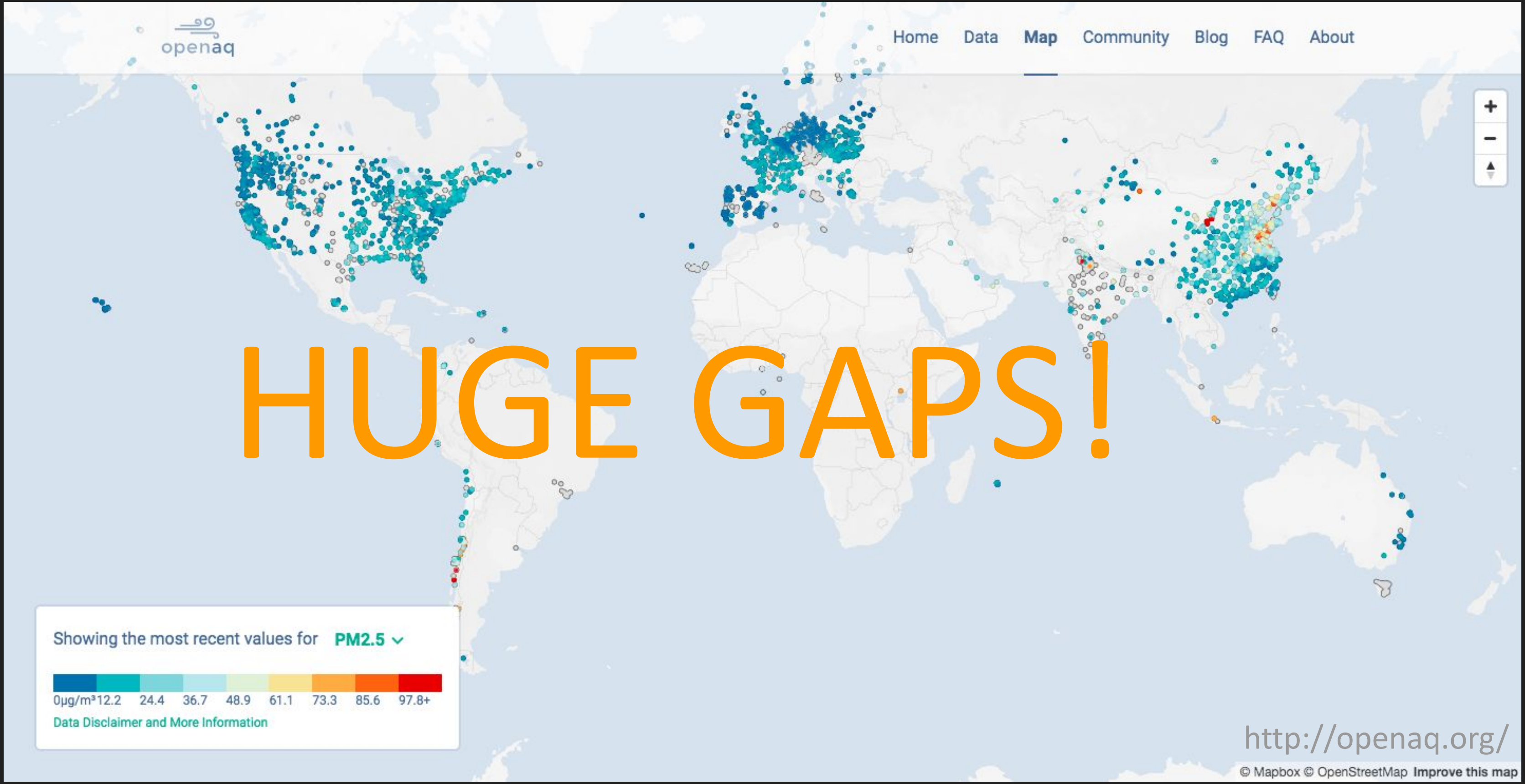
Investigating Use of Low-Cost Sensors to Increase Accuracy and Equity of Real-Time Air Quality Information

Ellen Considine
Mentors: Rachel Nethery, Danielle Braun, Priyanka deSouza
ASIC, 2022

