Machine learning for spatiotemporal $PM_{2.5}$ estimates across the western US, 2008-2014

Melissa May Maestas, Ph.D.
Research Associate (Post-Doctoral)

Why Wildfires?

- Globally and regionally, wildfire risk is projected to increase under various potential future climate scenarios.
- The percent of our air pollution due to wildfires will likely increase, not just from climatic changes, but also because of declines in other sources of air pollution

Image: National Interagency Fire Center, Fox Fire East, Glacier 2000

Exposure assessment difficulties

- Sparse air pollutant monitoring network
- Many $PM_{2.5}$ monitors only measure every sixth or third day
- Leads to spatial and temporal averaging of exposure measurements
- But, smoke plumes migrate quickly, changing exposures over smaller spatial and temporal scales

Previous Machine Learning (ML) work

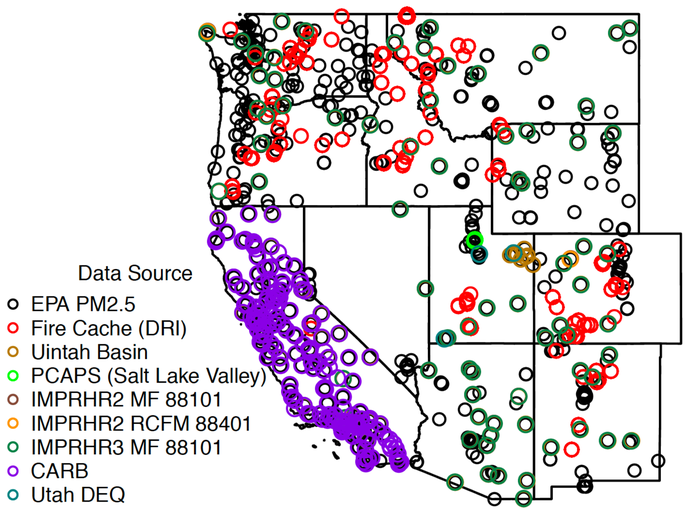
- Dr. Colleen E. Reid modeled Northern CA 2008 wildfire season for her dissertation
- Lightning storm June 20-21, 2008 started fires
- 10-12 million people exposed to smoke
- Reid et al., Spatiotemporal Prediction of Fine Particulate Matter During the 2008 Northern California Wildfires Using Machine Learning. Environ. Sci. Technol. 2015, 49, 3887−3896

Current Project – Scaling up:

- Multi-year, 11 western states
- Similar modeling (e.g., Di et al., 2016) don't perform well in western US and fires were left out
- More than EPA monitors (Forest Service, TEOM, states, field campaigns, etc.)
- Plan to do ensemble of different machine learning algorithms

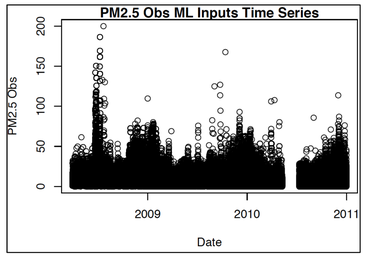
Di et al. Environ. Sci. Technol. 2016, 50, 4712−4721

All PM2.5 Observation Locations

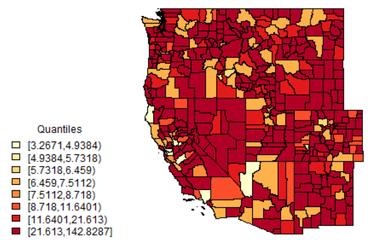


Project Status

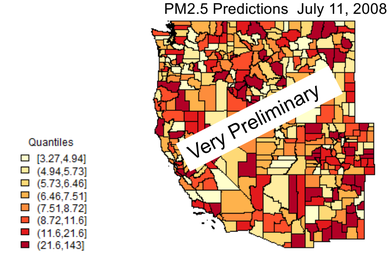
- Assembling data
- Starting to run ML models



PM2.5 Observations July 11, 2008

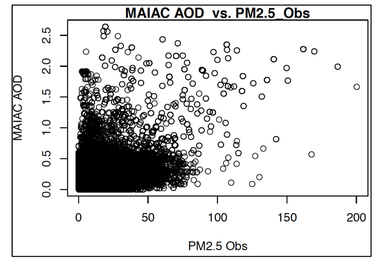


PM2.5 Predictions July 11, 2008



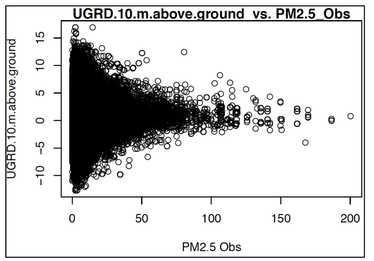
ML Predictors

Variables	Data Source	Temporal Resolution	Spatial Resolution	Buffer Size
Dependent Variable				
PM2.5 from monitoring stations	US EPA, states Federal Land Manager Environmental Database, Fire Cache Smoke Monitor Archive, IMPROVE Network, academic research groups	Daily or hourly	point	
Spatiotemporal Variables				
GOES Aerosol and Smoke Product (GASP) AOD	NOAA	Hourly	4 km	
Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD	NASA	Daily	1 km	
MODIS Active Fire Detection	NASA	Daily	1 km	
VIIRS Fire Occurence	NASA	Daily	375 m	
Hazard Mapping System (HMS) Smoke and Fire Product	NOAA	Daily	4km	25km, 50, 100km, 500km, 1000km, 2000km
MODIS Normalized Difference Vegetation Index (NDVI)	NASA	Monthly	1 km	
14 Meteorological Variables	NOAA: North American Mesoscale (NAM) Forecast System	6-Hourly	12 km	
Spatial Variables				
Elevation (m)	USGS	Nominal 2-month cycle	1 arc-second	
Percentages of land cover types	National Land Cover Database 2011	Every 5 years	30 m	1km
Kilometers of highway within buffer zones	National Highways Planning Network, US DOT			100, 250, 500, and 1000 m
Temporal Variables				
Julian Date		Daily		
Weekend		Daily		



Methods: Adapt land use regression modeling with machine learning

- Include novel spatiotemporal datasets
- Apply machine learning methods to
- Select from a long list of predictor variables
- Select from a variety of statistical algorithms
- 10-fold cross validation
  - May do spatial/temporal
- Minimize CV-RMSE
- Caret package in R



Next Steps

- Finish data assimilation
- ML modeling
  - Ensemble modeling
    - Random Forest, others
  - Cross validation by region/time instead of 10-fold
- Model by Region:
  - Predictor relationships could be different
- Pull in CTM model output
- Connect ML estimates to medical data for epidemiological analysis
- Flexible Geographic Resolution
  - County centroid
  - Population-weighted county centroid
  - ZIP code
  - Geo-coded address

Acknowledgements

Earth Lab Research Funding and Support: University of Colorado Boulder's Grand Challenge Initiative
Contributors: This poster represents the work of the Project Health team at Earth Lab: Colleen E Reid, PhD; Gina Li, MA-candidate; Ellen Considine, BS-candidate

Data Contributors/Sources: https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily; http://views.cira.colostate.edu/fed/DataWizard/Default.aspx; https://wrcc.dri.edu/cgi-bin/smoke.pl; https://www.arb.ca.gov/aqmis2/aqmis2.php; Dr. Seth Lyman at Utah State University; Dr. Geoff Silcox at University of Utah; http://www.airmonitoring.utah.gov/dataarchive/archpm25.htm; MODIS; GOES-West AOD; MAIAC; Landsat NLCD; USGS; NOAA;