



Using Machine Learning to Develop a Calibration Model for Low-Cost Air Quality Sensors Deployed During a Dust Event

By: Sean Hickey

Advisor: Dr. Lu Liang

Committee Members: Dr. Pinliang Dong, Dr. Chetan Tiwari

**Context &
Questions**

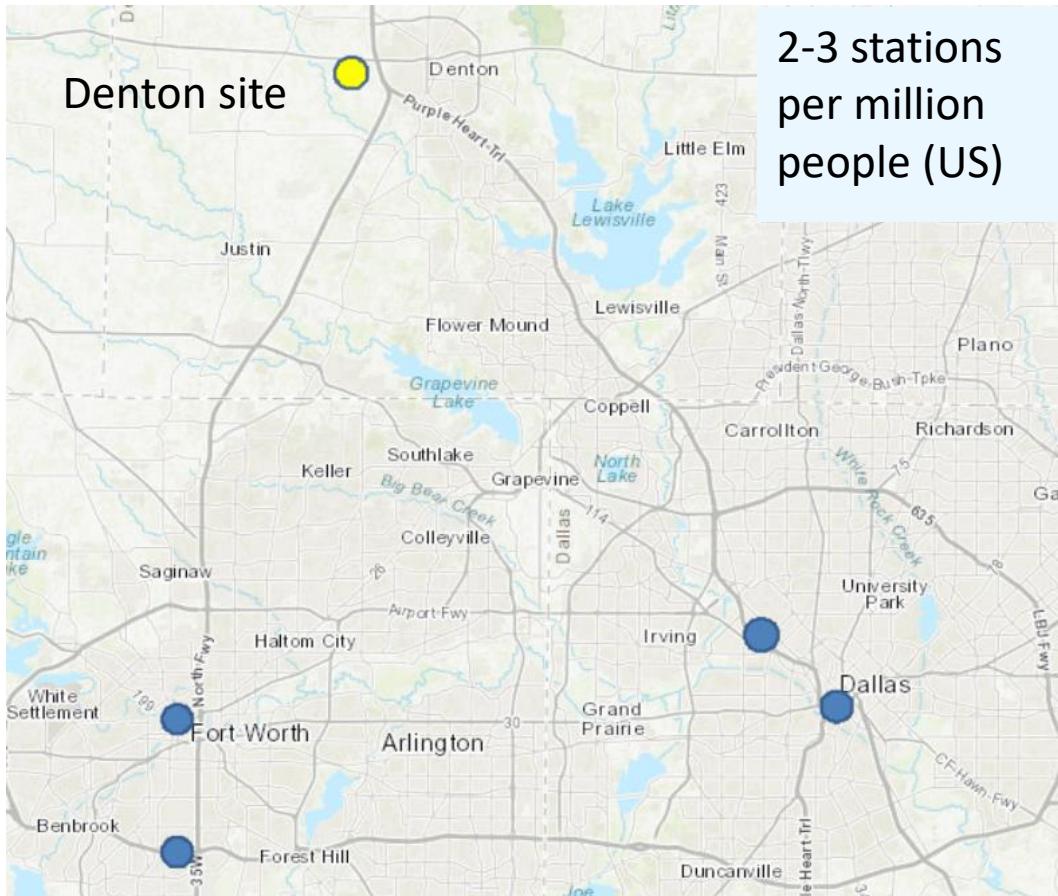
Methods

**Results &
Discussion**

Future Work

Pollution is more variable than what we can currently capture

Dallas-Fort Worth Regulatory Monitoring Sites



Dallas-Fort Worth Low-Cost Sensor Sites



Sensor Influences

Variable	Effect
Relative Humidity	Water droplets binds to particles
Temperature	Temperature influences particle velocity
Background Concentration	Sensors tend to be more accurate in high concentration environments than low-concentration

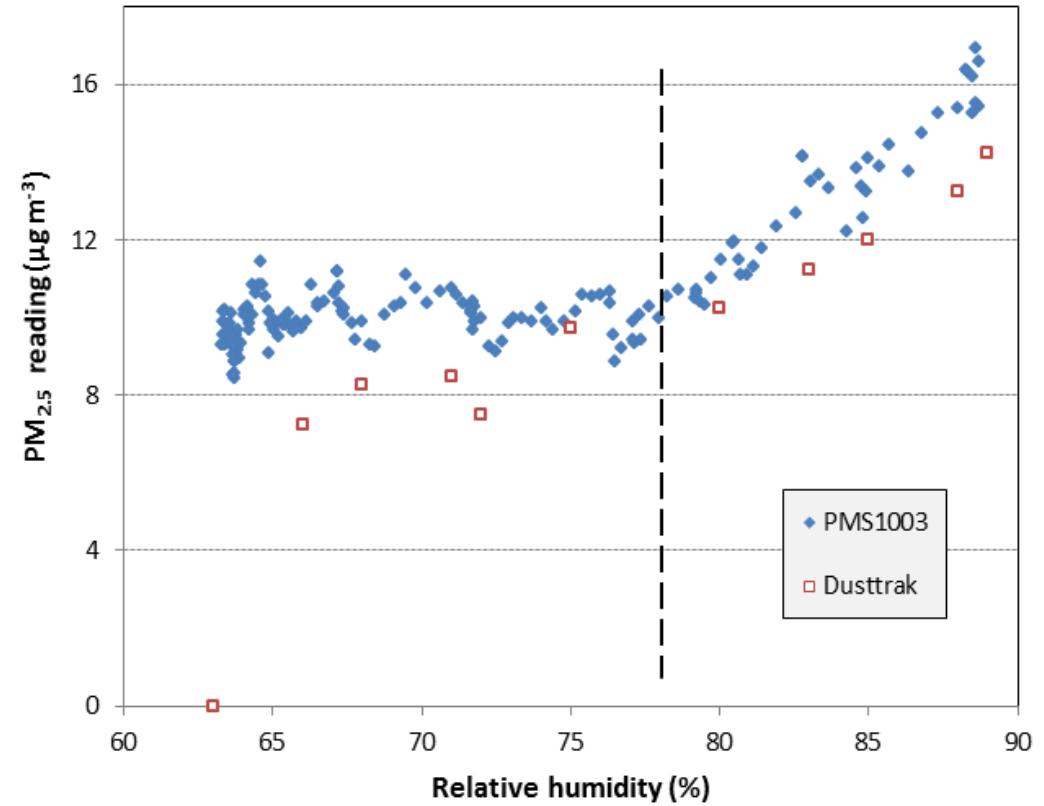
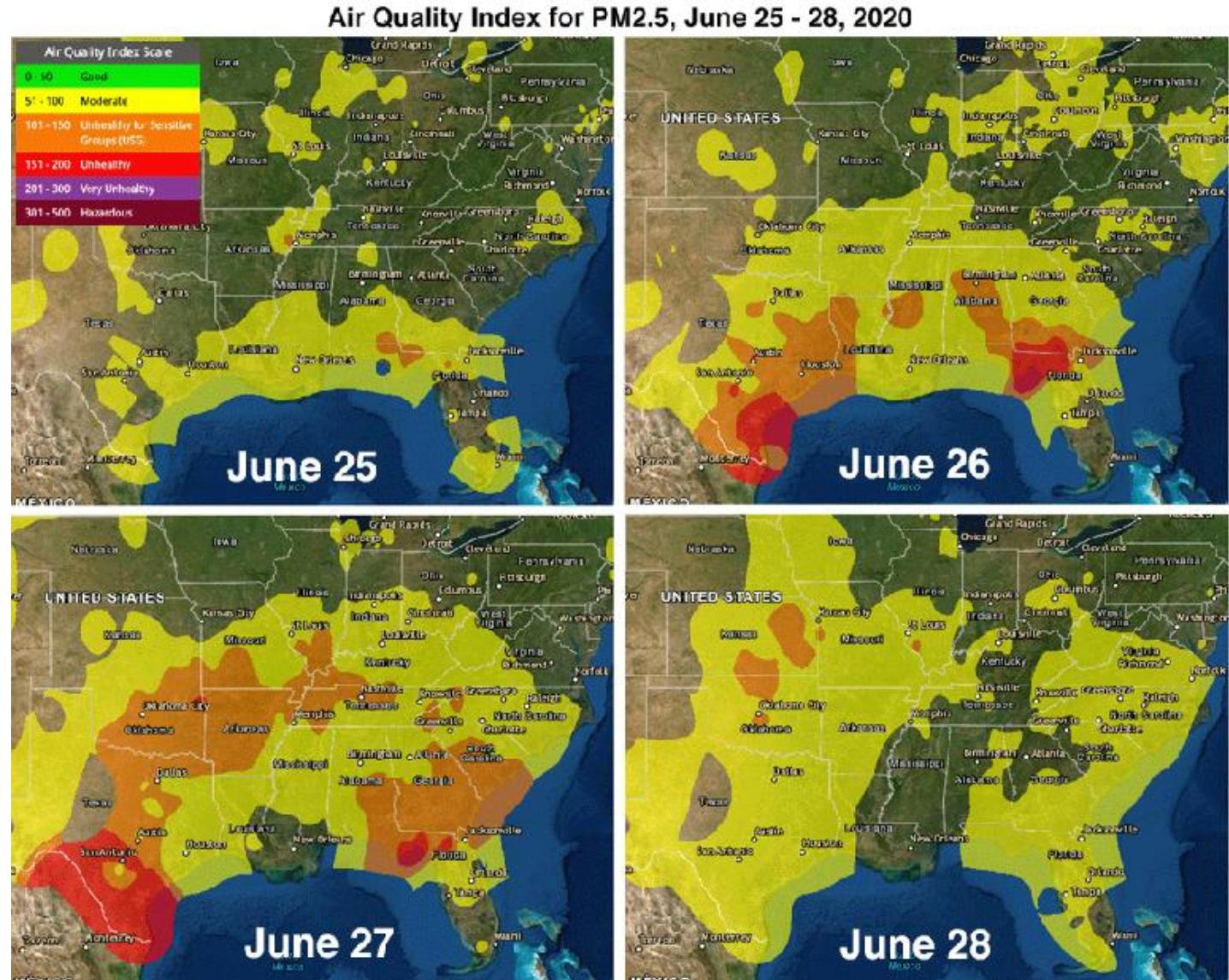


Figure 1. The PM_{2.5} concentrations reported by two optical-based particle sensors as the relative humidity was systematically increased in a laboratory chamber (Jayarantne et al. 2018).

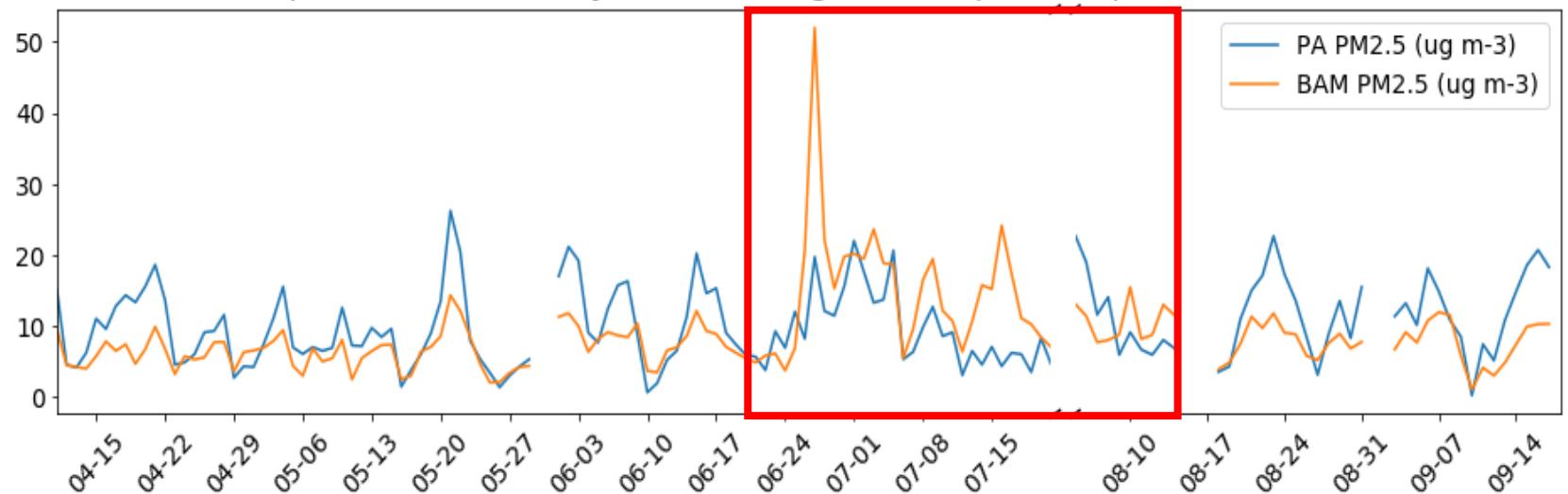
Does dust affect sensor performance?

- A mass of hot, dry and dusty air forming over the Sahara
- It moves across the North Atlantic every 3-5 days in late spring to early fall
- Largest event in decades came this past summer from ~June - August

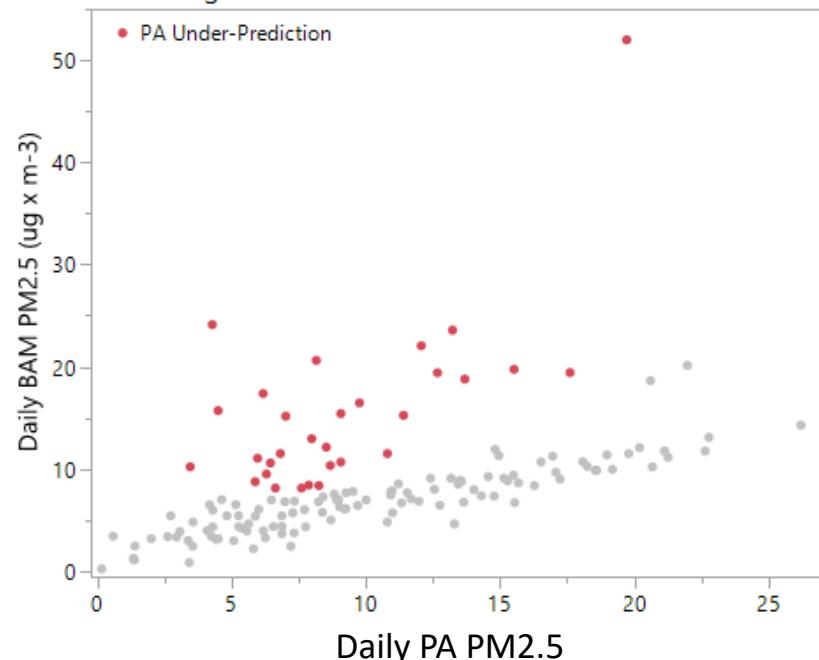


Does dust affect sensor performance?

