

PurpleAir PM_{2.5} U.S. Correction and Performance During Smoke Events

Karoline K. (Johnson) Barkjohn^{1,2}, Amara L. Holder², Samuel Frederick^{2,3}, Andrea L. Clements²

¹*Oak Ridge Institute for Science and Education Postdoctoral Fellow*

²*U.S. EPA Office of Research and Development, Center for Environmental Measurement & Modeling*

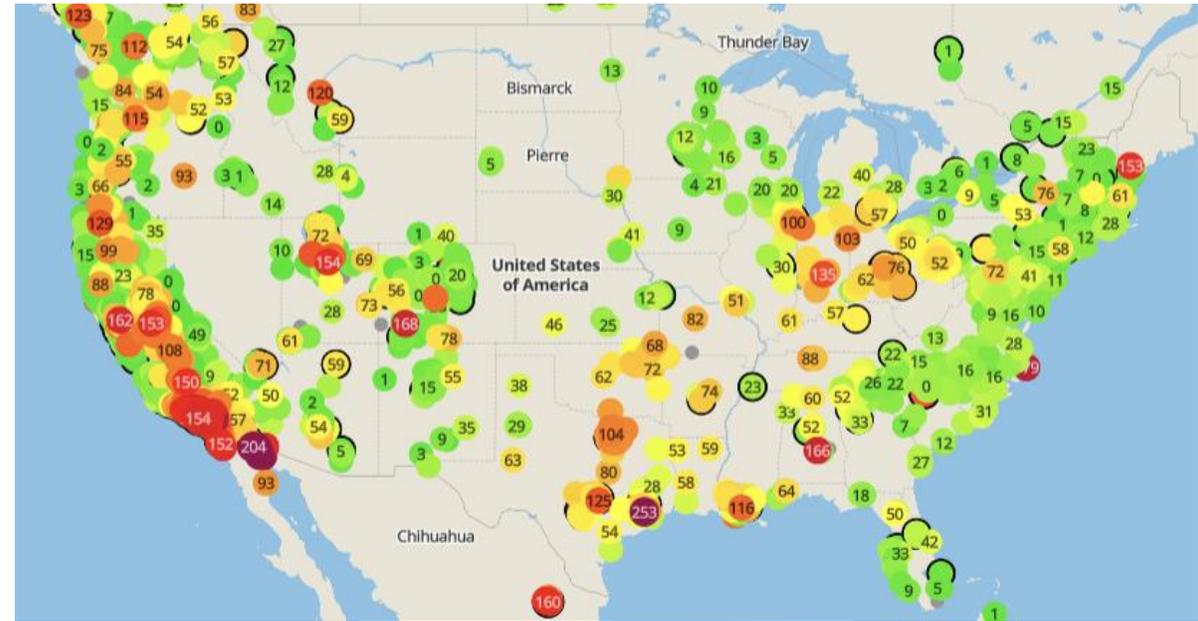
³*Oak Ridge Associated Universities*

Disclaimer: *The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the U.S. Environmental Protection Agency. The EPA does not endorse any commercial products, services, or enterprises.*

Questions

Can PurpleAir data be used to provide near real-time air quality index (AQI) comparable to the regulatory network especially during smoke impacts?

What correction or data processing steps are recommended to provide adequate data quality?



PurpleAir Sensors are deployed widely across the U.S. providing the opportunity for more dense air quality measurements

Image source: PurpleAir.com

Overview

1. Data cleaning steps and development of a U.S. wide correction
2. Application of correction to typical and smoke impacted datasets
3. Evaluation based on NowCast AQI category

Development of a U.S. Wide Correction & Data Cleaning Steps

U.S. Wide Correction: Dataset

Through partnership with state, local, and tribal agencies, received data from PurpleAir sensors collocated at regulatory monitoring sites across the U.S.

Averaged sensor and regulatory data to 24-hour averages with a 90% completeness threshold

Reference measurements from

- Federal Equivalent Method (FEM)
 - A variety of continuous instruments
- Federal Reference Method (FRM)
 - Filter based



Collocation sites across the U.S.

Channels & Correction Factors

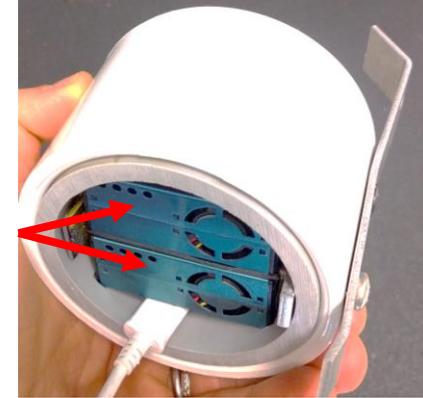
Channels:

- 2 Plantower PMS5003 (channels A & B)
- Sample for alternating 10-second intervals
- Provides many data parameters including temperature, RH, pressure, PM_{1} , $PM_{2.5}$, PM_{10} , and binned counts
 - Only $PM_{2.5}$ and RH values used in this work
- Generate 2-minute averages
 - previously 80-second

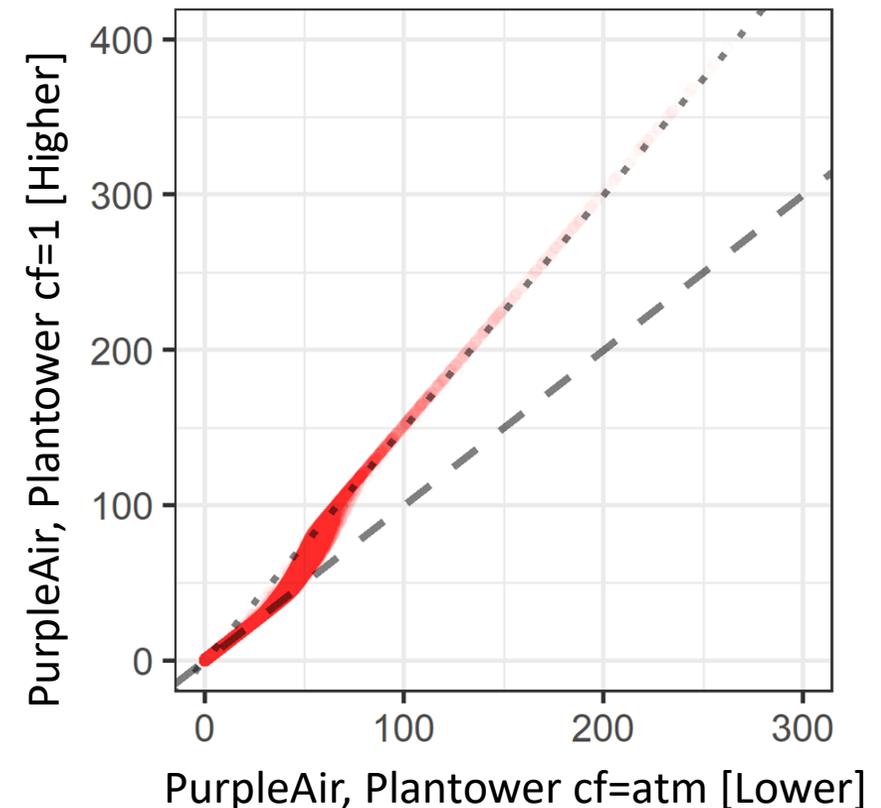
Correction Factors:

- PurpleAir provides PM data with two corrections
 - PurpleAir.com map: **cf=atm** [lower]
 - Also: **cf=1** [higher]
- Previously PurpleAir switched these labels
 - Easy check: cf=1 is higher
- **PurpleAir cf=1 selected for correction**
 - $R^2_{[cf=1]} = 0.65 > R^2_{[cf=atm]} = 0.64$

