Using a Remote Calibration Technique to Improve Data Quality for Large Networks of Particulate Matter Sensors

ASHLEY COLLIER-OXANDALE[1], BRANDON FEENSTRA[1], RANDY LAM[1], GEOFF HENSHAW[2], LENA WEISSERT[2], VASILEIOS PAPAPOSTOLOU[1], & ANDREA POLIDORI[1]

[1]*South Coast Air Quality Management District*
[2]*Aeroqual Ltd.*

AQ-SPEC
Air Quality Sensor Performance Evaluation Center

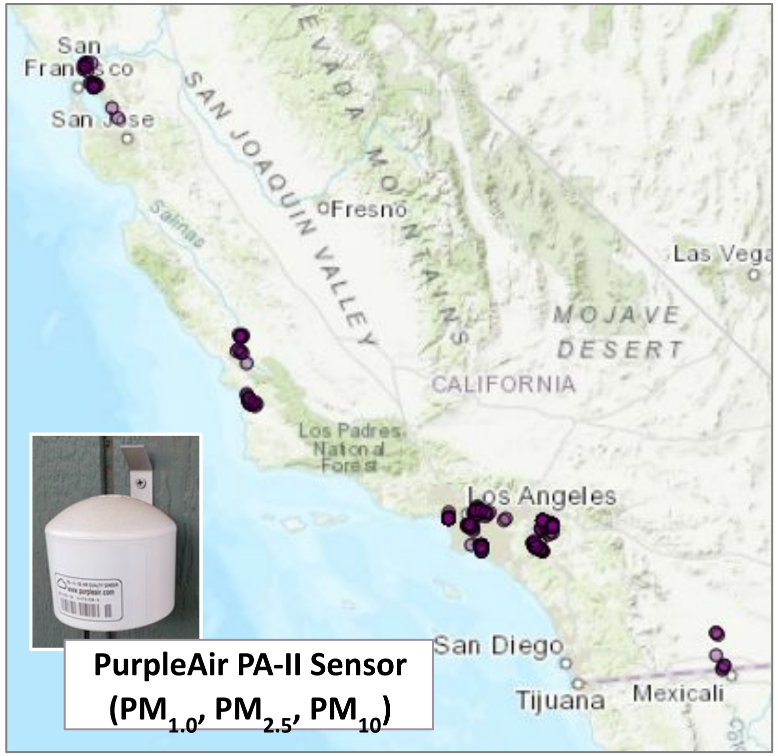
South Coast AQMD

Outline

- Background
  - PurpleAir PA-II Sensor Network & The STAR Grant Project
  - Sensor Calibration
  - MOMA Remote Calibration Technique
- Results
  - Data Access and Processing
  - Preliminary performance assessment for co-located sensors
  - Examples of spatial variability observed within the network
- Conclusions and Next Steps

PurpleAir PA-II Network & the STAR Grant

- PurpleAir PA-II sensors offer open-access data for $PM_{1.0}$, $PM_{2.5}$, and PM_{10}
- Widely used sensor, estimated that more than 10,000 have been deployed across the US
- In 2016, U.S. EPA funded Science to Achieve Results (STAR) Project was undertaken to “Engage, Educate, and Empower California Communities on the Use and Applications of Low-Cost Air Monitoring Sensors”
  - Large scale (included 14 different communities)
  - Multi-year deployment (some sensors operating > three years)
  - ~ 300 PA-II sensors were deployed
  - *dataset leveraged for this analysis*