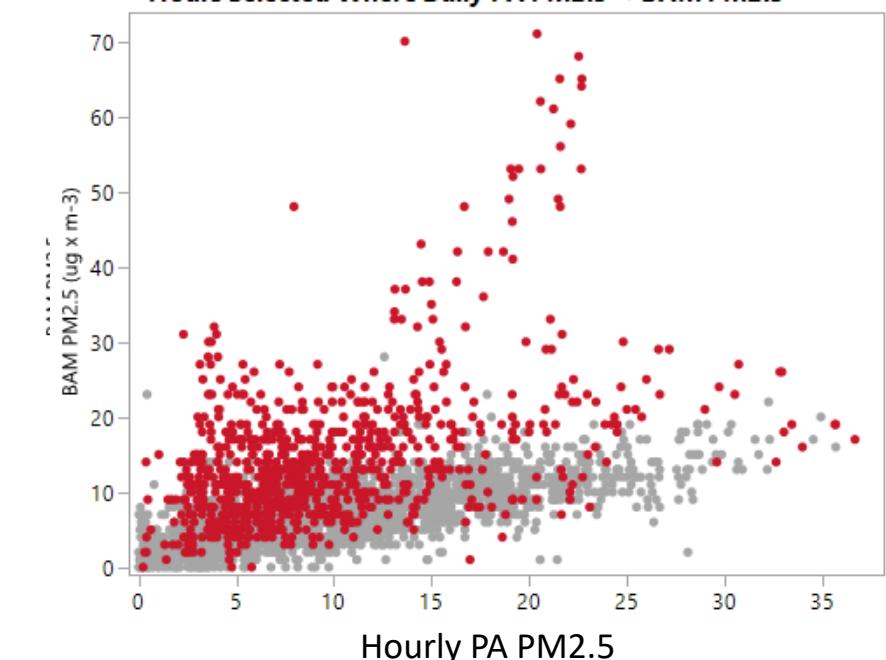
PurpleAir and BAM Daily PM2.5 Averages from April to September 2020



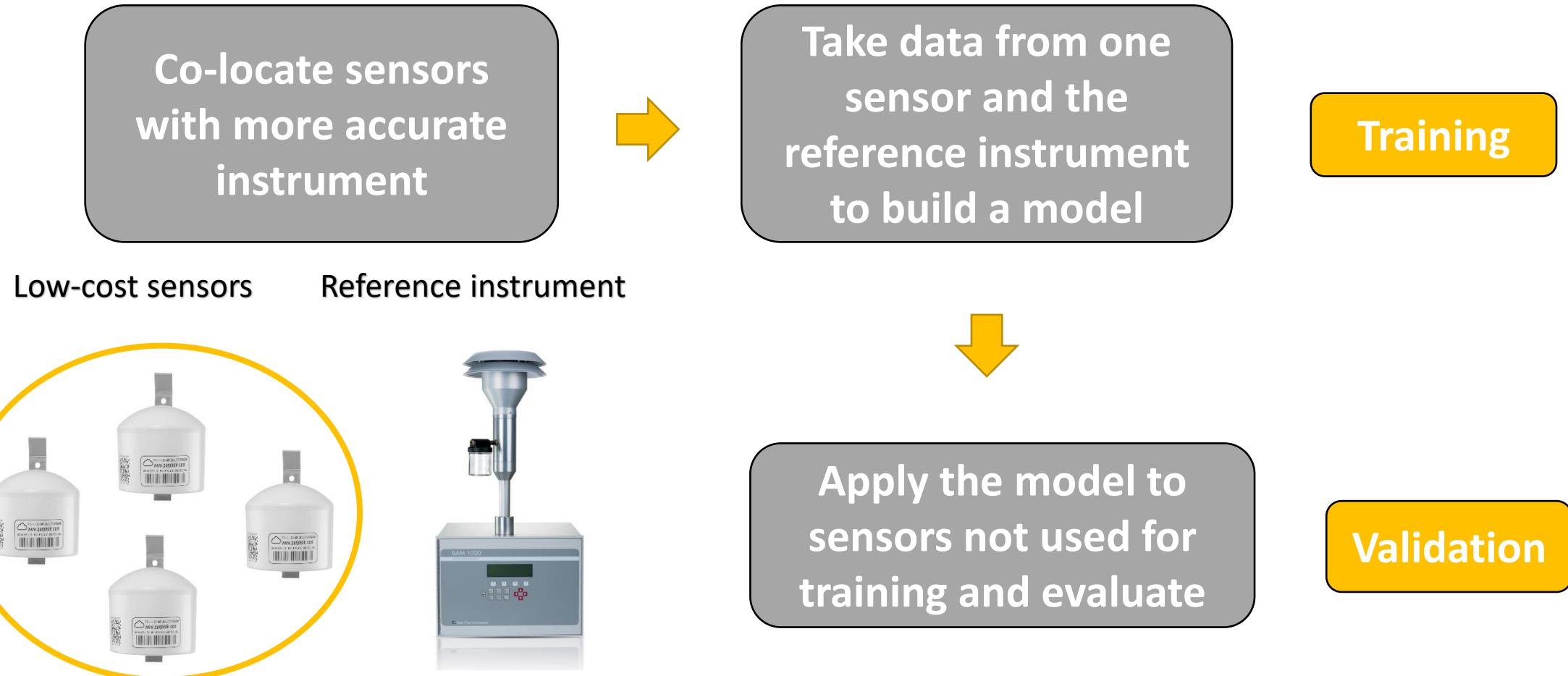
Estimating Dust as a Function of PA Under-Prediction



Hours Selected Where Daily PA PM2.5 < BAM PM2.5



What is calibration?



Sensor Variables



Actual PM_{2.5}



Cost: \$280

PurpleAir Sensor

- 2 PM sensors
 - PM₁, PM_{2.5}, PM₁₀
 - Particle Size Counts (P0.3, P0.5, ... P10)
- RH, T sensors



Cost: >\$20,000

MetOne BAM 1020

- Measures PM_{2.5}

Questions & Objectives:

- **Q1: Can a calibration model built from one sensor generalize to many sensors?**
 - Build a calibration model using ML algorithms
 - Test different variables
 - Use cross-validation to select the model that generalizes best
 - Best prediction accuracy
- **Q2: Can dust be identified from PurpleAir sensors?**
 - Create a filter to separate data from PurpleAir variables
- **Q3: Is partitioning data by dust, and subsequently building two models rather than one, a better strategy for calibrating low-cost sensors?**
 - Evaluate performance of split-model strategy vs. single model strategy

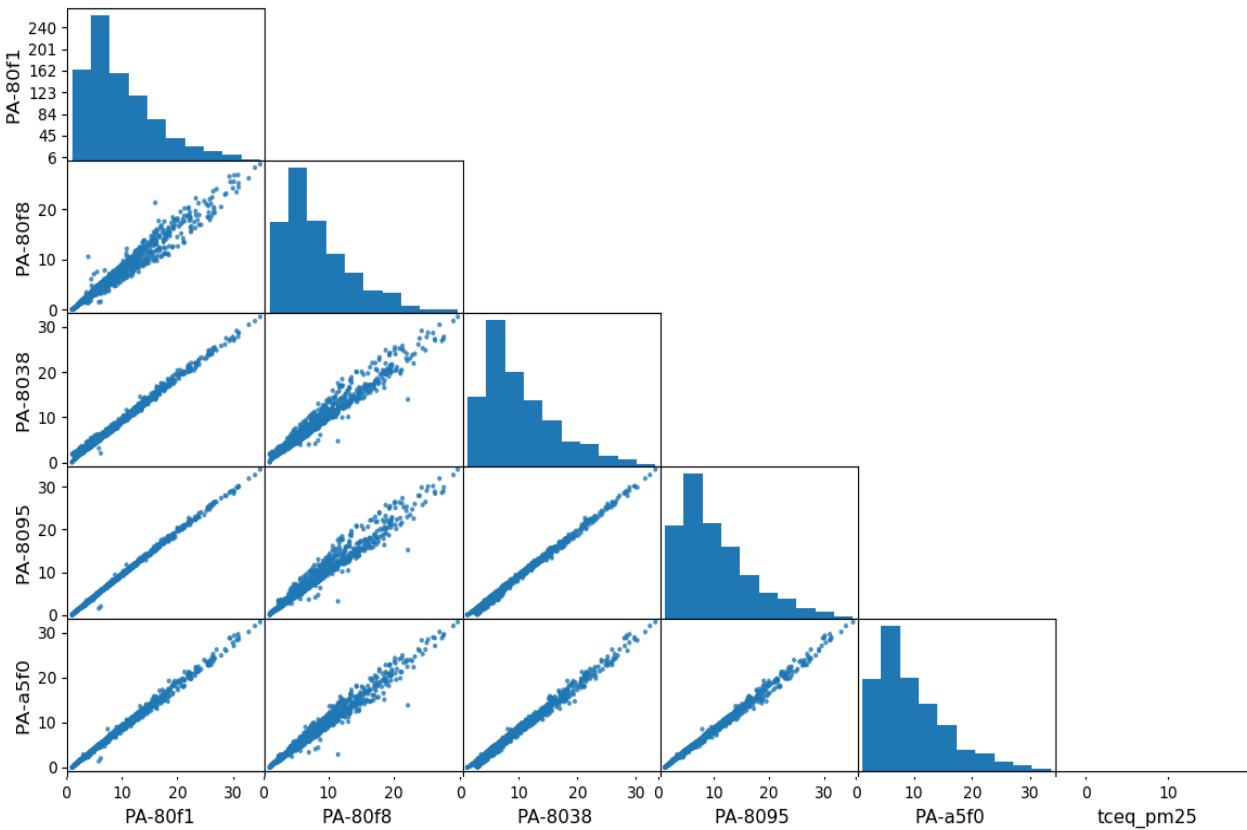
**Context &
Questions**

Methods

**Results &
Discussion**

Future Work

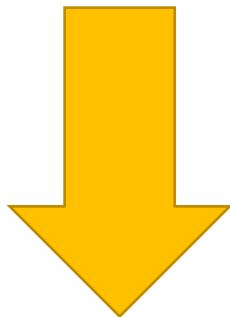
Data Collection



Deployment Period	Training Set 1	Validation Set 2	Validation Set 3	Validation Set 4
April - May	PA - Sensor	PA - Sensor	PA - Sensor	PA - Sensor
June	PA - Sensor	PA - Sensor	PA - Sensor	PA - Sensor
July	PA - Sensor	PA - Sensor	PA - Sensor	PA - Sensor
August - September	PA - Sensor	PA - Sensor	PA - Sensor	PA - Sensor

Objective: Build Calibration Models Using Two Strategies

Build a Single Model



Build a Dust and Non-Dust Model

Split each dataset according to a dust filter to create “dust” and “non-dust” datasets



Train algorithms on training data

Linear Regression & Random Forest

Evaluate on 3 validation datasets

Mean Absolute Error and Adj. R-Squared

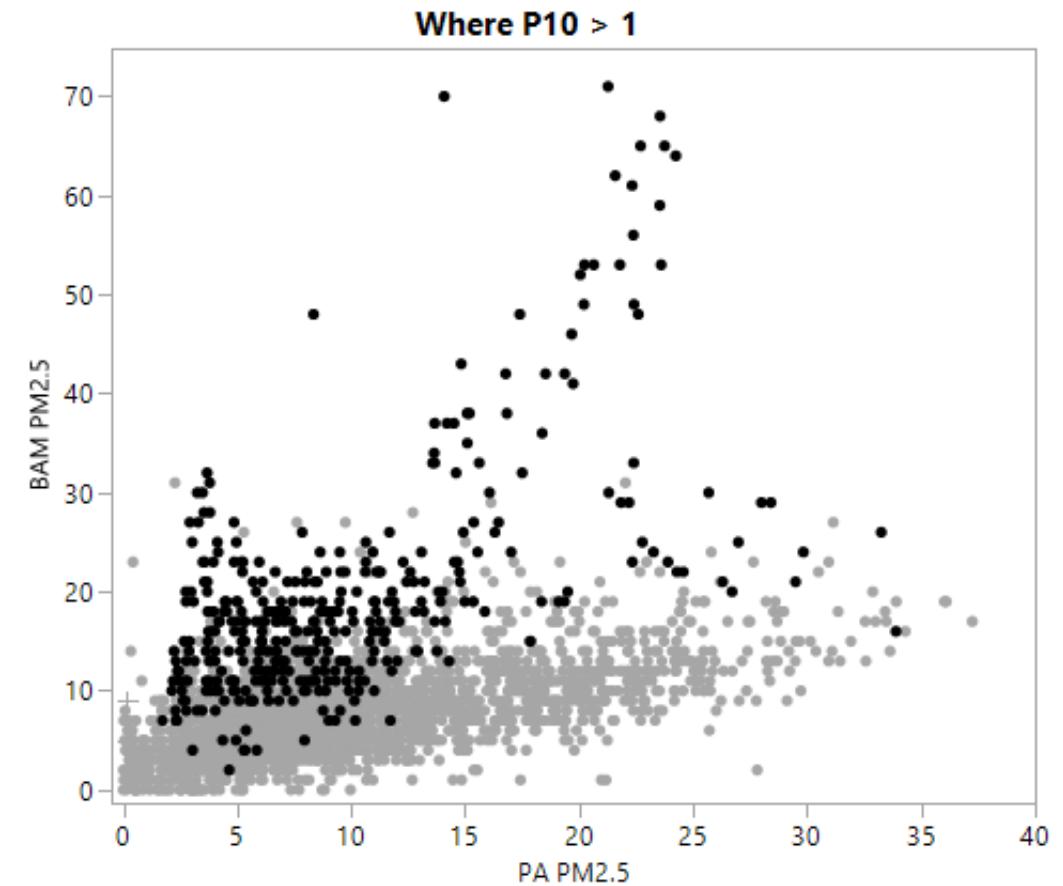
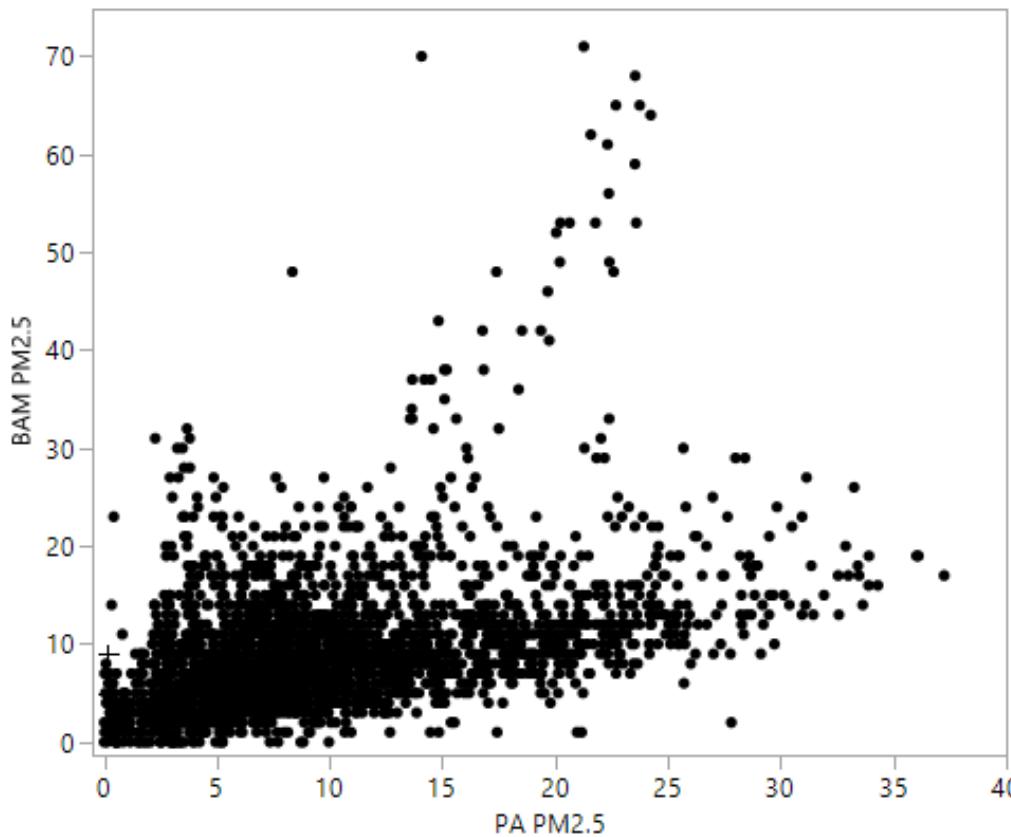
Objective: Evaluate Using Different Variables

- PM: Mass concentration ($\text{ug} \times \text{m}^{-3}$)
 - **PM2.5, PM1, PM10**
- Meteorological
 - **RH & T**
- New Variables: Ratios of PM
 - **PM1:PM2.5, PM2.5:PM10, PM1:PM10**
- P: Particle Size Count (μm)
 - **P0.3, P0.5, P1.0, P2.5, P5.0, P10**

