Performance of Correction Models for Accurate $PM_{2.5}$ Estimation from Purple Air Sensors Data Based on Distance

Dinushani Senarathna[1]

senarasd@clarkson.edu

Sumona Mondal [1], Suresh Dhaniyala [1], Shantanu Sur [1], Supraja Gurajala [2] , Vijay Kumar [1]

Clarkson University[1]

SUNY Potsdam University[2]

AAAR-2021 (18-22 October)

Clarkson™
SCHOOL OF ARTS & SCIENCES

AIR LAB
AEROSOL INSTRUMANTATION RESEARCH

Introduction

- **EPA** monitoring sites in the US are limited in number (~1350[1])
- Globally, EPA quality measurements are even less available
- Purple-Air (**PA**) sensor is one popular low-cost sensor in US
- Currently >1500 PA sensors are deployed in the US and >10,000 sensors around the world
- EPA and PA use different technology for $PM_{2.5}$ estimation and the measurements differ
- Can PA sensor data could be utilized to predict EPA quality measurements?

PA sensor

Ardon-Dryer, K., Atmos. Meas. Tech., 2020
https://www2.purpleair.com

Use of low-cost sensors to predict AQ with high accuracy

- Low –cost sensors data can be noisy and low in accuracy[1][2]
- PA sensor data are sensitive to air properties including temperature and relative humidity[1]
- Models built to better estimate $PM_{2.5}$ form PA sensor data use EPA measurements as gold-standard[1][2]
- Relative humidity (RH) measured by PA sensors is used to build correction models applicable US wide [2]
- Maximum distance between EPA and PA in these models are arbitrarily set as 10 km[1]

1.Ardon-Dryer, K. (2020), Measurements of $PM_{2.5}$ with Purple Air under atmospheric conditions, Atmos. Meas. Tech., 13, 5441–5458, 2020, https://doi.org/10.5194/amt-13-5441-2020
2.Barkjohn, K, K (2020), Development and Application of a United States wide correction for $PM_{2.5}$ data collected with the Purple Air sensor, https://doi.org/10.5194/amt-2020-413