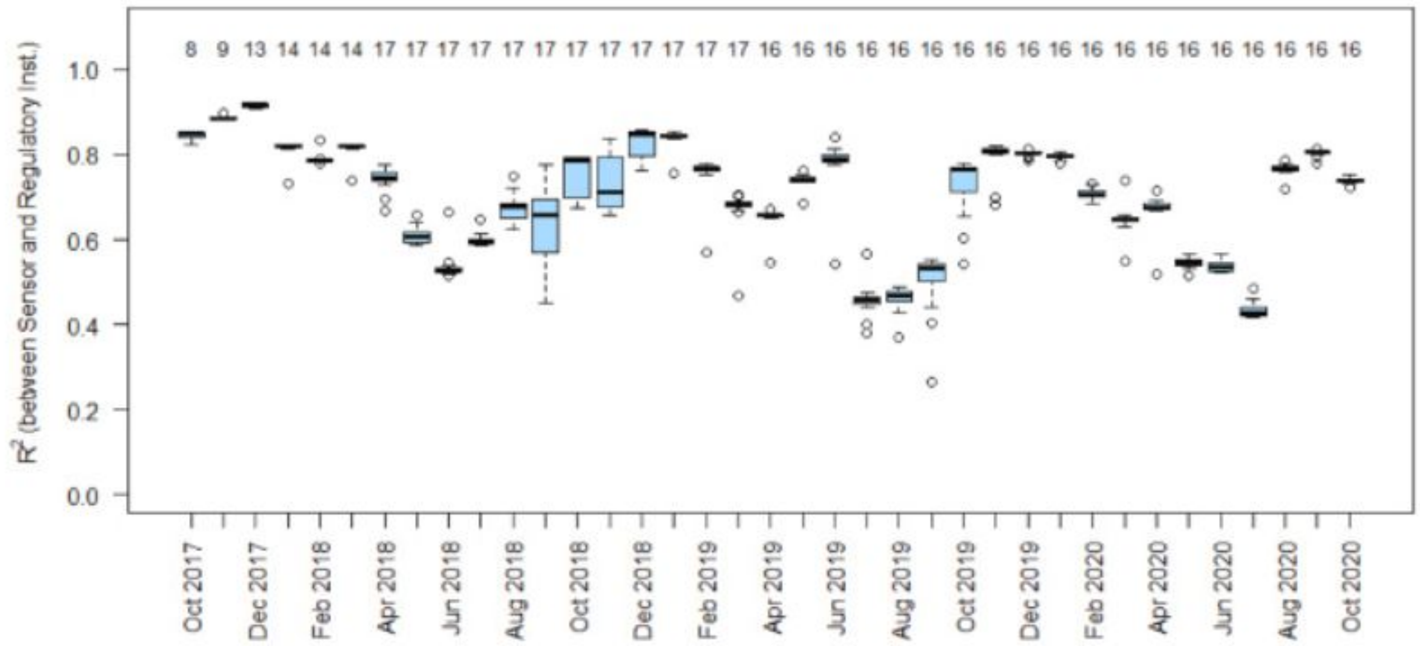


PurpleAir PA-II Sensor ($PM_{1.0}$, $PM_{2.5}$, PM_{10})

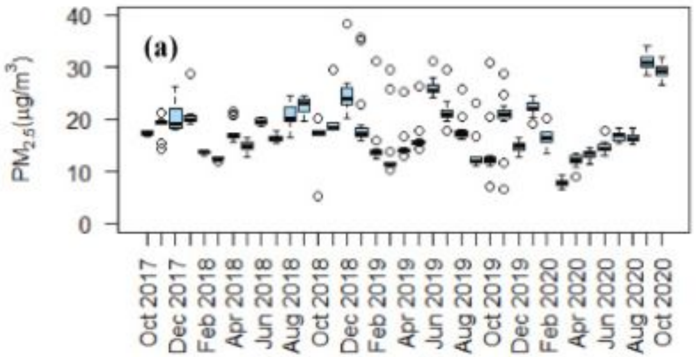
A Need for Calibration

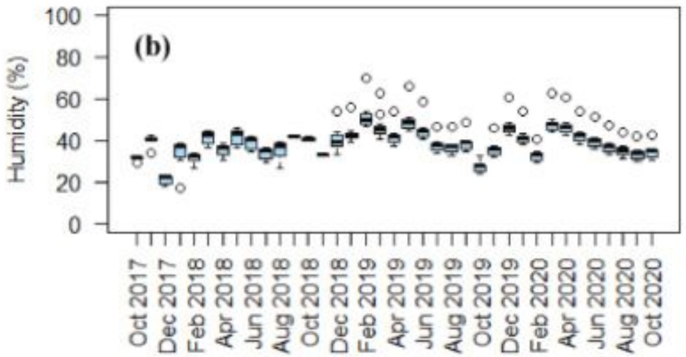
- Previous analysis revealed dramatic variability in sensor performance, seemingly driven by seasonal trends and PM-type, as opposed to a consistent decline or drift
- Plots depict monthly aggregate data from 17 sensors co-located at a regulatory monitoring site, collected over three years

(from Collier-Oxandale et al., 2021)



