

# Update: background concentrations lookup tool

Ranil Dhammapala

ranil.dhammapala@ecy.wa.gov



Sara Strachan

Sara.Strachan@deq.idaho.gov



Jill Schulte

jill.schulte@ecy.wa.gov



# Updating the bkg lookup tool

- Decided to use Jul 2014- June 2017 data
- Got all criteria pollutant data (thanks Sara, Bruce & Jill)
- Included IMPROVE sites for PM<sub>2.5</sub> and PM<sub>10</sub>
- No temporary monitors
- Worked through data quirks
  - Omitted sites with < 1 yr of data
  - No seasonal sites that didn't run all 3 yrs
  - Removed SO<sub>2</sub> monitors very close to large point sources (Pocatello, ID, Ferndale, WA & Trail, BC)
  - Combined data from sites that moved short distances

The screenshot shows the NW AIRQUEST website interface. At the top, there is a navigation bar with links for 'A-Z Index', 'Statewide', 'myWSU', 'WSU Home', and a search box. The main header identifies the site as the Northwest International Air Quality Environmental Science and Technology Consortium. The page title is 'Lookup 2009-2011 design values of criteria pollutants'. A blue box on the right contains links for 'Overview', 'Methodology', and 'FAQs'. The main content area provides instructions for single point queries, including input fields for latitude/longitude or UTM coordinates, UTM zone, and a 'Submit coordinates' button. Below this, there are instructions for multiple point queries, including a 'Choose File' button and a 'Submit file' button. At the bottom, a map of Washington and Idaho shows the locations of various air quality sites and reserves.



