Optimizing Field Co-location of Low-cost Sensors

Misti Levy Zamora

07/27/2023

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

UCONN
HEALTH

Introduction

- Instrument calibration is one of the main processes used to ensure instrument accuracy

- In the case of low-cost air pollution sensors, the raw output is often a voltage or resistance instead of a concentration

- Many sensor calibration protocols involved co-locating the low-cost sensors with reference instruments

- There is currently no standardized co-location duration
  - Reported co-location durations for low-cost sensors with reference instruments in recent work have varied from several days to several months

- Little discussion has focused on whether this period is ideal for the deployment period or whether the calibration period can be optimized.

- The goal of my recent work was to identify efficient field calibration practices by:
  - 1) Identifying the key factors that influence the sensor values (briefly shown)
  - 2) Determining if a there is an optimal co-location length

Assessing Ambient Levels and Personal Exposures in Baltimore: The SEARCH Project

- Air Climate & Energy (ACE) Center Grant funded by U.S. EPA
- **Kirsten Koehler (JHU) & Drew Gentner (Yale)**
- Monitor development: Lizi Xiong (Yale), Branko Kerkez (U. Mich.), Jordan Peccia (Yale), **Colby Buehler (Yale)**
- Monitor siting: Jesse Berman (UMN), Ben Zaitchik (JHU), Eddie Meade (JHU), Dorothy Clemons-Erby (JHU)
- Statistics: **Abhirup Datta (JHU)**
- Data Management: **Hao Lei (JHU),** Megan Wood (JHU), Yitong Chen (JHU)

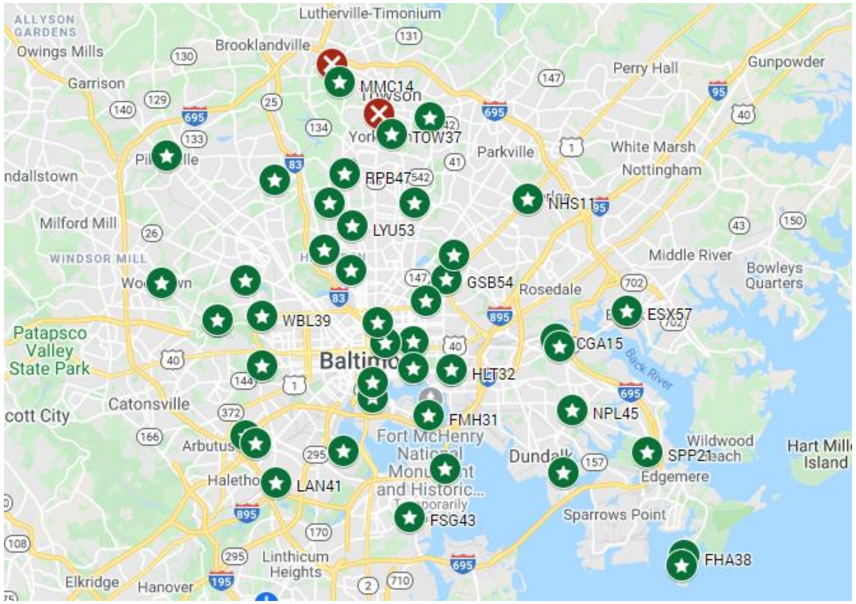
SEARCH

Solutions for Energy Air, Climate & Health

UCONN HEALTH

The SEARCH Low-cost Stationary Multipollutant Monitoring Network

- Co-located at four reference sites
  - Maryland Department of the Environment
    - CO (1)
    - NO (2)
    - NO_2 (2)
    - $PM_{2.5}$ (1)
    - O_3 (1)
  - National Institute of Standards and Technology (NIST)
    - CH_4 (2)
    - CO_2 (2)



- Used co-location data from February 1, 2019, to February 1, 2020

Step 1: Identify significant factors for each sensor

- Sensor data from the calibration period was used to determine the coefficients for multiple linear regression (MLR) models for each sensor

