# Residential segregation and summertime air temperature across 13 northeastern U.S. states: Potential implications for energy burden

Daniel Carrión[1], Johnathan Rush[1], Elena Colicino[1], and Allan Just[1]

[1]Affiliation not available

February 28, 2024



Daniel Carrión[1,2], Johnathan Rush[3], Elena Colicino[4], Allan C. Just[5,6]

[1] Department of Environmental Health Sciences, Yale University School of Public Health, New Haven, Connecticut 06510

[2] Yale Center on Climate Change and Health, Yale University School of Public Health, New Haven, Connecticut 06510

[3] Element 84, Alexandria, Virginia, 22314

[4] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York City, New York 10029

[5] Department of Epidemiology, Brown University School of Public Health, Providence, Rhode Island 02903[6] Institute at Brown for Environment and Society, Brown University, Providence, Rhode Island 02912

**Corresponding author:** Daniel Carrión

**Email:** daniel.carrion@yale.edu

**Author Contributions:** DC and ACJ designed the research; DC and JR performed the research; DC analyzed the data with input from EC and ACJ; DC and ACJ drafted the manuscript; and all authors approved the final version of the manuscript.

**Competing Interest Statement:** The authors have no competing interests to declare.

**Classification:** Social sciences; sustainability science

**Keywords:** environmental justice; extreme heat; energy insecurity; disparities

**Abstract**

High ambient summertime temperatures are an increasing health concern with climate change (1). This is a particular concern for minoritized households in the United States, for which differential energy burden may compromise their adaptive capacity to high temperatures (2-14). Using a fine-scaled spatiotemporal air temperature model and U.S. census data, we examined local (within-county) differences in warm season cooling degree days (CDDs) by ethnoracial group as a proxy for local energy demand across states of the northeast and mid-Atlantic U.S. in 2003–2019. Using state-specific regression models with fixed effects for year and county, we found that Black and Latino people consistently experienced more CDDs, non-Hispanic white people experienced cooler summers, and Asian populations showed mixed results. We also explored a concentration-based measure of residential segregation for each ethnoracial group as one possible pathway towards temperature disparities. The measure was constructed using a Gaussian kernel density smoothing

procedure of population-weighted census tract centroids per county. We included the segregation measure as a smooth term in a generalized additive model adjusted for county and year as well as a tensor smooth for census tract centroids. The results were nonlinear, but higher concentrations of white people were associated with lower annual CDDs and higher concentrations of Latino people were associated with higher annual CDDs than the county average. Concentrations for Black and Asian people were nonmonotonic, sometimes with bowed associations. These findings suggest that present-day residential segregation, as measured by ethnoracial subgroup concentrations, may contribute to summertime air temperature disparities and influence adaptive capacity.

**Significance Statement**

Energy poverty alleviation programs have focused primarily on cold season warming rather than warm season cooling. However, summertime energy burden may be an increasing concern for minoritized households nationwide. In this study we found that Latino and Black households consistently experience warmer summers, as measured by near-surface air temperature within the same county across 13 states of the northeast and mid-Atlantic United States. Present-day residential segregation represents a likely pathway. Segregation often overlaps with other place-based inequalities, so hotter summers for households burdened with poor housing quality, poverty, and health disparities indicate limited adaptive capacity for ethnoracial minorities and the need for housing and energy policy interventions.

**Main Text**

**Introduction**

Climate change has increased concern over warm season temperatures due to the health effects of extreme temperature exposures (1). In the United States, a small body of literature documents disparities in summertime temperatures experienced by ethnoracially minoritized subgroups (2, 3). This is a concern for several reasons. Minoritized subgroups have increased temperature vulnerability (4, 5), which may be due to higher levels of comorbidities (6–8). Further, air temperature estimates may systematically underestimate true exposures for vulnerable populations (9, 10), so the effects of these exposures also may be underestimated. In addition, minoritized populations experience higher energy burden and energy insecurity (11, 12). Given that air conditioning is the dominant individual-level adaptive strategy to heat in the U.S., minoritized populations may have reduced adaptive capacity to climate change (13, 14). Thus extreme temperature exposures represent a form of underestimated structural racism in climate impacts.

Social disadvantage and temperature exposure are related. Spatial temperature profiles are correlated with socioeconomic status, and land use/cover associated with the urban heat island effect is more prevalent in poor communities and communities of color (3, 15). More recently, Hoffman et al. identified an association between the Home Owners' Loan Corporation historical housing practices, often referred to as "redlining", and higher summertime land surface temperatures within cities (2). This work is bolstered by evidence that places benefiting from these lending practices now have attributes associated with lower summertime temperatures, such as increased vegetation and less impervious surfaces (16, 17). However, populations and developed land area have grown substantially since the 1940s. Thus while previous studies are integral to understand the formation of environmental inequities, present-day interventions would benefit from analyses associating current measures of segregation with air temperature. Further, many previous studies have used land surface temperatures, but near-surface air temperatures measured at ground-based monitors are more relevant for human thermoregulatory capacity and energy policy.

In the U.S., there is growing recognition of energy burden and energy insecurity (18, 19), yet few policies that explicitly protect disadvantaged groups that experience these hardships (20). The Low-Income Home Energy Assistance Program (LIHEAP) is one such protection program, but it largely distributes funding based on cold season energy demand and provides less support for the warm season (21). Indirect measures of energy demand are used to calculate LIHEAP funding allocations, specifically heating and cooling degree days. Although allocations assume that everyone in a state or region is exposed to roughly equal temperatures, residential segregation and temperature both vary over fine spatial scales. Thus air temperature disparities

may mean that minoritized groups are exposed to systematically higher summer temperatures. Here we used a fine spatial resolution air temperature product to assess the potential relationship between area-level ethnoracial composition and warm season air temperatures. We asked: *Do minoritized groups experience hotter summers than the area average, and do non-Hispanic white people experience cooler summers?* We then tested whether present-day measures of segregation are associated with summertime temperatures, exploring a likely pathway between exposure and temperature vulnerability.

## Results

We examined the relationships of ethnoracial composition, residential segregation, and air temperature in 13 states of the northeast and mid-Atlantic U.S., in addition to Washington D.C., using a metric of excess heat relevant to energy policy and planning for cooling needs: cooling degree days (CDDs; i.e., the sum of degrees Fahrenheit >65 °F for each daily mean temperature) in 2003–2019. This region represents 17,733 census tracts nested in 434 counties, representing approximately 72.4 million people according to the 2010 decennial census.

### *Descriptive analysis*

Ethnoracial groups lived in varying concentrations throughout the study region—white people were overrepresented in New England (cooler climate), and Black people were overrepresented in the South Atlantic (warmer climate). Therefore, we first examined within-region CDDs by state and year. For example, **Figure S1** shows distinct temperature profiles by ethnoracial groups in some states. We explored the cooling season of 2010, the warmest in the dataset. In New York state, white people experienced a wide distribution of exposures for which the mean was considerably lower than for Asian, Black, and Latino people. Other states showed more overlapping distributions but distinct bimodal patterning, as in West Virginia, potentially due to urban/rural development patterns. **Table S1** shows population-weighted statewide CDDs per year.