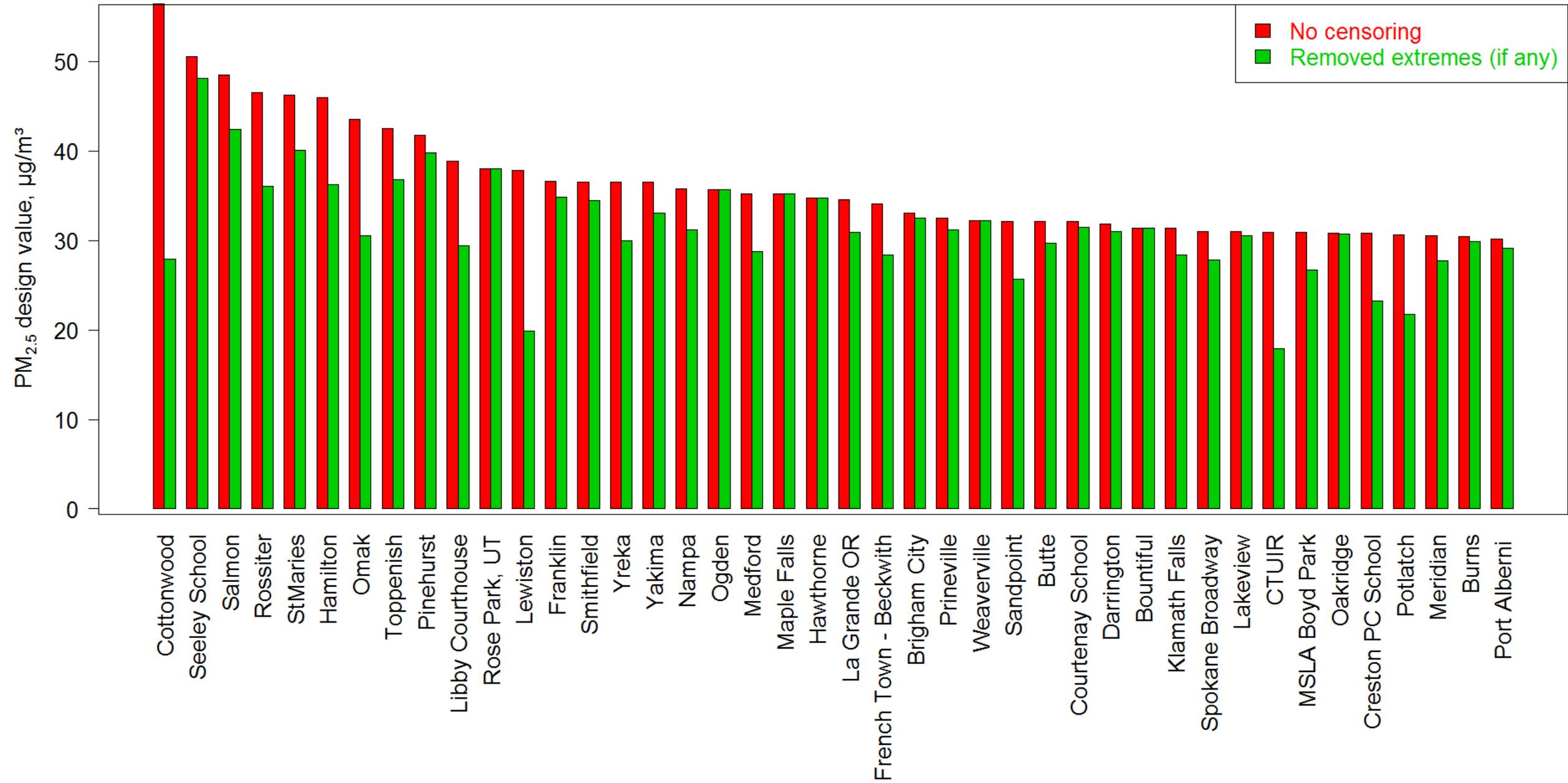
# Extremophiles



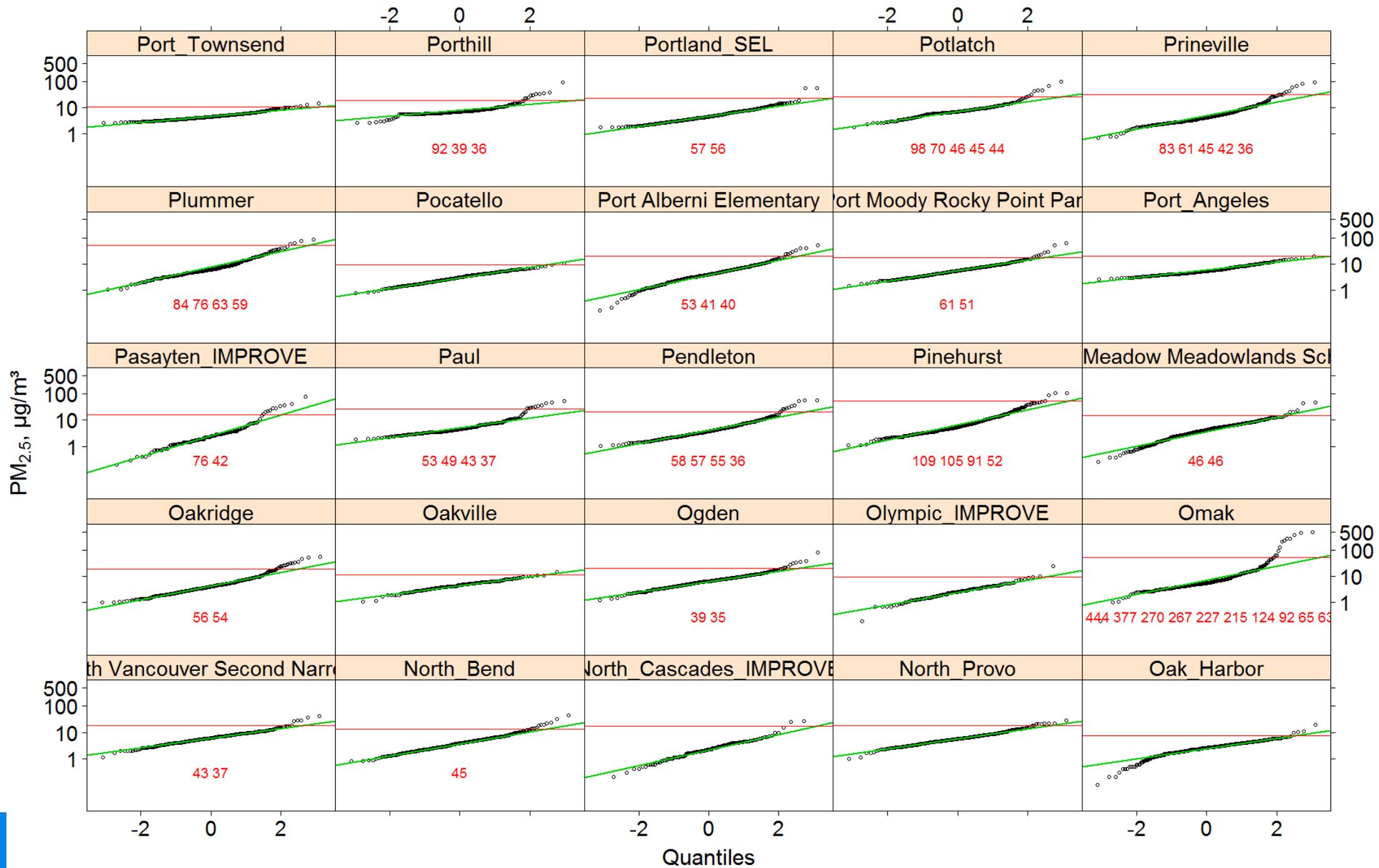
- Like their bacterial cousins, they show up during extreme events and make for “colorful” DVs
- “Adjusted boxplot” method for removing wildfire-affected  $PM_{2.5}$  &  $PM_{10}$ , and windstorm- effected  $PM_{10}$
- Adjusted boxplots can set cutoff too low sometimes
  - Set  $PM_{2.5}$  cutoff floor to  $35 \mu\text{g}/\text{m}^3$ . Only censored Jun- Sep 2015 data
  - Only removed  $PM_{10} > 200$  at non- IMPROVE sites
- Used  $PM_{10}$  cutoff at IMPROVE or FRM sites to avoid max- based DV



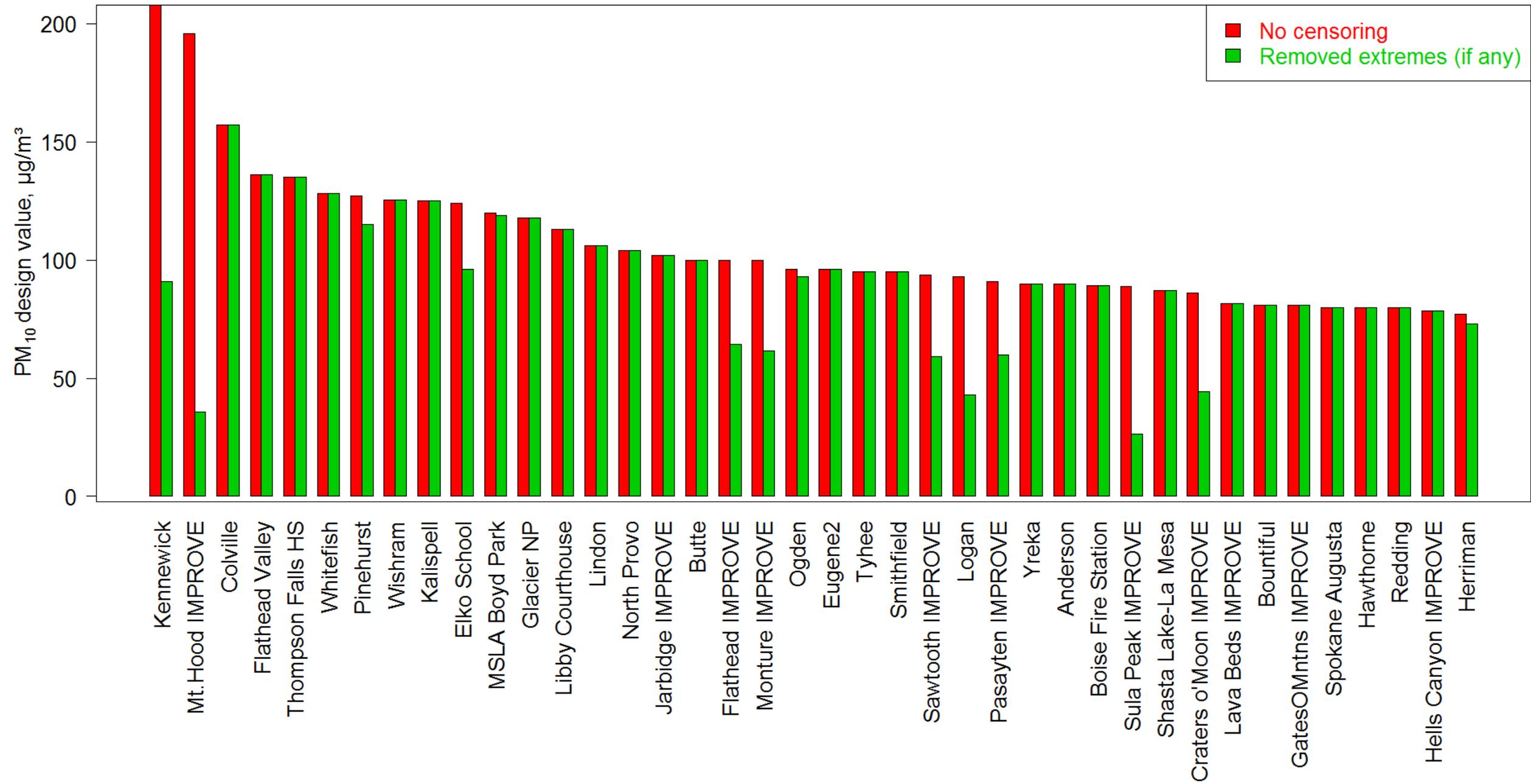
PM<sub>2.5</sub> DVs before and after censoring extremes. Only sites > 30 shown