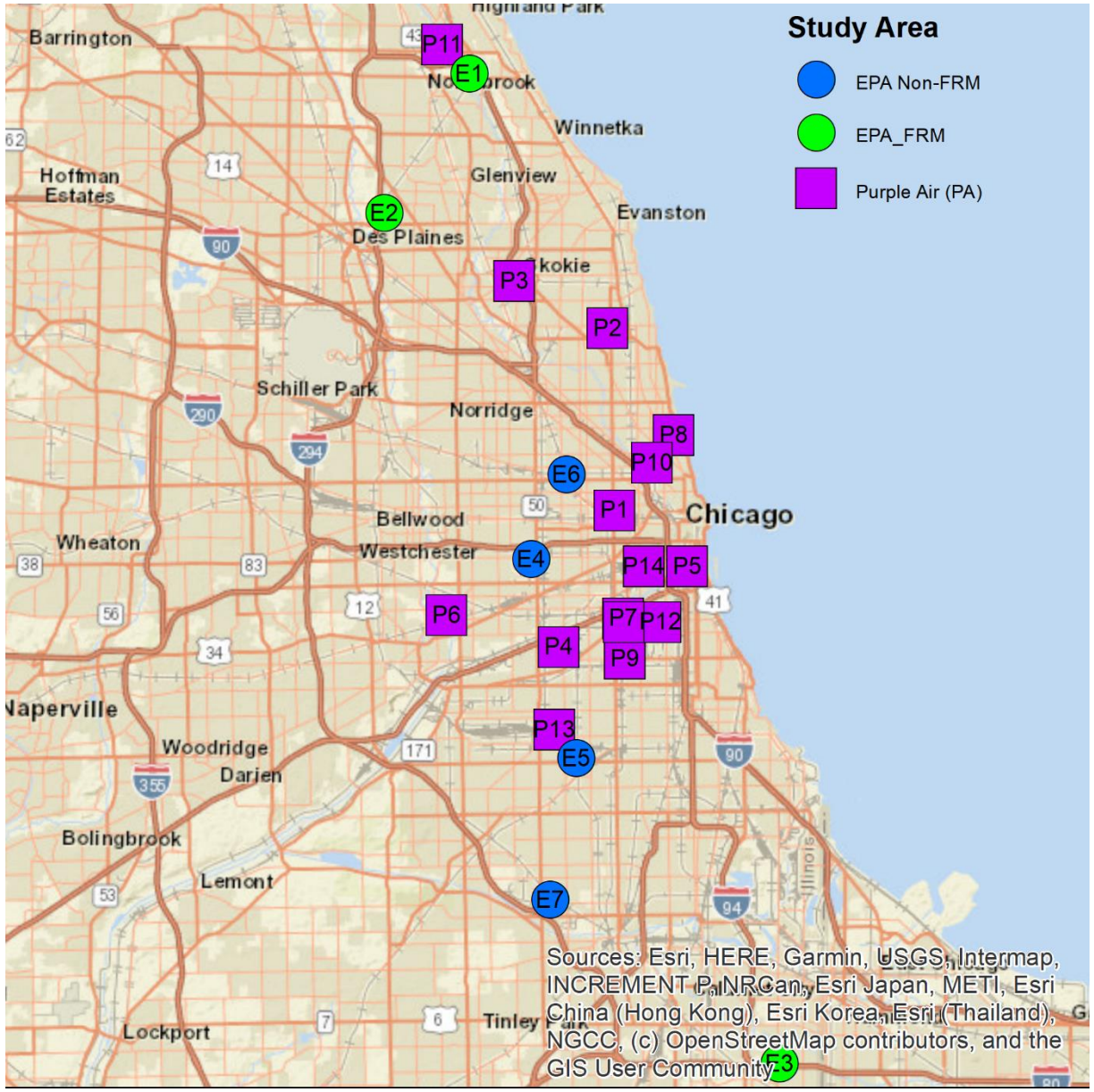
Research objectives

- Investigate how the $PM_{2.5}$ prediction model built on US-wide data perform in a local network of PA sensors [1]
- Study how the distance between PA sensor and prediction location impacts the model accuracy
- Determine if models built on data from multiple PA sensors achieve higher prediction accuracy

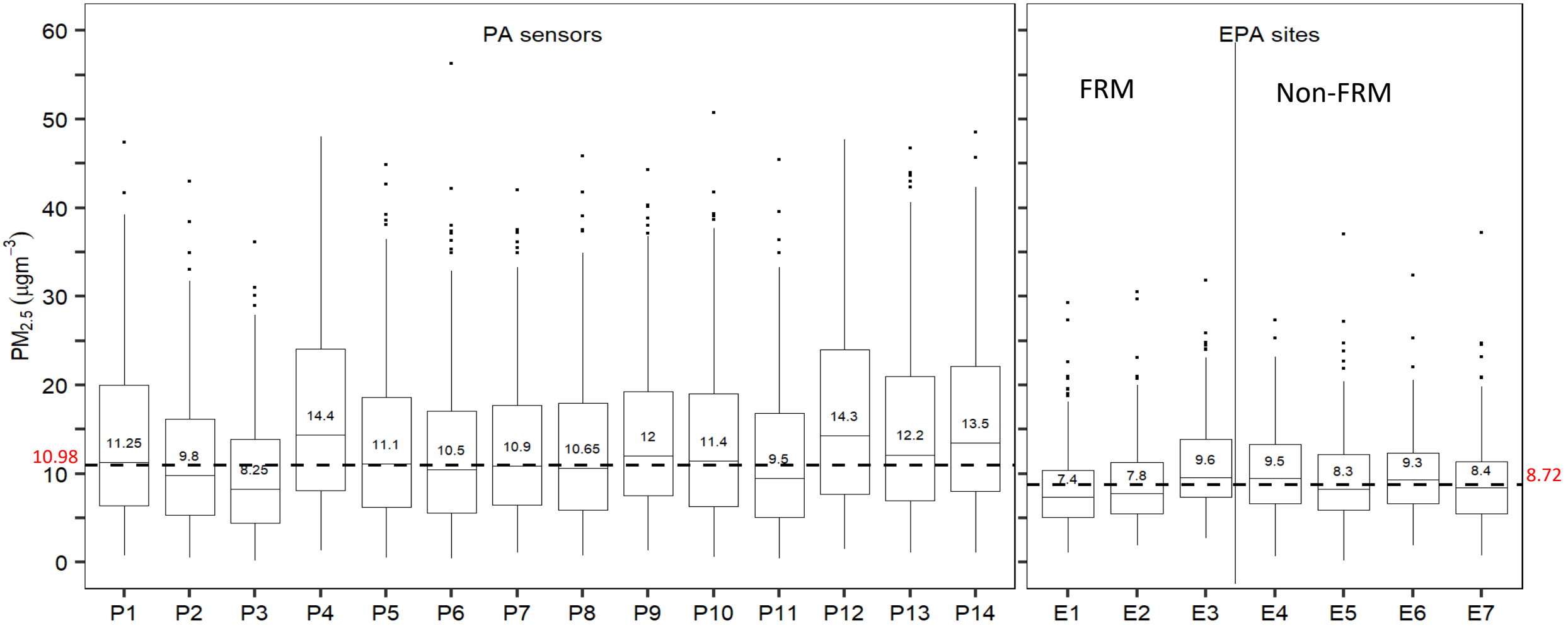
Barkjohn, K, K (2020), Development and Application of a United States wide correction for $PM_{2.5}$ data collected with the Purple Air sensor

Methodology

- Data source: Cook county, Illinois, USA; 2019 August to 2020 July
- A total of 15 PA sensors and 7 EPA sites considered
- Considered and compared the Federal Reference Methods/ Federal Equivalent Methods data (abbreviated as FRM here) and non-FRM data from EPA on model prediction
- Prediction models were built using PA-measured temperature (T) and relative humidity (RH), considered separately and together
- Prediction accuracy of models built using single PA sensor and multiple PA sensor data were compared