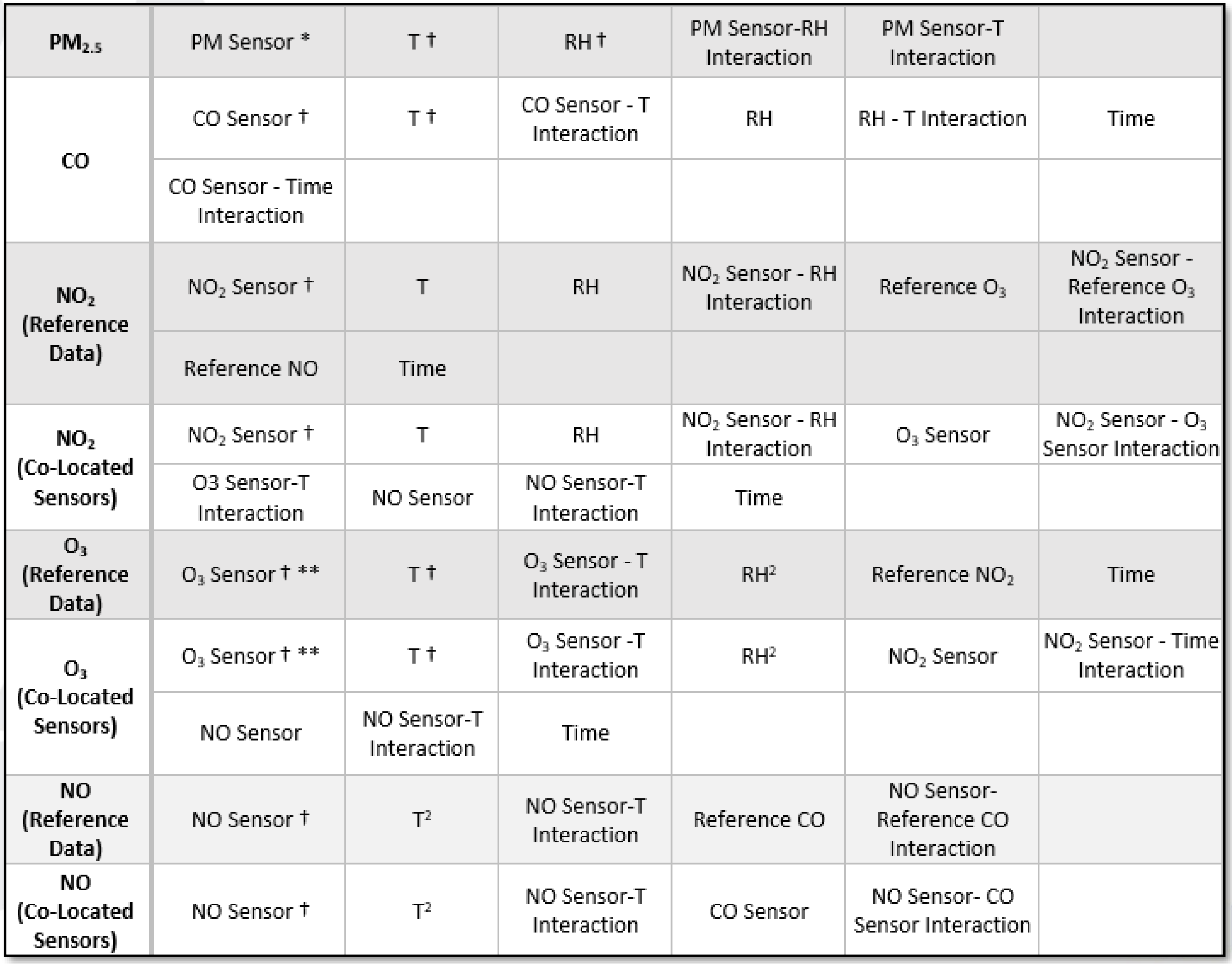
- A generic MLR model is given by:

$$Reference_{Pollutant}(t) = \beta_o + \beta_1 * Sensor_{Pollutant}(t) + \sum_1^n \; \beta_n * Predictor_n(t)$$

UCONN HEALTH

Levy Zamora, Misti, et al. "Evaluating the performance of using low-cost sensors to calibrate for cross-sensitivities in a multipollutant network." ACS ES&T Engineering 2.5 (2022): 780-793.

Step 1: Identify significant factors for each sensor



$PM_{2.5}$	PM Sensor *	T †	RH †	PM Sensor-RH Interaction	PM Sensor-T Interaction	
CO	CO Sensor †	T †	CO Sensor - T Interaction	RH	RH - T Interaction	Time
	CO Sensor - Time Interaction					
NO_2 (Reference Data)	NO_2 Sensor †	T	RH	NO_2 Sensor - RH Interaction	Reference O_3	NO_2 Sensor - Reference O_3 Interaction
	Reference NO	Time				
NO_2 (Co-Located Sensors)	NO_2 Sensor †	T	RH	NO_2 Sensor - RH Interaction	O_3 Sensor	NO_2 Sensor - O_3 Sensor Interaction
	O3 Sensor-T Interaction	NO Sensor	NO Sensor-T Interaction	Time		
O_3 (Reference Data)	O_3 Sensor † **	T †	O_3 Sensor - T Interaction	RH^2	Reference NO_2	Time
O_3 (Co-Located Sensors)	O_3 Sensor † **	T †	O_3 Sensor -T Interaction	RH^2	NO_2 Sensor	NO_2 Sensor - Time Interaction
	NO Sensor	NO Sensor-T Interaction	Time			
NO (Reference Data)	NO Sensor †	T^2	NO Sensor-T Interaction	Reference CO	NO Sensor- Reference CO Interaction	
NO (Co-Located Sensors)	NO Sensor †	T^2	NO Sensor-T Interaction	CO Sensor	NO Sensor- CO Sensor Interaction	

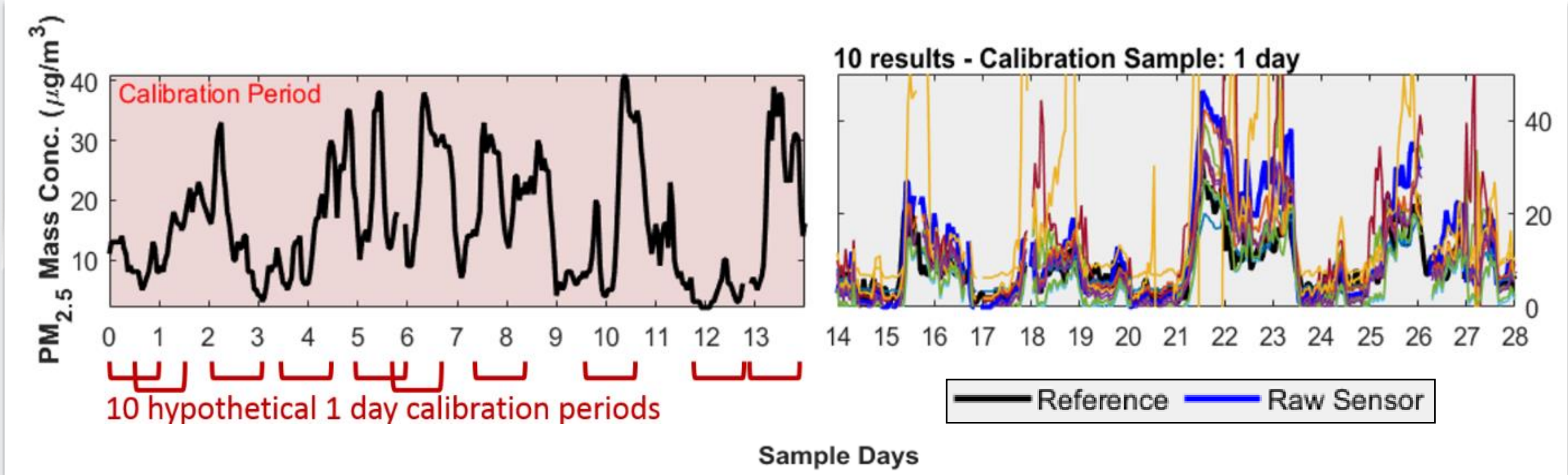
Levy Zamora, Misti, et al. "Evaluating the performance of using low-cost sensors to calibrate for cross-sensitivities in a multipollutant network." ACS ES&T Engineering 2.5 (2022): 780-793.

