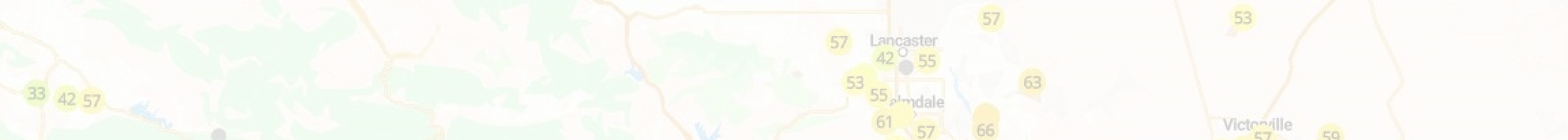
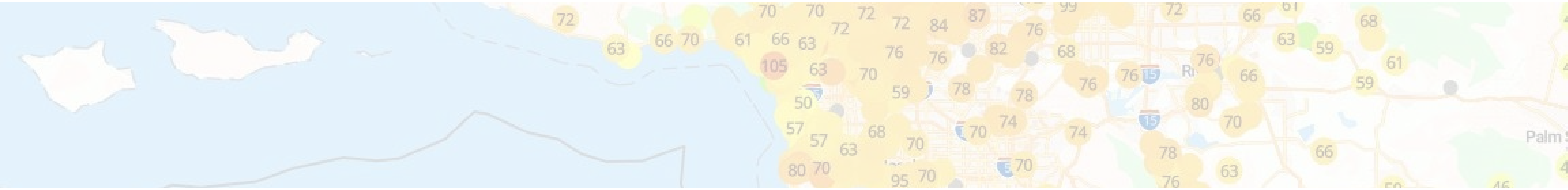


Using Crowd-Sourced Low-Cost Sensors in a Land Use Regression of $PM_{2.5}$ in 6 US Cities



Tianjun Lu[1], Matthew J Bechle[2], Albert A Presto[3], Steve Hankey[4]

[1]California State University, Dominguez Hills; [2]University of Washington

[3]Carnegie Mellon University; [4]Virginia Tech

ASIC 2022

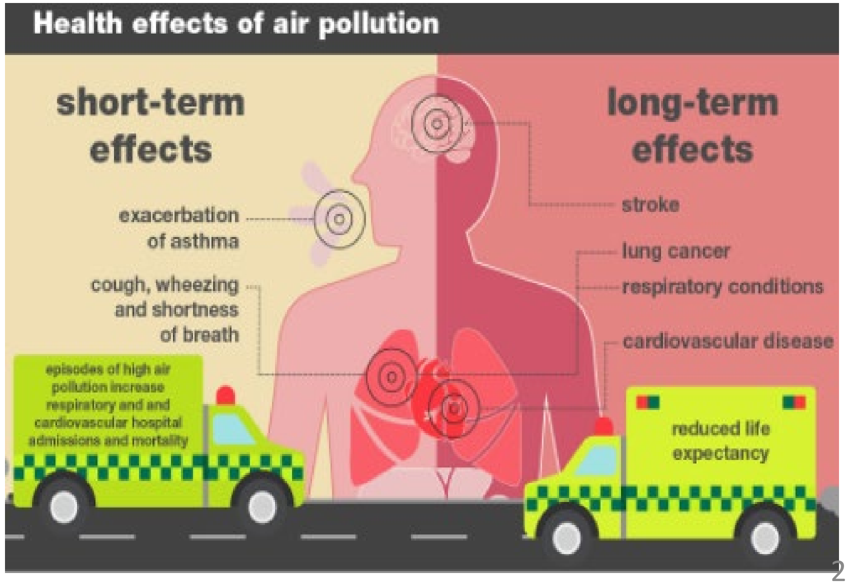
May 11, 2022

CSUDH
CALIFORNIA STATE UNIVERSITY, DOMINGUEZ HILLS

Background and Motivation

- **Health-promoting** cities and air quality.
- Health effects; policy; air quality **monitoring**.
- **Valuable** regulatory monitors.
- Growing global interest in **public data collection**.

EPA United States Environmental Protection Agency



Crowd-sourced efforts in exposure assessment

Low-Cost Sensing

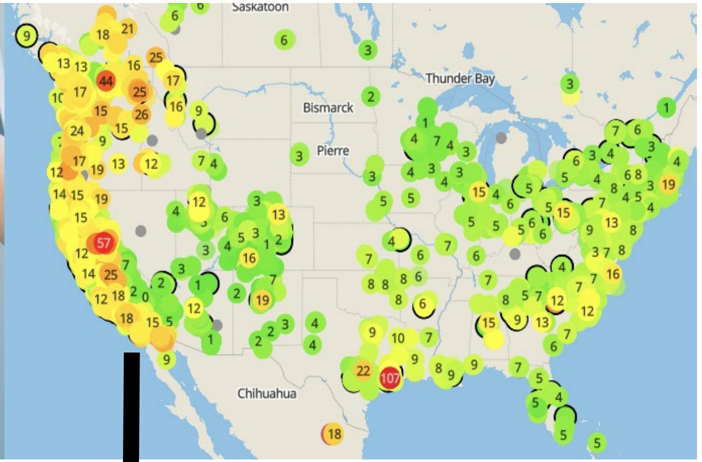
Low-cost air quality sensing

- **Dense** fixed sensor network.
- **Community** engagement.
- **“Open” data**.

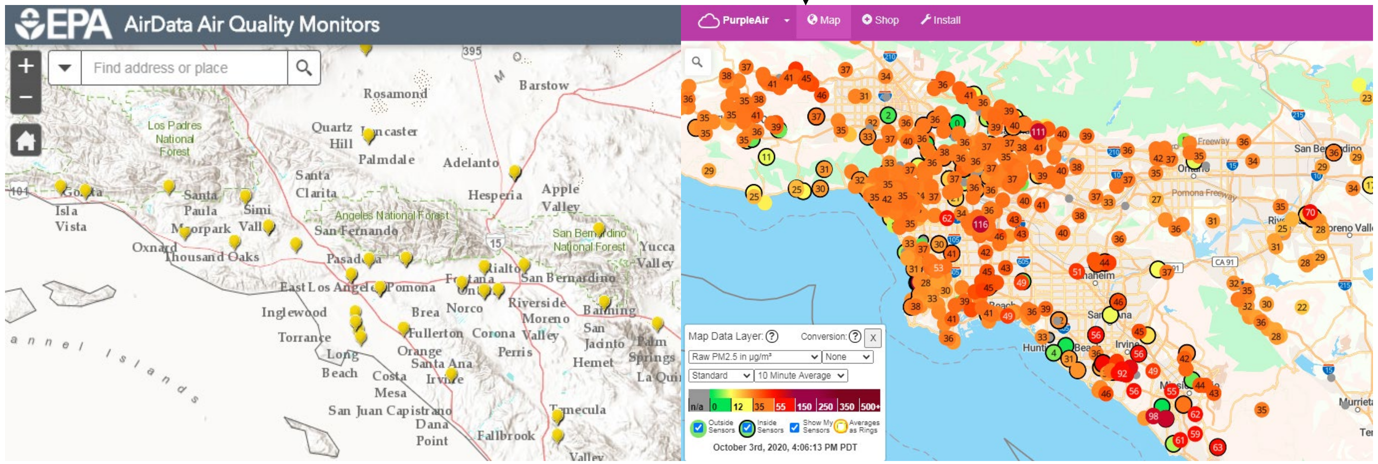
Data quality

- Relative humidity, temperature.
- Careful lab and in-field **calibrations**.
- **Well correlation** with reference measurements.
- **Emerging calibration** efforts.