### *Within-state disparities*

To examine state-level air temperature differences, we conducted regressions stratified by state or district in 2003–2019, with fixed effects for county and year and standard errors constructed using Conley equations to account for spatial dependence, calculated based on tract centroids (**Table 1** ). Asian people experienced notably higher temperature exposures in nine states compared to the county-level average, except for Delaware where CDDs were lower (mean difference [MD]: -20.4; 95% CI: -36.5, -4.3). Maryland, New Jersey, Virginia, and Washington D.C. showed no difference for Asian people, and the highest exposure difference was in Rhode Island (MD: 27.0; 95% CI: 11.4, 42.7). Black people experienced higher CDDs in all states except Maryland and Washington D.C.; the largest exposure difference was in Rhode Island (MD: 43.6; 95% CI: 30.3, 56.9). Latino people experienced higher CDDs in all states except for Washington D.C.; the largest exposure difference was in Rhode Island (MD: 52.1; 95% CI: 37.3, 66.9). White people experienced fewer CDDs in all states except Maryland and Washington D.C.; the largest exposure difference was in New Jersey (MD: -17.0; 95% CI: -21.2, -12.8). Model predictions are visualized in**Figure 1** , which plots the results for one city from each census region. Results showed systematically higher temperature exposures for Black and Latino people, varying results for Asian people, and exposures consistently at or below the county averages for white people.

### *Regional residential segregation*

Numerous measures of residential segregation are aligned with the various dimensions of segregation. We explored the relationship between a measure of concentration calculated from a Gaussian kernel smoothing of ethnoracial composition at the census tract level within counties. Analyses were stratified by census regions (**Figure 2** ). Residential segregation measures varied substantially by subgroup, with the smallest range of values for Asian people and the largest range for white people. In New England, we observed a roughly monotonic increase in the number of CDDs with higher concentrations of Asian and Latino people. The relationship was non-monotonic for Black people, with an increase until an inflection point, and then fewer CDDs in the highest concentration tracts. In the mid-Atlantic, higher concentrations of Black and Latino

people were associated with higher CDDs. In the South Atlantic, a non-monotonic relationship was apparent for Black people, with concentrations of 25%–75% Black having higher CDDs, and confidence intervals >90% crossing zero, meaning that the most segregated areas for Black people were no different than their county average. A positive association was roughly monotonic for Latino people. While non-monotonic in the South Atlantic, negative associations were apparent in all three regions with high concentrations of white people. There was some heterogeneity in the shape of the associations based on analyses by state (**Figures S2–S4** ).

*Sensitivity analyses*

In a sensitivity analysis we reran our segregation analyses using the index of concentration at the extremes as our segregation measure, where the "disadvantaged" group included Asian, Black, and Latino people as well as everyone who did not identify as white; the "privileged" group included white people (**Figure S5** ). Temperatures tended to increase slightly moving from -1 (most disadvantaged people) and approaching 0 (least segregation). In all cases, moving from 0 to 1 (most white) was associated with fewer CDDs, consistent with our main analysis. We also reran our main results using varying degrees of freedom, which did not change our interpretations.

## Discussion

We used a fine resolution geostatistical prediction model to understand potential disparities in warm season ambient air temperature exposures for ethnoracially minoritized groups using measures of residential segregation in 2003–2019. We found that non-Hispanic Black and Latino people were consistently exposed to hotter warm seasons than would be expected if simply using the county average. Asian people also tended to experience higher average temperatures, but with a notable exception in Delaware where they experienced cooler summers. Finally, white people consistently tended to experience cooler warm seasons compared to the county averages. Spatial segregation analyses suggested that higher levels of minoritized groups are associated with hotter warm seasons and, relatedly, higher concentrations of white people are associated with cooler warm seasons. These results have potential ramifications for climate, health, and energy research and policy.

Many temperature-related studies have used coarse exposure assessments, often averaged to the county level or some other large administrative scale. However, we found systematically different exposure profiles for populations within counties in this region. This represents potential differential exposure misclassification, which could be a concern for temperature exposure and epidemiology studies if not properly addressed. With regard to policy, energy poverty alleviation programs like LIHEAP use statewide CDDs among measures of energy demand, implicitly assuming that all subgroups in the state are exposed to the same magnitude of season. Yet we found that throughout the study region, minoritized groups were exposed to higher warm season temperatures. Therefore, energy demand is likely underestimated for these populations. This is bolstered by other evidence that Black people are most likely to experience energy insecurity nationwide year-round (11), but Latino people experience the highest rates of warm season energy insecurity (12). Given that we found some of the most prominent warm season temperatures for Latino people, this may be one possible pathway explaining this energy insecurity disparity. Finally, other studies have found that historical redlining is associated with present-day land cover characteristics associated with the urban heat island effect (16, 17). Those studies were limited to cities with documented redlining, but our study covers entire states in our region using present-day measures of segregation and air temperature. Given that we found statewide temperature disparities beyond just formerly redlined cities, this suggests that the historical processes associated with land use and land cover disparities were not restricted to redlined cities. This is an area for future investigation but speaks to the potential importance of targeted greening initiatives in urban areas nationwide.

Our study has many strengths. We leveraged fine-resolution air temperature predictions to understand temperature disparities, while many other studies have used either coarser models or land surface temperatures, which we believe more accurately reconstruct neighborhood-level exposures. We also used metrics and me-

thods to enhance interpretability and policy relevance—our use of weighted effect coding provides an intuitive comparison of the ethnoracial average compared to the county average, and our use of CDDs is relevant to U.S. energy policy. Our study also covered a large region of more than 72 million people and more than 17 years. Finally, our data and programmatic code are freely available to enhance reproducibility of our findings.

Nonetheless, our study also has limitations. First, we examined temperature exposures as assigned by residential location, but there may be exposure disparities due to occupational exposures. Second, all confidence intervals for Washington D.C. crossed zero, and thus we concluded there were no apparent disparities. However, it is possible that our spatial inference methods were overly conservative for a small spatial area. Third, we examined two measures of residential segregation based on concentration, but residential segregation is a multidimensional construct (22). Relatedly, we examined segregation measures concurrent to temperature exposures. While these and other research suggest that residential segregation is a pathway towards temperature exposure disparities, we need to examine the sequence of segregation and later land development and temperature to make those conclusions. Further, while we adjusted for year, we did not assess time trends in these results across the study period. Finally, our temperature prediction model covered the northeast and mid-Atlantic regions of the U.S., but future studies should examine other regions or the entire U.S.

Overall, our findings suggest that ethnoracially minoritized groups experienced hotter average summers across the 13 states in the Northeastern U.S. in 2003–2019. This study highlights the importance of energy poverty alleviation programs, specifically for these subgroups. These findings are critical to target interventions that enhance the adaptive capacity of systematically marginalized groups in a warming climate.