Hypothetical Co-location Periods

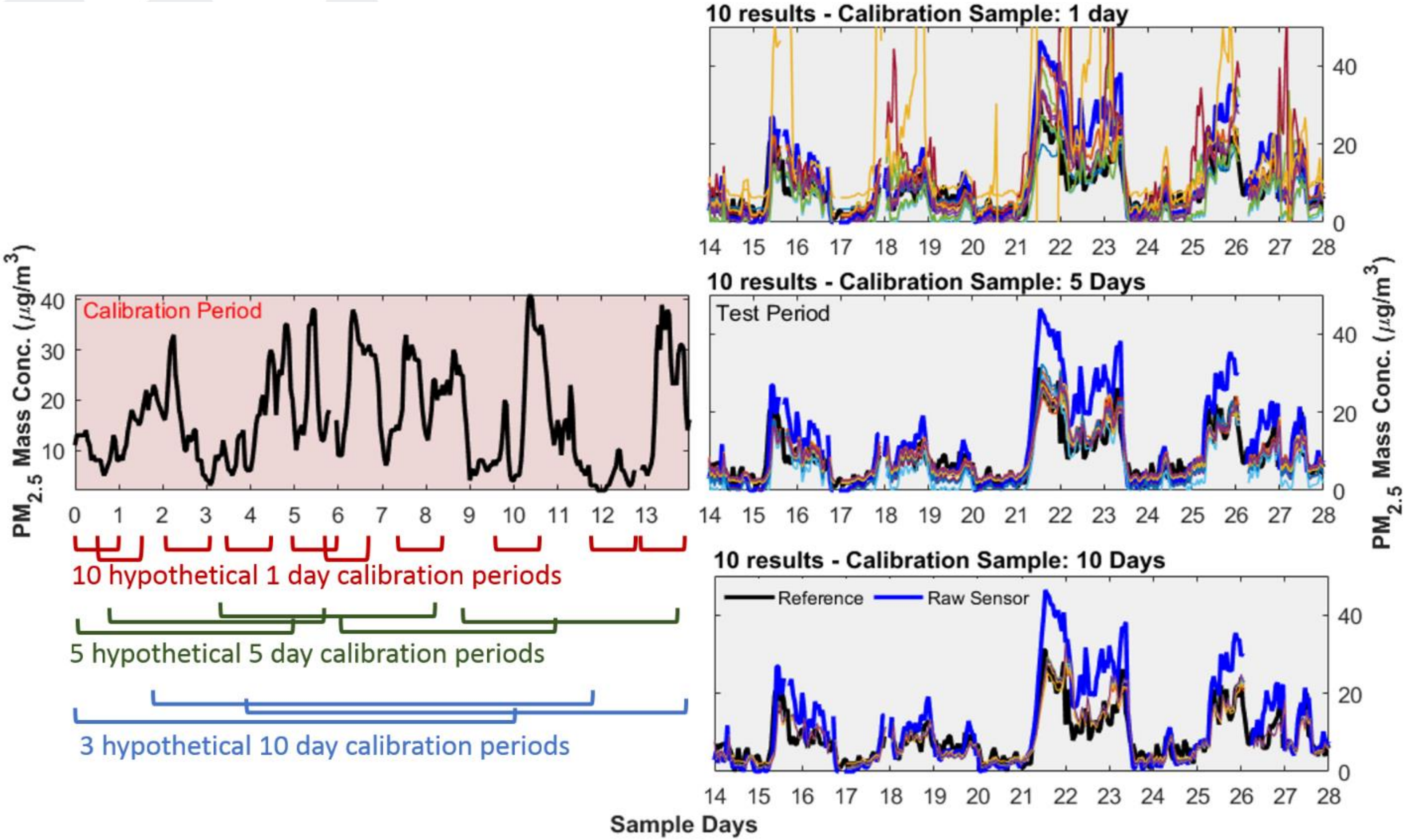
- First, we developed calibration equations using randomly selected co-location subsets spanning 1 to 180 consecutive days out of the 1-year period
  - 250 sample calibration periods were randomly selected for each duration

- Hourly concentrations for the evaluation period were produced using the equations created using that randomly selected calibration period