	Number	Variables In Group
VG1	1	PM2.5
VG2	3	PM2.5, RH, T
VG3	5	PM2.5, RH, T, PM1.0, PM10
VG4	8	PM2.5, RH, T, PM1.0, PM10, PM1:PM2.5, PM2.5:PM10, PM1:PM10
VG5	14	PM2.5, RH, T, PM1.0, PM10, PM1:PM2.5, PM2.5:PM10, PM1:PM10, P0.3, P0.5, P1.0, P2.5, P5.0, P10.0

Objective: Creating a Filter

Dust → WHERE **PM2.5:PM10 < Threshold**
OR WHERE **P10 > Threshold**



**Context &
Questions**

Methods

**Results &
Discussion**

Future Work

Single Model

Average Performance Across 3 Sets

RF

MAE = 3.1
R2 = 0.5

MLR

MAE = 3.27
R2 = 0.44

Variables Added

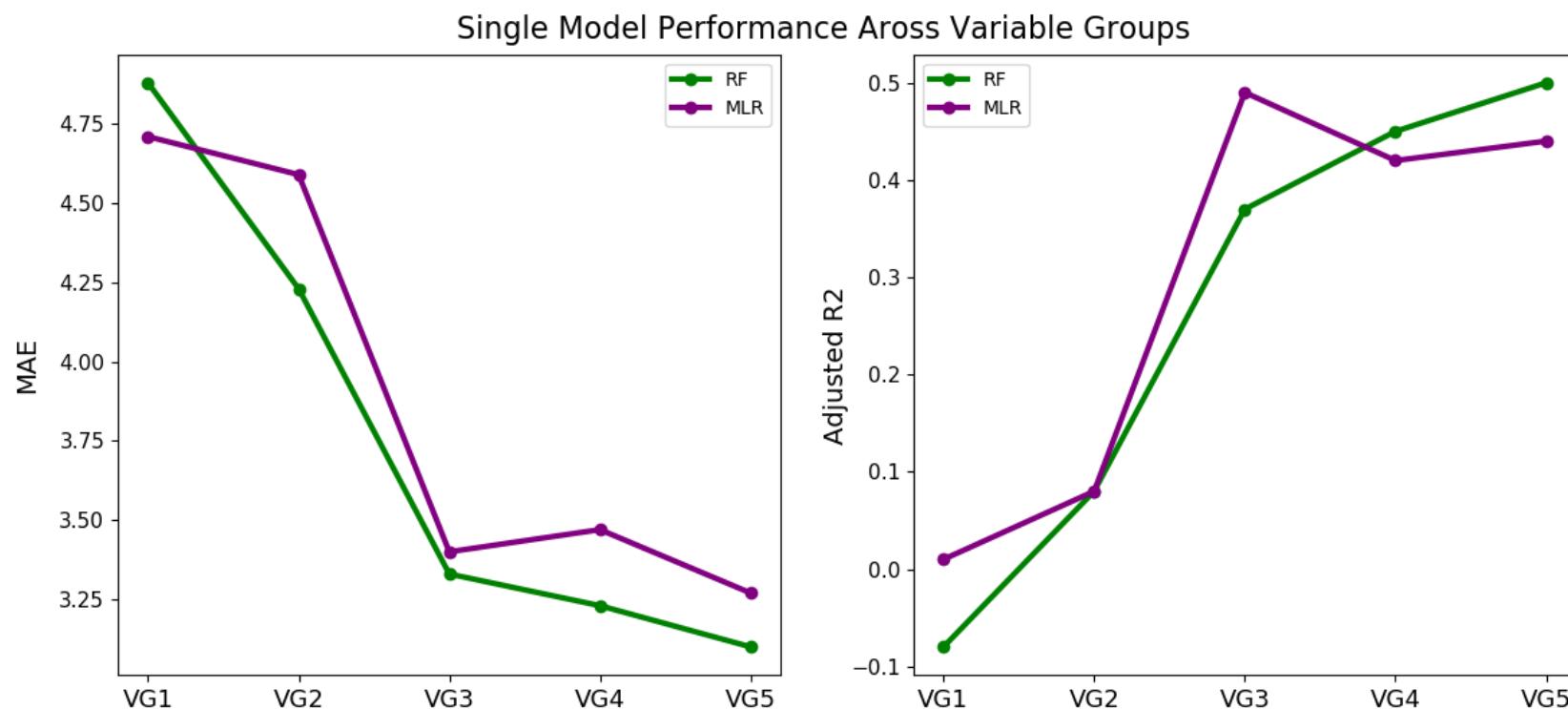
VG1 PM2.5

VG2 +RH, T

VG3 +PM1.0, PM10

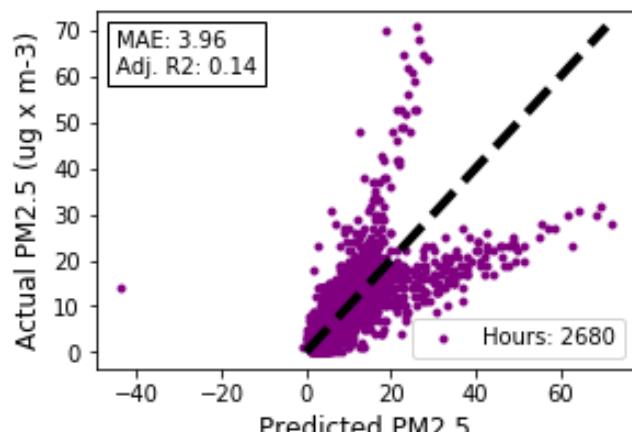
VG4 +Ratios (PM2.5:PM10, ...)

VG5 +Particle Size Counts (P0.3, ...)

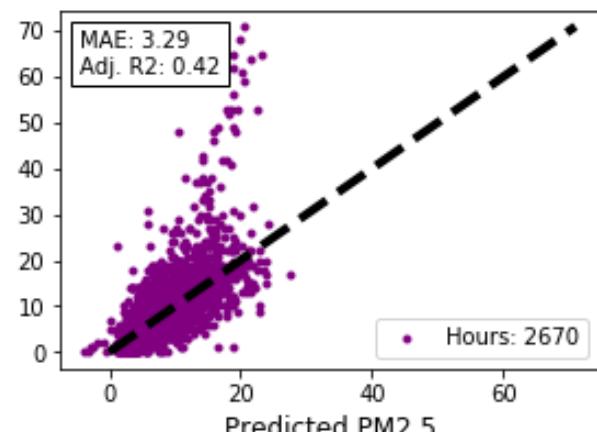


Single Model MLR Performance Across Validation Sets for VG5

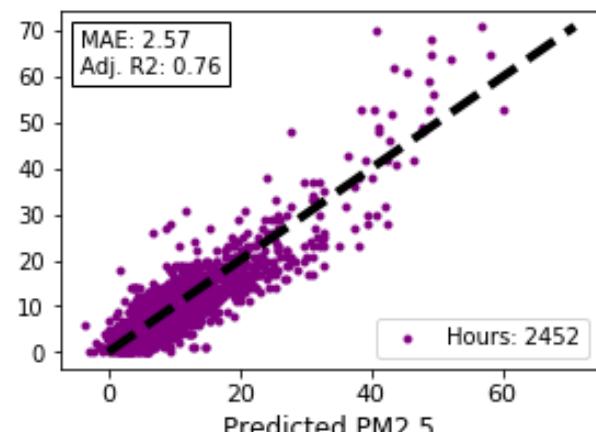
Validation: Set 2



Validation: Set 3



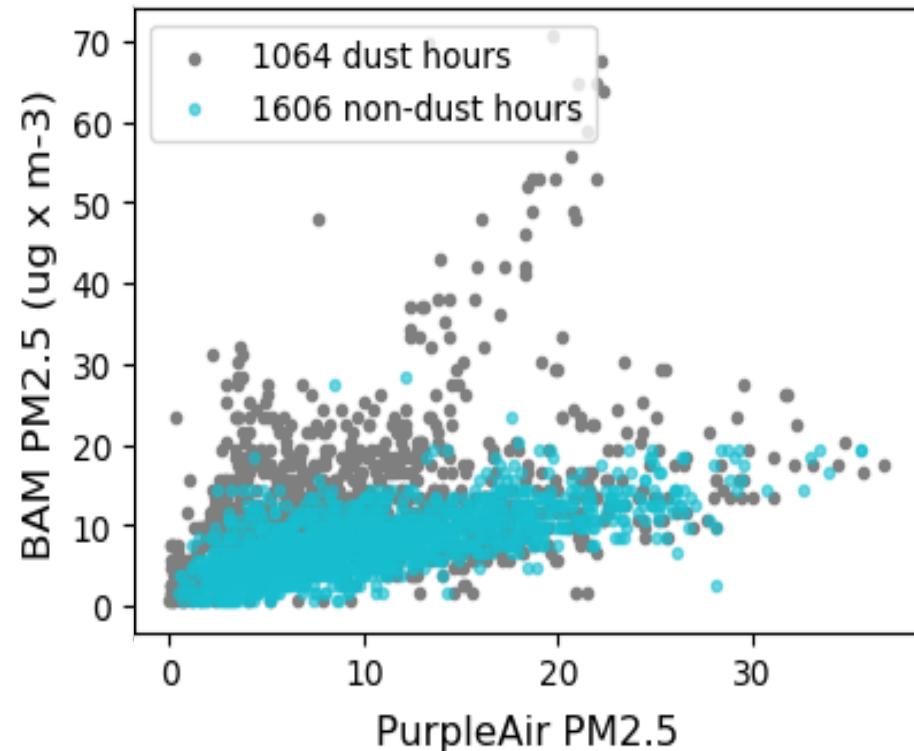
Validation: Set 4



Separating the Data by Filter

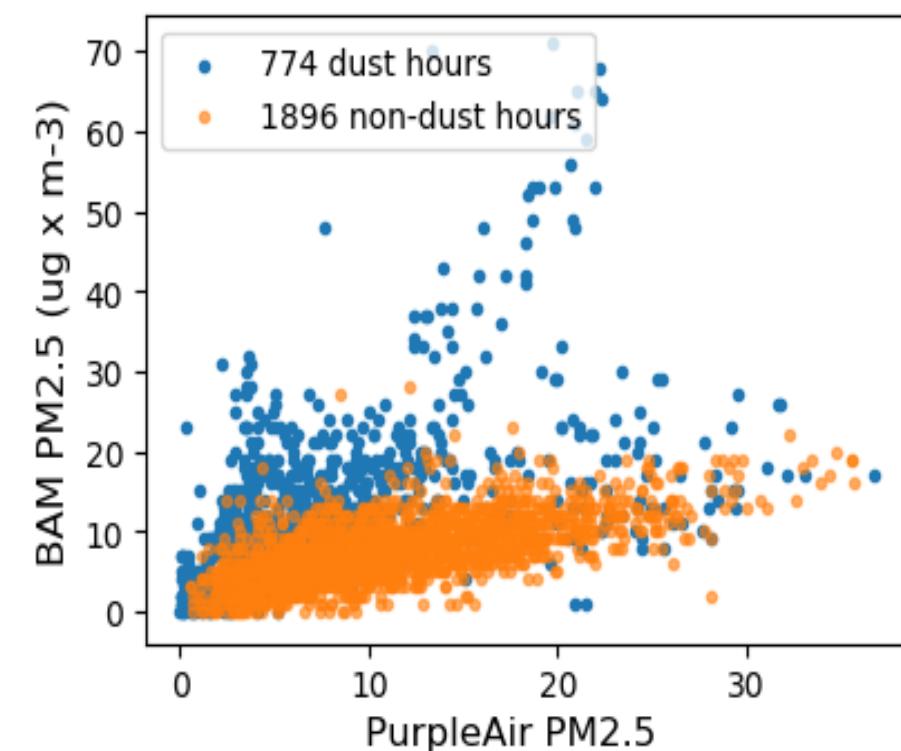
Random Forest

Dust WHERE $\text{PM2.5:PM10} < 0.84$ OR $\text{P10} > 0.4$



Multiple Linear Regression

Dust WHERE $\text{PM2.5:PM10} < 0.86$ OR $\text{P10} > 0.8$



Split-Model: Dust

Average Performance Across 3 Sets

RF

nMAE = 0.29
R2 = 0.64

MLR

nMAE = 0.22
R2 = 0.82

Variables Added

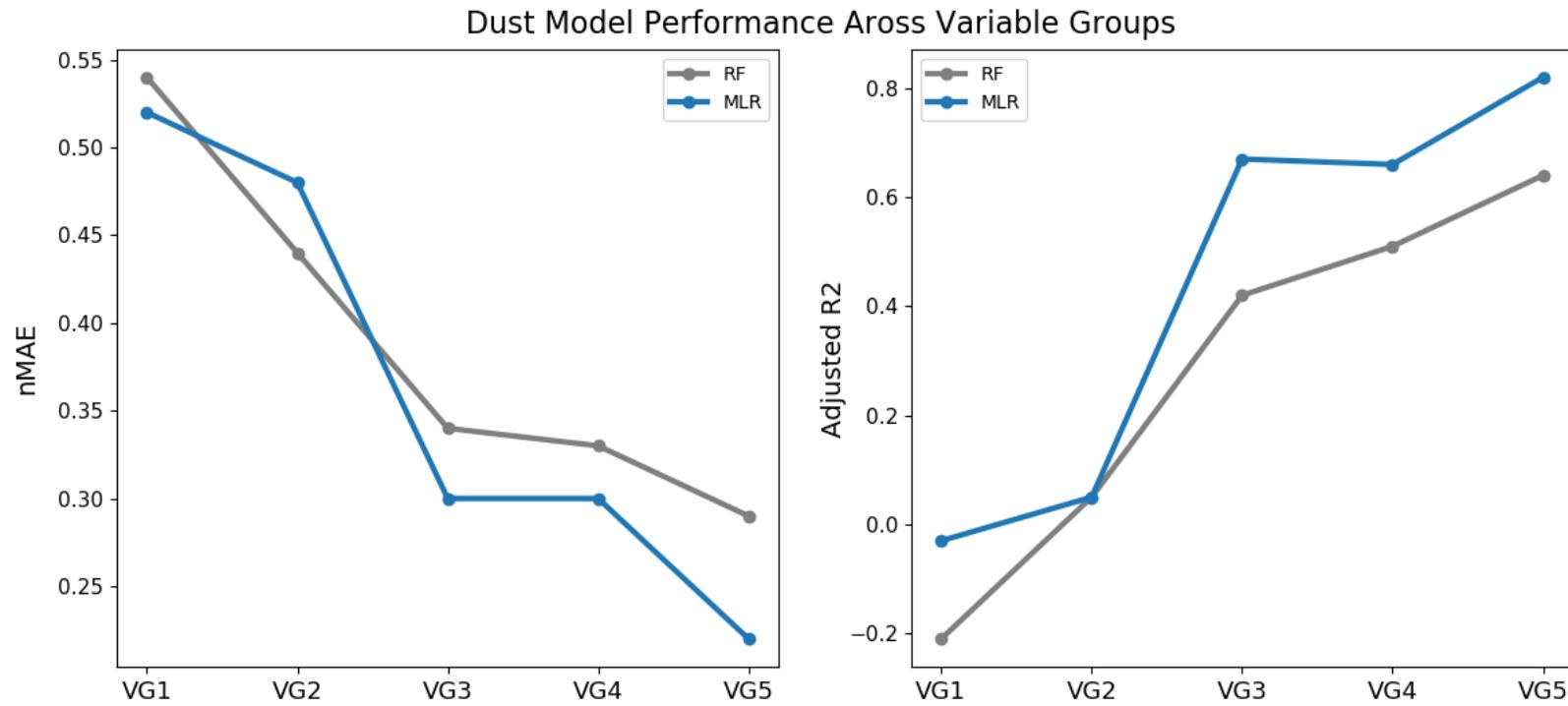
VG1 PM2.5

VG2 +RH, T

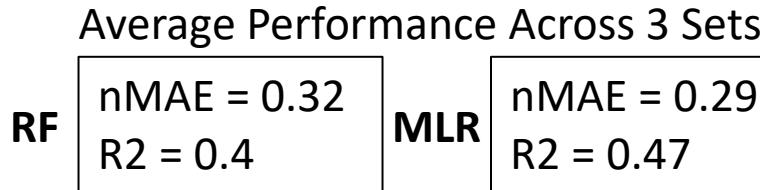
VG3 +PM1.0, PM10

VG4 +Ratios (PM2.5:PM10, ...)

