Calibration of Spokane PM 2.5 Network Using Clarity Sensors

Matthew Roetcisoender

6/11/20

Outline

- Introduction

- Sensor Overview

- Spokane Network

- Comparison of Calibration Approaches

- Preliminary Monitoring Results

- Continuing Work

Introduction

- Study Area

  - City of Spokane ~20 mi^2

- Air Quality Concerns

  - Seasonal increases in PM 2.5 concentrations

  - Wildfires

  - Wintertime conditions

- Goals

  - Evaluate low-cost sensors as suitable option for air quality network

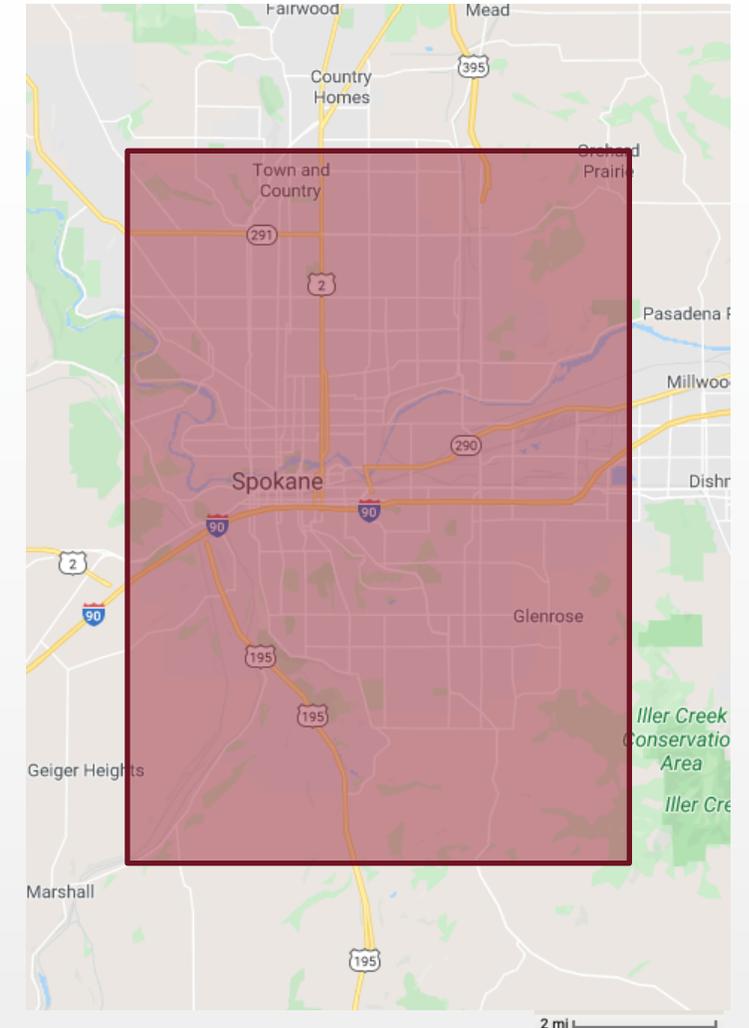 - Discern spatial differences between PM 2.5 concentrations within city

Figure 1: Approximate spatial coverage of Spokane PM 2.5 sensor network

Clarity Units

- Plantower PMS6003

  - Uses light scattering to estimate particle counts and concentrations

  - Reports PM10, PM2.5 and PM 1

  - Cellular data transmission

  - Variable sampling rates

- 10 solar powered nodes

- 2 reserve wired nodes

- Accuracy

  - +/- 10 ug/m^3 for measurements less than 100 ug/m^3 and 10% for concentrations greater than 100 ug/m^3

https://clarity.io/solution

http://www.aqmd.gov/aq-spec/product/clarity-node

Indoor Units

- Components
  - Raspberry Pi 3 B+
  - Plantower PMS5003
  - BME 280
    - Temp, pressure, relative humidity

- Low price-point ~ $100

- Simple assembly

- Easy install/maintenance

Figure 2: Assembled indoor PM 2.5 measurement units

Spokane PM 2.5 Monitoring Network

- 10 sampling locations at schools within the Spokane Public Schools District
    - ~20 km^2
    - Range of elevations and socioeconomic conditions
- Sensors placed inside and on exterior of school building at each location
- 2 wired nodes originally at our EBAM site (Eastern WA University Campus), moved to SRCAA Augusta Ave site
- All sensors installed by 8/29/19 (start of school year)

Summer Collocation Data

- Sensor collocation with EBAM on roof of PACCAR Environmental Technology Building at WSU

- Two periods of sensor overlap due to Clarity units arriving at separate times
  - Overlap period 1: 7/12/19 -7/31/19
    - EBAM, 5 solar Clarity units, 2 wired Clarity units, 10 indoor units
  - Overlap period 2: 8/19/19 – 8/22/19
    - EBAM, 5 solar Clarity units, 2 wired Clarity units
- Indoor unit software issue – not detected until Feb 8, 2020
  - Collocation data for indoor units unusable
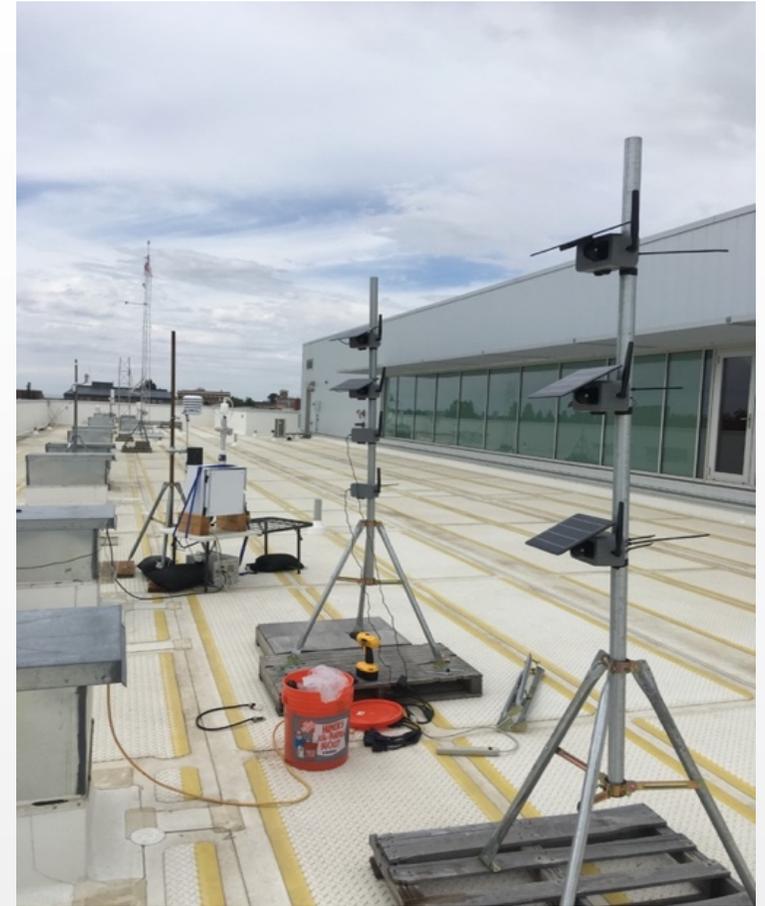 - Issue has been addressed, needs verification from another calibration