Air Quality Data Reported by Countries



Motivations

- Low-cost sensors (LCS) could fill information gaps between reference monitors
  - Cost per instrument: $250 vs. $10,000+
- Concerns about measurement error from LCS
- Persistent air inequality between socioeconomic and demographic groups
  - Discrepancy increasing as overall AQ improves in US (Jbaily et al., 2022)
- deSouza & Kinney (2021): PurpleAir sensors (most common $PM_{2.5}$ LCS) tend to be in more privileged areas
- US EPA recently allocated $20 million for enhanced community AQ monitoring
  - Priority: environmental justice (EJ) communities

Research Questions

1. How does the distribution (density and placement) of low-cost sensors affect real-time air quality (AQ) information, in terms of both accuracy and equity?
2. What mechanisms drive inaccurate AQ reporting, and how important are these mechanisms relative to each other?
3. How does the type and amount of sensor measurement error affect the usefulness of LCS in real-time AQ reporting?

Overview of Simulation Method

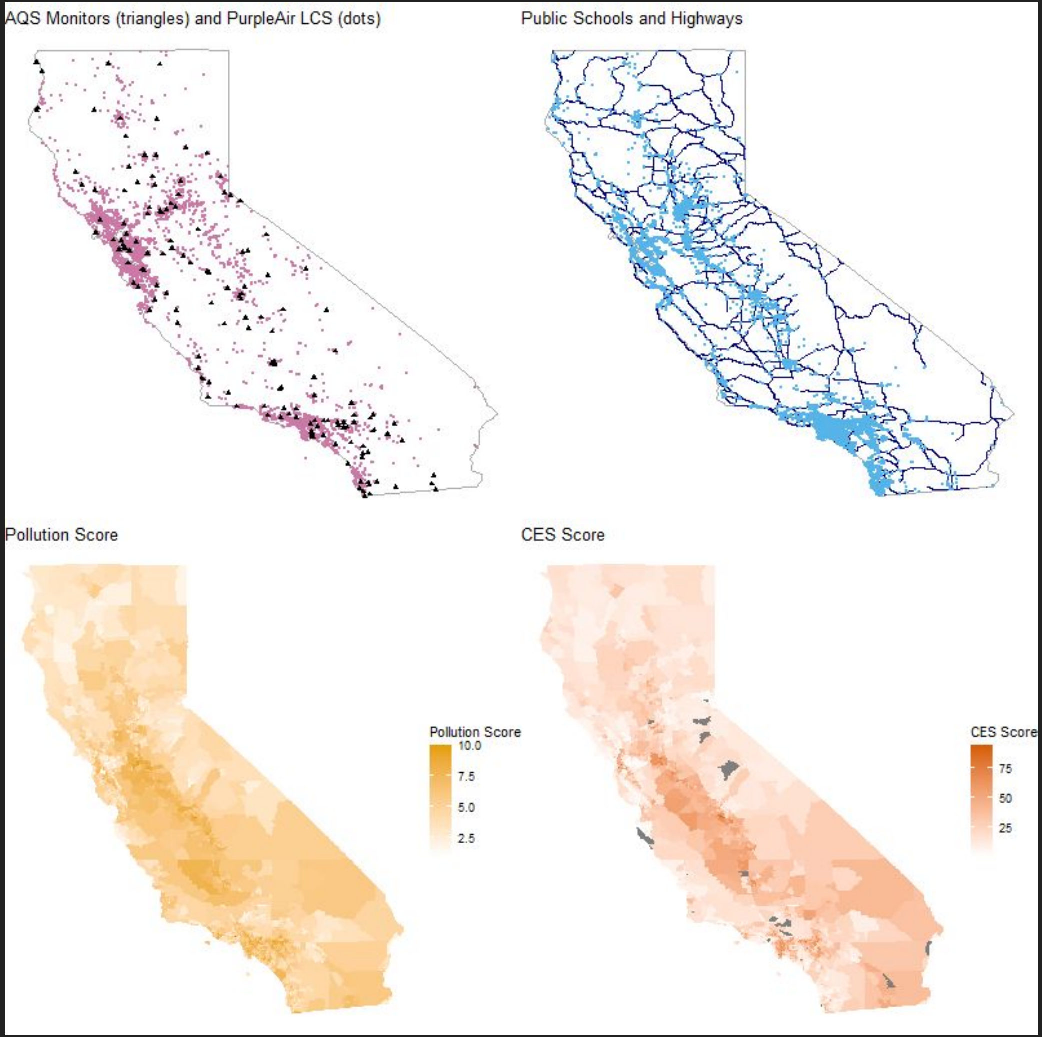
- Setting: California, 2016
- Consider daily 1x1km Di et al. $PM_{2.5}$ estimates “truth”
- Assume AQS reference monitors are fixed and have no measurement error
- Select LCS locations (based on one placement strategy), x 50
- Simulate measurement error at “sensor” locations
- Each grid cell “sees” AQ info from the nearest monitor/sensor
- Calculate performance metrics (comparing the “true” AQ to the “shown” AQ), overall and by subgroups, averaged across the 50 trials
- Main analysis: weighted by population density
  - Secondary analysis: unweighted

