Supplementary Information

Assessing disparities in air conditioning operation across Southern California

Diego Ramos Aguilera, Yishi Wang, Lu Zhang, Jiachen Zhang, and Kelly T. Sanders

SI.1. Covariates employed in the multilinear regressions

Variable	Description (units)	Source	10th p.	Mean	Median	90th p.
Age	Age of housing (years)	COT	13	43.2	41	70
Sqft	Size of housing (square footage)	COT	1,120	1,930	1,740	2,970
Bedrooms	Number of bedrooms	COT	2	3.44	3	4
Value	Value of housing (thousand US$)	COT	146	374	303	638
MedianInc	Census tract median income (thousand US$)	ACS	45.9	81.1	77.4	120
Unemployment	Unemployed population in tract (%)	ACS	3.33	8.31	7.27	14.5
Renting	Renting population in tract (%)	ACS	11.4	32.2	29.1	59.0
LivingAlone	Households in tract with a single occupant (%)	ACS	8.08	17.6	15.8	29.6
Crowding	Households in tract with more than 1.5 occupants per room	ACS	0.00	2.18	1.04	6.13
Elderly	Population over 65 years of age in tract (%)	ACS	7.02	14.4	12.6	22.4
OutdoorWorkers	Outdoor workers in tract (%)	ACS	2.74	8.21	7.54	14.3
ForeignBorn	Foreign-born population in tract (%)	ACS	11.6	25.2	22.8	42.7
NonWhite	Tract population who does not identify as White (%)	ACS	18.5	41.4	38.9	68.3
HispanicLatino	Tract population who identifies as Hispanic or Latino (%)	ACS	13.5	41.7	36.5	77.3
maxT	Maximum daily temperature (°C)	E5L	15.0	24.4	24.5	33.5
CDH205	Daily cooling degree-hours above 20.5 °C (°C-hours)	E5L	0.0	47.0	20.6	135
meanRH	Mean daily relative humidity (%)	E5L	31.5	58.2	61.1	79.9

Table S1: Description, sources, and descriptive statistics (mean, median, and 10th and 90th percentiles) of the covariates employed in the multilinear regression. Source codes: COT = Cotality, E5L = ERA5-Land, ACS = American Community Survey.

SI.2. Descriptions of building climate zones (BCZ)

BCZ	Reference city	Description
6	Santa Monica, CA	Coastal, flat regions with mild climate year-round.
8	Gateway Cities, CA	Influenced by marine air, but warmer summers than BCZ 6.
9	Covina, CA	Influenced by both coastal and interior weather. Inland winds bring hot and dry air, while marine winds bring cool and moist air.
10	San Bernardino, CA	Interior valleys with a more extreme annual temperature swing. Heating and cooling are needed to maintain thermal comfort.
14	Lancaster, CA	Medium- to high-desert zones with wide diurnal and seasonal temperature swings. A high energy-consuming climate.
15	Palm Springs, CA	Low desert with extremely hot summers and moderately cold winters. Large diurnal temperature ranges due to very low humidity.
16	Big Bear Lake, CA	High mountainous and semiarid region with a very extreme temperature range. Very cold winters and mild summers.

Table S2: Description of building climate zones (BCZ), adapted from the California Energy Commission [68].

SI.3. Comparison of model performance with different temperature variables

Temp. var.	Dependent var.	Time of day	Adj. R^2	AIC ($\times 10^6$)	BIC ($\times 10^6$)	RMSE
maxT	Cooling load	Daytime	0.296	12.59	12.59	0.973
CDH205	Cooling load	Daytime	0.297	12.54	12.54	0.972
maxT	Cooling load	Nighttime	0.221	5.551	5.552	0.903
CDH205	Cooling load	Nighttime	0.223	5.527	5.528	0.902
maxT	Non-cooling load	Daytime	0.181	-2.482	-2.481	0.504
CDH205	Non-cooling load	Daytime	0.179	-2.282	-2.282	0.505
maxT	Non-cooling load	Nighttime	0.158	-14.43	-14.43	0.463
CDH205	Non-cooling load	Nighttime	0.158	-14.39	-14.39	0.463

Table S3: Comparison of ordinary least squares (OLS) model performance between maximum daily temperature (*maxT*) and cooling degree-hours (*CDH205*), showing adjusted R^2, AIC, BIC, and root mean squared error (RMSE) statistics.

Temp. var.	Dependent var.	Time of day	Pseudo R^2	AIC ($\times 10^6$)	BIC ($\times 10^6$)	AUC
maxT	Cooling hours	Daytime	0.286	76.77	76.77	0.842
CDH205	Cooling hours	Daytime	0.275	77.95	77.95	0.837
maxT	Cooling hours	Nighttime	0.335	51.62	51.63	0.881
CDH205	Cooling hours	Nighttime	0.332	51.79	51.79	0.880

Table S4: Comparison of binary logit model performance between maximum daily temperature (*maxT*) and cooling degree-hours (*CDH205*), showing pseudo R^2, AIC, BIC, and area under the curve (AUC) statistics.

Temp. var.	Dependent var.	Time of day	LL ($\times 10^6$)	AIC ($\times 10^6$)	BIC ($\times 10^6$)
maxT	Cooling hours	Daytime	-60.39	120.8	120.8
CDH205	Cooling hours	Daytime	-60.16	120.3	120.3
maxT	Cooling hours	Nighttime	-25.66	51.32	51.32
CDH205	Cooling hours	Nighttime	-25.53	51.07	51.07

Table S5: Comparison of ordered logit model performance between maximum daily temperature (*maxT*) and cooling degree-hours (*CDH205*), showing log-likelihood (LL), AIC, and BIC statistics.

