Supplementary Information for

Location-specific strategies for eliminating US national racial-ethnic $\mathrm{PM}_{2.5}$ exposure inequality

Yuzhou Wang, Josh S. Apte, Jason D. Hill, Cesunica E. Ivey, Regan F. Patterson, Allen L. Robinson, Christopher W. Tessum and Julian D. Marshall

Julian D. Marshall.
E-mail: jdmarsh@uw.edu

This PDF file includes:

Supplementary text
Figs. S1 to S17
Table S1
SI References

Supporting Information Text

S1. Sensitivity analyses comparing three approaches. In addition to the base case for the three approaches, we also perform extensive sensitivity analyses.

To compare the "location" and "sector" approaches we (a) considered relative inequality rather than absolute inequality (Fig. S1); (b) used two alternative reduction scenarios (only reducing 50% or 90% of each emission location or sector, such that emissions are not completely eliminated; Fig. S2); (c) optimized to reduce inequality in each geographic region or in each Urban Area instead of nationally (Fig. S3); (d) reduced inequality for HV instead of racial-ethnic groups (Fig. S4); (e) considered urban disparities instead of national disparities (Fig. S5); (f) optimized to reduce average concentrations rather than concentration-disparities (Fig. S6); and (g) optimized location-sectors and sector-regions (i.e., modified the "sector" approach to make it more similar to the "location" approach; Fig. S7).

For sensitivity analysis [b] (i.e., only partial emission-reductions), we apply 90% (in another case, 50%) emission reductions for each location-pollutant or sector-pollutant parings. In these two scenarios, the emission reductions end once 10% (or, 50%) of the original total concentrations and concentration inequalities has been reached, respectively. If no further emission reduction could reduce the disparity (i.e., if the marginal concentration differences become negative for all the rest of parings), one could decide to halt the optimization at that point. (Further emission-reductions beyond that point will increase racial-ethnic disparities; the reason is that further emission-reductions reduce concentrations more for less-exposed groups than for more-exposed group.) However, to shed additional light on these scenarios, our optimization algorithm instead continues its simulation: it simulates the local optimum to keep the disparity low and non-negative (see green lines for 50% reduction scenario in Fig. S2: the line slopes upward after 5 MT/y emission reduction). Compared with the main approach, partial emission-reductions require more emission reduction amounts to reach the same disparity reductions.

For sensitivity analysis [c] (optimize regionally and locally instead of nationally), we reduce the emissions within each EPA region or Urban Area, separately. For the regional optimization, the median emission reductions to reduce 90% (from 1.5 to 0.15 μg/m^3) of the median regional disparity are 0.03 and 0.8 MT/y (a 27-fold difference) for "location" and "sector" approaches, respectively. For the local optimization, the median disparity is 0.4 μg/m^3; to reduce 90% of the disparity, the required emission reductions for the two approached are 0.001 and 0.01 MT/y (a 11-fold difference), for "location" and "sector" approaches, respectively. Considering large, medium, and small urban areas (UAs) separately for the local optimization, the median disparities are 1.0, 0.6, and 0.2 μg/m^3 (relative disparities: 11%, 9%, and 3%), respectively. (Large/medium/small UAs are defined following (1), as population tertiles: n=10 large UAs, population >4m; n=44 medium UAs, population 728k – 4m; n=177 small UAs, population <728k; restricted to UAs with >20 ISRM grid cells.) To reduce 90% of the disparity, the required emission reductions for "location" and "sector" approaches are 0.003 and 0.06 MT/y (a 17-fold difference) for large urban areas, 0.002 and 0.03 MT/y (a 14-fold difference) for medium urban areas, and 0.0003 and 0.003 MT/y (a 12-fold difference) for small urban areas. Thus, our findings (the greater efficiency of the "location" approach relative to the "sector" approach) are consistent across large, medium, and small urban areas.

For sensitivity analysis [d] (social vulnerability), the population-average modeled $PM_{2.5}$ concentration is 7.9 μg/m^3 for HV locations and 6.9 μg/m^3 for non-HV locations (using our main definition for HV: 10% of CDC's SVI). The estimated disparity for HV is 0.9 μg/m^3 (13%) (i.e., the average $PM_{2.5}$ level for HV locations versus the population average). Using two alternative definitions for HV locations (20% of CDC's SVI; 10% $PM_{2.5}$ EJ index in EJScreen), the disparities for HV are 0.8 μg/m^3 (11%) and 1.8 μg/m^3 (25%), respectively. The overall reduction efficiency for HV is higher compared with the racial-ethnic disparities, using all three HV definitions. For all the three HV definition, "location" approach is much more efficient than "sector" approach in reducing the disparity for HV locations. Both approaches (i.e., "location" and "sector") reduce disparities by 90% or more, at less than 1.5 MT/y emission-reduction. In summary, the optimization can dramatically reduce disparities at far less emission-reduction for HV than by race-ethnicity; in all sensitivity analyses (as is also true for the main case), the "location" approach is more efficient than "sector" approach at reducing disparities.