Repeat for different types and amounts of sensor measurement error.

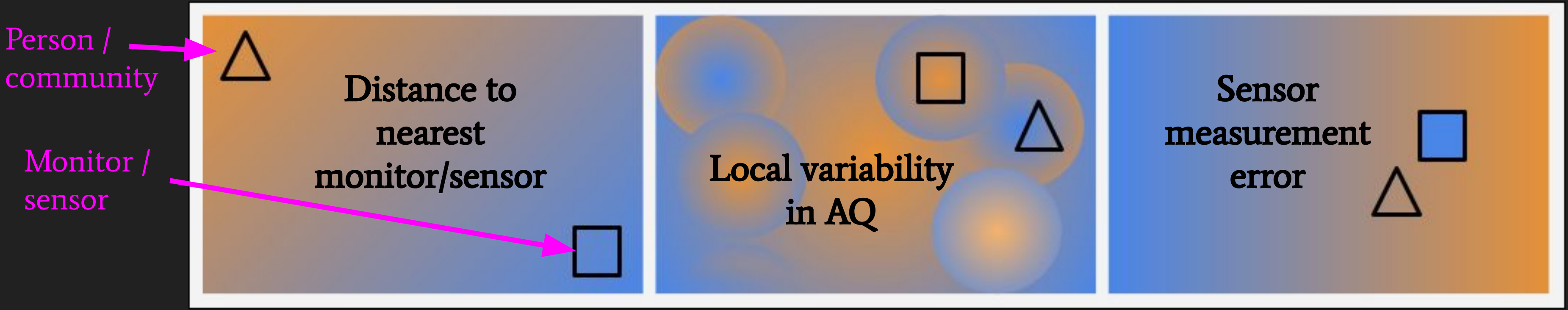
Placement Strategies

Randomly-selected n = 50, 100, 250, 500, and 1,000:

- Current PurpleAir locations
- Schools
- Weighted by lengths of major roads within a 500m buffer
- Weighted by environmental and socioeconomic marginalization:
    - CalEnviroScreen: Pollution Score and compound CES Score



Investigating Mechanisms Driving Errors in Reported AQ

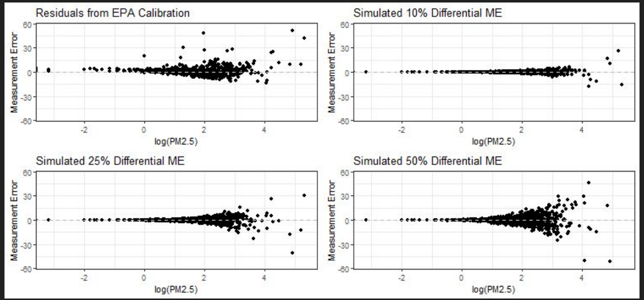


Calculations:

- Distance to nearest monitor/sensor
  - Overall and among misclassified observations
    - “Classes” from EPA AQI
- Sensitivity analysis: simulations without adding sensor measurement error