$PM_{2.5}$ measurements by EPA and PA

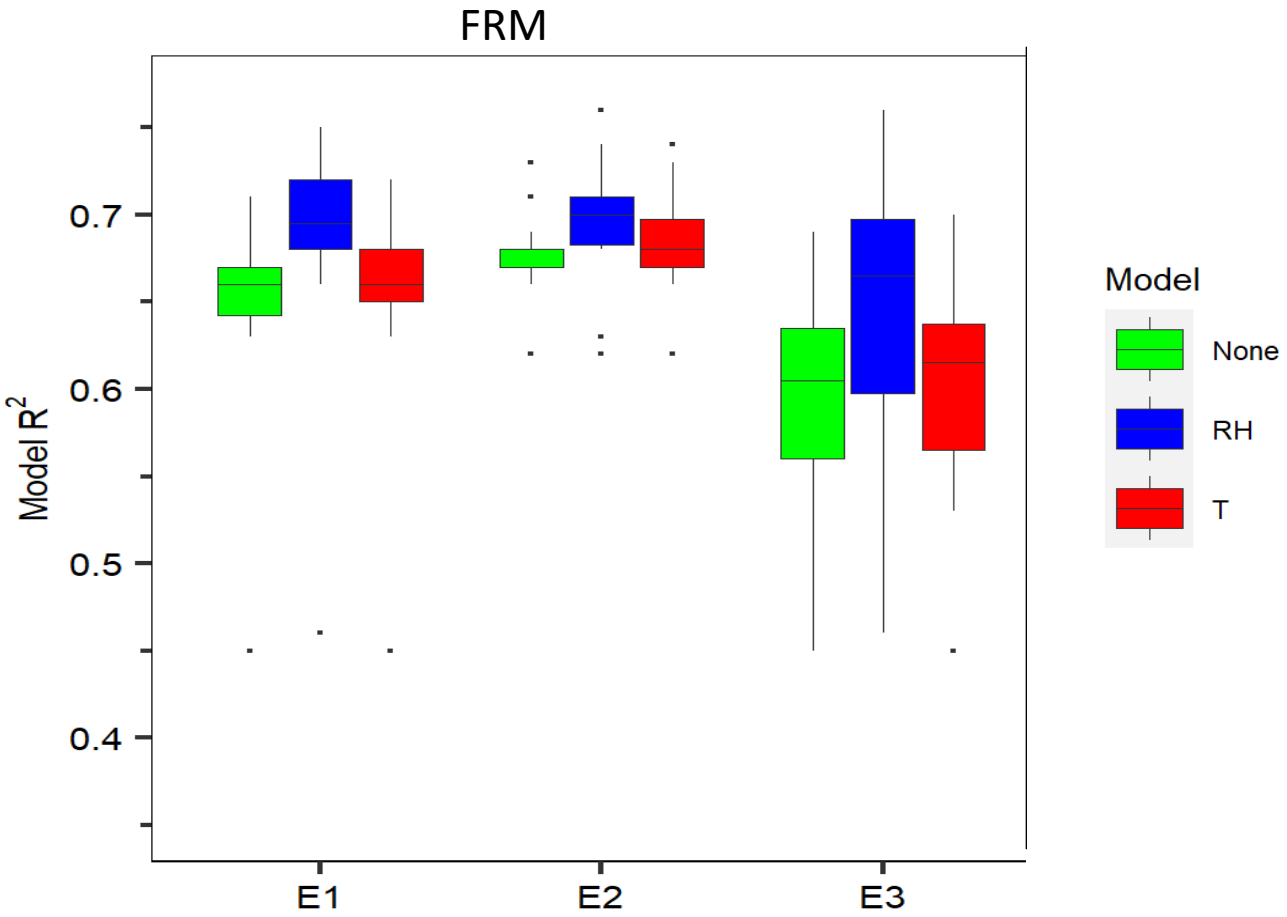


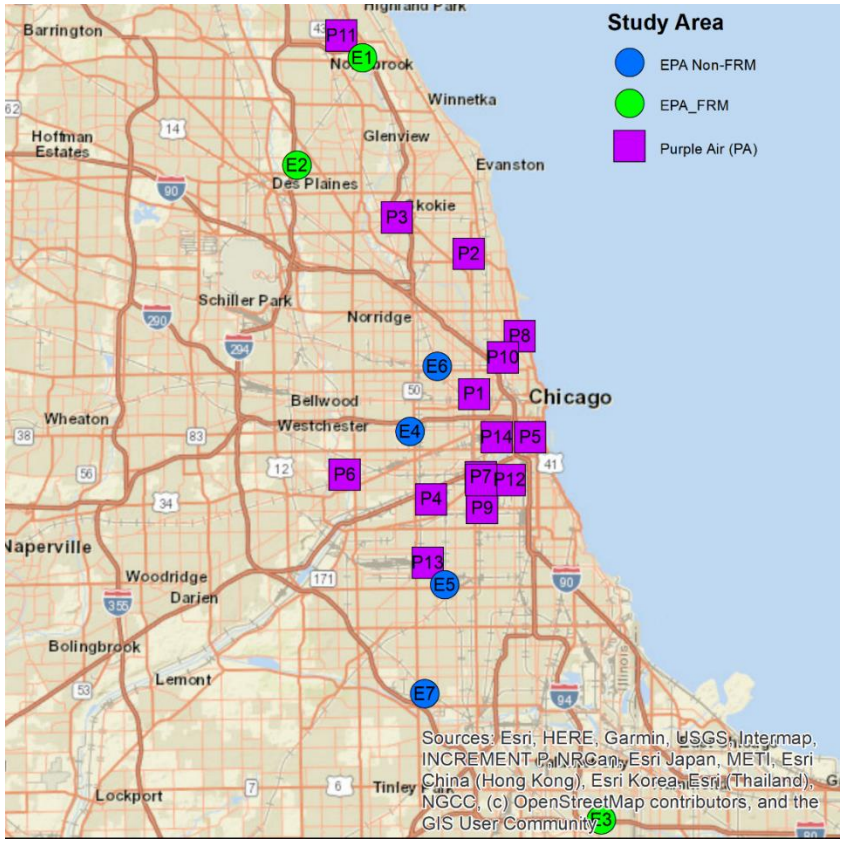
$PM_{2.5}$ estimate by PA has broader distribution and ~20% higher value than EPA

Clarkson
SCHOOL OF ARTS & SCIENCES

Effects of temperature and relative humidity on model accuracy

$$PM_{2.5\,(EPA)} = \beta_0 + \beta_1 PM_{2.5\,(PA)} + \beta_2 RH_{\,(PA)} + \beta_3 T_{\,(PA)}$$