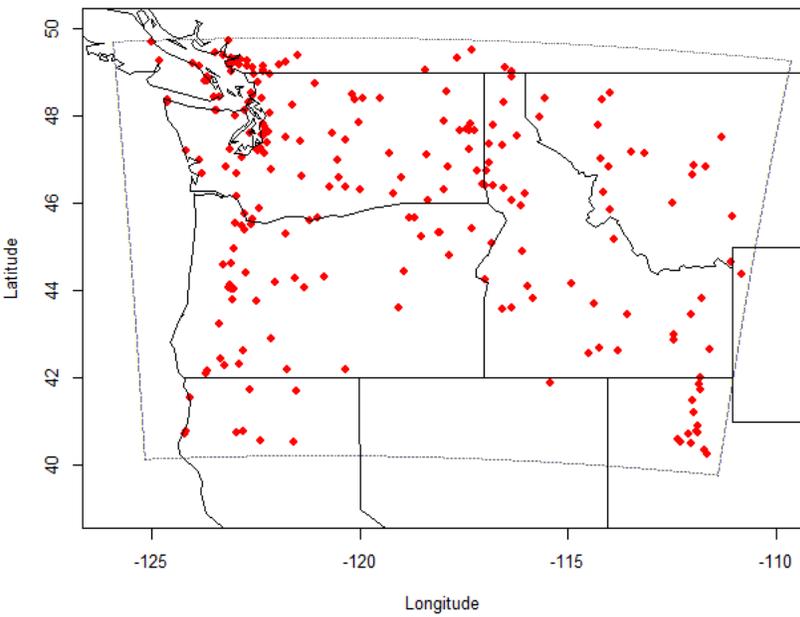
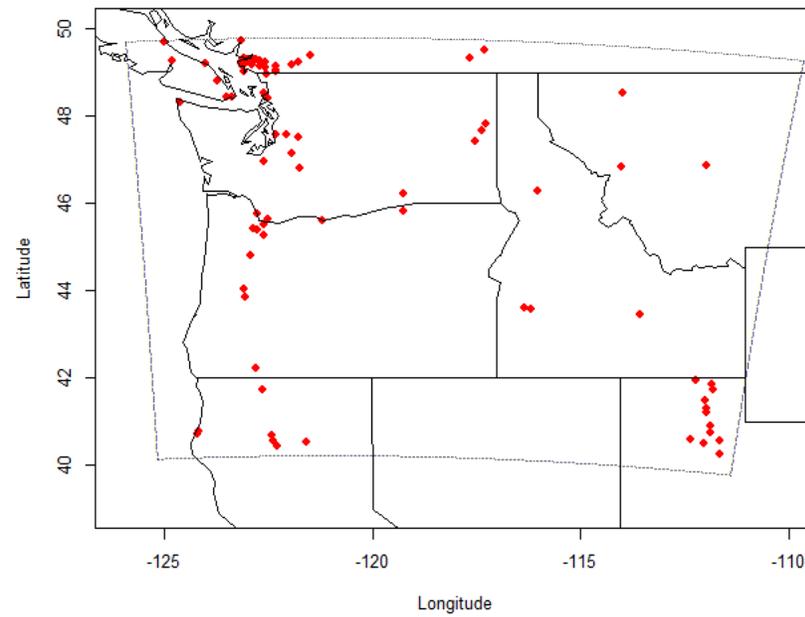
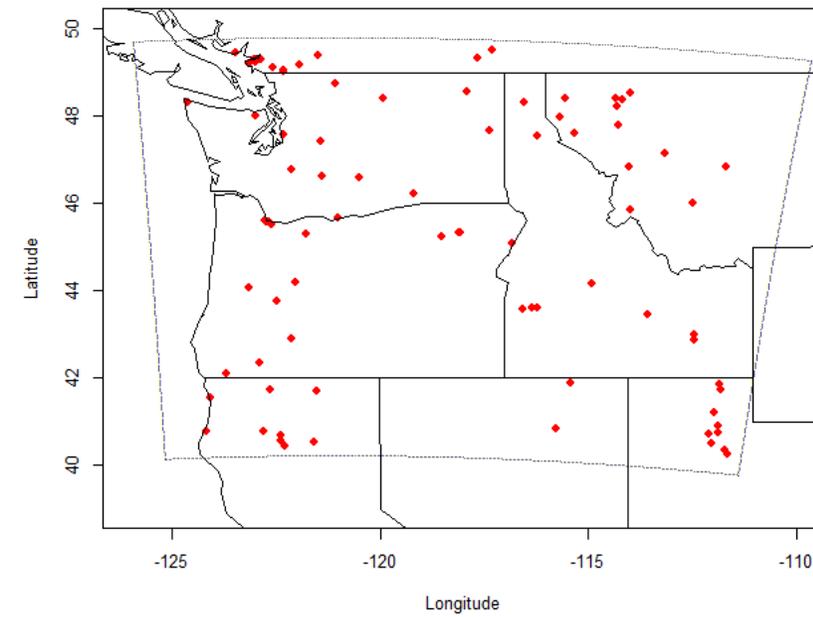
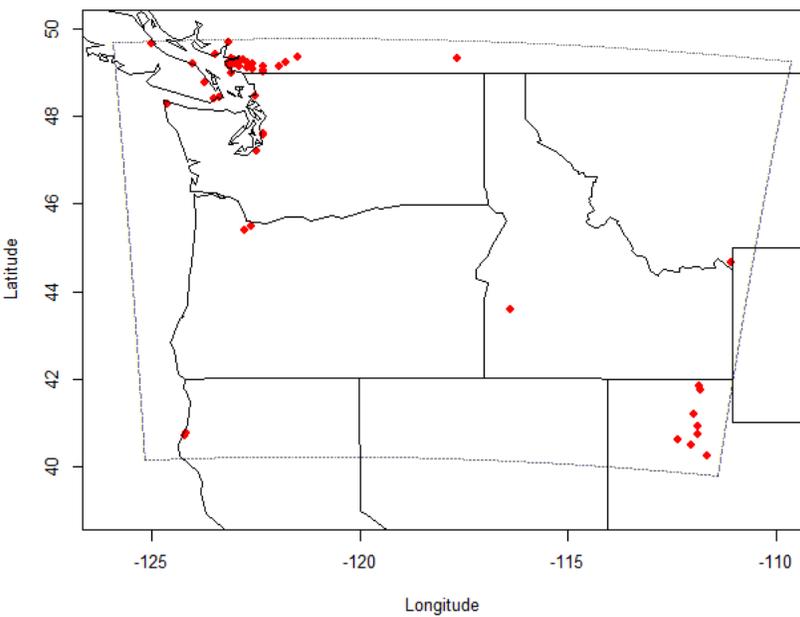
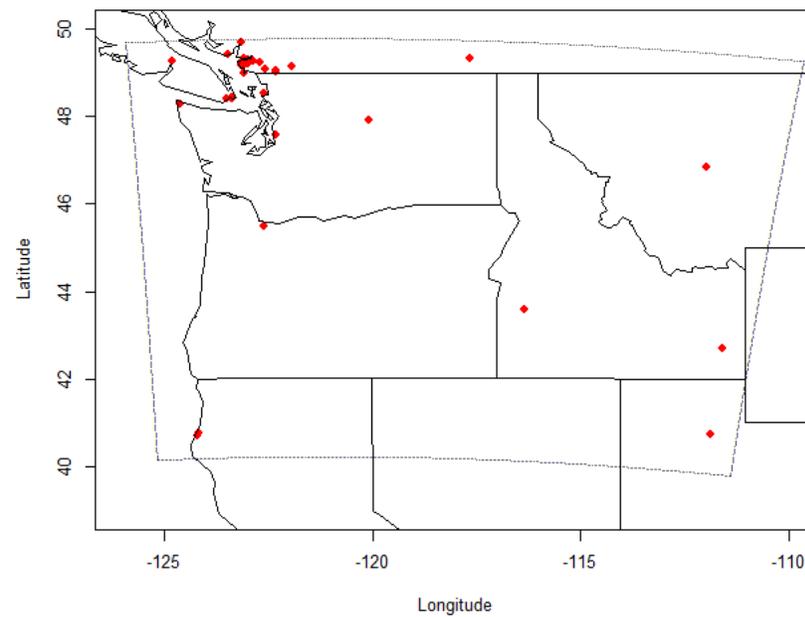
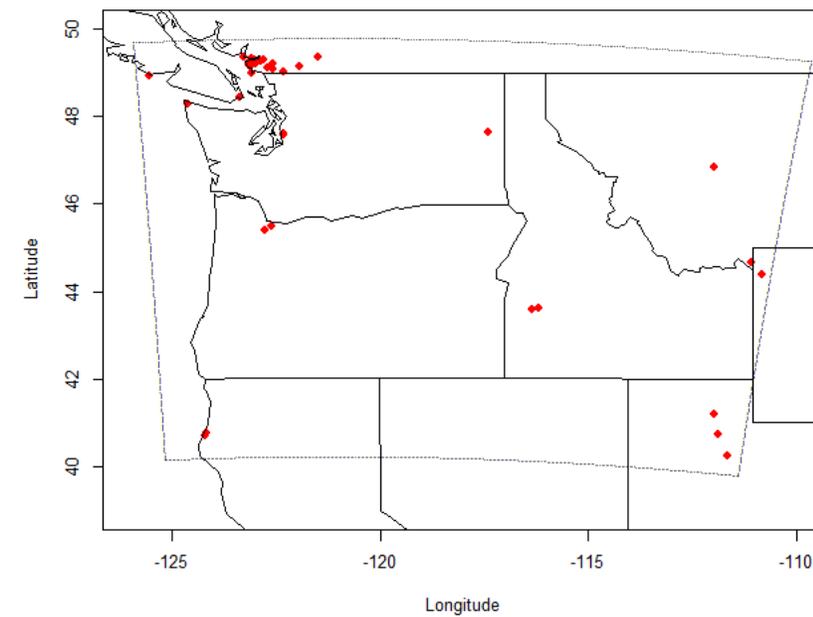