Channels
A&B

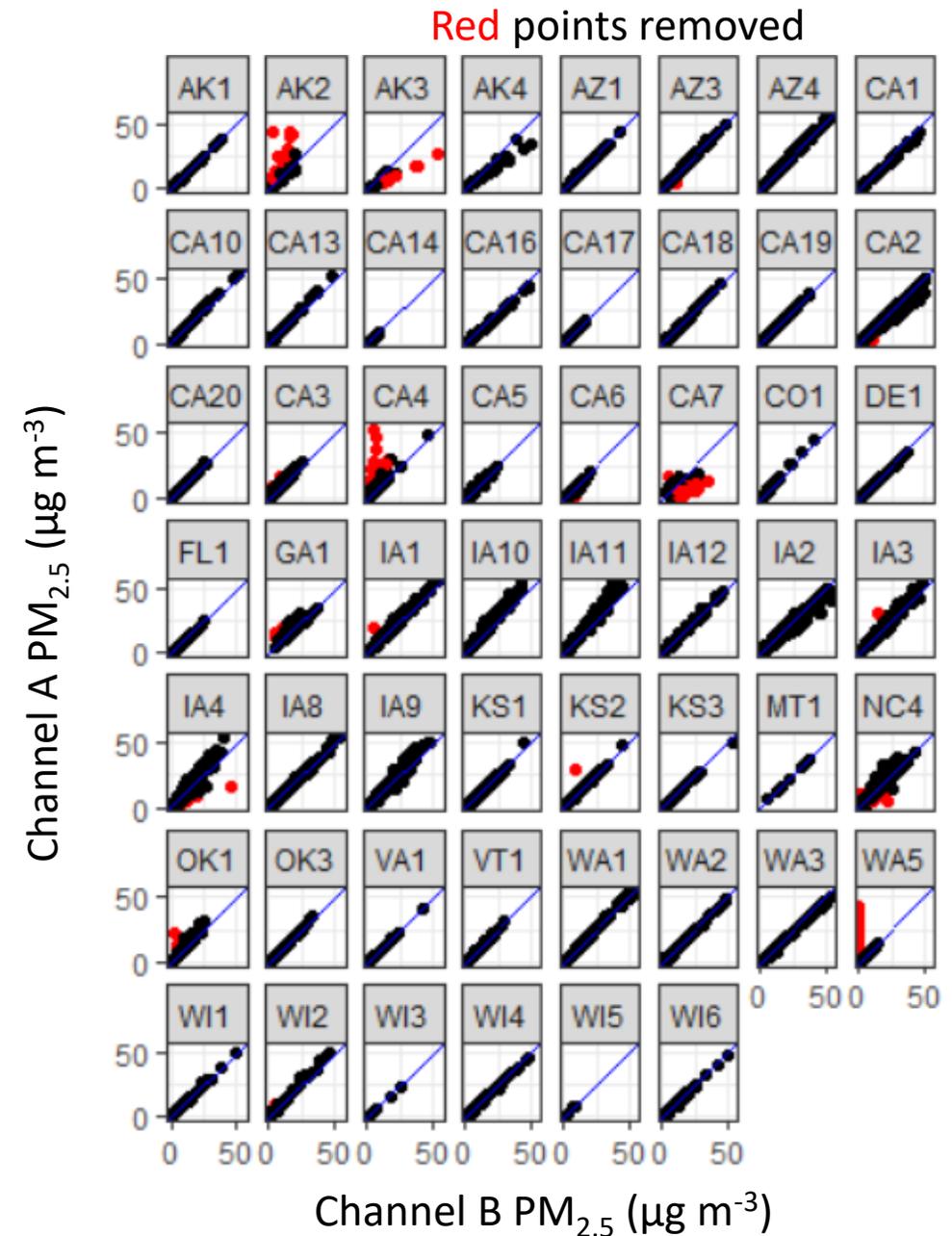


PurpleAir underside view



Data Cleaning Steps

- Agreement between A and B channels provides confidence in measurements
- Points removed if 24-hr averaged A & B $PM_{2.5}$ differ by
 - $\geq \pm 5 \mu\text{g m}^{-3}$ AND
 - $\geq \pm 2 * \text{sd}(\% \text{ error})$
 - 56% for 24-hr data
 - 70% for 1-hr data
 - Removed
 - 0-12% for individual sensors
 - 2% overall
- A & B channels averaged



U.S. Wide Correction: Equation

- Considered data collected by the sensor
- Explored several equation forms
- Balanced accuracy and complexity

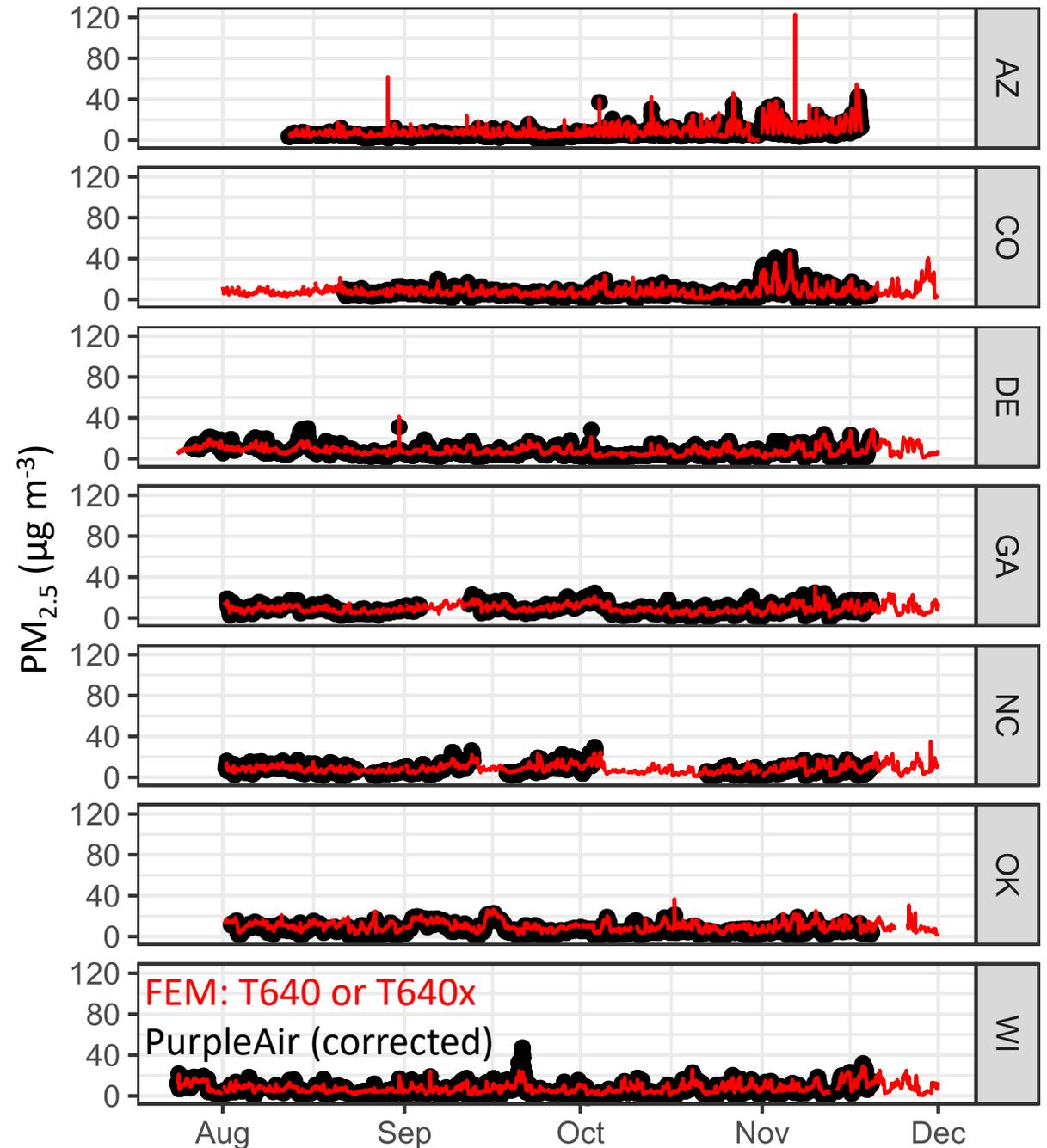
$$\text{PM}_{2.5} \text{ corrected} = 0.52 * [\text{PA_cf1}(\text{avgAB})] - 0.085 * \text{RH} + 5.71$$

