Contribution of Low-Cost Sensor Measurements to the Prediction of $PM_{2.5}$ Levels

Yang Liu

Co-investigators: Jianzhao Bi, Jennifer Stowell, Howard Chang, Avani Wildani, Edmund Seto, Paul English, Frank Freedman, Mohammad Al-Hamdan, Patrick Kinney

HAQAST

ROLLINS SCHOOL OF PUBLIC HEALTH

EMORY

Outline

- $PM_{2.5}$ exposure modeling framework to evaluate low-cost sensor data
- Regulatory and low-cost $PM_{2.5}$ monitoring
- Results of pilot and ongoing study in CA
- Summary

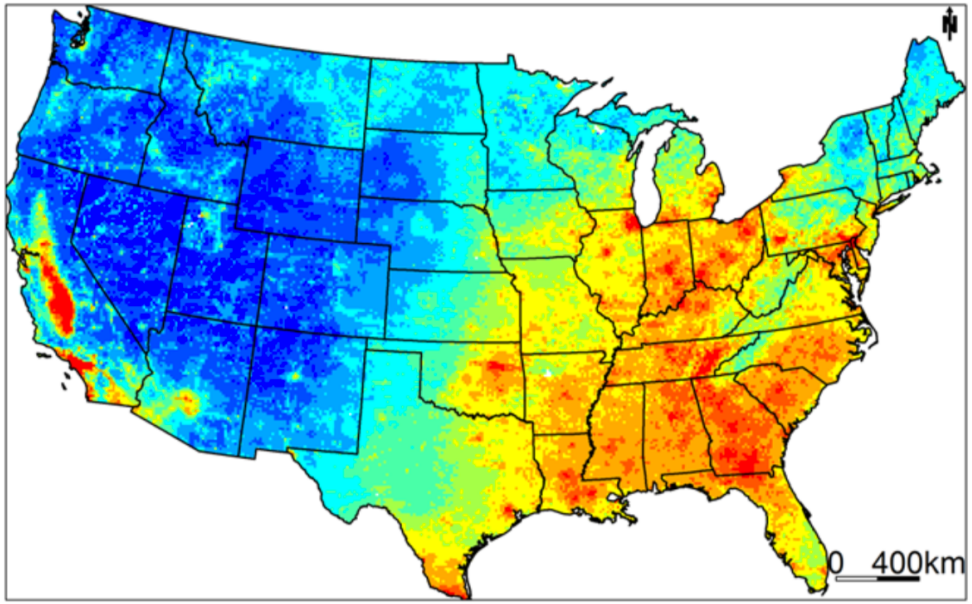
Satellite-driven spatial $PM_{2.5}$ modeling

- Empirical models built on predictors/indicators of $PM_{2.5}$
  - Satellite data：AOD，NDVI，elevation, etc.
  - Met fields：weather station data, gridded reanalysis
  - Land cover, population, traffic, etc.
- Generate continuous $PM_{2.5}$ fields in space and time
- High spatial (1 km) and temporal (hourly to daily) resolution
- High accuracy, have been increasingly used in both acute and chronic health effects research

- **Model development / performance depends on quality and distribution of ground $PM_{2.5}$ measurements**

Model Structures

- Statistical models
  - Multivariate regression
  - Mixed-Effects Models
  - GAM, GWR, hybrid
  - …
- Machine learning models
  - Random Forests
  - Gradient boosting
  - Neural network
  - …



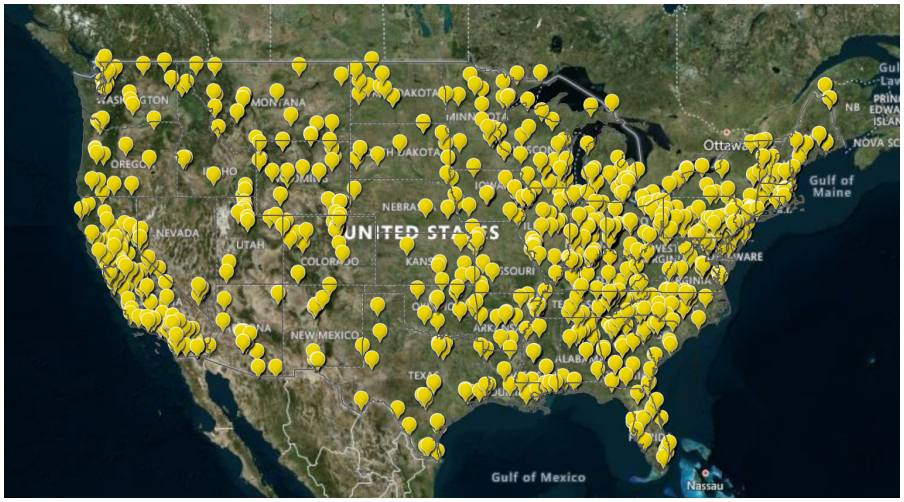
Predicted $PM_{2.5}$ annual mean concentrations in the US (based on a daily random forest model) (Hu et al., 2017)