**Q-Q  
norm  
plots for  
removing  
wildfire-  
impacted  
PM2.5  
data**



PM<sub>10</sub> DVs before and after censoring extremes. Only sites > 75 shown



**PM2.5****O3****PM10****NO2****SO2****CO**

Drumroll please

**OVER TO JILL**

# Goal: Interpolated DVs at 4km Grid Cells

- Challenges:
  - Monitor density is highly variable
  - Model/monitor agreement varies by location and pollutant
  - Model accuracy is not great at the 98<sup>th</sup>/99<sup>th</sup> percentiles
    - But that's where we need design values



# Enter Empirical Bayesian Kriging

- EBK subsets data and estimates semivariograms through repeated simulations
  - Other interpolation tools (e.g. ordinary kriging/IDW) require manual parameter inputs and more assumptions



# Even Better: EBK Regression Prediction

- Combines EBK with predictor (independent) variables
- Monitors = dependent variables
- Model points = independent variables
- Output = Interpolated model/monitor surface



# Benefits of EBK Regression Prediction

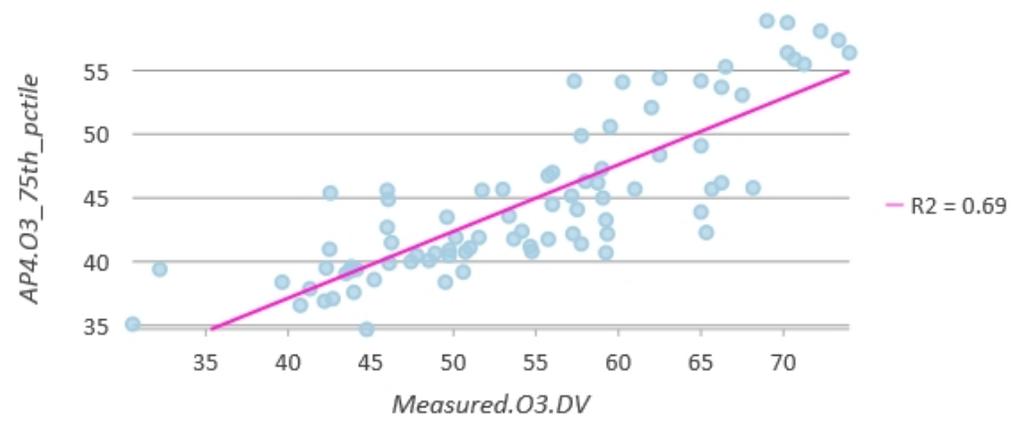
- Previous work interpolated model/monitor ratio. EBK-RP bypasses that step.
- It's the best of both worlds:
  - Uses model output to capture finer-scale spatial variation from topography, meteorology, etc.
  - Uses monitor design values to ensure accuracy at the upper percentiles
- Allows model/monitor relationships to vary by location



