

ENERGY FINDINGS

How Differential Privacy Will Affect Estimates of Air Pollution Exposure and Disparities in the United States

Madalsa Singh¹ ¹ Energy Science and Engineering, Stanford University

Keywords: environmental justice, air quality, census data, differential privacy, pollution

<https://doi.org/10.32866/001c.74975>

Findings

Census data is crucial to understand energy and environmental justice outcomes such as poor air quality which disproportionately impact people of color in the U.S. With the advent of sophisticated personal datasets and analysis, Census Bureau is considering adding top-down noise (differential privacy) and post-processing 2020 census data to reduce the risk of identification of individual respondents. Using 2010 demonstration census and pollution data, I find that compared to the original census, differentially private (DP) census significantly changes ambient pollution exposure in areas with sparse populations. White Americans have lowest variability, followed by Latinos, Asian, and Black Americans. DP underestimates pollution disparities for SO₂ and PM_{2.5} while overestimates the pollution disparities for PM₁₀.

1. Questions

Researchers studying energy systems and decarbonization rely heavily on census data to understand environmental justice outcomes (Brockway, Conde, and Callaway 2021; Burger 2019; Tessum et al. 2021; Thind et al. 2019). Population counts of different racial and ethnic groups at various spatial resolutions – state, county, census tract, block groups, and blocks – are used to find out which populations are adversely affected and where do they reside. Finer data resolution can help identify specific neighborhoods and communities for targeted energy and environmental policies by government. At the same time, Census Bureau is required by law to protect the privacy of Census participants and has implemented various disclosure avoidances systems (DAS) since 1960s. The bureau introduced a new DAS called differential privacy (DP) for the Census data of 2020. DP injects top-down random noise to Census tabulations. Noise is smallest at the national or state level and highest for smaller spatial units such as block group or blocks. Various post-processing steps, though not formally part of differential privacy, are required to maintain the facial validity of census products (Kenny et al. 2021). The question is whether the infusion of random noise coupled with post-processing adjustments lead to unintentional systemic deviations in understanding environmental justice outcomes.

Air quality is a useful case-study to investigate in this context. Air pollution can vary significantly across small distances. Estimates of exposure disparities are impacted by the spatial resolution at the level of input (spatial unit at which outcome is observed) as well as level of aggregation (spatial unit at which outcome is reported). Pollution exposure, which can change over short distances, is more accurately observed at finer spatial scale, usually block or

block-group level data, while increasing the level of aggregation to larger spatial resolution (state or county level) underestimates disparities compared to census tract or block group level aggregation (Clark et al. 2022; Paolella et al. 2018). Noise and adjustments in census data can significantly alter these estimates.

In this piece, I answer how introducing differential privacy in Census data impact:

1. Air pollution exposure of different race and ethnicity in the United States
2. Exposure disparities when aggregated to county and census tract levels.

2. Methods

I use population data at census block group level (CBG) from the original 2010 Census and from the latest experimental runs of differential privacy algorithm applied to the original 2010 Census (Vintage 2022-08-25) from IPUMS NHGIS (Manson et al. 2022). Americans who identify as non-Hispanic blacks only, non-Hispanic whites only, non-Hispanic Asians only, non-Hispanic native American and American Indian only are referred to as Blacks, Whites, Asians, and Native Americans in this work. Latinos include all Americans who identify as Latinos or Hispanics. Americans who identify as mixed race aren't included in this analysis. I use census block group level (CBG) ambient pollution estimates of four air pollutants ($PM_{2.5}$, PM_{10} , NO_2 , SO_2) for the year 2010 from the Center for Air, Climate and Energy Solutions ("The Center for Air, Climate, and Energy Solutions" n.d.) as described in published work (Kim et al. 2020).

Exposure of pollutant i by race and ethnicity j is aggregated to census tract level and county level is given as:

$$\begin{aligned}
 & \textit{Exposure}_{i,j, \textit{census tract or county}} \\
 &= \frac{\sum_{\textit{census tract or county}} \textit{over all CBG in a } \textit{Exposure}_{i,CBG} \times \textit{Population}_{j,CBG}}{\sum_{\textit{census tract or county}} \textit{over all CBG in a } \textit{Population}_{j,CBG}}
 \end{aligned}$$

Where $\textit{Exposure}_{i,CBG}$ denotes the ambient pollution estimate of pollutant i in each census block group (CBG) and $\textit{Population}_{j,CBG}$ denotes the total population or population of race/ethnicity j in each census block group (CBG) summed over all census block groups in a census tract or county. Figure 1 and 2 plot the percentage difference of exposure of pollutants experienced by total population and different race and ethnicity in the differentially private census compared to the original census aggregated at county and census tract level respectively. Census tracts or counties with any population count of 0 in either original or differentially private census are removed.