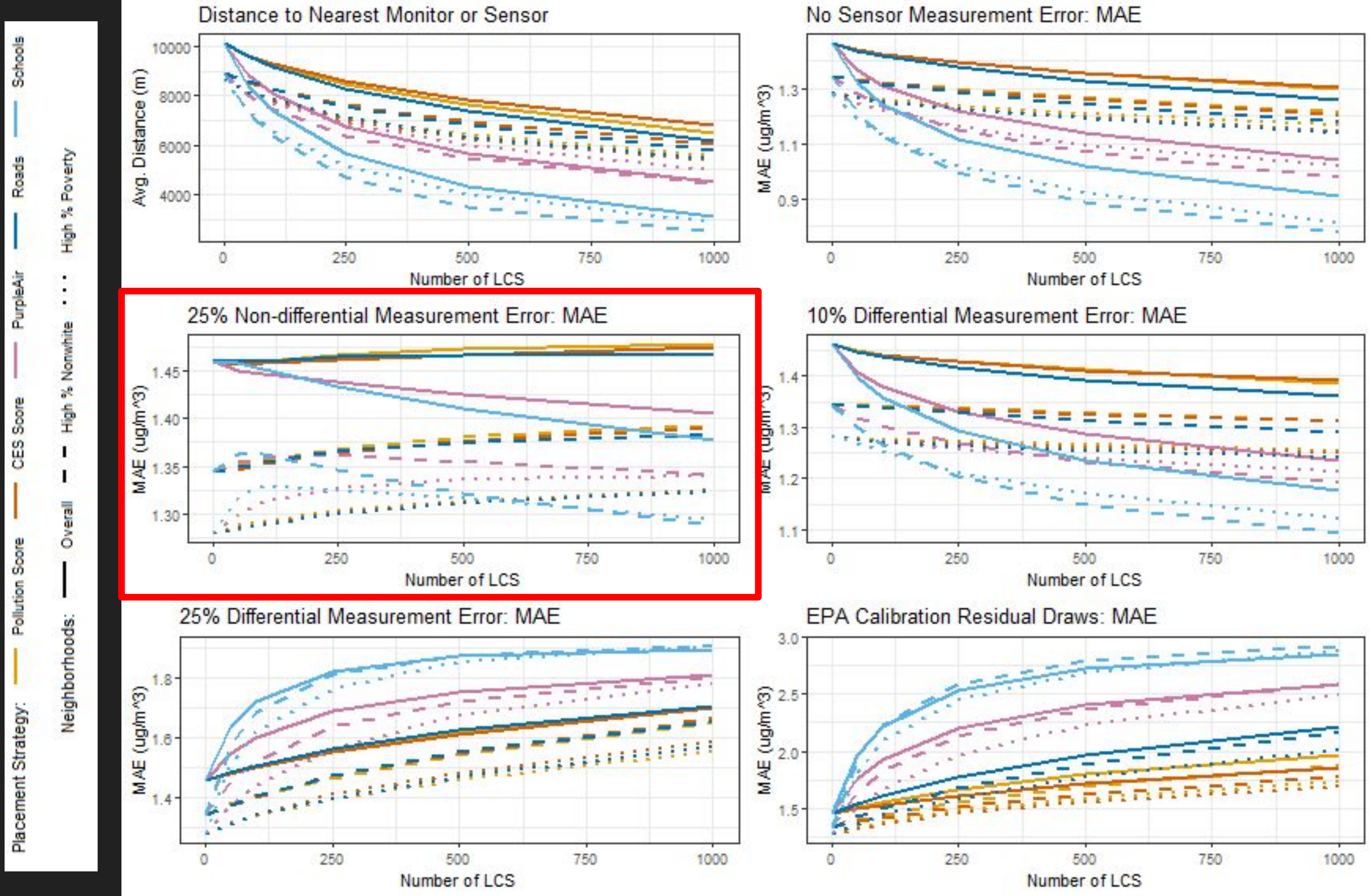
LCS Measurement Error Simulation

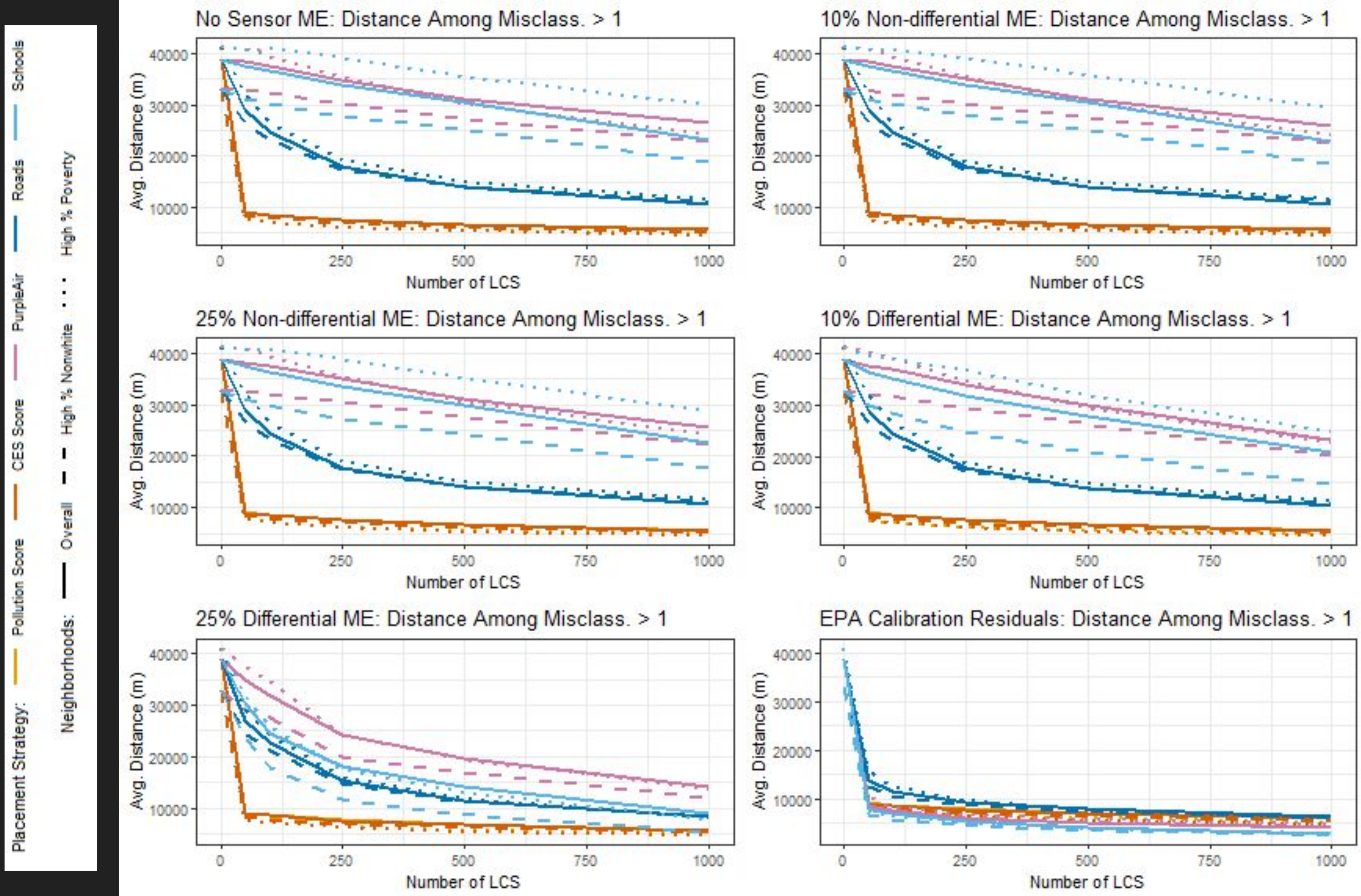
- Williams, et al. (2019):
  - Regulatory monitoring: ±10% of the average AQ levels in an area
  - Mapping spatial gradients and monitoring microenvironments: ±25%
  - ±50% accuracy still useful for tracking large sources of air pollution
- EPA developed linear correction for PurpleAir in US – Barkjohn, et al. (2021)
- Our simulations:
  - No LCS measurement error
  - Non-differential measurement error, 10% and 25% of avg.
  - Differential measurement error, 10% and 25% of “truth”
  - Sampled empirical residuals from EPA correction
    - Obtained from collocated monitor-sensor pairs