Low-cost sensors



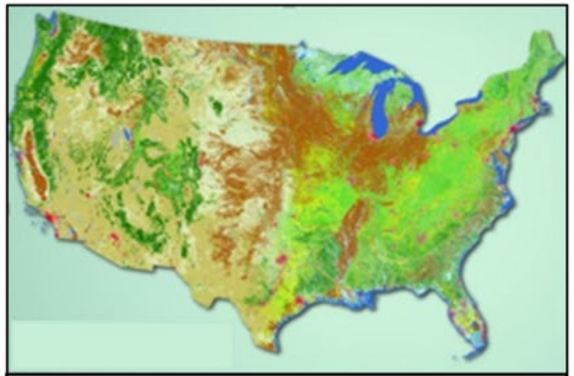
PurpleAir network



EPA air quality monitors vs. PurpleAir sensors

How Low-cost Sensing Help?

Regulatory monitors



Geographic variables

Traditional air quality models

+

Low-cost sensors

Crowd-sourced data

- Little research assessed the **utility** of such growing network from **multiple** cities in land use regression (LUR).
- Possibility to **improve** the LUR model **to capture spatial variability**?

Existing National LUR Model

CACES LUR

PLS-UK partitions annual average concentrations into

- (1) a **variance** component that accounts for spatial and non-spatial variability.
- (2) a **mean** component based on a small number of reduced dimension variables from partial least squares of a large number of independent variables (Kim et al., 2020).

Category	Measure	Note[a]
Traffic	Distance to the nearest road (0.05-15 km)	Any available road
Population	Sum (0.5-3 km)	Population in block groups
Land use/land cover (Urban)	Percent (0.05-15 km)	Urban or built-up land, etc.
Land use/land cover (Rural)	Percent (0.05-15 km)	Agriculture, forest, water, etc.
Position	Coordinates	Longitude, latitude
Source	Distance to the nearest source	Coastline, railroad, airport, etc.
Emission	Sum of cite-specific facility emissions (3-30 km)	$PM_{2.5}$
Vegetation	Quantiles (0.5-10 km)	Normalized Difference Vegetation Index
Imperviousness	Percent (0.05-5 km)	Impervious surface value
Elevation	Counts of points above/below a threshold (1-5 km)	Elevation value
Satellite estimate	Grid-level estimates	$PM_{2.5}$

[a]Detailed information can be found from the CACES LUR modeling study (Kim et al., 2020).

11 categories of geographic variables
339 independent variables
757 regulatory $PM_{2.5}$ monitoring sites

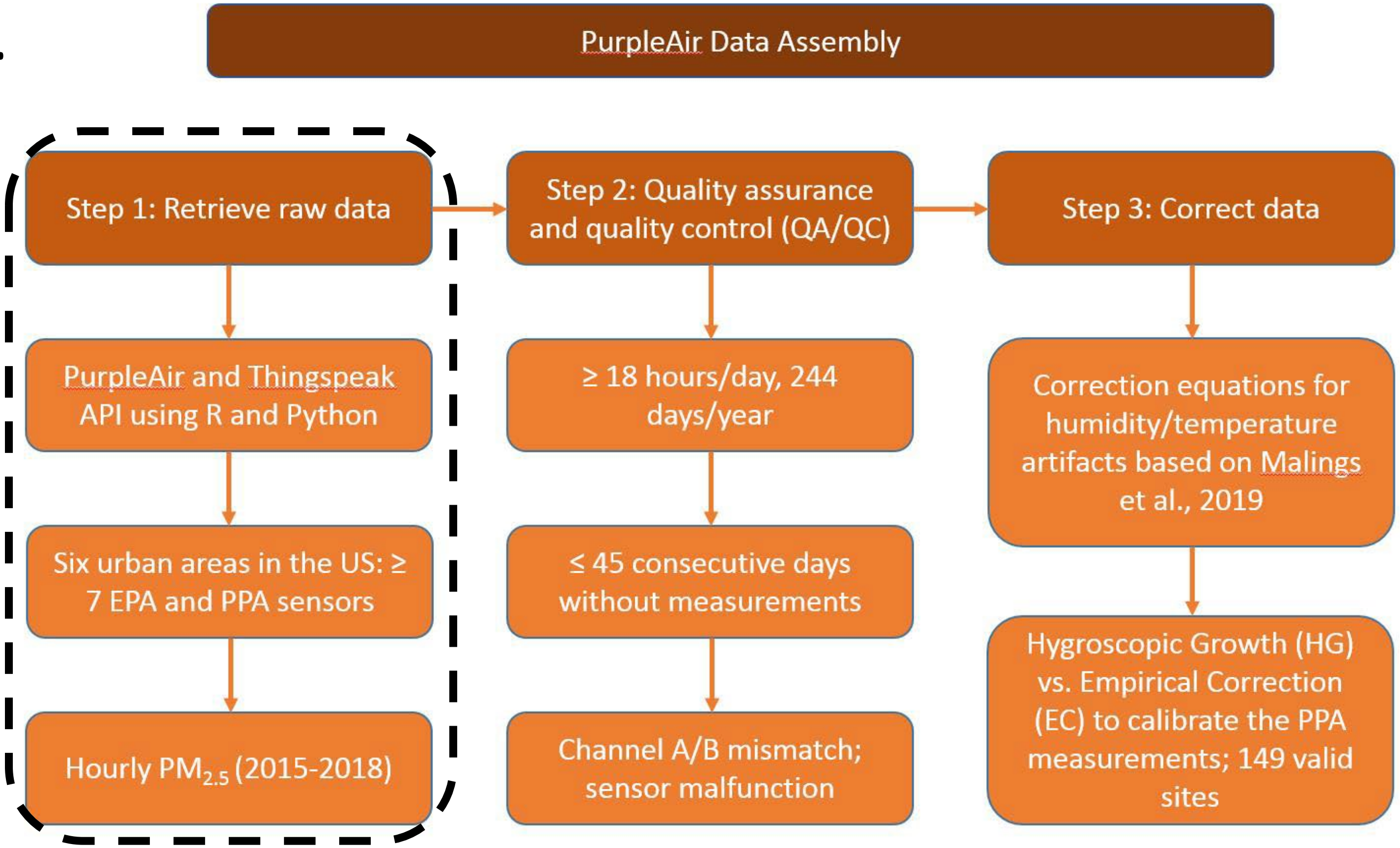
CACES LUR (random 10-fold CV: R^2 = 0.83; standardized RMSE = 0.13)

CACES: Center for Air, Climate, and Energy Solutions
PLS-UK: Partial Least Squares-Universal Kriging