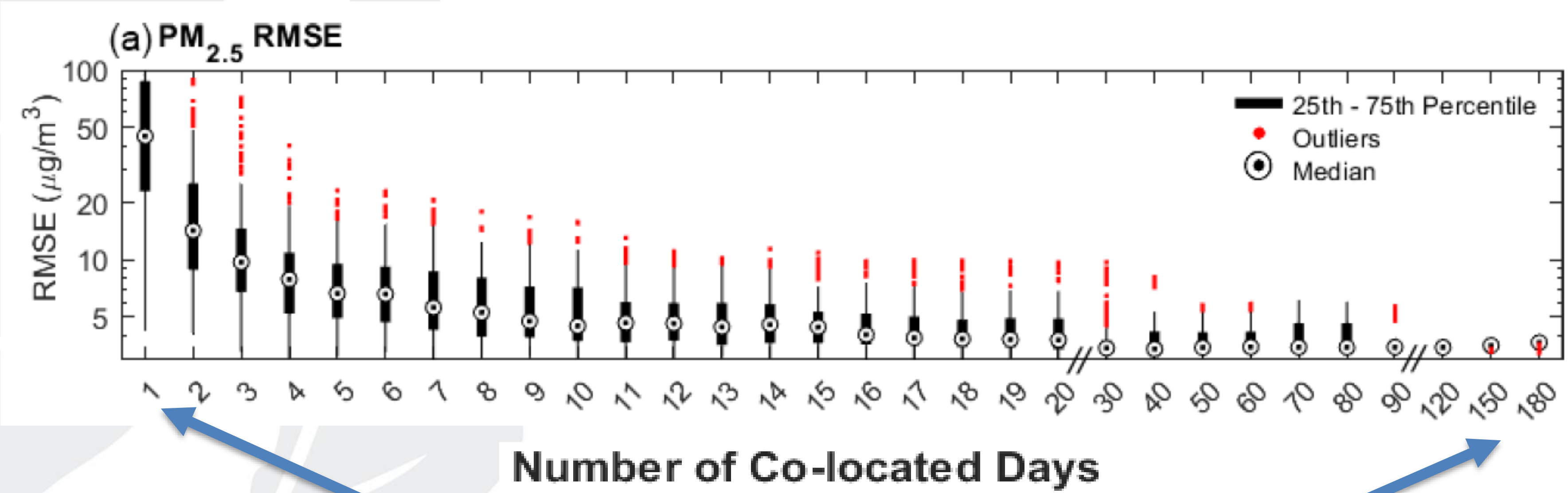
- We compared the potential root-mean-square error (RMSE) and Pearson correlation coefficient (r) values for each of the 250 calibrations



Co-location subsets spanning 1 to 180 consecutive days



Range of potential RMSE for $PM_{2.5}$

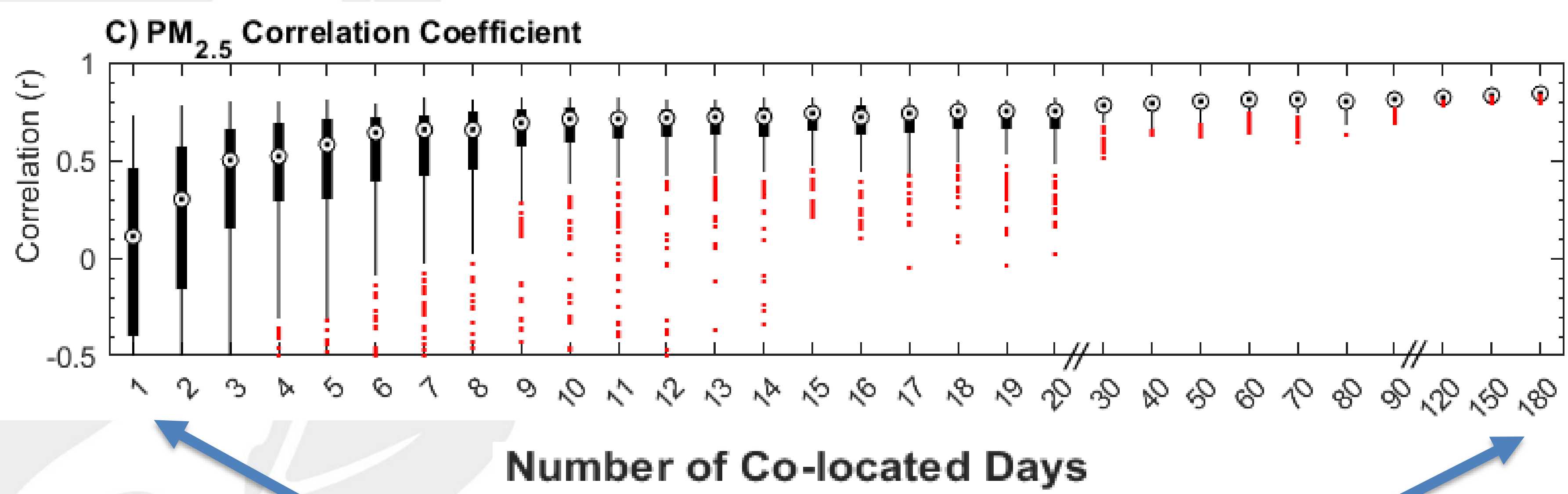


	1 Day	6 Months
$PM_{2.5}$ ($\mu g/m^3$)	44.9 (5.2 – 400)	3.6 (3.2 – 3.7)

UCONN HEALTH

Levy Zamora, Misti, et al. "Identifying optimal co-location calibration periods for low-cost sensors." *Atmospheric measurement techniques* 16.1 (2023): 169-179.