VG5 +Particle Size Counts (P0.3, ...)

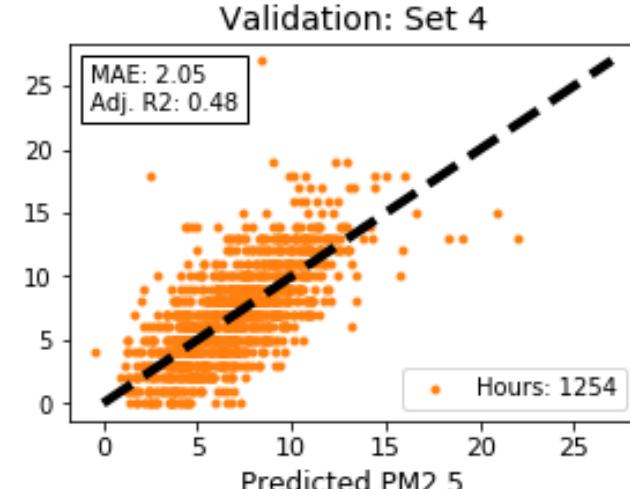
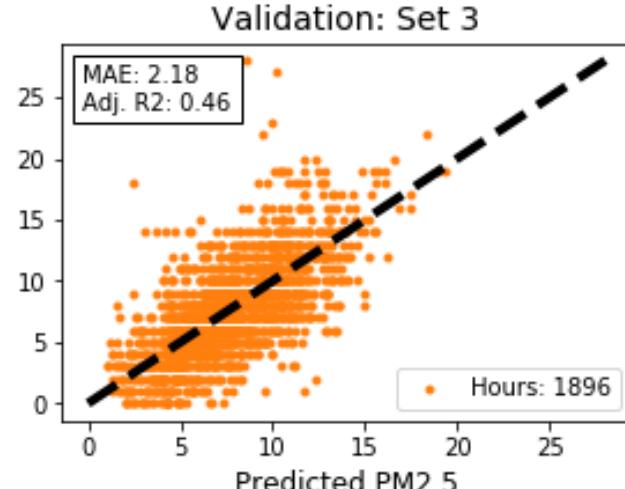
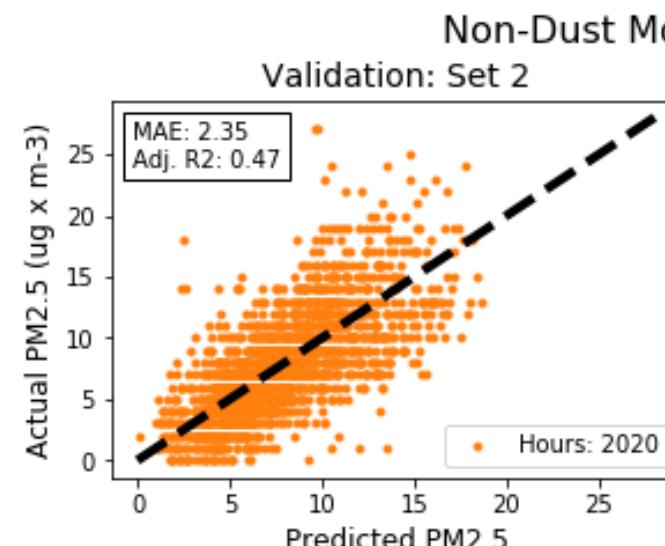
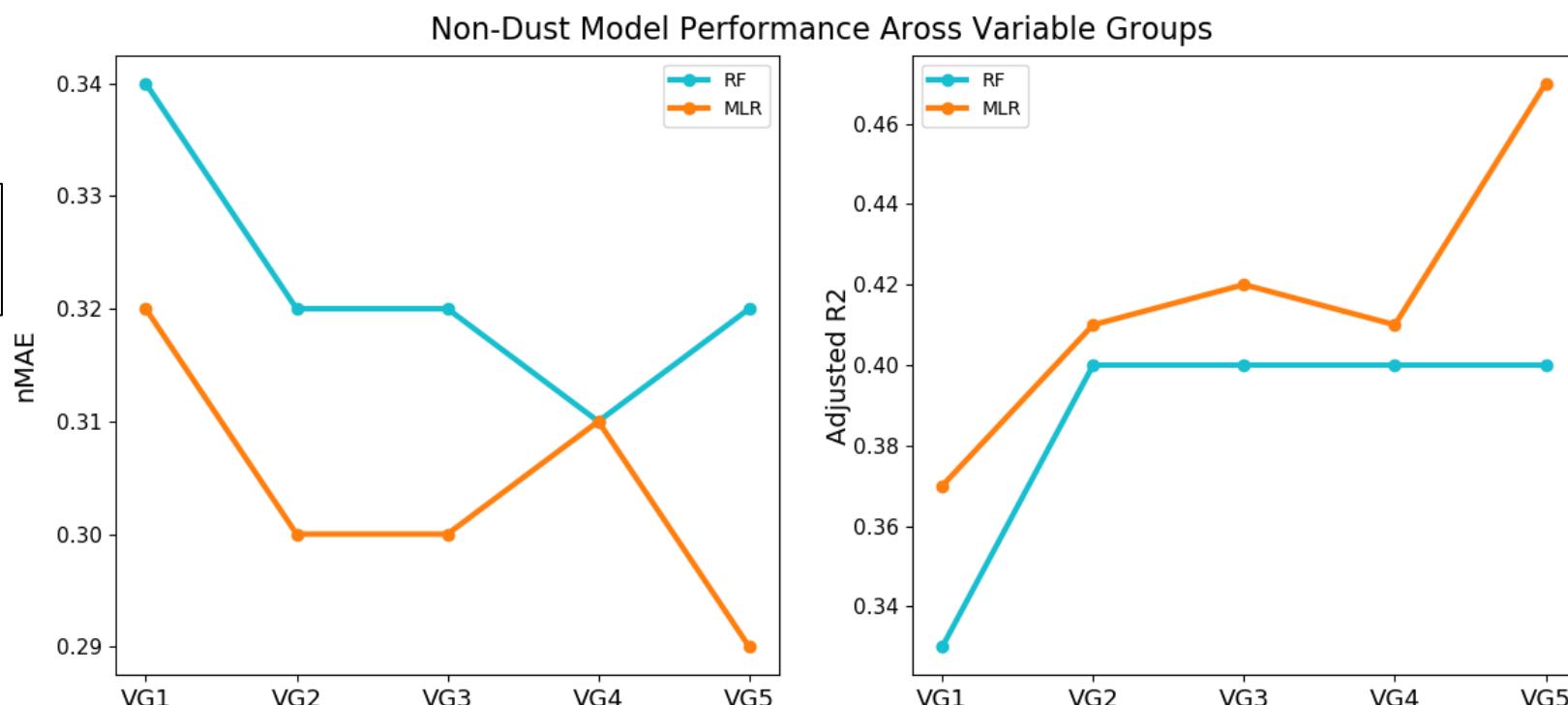


Split-Model: No dust



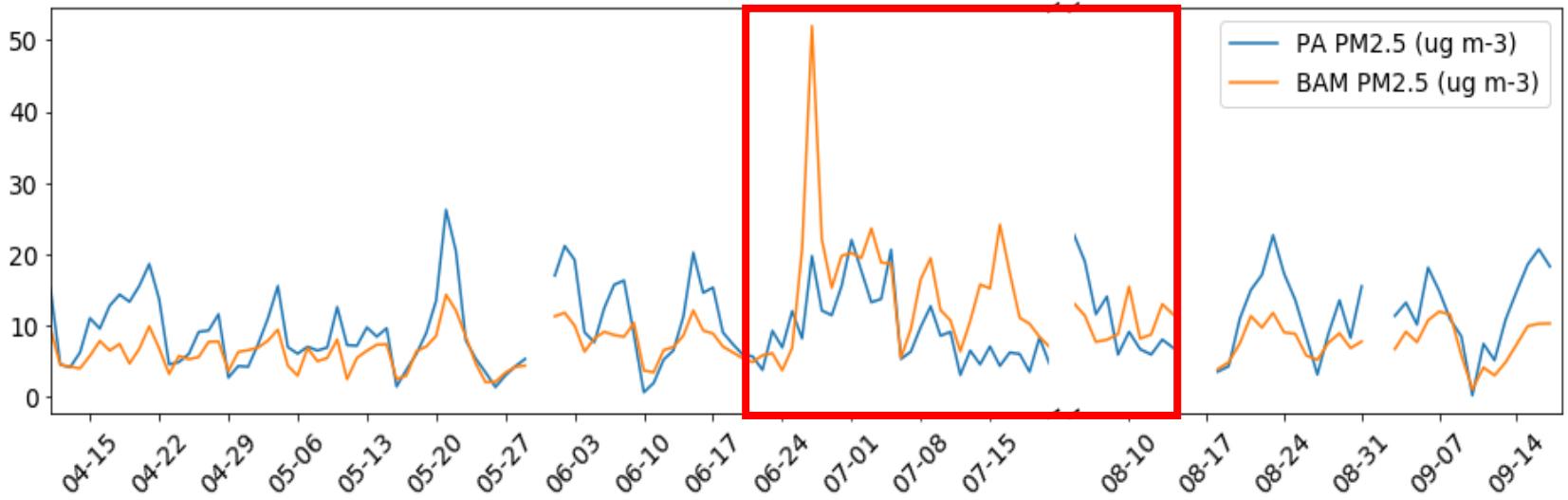
Variables Added

VG1	PM2.5
VG2	+RH, T
VG3	+PM1.0, PM10
VG4	+Ratios (PM2.5:PM10, ...)
VG5	+Particle Size Counts (P0.3, ...)

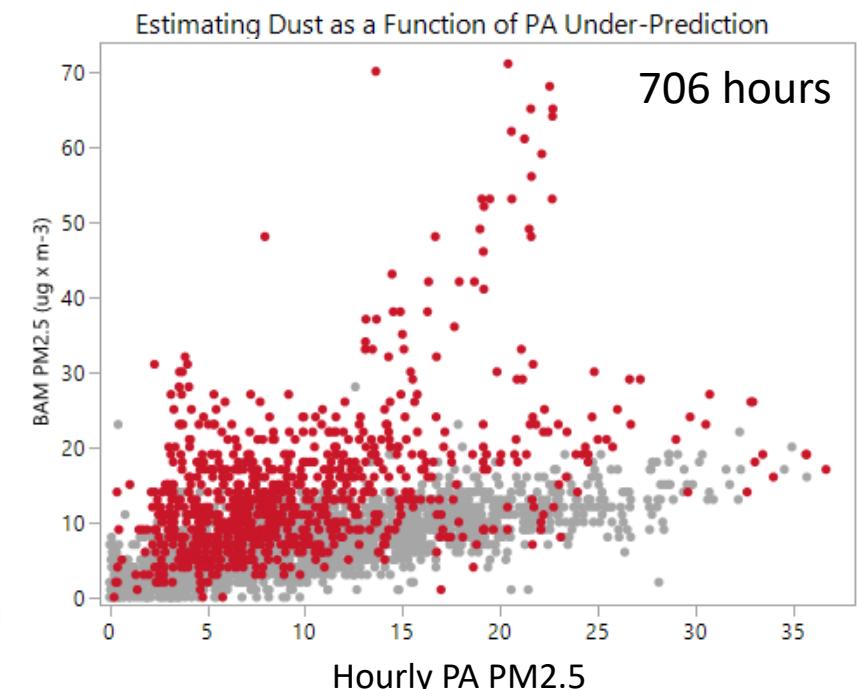
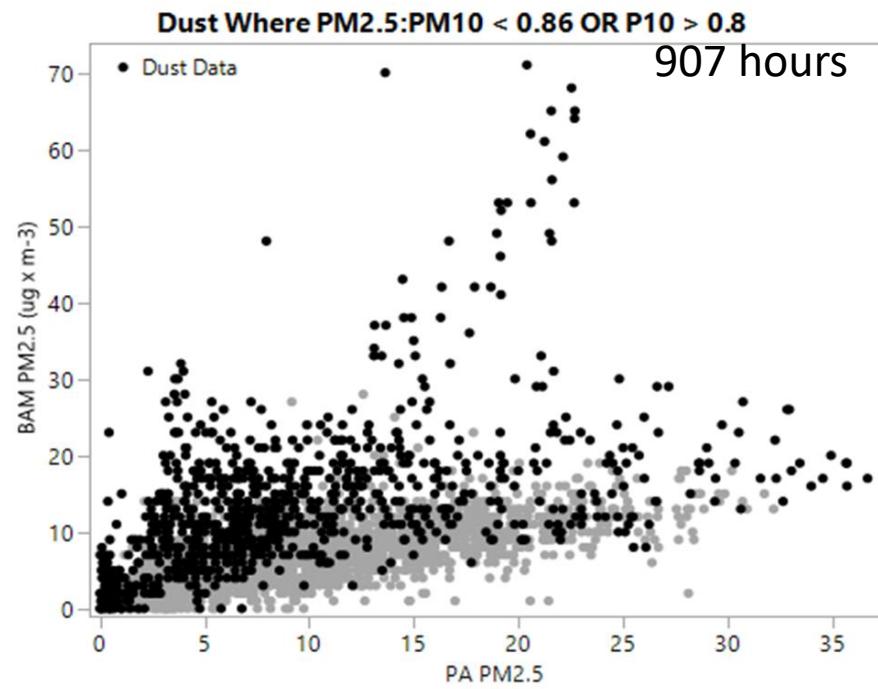


The split-model approach looks good, but does the filter make sense?