For sensitivity analysis [e] (i.e., urban disparities), we explore the within-CBSA disparity changes for the baseline national optimizations. The results indicate that "location" approach is slightly more efficient than "sector" approach at most of the emission reduction levels, and the required emission reductions to reduce 50% (from 0.83 to 0.41 μg/m^3) of the within-urban disparities are 0.7 MT/y and 1.3 MT/y for "location" and "sector" approaches, separately.

For sensitivity analysis [f], we adjust the optimization metric in "location" and "sector" approaches to reduce average concentrations rather than concentration-disparities (Fig. S6). The goal of this alternative optimization changes from addressing/minimizing exposure disparity to maximize the health benefit for the total population. Results from [f] also indicate "location" approach is more efficient (here, at reducing population-average exposure concentrations) than "sector" approach.

For sensitivity analysis [g], we explore two alternative emission reduction steps: sector & region & pollutant combinations (the "sector and geographic region" approach) and sector & location combinations (the "sector and location" approach). The "sector and geographic region" approach adds EPA region of the emission sources as a further dimension in the reduction steps compared with the "sector" approach, and has 595 combinations (i.e., 14 sectors; 5 pollutants; 10 EPA regions; of the 700 maximum possible sector-pollutant-region pairings, 105 have zero emissions and so are not considered here as an opportunity for emission-reduction.) in total; the "sector and location" approach changes the pollutant type to source sector compared with the "location" approach, and has 509,128 combinations (i.e., 14 sectors; 52,411 locations; of the 733,754 maximum possible sector-location parings, 172,215 have zero emissions.) in total. Results from [g] (Fig. S7) indicate that adding additional dimension of geographic region improve the efficiency of "sector" approach; the efficiency of the approach combining sector and location (i.e., "sector" approach modified to be similar to "location" approach) is almost the same as "location" approach.

This finding reflects that this sensitivity analysis modifies the "sector" approach by adding information on location (the EPA region), i.e., it is partially a "sector" - "location" hybrid method.

To examine the NAAQS-like scenario, we varied the NAAQS-like concentration standard (specifically 5, 6, 7, 8, 9, and 10 μg/m^3; Fig. S8) and, as a sensitivity analysis, considered urban and regional disparities rather than national disparities (Fig. S8, middle and bottom rows). None of the NAAQS-like scenarios explored eliminate the national, regional, and urban disparities.

S2. Relationship between reduction priority for the "location" approach and grid cell characteristics. To determine the relationships between grid cell characteristics and emission reduction priority for the "location" approach, we employ both unadjusted (univariate) and adjusted (multivariate) analyses. The grid cell characteristics included in the analysis are racial-ethnic composition, median household income, population density, pollution emission density, and racial segregation index. Racial-ethnic segregation is represented by dissimilarity index at each location (grid cell), which measures the percentage of the White population in an ISRM grid which would have to change census block to equalize the racial distribution between White and non-White (or a specified non-White, e.g., Hispanic) population groups across all blocks in the grid cell. The formula of segregation index is: $D_i = 0.5 \sum_{j=1}^{n} \left| \frac{w_{i,j}}{W_i} - \frac{n_{i,j}}{N_i} \right|$, where D_i is the dissimilarity index in ith ISRM grid; W_i is the total White population in ISRM grid i; N_i is the total non-White population in ISRM grid i; $w_{i,j}$ represents the White population in jth census block that within the boundary of ISRM grid i; $n_{i,j}$ represents the non-White population in jth census block that within the boundary of ISRM grid i.

For the unadjusted (univariate) analyses (Fig. S9), White percentage, Asian percentage, and median household income are negatively related with reduction priority and statistically significant ($p<0.05$). Black percentage, population density, emission density, and segregation index have positive relationships with reduction priority and statistically significant ($p<0.05$). Hispanic percentage is positively related with emission reduction priority, but the relationship is not statistically significant ($p = 0.32$). This result implies that, in general and averaged across the country, to optimally reduce disparities one would target emission-reductions in locations that have higher values for Black percentage, population density, emission density, and segregation. (An analogous result holds for Hispanic percentage, but the relationship is "noisier" (has more scatter).)