## Materials and Methods

*Demographic data*

All demographic data were accessed using the *tidycensus* package in R (23). Decennial census data were assigned based on year: 2003–2009 data were assigned to the 2000 Census, and 2010–2019 data were assigned to the 2010 Census. Ethnoracial groups were constructed using the following categories: non-Hispanic Asian alone (Asian), non-Hispanic Black alone (Black), Hispanic/Latino of any race (Latino), and non-Hispanic white alone (white). Census population-weighted tract centers of population were drawn from the census as well (24).

*Air temperature model and areal interpolation*

Air temperature predictions were drawn from a geostatistical model detailed previously (9). Briefly, we trained a machine learning model (XGBoost) on ground station data aggregated by the National Oceanic and Atmospheric Administration (NOAA) Meteorological Assimilation Data Ingest System (MADIS). The specific MADIS dataset we used was the National Mesonet, with >4,000 weather stations across the region. Predictors in the model included land surface temperature from the MODIS instrument on the Aqua and Terra satellites, an inverse-distance weighting interpolation of air temperature, enhanced vegetation index, landcover and landform characteristics, elevation, a topological position index, and temporal terms for seasonality. Careful attention was paid to avoid overfitting with spatial cross-validation methods, and the model was assessed and validated against an independent external dataset. The resulting model was approximately 1-km grid cells and an hourly timestep. For comparison, air temperature predictions from this model have a root mean square error of 1.6 Kelvin when compared to ground observations, whereas the North American Land Data Assimilation System-2 (NLDAS-2) model, used in the Centers for Disease Control and Prevention Heat and Health Tracker, has a root mean square error of 2.5 Kelvin.

An areal interpolation procedure was designed to align temperature data with census tracts. First, 1-km prediction cells were reprojected to align with NASA's Gridded Population of the World version 4.11 data product for 2000 and 2010 (25). We then used the *exactextractr* package in R to calculate area-weighted mean population density values in each cell (26). Finally, we used the population density and coverage area of each prediction cell to weight temperature predictions in each cell and compute weighted average by each tract.

The result was a population-weighted areal interpolation of temperature within each census tract.

Air temperature data were transformed to create CDDs, restricted to predictions in May–September of each year. We used Fahrenheit instead of Celsius to align with energy policies in the U.S. CDD calculations were adapted based on NOAA methods:

$$Cooling\ degree\ days = \sum_i^n \{ \begin{matrix} \left(\frac{\mathrm{Tmax}_i - \mathrm{Tmin}_i}{2}\right) - 65, if > 0, & otherwise \\ 0 \end{matrix} \quad (1)$$

where *Tmax* was the maximum hourly temperature and *Tmin* was the minimum on day *i,* summed over *n* days in May–September.

*Constructing a segregation index*

We leveraged a spatial and localizable segregation measure based on census tract data for each ethnoracial group, assigned to the mean population-weighted centroids of each tract (27, 28). Each county was projected to the appropriate state-plane projection system, and then we conducted a Gaussian kernel density smoothing procedure for the total population using the *spatstat* package in R (29). We used the same bandwidth to calculate the smooth for each subgroup. Each kernel-smoothed ethnoracial subgroup value was divided by the corresponding value for the total population to create a concentration measure per tract. Each value was multiplied by 100 to create a percent value. This approach relies on multiple tracts per county, which was not true for all counties, specifically in Virginia and West Virginia. When we encountered such errors, we merged the county with its nearest neighbor based on centroid distances.

*Statistical analyses*

The first analysis was to calculate average difference between summer temperatures experienced by ethnoracial group and the county average. To calculate mean difference per ethnoracial group, we used fixed effects regressions. To account for spatial variation in temperature and ethnoracial composition, we used three strategies. First, regressions were stratified by census regions, which closely align with NOAA climate regions. Second, we included fixed effects for county. Finally, we used Conley variance–covariance calculations to construct standard errors to account for spatial autocorrelation in the data, calculated based on population-weighted centroids of tracts. We also included a fixed effect for the year, and we weighted our regression models by the total population of the tract. These statistical procedures were conducted using the *fixest* package in R (30). Often regression models with categorical variables like ethnoracial groups use traditional dummy coding with one referent group, typically white people. This is problematic because it makes the referent result invisible and makes one group's experience the standard, norm, or aspirational depending on context (31). To avoid this, we implemented weighted effect coding, which functionally weights each category to represent deviation from the sample mean, in this case the county-averaged temperature (32).

Our second analysis was to associate our measure of residential segregation with local air temperatures. Associating tract-level temperature with the segregation index required a flexible regression framework to accommodate nonlinearities, so we used generalized additive models with smoothing splines. The segregation measure was modeled with a natural cubic spline with three knots. We included fixed effects for county and year and accounted for spatiotemporal dependence by modeling a tensor product smooth of the geocoordinates of population-weighted centroids by year. These regressions were implemented with the*mgcv* package in R, specifically using the *bam* function for computational efficiency in large datasets (33).

*Sensitivity analyses*

Our first sensitivity analysis was to rerun our segregation analyses with varying degrees of freedom for the natural cubic splines and inspect differences in the resulting shapes of the smooths. We then reran our segregation analyses with an alternative segregation measure, the index of concentration at the extremes (34, 35). This is a measure of social polarization used in numerous environmental, social, and health science

studies to assess the impacts of spatial social stratification. The index of concentration at the extremes was calculated as:

$$\text{ICE}_i = (D_i - A_i)/\text{Total}_i \quad (2)$$

where $D$ is the number of disadvantaged people, $A$ is the number of advantaged people, and *Total* is the total population of census tract $i$ . We repeated the regression four times, using Asian, Black, and Latino alone as the disadvantaged groups, and then all except white as the disadvantaged group.

*Computational environment and reproducibility*

All analyses were conducted in R version 4.3.0. To enhance reproducibility we used the *targets* and *renv* packages (36, 37). The *targets* package is a workflow and pipeline tool, while the *renv package* tracks the versioning of software dependencies. All code are available on GitHub (https://github.com/justlab/northeast_temperature_disparities). Data are downloaded automatically throughout the code, including for our temperature predictions, which can be found at https://doi.org/10.5281/zenodo.10557980.

**Acknowledgments**

DC was supported by NICHD T32HD049311 and NASA 80NSSC22K1666. ACJ was funded by NIEHS R01ES031295. EC was supported by NIEHS R01ES032242. We wish to acknowledge Xuezhixing Zhang for his help making the codebase reproducible.

**References**

1. Q. Zhao, *et al.* , Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study. *The Lancet Planetary Health***5** , e415–e425 (2021).

2. J. S. Hoffman, V. Shandas, N. Pendleton, The effects of historical housing policies on resident exposure to intra-urban heat: A study of 108 US urban areas. *Climate* **8** , 12 (2020).

3. S. A. Benz, J. A. Burney, Widespread race and class disparities in surface urban heat extremes across the United States. *Earth's Future* **9** , e2021EF002016 (2021).

4. J. Madrigano, K. Ito, S. Johnson, P. L. Kinney, T. Matte, A case-only study of vulnerability to heat wave–related mortality in New York City (2000–2011). *Environmental Health Perspectives* **123** , 672–678 (2015).