To understand the impact of differential private census products on pollution disparities, I estimate risk gap at county and census tract levels. Risk gap is defined as the difference between the pollution exposure of most burdened group, i.e., maximum value of exposure for a race and ethnicity as calculated above and the total population average exposure.

$$\begin{aligned} \text{Risk Gap}_{i, \text{census tract or county}} &= \max(\text{Exposure}_{i, j, \text{census tract or county}}) \\ &\quad - \text{Exposure}_{i, \text{population, census tract or county}} \end{aligned}$$

Where $\text{Exposure}_{i, j, \text{census tract or county}}$ is pollution exposure of pollutant i by race and ethnicity j in census tract or county and $\text{Exposure}_{i, \text{population, census tract or county}}$ is the pollution exposure of pollutant i for the entire population in census tract or county. Figure 3 plots the ratio of risk gap calculated using the DP and original census by the population average pollution exposure at census tract and county. Ratio above (below) 1 denotes that DP census shows larger (smaller) risk gap compared to the original census.

3. Findings

Differential privacy in census data significantly changes the ambient pollution exposure in small spatial units with sparse population of people of color ([Figure 1](#) and [2](#)). Census tracts have higher variations than counties. White American have the lowest variance in exposure, followed by Latinos, Asian, and Black Americans. This is, in part, due to post-processing procedure which gives priority to the accuracy counts for the largest racial group in an area. The changes in pollution exposure also depends on the pollutant. For example, in counties with sparse population of Asian and Black Americans, the NO₂ exposure changes can be as high as +/- 50%. Exposure differences nullify for larger population counts.

[Figure 3](#) displays the ratio of the risk gap calculated by DP census and original census with ambient pollutant levels for both county and census tract aggregation. Differentially private census underestimates (ratio less than 1) the disparity for SO₂ in both county and census tract aggregations. The ratio decreases with higher levels of ambient SO₂. DP overestimates the risk gap associated with PM₁₀ for both county and census tract compared to original census (ratio greater than 1), with the ratio increasing as the ambient pollution of PM₁₀ increases. The trends in risk gap ratio at the county level for NO₂ and PM_{2.5} are not significant, but DP significantly underestimates the disparity for PM_{2.5} at the census tract level, particularly in more polluted census tracts.

.....