For the adjusted (multivariate) analyses, we employ four groups of multiple linear regression models (Table S1; 13 models in total). The first group (model 1) has three independent variables: income, population density, and emission density (the "baseline" variables). The second group (models 2-5) has the three "baseline" variables, plus racial-ethnic compositions. The third group (models 6-9) has the second-group variables, plus segregation indexes. The fourth group (models 10-13) has the third-group variables, plus an interaction term between racial-ethnic composition and segregation index. The second, third, and fourth groups each contain four regression models: one for the combined non-White population and one for each of the three specified groups (Black, Hispanic, Asian).

In all of the regression models (and, consistent with the univariate analyses), population density and emission density have positive slopes ($p<0.001$) and median household income has negative slopes ($p<0.001$).

The slopes (i.e., the beta coefficients in the regression models) of racial-ethnic composition and segregation index have different patterns across racial-ethnic groups (non-White; Black; Hispanic; Asian). For the non-White group, both non-White percentage and segregation index have positive slopes in all the models (models 2, 6, and 10 for non-White percentage; models 6 and 10 for non-White/White segregation index). The interaction term between non-White percentage and segregation index (model 10) has a slight positive value, which indicates that with an increase of segregation level, the positive slope for non-White percentage becomes slightly steeper. The patterns (models 3, 7, and 11) for Black population are generally the same as combined non-White group. The only difference is that the interaction term of Black percentage and segregation index is negative (model 11), which indicates that with an increase of segregation level, the positive slope for Black percentage become less steep. Regression models for Hispanic (models 4, 8, and 12) and Asian (models 5, 9, and 13) groups have similar patterns, which are different from the non-White group and the Black group. In the regression models without interactions (models 4, 5, 8, and 9), Hispanic & Asian percentages have negative coefficients; segregation indexes have positive coefficients. However, in the models with interaction terms (models 12 and 13), the slopes of Hispanic & Asian percentages become positive, and the interaction terms are negative. The results for Hispanic and Asian population indicate that at a zero segregation levels, Hispanic and Asian percentages both have positive slopes. With an increase of segregation levels, the positive slopes become less steep, which eventually flip to be negative; at the average segregation levels, the slopes for Hispanic and Asian percentages are both negative (model 8 and 9). The p values for all the slopes in all 13 models are less than 0.001.

S3. Comparing five species for the "sector" approach. Comparing the five types of emissions that contribute to $PM_{2.5}$ – "primary" (directly-emitted) $PM_{2.5}$, and four precursor species that can form secondary $PM_{2.5}$ – primary $PM_{2.5}$ and NH_3 have the highest reduction priorities for all sectors. Reduction of primary $PM_{2.5}$ emissions causes the largest inequality reduction for most of the sectors (account for 57% of the total disparity). For addressing $PM_{2.5}$ disparities, the precursors VOC and NO_x have the lowest reduction priorities for all sectors.