5. M. Manware, R. Dubrow, D. Carrión, Y. Ma, K. Chen, Residential and race/ethnicity disparities in heat vulnerability in the United States. *GeoHealth* **6** , e2022GH000695 (2022).

6. A. G. Berberian, D. J. Gonzalez, L. J. Cushing, Racial disparities in climate change-related health effects in the United States. *Current Environmental Health Reports* **9** , 451–464 (2022).

7. J. Moon, The effect of the heatwave on the morbidity and mortality of diabetes patients; a meta-analysis for the era of the climate crisis. *Environmental Research* **195** , 110762 (2021).

8. R. V. Remigio, *et al.* , Association of Extreme heat events with hospital admission or mortality among patients with end-stage renal disease. *JAMA Network Open* **2** , e198904–e198904 (2019).

9. D. Carrión, *et al.* , A 1-km hourly air-temperature model for 13 northeastern US states using remotely sensed and ground-based measurements. *Environmental Research* , 111477 (2021).

10. C. J. Gronlund, Racial and socioeconomic disparities in heat-related health effects and their mechanisms: a review. *Current Epidemiology Reports* **1** , 165–173 (2014).

11. D. Hernández, Y. Jiang, D. Carrión, D. Phillips, Y. Aratani, Housing hardship and energy insecurity among native-born and immigrant low-income families with children in the United States. *Journal of Children and Poverty* **22** , 77–92 (2016).

12. M. Graff, D. M. Konisky, S. Carley, T. Memmott, Climate change and energy insecurity: a growing need for policy intervention.*Environmental Justice* (2021).

13. A. A. Williams, J. D. Spengler, P. Catalano, J. G. Allen, J. G. Cedeno-Laurent, Building vulnerability in a changing climate: indoor temperature exposures and health outcomes in older adults living in public housing during an extreme heat event in Cambridge, MA.*International Journal of Environmental Research and Public Health***16** (2019).

14. K. Lane, *et al.* , Extreme heat and COVID-19 in New York City: an evaluation of a large air conditioner distribution program to address compounded public health risks in summer 2020. *Journal of Urban Health* , 1–13 (2023).

15. A. Buyantuyev, J. Wu, Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology***25** , 17–33 (2010).

16. L. H. Schinasi, *et al.* , Associations Between historical redlining and present-day heat vulnerability housing and land cover characteristics in Philadelphia, PA. *Journal of Urban Health***99** , 134–145 (2022).

17. A. Nardone, K. E. Rudolph, R. Morello-Frosch, J. A. Casey, Redlines and greenspace: the relationship between historical redlining and 2010 greenspace across the United States. *Environmental Health Perspectives* **129** , 017006 (2021).

18. D. Hernández, Energy insecurity and health: America's hidden hardship. *Health Affairs* (2023) https:/doi.org/10.1377/hpb20230518.472953 (July 27, 2023).

19. D. Hernández, Understanding 'energy insecurity' and why it matters to health. *Social Science & Medicine* **167** , 1–10 (2016).

20. D. J. Bednar, T. G. Reames, Recognition of and response to energy poverty in the United States. *Nature Energy* **5** , 432–439 (2020).

21. L. Perl, "The LIHEAP Formula" (Congressional Research Service, 2019) (July 11, 2023).

22. D. S. Massey, N. A. Denton, The dimensions of residential segregation. *Social Forces* **67** , 281–315 (1988).

23. K. Walker, tidycensus: Load US Census boundary and attribute data as "tidyverse" and 'sf'-ready data frames (2020).

24. Bureau of the Census, Centers of Population Computation for the United States 1950-2010 (2011) (December 3, 2023). Accessed from: https://www2.census.gov/geo/pdfs/reference/cenpop2020/COP2020_documentation.pdf.

25. Center for International Earth Science Information Network, Columbia University, Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11 (2018).

26. D. Baston, exactextractr: Fast extraction from raster datasets using polygons. *R package version 0.5. 0* (2020).

27. D. O'Sullivan, D. W. Wong, A surface-based approach to measuring spatial segregation. *Geographical Analysis* **39** , 147–168 (2007).

28. A. M. Planey, S. C. Grady, R. Fetaw, S. L. McLafferty, Spaces of segregation and health: complex associations for Black immigrant and US-born mothers in New York City. *Journal of Urban Health***99** , 469–481 (2022).

29. A. Baddeley, R. Turner, Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12** , 1–42 (2005).

30. L. Berge, "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm" (Department of Economics at the University of Luxembourg, 2018).

31. S. S. Johfre, J. Freese, Reconsidering the reference category. *Sociological Methodology* **51** , 253–269 (2021).

32. R. Nieuwenhuis, M. te Grotenhuis, B. Pelzer, Weighted effect coding for observational data with wec. *The R Journal* **9** , 477 (2017).

33. S. N. Wood, *Generalized additive models: an introduction with R* (CRC press, 2017).

34. J. M. Feldman, P. D. Waterman, B. A. Coull, N. Krieger, Spatial social polarisation: using the index of concentration at the extremes jointly for income and race/ethnicity to analyse risk of hypertension. *Journal of Epidemiology and Community Health* **69** , 1199 (2015).

35. N. Krieger, P. D. Waterman, A. Gryparis, B. A. Coull, Black carbon exposure, socioeconomic and racial/ethnic spatial polarization, and the Index of concentration at the extremes (ICE). *Health & Place***34** , 215–228 (2015).

36. W. Landau, The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Social Structure***6** , 2959 (2021).

37. K. Ushey, Renv: Project Environments. R package version 0.10. 0 (2020).

**Figures and Tables**

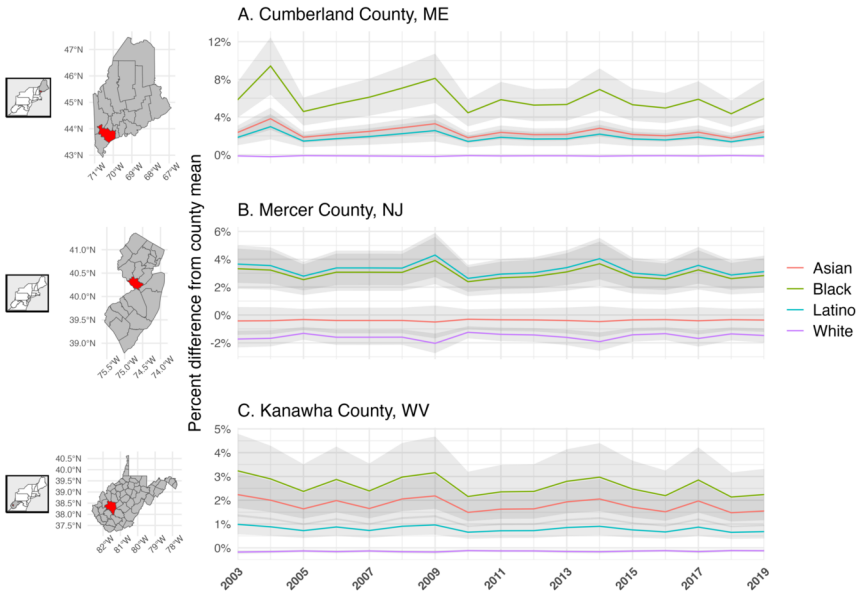


**Figure 1.** Temperature predictions from state-specific temperature disparity regressions. Maps identify the location of the counties, with corresponding time series of predictions. The y axis indicates percent difference between the county average temperature and the ethnoracial subgroup's prediction, with 95% confidence intervals over the study period (2003–2019).