PurpleAir and BAM Daily PM2.5 Averages from April to September 2020



The filter selected 584 hrs of under-predicted hours as dust (83%)



Key Takeaways: The Best Approach

Split MLR

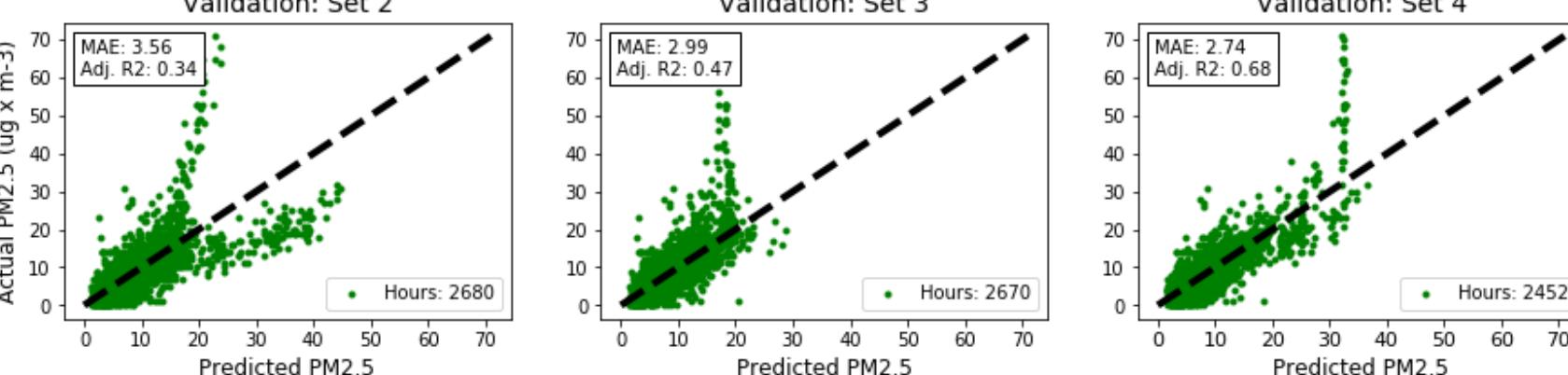
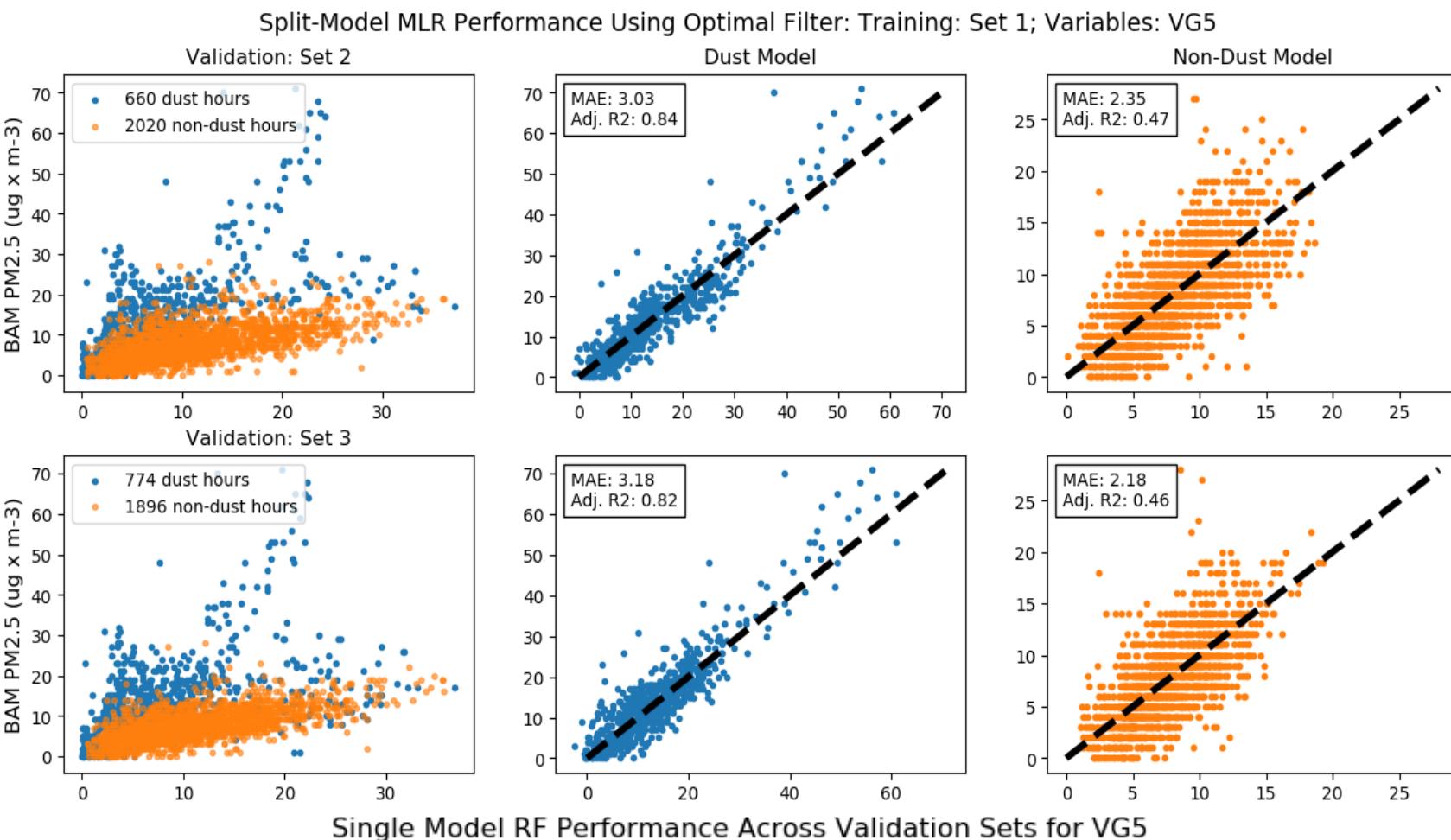
nMAE = 0.27
Adj. R2 = 0.65

Single-Model RF

nMAE = 0.32
Adj. R2 = 0.49

The split method does not hurt predictive performance. In fact, it improves it.

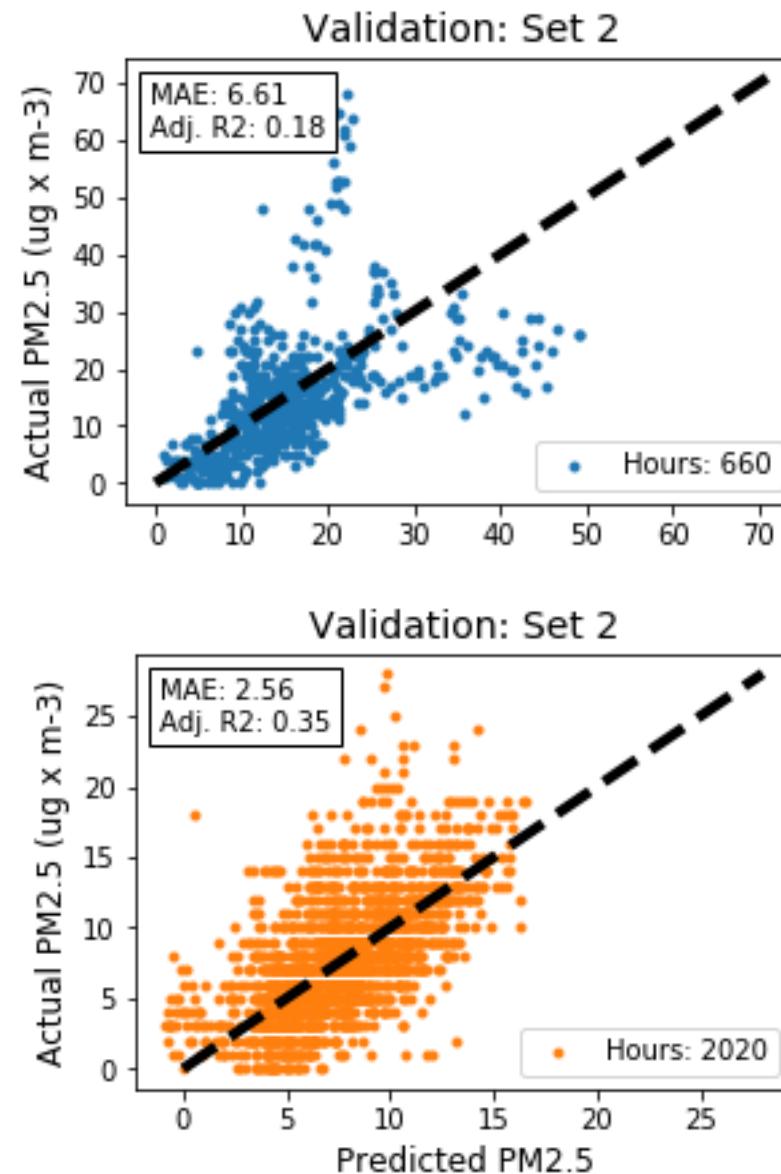
- Consistent across validation sets
- Higher accuracy
- Contextualizes the data by adding dust component



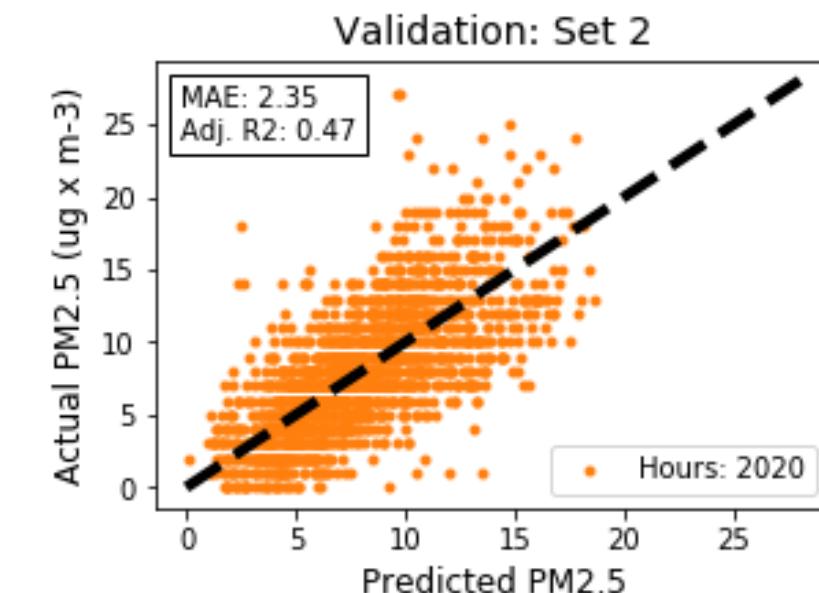
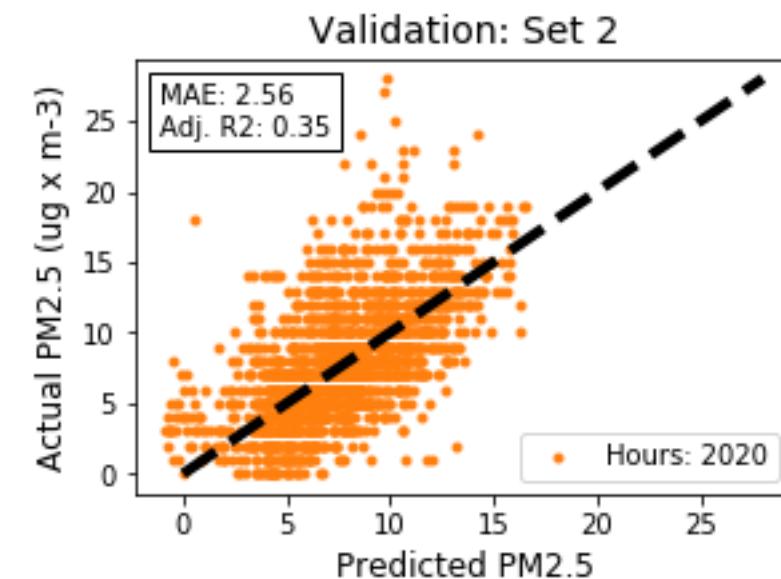
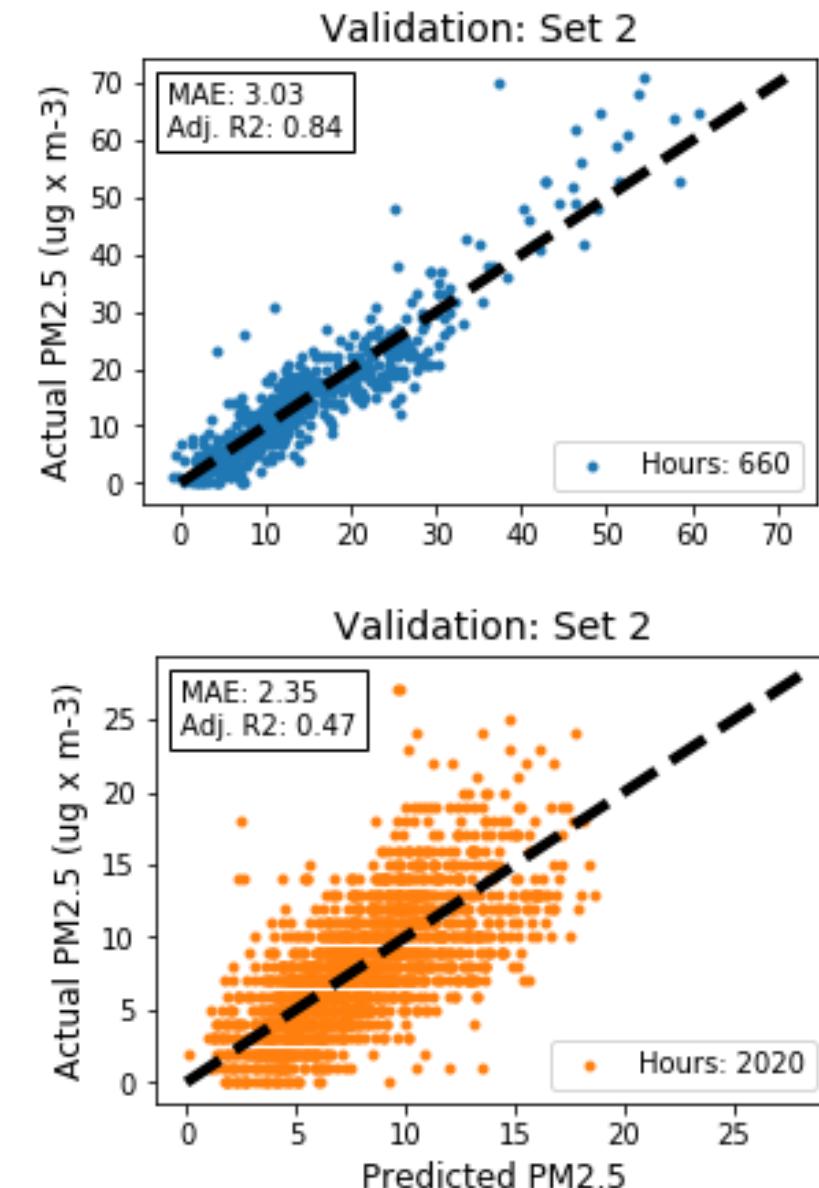
Key Takeaways: Model complexity

Relationships
between sensors and
actual PM_{2.5} more
complex during the
presence of dust