- $\text{PM}_{2.5} = \mu\text{g m}^{-3}$
- RH= Relative Humidity (%)
- $\text{PA_cf1}(\text{avgAB})$ =PurpleAir higher correction factor data averaged from the A and B channels

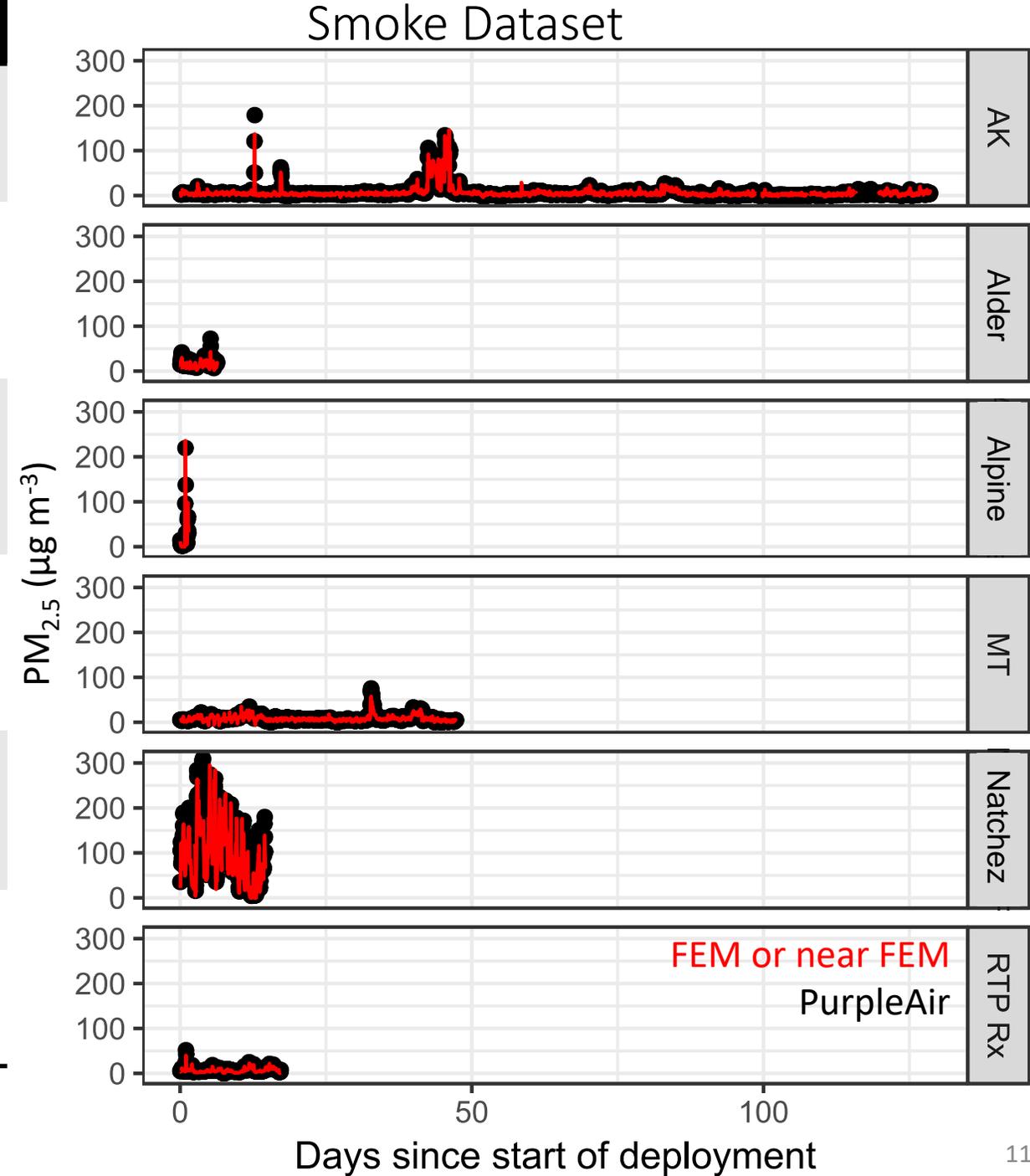
Application of Correction to Typical Ambient and Smoke Impacted Datasets

Typical Ambient Dataset

- Sites part of EPA's long-term sensor performance project
- 7 sites spread across the U.S.
- Urban and Neighborhood sites
- Different meteorological conditions represented
- Variation in particle types and sizes
- Work began in late July 2019



Event	Location	Dates	Reference	PM Type
AK Collocation	Anchorage, AK	May–Aug 2019	BAM 1020	Ambient - aged fire
Alder Fire	Pinehurst, CA	Oct 20–27 2018	BAM 1020	Ambient - aged fire
Alpine Burns	Oakley, UT	Oct 31–Nov 4 2019	E-Sampler	Prescribed fire + woodstove
MT Collocation	Missoula, MT	July–Sept 2019	BAM 1020	Ambient + prescribed fire
Natchez Fire	Happy Camp, CA	Aug 11–29 2018	E-BAM	Wildfire
RTP Initial Evaluation	EPA – RTP, NC	Aug–Dec 2018	Grimm EDM-180	Ambient + prescribed fire