Figure 3: Clarity node collocation setup at Paccar building, WSU

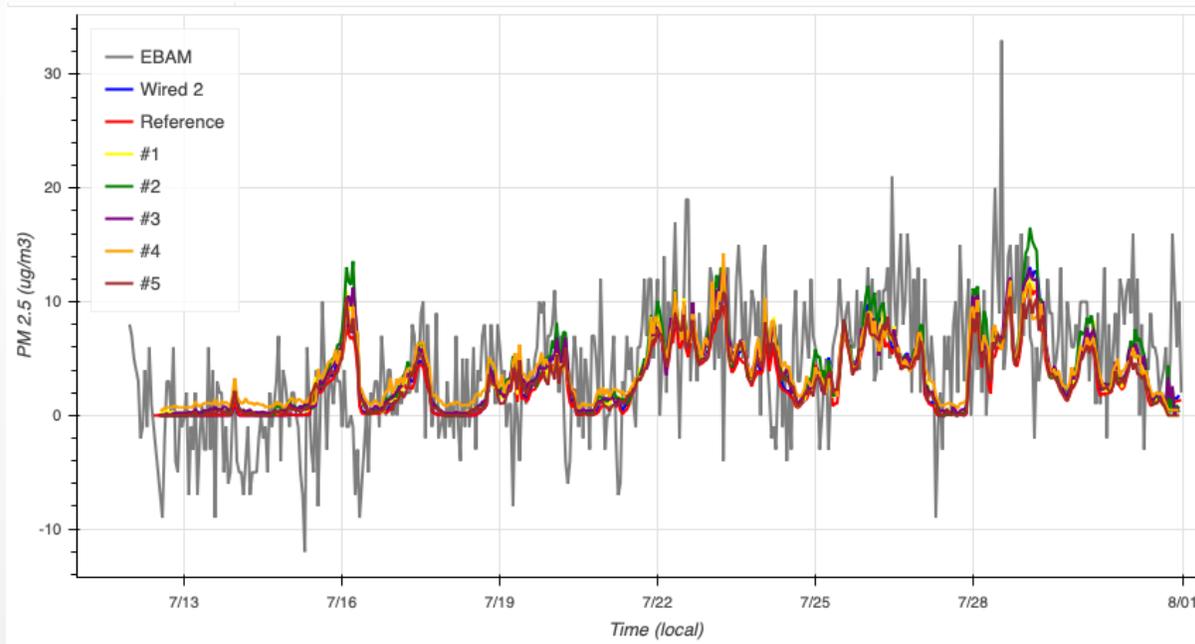# Summer Calibration – Clarity units and EBAM

Figure 4: Time series of 1st batch of Clarity units collocated with EBAM on PETB roof in Pullman, WA
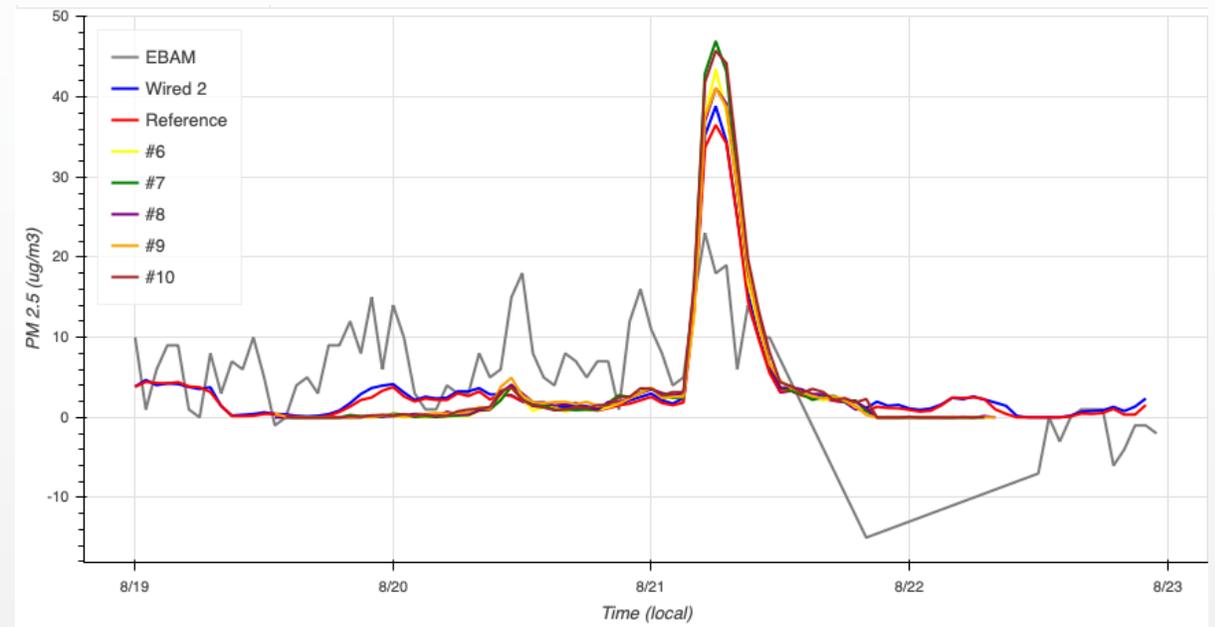

Figure 5: Time series of 2nd batch of Clarity units collocated with EBAM on PETB roof in Pullman, WA

Summer Calibration – Clarity Units

- Low Clarity inter-sensor variability
- Clarity units calibrated to "reference" Clarity node
- x = (y - b)/m
- Where,
  - y = raw Clarity node data
  - x = calibrated Clarity node data
  - m = slope of initial regression
  - b = intercept of initial regression

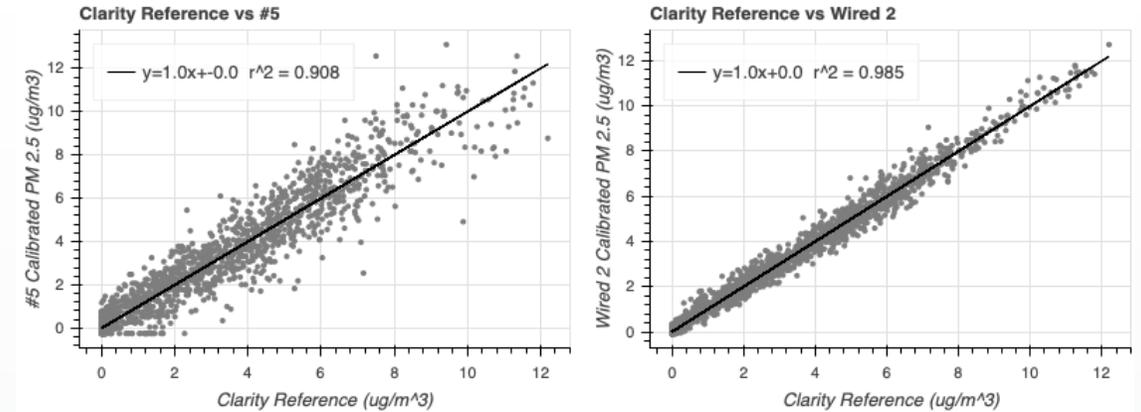| Sensor	Slope	Intercept	R^2
Reference	1	0	1
Wired 2	1.12	0.07	0.99
1	1.3	1.13	0.97
2	1.27	0.9	0.97
3	1.24	0.29	0.92
4	1.16	-0.94	0.96
5	1.16	0.70	0.96
6	1.15	0.67	0.95
7	1.11	0.47	0.93
8	1.09	0.36	0.93
9	0.98	1.06	0.84
10	0.98	0.25	0.91