Range of potential correlations for $PM_{2.5}$

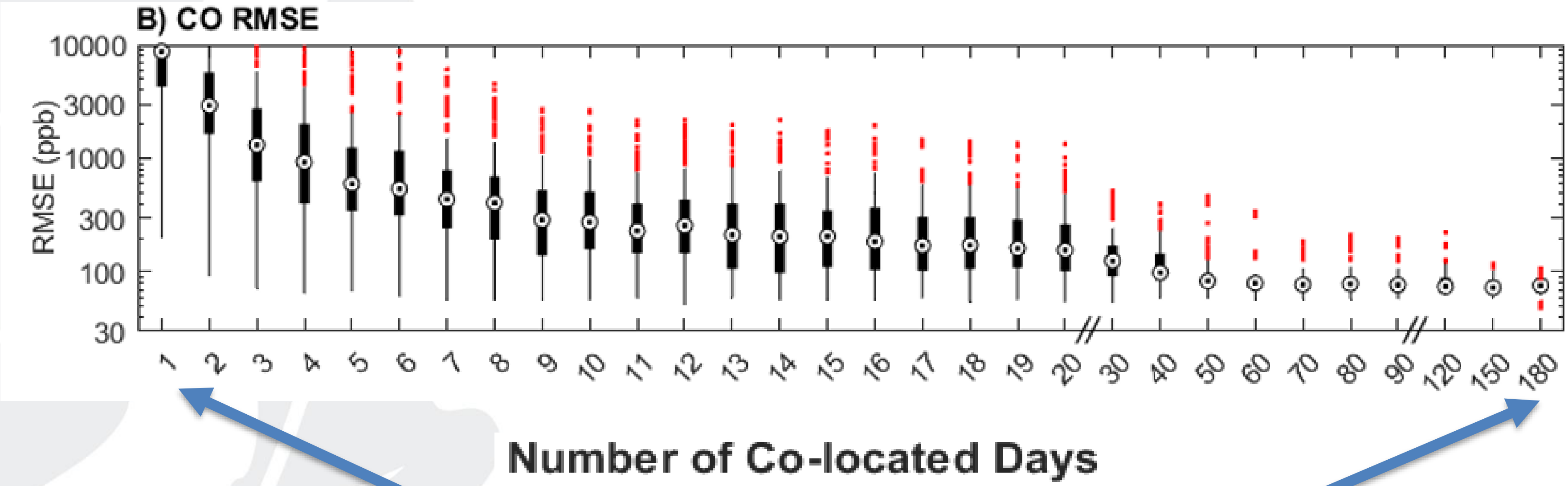


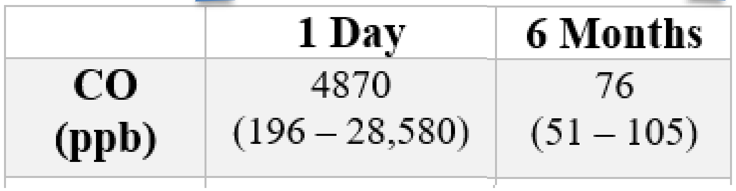
	1 Day	6 Months
$PM_{2.5}$	0.11 (-0.78 – 0.70)	0.84 (0.78 – 0.87)

UCONN HEALTH

Levy Zamora, Misti, et al. "Identifying optimal co-location calibration periods for low-cost sensors." *Atmospheric measurement techniques* 16.1 (2023): 169-179.

Range of potential RMSE for CO





	1 Day	6 Months
CO (ppb)	4870 (196 – 28,580)	76 (51 – 105)

Levy Zamora, Misti, et al. "Identifying optimal co-location calibration periods for low-cost sensors." *Atmospheric measurement techniques* 16.1 (2023): 169-179.

Thoughts on Co-location Length

- Longer calibrations tended to result in more accurate numbers

- Scenario: a user wanted all 250 potential co-location periods for the $PM_{2.5}$ sensor to have an RMSE below 4 μg/m^3 and an r > 0.6
  - The minimum co-location duration that would ensure all calibration periods satisfied these two requirements would be 108 days at this site

- It was possible to obtain a qualifying calibration in as little as one week
  - 22% of the 7-day co-locations also produced calibrations that satisfied these two requirements

	1 Day	1 Week	1 Month	6 Weeks	3 Months	6 Months
$PM_{2.5}$ (μg/m^3)	44.9 (5.2 – 400)	6.6 (3.1 – 18.3)	3.4 (3.1 – 9.1)	3.4 (3.2 – 7.9)	3.5 (3.2– 5.6)	3.6 (3.2 – 3.7)

- We hypothesized that the most important predictor is not length of co-location, but if the conditions during the co-location period represent the full measurement period

- We analysed the environmental factors during one-week calibrations that led to low and high RMSE

Coverage

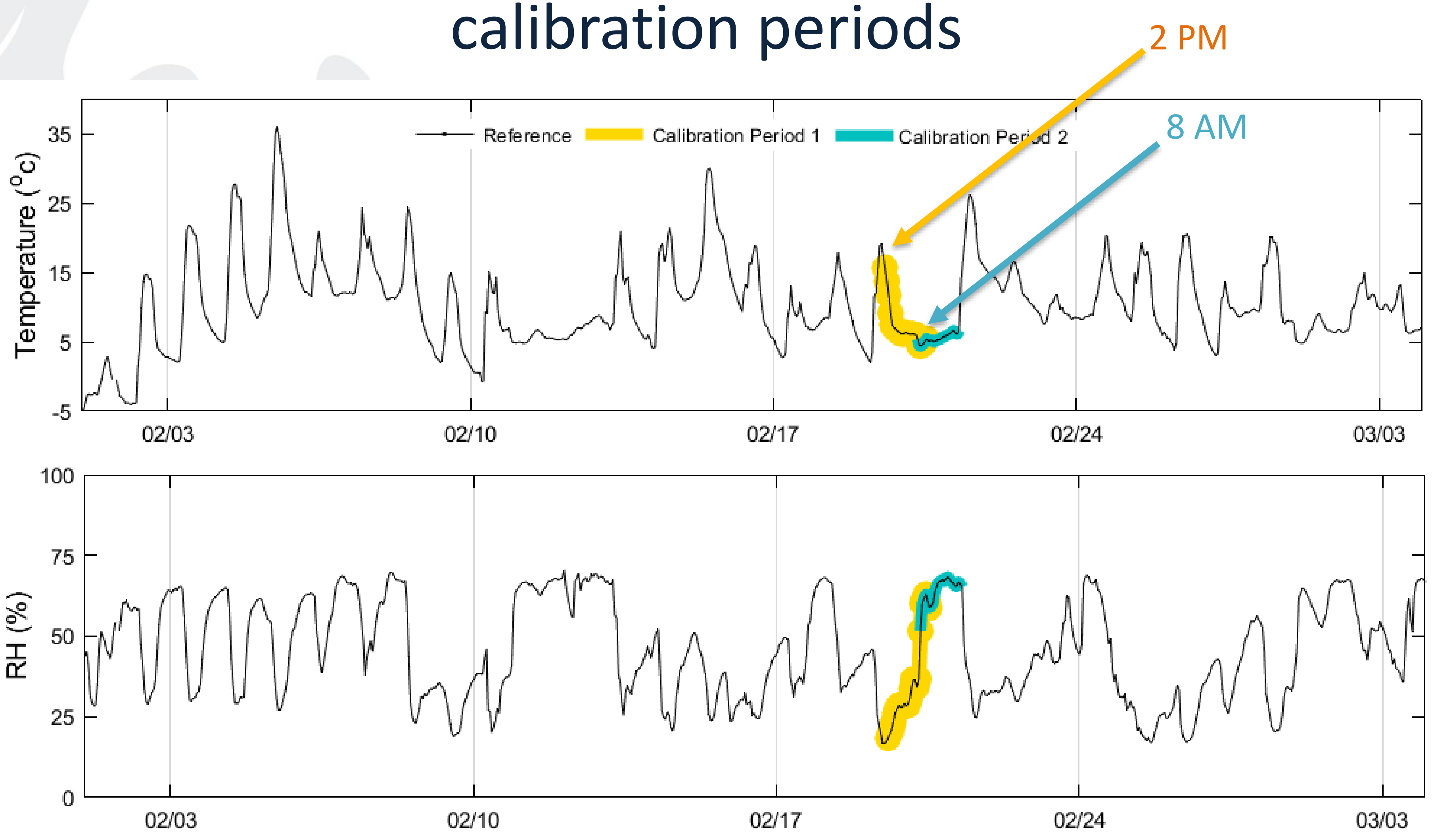
- A term to quantity how similar the co-location period was to the test period.