Now for some results…

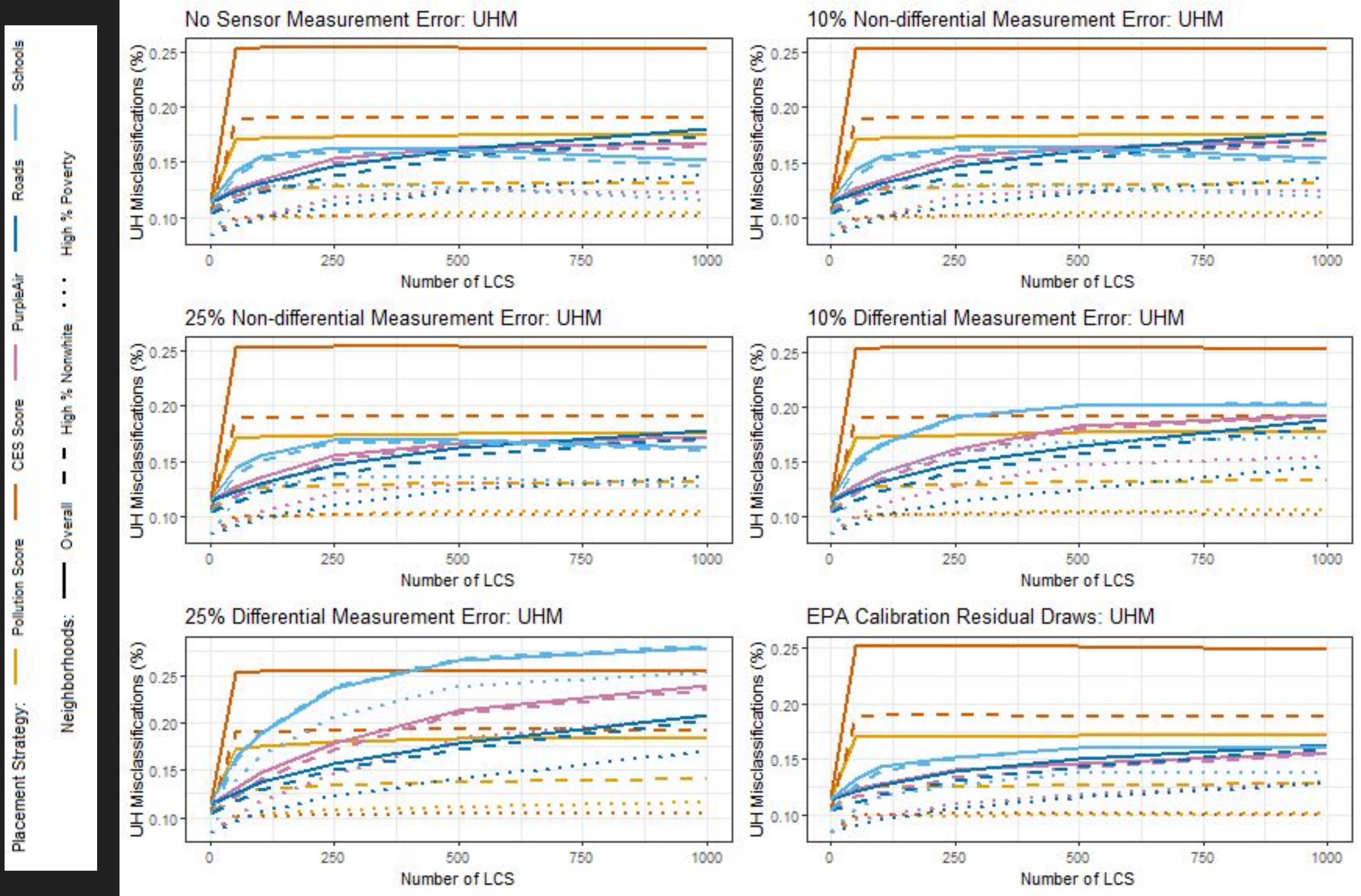
Weighted by Population Density





Weighted by Population Density

Weighted by Population Density



Overall Conclusions

- The value of using LCS for real-time AQ reporting depends strongly on type and amount of sensor measurement error (ME)
    - 25% non-differential ME appears to be workable for some placement strategies, but not others
    - LCS corrections may need to be more localized / advanced than a national linear correction
- With low-to-moderate amounts of ME (depends on type): placing LCS at schools results in the greatest decrease in MAE, for all demographics considered, as the number of LCS increases
- Placing LCS in EJ hotspots may help the immediate community, but can cause issues when integrated into wider AQ reporting platforms
- Balancing policy priorities will be tricky (e.g. urban vs. rural)

Limitations / Future Directions

- Real-time AQ info is only a first step for public health
  - Behavioral insights...
- Looking beyond location of residence
  - Other sensors and activity patterns...
- Results from California in 2016 not necessarily generalizable
  - Value of adding LCS depends on coverage of existing reference monitors
- More advanced placement strategies, more targeted EJ placements
- Spatiotemporal resolution of truth surface…
  - Other air pollutants may exhibit even more local variability
- More advanced / regional calibrations or simulations for measurement error

Unweighted by Population Density