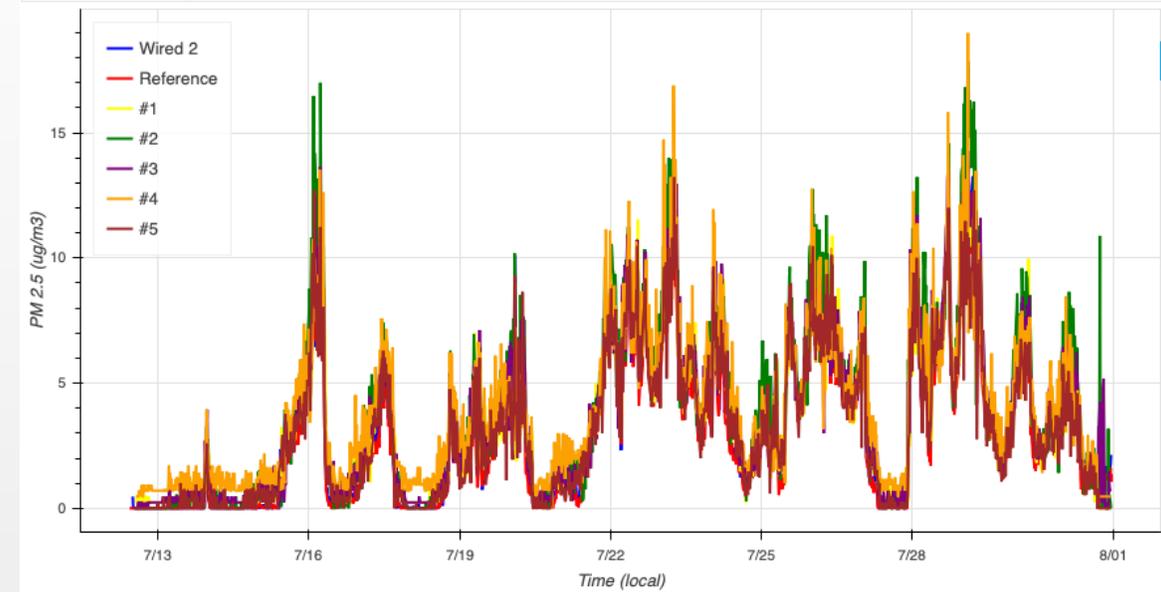Figure 6: Selected Clarity nodes calibrated to "reference" Clarity unit

Figure 7: Clarity "Batch 1" overlap

Collocation with SRCAA BAM at Augusta site

- WSU's EBAM malfunctioned in September 2019

- Needed alternative solution for data quality checks and calibration data for Clarity nodes after less than ideal summer conditions

- Two wired Clarity nodes collocated with SRCAA BAM at the Augusta site from 12/17/19 to 3/5/20
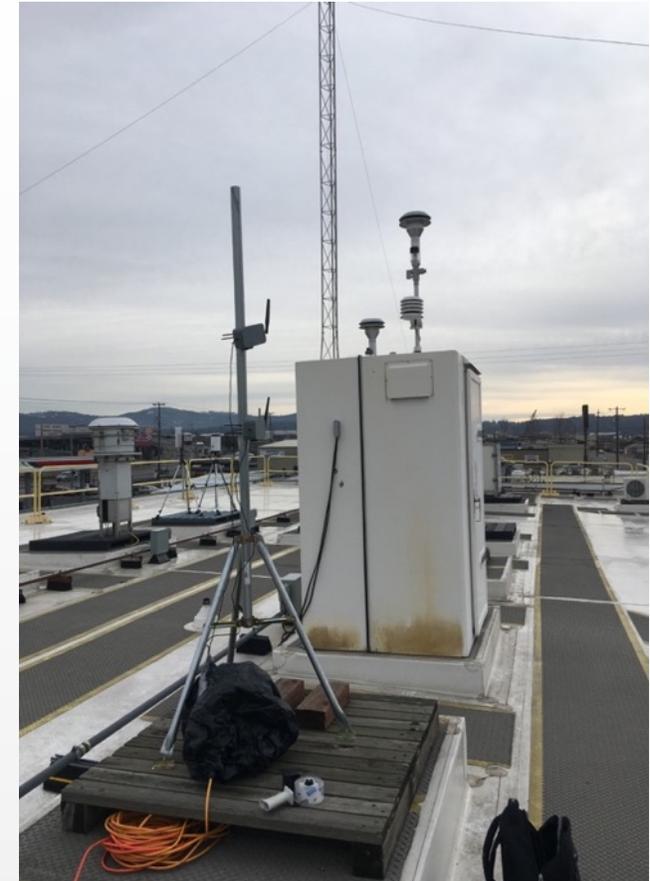
- 1,904 hourly measurements

Figure 8: Collocation of Clarity nodes and SRCAA BAM from 12/17/19 – 3/5/20

Calibration Approaches

- Linear regression

- Multiple linear regression (MLR)

- Minimizing Mean Absolute Error (MAE)