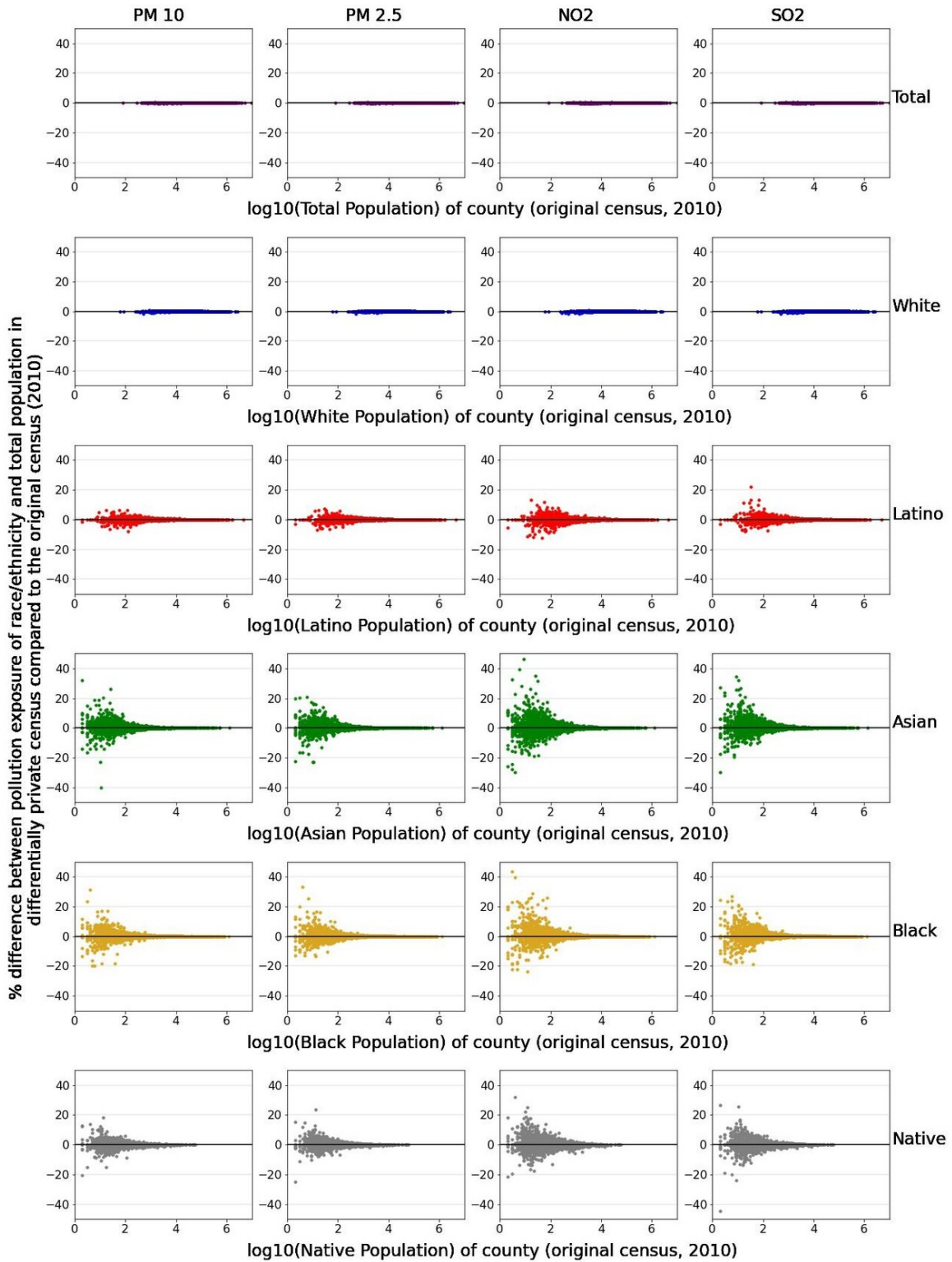


Figure 1. Percentage change in air pollution exposure (PM₁₀, PM_{2.5}, NO₂, and SO₂) of total population and different racial and ethnic groups in differentially private census compared to the original census for year 2010 aggregated at county level.

X axis for each plot shows the logarithm (base 10) of population count of specific racial and ethnic group or the total population.