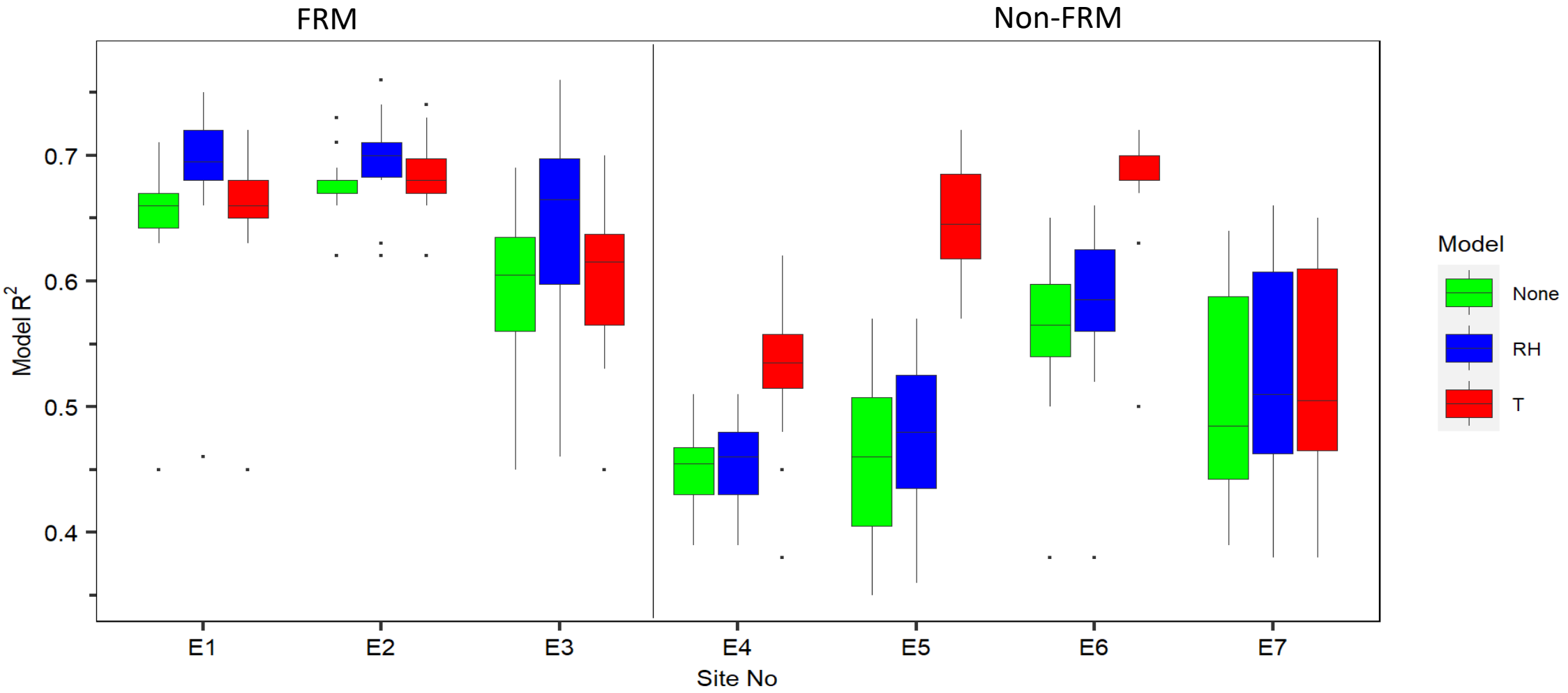




Relative humidity has a stronger influence in models using FRM data

Comparison between models from FRM and non-FRM data

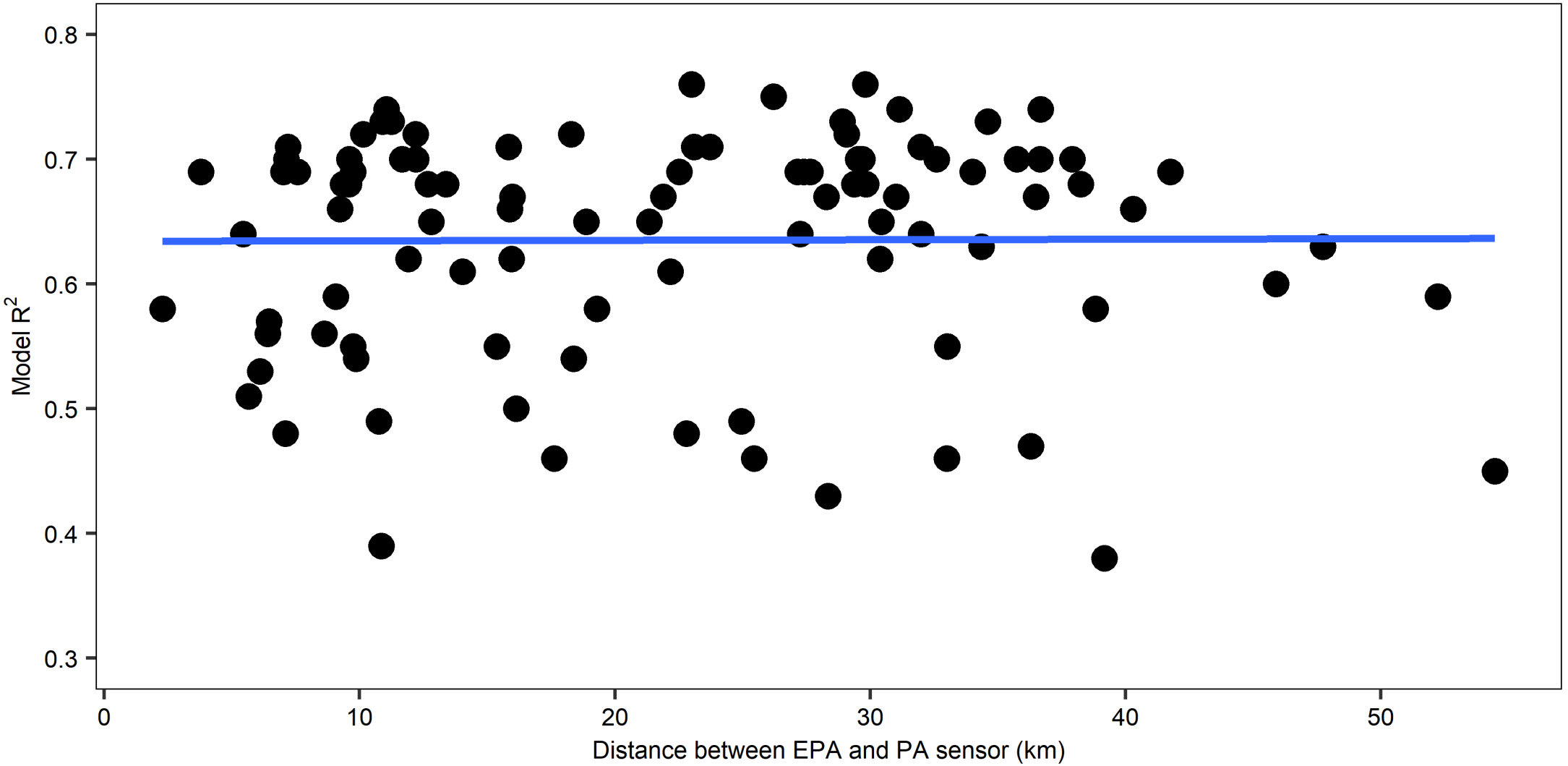
$$PM_{2.5\,(EPA)} = \beta_0 + \beta_1 PM_{2.5\,(PA)} + \beta_2 RH_{(PA)} + \beta_3 T_{(PA)}$$