- Random Forest

Linear Regression

- Same approach as initial summer overlap calibration
- $x = (y - b)/m$
- Where,
  - y = raw Clarity node data
  - x = calibrated Clarity node data
  - m = slope of initial regression
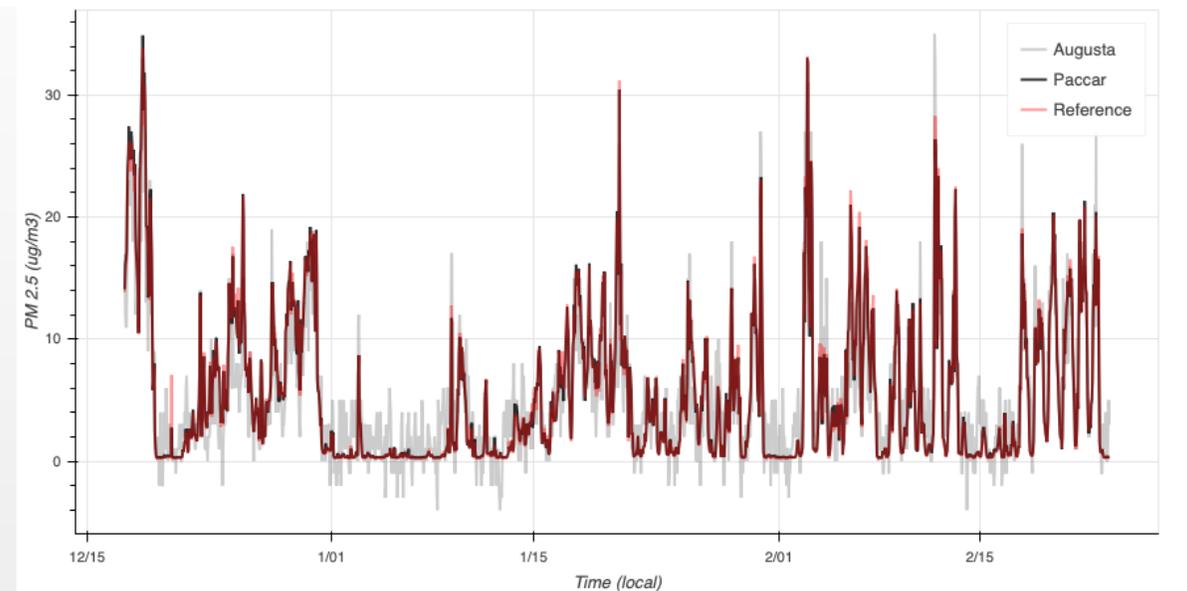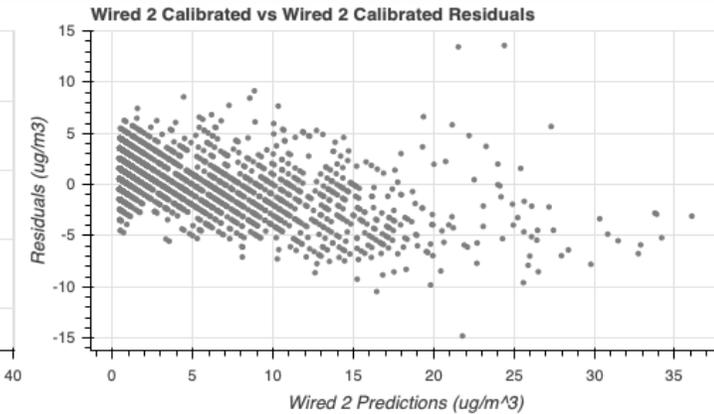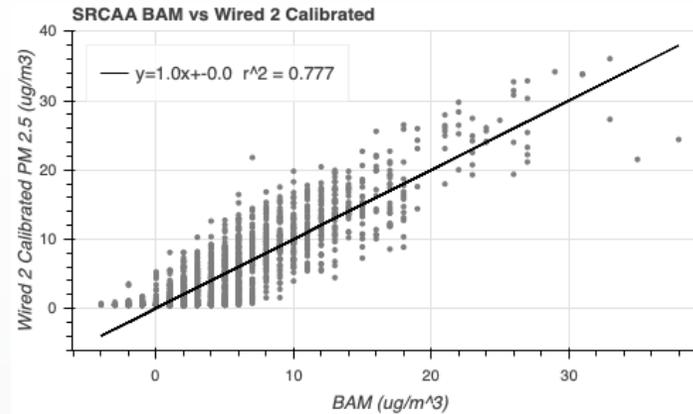 - b = intercept of initial regression

Sensor	Slope	Intercept	R^2	RMSE	%Res < 5 ug/m^3
Reference	1	0	0.774	2.87	92
Wired 2	1	0	0.777	2.84	92

Figure 9: Calibrated Clarity nodes using linear regression (complete overlap dataset)

Multiple Linear Regression

- Additional parameters
  - Temperature
  - Relative humidity

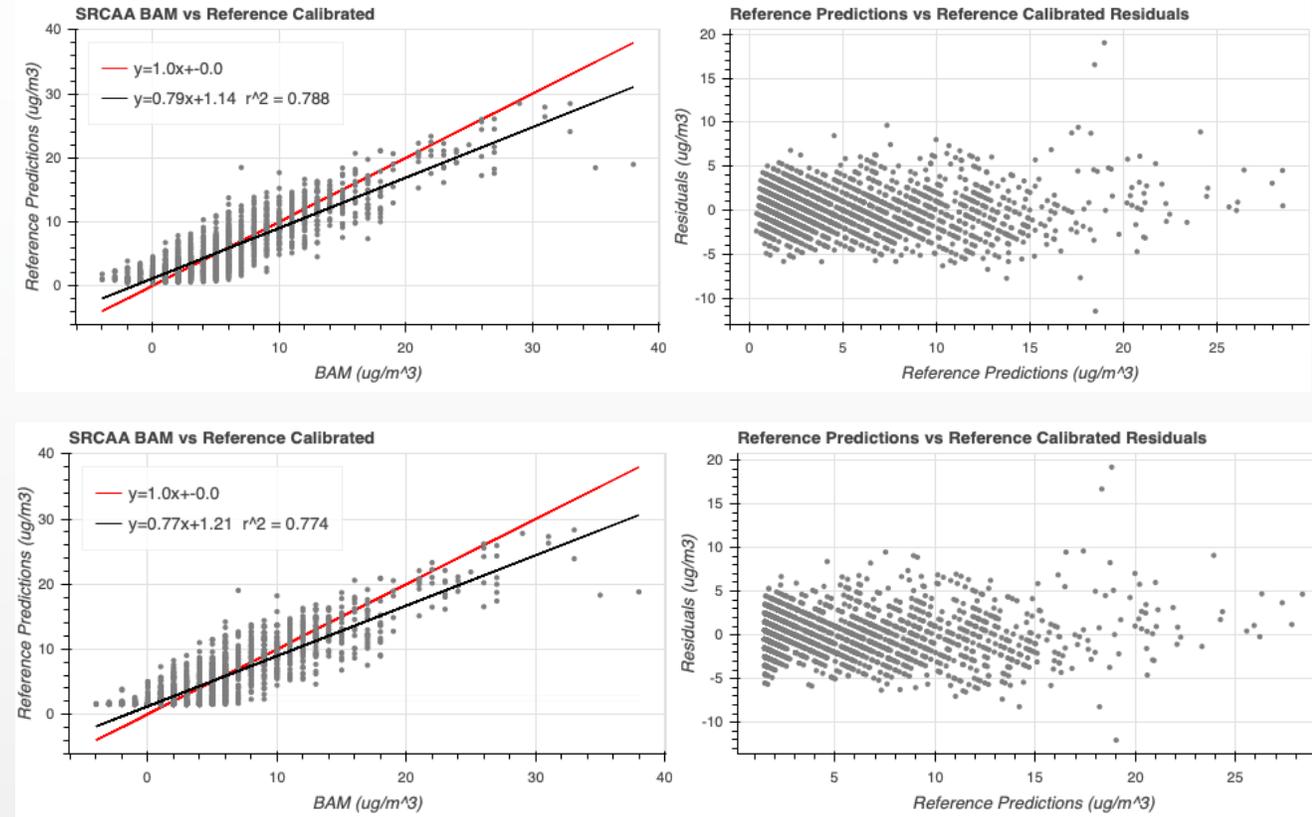| Sensor	Model	Slope	Intercept	R²	MAE
Reference	MLR	0.79	1.14	0.79	1.87
	Standard	0.77	1.21	0.77	1.92
Wired 2	MLR	0.79	1.12	0.79	1.86
	Standard	0.78	1.20	0.78	1.91