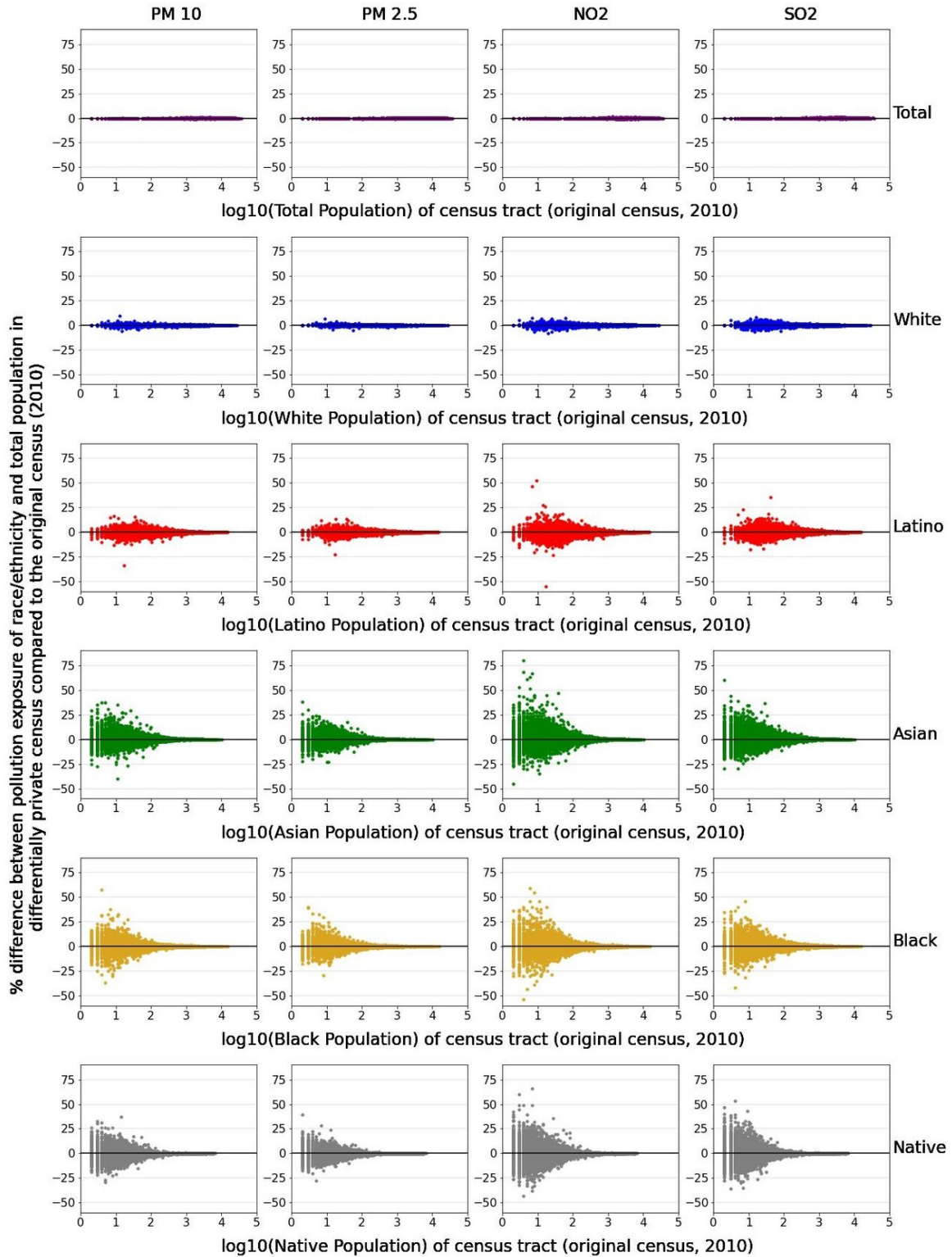


Figure 2. Percentage change in air pollution exposure (PM₁₀, PM_{2.5}, NO₂, and SO₂) of total population and different racial and ethnic groups in differentially private census compared to the original census for year 2010 aggregated at census tract level.

X axis for each plot shows the logarithm (base 10) of population count of specific racial and ethnic group or the total population.

