SCIENTIFIC DATA

Overleaf

Daily wildfire and non-wildfire $PM_{2.5}$ concentration estimates by ZIP code in western US states, 2008-2018

Melissa M Maestas[1], Colleen Reid[2*], Ellen Considine[1], Gina Li[1]

June 17, 2019

1. Cooperative Institute for Research in Environmental Sciences, Earth Lab and 2. Department of Geography, University of Colorado Boulder, Boulder, Colorado, USA *corresponding author(s): Colleen Reid (Colleen.Reid@Colorado.edu)

Abstract

[Insert abstract text here.]

Background & Summary

[Insert Background & Summary text here.]

[insert Figure 1: monitor locations (points) and state boundaries]

[insert Table 1: list variables]

Methods

Study Area

Our study area includes 11 western US states: Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.[What other descriptions should we put? - square kilometers? climate zones? topography? other?]

Example of citation: [1]

$PM_{2.5}$ Measurements

[Write short description of each $PM_{2.5}$ data source.]

We downloaded the 2008-2018 pre-generated daily summary files for $PM_{2.5}$ (88101 and 88502 parameter codes) (https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily) as well as the spreadsheet listing all AQS monitors with datums (https://aqs.epa.gov/aqsweb/airdata/aqs_monitors.zip) from the United States Environmental Protection Agency (US EPA).

All available $PM_{2.5}$ data in the Fire Cache Smoke Monitor Archive (https://wrcc.dri.edu/cgi-bin/smoke.pl) was downloaded for the years 2008-2018.

$PM_{2.5}$ data from the Uintah Basin, Utah were provided by Seth Lyman at Utah State University (personal communication).

Predictors

[Write short description of each predictor data set and refer to Table 1]

Machine learning modelling and mapping

[Write description of ML modelling approach]

Code availability

[Insert brief description of how to access code on GitHub.] The code was written and annotated in R [version number] and Python [version number] and is available from GitHub [doi citation link]. The key package for implementing the ML model was [caretEnsemble?].

Data Records

All data are freely available from [repository name, data doi citation]. We provide ... [reference Figure 2]

[insert Figure 2: choropleths at zip code level - 4-panel: a) highest year PM2.5, Aug or Sept, b) highest year PM2.5, Jan/Feb, c) lowest year PM2.5, Aug or Sept, d) lowest year PM2.5, Jan/Feb.]

[Insert Table 3: list of files]

Technical Validation

[Write description of goodness of fit methods/metrics - out-of-bag data, RMSE, R2, models run on subsets of data, etc.]

[Insert Figure 4: a) out-of bag observed PM2.5 vs predicted, b) full model observed PM2.5 vs predicted, c-j) various subsets of data - oob and full model plots (see figure 5 of example paper)]

[Write discussion about variable importance, possibly referring to the suggested figure of variable importance panel figure. Could make an observation or two about the complexity of the variables, e.g., PM2.5 can be highest at highest and lowest temperatures (summer fire season and winter inversions), etc.]

[Thoughts - insert figure of predicted PM2.5 vs predictor variable for the 8 (or so) most important variables (panel figure)]

Thoughts: compare to PM2.5. Concerned comparing to HMS will take too long?

Usage Notes

[Write brief description of things the provided code can be adapted to do, such as making plots of specific years, use in health/pollution studies.]

Acknowledgements

[Write acknowledgements text here.]

Author contributions

[Write brief description of contribution from each author.]

Competing interests

The authors declare not competing interests.

Figures and figures legends

[All figures go here and are referred to in the text]

Tables

[All tables go here and are referred to in the text - read template text for tables]

References

[1] Y. Liu, J. A. Sarnat, V. Kilaru, D. J. Jacob, and P. Koutrakis. Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. *Environ Sci Technol*, 39(9):3269–78, May 2005.