Figure 10: Comparison of MLR (top) and standard linear regression (bottom) of "reference" Clarity node

Minimizing Mean Absolute Error (MAE)

- Determined the slope and intercept that minimized the mean absolute error between calibrated data and SRCAA BAM measurement

$$MAE = \frac{\sum \left| BAM - \left(\frac{Clarity + b}{m} \right) \right|}{n}$$

- Where,
  - MAE = mean absolute error
  - BAM = Augusta BAM measurement
  - Clarity = hour resample of raw Clarity measurement
  - m = slope adjustment
  - b = intercept adjustment
  - n = number of measurements

- Applied these values to the previously used regression equation
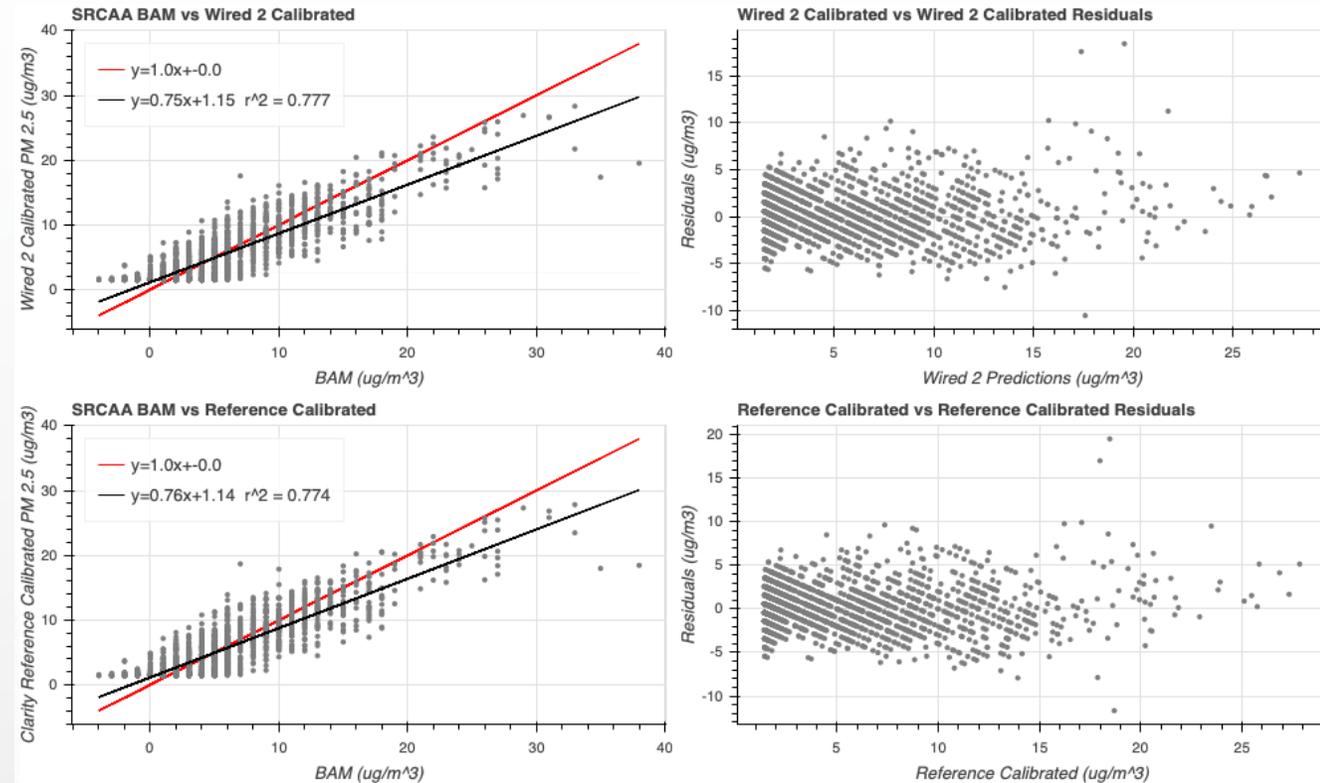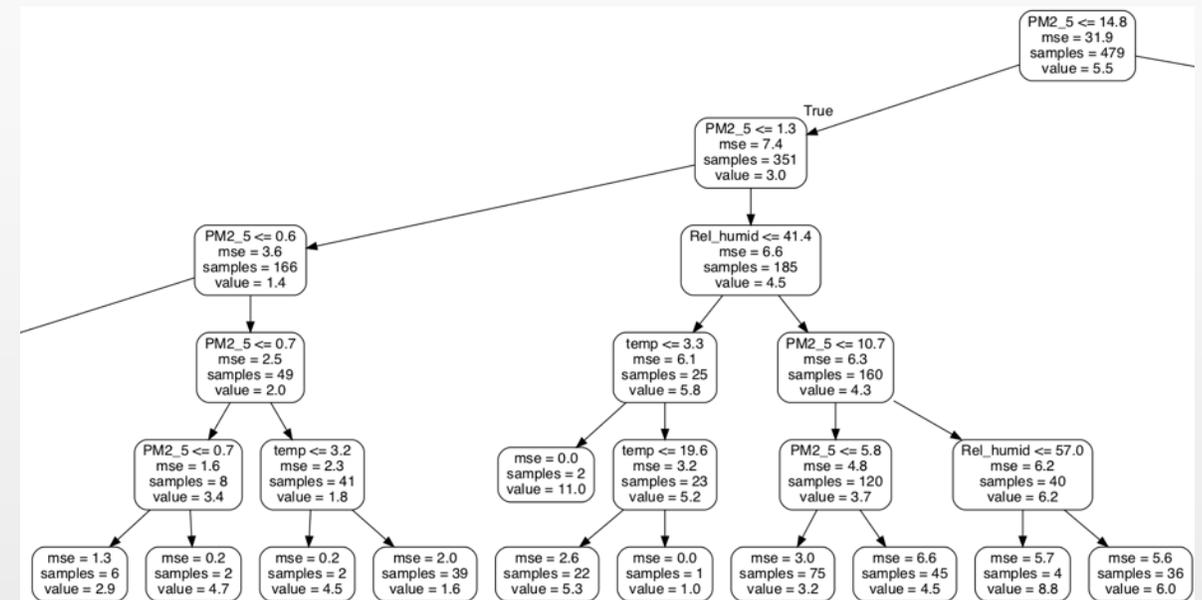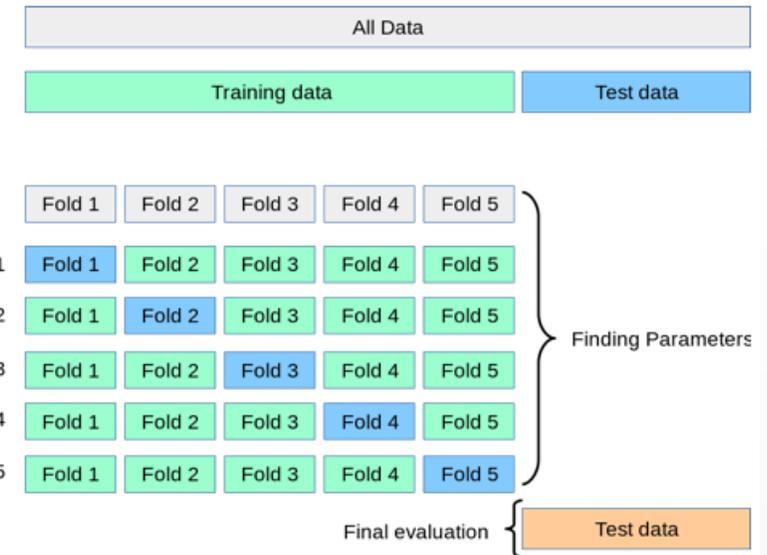
$$calibrated = (RC + b)/m$$