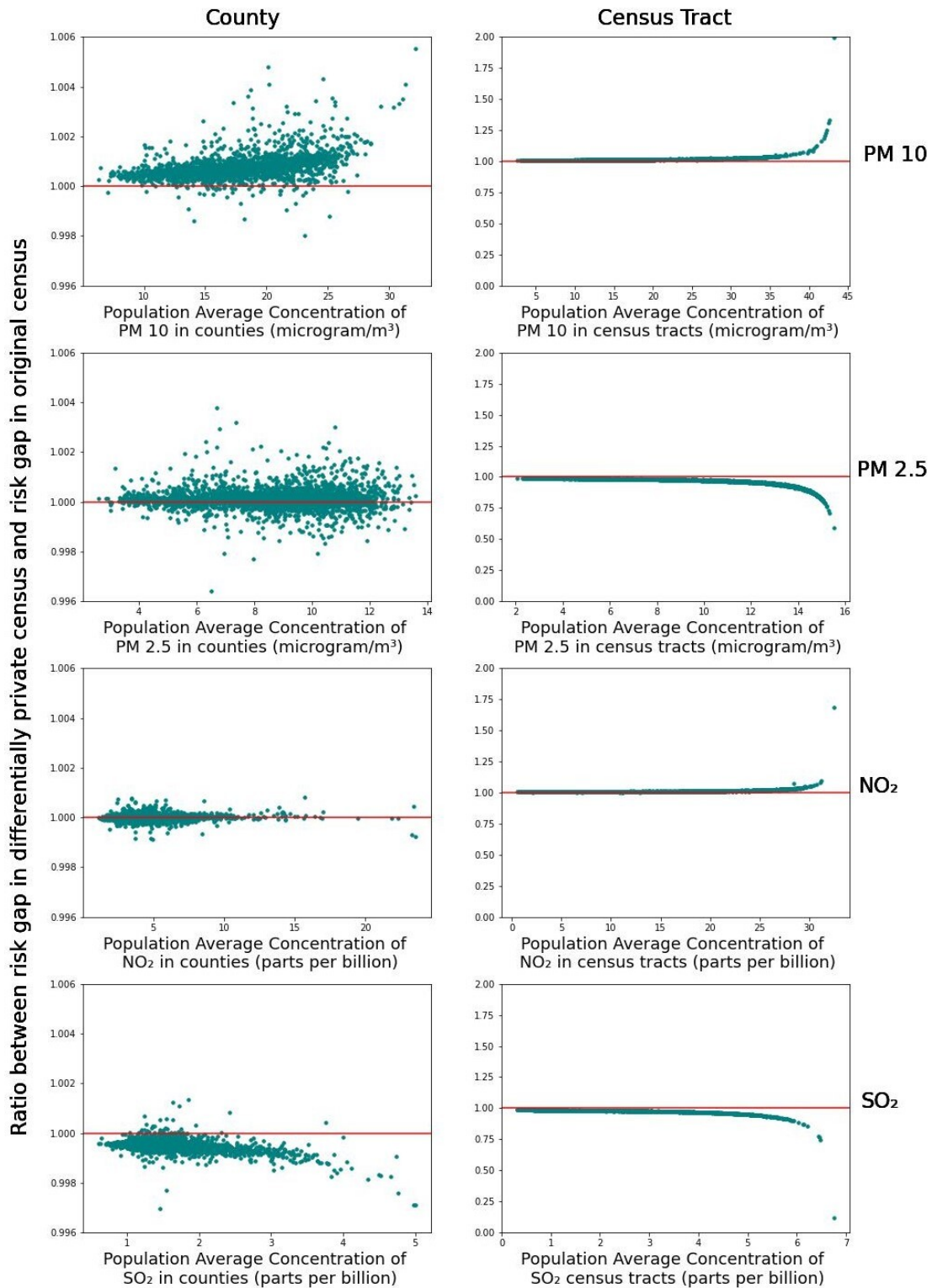


Figure 3. Ratio of risk gap for pollutants at county (left) and census tract (right) level of DP census compared to the original census.

X axis shows the population average concentration of pollutant in a county or census tract. Risk gap is defined as the difference between the pollution exposure of the most burdened race and ethnicity compared to the population average exposure. Ratio less than 1 indicates that DP underestimates pollution disparity compared to the original census and vice-versa.

ACKNOWLEDGMENTS

I acknowledge that I received no funding in support of this research. I thank Josh Apte and Inês Azevedo for their feedback on an earlier version of this work. I thank the two anonymous reviewers that greatly improved the quality of this manuscript.

Submitted: March 01, 2023 AEST, Accepted: May 02, 2023 AEST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Brockway, Anna M., Jennifer Conde, and Duncan Callaway. 2021. “Inequitable Access to Distributed Energy Resources Due to Grid Infrastructure Limits in California.” *Nature Energy* 6 (9): 892–903. <https://doi.org/10.1038/s41560-021-00887-6>.
- Burger, Scott P. 2019. “Rate Design for the 21st Century: Improving Economic Efficiency and Distributional Equity in Electricity Rate Design,” 257.
- Clark, Lara P., Maria H. Harris, Joshua S. Apte, and Julian D. Marshall. 2022. “National and Intraurban Air Pollution Exposure Disparity Estimates in the United States: Impact of Data-Aggregation Spatial Scale.” *Environmental Science & Technology Letters* 9 (9): 786–91. <https://doi.org/10.1021/acs.estlett.2c00403>.
- Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan T. R. Rosenman, Tyler Simko, and Kosuke Imai. 2021. “The Use of Differential Privacy for Census Data and Its Impact on Redistricting: The Case of the 2020 U.S. Census.” *Science Advances* 7 (41): eabk3283. <https://doi.org/10.1126/sciadv.abk3283>.
- Kim, Sun-Young, Matthew Bechle, Steve Hankey, Lianne Sheppard, Adam A. Szpiro, and Julian D. Marshall. 2020. “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. <https://doi.org/10.1371/journal.pone.0228535>.
- Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. “National Historical Geographic Information System: Version 17.0.” Minneapolis, MN: Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D050.V17.0>.
- Paoletta, David A., Christopher W. Tessum, Peter J. Adams, Joshua S. Apte, Sarah Chambliss, Jason Hill, Nicholas Z. Muller, and Julian D. Marshall. 2018. “Effect of Model Spatial Resolution on Estimates of Fine Particulate Matter Exposure and Exposure Disparities in the United States.” *Environmental Science & Technology Letters* 5 (7): 436–41. <https://doi.org/10.1021/acs.estlett.8b00279>.
- Tessum, Christopher W., David A. Paoletta, Sarah E. Chambliss, Joshua S. Apte, Jason D. Hill, and Julian D. Marshall. 2021. “PM_{2.5} Polluters Disproportionately and Systemically Affect People of Color in the United States.” *Science Advances* 7 (18): eabf4491. <https://doi.org/10.1126/sciadv.abf4491>.
- “The Center for Air, Climate, and Energy Solutions.” n.d. CACES. Accessed February 27, 2023. <https://www.caces.us>.
- Thind, Maninder P. S., Christopher W. Tessum, Inês L. Azevedo, and Julian D. Marshall. 2019. “Fine Particulate Air Pollution from Electricity Generation in the US: Health Impacts by Race, Income, and Geography.” *Environmental Science & Technology* 53 (23): 14010–19. <https://doi.org/10.1021/acs.est.9b02527>.