(left) $PM_{2.5}$ concentrations (right) humidity values observed by month





Typical Approaches to Calibration

Types	Overview	Pros	Cons
Factory Calibration	High throughput batch calibrations, resulting in correction factors (often linear)	All sensors in a batch calibrated under the same conditions	Occurs once by manufacturer
Laboratory Calibration	Calibration by end users in chamber systems designed to mimic real world conditions	Relatively quick, replicable, sensors can be calibrated in batches	May not fully capture the field conditions (e.g., dynamic changes in environment conditions or the background pollutant mixtures)
Field Calibration	Sensors are co-located with high quality reference instrumentation for a defined period, calibration models typically developed through linear regression, multiple linear regression, or machine learning techniques using the co-located dataset	Able to account for typical field conditions (environmental conditions and background pollutant mixtures)	Time and labor intensive, will likely need to be repeated at regular intervals or before and after a field deployment

- *None of these approaches are well-suited to be applied to large-scale, long-term sensor networks*
- *Newer approaches include global correction equations, remote calibration, and calibration using mobile platforms*

The “MOMA” Approach

***MOMA Assumption:** given a suitable proxy site, the distribution of values is the same and a remote calibration of an AQY to that proxy site can correct for baseline drift (offset) and sensitivity changes (gain) and for PM sensors this approach can correct for seasonal composition/weather impacts as well*

MOMA, PA-II Pilot (procedure)

- Two approaches were used – applying MOMA on a monthly basis and driven by a drift detection algorithm that tracks the difference between the sensor data and the proxy site
- Calibration periods were identified where the data was expected to be most similar between the sensor and the proxy site, based on meteorological conditions and the similarity of pollutant trends between the sensor and proxy data
- The MOMA algorithm was then used to determine updated gains and offsets, which were applied to the PA-II sensor data

Key publications detailing the development and evaluation of the MOMA Approach:

- Miskell, G., et al., (2018). Solution to the problem of calibration of low-cost air quality measurement sensors in networks. ACS sensors, 3(4), 832-843.
- Miskell, G., et al., (2019). Reliable data from low cost ozone sensors in a hierarchical network. Atmospheric Environment, 214, 116870.
- Weissert, L., et al., (2020). Low-cost sensor networks and land-use regression: Interpolating nitrogen dioxide concentration at high temporal and spatial resolution in Southern California. Atmospheric Environment, 223, 117287.
- Weissert, L., et al., (2020). Hierarchical network design for nitrogen dioxide measurement in urban environments. Atmospheric Environment, 228, 117428.

Data Access & Processing

- Data accessed: CF_1 data (due to the higher linearity)
- Quality Assurance and Quality Control
  1. function applied to filter "out-of-spec" values (or values outside of manufacturer specified bounds)
  2. "PurpleAirQC_hourly_AB_03" algorithm applied, involves the following steps:
     - Invalidate data where: A/B hourly difference > 5 AND A/B hourly percent difference > 70%
     - Invalidate data where: A/B hourly data recovery < 90%
     - Hourly-average valid data