$PM_{2.5}$ Regulatory Monitoring in the US - AQS

HAQAST

ROLLINS SCHOOL OF PUBLIC HEALTH
EMORY

- ~1,000 monitors operated by regulatory agencies in ~3000 counties. Distribution follows population. Most rural and many suburban regions have no coverage.
- Insufficient spatial coverage and sampling frequency for extensive model training (1-in-3 days or 1-in-6 days, a subset of continuous monitors)
- Often inadequate to address EJ issues



Low-cost $PM_{2.5}$ sensors

- Mostly commercially available optical sensors converting scattering to mass concentration
- Low instrument cost (<$2,500) , almost no maintenance cost (i.e., no support from vendor)
- High sampling frequency, portable, easy to install
- Much more rapid expansion than gold-standard monitoring networks
- **High and unknown measurement errors**



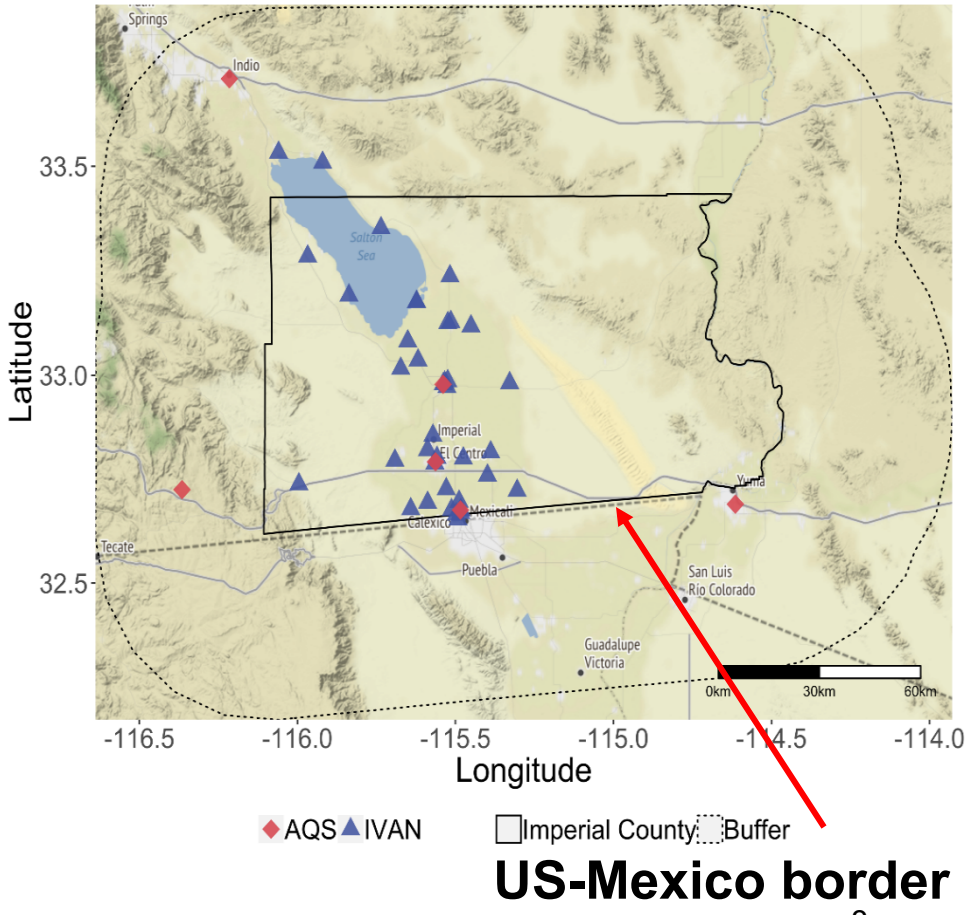

Study Questions

- Are low-cost $PM_{2.5}$ sensor data valuable in AQ and PH management and research?
  - Does it provide new information?

- If so, how do we use them?
  - Can we combine them with satellite data/gold-standard measurements/CTM results?