VG2: PM2.5, RH, T

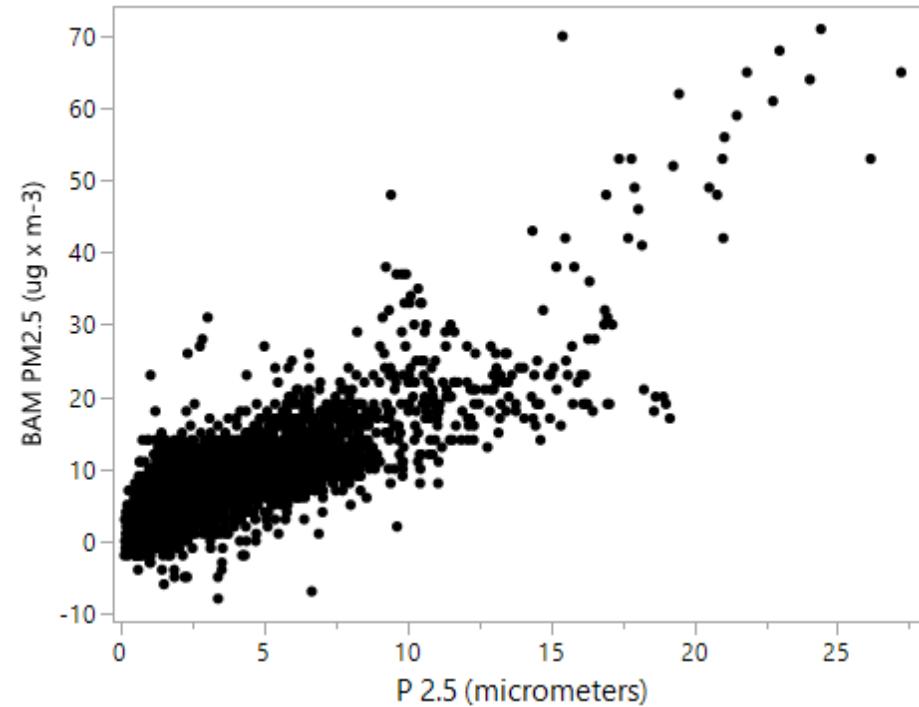
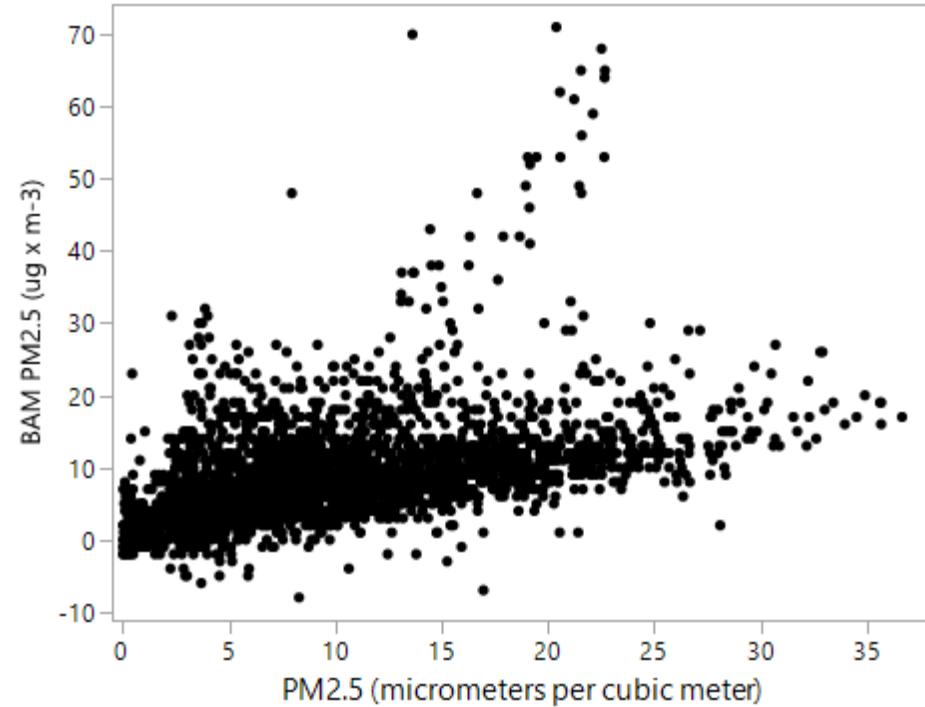


VG5: All variables



Key Takeaways: PM vs. P

Particle size counts may be better predictors of actual PM2.5 than sensor PM concentrations in some environments



Why is this significant?

PA-sensors convert particle size counts into mass concentrations according to factory calibration

**Context &
Questions**

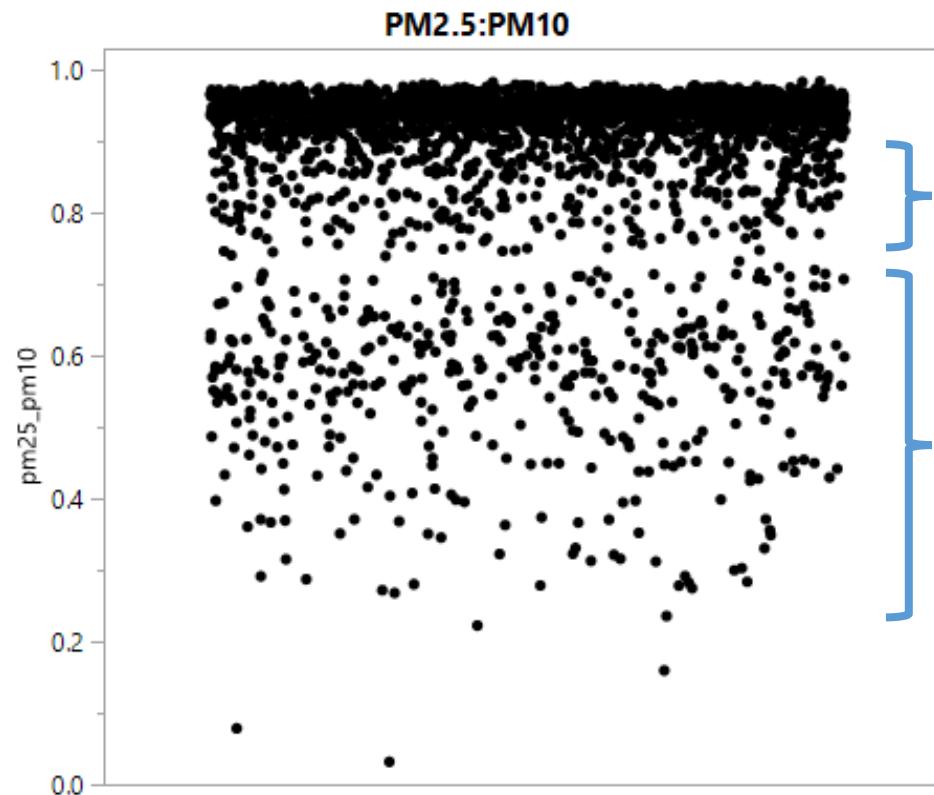
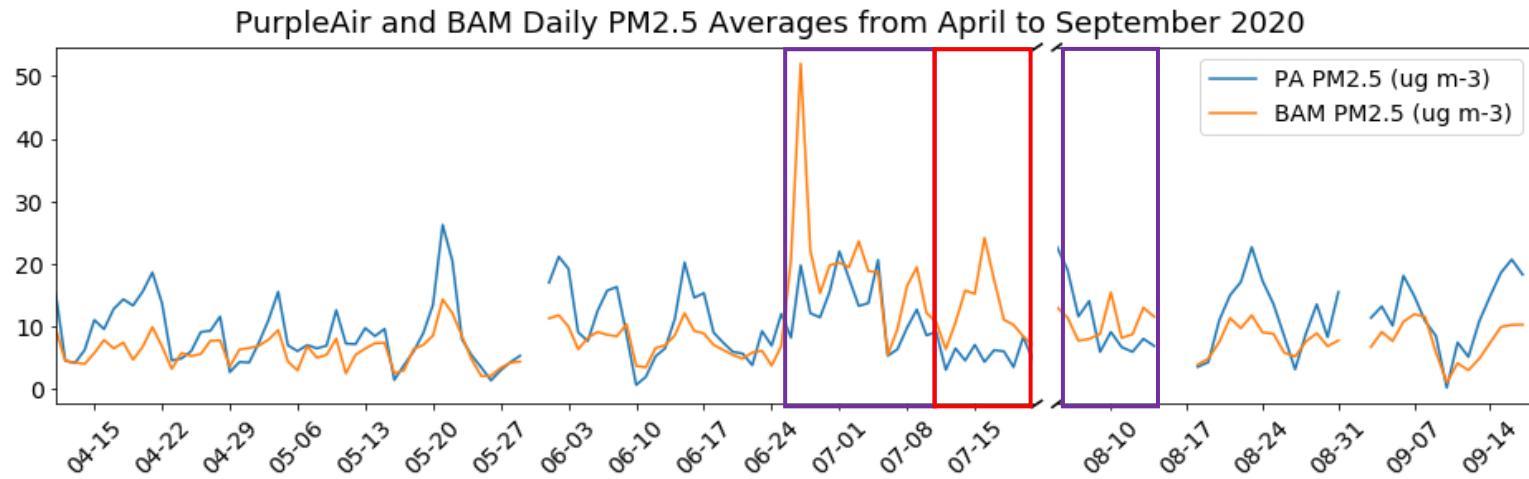
**Methods
& EDA**

**Results &
Discussion**

Future Work

Future Work

1. Apply correction model to other sensors deployed to locations in Denton to perform spatial and temporal analyses.
2. Validate model at another site and at future times.
3. Validate filter using another data source.
4. Can PA PM2.5:PM10 ratio tell us about the source of the Saharan dust?



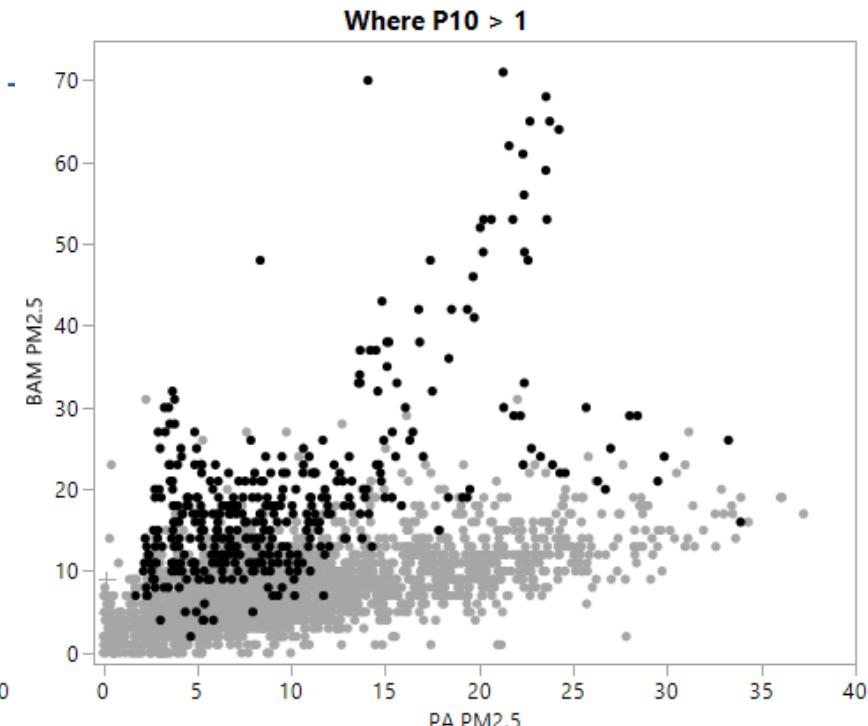
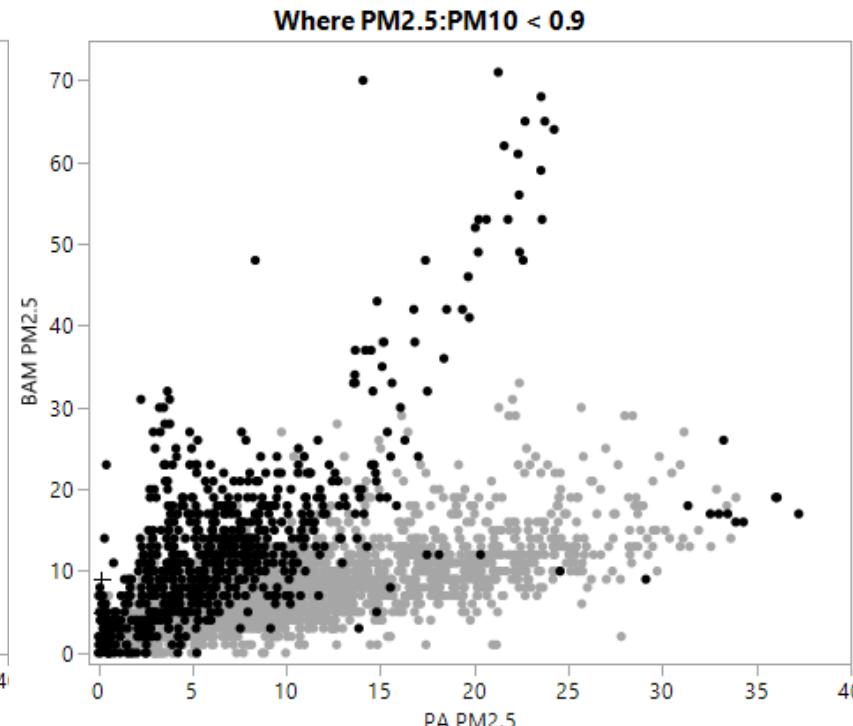
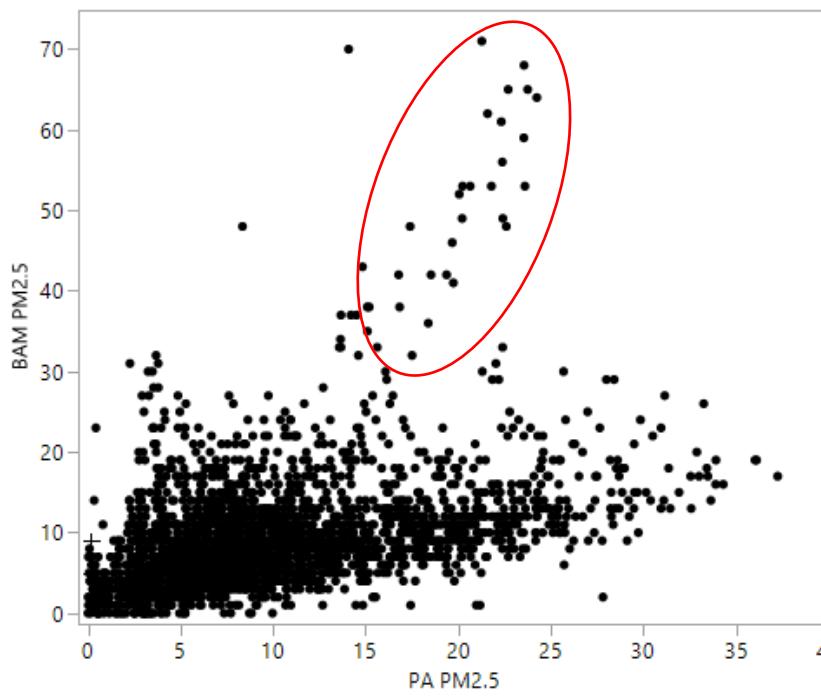
Thank you for
listening!

Questions?