Evaluating Corrected Data

Apply correction to 1-hr averaged data

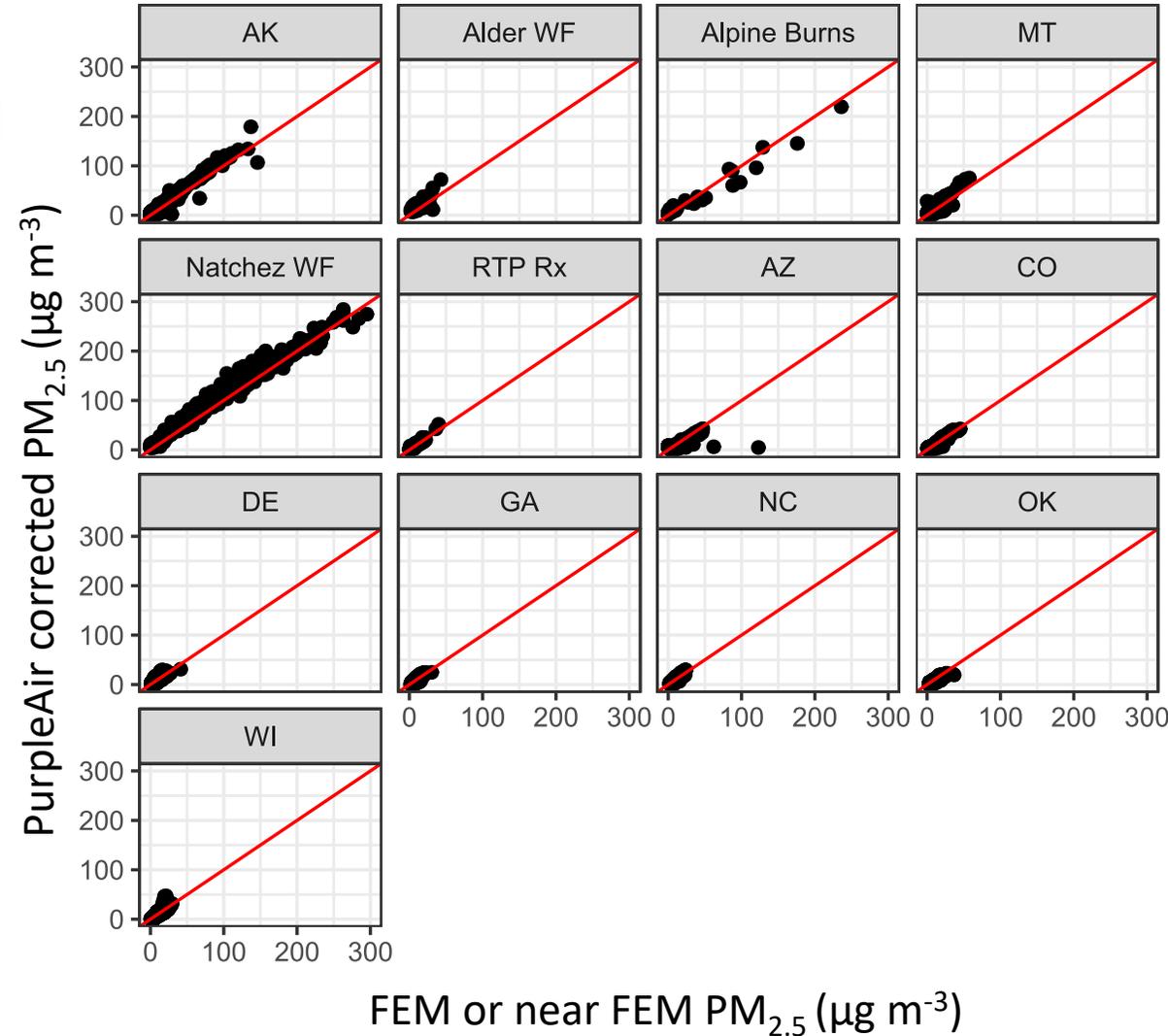
- Typically reduces bias across states

Next apply NowCast algorithm

- AQI value generated **every hour** based on the previous 12-hours
- Weighted more heavily to the recent data if concentrations are changing quickly

C_{low}	C_{high}	I_{low}	I_{high}	Category
0	12.0	0	50	Good
12.1	35.4	51	100	Moderate
35.5	55.4	101	150	Unhealthy for Sensitive Groups
55.5	150.4	151	200	Unhealthy
150.5	250.4	201	300	Very Unhealthy
250.5	350.4	301	400	Hazardous
350.5	500.4	401	500	Hazardous

AQI categories



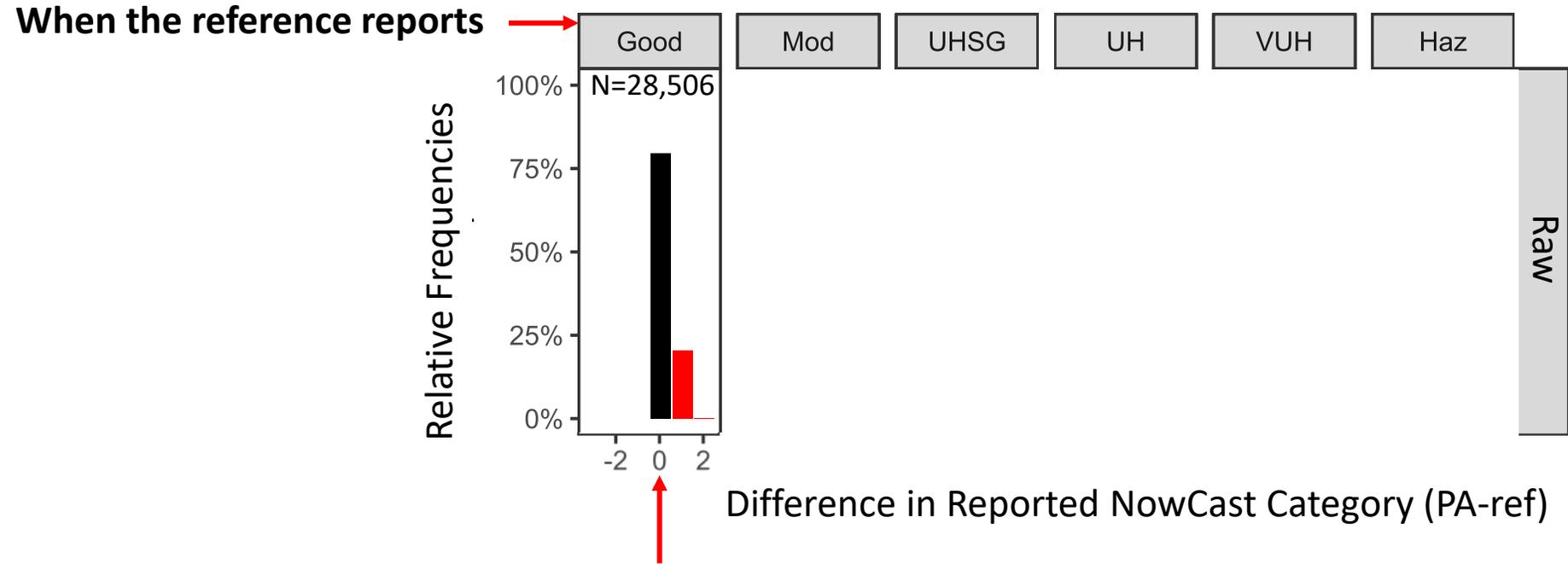
Additional details on NowCast calculation:

https://www3.epa.gov/airnow/ani/pm25_aqi_reporting_nowcast_overview.pdf

Evaluation based on NowCast AQI Category

NowCast Performance: Full Dataset

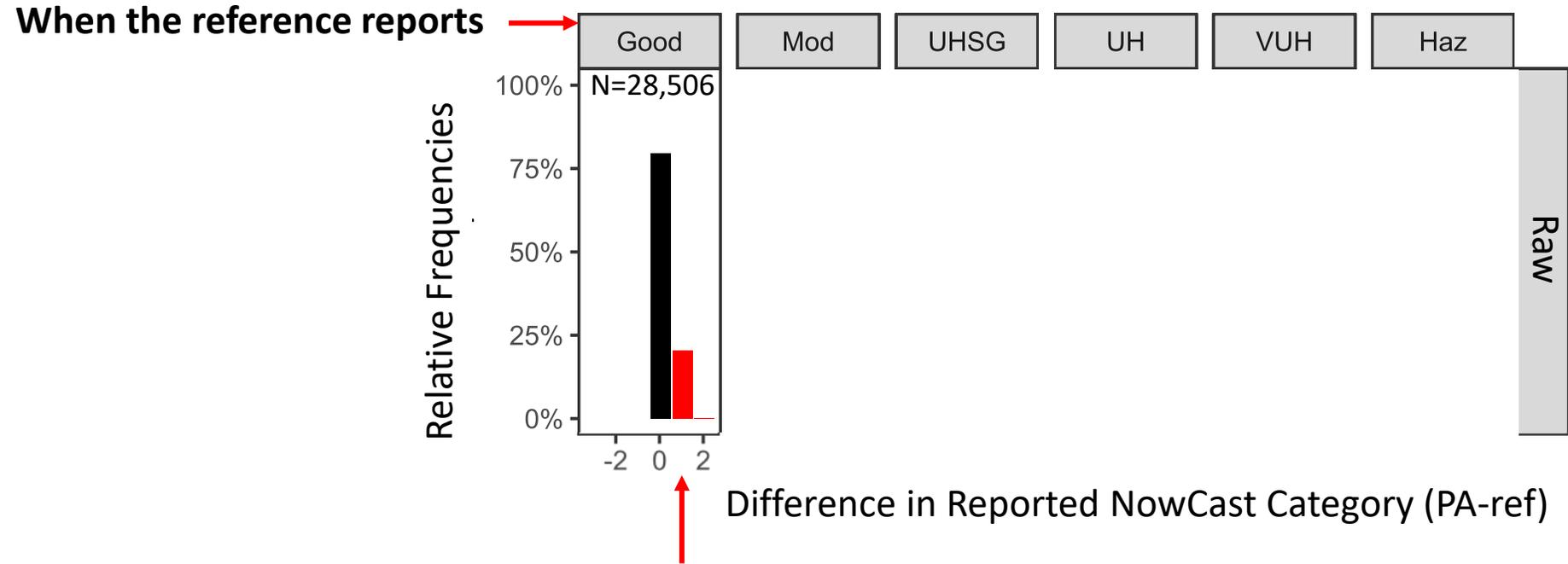
-  PurpleAir underpredicts
-  PurpleAir correct
-  PurpleAir overpredicts