PurpleAir (PPA) Data Preparation

PPA data assembly

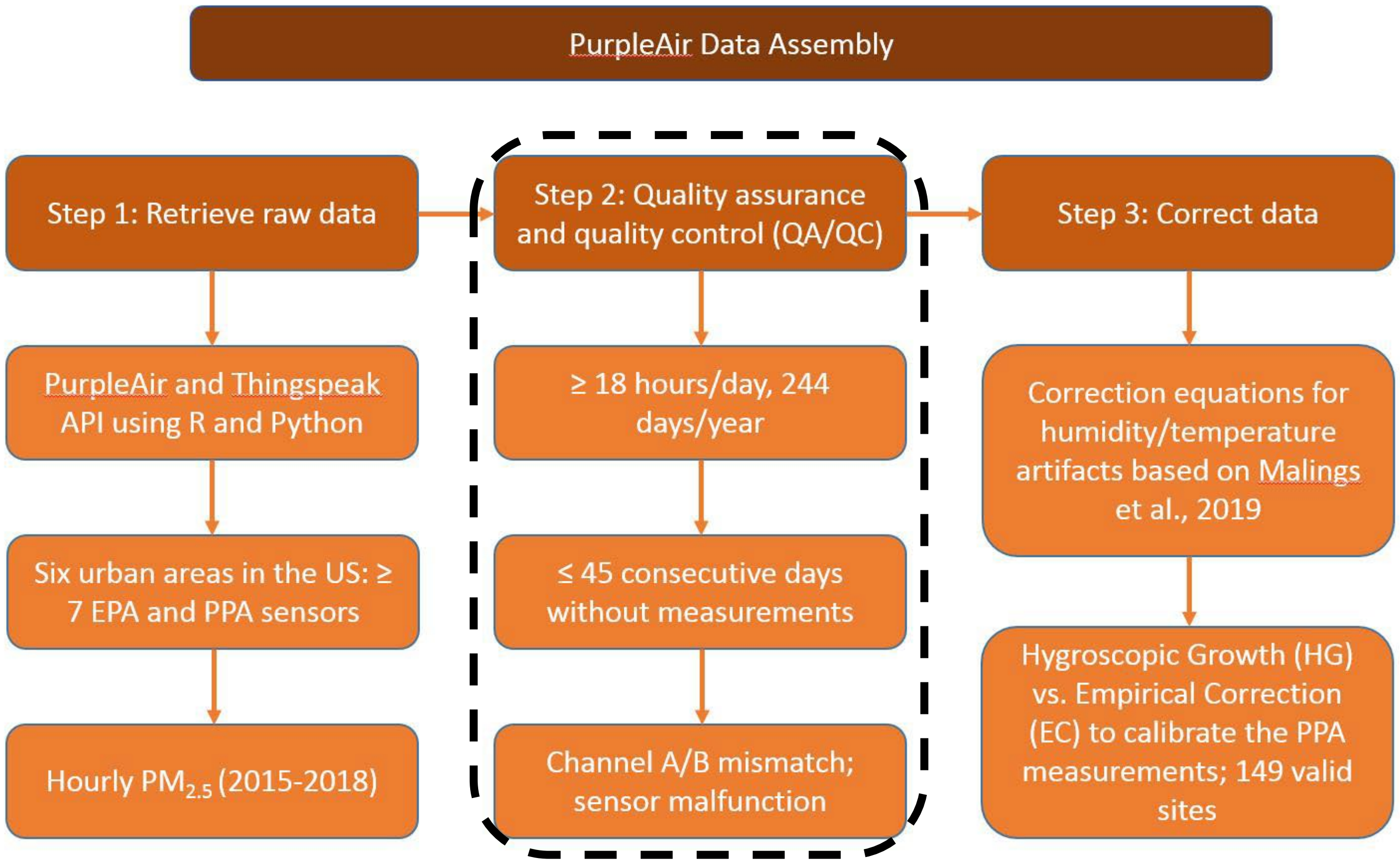
- **Six cities**: ≥ 7 EPA and PPA sensors.



PurpleAir (PPA) Data Preparation

PPA data assembly

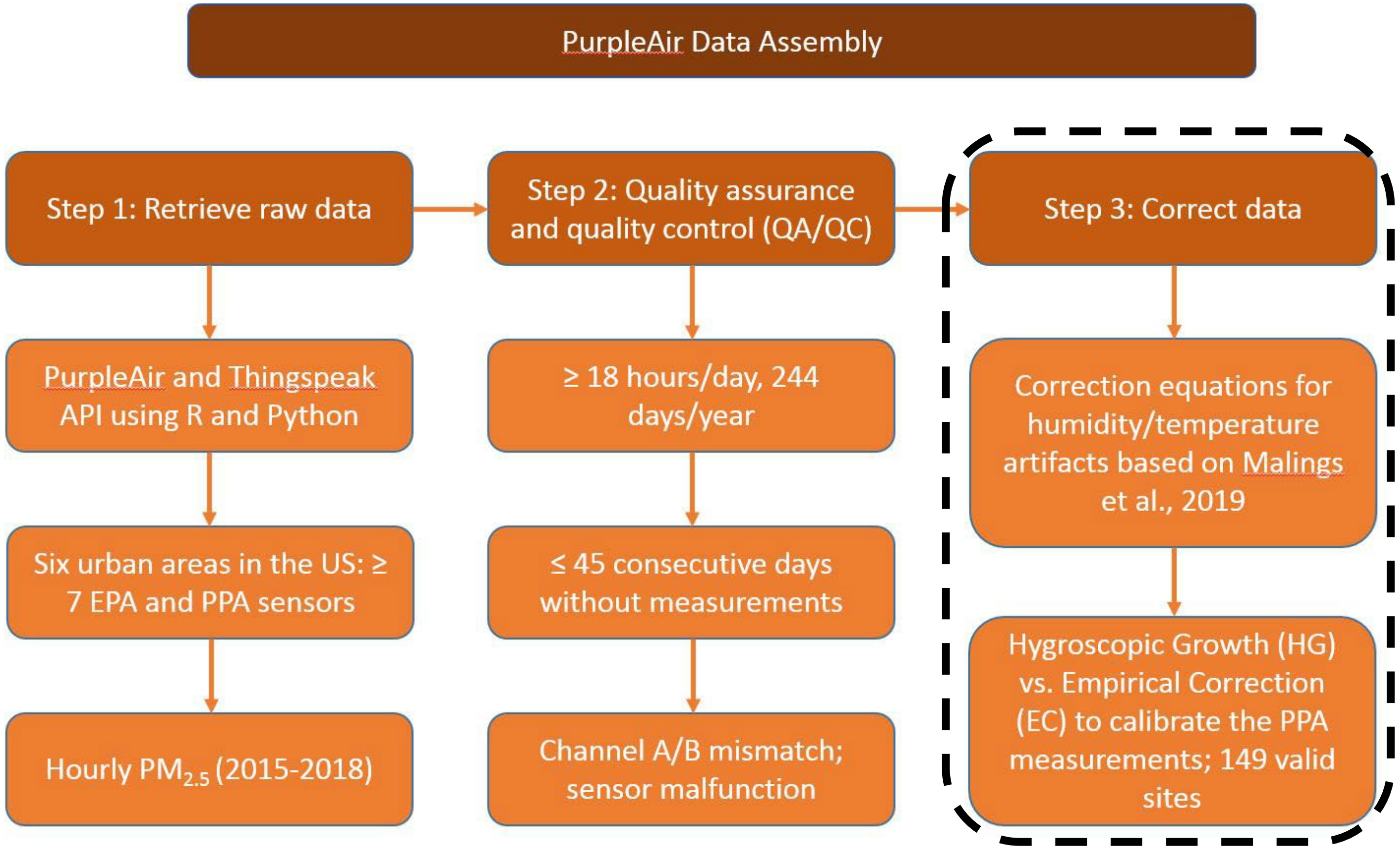
- **Six cities**: ≥ 7 EPA and PPA sensors.
- **QA/QC**:
  - same criteria as the CACES LUR
  - channel mismatch (removing hours when the absolute difference was larger than **3 μg/m³** or **20%** of the maximum channel readings, whichever is greater (Malings et al., 2019).