Figure 11: MLR calibration of "reference" and Wired 2 Clarity nodes

Random Forest (Using Sci-kit Learn)

- Lower chance of overfitting model to data
- 70/30% split on Training/Test datasets
- Features used
  - PM 2.5, temperature, RH
    - Augusta site only: Wind speed, wind direction
    - Potential parameters: inversion height, lapse rate
- Tuning Parameters
  - Number of trees in forest
  - Number of features used
  - Max tree depth
  - Minimum node split
  - Minimum leaf node samples
  - Max sample size
- K-fold cross validation used during tuning
  - 5 folds, 3 repeats
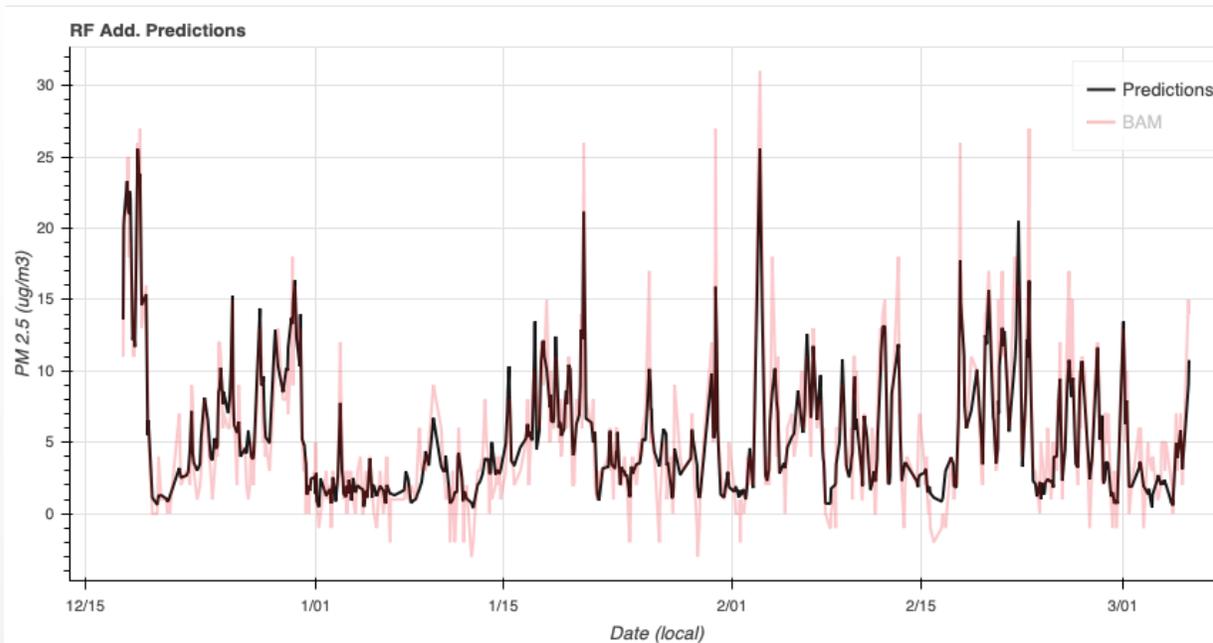
Random Forest Calibration

Figure 12: Time series of calibrated Random Forest model applied to test data for Clarity "reference" node and using additional features
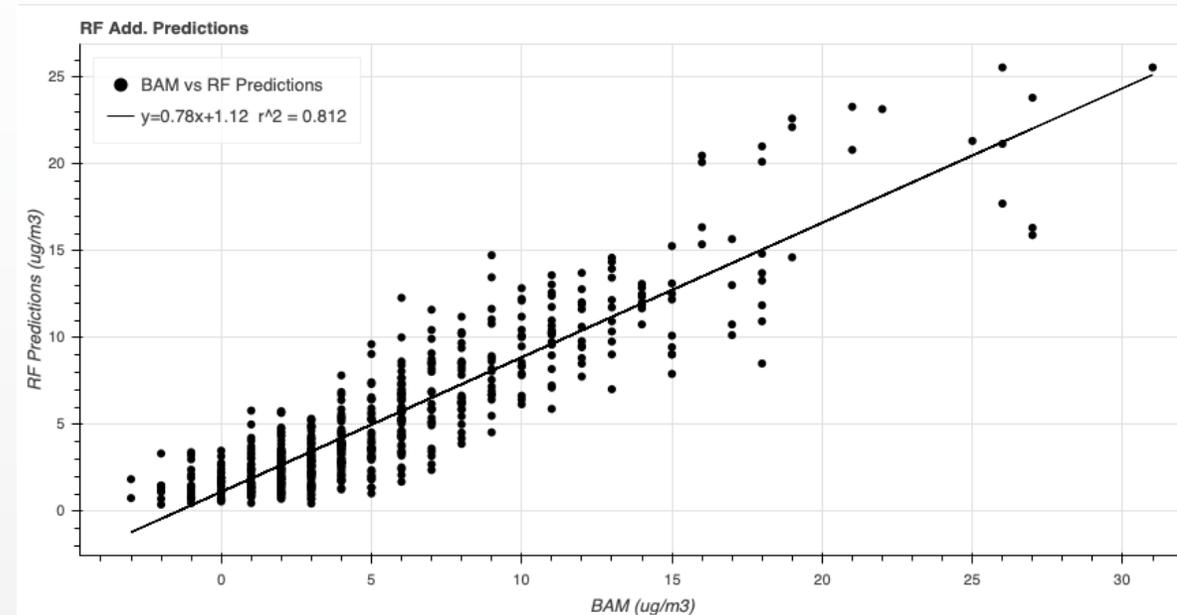
Figure 13: Random Forest model applied to test data set during calibration overlap period at Augusta site for Clarity "reference" node

Summary of Calibration Model Performances

Method	Slope	Intercept	R^2	MAE	RMSE	Res<5 ug/m^3 (%)
Linear	1	0	0.77	2.2	2.87	92
MLR	0.79	1.14	0.79	1.87	2.45	96
Min. MAE	0.75	1.15	0.78	1.92	2.53	95
RF "Ref"	0.83	1.01	0.79	1.84	2.45	97
RF add. "Ref"	0.77	1.16	0.81	1.75	2.31	97

