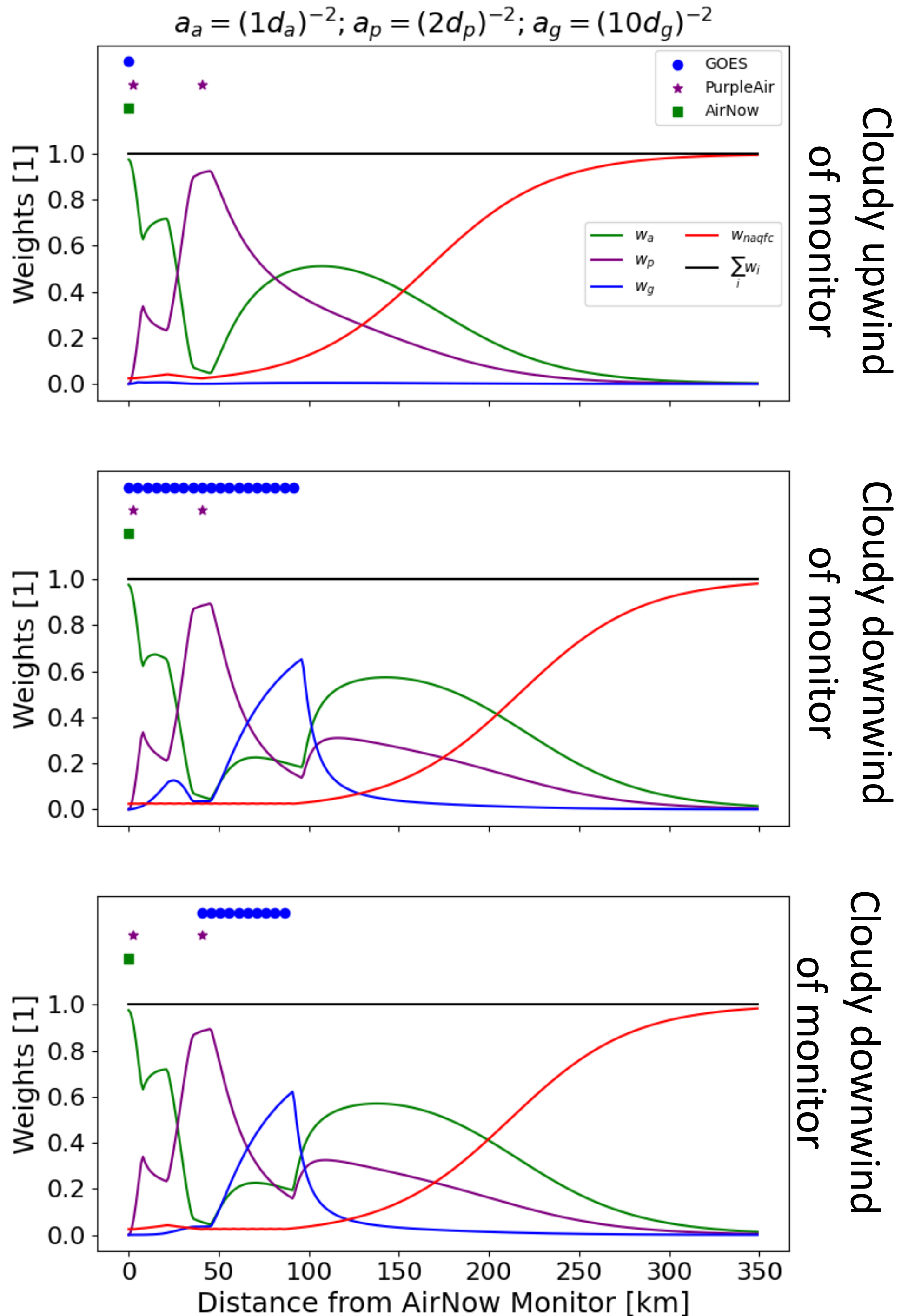
- Login to ANT
- Choose Navigator
- On Navigator, choose the Data Fusion tab.
- Select an AirFuse or GOES Layer



Thanks AirNow team!

Weight the ensemble of surfaces on distance

- Three scenarios to illustrate weights
  - AirNow on at the left.
  - PurpleAir near and a bit downwind.
  - GOES-PM25 coverage varies
- Including satellite using the same functional form
- $Y = w_A Y_A + w_P Y_P + w_G Y_G + w_N Y_N$
  - $a_a = (1 \times d_{AN})^{-2}$
  - $a_p = (2 \times d'_{PA})^{-2} : d'_{PA} = \max(d_{PA}, 3.6)$
  - $a_g = (10 \times d'_G)^{-2}: d'_G = \max(d_G, 3.6)$
  - Normalize
    - $w_n = 1 / (1 + \exp(k * (d_{apg} - x_0)))$
    - $w_i = a_i (1 - w_n) / (a_a + a_p + a_g)$
- Performance
  - Statistical performance is better with satellite!
  - Created artificial "hard edges" when satellite and AirNow/PurpleAir diverge.
- Does it matter? Need method of identifying artifacts and then a new method to reduce artifacts.