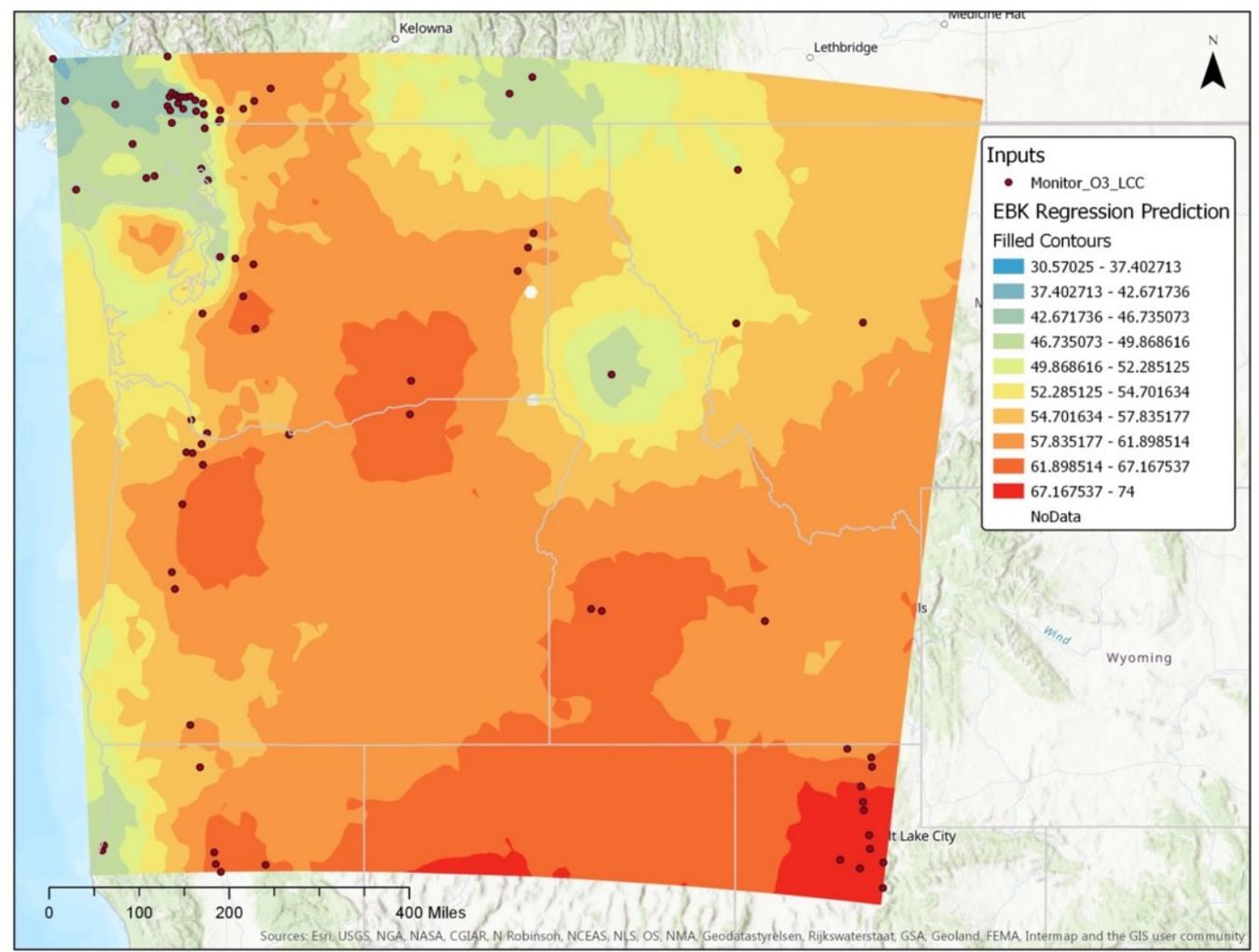
**BUT it's only available with ArcGIS Pro**

# EBK Regression Prediction Results: O<sub>3</sub>

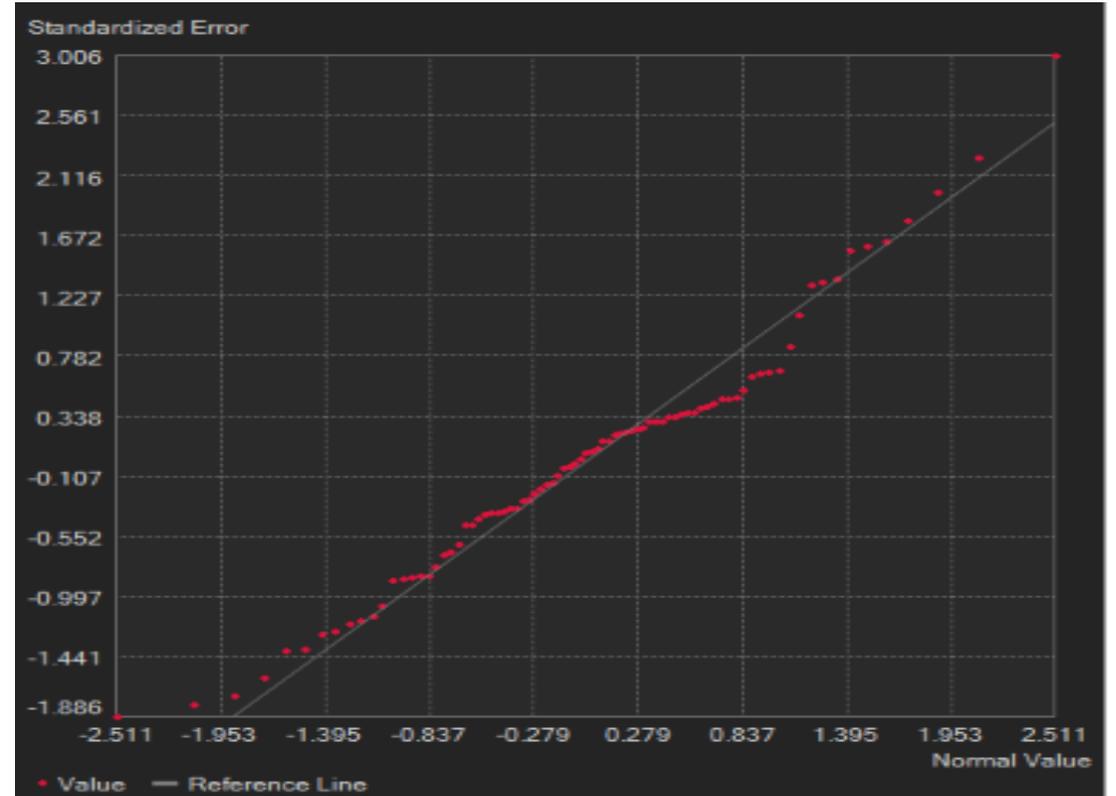
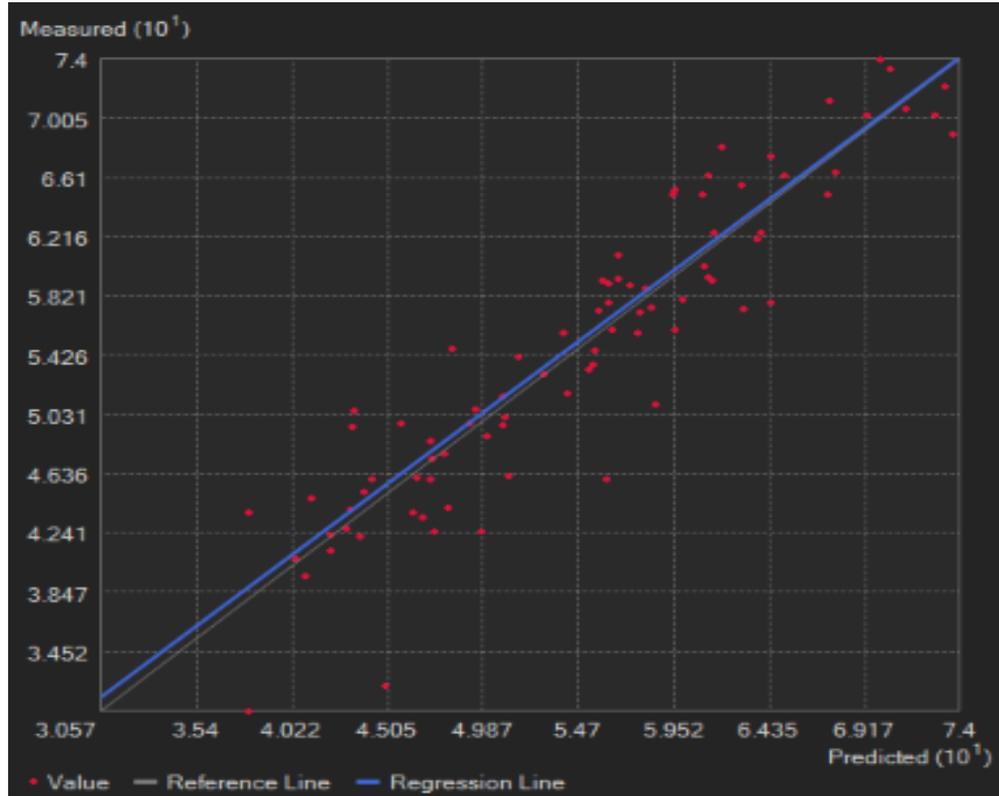
Relationship between Measured.O3.DV and AP4.O3\_75th\_pctile