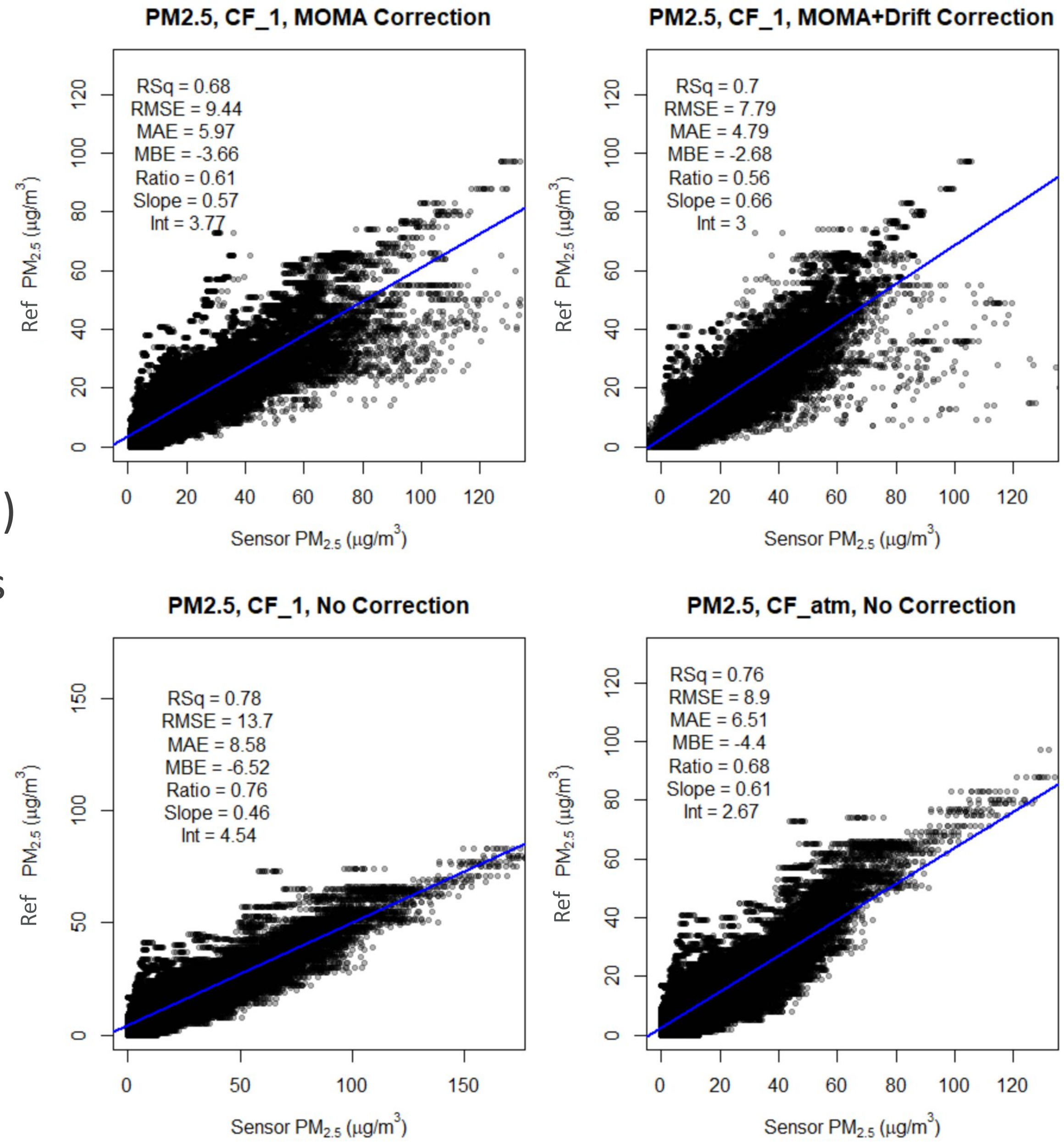
Facilitated via the **AirSensor package** (open-source R package), intended to support data access, processing, analysis and visualization for PurpleAir sensor data

Developed under the STAR Grant Project in collaboration with Jonathan Callahan (Desert Research Institute)

- Future AirSensor enhancements: add functions to enable the application of corrections to the sensor data, which could support the scalability various correction approaches