Temperature has a stronger influence in models using non-FRM data

Effect of distance on model prediction accuracy

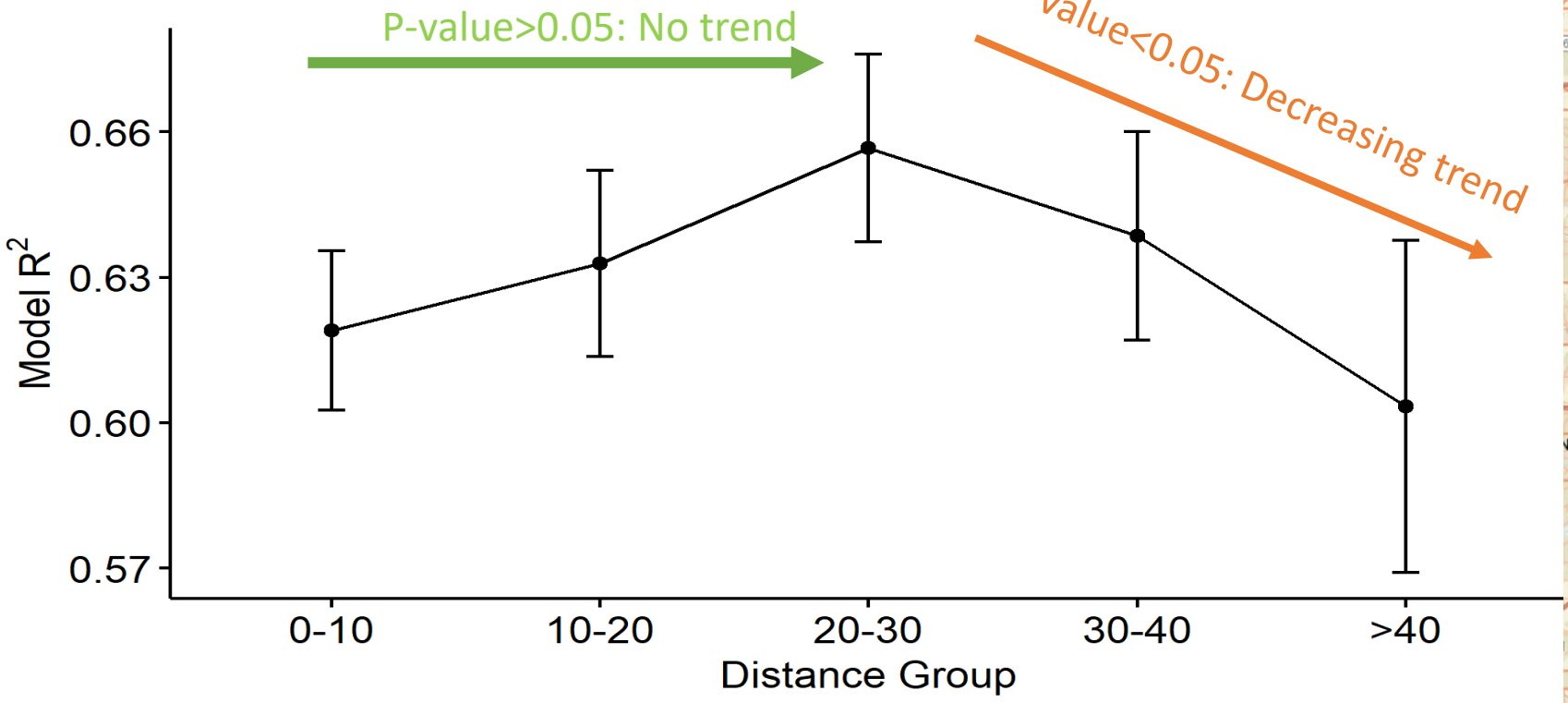
$$PM_{2.5\,(EPA)} = \beta_0 + \beta_1 PM_{2.5\,(PA)} + \beta_2 RH_{\,(PA)} + \beta_3 T_{\,(PA)}$$