Radiosonde Data

- Two atmospheric profiles/day
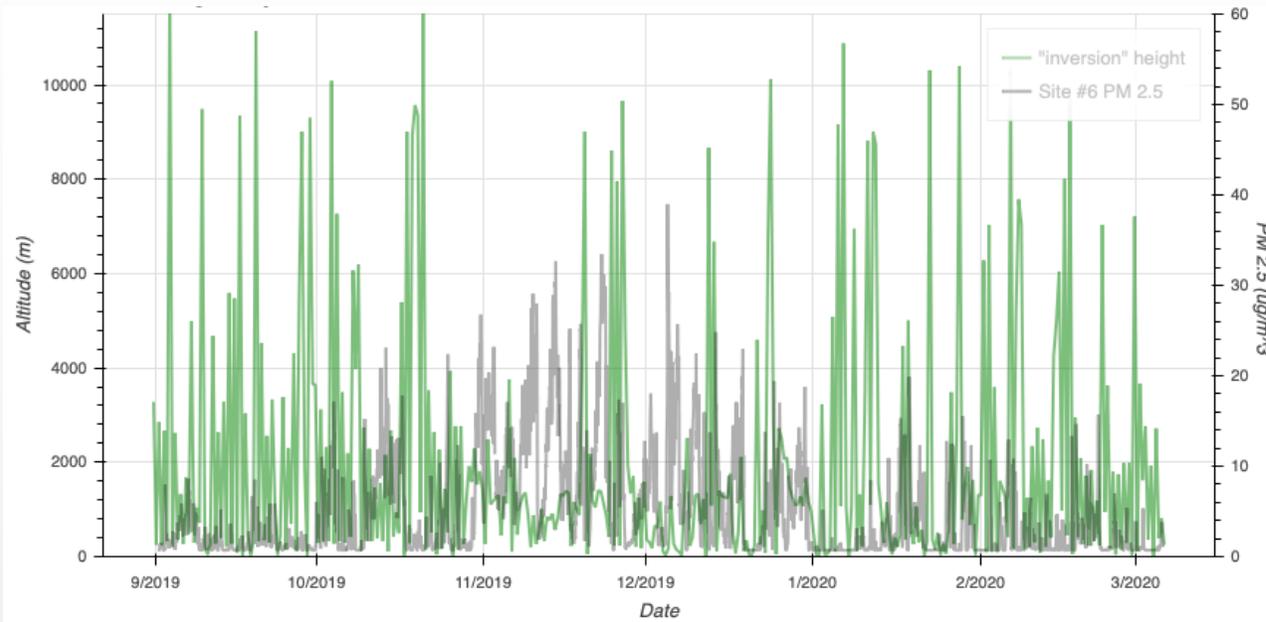- ~ 10 miles from study area

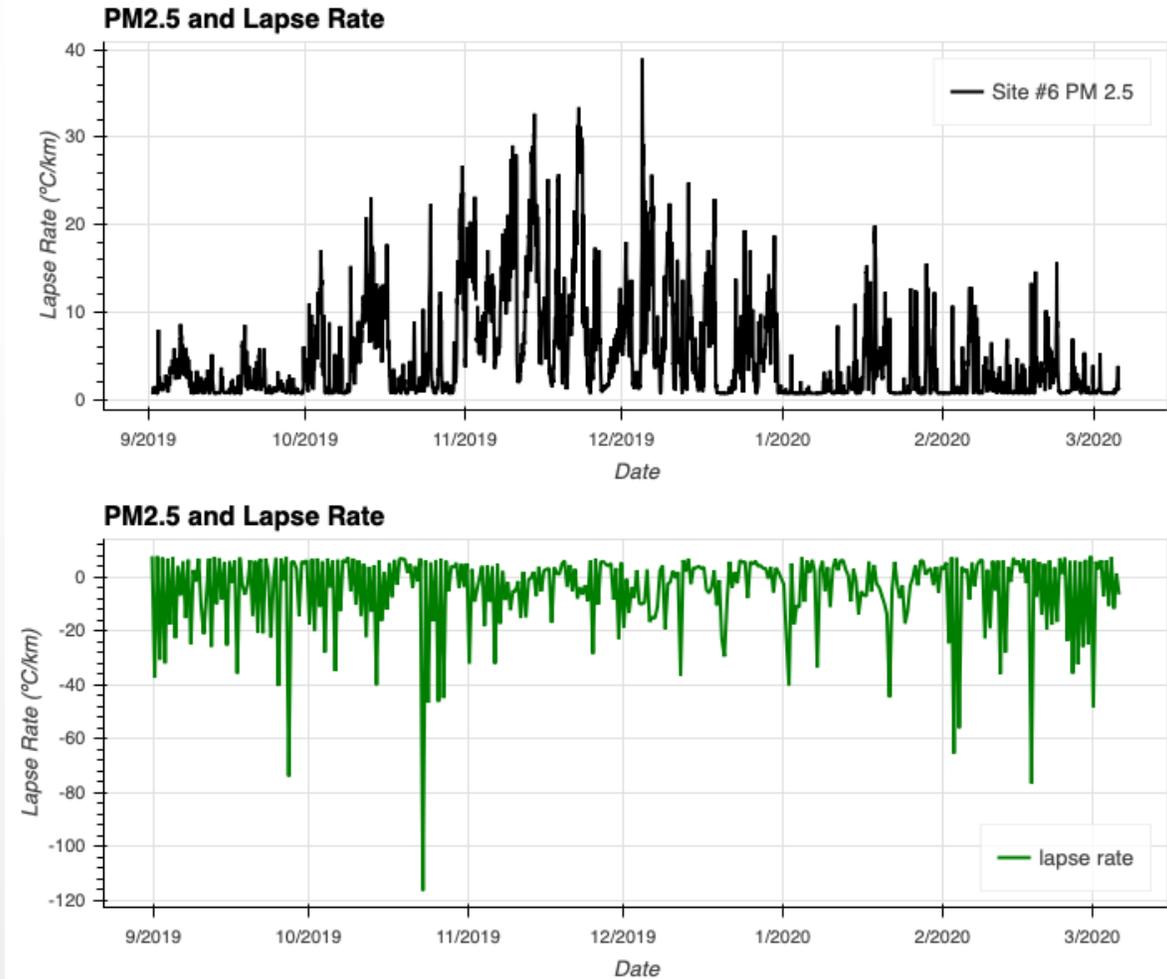Figure 14: Correlation between site #6 PM 2.5 concentrations and inversion layer height

Figure 15: Comparison of site #6 PM 2.5 concentrations and estimated atmospheric lapse rate

Preliminary Results – Basic Metrics

9/1/19 – 5/28/20

Location	Avg. (ug/m^3)	Median (ug/m^3)	Stdev (ug/m^3)
9	6.41	3.17	7.41
10	5.99	3.1	6.7
7	5.88	2.59	7.01
5	5.21	2.54	6.35
2	5.12	2.25	6.55
4	4.74	2.53	5.4
8	4.68	2.14	5.92
6	3.99	1.73	5.15
3	3.94	1.77	5.12
1	3.02	1.45	3.82