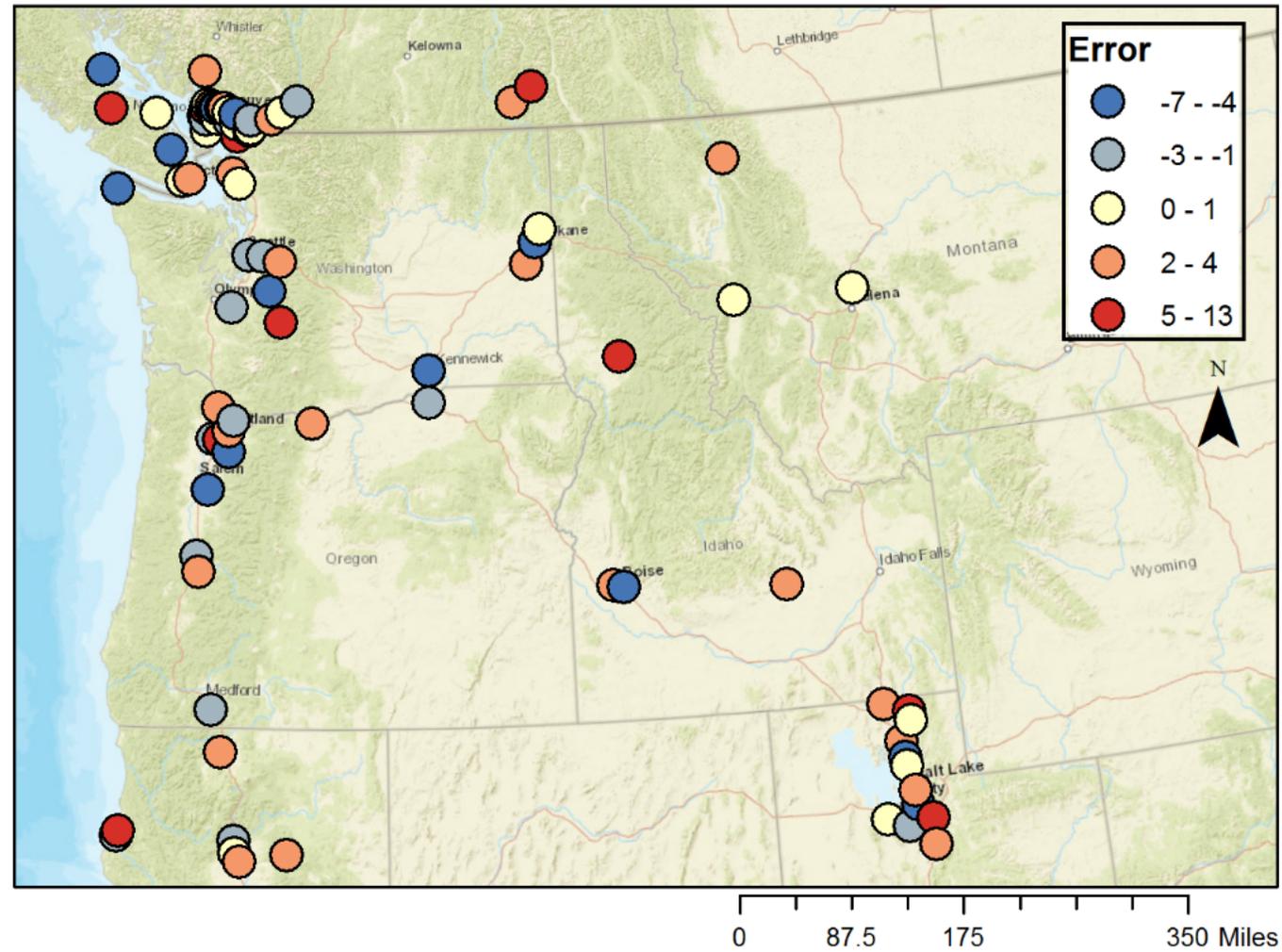
Variogram subset size: 50  
Neighbors: 10-50  
RMSE: 3.73 ppb  
R<sup>2</sup>: 0.93  
Cross-validated R<sup>2</sup>: 0.86