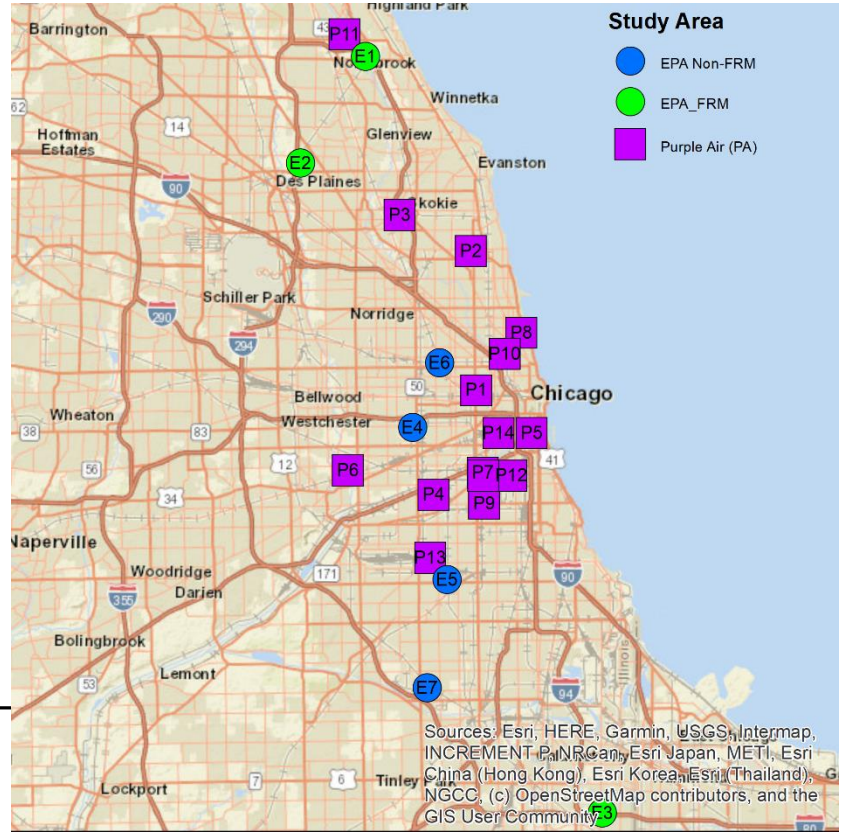


Distances between EPA and PA sites range between 2 - 55 km

Changes of model accuracy across various distance groups

- Jonckheere-Terpstra test was conducted to analyze the trend of model R^2 between PA sensors in five distance groups