PurpleAir (PPA) Data Preparation

PPA data assembly

- **Six cities**: ≥ 7 EPA and PPA sensors.
- **QA/QC**:
  - same criteria as the CACES LUR
  - channel mismatch (removing hours when the absolute difference was larger than **3 μg/m³** or **20%** of the maximum channel readings, whichever is greater (Malings et al., 2019).
- **Data correction**:
  - humidity and temperature artifacts;
  - colocation calibrations.



LUR Model Development

<u>Dependent variables (annual averages)</u>

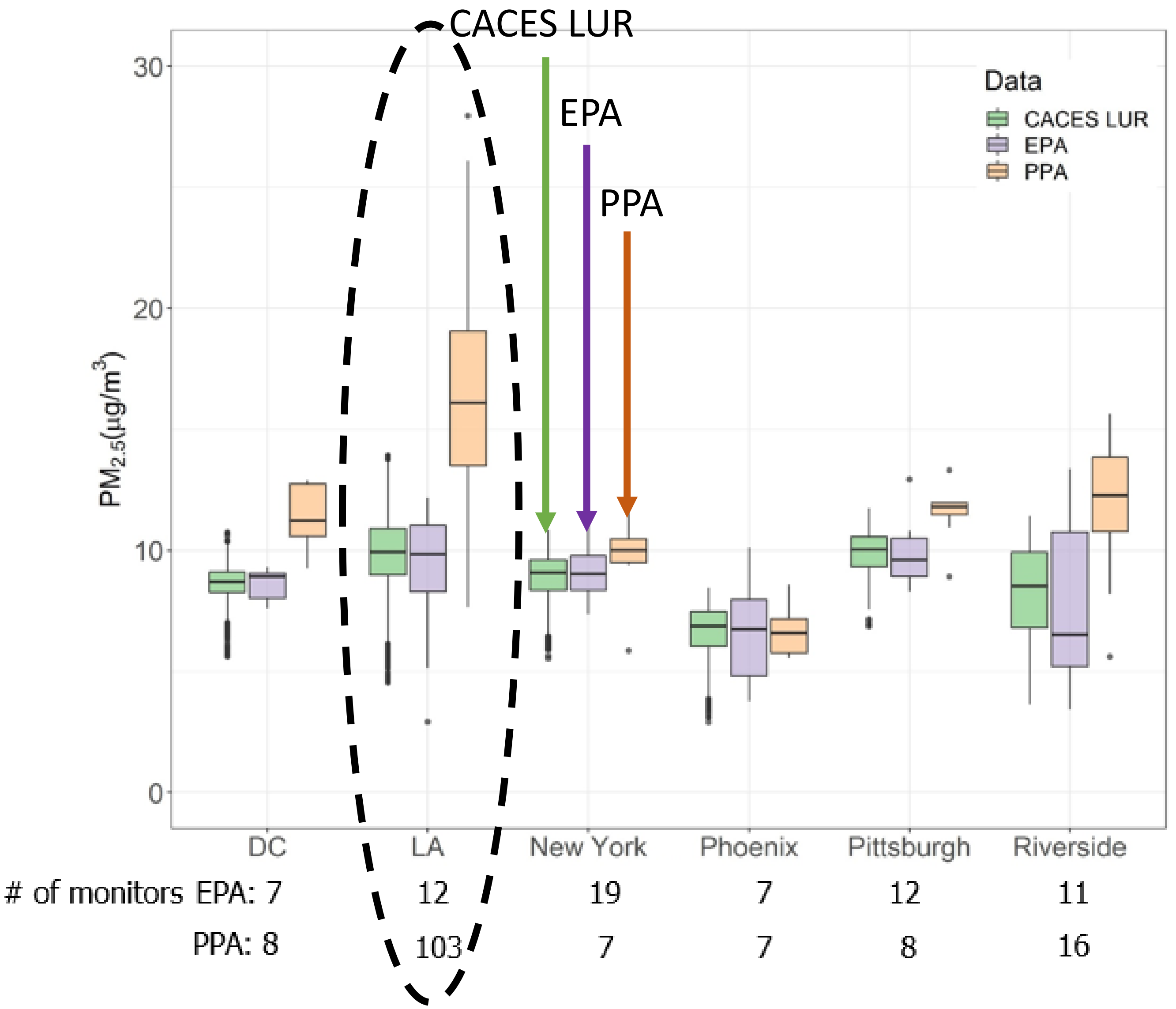
- EPA data (national and 6 cities).
- PPA data (6 cities).
- Hybrid (EPA + PPA data).

<u>Independent variables</u>

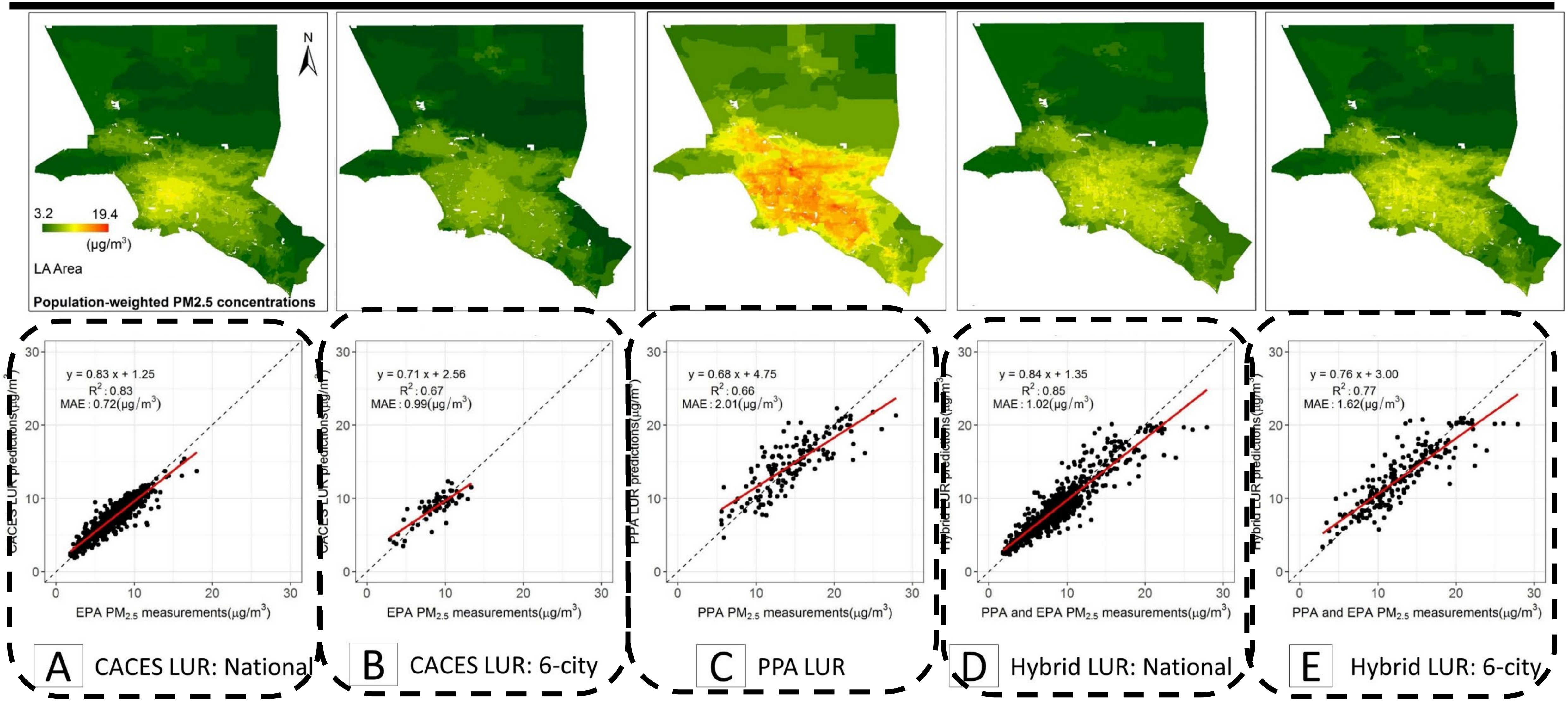
- 11 categories (e.g., traffic, population, land use).