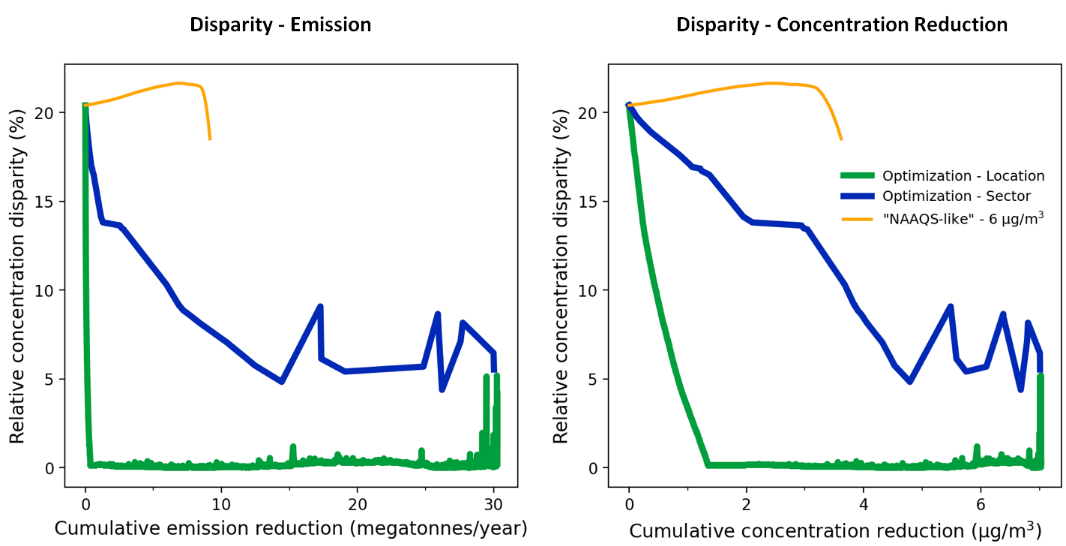

Fig. S1. Relative $\mathrm{PM}_{2.5}$ exposure disparity changes with emission reduction and concentration reduction.

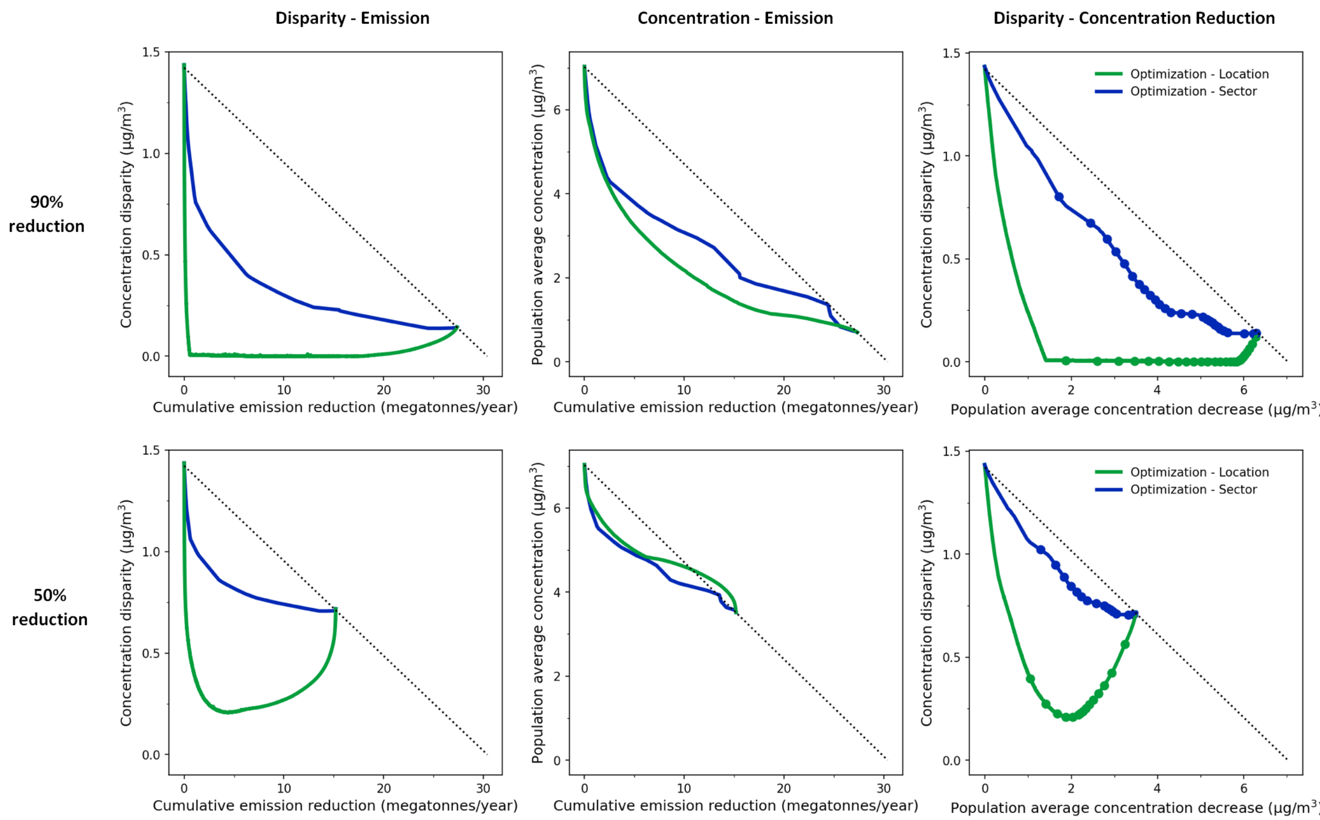

Fig. S2. $\mathrm{PM}_{2.5}$ exposure disparity and concentration reduction curves for the alternative conditions of (90% and 50%) emission reduction. Where a line trends upward (i.e., has a positive slope), any emission reduction would increase the exposure disparity (between the most-exposed racial-ethnic groups and the population average); here, the optimization procedure priorities emission-reductions with the lowest marginal increase in the exposure disparity.