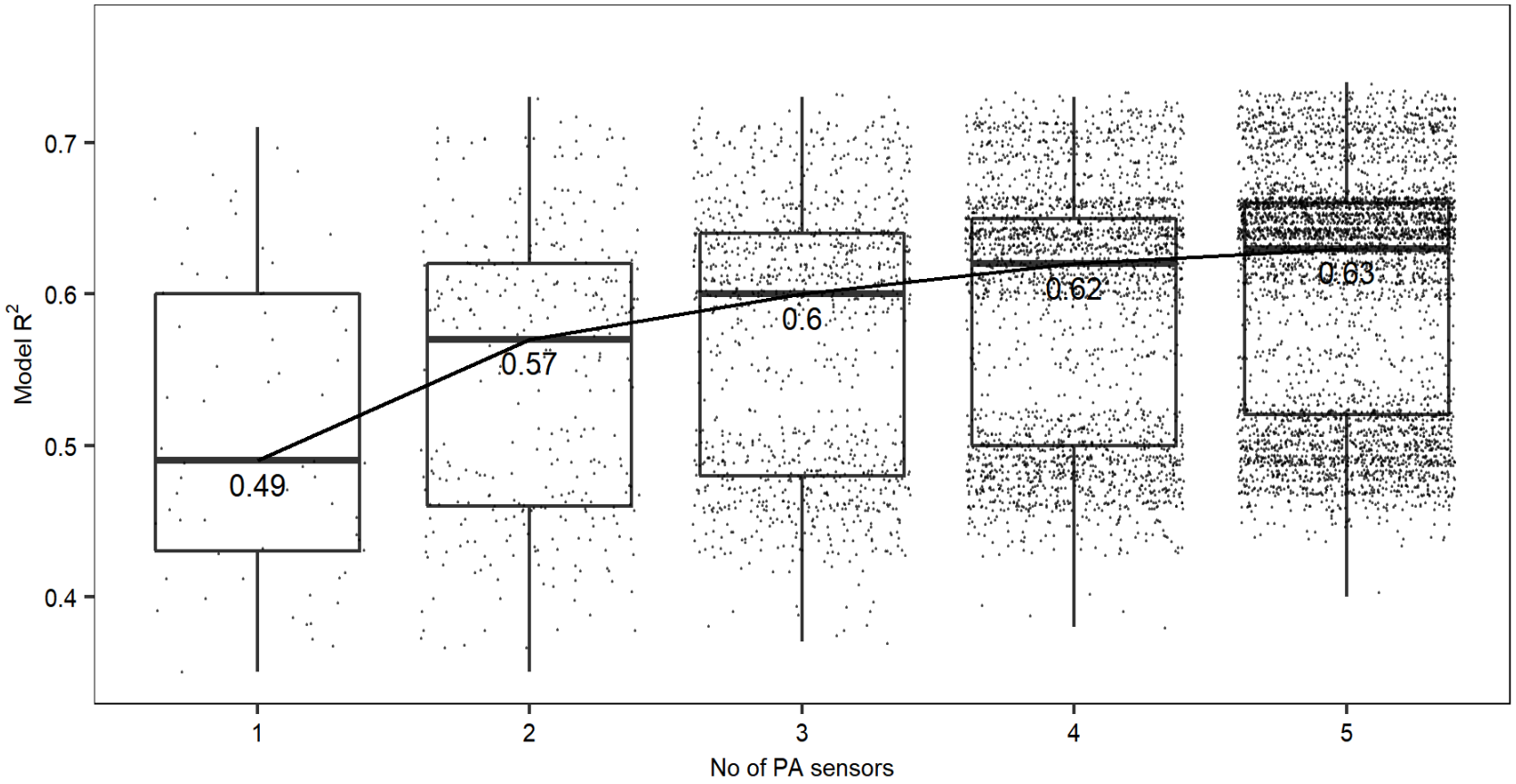


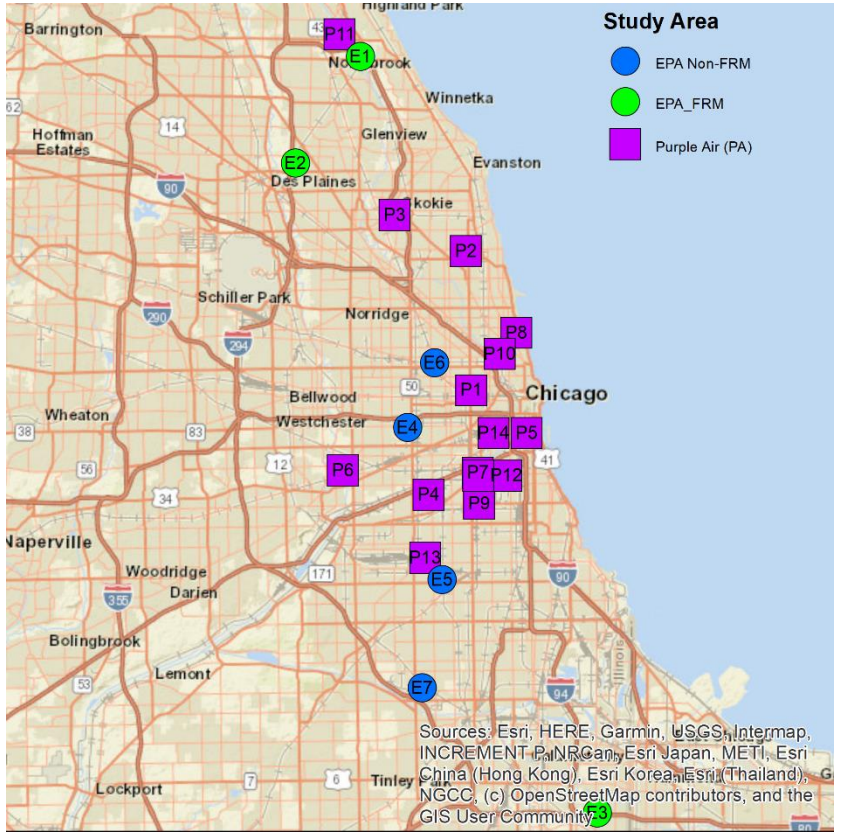


Model accuracy decreased for groups when distance is >30 km

Correction models using multiple PA sensors

$$PM_{2.5\,(EPA)} = \beta_0 + \beta_1 T_{(PA)} + \beta_2 RH_{(PA)} + \sum_{i=3}^{7} \beta_i PM_{2.5\,(PAj)} \text{ where } j=1,..,5$$





- Any PA sensor within 30km can be used for better model performance
- Models built using two to three PA sensors result in optimum model R^2

Conclusion and future work

- Correction models, built on PA sensors' data considering temperature provides higher accuracy using non FRM/FEM EPA data while relative humidity provide better prediction for FRM/FEM EPA data since data distribution is same for two types of EPA
- Model R^2 value decreases significantly when the distance between EPA and PA sensors are > 30 km
- Models using multiple PA sensors performed better than using a single PA sensor, however improvement was minimal for more than two/three PA sensors
- Consideration of additional parameters, wind speed, wind direction may help to obtain higher accuracy for the models

Reference

- Ardon-Dryer, K. (2020), Measurements of $PM_{2.5}$ with Purple Air under atmospheric conditions, Atmos. Meas. Tech., 13, 5441–5458, 2020, https://doi.org/10.5194/amt-13-5441-2020