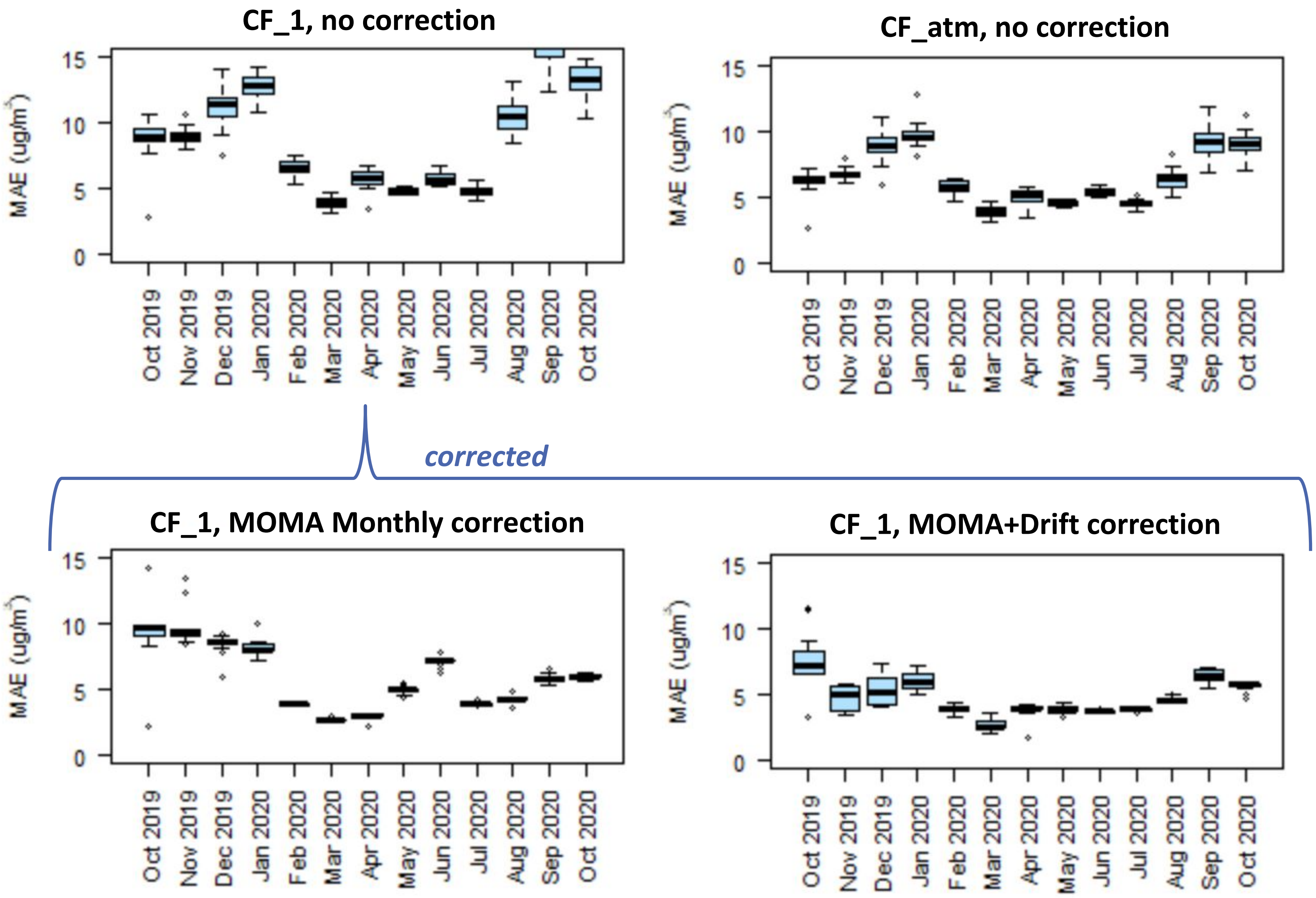
Preliminary Results

- Applying MOMA on a monthly basis and using the drift detection algorithm results reduces the overall error (e.g., MAE and MBE)
- MOMA+drift results in greater improvements that MOMA applied on a monthly basis
- MOMA was applied using data from a proxy site ~ 5 miles away
- Plots depict one year of data, hourly-averaged, aggregate of 15 sensors co-located at a regulatory monitoring site (labeled Ref)



Preliminary Results

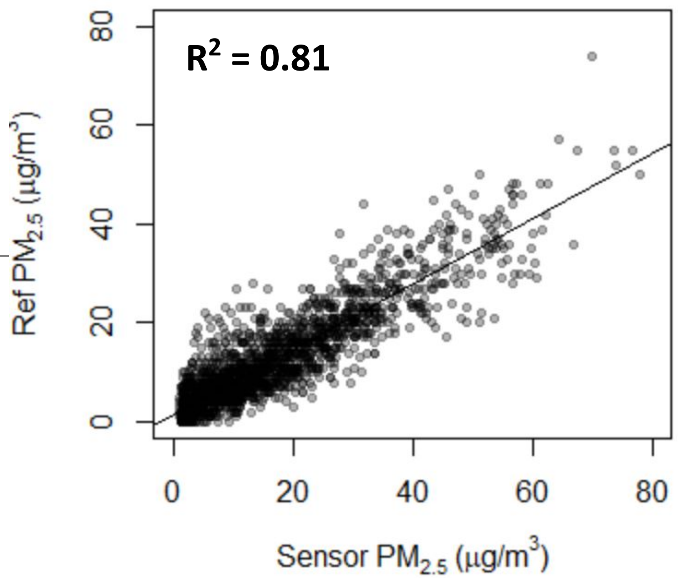
- Comparing uncorrected CF_1 sensor data to CF_1 data corrected using the MOMA+drift approach, not only is the overall error reduced, but the correction seems to mitigate some of the seasonal effects
- Plots depict monthly aggregate MAE from 15 sensors co-located at the reference site