SI.4. Results of the logistic regression model

Variable	**Delta**	ΔP_{USE}	ΔE_{POS}	ΔE_{TOT}
Age	10 years	0.0026 (0.0025, 0.0027)	0.0089 (0.0083, 0.0093)	0.0165 (0.0161, 0.0168)
Sqft	500 sqft	-0.0031 (-0.0032, -0.0030)	-0.0442 (-0.0451, -0.0435)	-0.0393 (-0.0399, -0.0387)
Bedrooms	1 bedroom	0.0014 (0.0013, 0.0016)	0.0394 (0.0384, 0.0406)	0.0296 (0.0289, 0.0305)
Value	$250 k	-0.0005 (-0.0006, -0.0004)	-0.0024 (-0.0034, -0.0015)	-0.0035 (-0.0042, -0.0028)
MedianInc	$10 k	0.0015 (0.0015, 0.0016)	-0.0142 (-0.0147, -0.0135)	-0.0019 (-0.0023, -0.0014)
Unemployment	10 pp	-0.0074 (-0.0076, -0.0071)	0.0868 (0.0848, 0.0888)	0.0188 (0.0174, 0.0203)
Renting	10 pp	-0.0030 (-0.0031, -0.0029)	-0.0214 (-0.0223, -0.0205)	-0.0253 (-0.0259, -0.0246)
LivingAlone	10 pp	0.0016 (0.0015, 0.0018)	-0.0034 (-0.0051, -0.0019)	0.0050 (0.0035, 0.0064)
Crowding	10 pp	-0.0016 (-0.0020, -0.0011)	-0.1061 (-0.1110, -0.1024)	-0.0696 (-0.0731, -0.0666)
Elderly	10 pp	-0.0050 (-0.0052, -0.0049)	-0.0538 (-0.0555, -0.0517)	-0.0534 (-0.0544, -0.0519)
OutdoorWorkers	10 pp	0.0094 (0.0091, 0.0097)	0.0036 (0.0018, 0.0065)	0.0423 (0.0409, 0.0445)
ForeignBorn	10 pp	-0.0063 (-0.0064, -0.0061)	-0.0420 (-0.0438, -0.0407)	-0.0517 (-0.0529, -0.0508)
NonWhite	10 pp	-0.0064 (-0.0065, -0.0063)	-0.0316 (-0.0322, -0.0309)	-0.0459 (-0.0463, -0.0454)
HispanicLatino	10 pp	-0.0043 (-0.0044, -0.0043)	-0.0127 (-0.0133, -0.0121)	-0.0261 (-0.0265, -0.0257)
CDH205	10 CDH	0.0235 (0.0235, 0.0235)	0.1486 (0.1483, 0.1488)	0.1916 (0.1914, 0.1918)
meanRH	10 pp	-0.0107 (-0.0108, -0.0107)	0.1285 (0.1276, 0.1293)	0.0283 (0.0277, 0.0288)

Table S6: Daytime average discrete changes (ADC) in probability of any cooling use (ΔP_{USE}), expected hours given some cooling (ΔE_{POS}), and total expected hours (ΔE_{TOT}). The 95th percentile interval is given in parenthesis. "pp" = percentage points.

Variable	**Delta**	ΔP_{USE}	ΔE_{POS}	ΔE_{TOT}
Age	10 years	-0.0025 (-0.0026, -0.0024)	0.0255 (0.0250, 0.0261)	0.0055 (0.0052, 0.0059)
Sqft	500 sqft	-0.0063 (-0.0064, -0.0061)	0.0233 (0.0224, 0.0244)	-0.0064 (-0.0070, -0.0057)
Bedrooms	1 bedroom	0.0002 (0.0000, 0.0004)	-0.0047 (-0.0060, -0.0033)	-0.0017 (-0.0026, -0.0008)
Value	$250 k	0.0054 (0.0053, 0.0055)	0.0418 (0.0409, 0.0427)	0.0366 (0.0359, 0.0372)
MedianInc	$10 k	0.0055 (0.0054, 0.0055)	0.0328 (0.0321, 0.0333)	0.0321 (0.0316, 0.0324)
Unemployment	10 pp	-0.0269 (-0.0272, -0.0266)	-0.0744 (-0.0771, -0.0724)	-0.1123 (-0.1142, -0.1110)
Renting	10 pp	0.0009 (0.0008, 0.0010)	0.0287 (0.0277, 0.0297)	0.0170 (0.0164, 0.0177)
LivingAlone	10 pp	0.0159 (0.0157, 0.0162)	0.0439 (0.0420, 0.0455)	0.0680 (0.0666, 0.0691)
Crowding	10 pp	0.0014 (0.0008, 0.0020)	0.1439 (0.1404, 0.1489)	0.0759 (0.0734, 0.0792)
Elderly	10 pp	-0.0125 (-0.0127, -0.0122)	-0.0143 (-0.0162, -0.0124)	-0.0426 (-0.0439, -0.0413)
OutdoorWorkers	10 pp	0.0054 (0.0051, 0.0059)	0.0123 (0.0098, 0.0150)	0.0218 (0.0200, 0.0240)
ForeignBorn	10 pp	-0.0084 (-0.0085, -0.0082)	-0.0229 (-0.0242, -0.0218)	-0.0351 (-0.0359, -0.0344)
NonWhite	10 pp	0.0049 (0.0048, 0.0050)	0.0285 (0.0278, 0.0292)	0.0283 (0.0279, 0.0287)
HispanicLatino	10 pp	0.0027 (0.0026, 0.0028)	0.0424 (0.0416, 0.0430)	0.0290 (0.0285, 0.0294)
CDH205	10 CDH	0.0309 (0.0309, 0.0310)	0.0964 (0.0962, 0.0966)	0.1389 (0.1388, 0.1391)
meanRH	10 pp	0.0049 (0.0048, 0.0050)	0.1673 (0.1661, 0.1686)	0.0981 (0.0976, 0.0989)

Table S7: Nighttime average discrete changes (ADC) in probability of any cooling use (ΔP_{USE}), expected hours given some cooling (ΔE_{POS}), and total expected hours (ΔE_{TOT}). The 95th percentile interval is given in parenthesis. "pp" = percentage points.