**The PurpleAir reports the correct AQI 80% of the time
(0 categories different)**

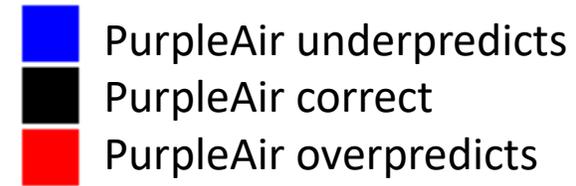
NowCast Performance: Full Dataset

- PurpleAir underpredicts
- PurpleAir correct
- PurpleAir overpredicts



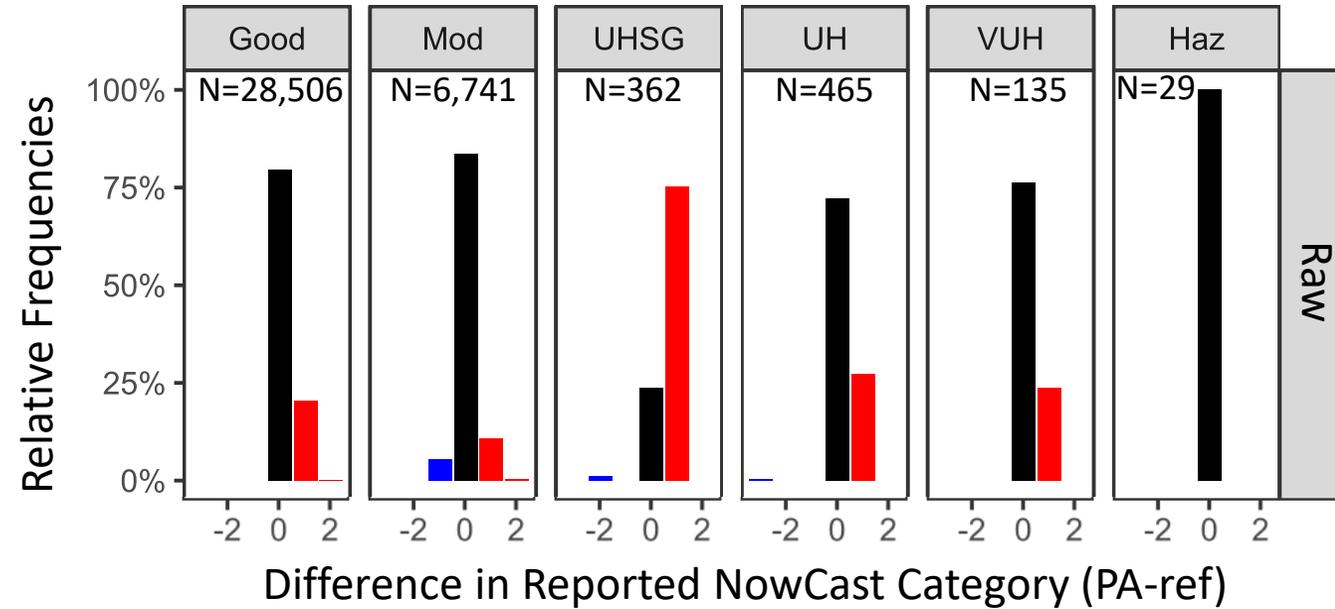
**The PurpleAir reports 1 AQI higher (Moderate)
20% of the time**

NowCast Performance: Full Dataset

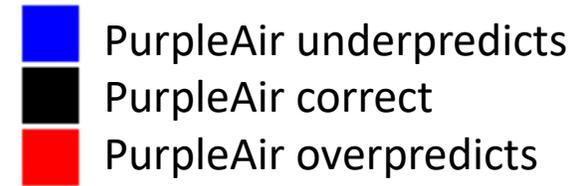


Raw: PurpleAir data

- Often over predicts NowCast AQI
- Typically within 1 category except a few outliers