*Optimized parameters

Finding anomalies in hourly surfaces

Current anomaly detection criteria:

Step 1:

1.1 Large AN-GOES difference (> 55.5 μg/m^3)

1.2 Edge detection applied on FUSED results and GOES_WGT

Step 2: Morphological image processing

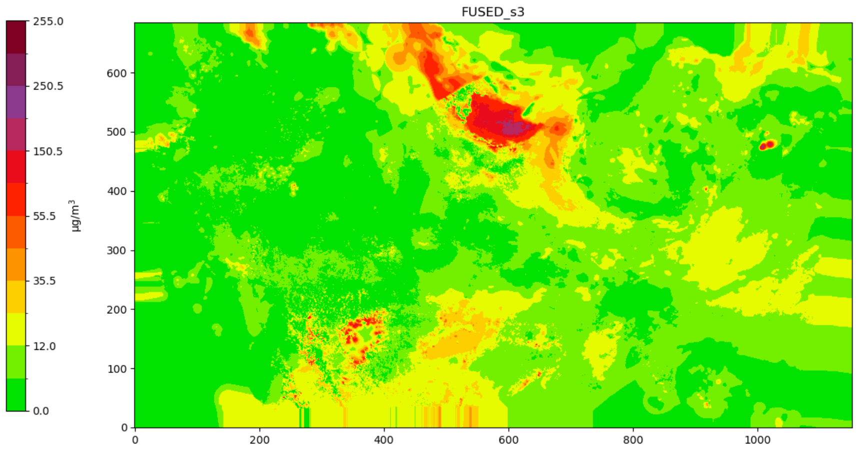
- fill the edges to capture the anomaly area

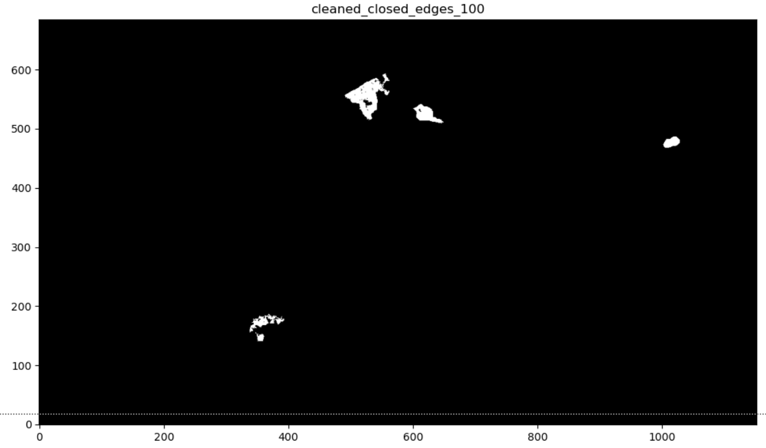
Step 3: filter the closed edge by large AN-GWR difference again