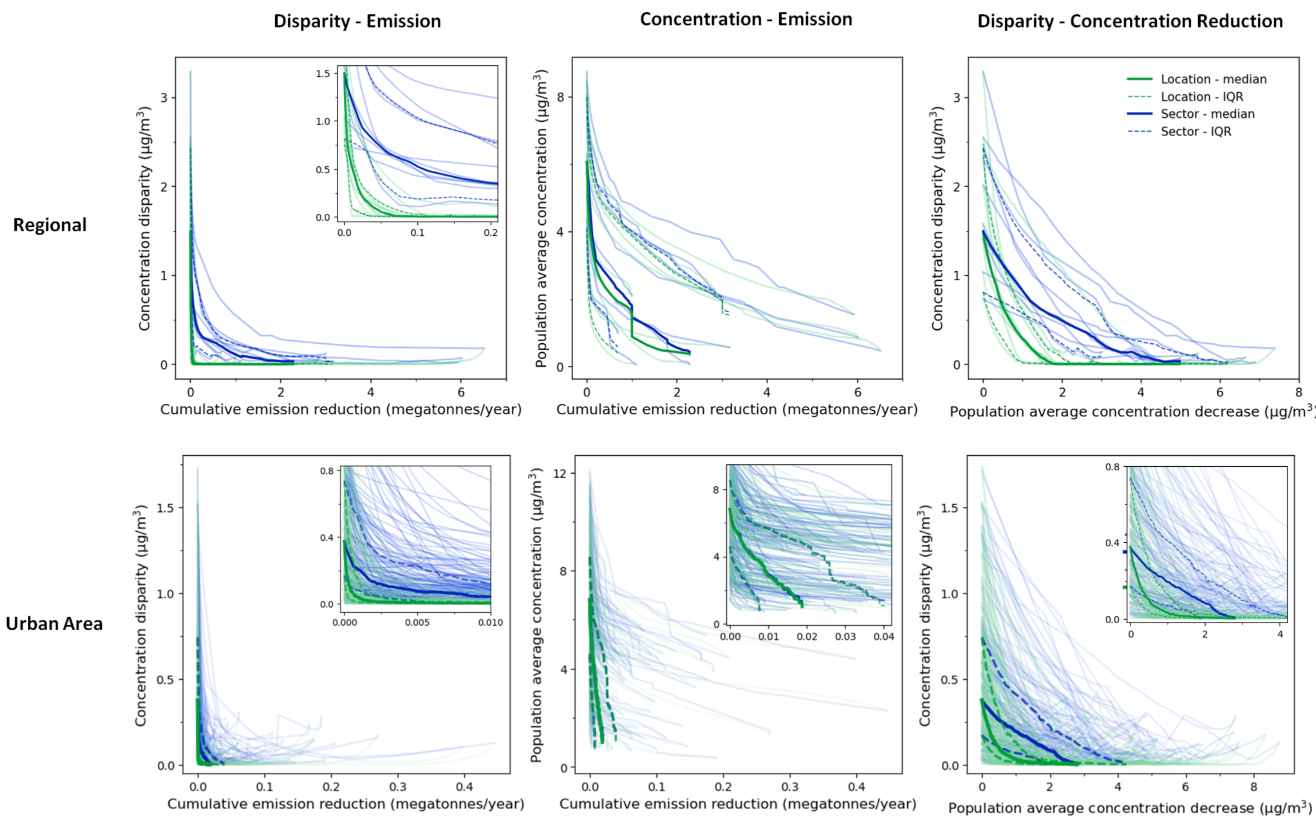

Fig. S3. $\mathrm{PM}_{2.5}$ exposure disparity and concentration reduction curves. Top row: within-region results, reflecting each region's emission-reductions to optimally reduce disparities in that region. Each light-color line reflects one US EPA region (n=10); median and interquartile range (IQR) are dark-color lines. Bottom row: within-urban results, reflecting each Urban Area's emission-reductions to optimally reduce disparities in that Urban Area [UA]. Each light-color line reflects one UA (n=171); median and IQR are dark-color lines. Some panels display zoom-in results in a sub-panel. For both rows, the location-based approach eliminates racial-ethnic disparities in exposure well before the source-based approach (i.e., the green line is below the blue line). For example, at the regionally level the location-based approach rapidly reduces disparities to zero; the source-based approach does not.