- Barkjohn, K, K (2020), Development and Application of a United States wide correction for $PM_{2.5}$ data collected with the Purple Air sensor, https://doi.org/10.5194/amt-2020-413
- Feenstra, B.(2019), Performance evaluation of twelve low-cost $PM_{2.5}$ sensors at an ambient air monitoring site, Atmospheric Environment, https://doi.org/10.1016/j.atmosenv.2019.116946
- Datta, A. (2020),Statistical field calibration of a low-cost $PM_{2.5}$ monitoring network in Baltimore, Atmospheric Environment, https://doi.org/10.1016/j.atmosenv.2020.117761
- Liu, H, Y. (2019), Performance Assessment of a Low-Cost $PM_{2.5}$ Sensorfor a near Four-Month Period in Oslo, Norway, http://dx.doi.org/10.3390/atmos10020041
- Kelleher, S (2018), A low-cost particulate matter ($PM_{2.5}$ monitor for wildland fire smoke, Atmos. Meas. Tech., 11, 1087–1097, 2018, https://doi.org/10.5194/amt-11-1087-2018
- Jayaratne, R (2018), The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog, https://doi.org/10.5194/amt-11-4883-2018
- Hopke P. K,(2018), Hourly land-use regression models based on low-cost PM monitor data, Environmental Research 167(2018) 7-14, https://doi.org/10.1016/j.envres.2018.06.052

Thank you