# EBK Regression Prediction Results: O<sub>3</sub>

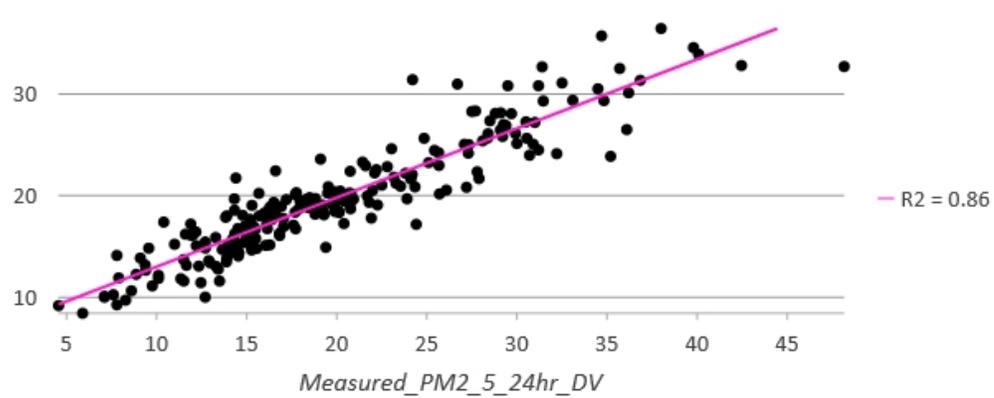


# EBK Regression Cross-Validated Prediction Results: O<sub>3</sub>



# EBK Regression Prediction Results: PM<sub>2.5</sub>

Measured vs. EBKRegression\_PM25\_3yrmedian\_censored



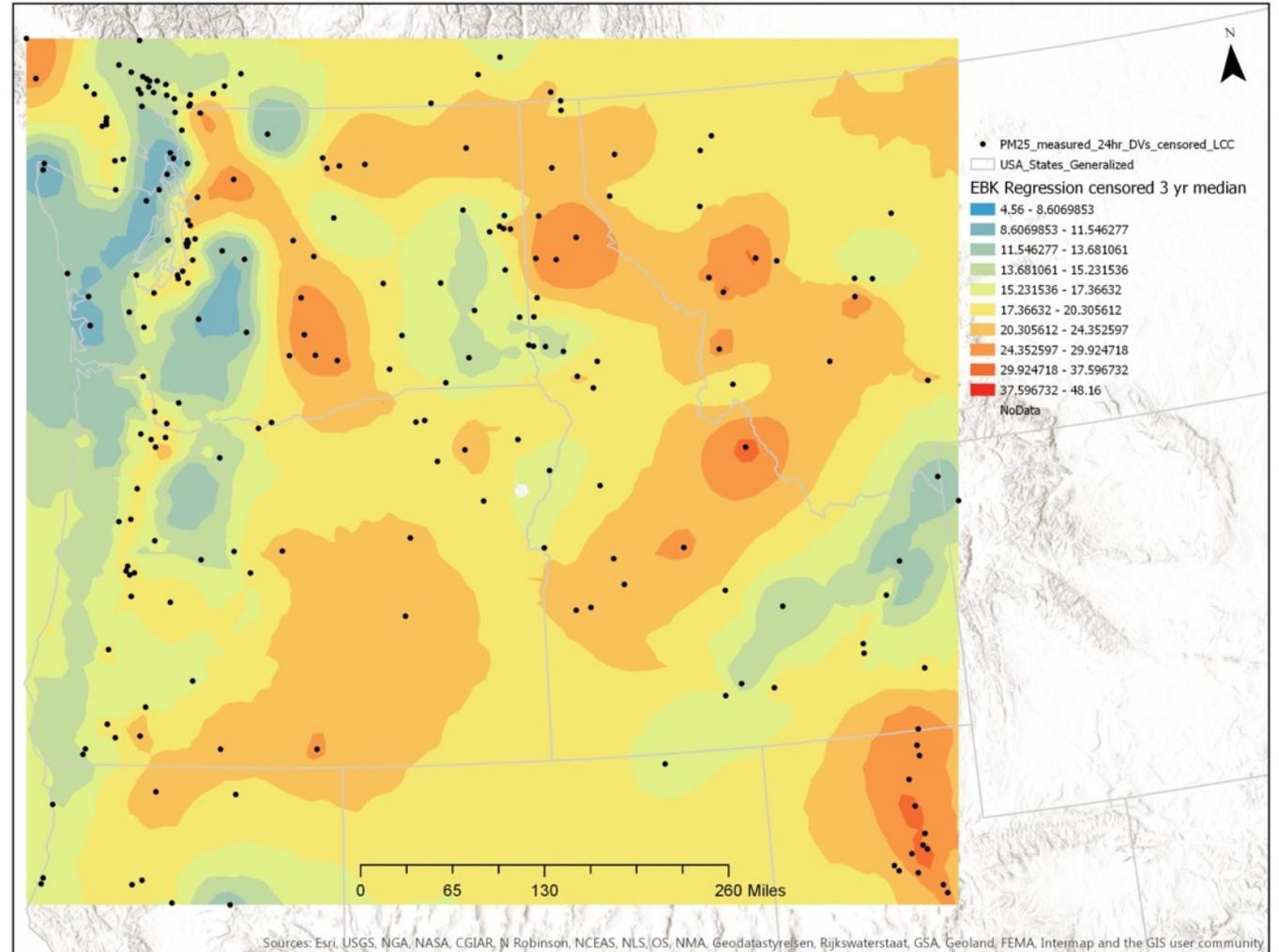
Variogram subset size: 100

Neighbors: 10-50

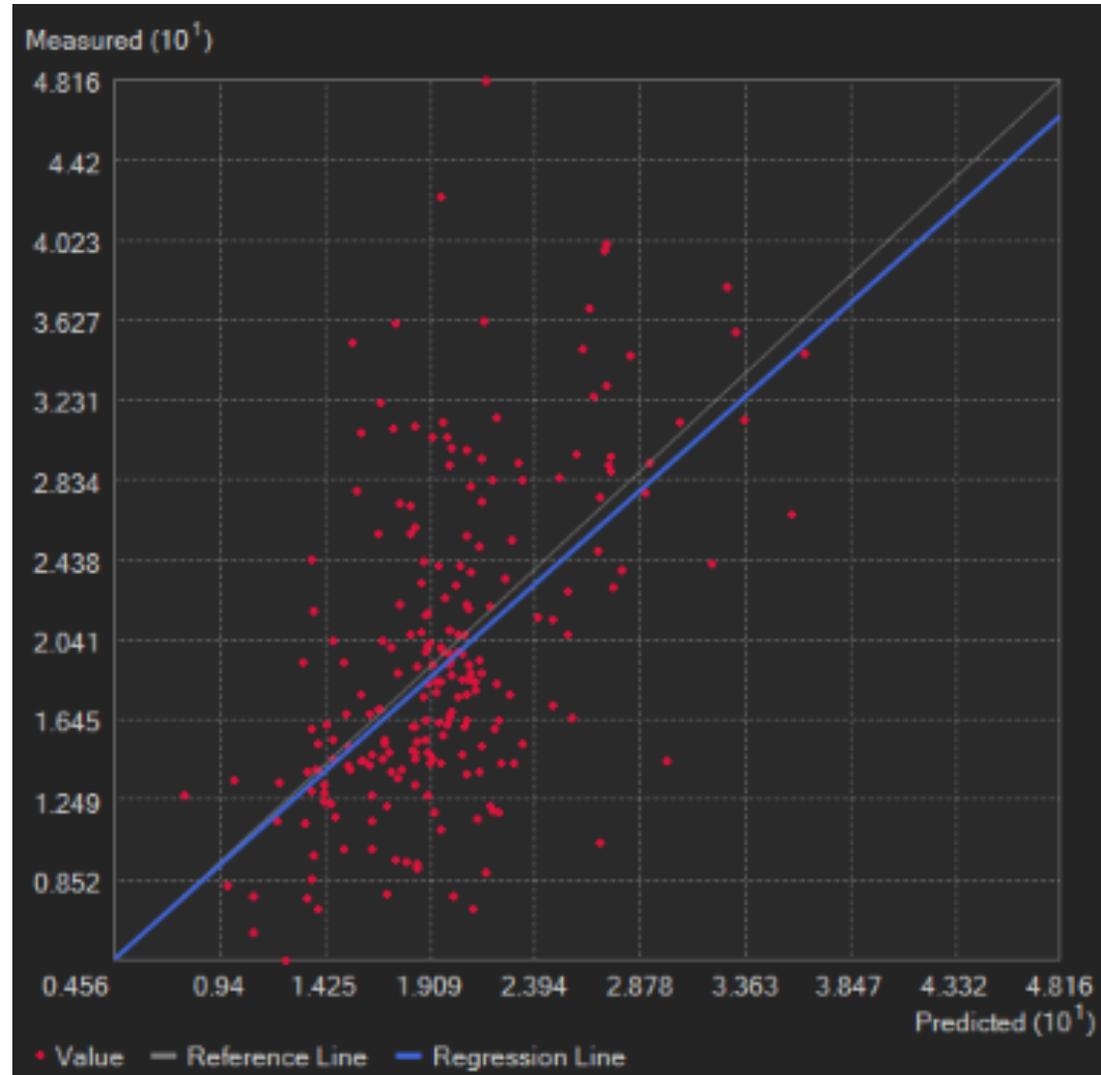
RMSE: 6.53  $\mu\text{g}/\text{m}^3$

R<sup>2</sup>: 0.86

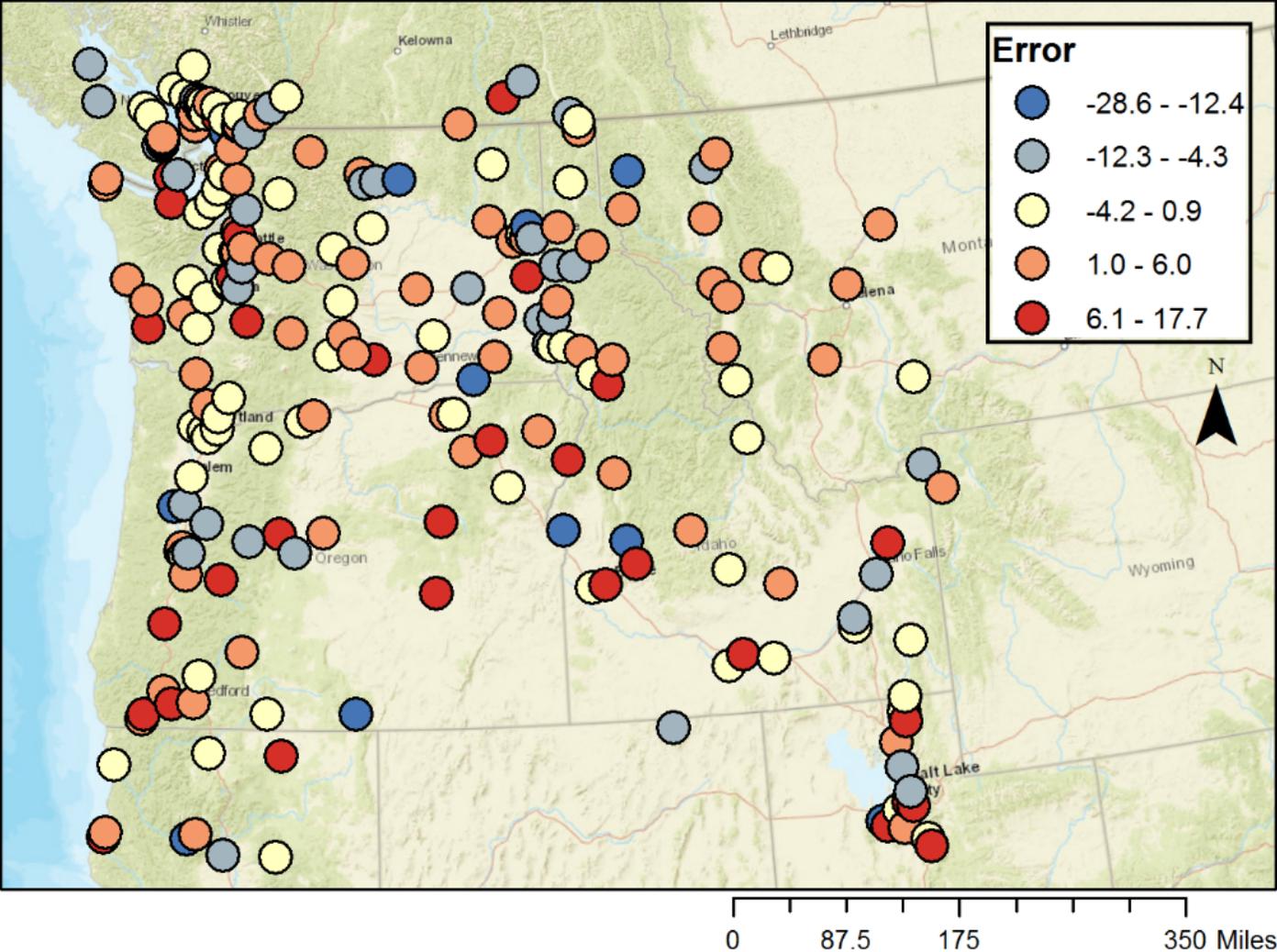
Cross-validated R<sup>2</sup>: 0.30



# EBK Regression Prediction Results: PM<sub>2.5</sub>



# EBK Regression Cross-Validated Prediction Results: PM<sub>2.5</sub>



# Next Steps

- Improve performance by tweaking inputs
- Expand to other criteria pollutants
- Daily interpolation?
  - No comparable R package (yet)
  - Could iterate in ArcGIS Pro using Python
    - Depends on computational time and Sara's patience
- Host results in ArcGIS Online web map app
  - Users can select and analyze data
  - Users can subset area of interest



# Should be done by December

ODEQ on tap to develop documentation