<u>Modeling approach</u>

- PLS-UK.



LUR Model Comparison (Pop-weighted)



Population-weighted $PM_{2.5}$ concentration maps

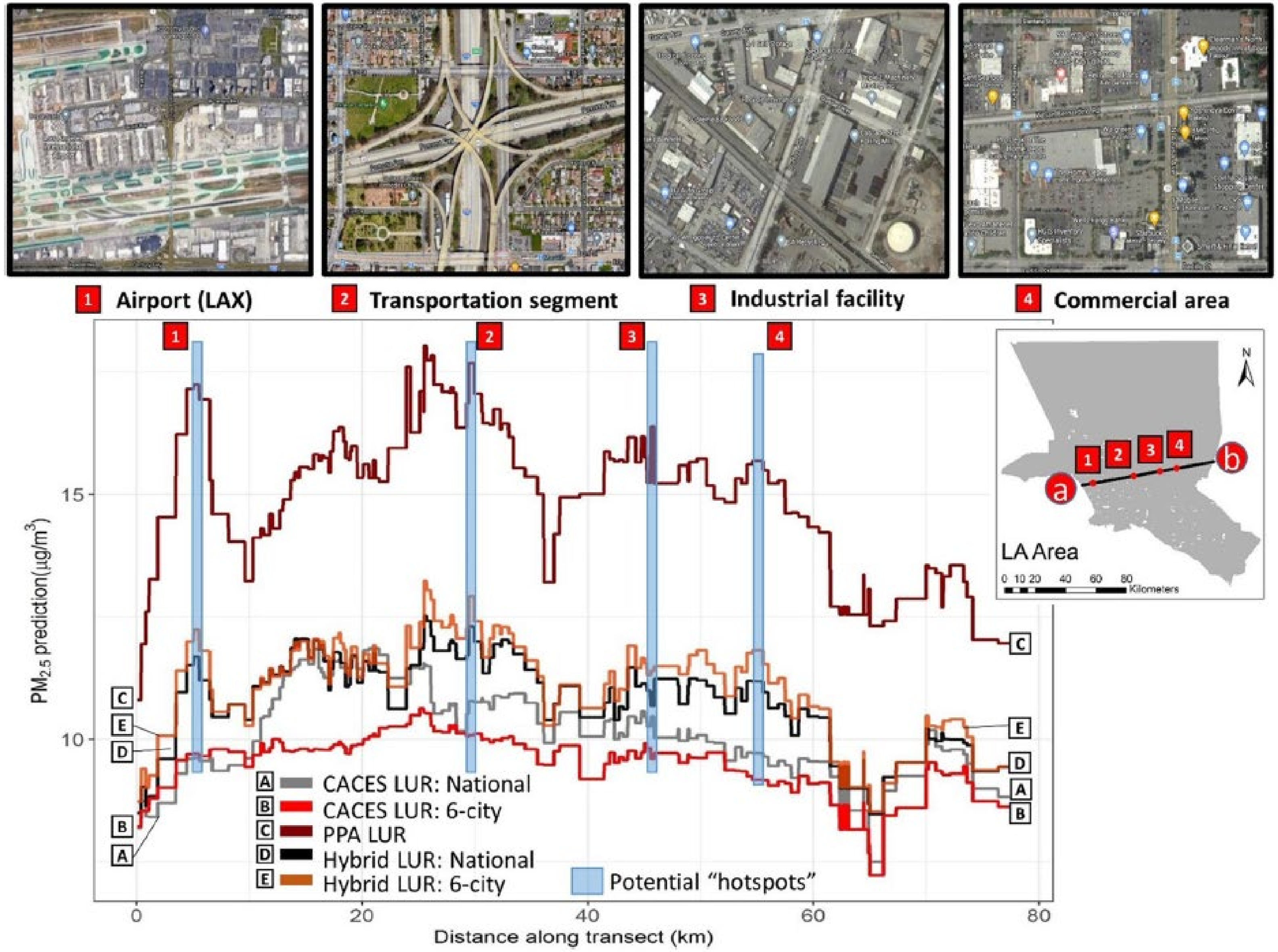
LUR Model Comparison (Transect Plots)

- Transect plot of the five LUR predictions.

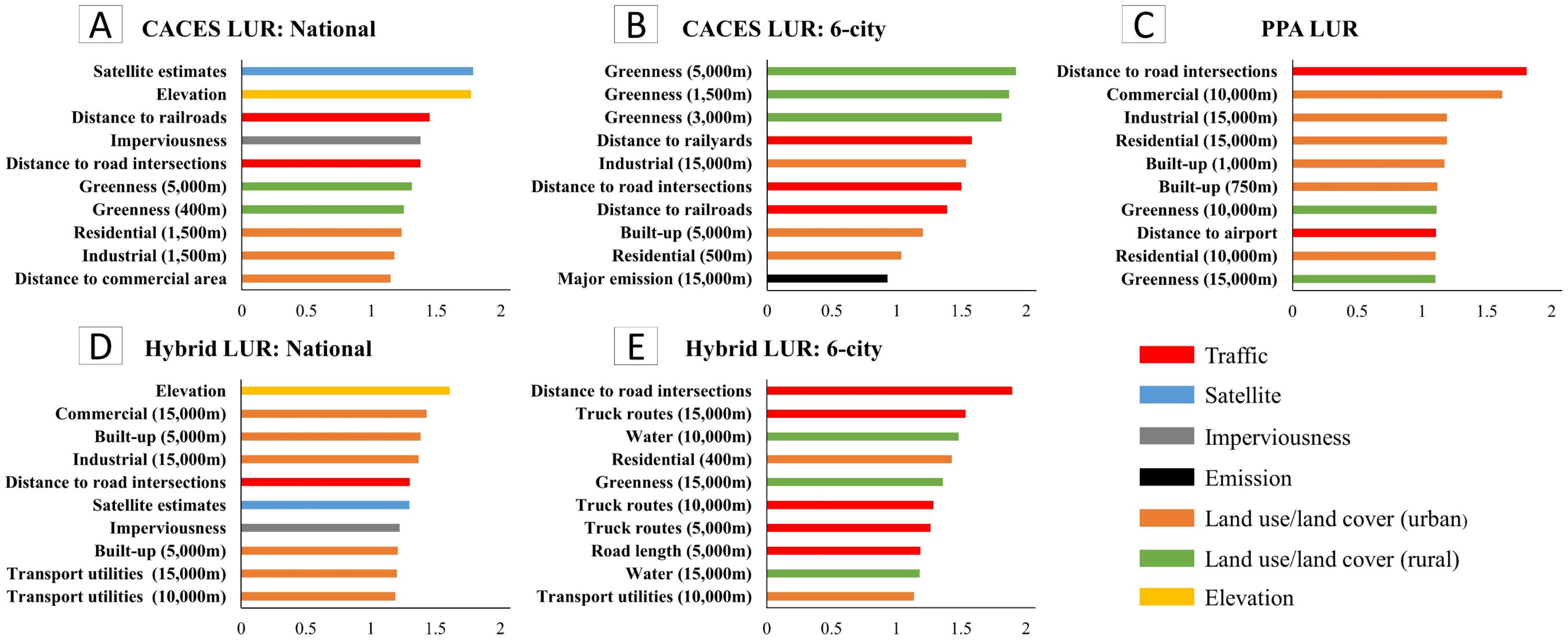
<u>Advantages</u>

- Models with the PPA data were **more spatially variable** than models without.