SI.5. Results of the OLS regression model

	CL, D		CL, N		non-CL, D		non-CL, N	
Variable	β	$P > \|z\|$	β	$P > \|z\|$	β	$P > \|z\|$	β	$P > \|z\|$
Age	-0.001	0.000	-0.001	0.000	0.001	0.000	0.001	0.000
Sqft	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bedrooms	-0.006	0.003	-0.004	0.060	-0.002	0.382	-0.001	0.523
Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MedianInc	0.000	0.000	0.001	0.000	0.000	0.780	0.000	0.003
Unemployment	-0.021	0.374	-0.078	0.002	0.051	0.002	0.013	0.358
Renting	-0.108	0.000	-0.094	0.000	-0.103	0.000	-0.090	0.000
LivingAlone	0.034	0.082	0.025	0.247	-0.074	0.000	-0.048	0.000
Crowding	-0.634	0.000	-0.418	0.000	0.028	0.383	0.073	0.014
Elderly	-0.174	0.000	-0.154	0.000	-0.120	0.000	-0.168	0.000
OutdoorWorkers	-0.061	0.038	-0.004	0.902	0.202	0.000	0.200	0.000
ForeignBorn	-0.086	0.000	-0.067	0.000	-0.106	0.000	-0.099	0.000
NonWhite	-0.036	0.000	0.005	0.620	-0.090	0.000	-0.027	0.000
HispanicLatino	-0.086	0.000	-0.034	0.000	-0.054	0.000	-0.004	0.389
meanRH	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000

Table S8: Coefficients (β) and significance ($P > |z|$) of daytime ("D") and nighttime ("N") cooling load ("CL") and non-cooling load ("non-CL"), obtained from the OLS regression.

SI.6. CDH response curves from the OLS regression

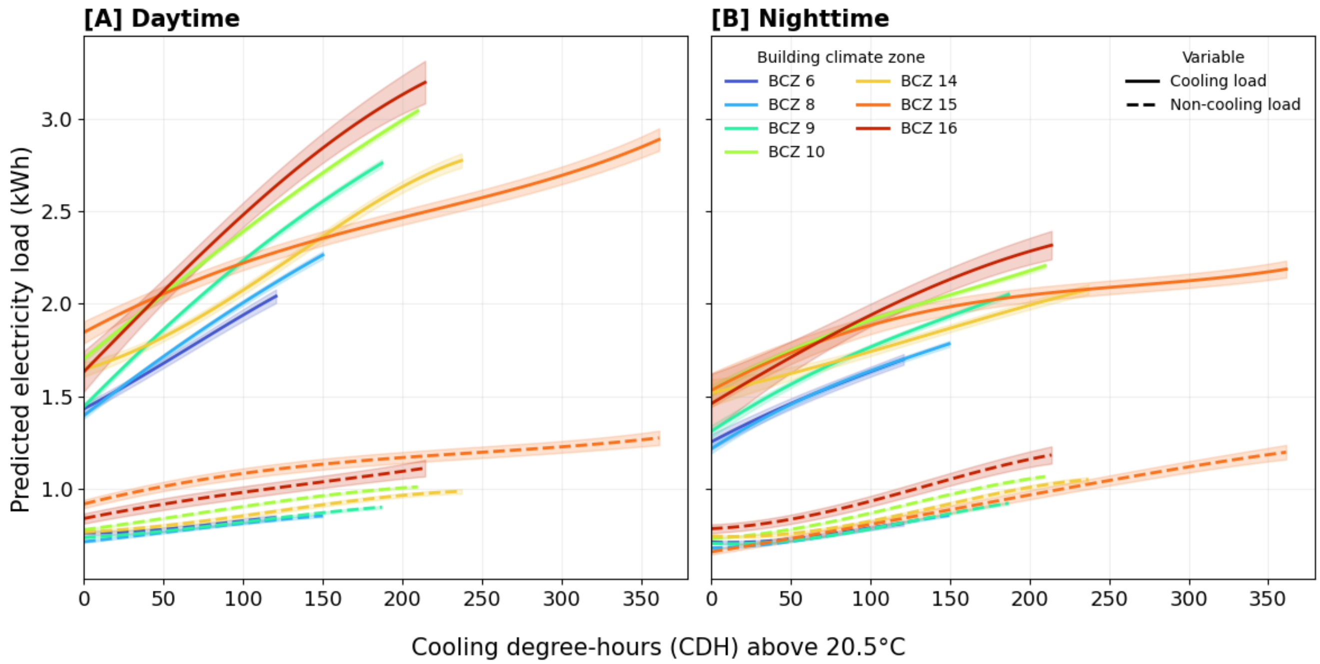

Figure S1: Predicted (a) daytime and (b) nighttime electricity loads as a function of cooling degree-hours (CDH) above 20.5°C, split by building climate zone (BCZ). Solid lines denote cooling loads and dashed lines denote non-cooling loads. Curves are shown for the central 95% of the observed CDH distribution for each BCZ. The shaded region indicates the 95% confidence interval in the fitted mean.

