

Overview of CACES LUR Models

This readme is divided into three parts: (1) Models of six pollutants, through 2015; (2) Models of six pollutants, 2016-2020; (3) Models for Particle Number Concentration (PNC), 2017.

For all of the models, data are free and publicly available for non-commercial purposes. Data may not be used for commercial purposes without prior, written permission.

1. Models of six pollutants, through 2015

These models provide estimates of outdoor concentrations, through 2015, for six pollutants (four gases: O₃, CO, SO₂, NO₂; two aerosols: PM₁₀, PM_{2.5}) throughout the contiguous U.S. Model estimates, with the exception of ozone, are annual-average values, for years with available monitoring data. Ozone model estimates are the average during May through September of the daily maximum 8-hour moving average, for years with available monitoring data.

Pollutant	Units	Available years
O ₃	ppb	1979-2015; 2016-2020
SO ₂	ppb	1979-2015; 2016-2020
NO ₂	ppb	1979-2015; 2016-2020
PM ₁₀	µg/m ³	1988-2015; 2016-2020
CO	ppm	1990-2015; 2016-2020
PM _{2.5}	µg/m ³	1999-2015; 2016-2020

Data are available at national, state, county, census tract*, and census block group* levels, based on population-weighted averages of census block level predictions using 2010 census geographies.

*Our models only provide estimates for tracts and block groups with a population greater than zero.

These models are derived from publicly available concentration measurements from U.S. EPA regulatory monitors, and use information about land use (for example, locations of major and minor roads; elevation; and whether an area is urban or rural) and satellite-derived estimates of air pollution to predict concentrations at locations without measurements. Researchers generally do not use estimates of the types given here to draw conclusions about air quality at one location or a small number of specific locations; as such, we do not recommend that type of use.

File Format

Parameter	Description
Fips	Unique code for geographic unit with the following structure: SSCCTTTTTTG ¹ ¹ S = 1-2 digit state code; C = 3 digit county code; T = 6 digit tract code; G = 1 digit block group code This code can be used to join these data to Census data and shapefiles

pollutant	Pollutant for a given observation (co, no2, o3, pm10, pm25, so2)
year	Prediction year for a given observation
pred_wght	Population-weighted concentration based on block level centroid predictions, units vary by pollutant (see above)
state_abbr	State abbreviation for a given observation
lat/lon	Population-weighted latitude and longitude based on block level centroid

Acknowledgment

Data are free and publicly available for non-commercial purposes. Data may not be used for commercial purposes without prior permission. Please cite as follows:

"This article includes concentration estimates developed by the Center for Air, Climate and Energy Solutions using v1 empirical models as described in Kim et al., 2018."

Suggested citation: Kim S.-Y.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. 2018. "Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: Role of prediction model parsimony in integrated empirical geographic regression." PLoS ONE 15(2), e0228535. DOI: 10.1371/journal.pone.0228535

2. Models of six pollutants, 2016 - 2020

Models for the six pollutants for 2016 – 2020 are generated the same as for 2015 and earlier years. (For brevity, this section of the README file refers to the 2016-2020 models as "post-2015" and the models for 2015 and earlier as "pre-2015"; the year 2015 itself is within "pre-2015".)

For a small number of locations, the datasets for predictor variables were missing data post-2015. For those locations, we do not make a concentration prediction for that geography. (Internally, predictions are made at the block level, and then averaged up, with population-weighting, to the block-group, tract, and county -level results given on the website.) Specifically, the post-2015 models predict 4,997 fewer census block groups, 1,604 fewer census tracts, and one fewer county, compared to pre-2015 models (see table below).

Spatial resolution	Predictions pre-2015	Predictions post-2015	Number missing	Percent missing
Census block group	215,491	210,494	4,997	2.3%
Census tract	72,043	70,439	1,604	2.2%
County	3,109	3,108	1	0.03%

Acknowledgment

Data are free and publicly available for non-commercial purposes. Data may not be used for commercial purposes without prior, written permission. Please cite as follows:

"This article includes concentration estimates developed by the Center for Air, Climate and Energy Solutions using v1 empirical models as described in Kim et al., 2018 and Lu et al., 2024."

Suggested citation: Kim S.-Y.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. 2018. "Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: Role of prediction model parsimony in integrated empirical geographic regression." PLoS ONE 15(2), e0228535. DOI: 10.1371/journal.pone.0228535

Lu, T.; Kim, S.-Y.; Marshall, J. D. 2024. "High-resolution geospatial database: national criteria-air-pollutant concentrations in the contiguous U.S., 2016 – 2020." Geoscience Data Journal, under review.

3. Overview of Ultrafine (Particle Number Concentration) Model

Model predictions: Our model provides estimates of outdoor concentrations for particle number concentrations (PNC; unit: particles number / cm^3) throughout the contiguous U.S. *Model estimates are annual-average values for 2017.* The model is derived using particle number concentration measurements from mobile monitoring and fixed sites across the U.S. and a land-use regression modeling framework to predict concentrations at locations without measurements. It is noted that we only made predictions at census blocks with predictor variable values within the 1st and 99th percentile range of the measurement dataset used for model development. This means that model predictions do not extrapolate in covariate space. After applying this constraint, the model predicts outdoor concentration at 6,056,703 residential census block centroids in the contiguous US, which corresponds to 98.6% of the population. This constraint censors predictions for 1.9% of census blocks (1.4% of the population) in the contiguous US. Censored census blocks are generally in extremely urban (e.g., in Manhattan, New York City) and rural locations.

Population-weighted average concentrations: Estimates available here are at national, state, county, census tract, and census block group levels, based on population-weighted averages of census block level predictions using 2010 census geographies. You may get the census block level estimates by contracting Albert A. Presto (albert.presto@gmail.com) or Provat K. Saha (sahaprovat@gmail.com).

File format

fips: Unique code for geographic unit with the following structure: SSCCTTTTTTG

(S = 1–2-digit state code; C = 3-digit county code; T = 6-digit tract code; G = 1 digit block group code).

PNC: Population-weighted particle number concentration based on block level centroid predictions, unit (particle number / cm^3)

lat/lon: Population-weighted latitude and longitude based on block level centroid.

Suggested citation:

Data are free and publicly available for non-commercial purposes. Data may not be used for commercial purposes without prior, written permission. Please cite as follows:

"This article includes ultrafine particle concentration estimates developed by the Center for Air, Climate and Energy Solutions using an empirical model, as described in Saha et al., 2021."

Saha, P. K.; Hankey, S.; Marshall, J. D.; Robinson, A. L.; Presto, A. A. High-Spatial- Resolution Estimates of Ultrafine Particle Concentrations across the Continental United States. *Environ. Sci. Technol.* 2021, 55 (15), 10320–10331. <https://doi.org/10.1021/acs.est.1c03237>.