**Hosted file**

`image2.emf` available at https://authorea.com/users/545090/articles/717937-residential-segregation-and-summertime-air-temperature-across-13-northeastern-u-s-states-potential-implications-for-energy-burden

**Figure 2.** Regression results for the association between ethnoracial concentration and cooling degree days (CDDs). Solid black lines indicate partial effects from a natural cubic spline with three knots, and dashed lines indicate 95% confidence intervals. All models are adjusted with fixed effects for county and year as well

as a tensor product smooth of the latitude and longitude of the census tract's population-weighted centroid by year. Binned hexagons of partial residuals are depicted as light blue (low values) to charcoal (high values).

**Table 1.** Results from fixed effect regression analyses: adjusted mean difference in cooling degree days (95% confidence intervals) from the population mean. Each row represents one regression, and columns are coefficients from weighted effect coding of tract-level ethnoracial composition. Regression models included fixed effects for county and year, and standard errors constructed based on spatially dependent observations. Bolded values indicate confidence intervals that do not cross or include zero.

| | Asian | Black | Latino | White |
|---|---|---|---|---|
| Connecticut | **13.0 (5.7, 20.3)** | **40.0 (31, 49)** | **45.1 (35.7, 54.6)** | **-14.7 (-17.5, -11.8)** |
| Delaware | **-20.4 (-36.5, -4.3)** | **31.6 (14.3, 49)** | **18.7 (7, 30.4)** | **-11.0 (-16.8, -5.2)** |
| Maine | **8.8 (6.5, 11.1)** | **21.6 (12.8, 30.4)** | **6.8 (3.9, 9.7)** | **-0.5 (-0.7, -0.3)** |
| Maryland | -2.9 (-9.9, 4.1) | 4.8 (-3.2, 12.7) | **16.1 (6.9, 25.2)** | -4.4 (-8.8, 0) |
| Massachusetts | **22.4 (16.4, 28.4)** | **22.1 (5.4, 38.8)** | **38.0 (24.6, 51.4)** | **-8.5 (-11.1, -5.9)** |
| New Hampshire | **30.0 (19.1, 41)** | **37.4 (26.8, 48.1)** | **46.1 (29.1, 63.1)** | **-2.3 (-2.8, -1.7)** |
| New Jersey | -4.2 (-12.3, 3.9) | **32.7 (21, 44.3)** | **35.9 (25.3, 46.6)** | **-17 (-21.2, -12.8)** |
| New York | **8.2 (1.3, 15)** | **17.4 (9.2, 25.7)** | **17.2 (10.7, 23.7)** | **-10.7 (-14.2, -7.2)** |
| Pennsylvania | **17.5 (7.8, 27.2)** | **29.2 (13.6, 44.8)** | **47.4 (34.5, 60.2)** | **-7.7 (-10.1, -5.2)** |
| Rhode Island | **27.0 (11.4, 42.7)** | **43.6 (30.3, 56.9)** | **52.1 (37.3, 66.9)** | **-11.9 (-15.3, -8.5)** |
| Vermont | **23.2 (14, 32.3)** | **29.8 (20.4, 39.1)** | **13.1 (8.5, 17.7)** | **-0.8 (-1, -0.5)** |
| Virginia | 2.4 (-4.2, 9) | **9.1 (6.1, 12.1)** | **11.3 (7.5, 15.1)** | **-4.2 (-5.2, -3.2)** |
| Washington D.C. | 10.3 (-13.2, 33.8) | 4.1 (-8.5, 16.7) | 1.1 (-14.3, 16.5) | -9.6 (-32.1, 12.9) |
| West Virginia | **21.0 (12.5, 29.5)** | **30.3 (17.4, 43.3)** | **9.2 (5.9, 12.6)** | **-1.7 (-2.3, -1.1)** |

# Supplementary Information for

Residential segregation and summertime air temperature across 13 northeastern U.S. states: Potential Implications for energy burden