SI.7. Income modification of the CDH-cooling relationship

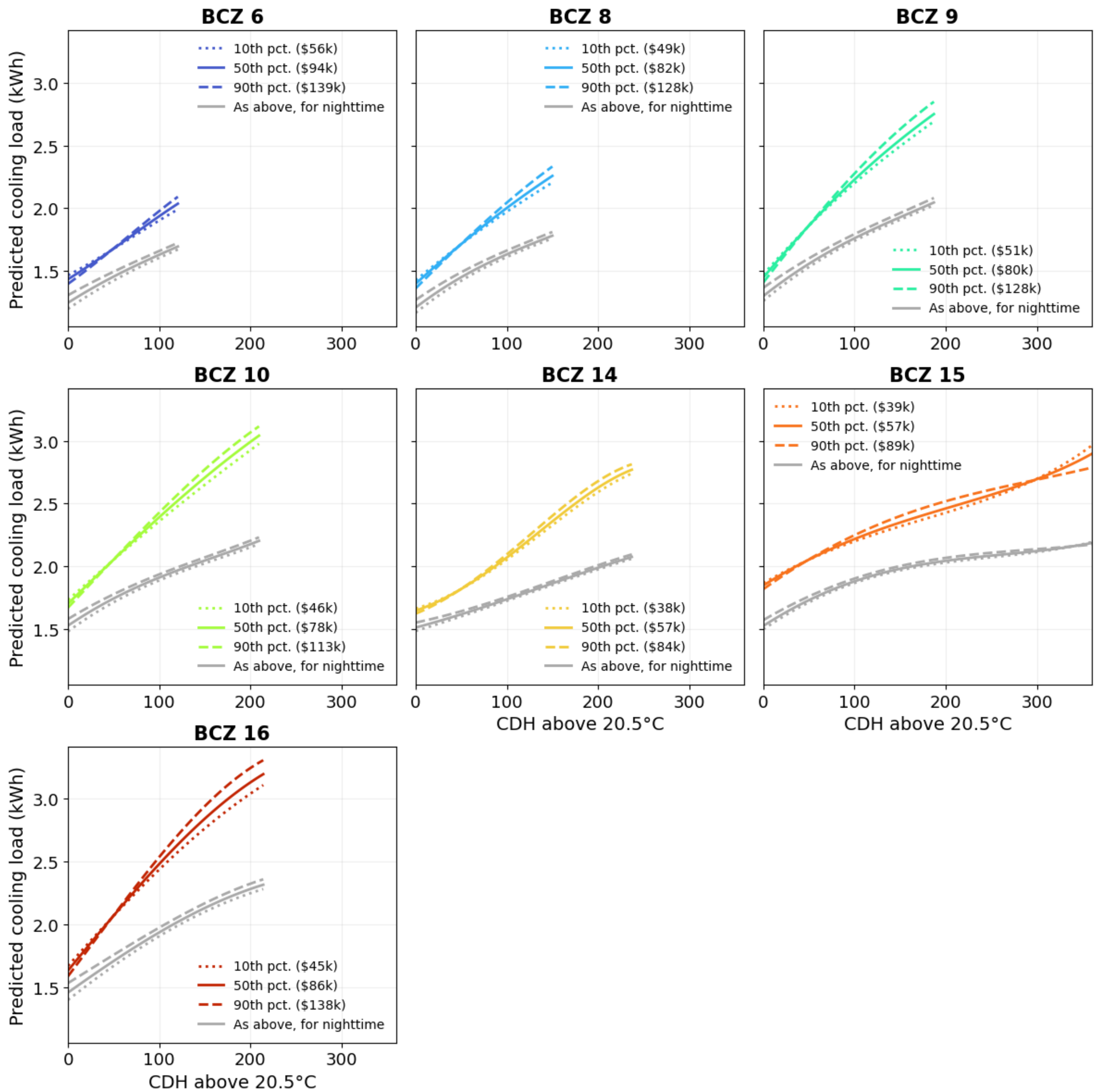

Figure S2: Predicted cooling electricity load as a function of cooling degree-hours (CDH) above 20.5°C, by climate zone and income levels. Dotted lines denote the 10th percentile by income, solid lines denote the median, and dashed lines denote the 90th percentile. Colored lines indicate daytime cooling loads, and gray lines indicate nighttime cooling loads. Curves are shown for the central 95% of the observed CDH distribution for each BCZ.

SI.8. Lorenz curves and Gini coefficients at different scales and across BCZs

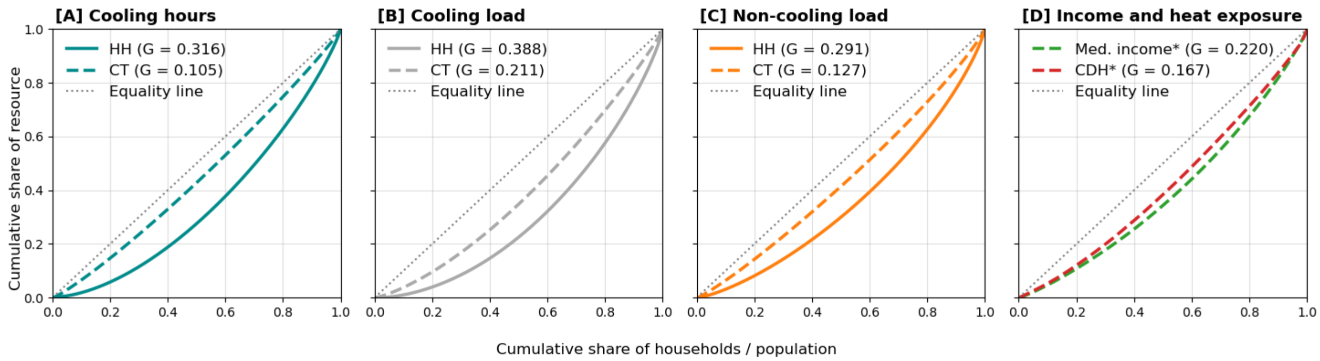

Figure S3: Lorenz curves for (*a*) cooling hours, (*b*) cooling load and (*c*) non-cooling load, computed at the household (HH) and census tract (CT) levels. For reference, panel *d* shows median income and cooling degree-hours (CDH) (the asterisk denotes that these are both computed at the CT-level). The Gini index (G) for each curve is shown in the legend.