Type 1: AN_WGT < 0.4
- Type 1.1: AN >> GOES
- Type 1.2: AN << GOES

Type 2: AN_WGT > 0.8

Step4: Only keeping detected areas with connected size larger than 100



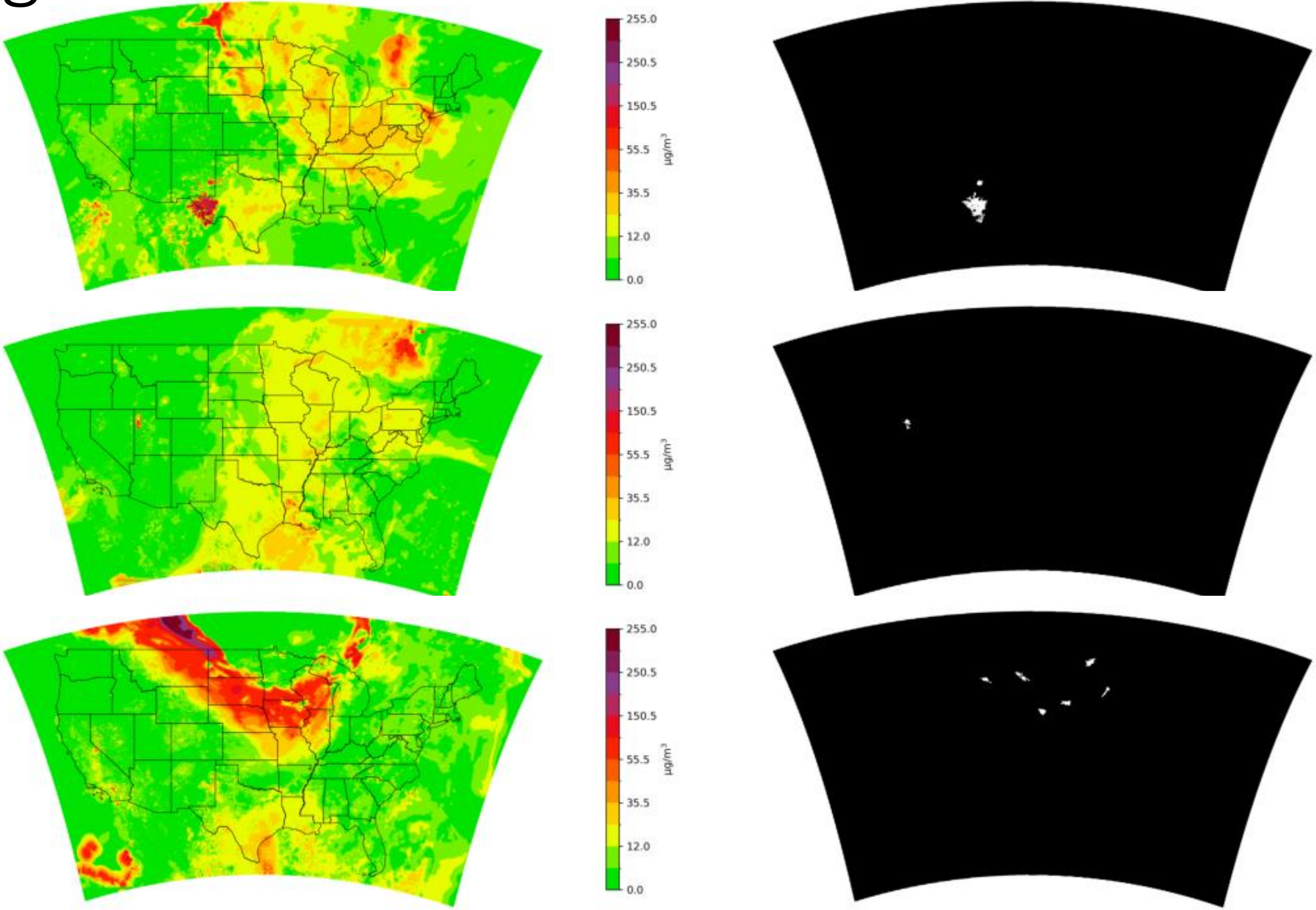


source: Meng Qi at Emory University

source: Meng Qi at Emory University

Example Edge Detections

- Emory University developed an edge detection algorithm to process years of data.
- Plume in Texas Pan Handle
  - 2023-06-18T01Z
- Interesting feature in west Utah
  - 2023-06-20T18Z
- Fire plume from Alberta Canada with missing data
  - 2023-07-15T23Z



Number of anomalies detected

Type 1: AN_WGT < 0.4
- Type 1.1: AN >> GOES
- Type 1.2: AN << GOES

Type 2: AN_WGT > 0.8