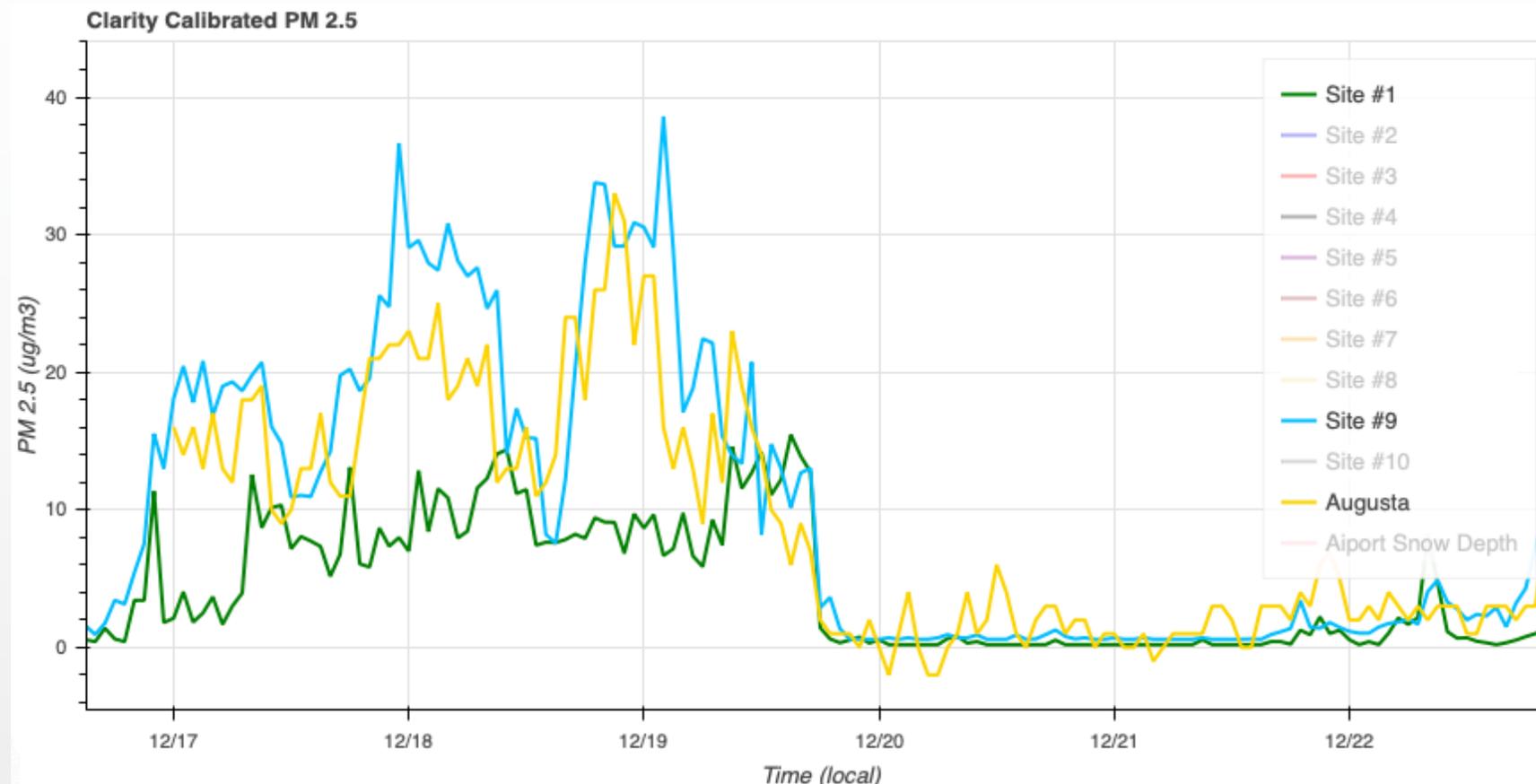# Preliminary Results – Site Variation

Figure 15: Sample PM 2.5 event for selected monitors

Preliminary Results – Indoor/Outdoor Comparison

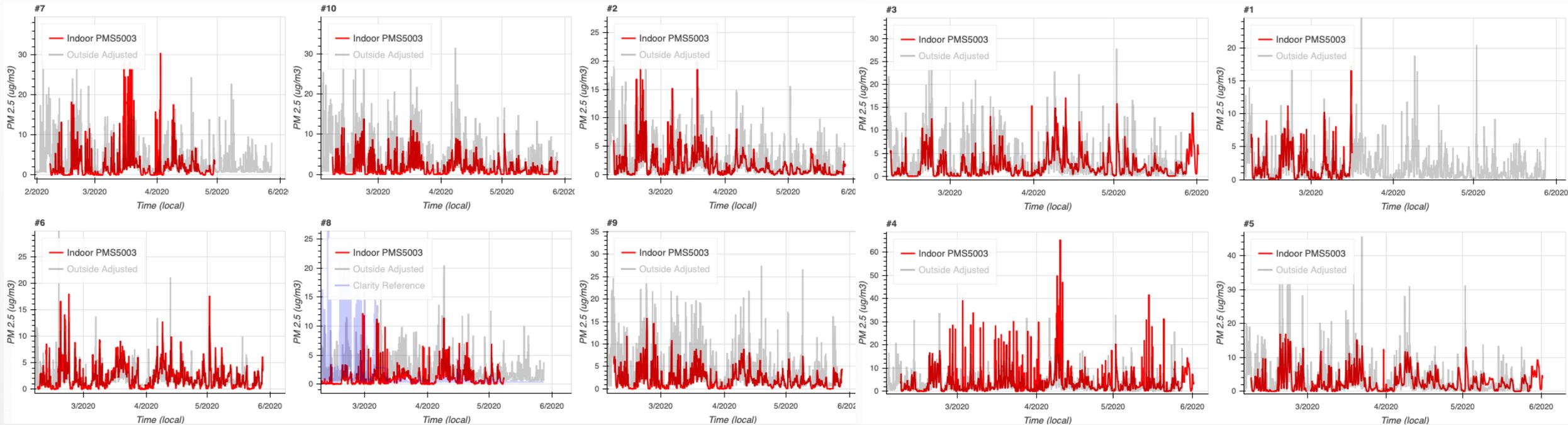- Uncalibrated indoor data

Figure 16: Comparisons of uncalibrated indoor and calibrated outdoor PM2.5 sensors

Initial Conclusions/Continuing Work

- Network has been simple to operate, maintain, and collect data from

- Random Forest appears to provide optimal calibration

- Data looks reasonable, extent of spatial differences still unclear due to uncertainty of each sensor
  - Effect of site proximity to I-90 on PM 2.5 concentrations
  - Clear diurnal pattern

- Continuing Work
  - Monitoring of wildfire season 2020 – as well as indoor/outdoor comparison to determine efficacy of building air handling capability
  - Post-experiment calibration

Thank you!

Questions?