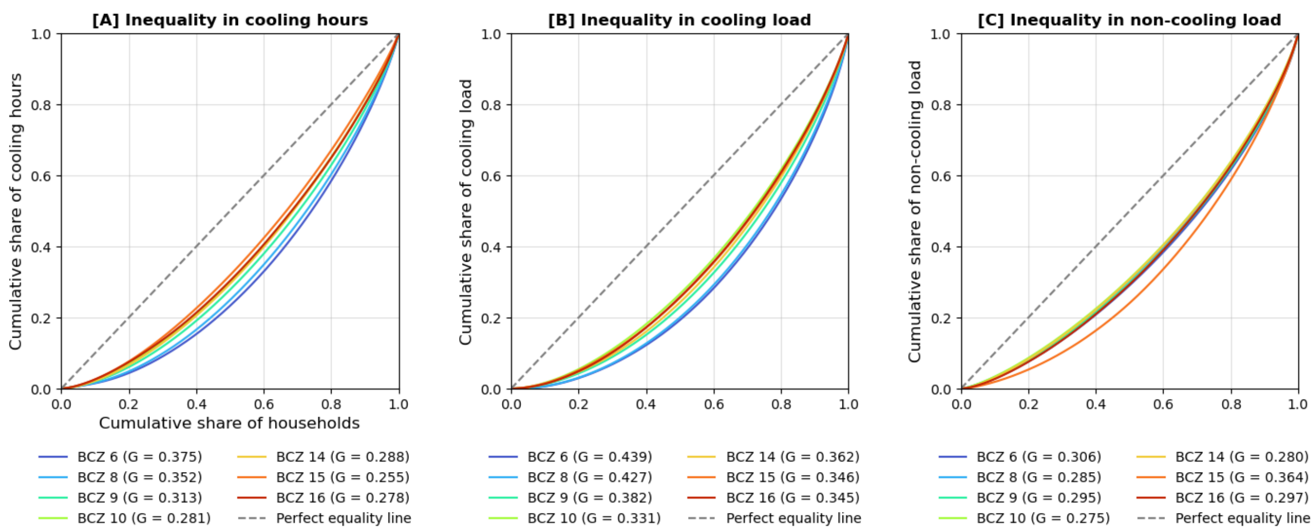

Figure S4: Lorenz curves for (*a*) cooling hours, (*b*) cooling load and (*c*) non-cooling load, stratified by building climate zone (BCZ). The Gini index (G) for each curve is shown in the legend.

SI.9. *Demographic composition of quintiles of cooling metrics*

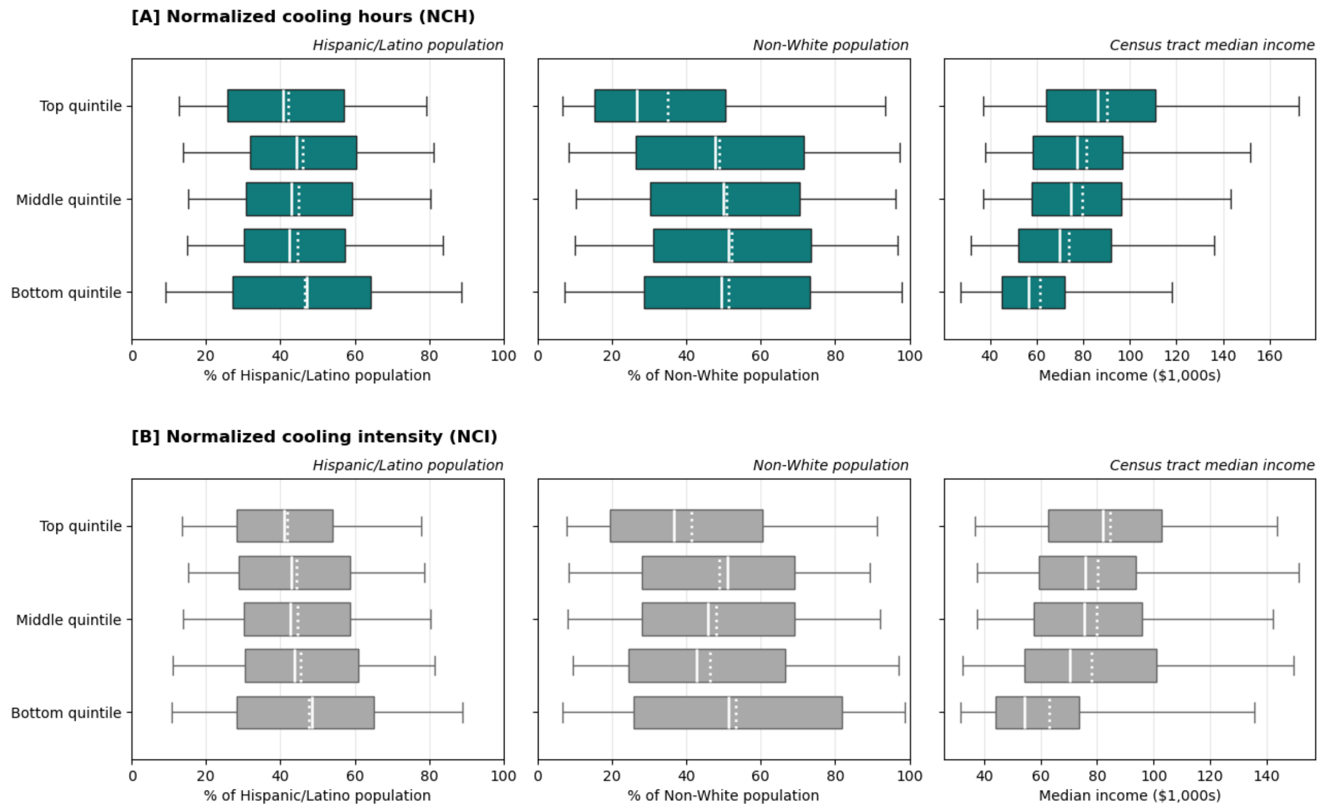

Figure S5: Boxplots show the percentage of non-White population (first column), percentage of Hispanic/Latino population (second column), and median income (third column) for quintiles of (*a*) normalized cooling hours and (*b*) normalized cooling intensity. Boxplot whiskers represent the 95th percentile.

Cooling metric	**Demographic variable**	U	n_1	n_2	p
NCH	Non-White population	433,036	455	1,818	4.16×10^{-2}
NCH	Hispanic/Latino population	454,899	455	1,818	1.17×10^{-21}
NCI	Non-White population	447,225	455	1,817	2.34×10^{-4}
NCI	Hispanic/Latino population	466,647	455	1,817	1.18×10^{-6}

Table S9: Results of one-tailed Mann-Whitney U-tests comparing the demographic composition of census tracts in the lowest quintile of normalized cooling hours (NCH) and normalized cooling intensity (NCI) (n_1) with that of all remaining tracts (n_2).