- In 2020, anomalies increase in the fire season when AOD retrievals are difficult.
  - Total detections 1472
  - NOAA improved QA and updated DNN
- In 2023, fewer detections (532)
  - More often during spring and early summer.
- Path forward
  - Use edge detection to identify the types of situations that cause artifacts.
  - Likely use edge detections to constrain interpolations.
  - Find weights that smoothly transition even with large differences between products.

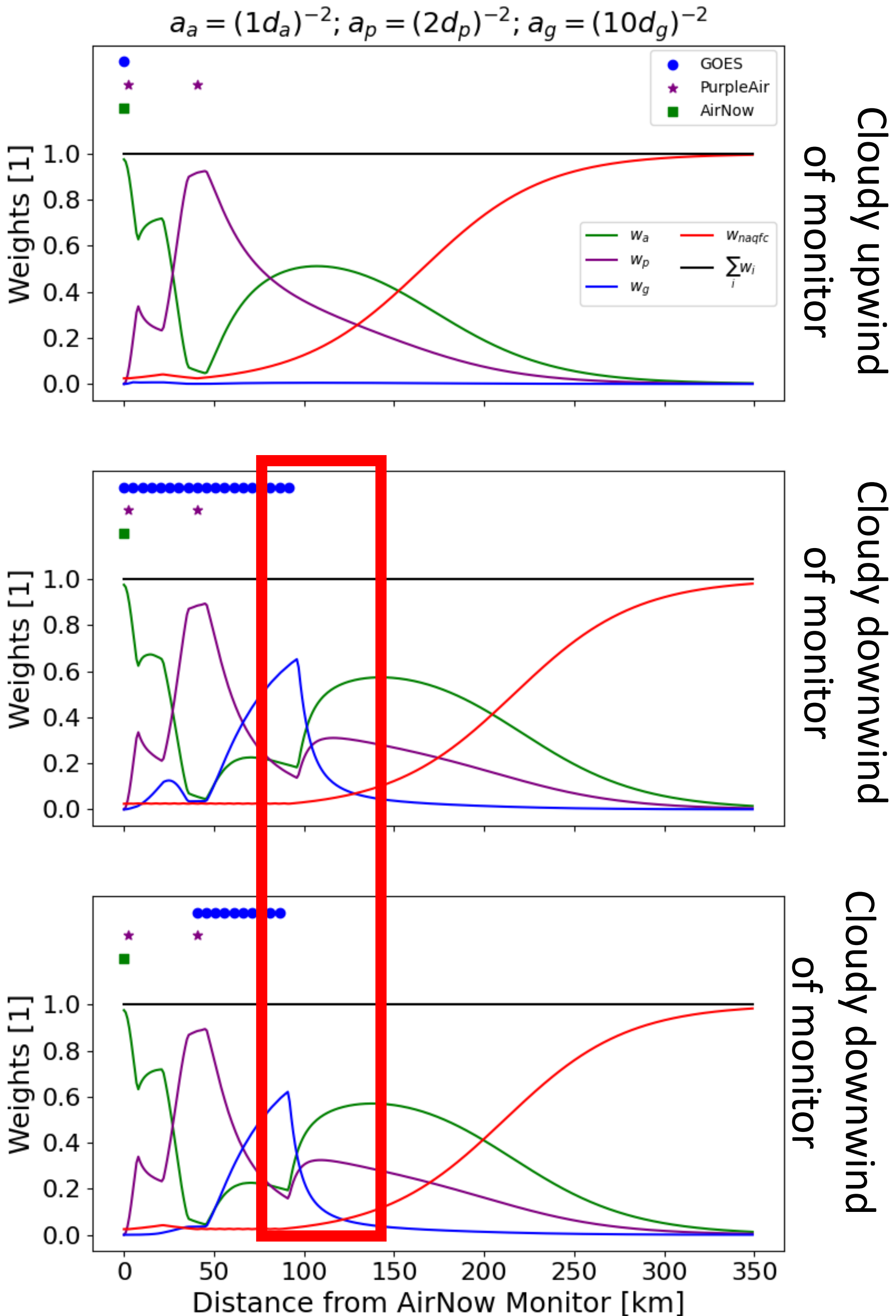
Month	**2020**			**2023**		
	Type 1.1	Type1.2	Type 2	Type 1.1	Type1.2	Type 2
5	0	2	29	108	130	77
6	3	116	21	124	195	56
7	22	67	9	79	186	72
8	291	401	64	192	192	40
9	1021	615	497	26	117	26
10	135	120	24	3	32	13