Extra

Objective: Creating a Filter

Dust → WHERE **PM2.5:PM10** < Threshold
OR WHERE **P10** > Threshold



Average validation metrics for MLR Split

Dust

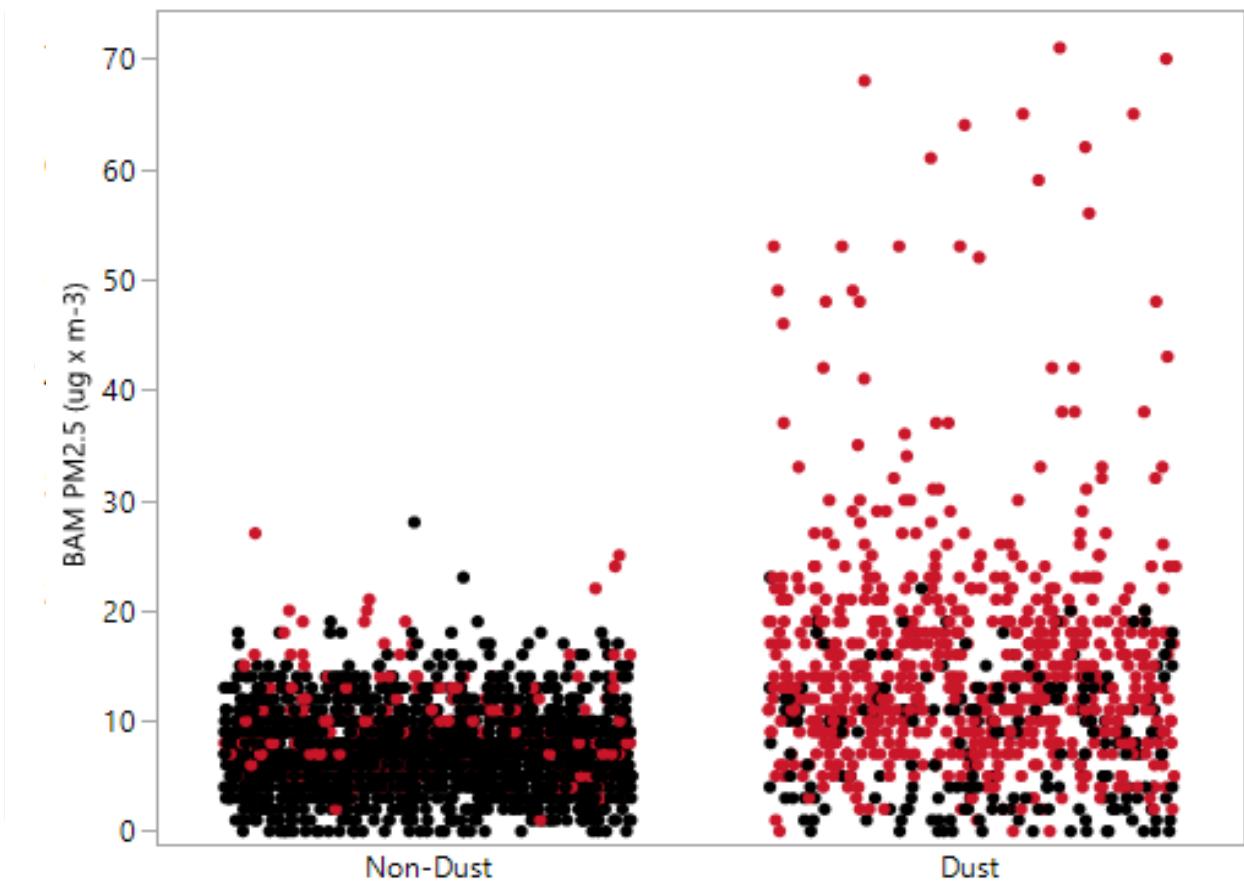
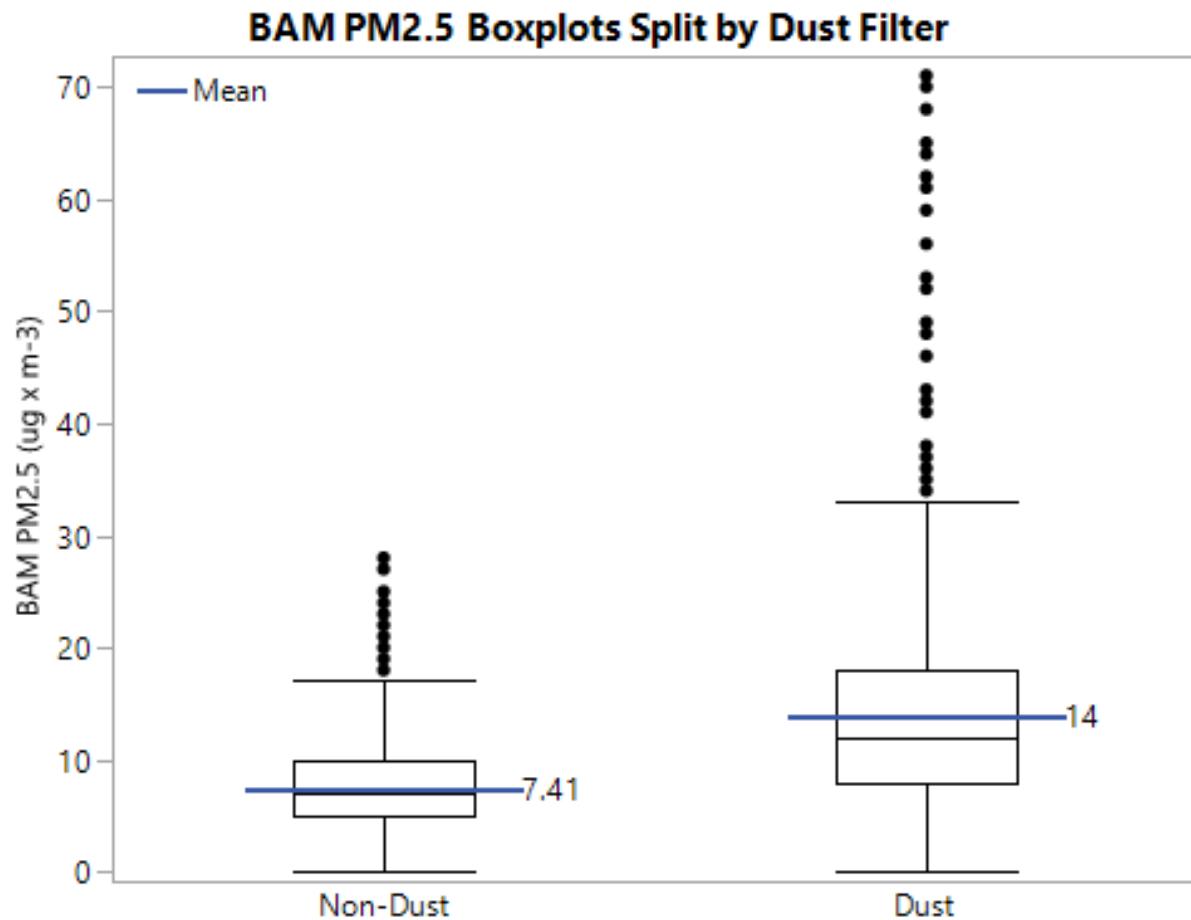
MAE = 3.01; Adj. R2 = 0.82

Non-Dust

MAE = 2.19; Adj. R2 = 0.47

The split-model results look good, but what about the filter?

584 hrs of overlap with under-prediction hours (83%)



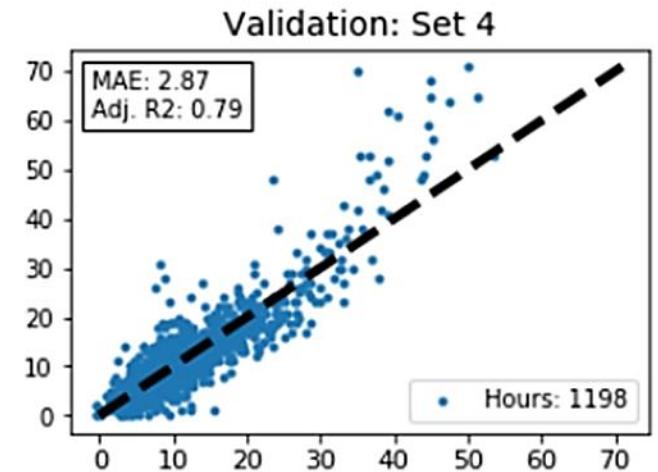
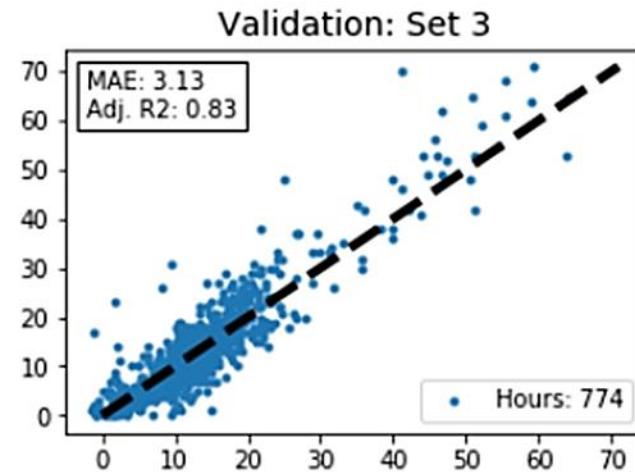
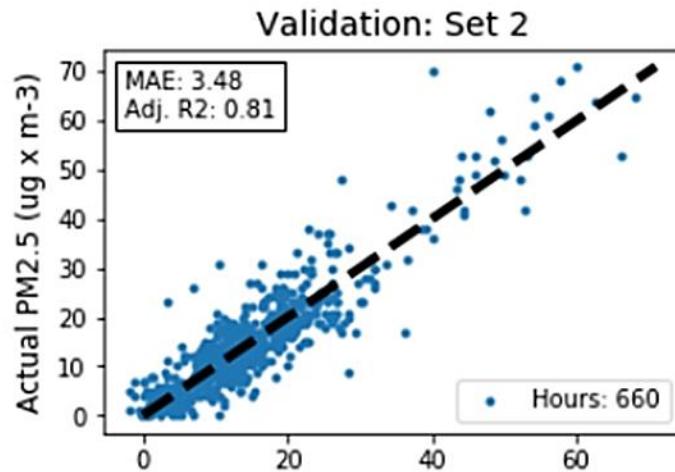
Particle Size Counts vs All Variables

Variables Added	
VG1	PM2.5
VG2	+RH, T
VG3	+PM1.0, PM10
VG4	+Ratios (PM2.5:PM10, ...)
VG5	+Particle Size Counts (P0.3, ...)

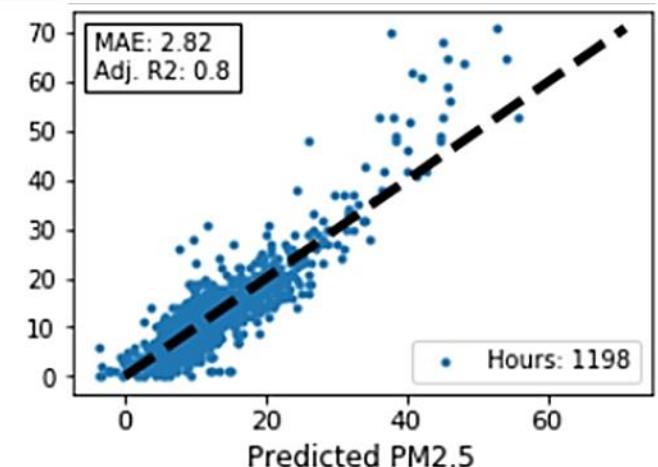
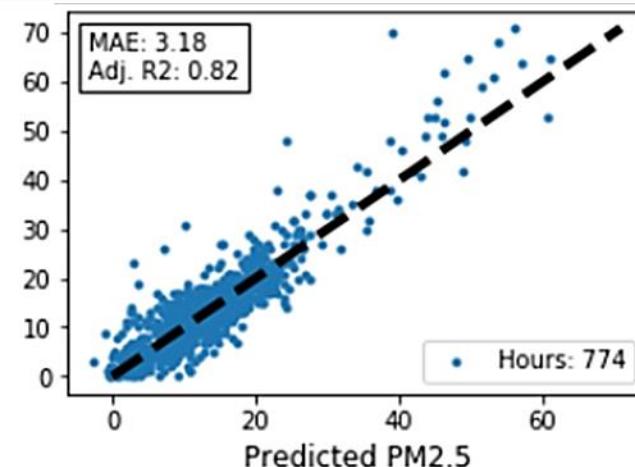
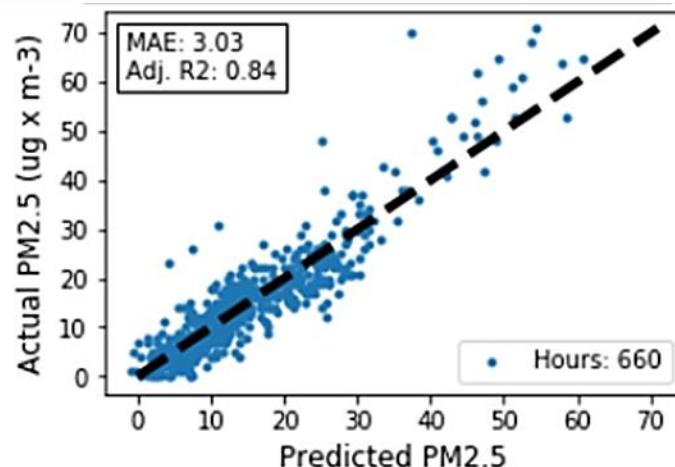
	Dust MAE	Non-Dust MAE	Dust R2	Non-Dust R2
P Count + RH, T	3.16	2.21	0.81	0.46
VG5	3.01	2.19	0.82	0.47

Using particle size counts + RH & T almost as strong as using all possible variables

RH, T +
Particle
Size
Counts

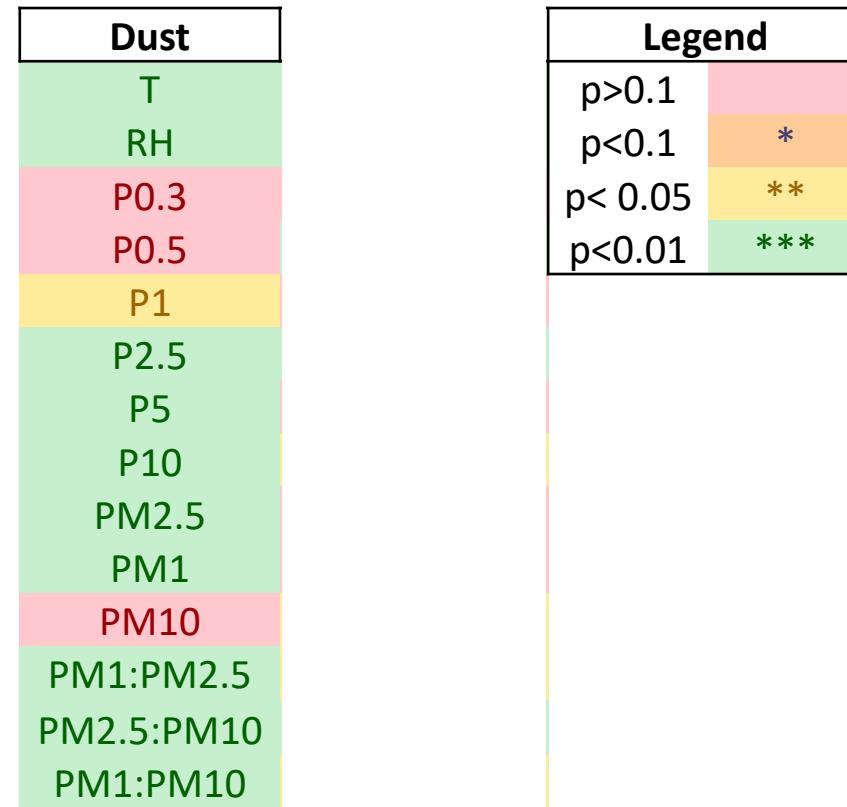


VG5: All
Variables
(adding PM
variables)



Particle Size Count Might Be a Better (or at least good of a) Predictor as Manufacturer Derived PM

- Dust
 - 9/14 significant at 99% level
 - 1 significant at the 95% level
 - Complex model needed
- Non-Dust
 - 5/14 significant at 99% level
 - 4 significant at the 95% level
 - Less complex model needed