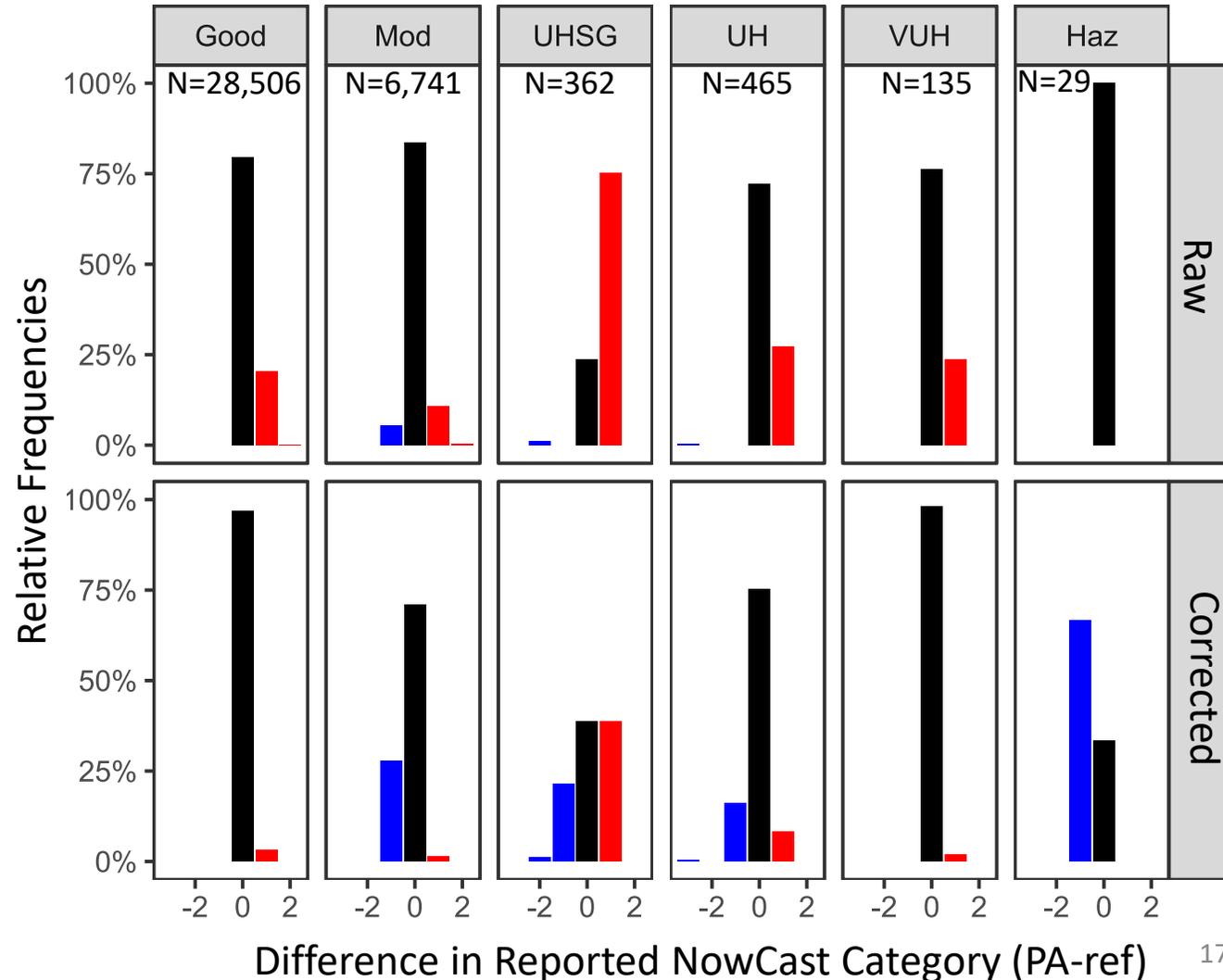
NowCast Performance: Full Dataset



Raw: PurpleAir data

Corrected: PurpleAir data with QC steps and U.S. correction

- Predicts correctly more often except at moderate and hazardous (limited hazardous data)
- Some underprediction but typically only by 1 category so behavior changes may be similar



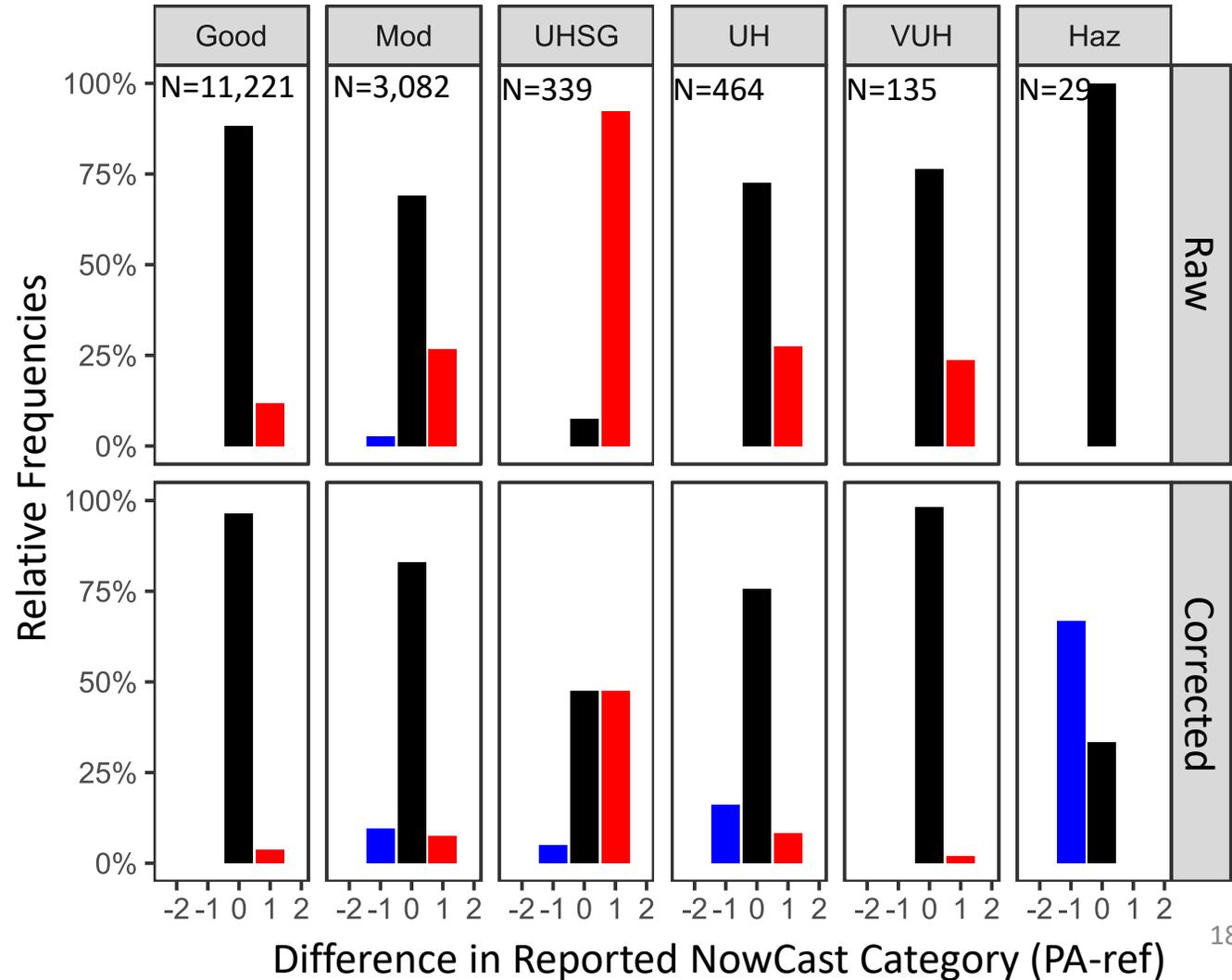
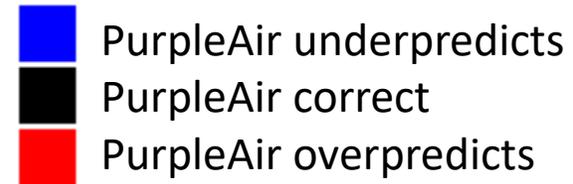
NowCast Performance: Smoke only

Almost all data >Mod is from the Smoke impacted datasets

Corrected: PurpleAir data with QC steps and U.S. correction

- Some underprediction
- Less over prediction than raw data
- Typically within 1 AQI category