source: Meng Qi at Emory University

Reminder of weighting scheme

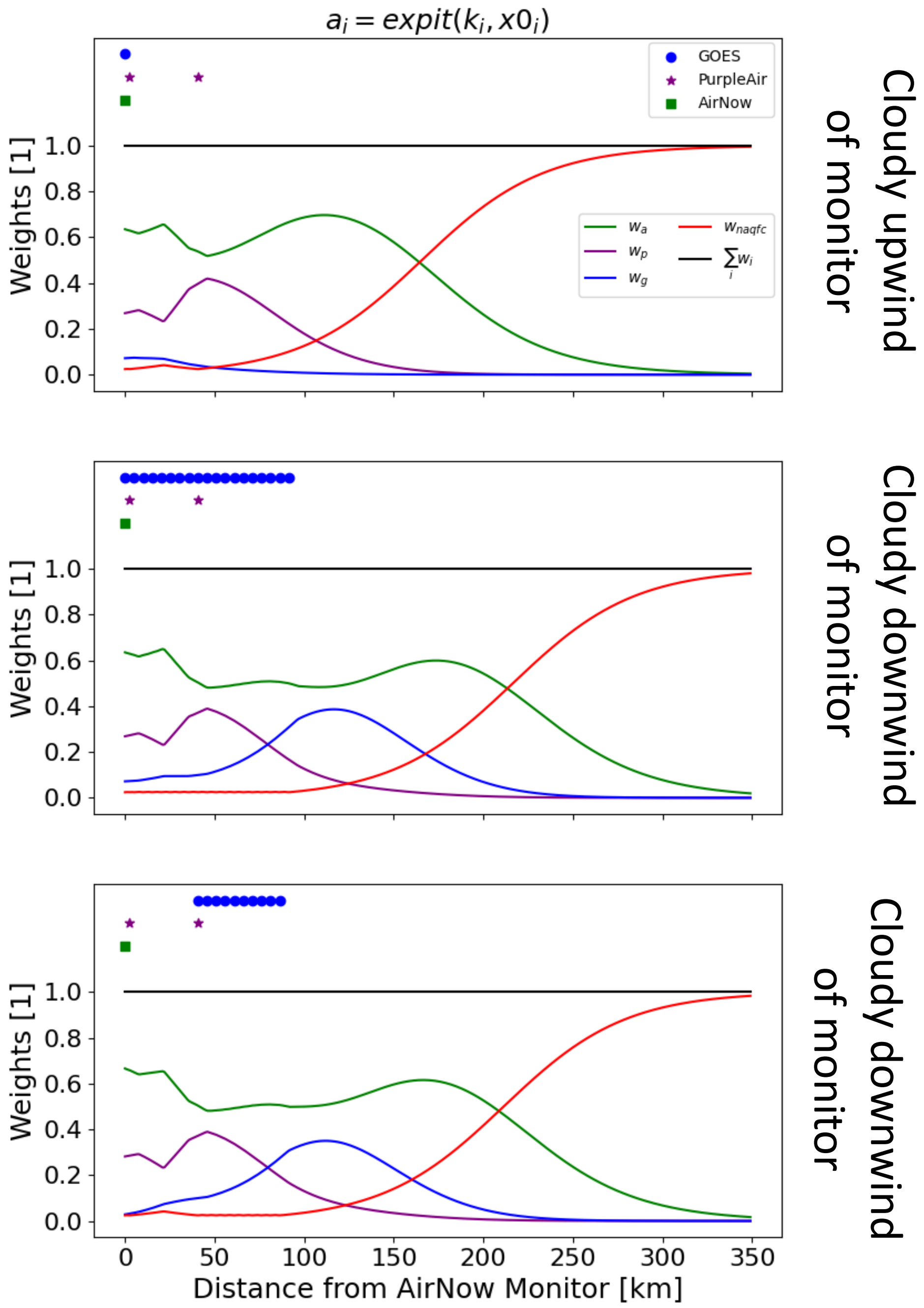
- Three scenarios to illustrate weights
  - AirNow on at the left.
  - PurpleAir near and a bit downwind.
  - GOES-PM25 coverage varies
- Including satellite using the same functional form
- $Y = w_A Y_A + w_P Y_P + w_G Y_G + w_N Y_N$
  - $a_a = (1 \times d_{AN})^{-2}$
  - $a_p = (2 \times d'_{PA})^{-2} : d'_{PA} = \max(d_{PA}, 3.6)$
  - $a_g = (10 \times d'_G)^{-2}: d'_G = \max(d_G, 3.6)$
  - Normalize
    - $w_n = 1 / (1 + \exp(k * (d_{apg} - x_0)))$
    - $w_i = a_i (1 - w_n) / (a_a + a_p + a_g)$
- Performance
  - Statistical performance is great without satellite and satellite improves performance!
  - Created artificial "hard edges" when satellite and AirNow/PurpleAir diverge.
- ***Need method of identifying artifacts and then a new method to reduce artifacts.***