- Models with the PPA data alone is **not recommended**.



Variable Importance



- Traffic and land use variables were important variables for models with the PPA data; strength of capturing “hotspots”.

Summary and Implications

- Hybrid models may capture **small-scale variations** that may be **missed** by the regulatory-based models
- Valuable dataset for LUR if data is **carefully** cleaned and calibrated.

- With available national correction approaches (Barkjohn et al., 2021), additional cities would help assess tradeoffs in **national vs. local corrections**.
- Calibrations by **co-locating** PPA sensors with regulatory-grade monitors in additional cities may help reduce bias.
- Further empirical investigation is warranted in hybrid models with **additional sensors** from **larger areas and multiple cities**.
- Neighborhood **planning and design**; clean streets; guidance on **outdoor** activities; interventions.

Acknowledgement and Contact

Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-022-01162-7

Check for updates

Using crowd-sourced low-cost sensors in a land use regression of $PM_{2.5}$ in 6 US cities

Tianjun Lu[1] · Matthew J. Bechle[2] · Yanyu Wan[3] · Albert A. Presto[3] · Steve Hankey[4]

Received: 8 September 2021 / Accepted: 19 January 2022

[2] Department of Civil & Environmental Engineering, University of Washington, 201 More Hall, Seattle, WA 98195, USA

[3] Department of Mechanical Engineering, Carnegie Mellon University, 2115 Doherty Hall, Pittsburgh, PA 15213, USA

[4] School of Public and International Affairs, Virginia Tech, 140 Otey Street, Blacksburg, VA 24061, USA

Tianjun (Luke) Lu

tilu@csudh.edu

https://tianjunlu.weebly.com/

EPA United States Environmental Protection Agency

CACES

VIRGINIA TECH™

CARNEGIE MELLON UNIVERSITY PITTSBURGH PENNSYLVANIA 1900

W

CSUDH CALIFORNIA STATE UNIVERSITY, DOMINGUEZ HILLS

Supplemental Material

Hygroscopic Growth (HG) Correction

HG correction

- Adjusted to be "Beta Attenuation Monitors (BAM) equivalent".
- **Over** prediction at high RH and **under** prediction of particles < 300 nm.
- Cities **with/without** co-located PPA sensors.
- Either the **Pittsburgh** (New York, DC) or the **Riverside** regression (LA, Phoenix) based on similarities in climate and $PM_{2.5}$ composition.

Water activity:

$$a_w(T, RH) = RH \exp\left(\frac{4\sigma_w M_w}{\rho_w RT D_p}\right)^{-1}$$

Hygroscopic growth factor:

$$\mathrm{fRH}(T, RH) = 1 + \kappa_{\mathrm{bulk}} \frac{a_w(T, RH)}{1 - a_w(T, RH)}$$

Additional linear correction:

$$[\text{corrected } \mathrm{PM}_{2.5}] = \theta_1 \left(\frac{[\mathrm{PM}_{2.5} \text{ as reported}]}{\mathrm{fRH}(T,\ RH)}\right) + \theta_0$$

We applied the best available city-specific composition, thus hygroscopicity for each city:

- For Pittsburgh, Riverside, and LA (where we had co-located regulatory-grade monitors in/surround that city), we calculated hygroscopicity and coefficients based on the co-located sensors.
- For other cities, we used the interagency monitoring of protected visual environments (IMPROVE) samples for the monitors closest to each city.

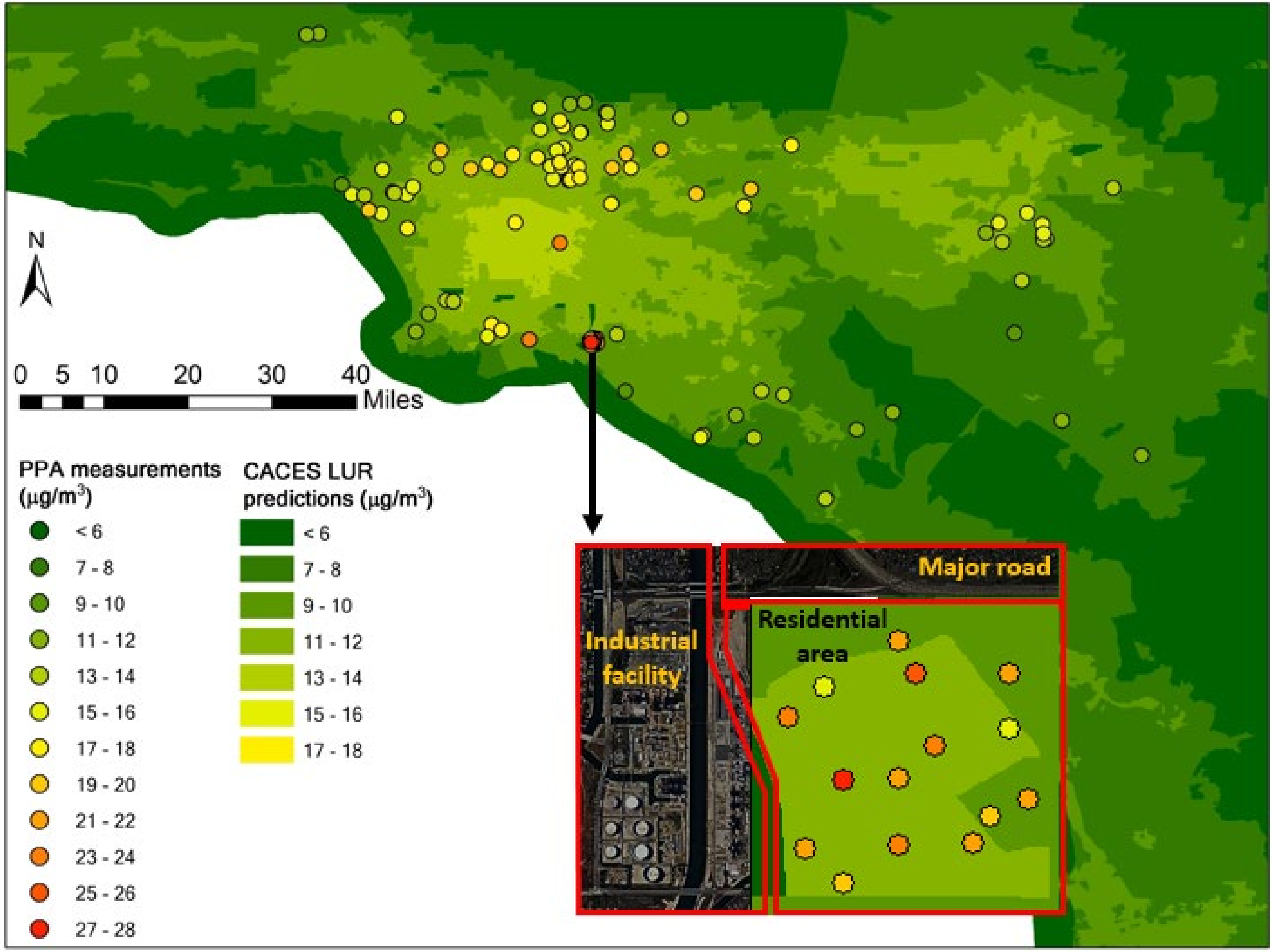
Where σ_w, M_w, ρ_w, T, R, D_p, RH, K_{bulk} denote the surface tension, molecular weight, density of water, absolute temperature, ideal gas constant, particle diameter, ambient relative humidity, and hygroscopicity of bulk aerosol, respectively.