$$\text{Coverage} = \frac{Maximum\ Value_{Calibration\ Period} - Minimum\ Value_{Calibration\ Period}}{Maximum\ Value_{Full\ Year} - Minimum\ Value_{Full\ Year}} \text{ X } 100$$

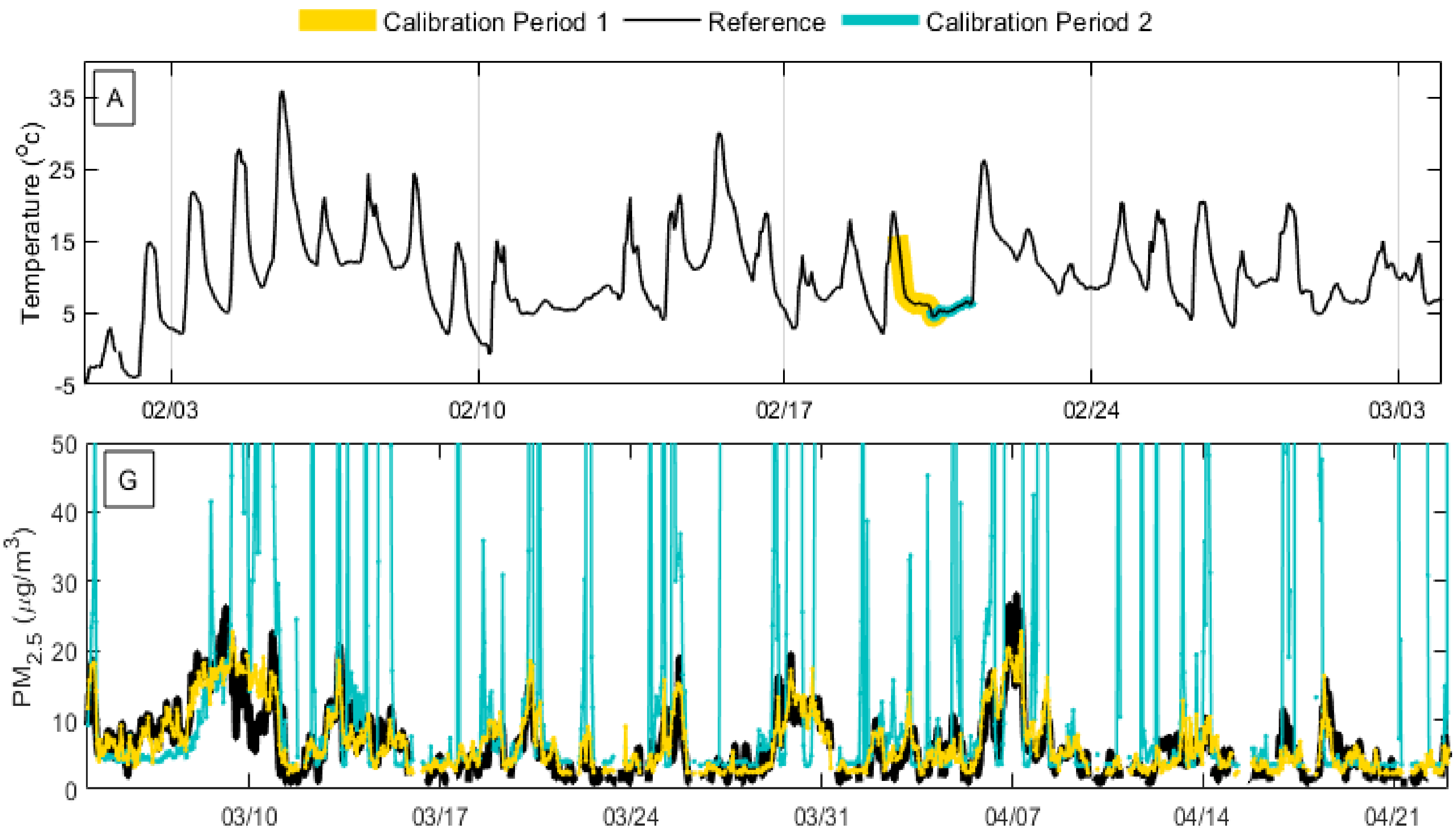
- For example:
  - Co-location temperatures ranged between -6 and 0^{o}C for one week
    - ΔTemperature = 6
  - Temperatures ranged between -6 and 47 ^{o}C during the full period
    - ΔTemperature = 53
- The coverage for that week would be about 11% (6/53*100)
- Calculate for every significant factor

Levy Zamora, Misti, et al. "Identifying optimal co-location calibration periods for low-cost sensors." *Atmospheric measurement techniques* 16.1 (2023): 169-179.