*Optimized parameters



Logistic alternative weight scheme

- Three scenarios to illustrate
  - AirNow on at the left.
  - PurpleAir near and a bit downwind.
  - GOES-PM25 coverage varies
- $Y = w_A Y_A + w_P Y_P + w_G Y_G + w_N Y_N$
  - $a_A = \text{expit}(d_A, k_A, r_A)$
  - $a_P = \text{expit}(d_P, k_P, r_P)$
  - $a_G = \text{expit}(d_G, k_G, r_G)$
  - Normalize
    - $w_n = \text{expit}(d_{APG}, -k_N, r_N)$
    - $w_a = a_a (1 - w_n) / (a_a + a_p + a_g)$
- Ideally, optimize ***k*** and ***r*** parameters
- Need to test with edge detection and categorize outliers.



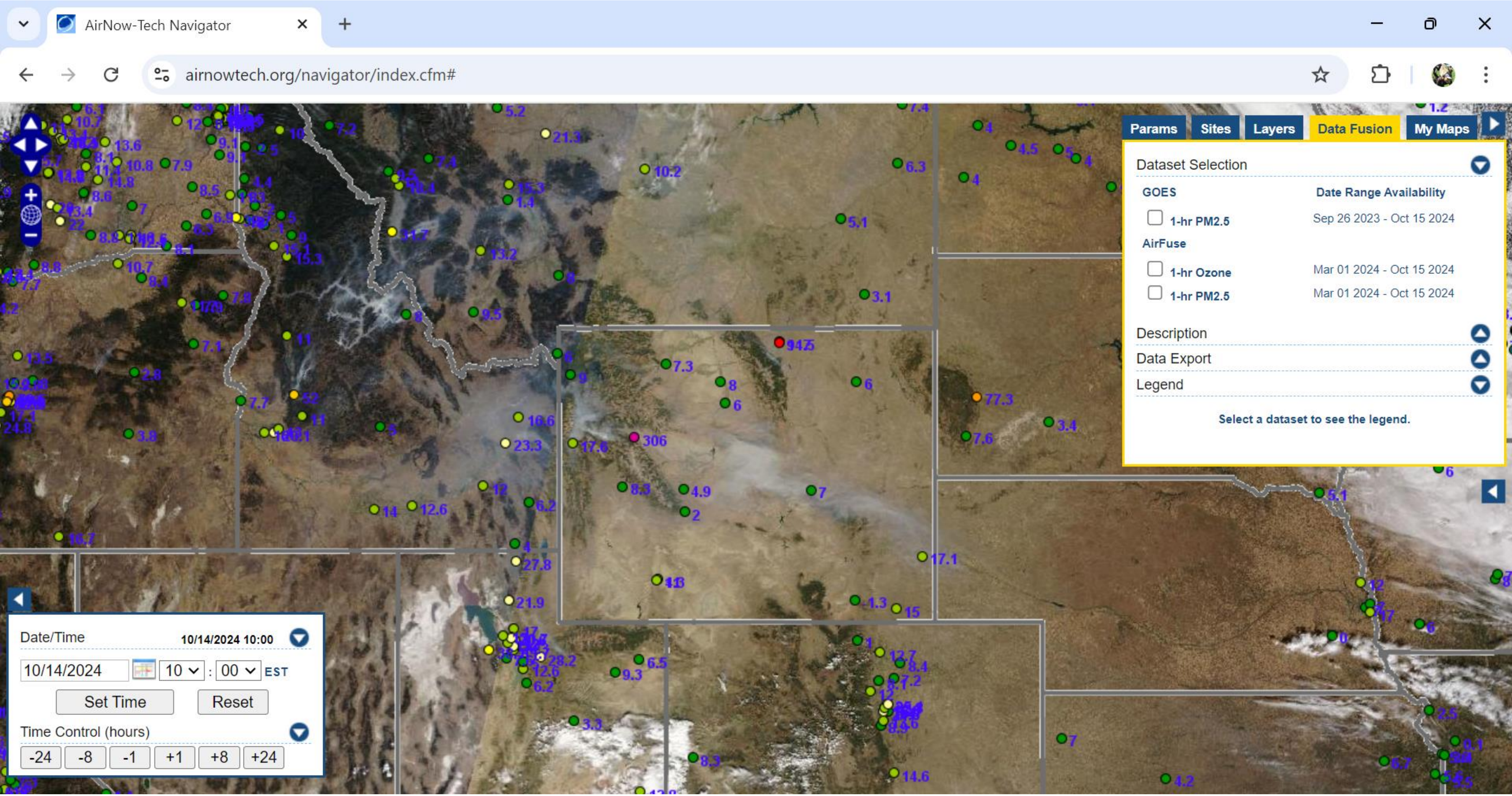
$\text{expit}(d, k, r) = 1 / (1 + \exp(k * (d – r)))$