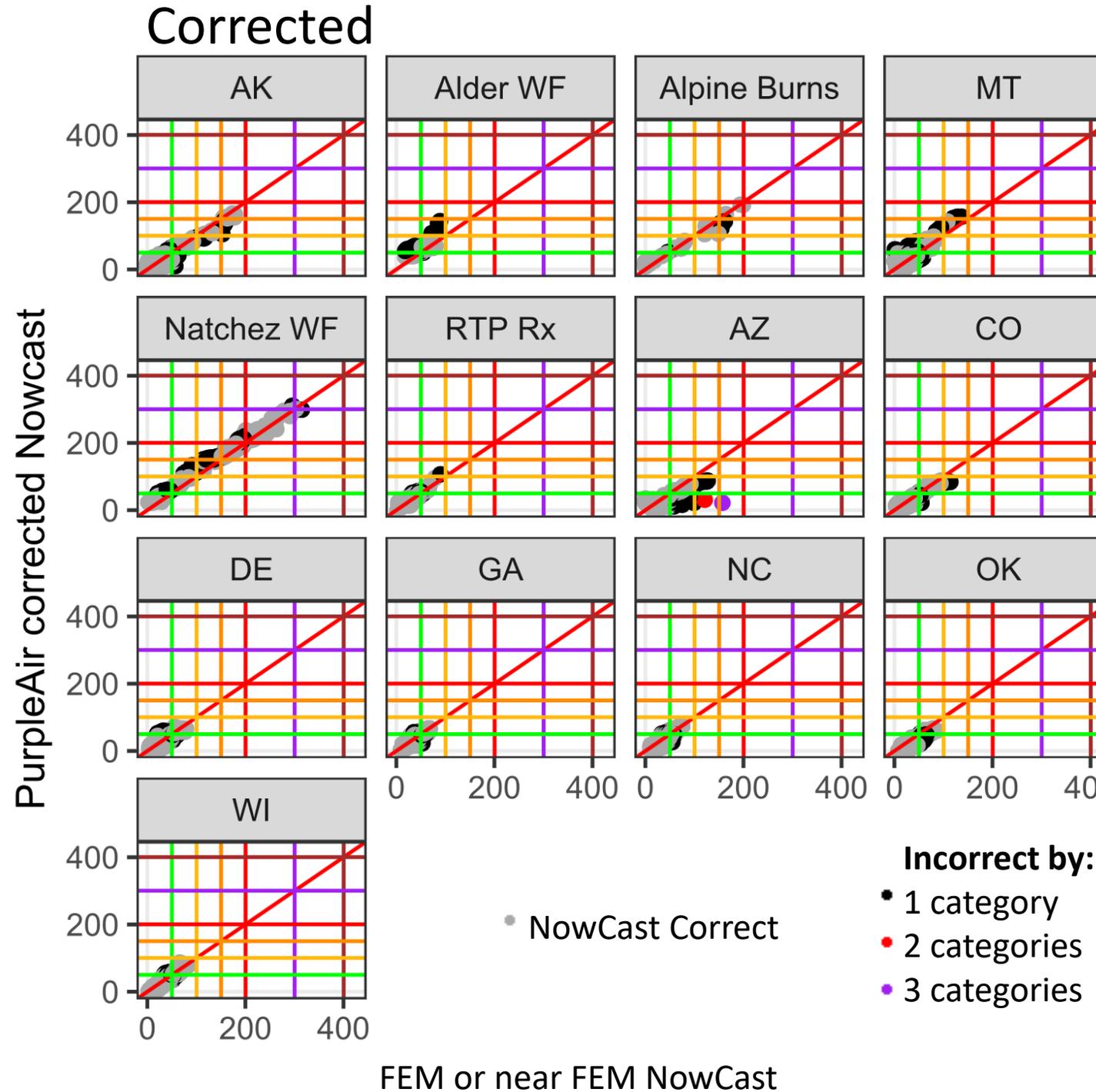
Actions taken at Unhealthy and >AQI may be similar limiting the need for accuracy at higher NowCast AQI levels



NowCast Summary

Most disagreement is at the boundary between AQI levels

- Shows that though the categories are different, the AQI values may not be largely different



Conclusions

Without correction PurpleAir sensors

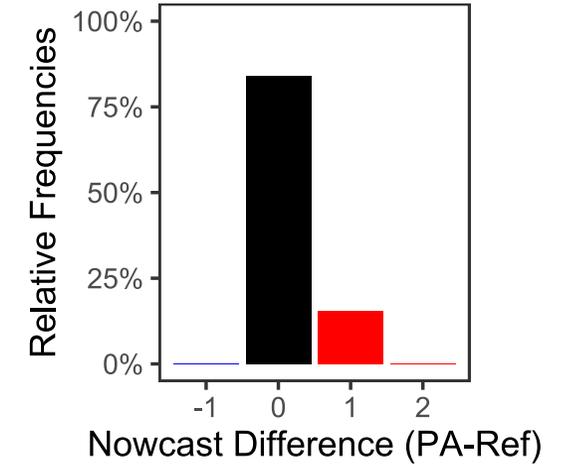
- Overestimate AQI category 20% of the time
- Overestimate most often in the unhealthy for sensitive groups category

A U.S. correction can improve NowCast AQI reporting

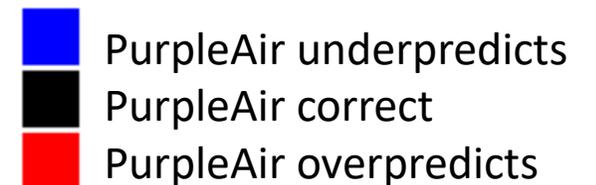
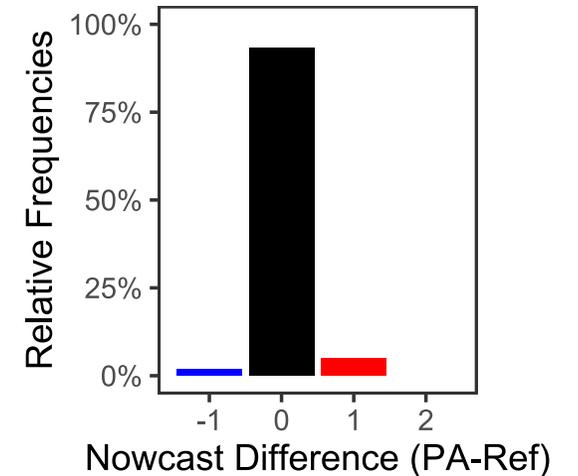
- AQI category correctly reported 93% of the time
- AQI reported within 1 category 100% of the time

Smoke only dataset

Raw



US correction



Acknowledgements

AK: State of Alaska, Citizens for Clean Air

AZ: Maricopa County Air Quality Department

CA: San Luis Obispo County Air Pollution Control District, Mojave Desert Air Quality Management District, Antelope Valley Air Quality Management District, California Air Resources Board, Santa Barbara County Air Pollution Control District, Air Quality Sensor Performance Evaluation Center, Ventura County Air Pollution Control District

CO: Colorado Department of Public Health and Environment

DE: Delaware Division of Air Quality

FL: Sarasota County Government

GA: EPA Region 4, Georgia Environmental Protection Division

IA: Iowa Air Quality Bureau

MT: Missoula County, Montana Department of Environmental Quality

NC: Forsyth County Office of Environmental Assistance & Protection, Clean Air Carolina, UNC Charlotte, North Carolina Department of Environmental Quality

OH: Akron Regional Air Quality Management District

OK: Quapaw Nation, Oklahoma Department of Environmental Quality

UT: University of Utah, Utah Department of Environmental Quality

VA: Virginia Department of Environmental Quality

VT: State of Vermont

WA: Washington Department of Ecology, Puget Sound Clean Air Agency

WI: Wisconsin Department of Natural Resources

Federal: Forest Service, Wildland Fire Air Quality Response Program, National Park Service, EPA Region 9, EPA Region 10, Lauren Maghran, Heidi Vreeland, Gayle Hagler

Questions?

Contacts:

Johnson.karoline@epa.gov

Holder.Amara@epa.gov

Clements.Andrea@epa.gov

Additional Slides

NowCast Summary

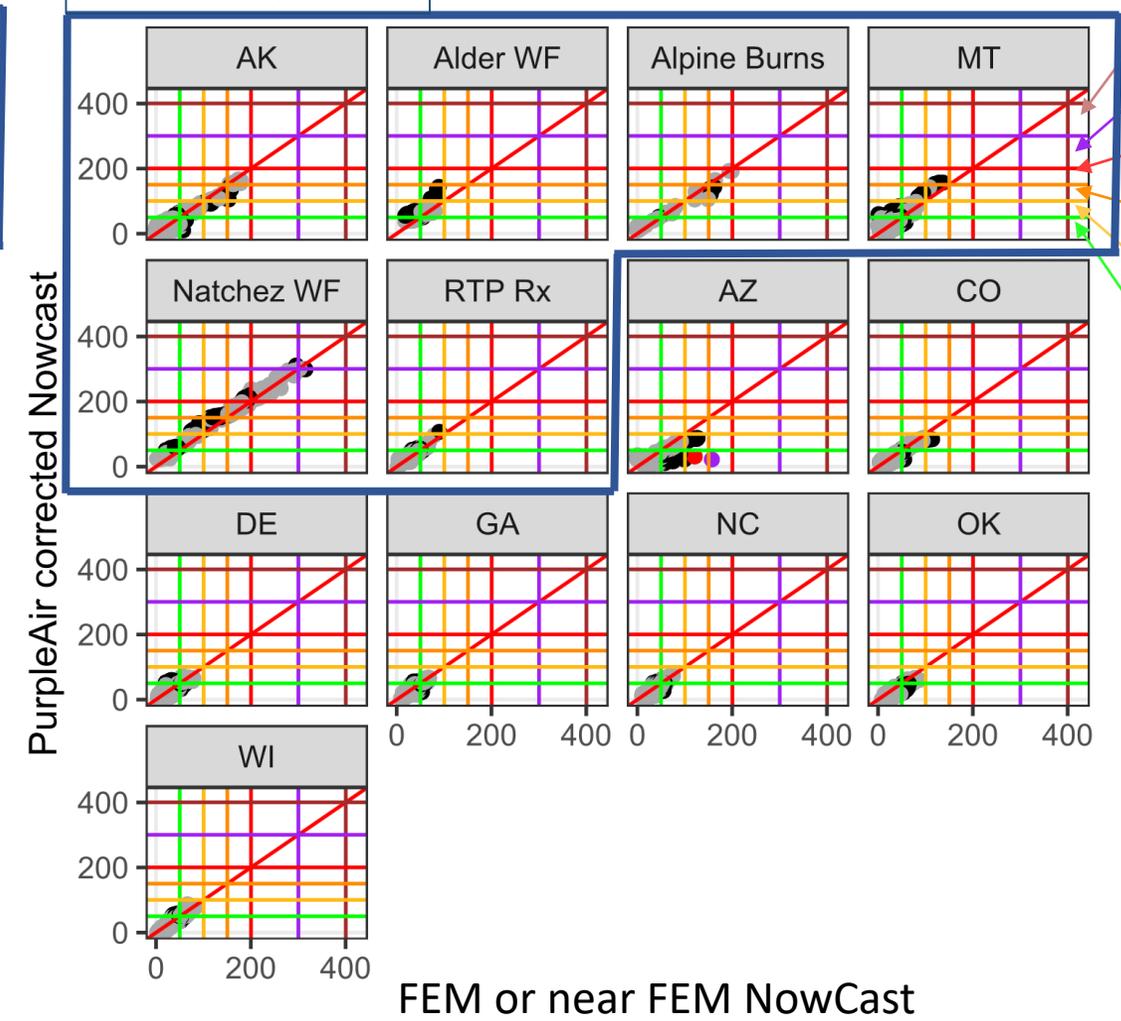
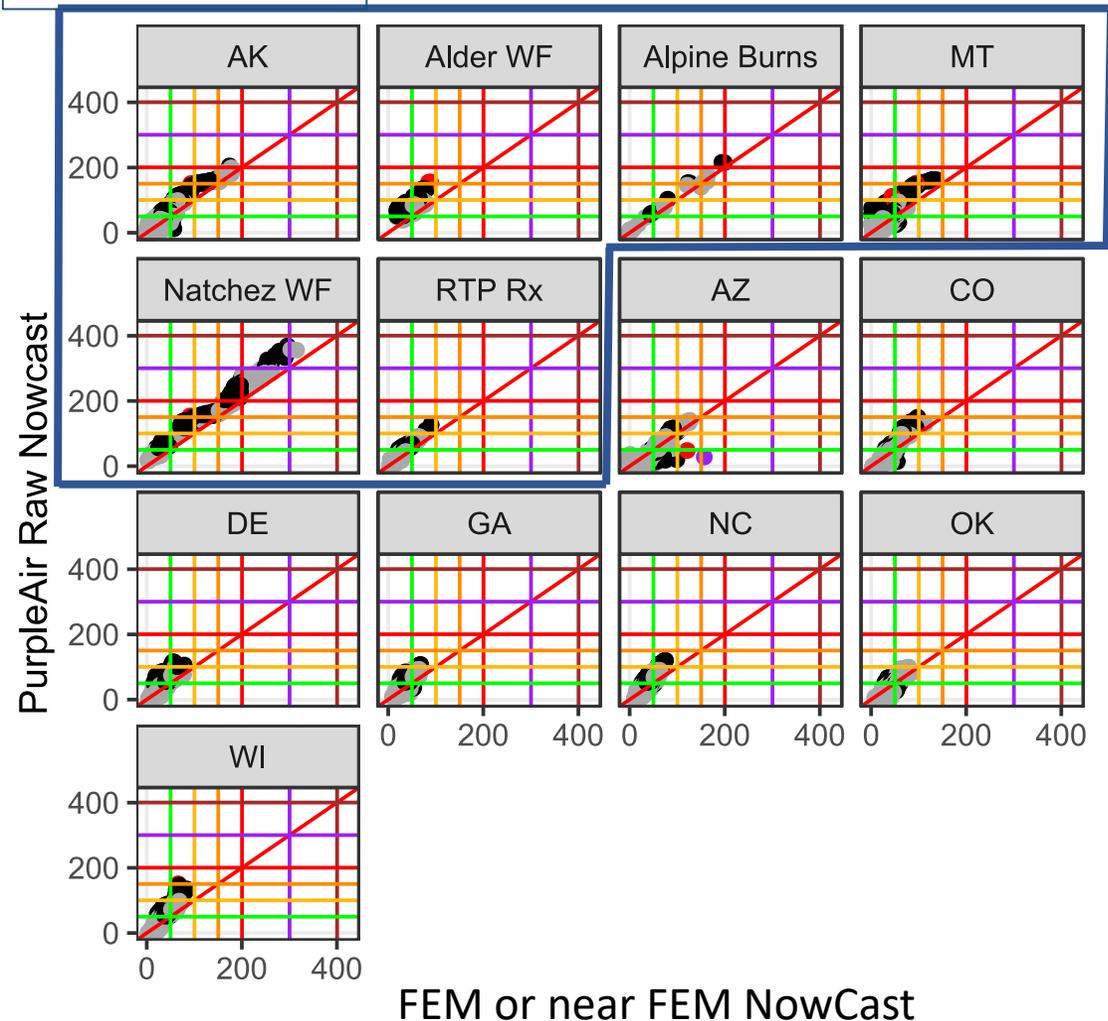
- NowCast Correct

Incorrect by:

- 1 category
- 2 categories
- 3 categories

Smoke impacted Raw: more overestimation

Smoke impacted Corrected



Step-by-step Procedures to Correct PA Data

Steps utilized in this study, **comments in purple on operational use considerations**

1. Receive 2-minute Raw PurpleAir data

2. Extract columns:

- Time stamp
- Channel A: PM2.5_CF1_ug/m3 (higher correction factor)
- Channel B: PM2.5_CF1_ug/m3 (higher correction factor)
- Humidity*

3. Average each column to 1-hr  *Skip if input data are hourly resolution*

4. If less than 90% of the measurements are available in that hour average exclude  *Skip if input data are hourly resolution*

5. If 1-hr A&B averages are different by BOTH: $5 \mu\text{g m}^{-3}$ & 70% exclude**

6. Apply U.S. Correction:

$$\text{PM}_{2.5} \text{ corrected} = 0.534 * [\text{PA_cf1}(\text{avgAB})] - 0.0844 * \text{RH} + 5.604$$

7. Calculate NowCast based on past 12 hours of data

*If data is from offline sensor also save uptime column

**If data is from offline sensor also remove data when uptime resets (indicating searching for WiFi)

NOTE: this was a rare event in our data set – only <1% of hours were flagged for having incomplete data (>0 data points, <26 datapoints). Further, this was a strict requirement to develop the algorithm and operationally lower completeness may be acceptable (e.g., 75%)