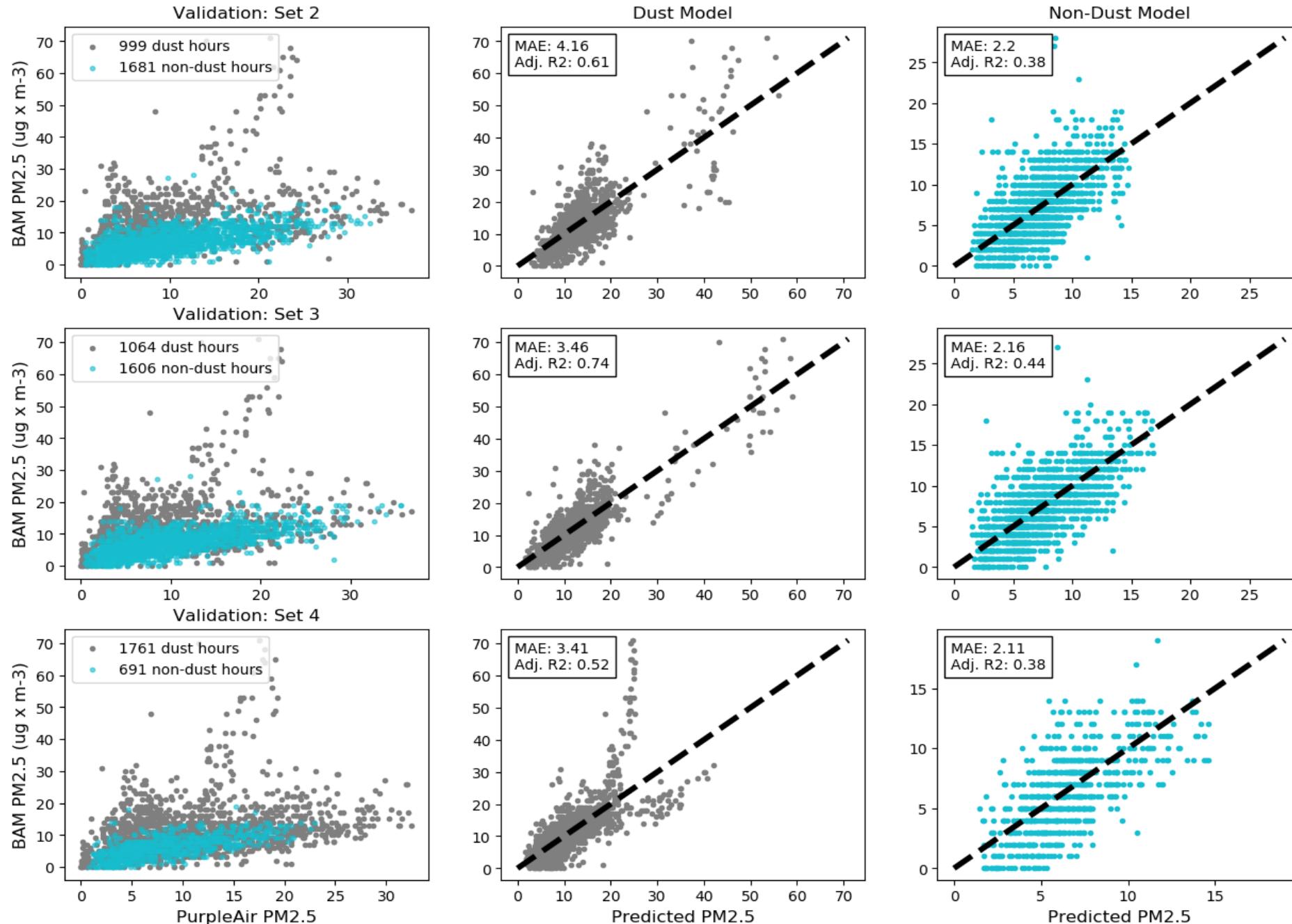
Deployment Period	Training Set 1	Validation Set 2	Validation Set 3	Validation Set 4
April - May	PA-a5f0	PA-80f1	PA-8095	PA-8038 B
June	PA-c074	PA-8079	PA-adb9	PA-8d18
July	PA-d5ec	PA-d147	PA-16a0*	PA-38da*
August - September	PA-e4f6	PA-198	PA-e396	PA-8060

Evaluating the Dust Filter

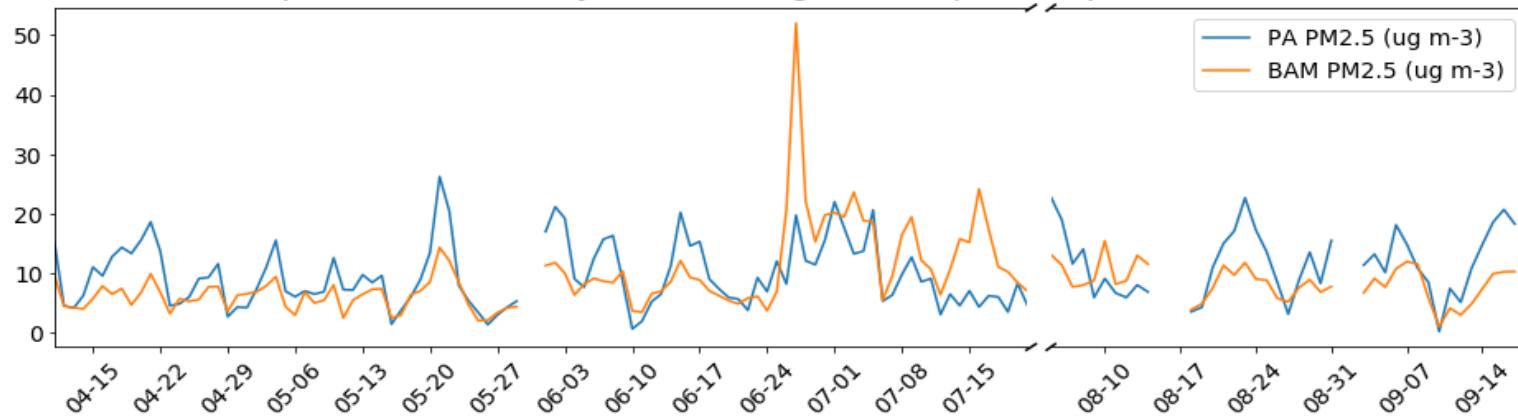
- Performance varied according to the parameters of the filter used
- A filter that optimized the dust data did not necessarily optimize the non-dust data
- A filter that optimized MAE did not necessarily optimize R2
- Filters were ranked according to MAE and R2 for both dust and non-dust models
- The ranks were summed to find

P10 Count	PM2.5:PM10 Ratio	Dust MAE	Dust R2	Non-Dust MAE	Non-Dust R2	Sum of Ranks
0.8	0.86	4	6	55	5	70
0.8	0.85	18	16	71	17	122
0.8	0.75	11	4	87	23	125
0.8	0.88	0	0	59	67	126
0.8	0.74	17	5	80	25	127
...
0.8	0.94	80	93	63	101	337
0.6	0.94	95	99	52	102	348
0.9	0.94	88	96	66	100	350
0.7	0.94	89	97	67	104	357
0.5	0.94	101	104	49	103	357

Split-Model RF Performance Using Optimal Filter: Training: Set 1; Variables: VG5



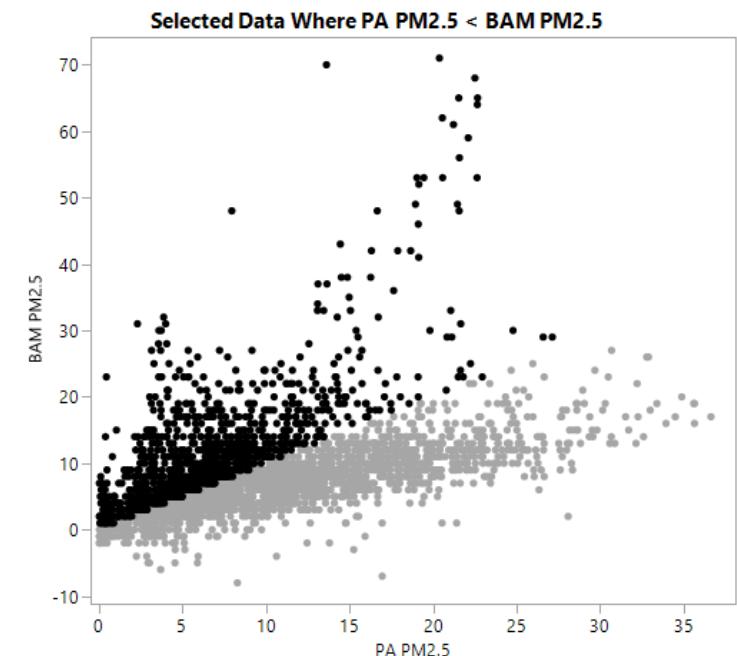
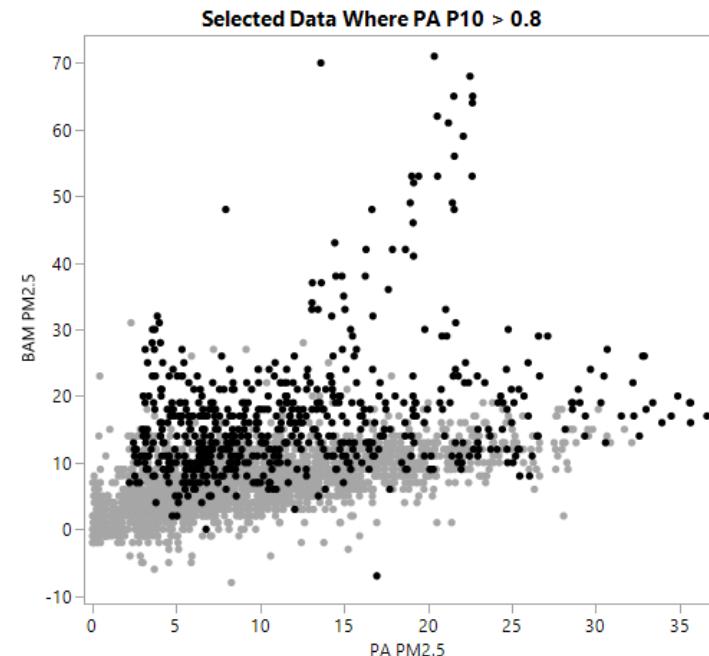
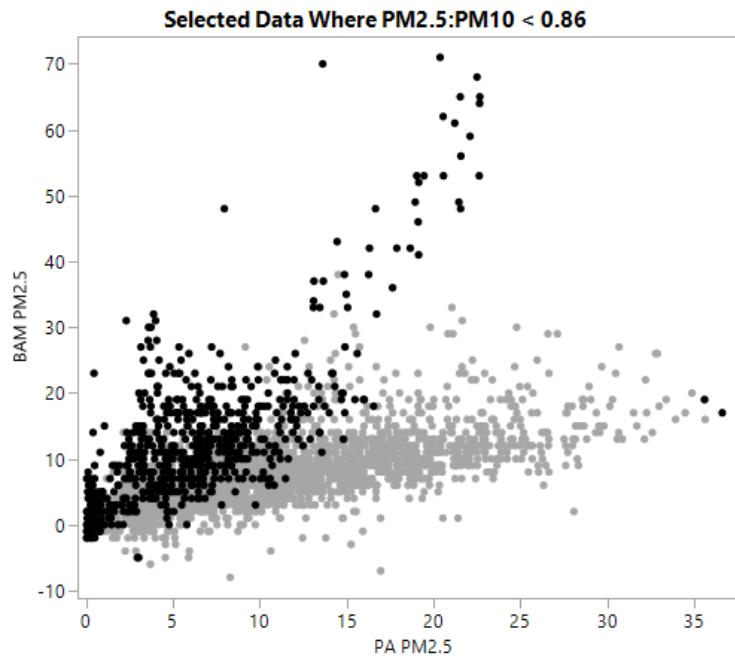
PurpleAir and BAM Daily PM2.5 Averages from April to September 2020

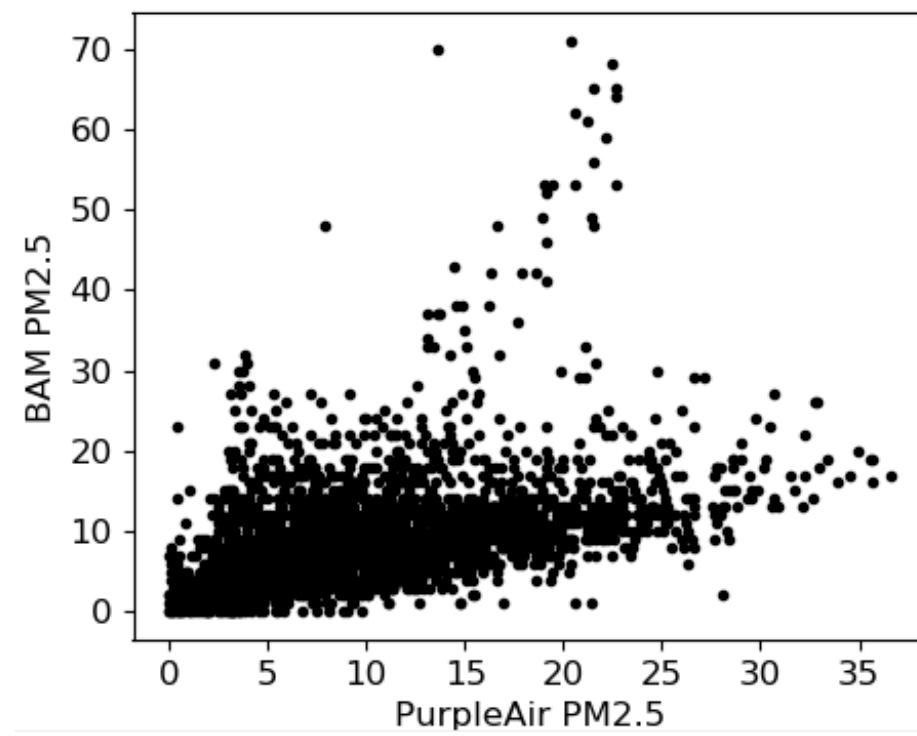
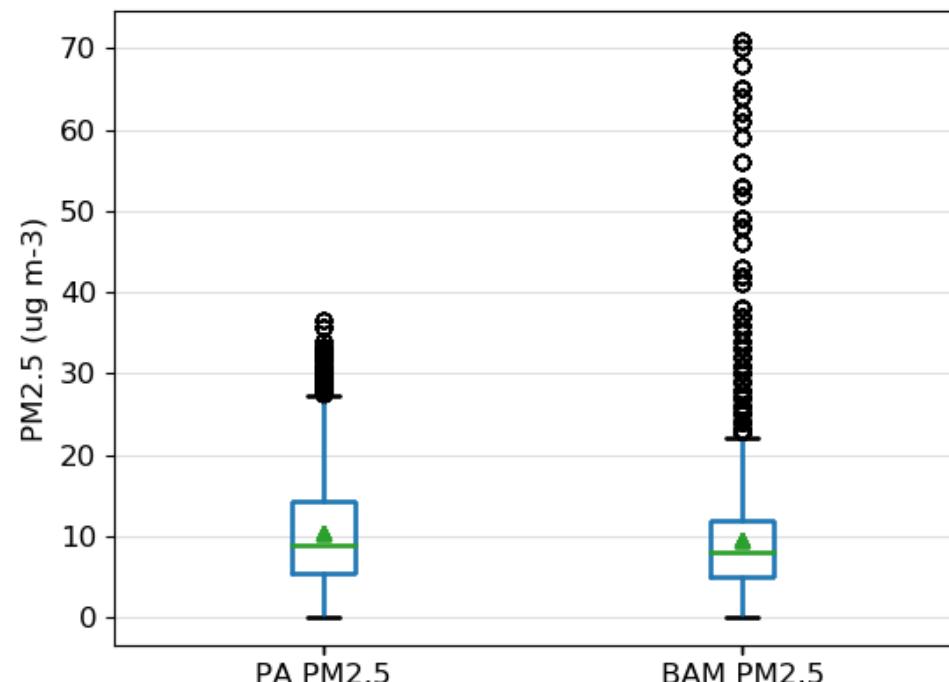


666 hours

625 hours

1108 hours





Abbreviation	Number	Variable Type	Variables In Group
VG1	1		PM: PM2.5
VG2	3	PM: Meteorological:	PM2.5 RH, T
VG3	5	PM: Meteorological:	PM2.5, PM1.0, PM10 RH, T
VG4	8	PM: Meteorological: PM Ratio:	PM2.5, PM1.0, PM10 RH, T PM1:PM2.5, PM2.5:PM10, PM1:PM10
VG5	14	PM: Meteorological: PM Ratio: Particle Size Count:	PM2.5, PM1.0, PM10 RH, T PM1:PM2.5, PM2.5:PM10, PM1:PM10 P0.3, P0.5, P1.0, P2.5, P5.0, P10.0



A New Approach: Calibrating Low-Cost Air Quality Sensors According to Pollutant Source

By: Sean Hickey

Advisor: Dr. Lu Liang

Committee Members: Dr. Pinliang Dong, Dr. Chetan Tiwari