Daniel Carrión, Johnathan Rush, Elena Colicino, and Allan C. Just

**Corresponding author:**
Daniel Carrión
**Email:** daniel.carrion@yale.edu

**This PDF file includes:**
Supplementary table 1
Supplementary Figs 1 - 5

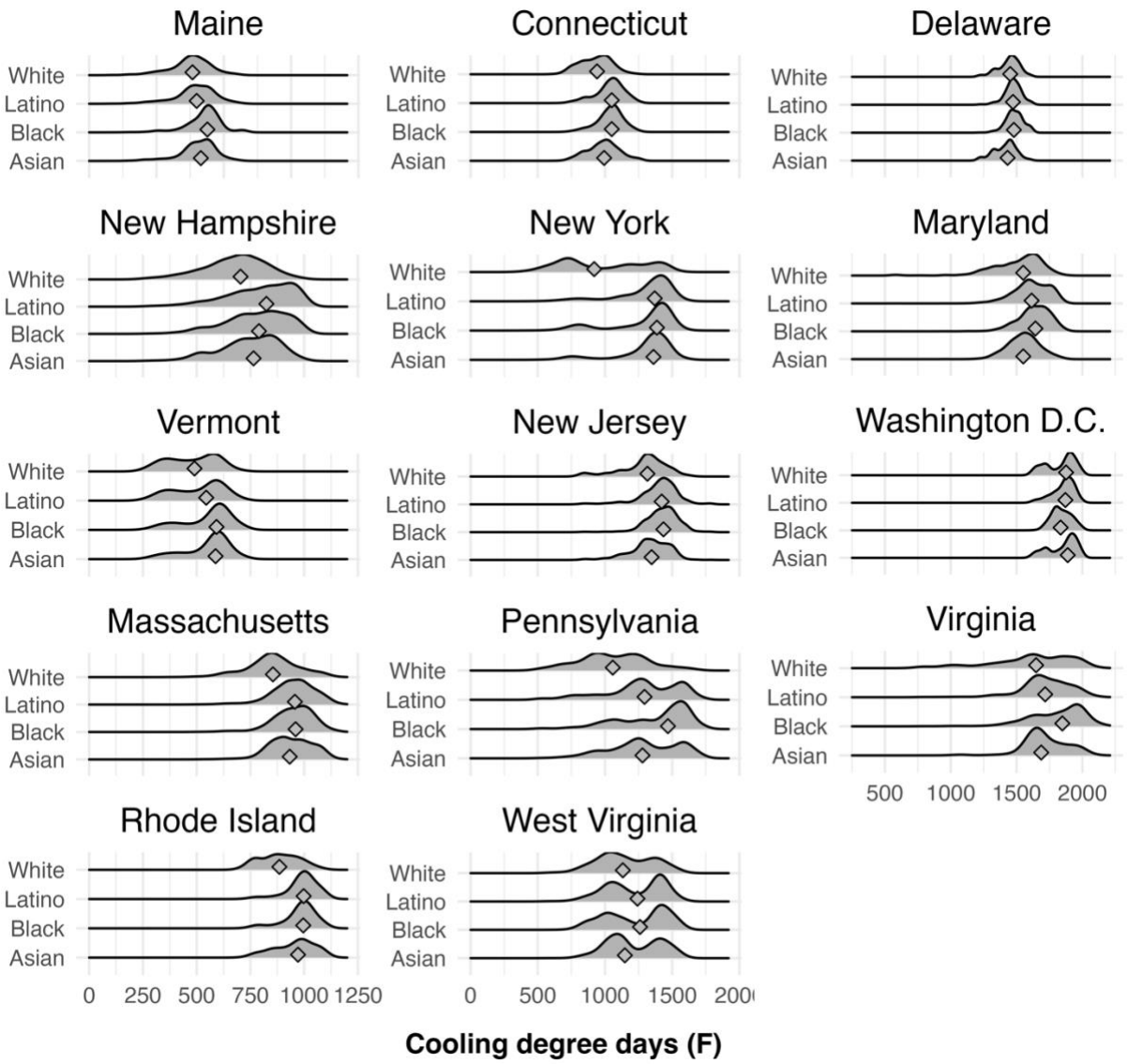


**Figure S1:** Kernel density distributions of census tract level cooling degree days by ethnoracial group and by state in 2010. Columns group states with similar cooling degree day ranges.

**Table S1:** Population-weighted cooling degree days from the XGBoost spatiotemporal prediction model per state per year.

| STATE | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CT** | 684 | 628 | 926 | 756 | 757 | 715 | 527 | 962 | 856 | 825 | 782 | 589 | 831 | 915 | 696 | 926 | 758 |
| **DE** | 1030 | 1166 | 1324 | 1158 | 1206 | 1090 | 957 | 1457 | 1335 | 1286 | 1087 | 950 | 1286 | 1355 | 1113 | 1363 | 1355 |
| **DC** | 1178 | 1411 | 1583 | 1469 | 1656 | 1441 | 1296 | 1846 | 1661 | 1547 | 1285 | 1221 | 1544 | 1625 | 1332 | 1591 | 1658 |
| **ME** | 330 | 187 | 423 | 354 | 311 | 263 | 227 | 437 | 322 | 367 | 359 | 270 | 358 | 388 | 319 | 449 | 313 |
| **MD** | 1068 | 1247 | 1334 | 1228 | 1389 | 1183 | 1078 | 1565 | 1422 | 1328 | 1128 | 998 | 1301 | 1409 | 1141 | 1370 | 1410 |
| **MA** | 642 | 530 | 809 | 698 | 650 | 601 | 457 | 876 | 751 | 735 | 719 | 555 | 758 | 829 | 645 | 849 | 676 |
| **NH** | 463 | 337 | 620 | 524 | 493 | 408 | 331 | 677 | 556 | 570 | 543 | 391 | 555 | 601 | 467 | 640 | 492 |
| **NJ** | 954 | 983 | 1255 | 1035 | 1031 | 1037 | 806 | 1341 | 1198 | 1155 | 1023 | 859 | 1164 | 1239 | 979 | 1222 | 1128 |
| **NY** | 794 | 768 | 1088 | 890 | 893 | 871 | 632 | 1105 | 1019 | 1003 | 887 | 732 | 968 | 1074 | 826 | 1068 | 894 |
| **PA** | 742 | 790 | 1086 | 880 | 915 | 819 | 652 | 1112 | 1029 | 993 | 836 | 704 | 950 | 1102 | 809 | 1053 | 965 |
| **RI** | 714 | 558 | 869 | 757 | 727 | 680 | 496 | 907 | 813 | 767 | 747 | 578 | 816 | 864 | 667 | 904 | 745 |
| **VT** | 315 | 202 | 506 | 379 | 327 | 264 | 226 | 462 | 392 | 423 | 407 | 278 | 385 | 440 | 314 | 506 | 323 |
| **VA** | 1155 | 1280 | 1426 | 1287 | 1455 | 1272 | 1207 | 1676 | 1500 | 1377 | 1180 | 1168 | 1423 | 1508 | 1284 | 1544 | 1556 |
| **WV** | 686 | 798 | 1032 | 808 | 1025 | 777 | 711 | 1154 | 1043 | 1032 | 835 | 773 | 976 | 1139 | 821 | 1181 | 1111 |