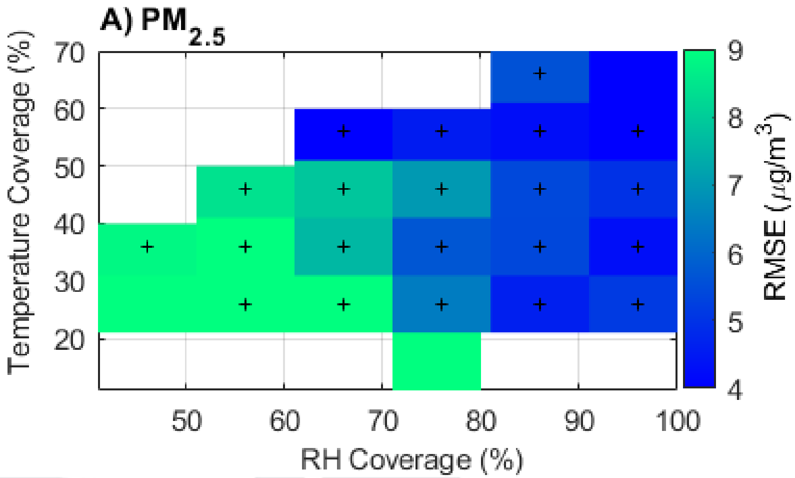
Example: Assessing the coverage of two calibration periods

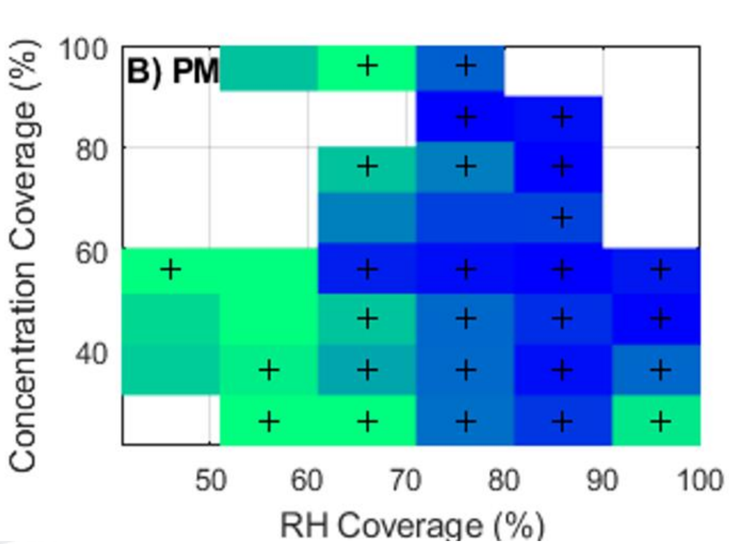


Which period does better?



Better Coverage = Better results



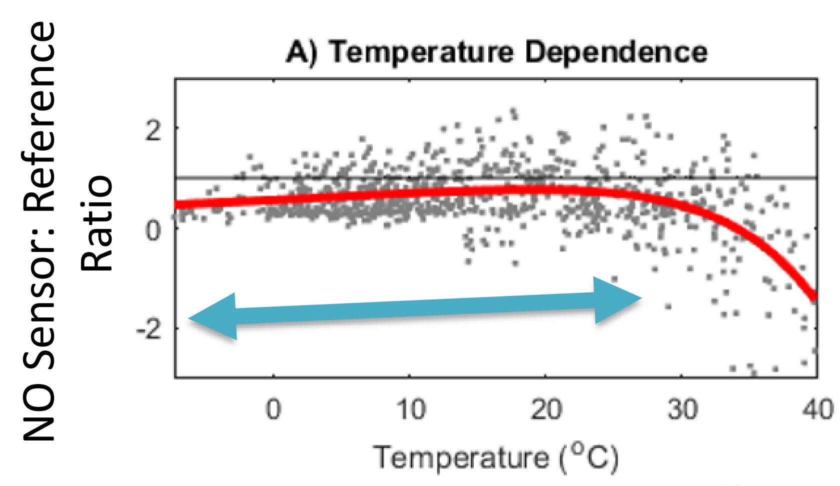


- Median RMSE values for $PM_{2.5}$ are shown as a function of RH and Temperature coverage for 1-week calibration periods
- Bluer colors indicate better calibration results with lower RMSE
- The + markers indicate where there were at least 25 calibration runs that fell within that box

Levy Zamora, Misti, et al. "Identifying optimal co-location calibration periods for low-cost sensors." *Atmospheric measurement techniques* 16.1 (2023): 169-179.

Other Considerations

- If a response is non-linear to any of these factors, you must ensure you capture that transition



- The duration of the full deployment (i.e., within a season or spanning multiple seasons).

Levy Zamora, Misti, et al. "Evaluating the performance of using low-cost sensors to calibrate for cross-sensitivities in a multipollutant network." ACS ES&T Engineering 2.5 (2022): 780-793.

Conclusions

- With optimal conditions it was possible to obtain an accurate calibration in as little as 1 week for all five sensors, suggesting that co-location can be minimized if the period is strategically selected and monitored so that the calibration period is representative of the desired measurement setting.