Hygroscopic Growth (HG) Correction Method

CACES LUR Estimates vs. PPA Measurements

<u>Differences</u>

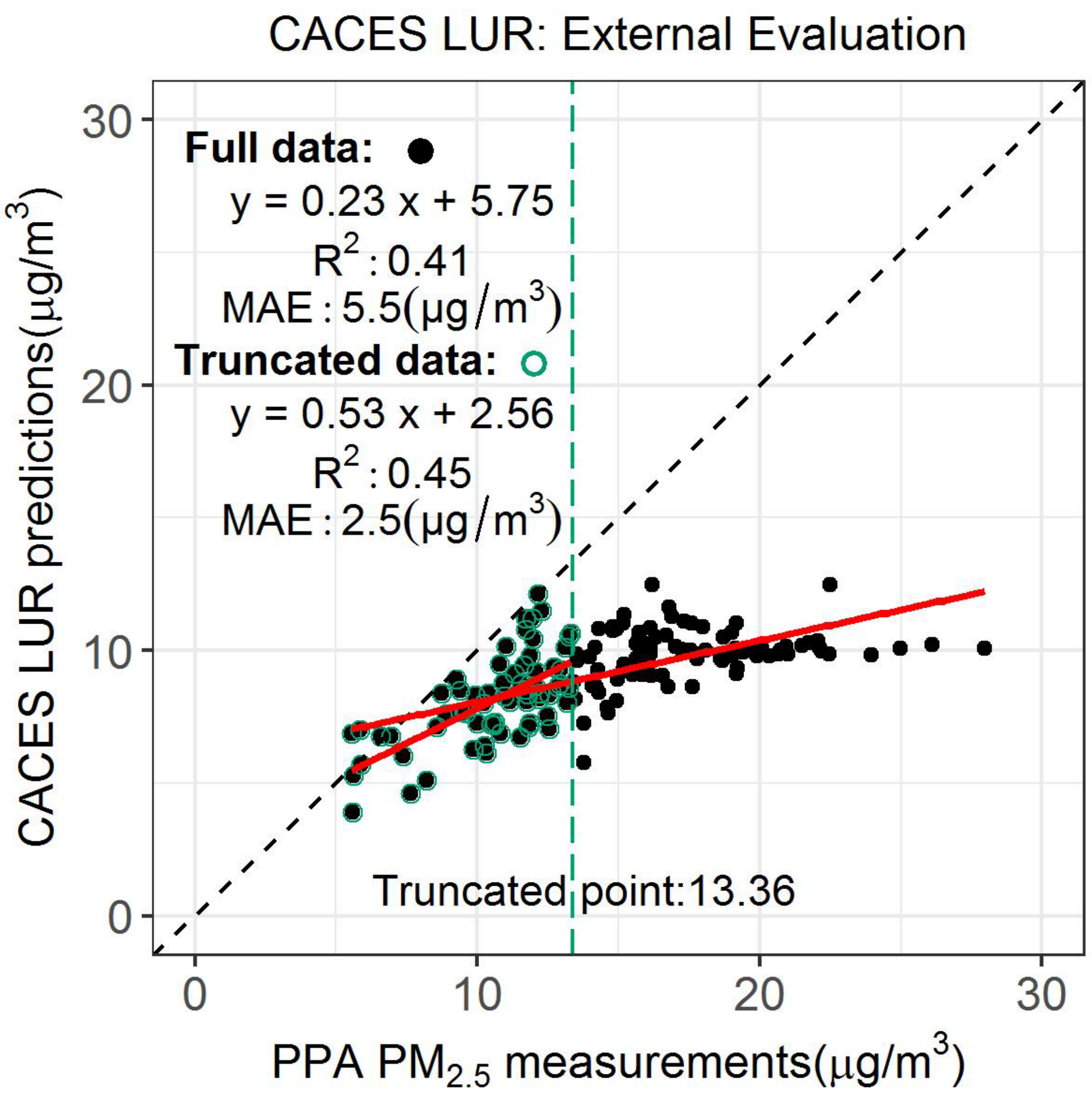
- Spatial **mismatch**.
- Uncaptured "**hotspots**": industrial facilities and highway.



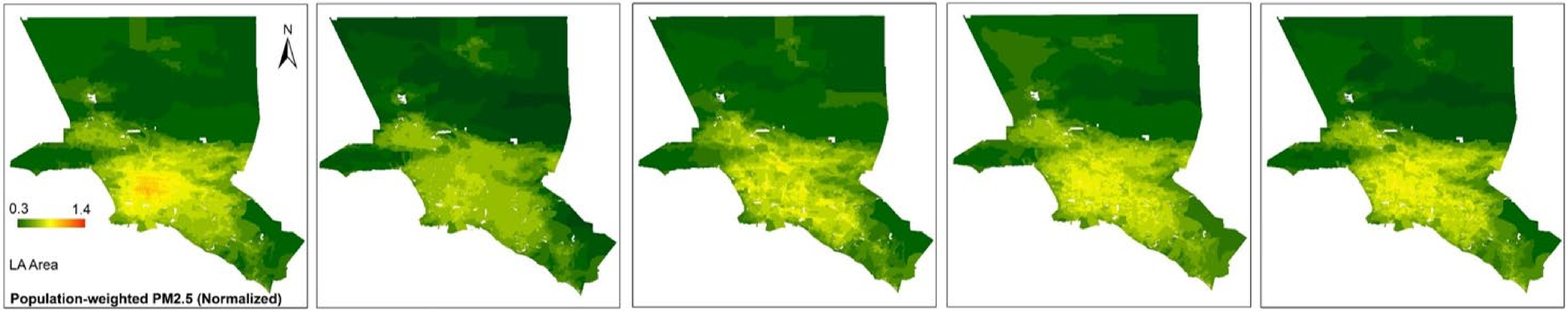
External Evaluation of CACES LUR Estimates

<u>Uncaptured</u>

- Miss “**hotspots**”.



LUR Model Comparison (Normalized Pop-weighted)