TT1: Pilot study in Imperial County, CA

- Identifying Violations Affecting Neighborhoods (IVAN)
- 6 AQS, 39 modified Dylos monitors
- Modeling period: 09/16 – 11/17
- Modeling domain: 41,000 km^2, 1x1 km grid cell



Model Development

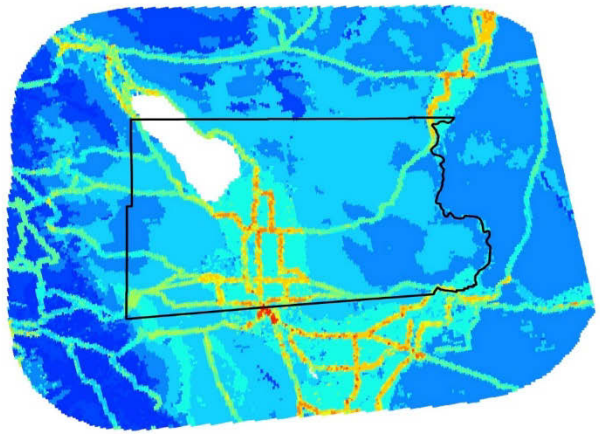
- Random forest model structure
- Daily $PM_{2.5}$ concentrations
  - AQS
  - RH-calibrated Dylos monitors from IVAN - no further correction attempted
- Model configurations
  - AQS only
  - AQS + IVAN

Model Prediction

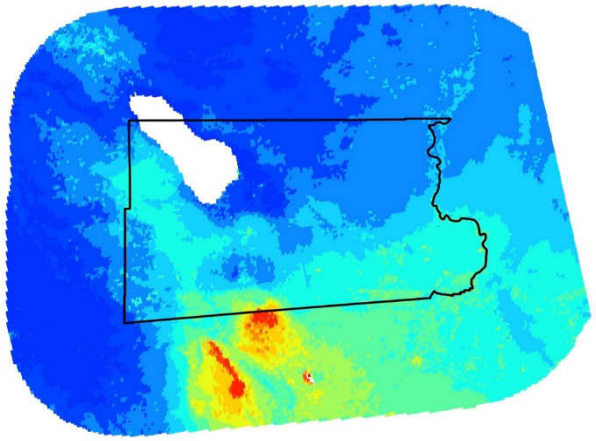
HAQAST

ROLLINS SCHOOL OF PUBLIC HEALTH EMORY

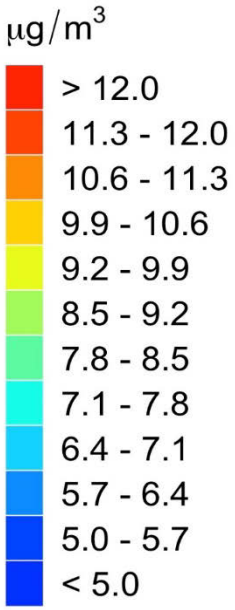
Model	N	Random CV R^2	Spatial CV R^2	Temporal CV R^2
AQS	1,617	0.53	0.24	0.55
AQS/IVAN	12,902	0.73	0.63	0.70



AQS-Only Model



AQS/IVAN Model



Observations

- Introducing low-cost sensor data can significantly change the predicted $PM_{2.5}$ distribution, and it appears to provide us with additional information
- Low-cost sensor data bring noise into model predictions, and formal treatment of measurement errors is necessary for quantitative characterization of $PM_{2.5}$

Year 3 Progress Update, PI Yang Liu

HAQAST

ROLLINS SCHOOL OF PUBLIC HEALTH
EMORY

Core Project 1: Effect of Wildfire Smoke on ED Visits in Colorado

- Region-specific C-R functions of various health endpoints during the fire season
- Epidemiologic analysis complete.
- 1 manuscript submitted

Core Project 2: Linking Air Pollution Exposure with CDC's Nationwide Asthma ED Database

- A national scale epidemiological study to link county-level ED visits with air pollution exposure.
- 1 manuscript accepted with CDC co-authors
- 1 manuscript accepted on national PM2.5 speciation model, 1 manuscript accepted on national Bayesian PM2.5 model
- 1 manuscript submitted on national machine learning PM2.5 model

Tiger Team Participation

TT1 led by Duncan and West: Efficacy of Environmental Regulations in the Eastern U.S

- 1 paper led by Columbia accepted using Emory PM2.5 estimates.
- 1 paper led by Columbia in preparation

TT1 led by Kinney and Freedman: High Resolution Assessment of Community Health Impacts of Air Pollution

- Completed 100 m resolution modeling in New York City. 1 manuscript submitted.
- Completed the Southern California PM2.5 modeling. 1 group manuscript submitted. 1 manuscript in preparation
- Completed the IVAN PM2.5 modeling. 1 manuscript submitted.
- Completed CA statewide low-cost sensor exposure modeling. 1 manuscript in preparation.

TT2 led by Anenberg: Global indicators

- Delivered global wildfire exposure indicators to Lancet Countdown . 1group manuscript submitted.

6 academic talks

4 papers published, 4 papers submitted, 3 in prep