SI.10. Demographic composition of cooling clusters

Cooling metric	Demographic variable	Direction	U	n_1	n_2	p
NCH	Non-White population	$C_{LL} > C_o$	468,160	409	1,864	2.31×10^{-13}
NCH	Hispanic/Latino population	$C_{LL} > C_o$	487,603	409	1,864	2.64×10^{-16}
NCH	Median income	$C_{LL} < C_o$	262,019	416	1,882	2.08×10^{-26}
NCI	Non-White population	$C_{LL} > C_o$	444,836	413	1,860	2.39×10^{-7}
NCI	Hispanic/Latino population	$C_{LL} > C_o$	461,063	413	1,860	8.87×10^{-11}
NCI	Median income	$C_{LL} < C_o$	259,630	416	1,877	1.70×10^{-28}

Table S10: Results of one-tailed Mann-Whitney U-tests comparing the demographic composition of census tracts in the *Low-Low* cluster (C_{LL}) for normalized cooling hours (NCH) and normalized cooling intensity (NCI) (sample count n_1) with that of all remaining clusters (C_o; sample count n_2). The Direction column specifies the direction of the one-tailed test.

SI.11. Choropleth maps of spatial clusters for different BCZ

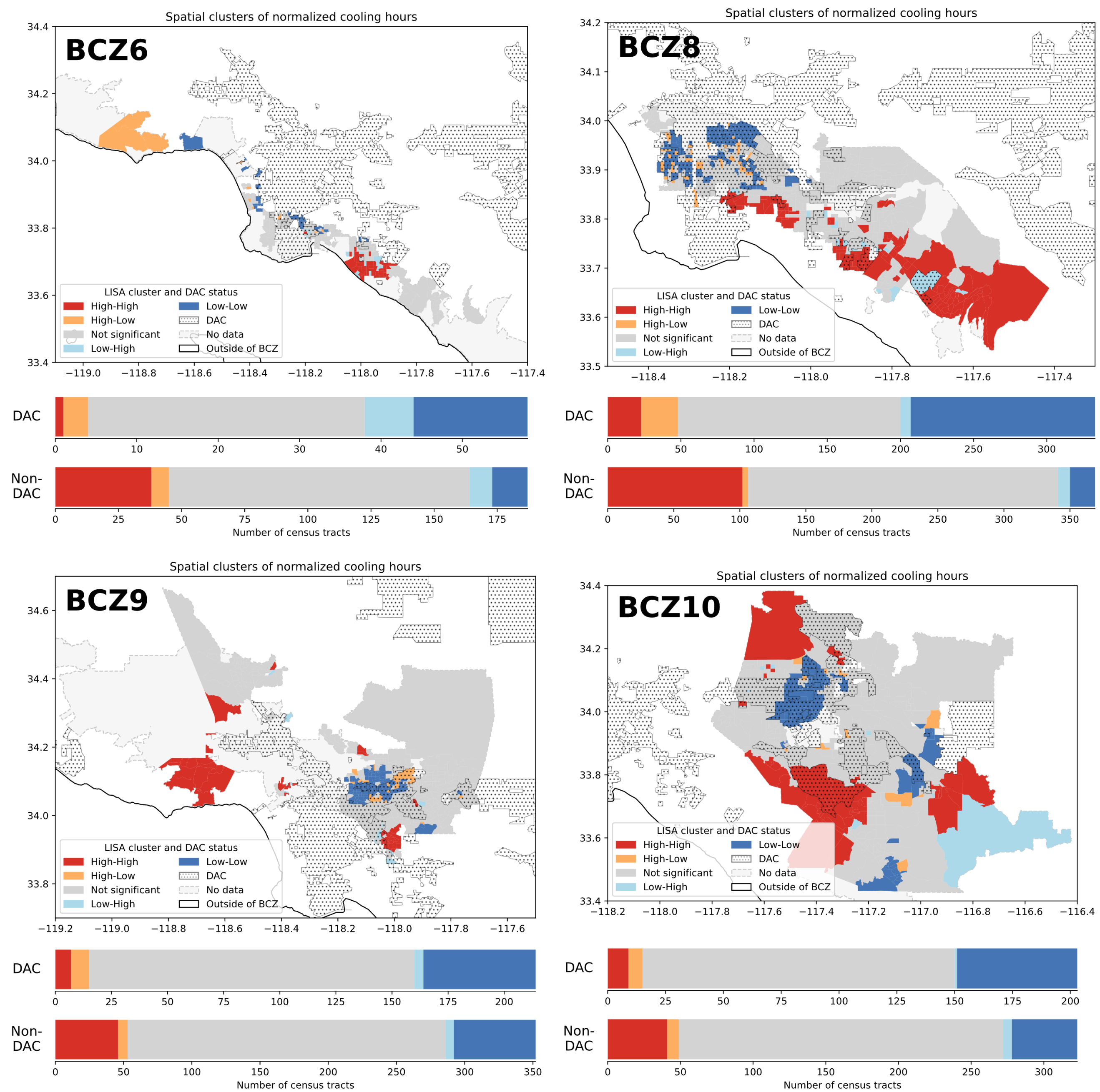

Figure S6: Maps showing spatial clusters of normalized cooling hours (NCH) identified in BCZ 6, 8, 9 and 10 (left to right, top to bottom). Disadvantaged communities (DACs) are represented with dotting. The two stacked bars under each map show the count of census tracts in DAC and non-DAC census tracts, split by cluster type.