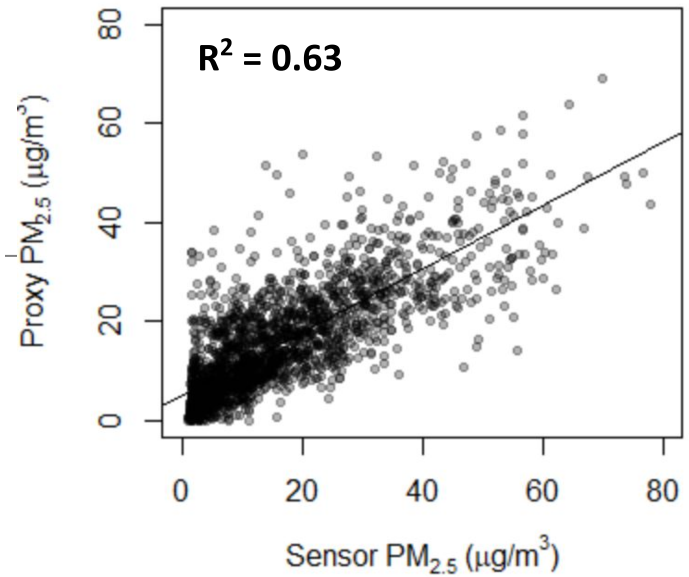
Preserving Local Variability

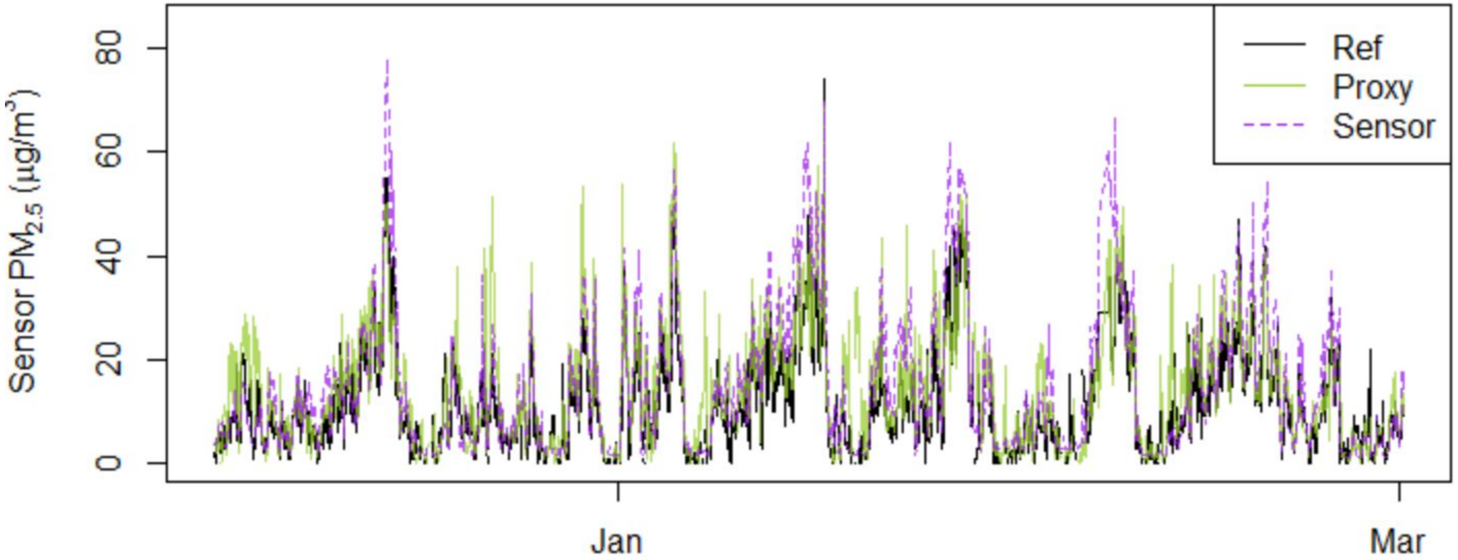
- A key concern with remote calibration techniques is the preservation of local variability and indications of short-term air quality events
- Initial results indicate that this approach to sensor correction reduces error, but does not impact local trends
- Note the higher R^2 between the sensor and the co-located Ref site as opposed to between the sensor and the proxy site

MOMA+drift Sensor vs. Reference (co-located)



MOMA+drift Sensor vs. Proxy Site





Conclusions and Next Steps

- Preliminary results indicate MOMA has the potential to improve PA-II sensor data and potentially mitigate variability in performance driven by seasonal factors
- The MOMA approach is scalable and feasible for implementation with large-scale stationary networks
- Publication in progress, assessing MOMA capabilities and potential over multiple years
  - Expand analysis to other validation sites
  - Optimize MOMA (esp. for events such as wildfires)
  - Compare to other established corrections (e.g., global corrections for PurpleAir PA-II sensors)
- Conduct pilots using other sensors, to explore the potential of this approach as a sensor-agnostic method for correcting and improving sensor data