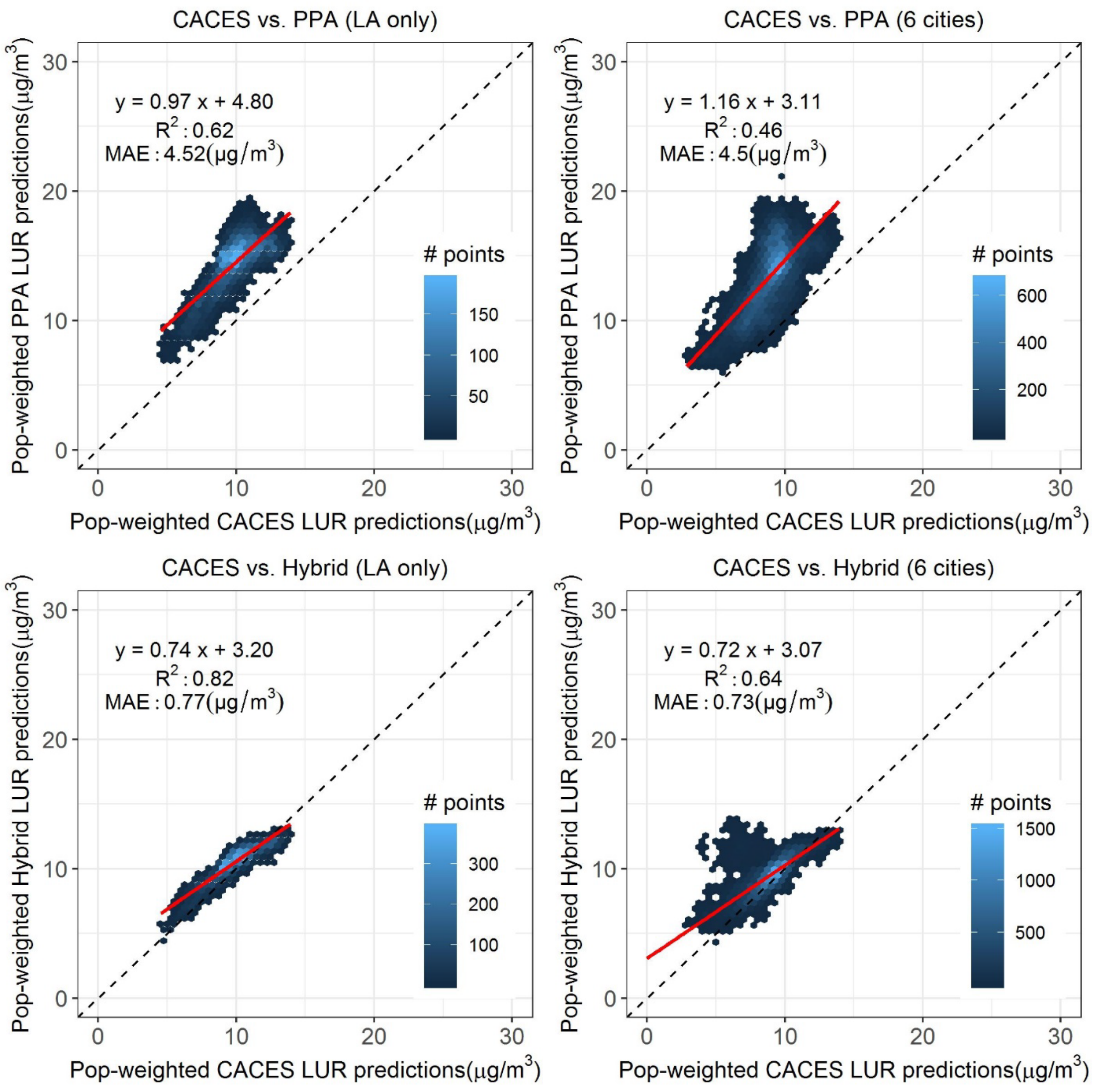
A CACES LUR: National | B CACES LUR: 6-city | C PPA LUR | D Hybrid LUR: National | E Hybrid LUR: 6-city

Hybrid models not only benefit from capturing "hotspots" but are also consistent with the regional spatial trends in the CACES LUR models.

Normalized Population-weighted $PM_{2.5}$ concentration maps

Hybrid LUR: Mitigating Uncertainty

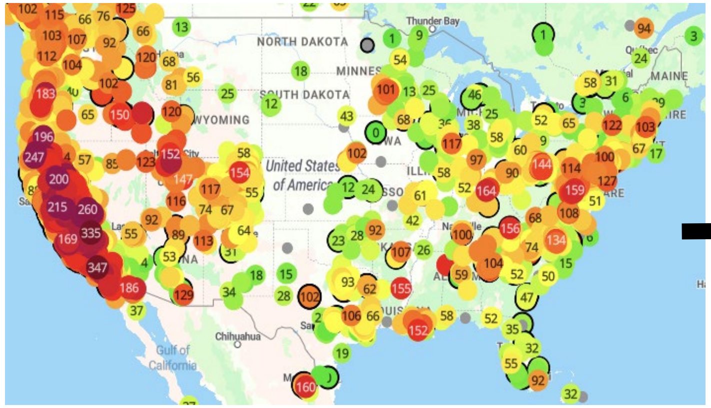
- LUR using only the PPA data may be reasonable; however, **consistently higher** predictions.
- Hybrid models suggest the **value** of combinations.
- Future LUR models: investigating **factors behind** model improvement.



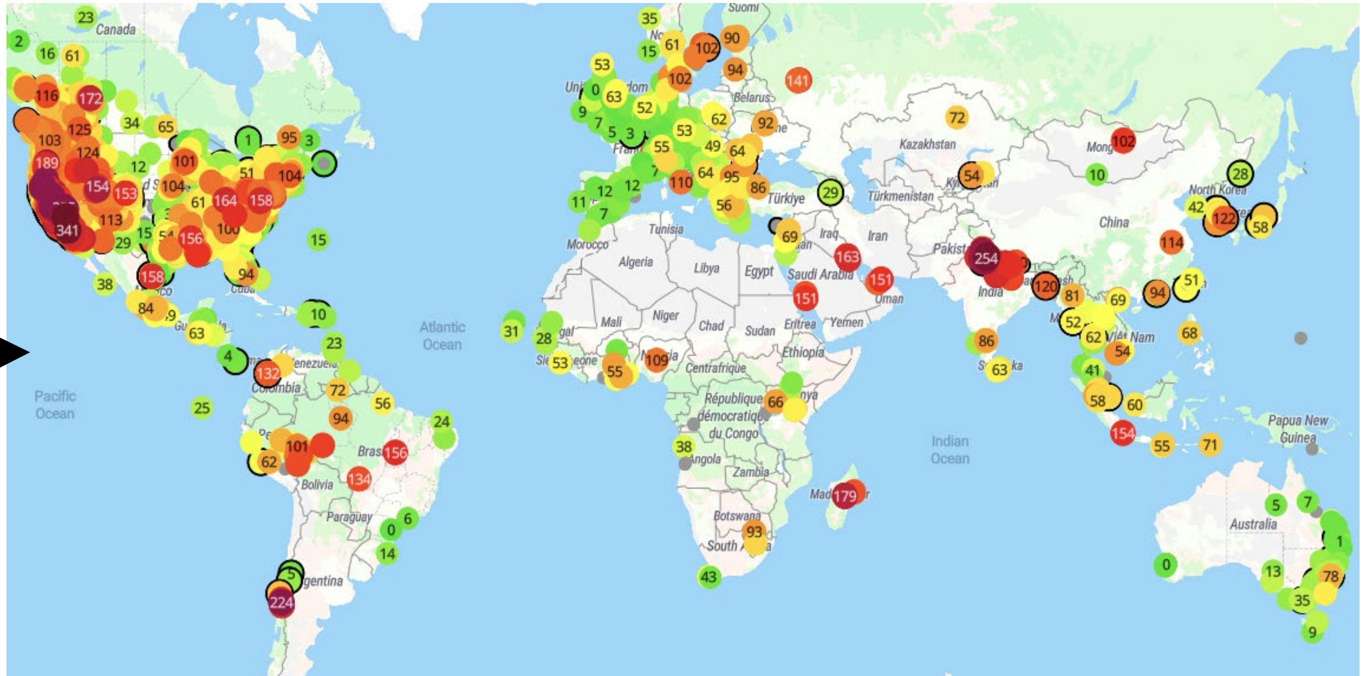
Low-Cost Sensing in Air Quality Models

- **Representative** samples.
- **Fast-growing** network.
- **Rural areas and low- and middle-income countries** (sparse regulatory monitors).

- Neighborhood **planning and design**; clean streets; guidance on **outdoor** activities; interventions



US PurpleAir network



World PurpleAir network