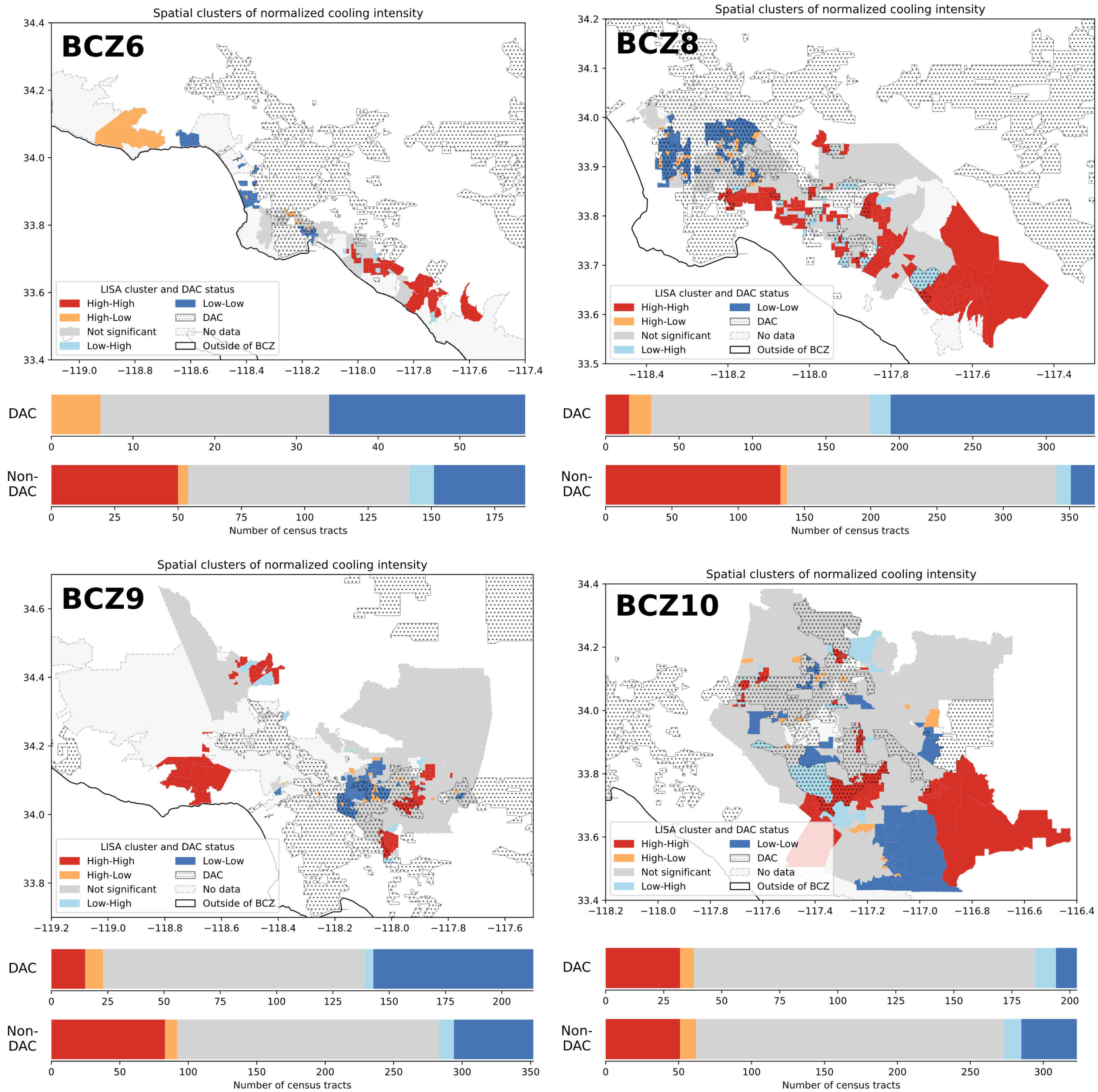

Figure S7: Maps showing spatial clusters of normalized cooling intensity (NCI) identified in BCZ 6, 8, 9 and 10 (left to right, top to bottom). Disadvantaged communities (DACs) are represented with dotting. The two stacked bars under each map show the count of census tracts in DAC and non-DAC census tracts, split by cluster type.

SI.12. AC operation data from Peplinski et al., 2025

AC operation data was obtained using multiple linear regression model described in Equation S1 (adapted from Equation 1 in Peplinski et al. [39]). This model allowed the classification of hourly smart-meter electricity data of households in Southern California into AC (cooling state = 1) and non-AC hours (cooling state = 0).

$$E_t = \mathbf{D}_{t\in h,w,s} + \begin{cases} 0 & \text{if } H_t = C_t = 0 \\ \beta_1 \times (SPT_H - T_t) + i & \text{if } C_t = 1 \\ \beta_2 \times (T_t - SPT_c) + j & \text{if } H_t = 1 \end{cases} \tag{S1}$$

where E_t is the electricity consumed in hour t, **D** is a vector of dummy variables for a given hour of day h, day type w (weekday versus weekend), and temperature group s (high or low temperature); β_1 and β_2 are the electricity-temperature sensitivity for heating and cooling, respectively; SPT_H and SPT_C are the setpoint temperatures for heating and cooling, respectively; i and j are the respective intercepts; T_t is the outdoor temperature at that hour; and H_t and C_t are binary variables indicating the heating and cooling state at that hour. For our purposes, we ignored $H_t = 1$ hours (when electric heating was occurring), therefore classifying each hour t as either an AC on hour ($C_t = 1$) or AC off hour ($C_t = 0$), irrespective of the heating state.

SI.13. Logistic regression to model cooling hours

We modeled the daily cooling hours using a two-stage (hurdle) logistic regression model. The first stage was a binary logistic regression, which predicts the probability $\pi_i = P(H_{C,i} > 0|X_i)$ of *any* cooling use for a given household-day i and a vector of independent variables X_i. Equation S2 shows the linear predictors of the binary logistic stage ($\eta_i^{(B)}$), such that $\pi_i = \sigma(\eta_i^{(B)})$, where σ is the logistic function.

$$\eta_i^{(B)} = \beta_0 + \beta^{(B)\intercal} X_i + \zeta^{(B)\intercal}\mathrm{BCZ}_i + \kappa^{(B)\intercal} S_i + \alpha^{(B)\intercal} b(\mathrm{doy}_i) \tag{S2}$$

In the above equation, β_0 is the intercept, $\beta^{(B)}$ is a column vector of coefficients on linear covariates X_i (see Table S1), $\zeta^{(B)}$ is the vector of coefficients on building climate zone (BCZ) indicator variables, $\kappa^{(B)}$ is the vector of coefficients on period of the year (S) indicator variables, and $\alpha^{(B)}$ is the vector of coefficients on $b(\mathrm{doy}_i)$, a spline basis function evaluated at day-of-year. The vector of covariates X_i contains all the variables listed in Table S1.