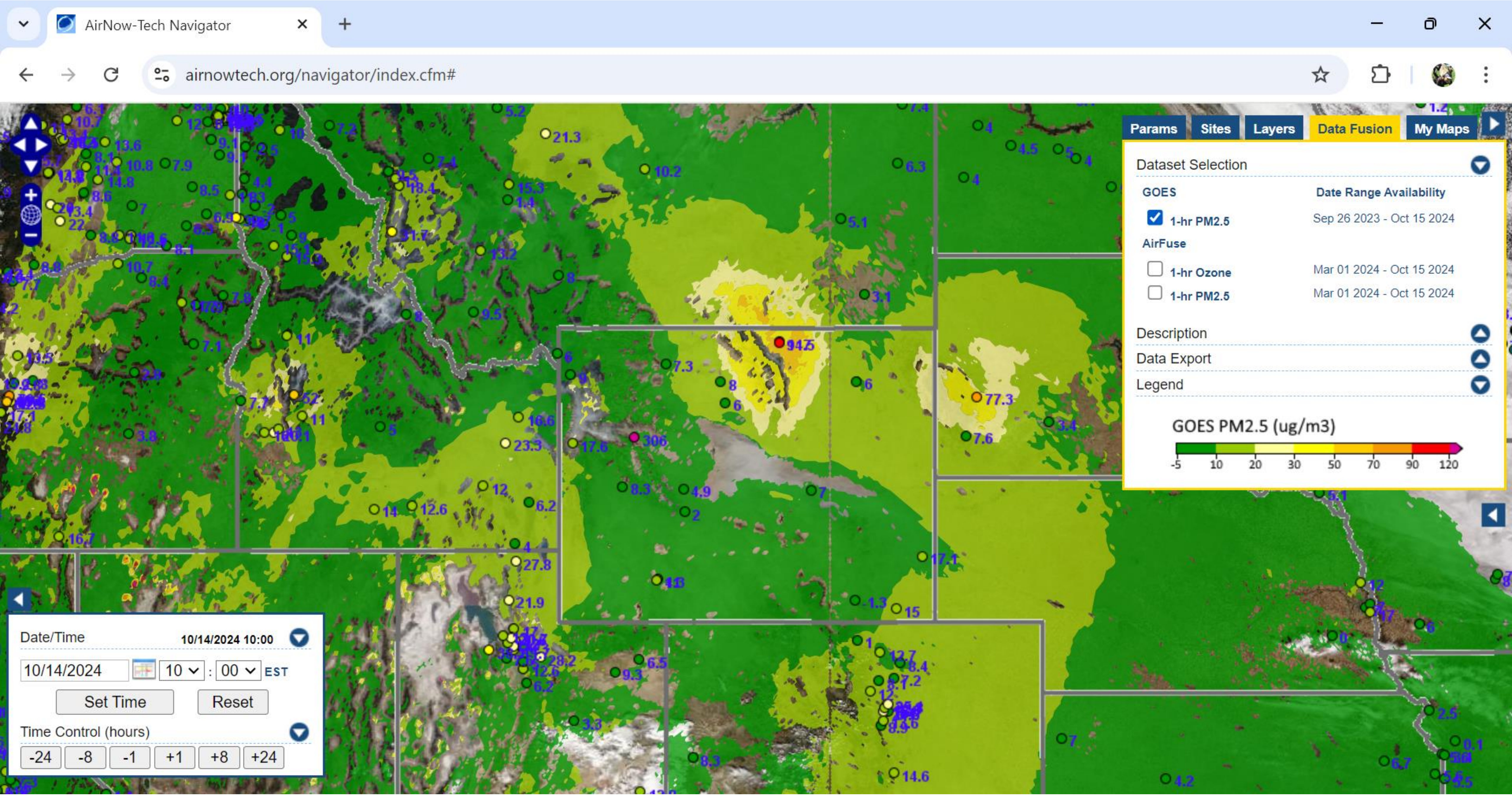
Summary

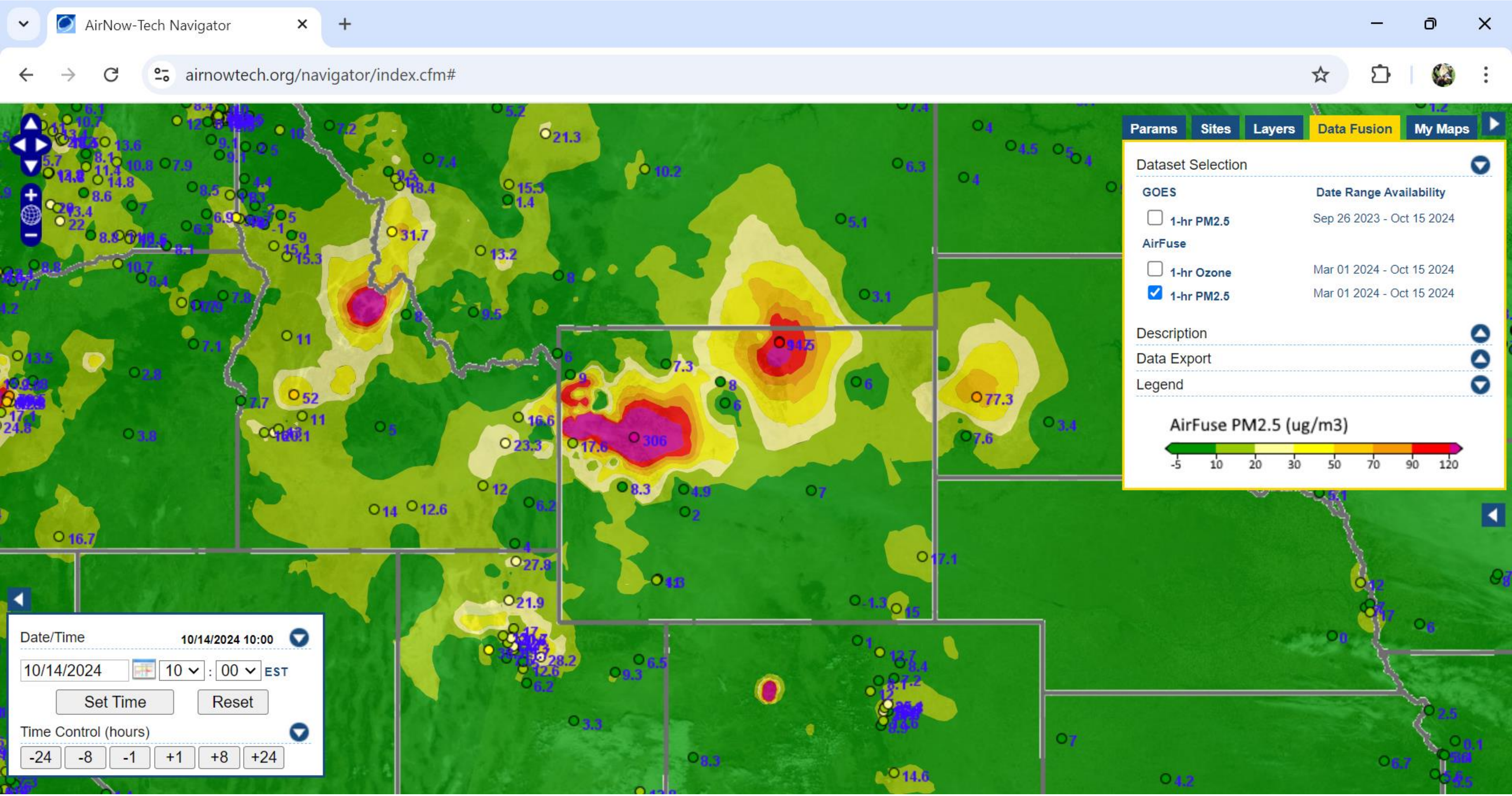
- Fusion with PurpleAir is running as a pilot without satellite
  - Schulte et al. demonstrated including models and PurpleAir improved on simple interpolations and applied it in an AirNow-like system.
  - Discontinuities are less stark than GOES because datasets are more spatially consistent (ie sparse in the same places).
  - Statistical value of PurpleAir in cross-validation is high because sensors are dense near monitors.
- Working on ensemble weighting with HAQAST team
  - HAQAST Tiger Team 2021 (Gupta) – now 2023 (Yang Liu)
  - Evaluated GOES PM25 for real-time-applications.
  - Developed edge detection algorithm for testing weighting schemes.
  - Finalizing weighting scheme and testing updated weighting methodology.
  - Harder to statistically quantify benefit because the value is further from monitors.
- Need your feedback on pilot!
  - Statistics will only tell us so much.
  - How does your area look?
  - When does AirFuse give weird answers?

Questions?