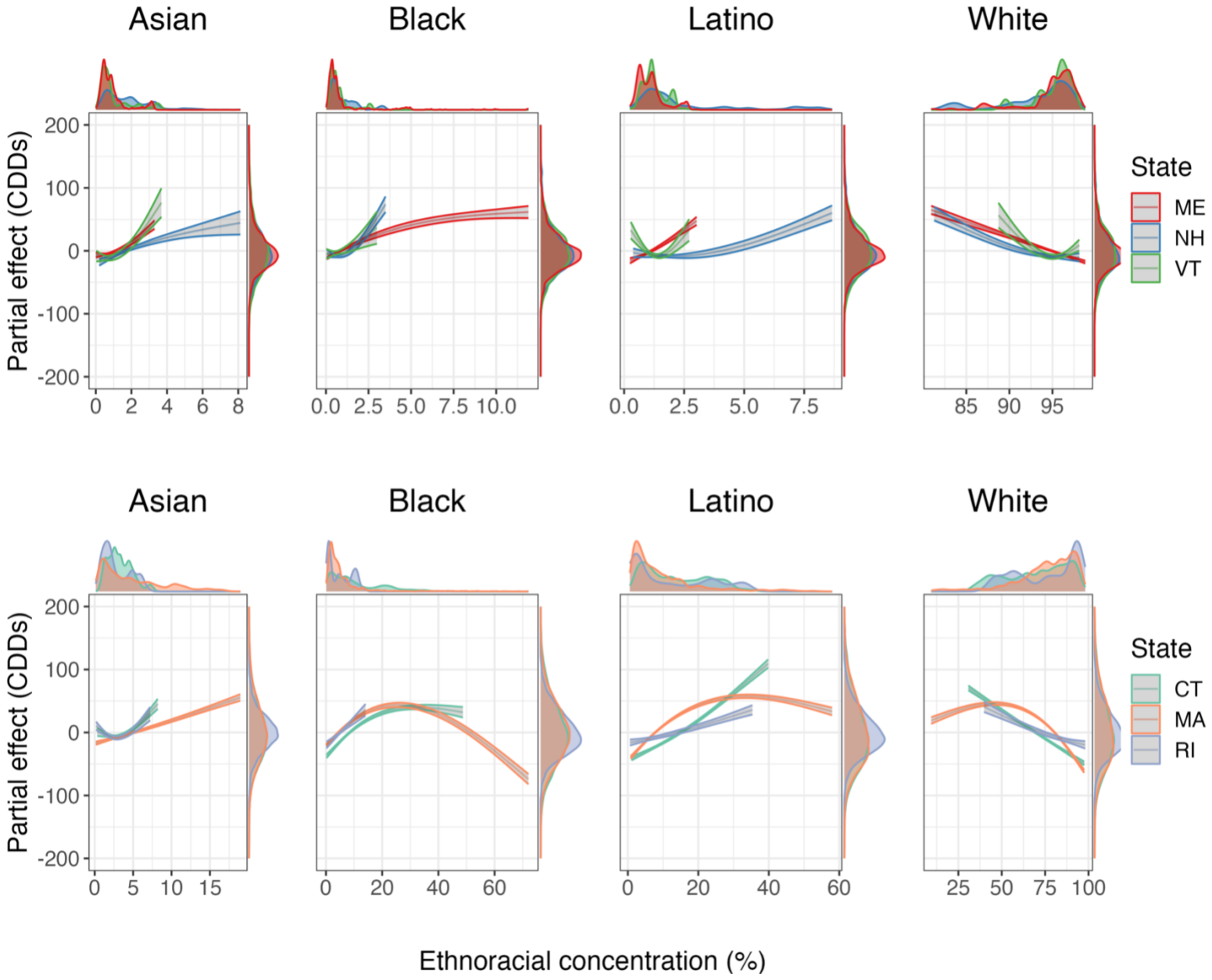


**Figure S2:** Regression results for the association between ethnoracial concentration and cooling degree days (CDDs) per state. Solid lines indicate partial effect from a natural cubic spline with three knots, and shaded areas are 95% confidence intervals. All models are adjusted with fixed effects for county and year as well as a tensor product smooth of the latitude and longitude of the census tract's population-weighted centroid by year.

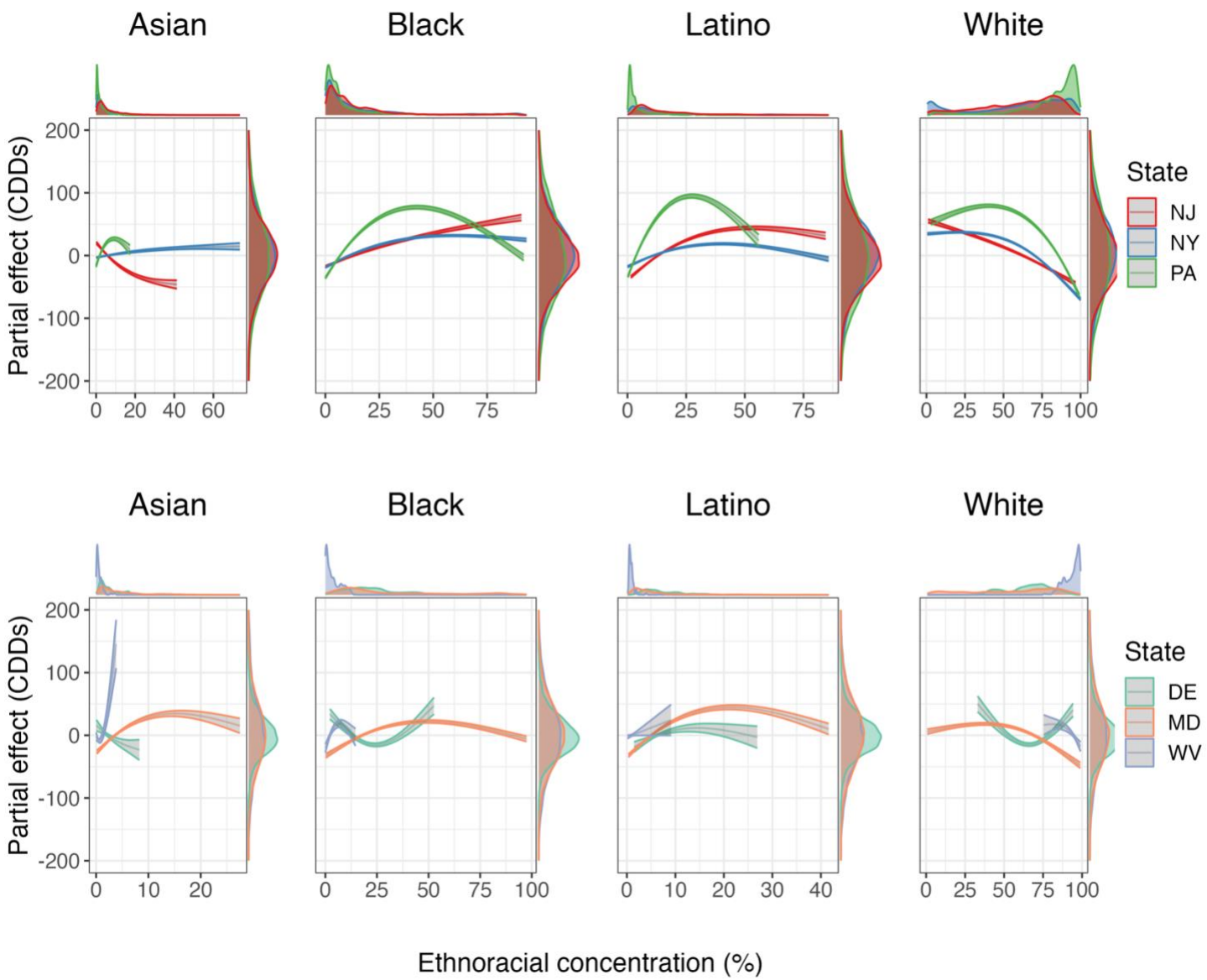


**Figure S3:** Regression results for the association between ethnoracial concentration and cooling degree days (CDDs) per state. Solid lines indicate partial effect from a natural cubic spline with three knots, and shaded areas are 95% confidence intervals. All models are adjusted with fixed effects for county and year as well as a tensor product smooth of the latitude and longitude of the census tract's population-weighted centroid by year.