The second stage of the logistic model was an ordered logistic regression for positive cooling hour counts, which determines the cumulative probability $\gamma_{i,k}$ for k cooling hours, for a given household-day i. Using thresholds (or *cutoff points*) τ_k for $k = 1, 2, ..., 11$ the logistic function can be written as $\gamma_{i,k} = \sigma(\tau_k - \eta_i^{(O)})$ where $\eta_i^{(0)}$ is the linear predictor of the ordered logistic model (Equation S3). Note that the structure of $\eta_i^{(0)}$ is the same as $\eta_i^{(B)}$, except in this case we drop the intercept.

$$\eta_i^{(O)} = \beta^{(O)\intercal} X_i + \zeta^{(O)\intercal}\mathrm{BCZ}_i + \kappa^{(O)\intercal} S_i + \alpha^{(O)\intercal} b(\mathrm{doy}_i) \tag{S3}$$

We obtained $q_{i,k}$, the probability of a specific category *k*, by comparing the cumulative probabilities of k and of $k-1$ (Equation S4). Then, we obtained the unconditional expected cooling hours, μ_i, by multiplying the probability of cooling π_i by the (conditional) expected values for each category of cooling hours m_i (Equation S5). Finally, we measured the effect on expected cooling hours resulting from a change Δx_j in a predictor j. We averaged this effect across all observations N to obtain the average discrete change (ADC) for the total expected cooling hours (Equation S6). We also obtained the ADC for the probability of cooling π_i and the conditional expected cooling hours m_i.

$$q_{i,k} = \gamma_{i,k} - \gamma_{i,k-1} \tag{S4}$$

$$\mu_i = \pi_i \cdot m_i = \pi_i \sum_{k=0}^{12} k \cdot q_{i,k} \tag{S5}$$

$$\mathrm{ADC}_j^{(\mu)} = \frac{1}{N} \sum_{i=1}^{N} \Delta\mu_i \tag{S6}$$

SI.14. *Ordinary least squares regression to model cooling load*

We modeled the average hourly cooling and non-cooling load ($\overline{E_C}$ and $\overline{E_{NC}}$) with ordinary least squares (OLS) regression. In order to satisfy the assumptions of OLS, we log-transformed the outcomes $\overline{E_C}$ and $\overline{E_{NC}}$. The linear predictor was defined based on a vector of linear covariates, spline basis functions, and indicator variables, as shown in Equation S7.

$$log(1 + Z_i) = \beta_0 + \beta^\intercal X_i + \theta^\intercal b(T_i) + \zeta^\intercal \text{BCZ}_i + \kappa^\intercal S_i + \alpha^\intercal b(\text{doy}_i) + \lambda^\intercal (\text{Inc}_i \cdot b(T_i)) + \varepsilon_i \,, \tag{S7}$$

where Z is the cooling or non-cooling load, β_0 is the intercept, β is a column vector of coefficients on linear covariates X_i (see Table S1); θ represents the coefficients on $b(T_i)$, the spline basis function for the temperature term; ζ is the vector of coefficients on building climate zone (BCZ) indicator variables; κ is the vector of coefficients on period of the year (S) indicator variables; α is the vector of coefficients on $b(\text{doy}_i)$, a spline basis function evaluated at day-of-year; λ is the coefficient of the interaction term between median income (Inc_i) and the temperature spline ($b(T_i)$); and ε is the error term. In this case, the temperature term was removed from X_i to avoid redundancy. Since the dependent variable was specified as $log(1+Z_i)$, we report average percent changes computed for a discrete change Δx_j, averaged across observations.

SI.15. *Locations of DACs within the study region*

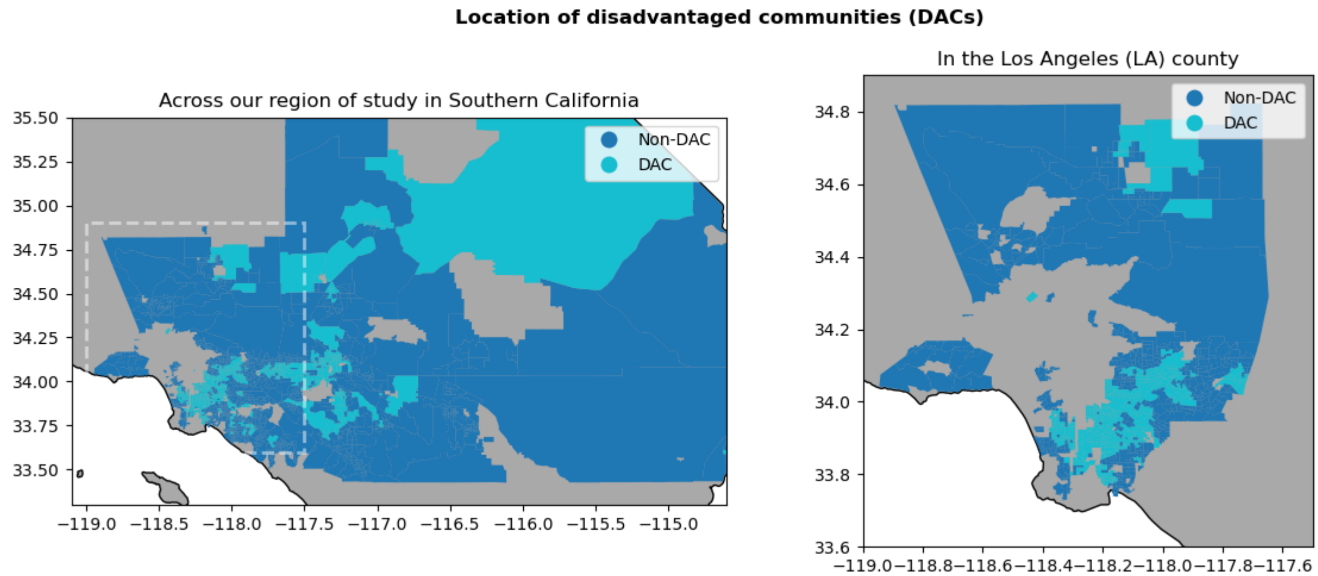

Figure S8: Maps showing locations of disadvantaged communities (DACs), as defined by CalEPA [48], across our region of study (left) and zoomed into the Los Angeles county (right).