- Using measurements from Baltimore, MD, where a broad range of environmental conditions may be observed over a given year, we found diminishing improvements in the median RMSE for calibration periods longer than about 6 weeks for all the sensors.

- Several factors increased the co-location duration required for accurate calibration, including the response of a sensor to environmental factors, such as temperature or relative humidity (RH), or cross-sensitivities to other pollutants.

- The best performing calibration periods were the ones that contained a range of environmental conditions similar to those encountered during the evaluation period (i.e., the true measurement period)

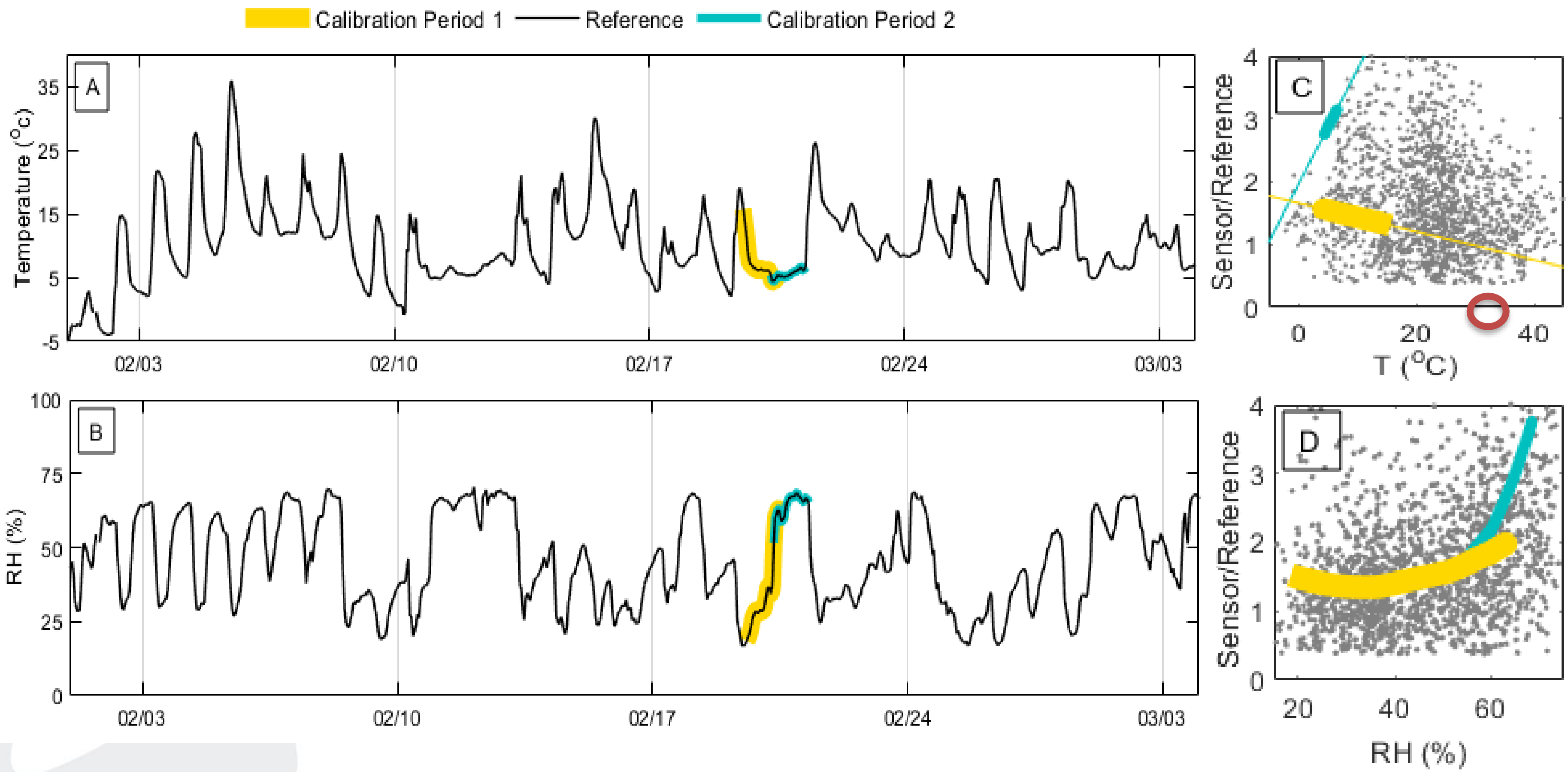
Conclusions & Recommendations

- A benefit of strategically identifying co-location needs is that it may permit users of sensor networks to co-locate each device in the network for shorter periods to get device-specific calibration equations.
  - By ensuring a minimum coverage of key factors for each device co-location period, calibration data between units would likely be more consistent even if the data were collected from different periods.
  - This would be particularly advantageous for sensor types that exhibit notable variability between units.
- If little information is known about key predictors at the measurement sites, which is likely at remote locations, it may be possible to use historical meteorological data and general information about pollutant patterns (e.g., emissions and seasonal concentration patterns) to determine a representative range of conditions.
- To yield the best performing calibration outcomes, highly influential cross-sensitives or environmental factors should have a minimum coverage of about 70% and secondary factors should have a minimum coverage of about 50%.
  - Prioritize the most significant factors
- It is advisable to increase the estimated co-location periods in case of data loss or unusual air quality events to increase the probability of well-performing calibrations.

Thank You

mzamora@uchc.edu
https://twitter.com/MistiLevyZamora

Example: Assessing the coverage of two calibration periods


Calibration Period 1
Reference
Calibration Period 2
A
Temperature (°c)
02/03
02/10
02/17
02/24
03/03
B
RH (%)
02/03
02/10
02/17
02/24
03/03
C
Sensor/Reference
T (°C)
D
Sensor/Reference
RH (%)