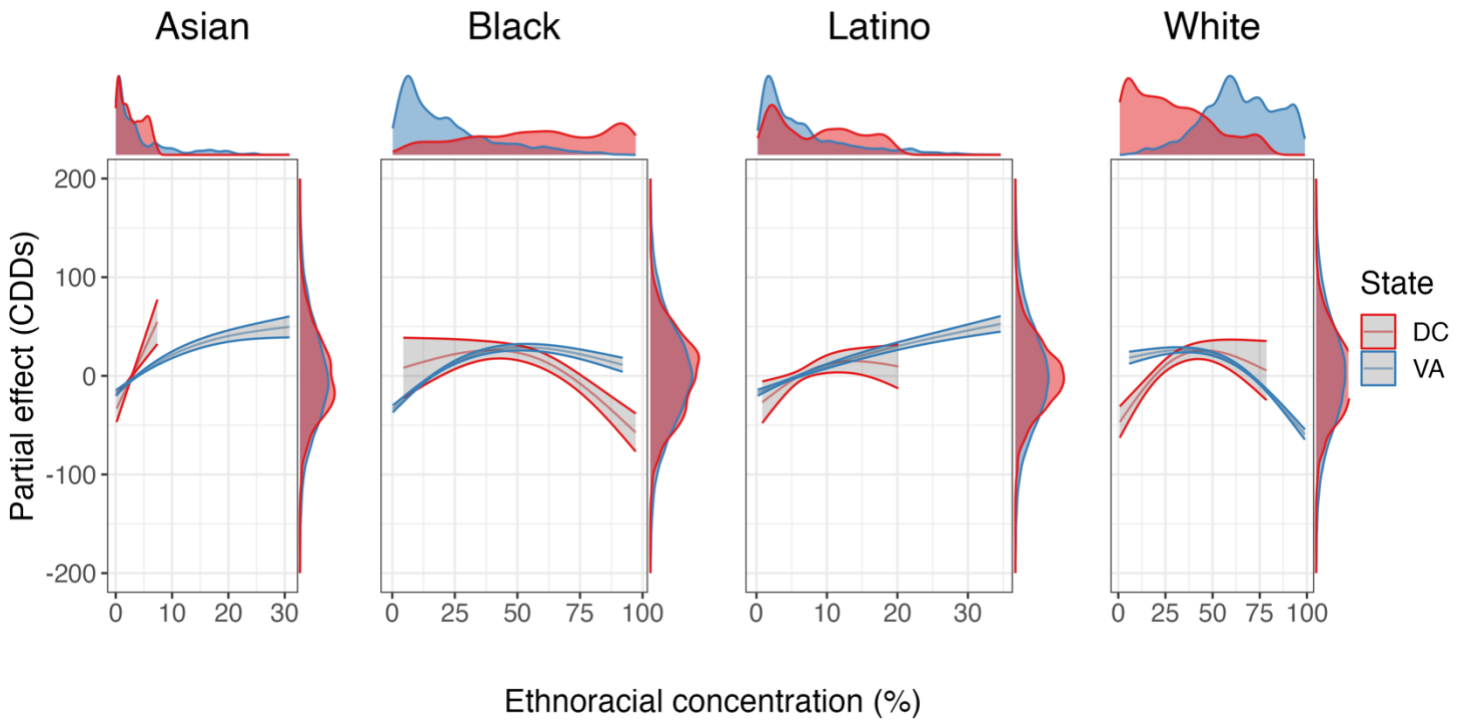


**Figure S4:** Regression results for the association between ethnoracial concentration and cooling degree days (CDDs) per state. Solid black lines indicate partial effect from a natural cubic spline with three knots, and shaded areas are 95% confidence intervals. All models are adjusted with fixed effects for county and year as well as a tensor product smooth of the latitude and longitude of the census tract's population-weighted centroid by year.

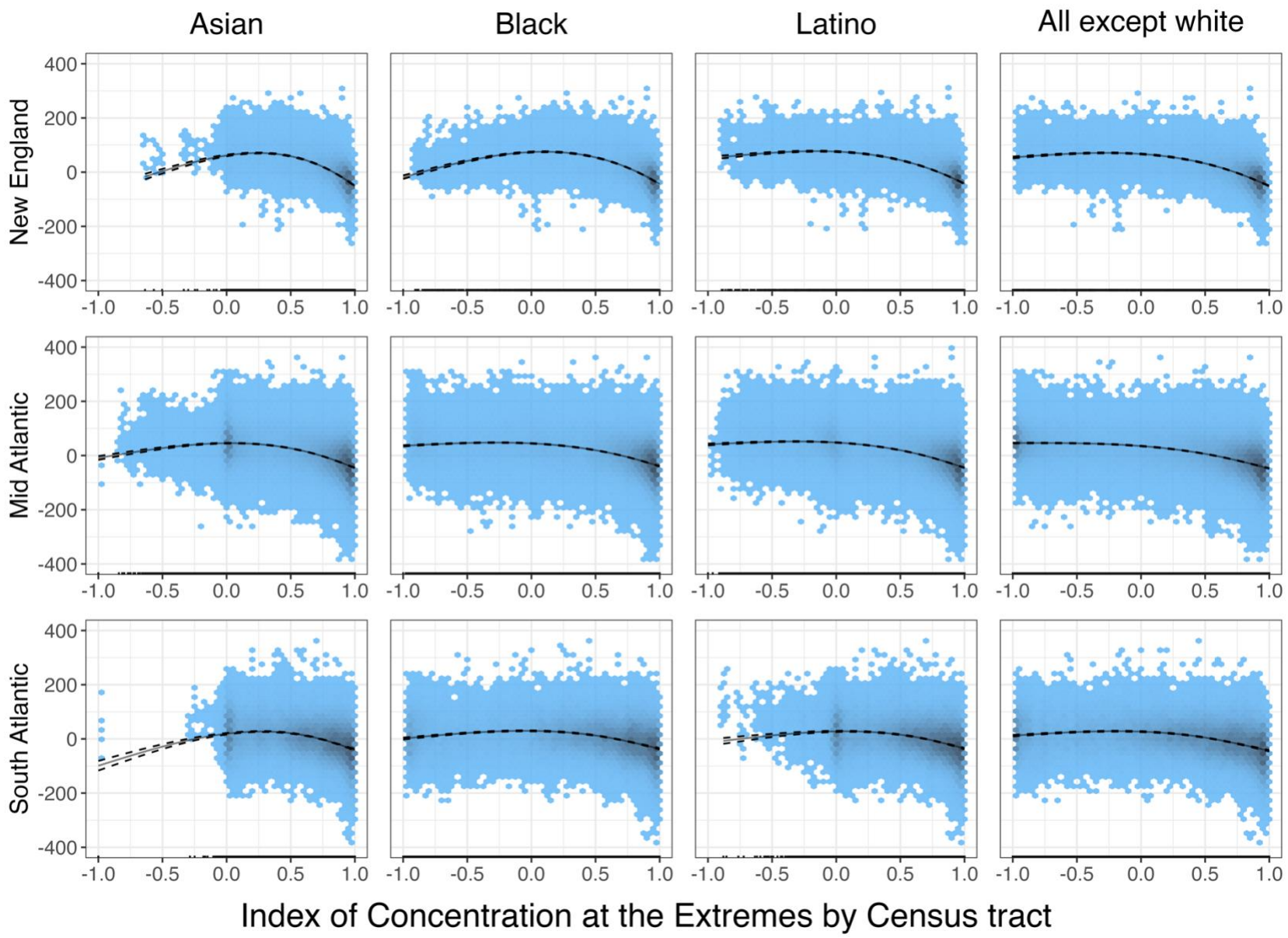


**Figure S5:** Regression results for the association between index of concentration at the extremes and cooling degree days (CDDs). Solid black lines indicate partial effect from a natural cubic spline with three knots, and dashed lines are 95% confidence intervals. All models are adjusted with fixed effects for county and year as well as a tensor product smooth of the latitude and longitude of the census tract's population-weighted centroid by year. Binned hexagons of partial residuals are depicted as light blue (low values) to charcoal (high values).