henderson.barron@epa.gov

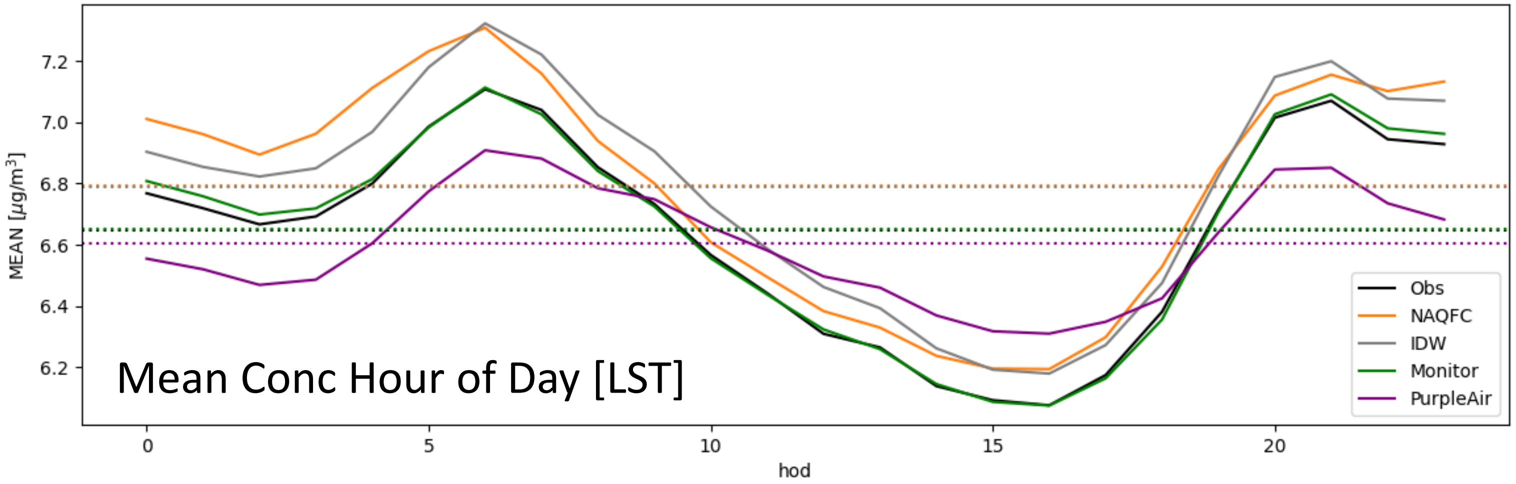

AirNow-Tech Navigator
airnowtech.org/navigator/index.cfm#
Params
Sites
Layers
Data Fusion
My Maps
Dataset Selection
GOES
Date Range Availability
1-hr PM2.5
Sep 26 2023 - Oct 15 2024
AirFuse
1-hr Ozone
Mar 01 2024 - Oct 15 2024
1-hr PM2.5
Mar 01 2024 - Oct 15 2024
Description
Data Export
Legend
Select a dataset to see the legend.
Date/Time
10/14/2024 10:00
10/14/2024
10
00
EST
Set Time
Reset
Time Control (hours)
-24
-8
-1
+1
+8
+24

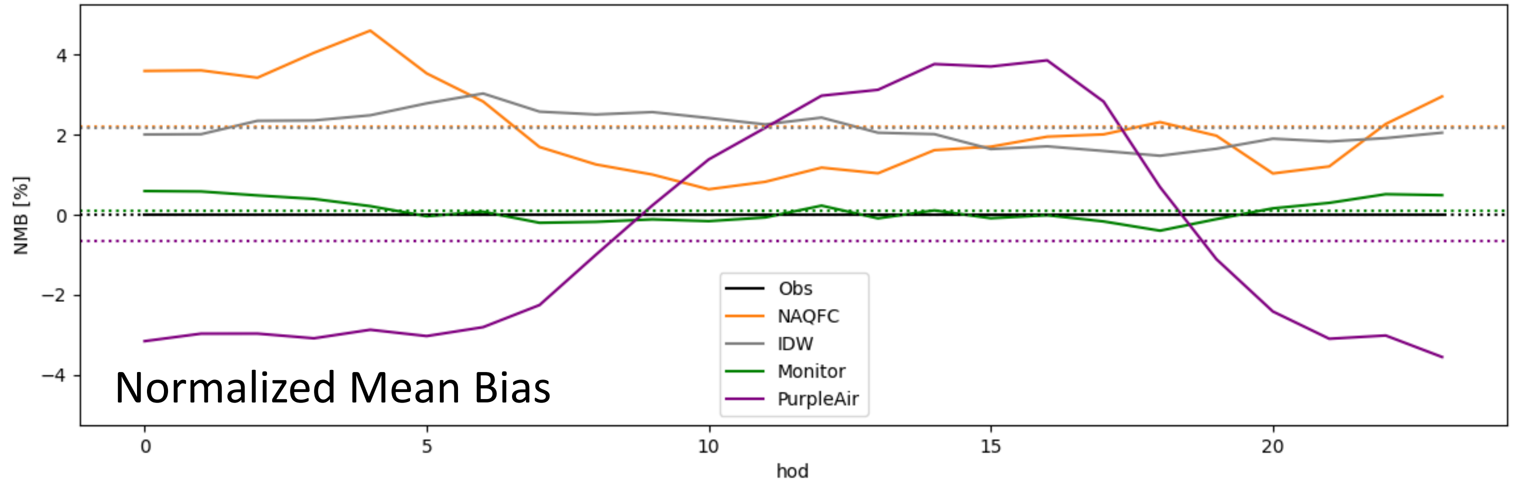

AirNow-Tech Navigator
airnowtech.org/navigator/index.cfm#
Params
Sites
Layers
Data Fusion
My Maps
Dataset Selection
GOES
Date Range Availability
1-hr PM2.5
Sep 26 2023 - Oct 15 2024
AirFuse
1-hr Ozone
Mar 01 2024 - Oct 15 2024
1-hr PM2.5
Mar 01 2024 - Oct 15 2024
Description
Data Export
Legend
GOES PM2.5 (ug/m3)
-5 10 20 30 50 70 90 120
Date/Time
10/14/2024 10:00
10/14/2024
10 : 00 EST
Set Time
Reset
Time Control (hours)
-24 -8 -1 +1 +8 +24


AirNow-Tech Navigator
airnowtech.org/navigator/index.cfm#
Params
Sites
Layers
Data Fusion
My Maps
Dataset Selection
GOES
Date Range Availability
1-hr PM2.5
Sep 26 2023 - Oct 15 2024
AirFuse
1-hr Ozone
Mar 01 2024 - Oct 15 2024
1-hr PM2.5
Mar 01 2024 - Oct 15 2024
Description
Data Export
Legend
AirFuse PM2.5 (ug/m3)
-5 10 20 30 50 70 90 120
Date/Time
10/14/2024 10:00
10/14/2024
10 : 00 EST
Set Time
Reset
Time Control (hours)
-24 -8 -1 +1 +8 +24

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.