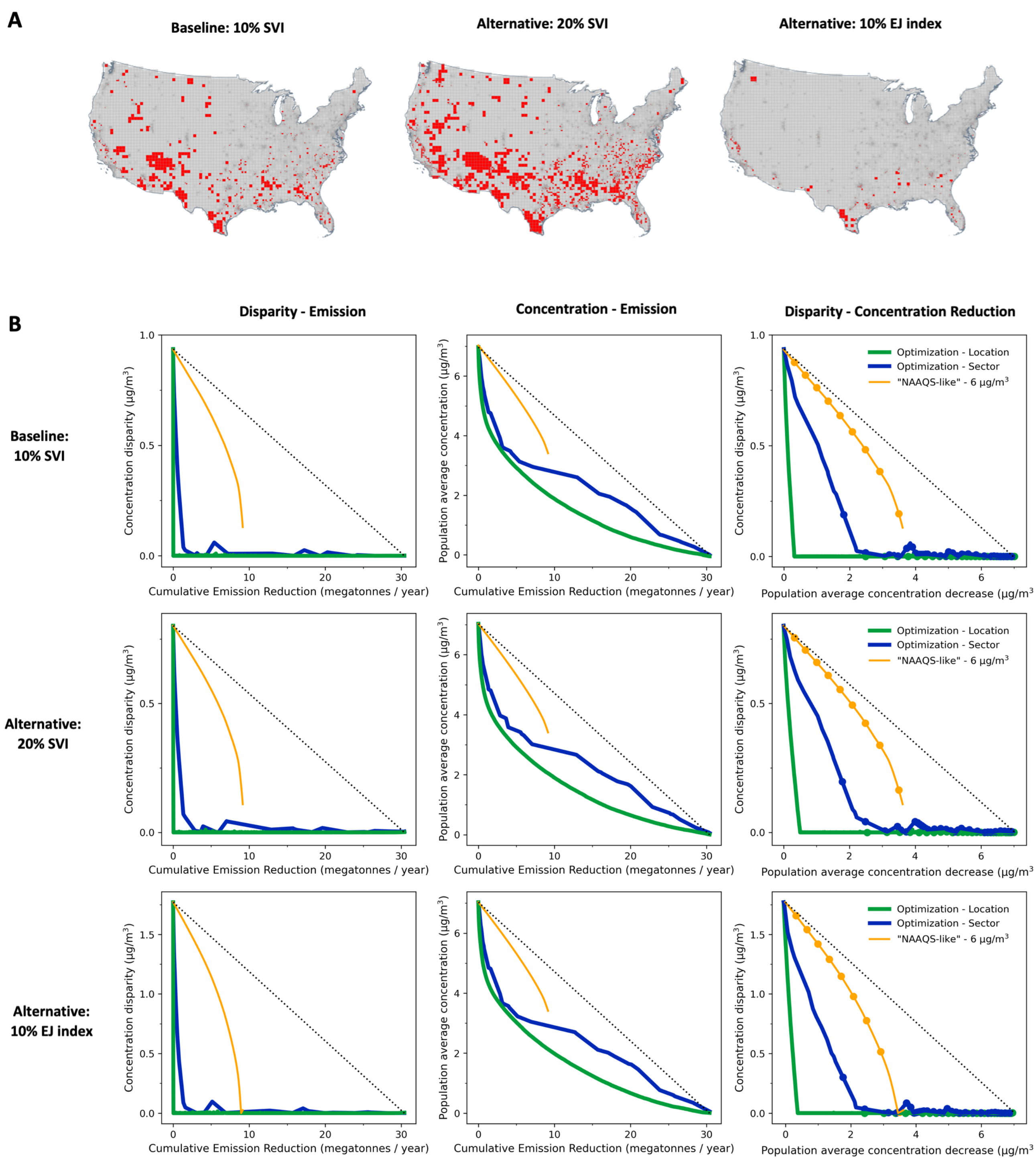

Fig. S4. Distribution map of "high vulnerability" locations, and $\mathrm{PM}_{2.5}$ exposure disparity and concentration reduction curves for HV locations.

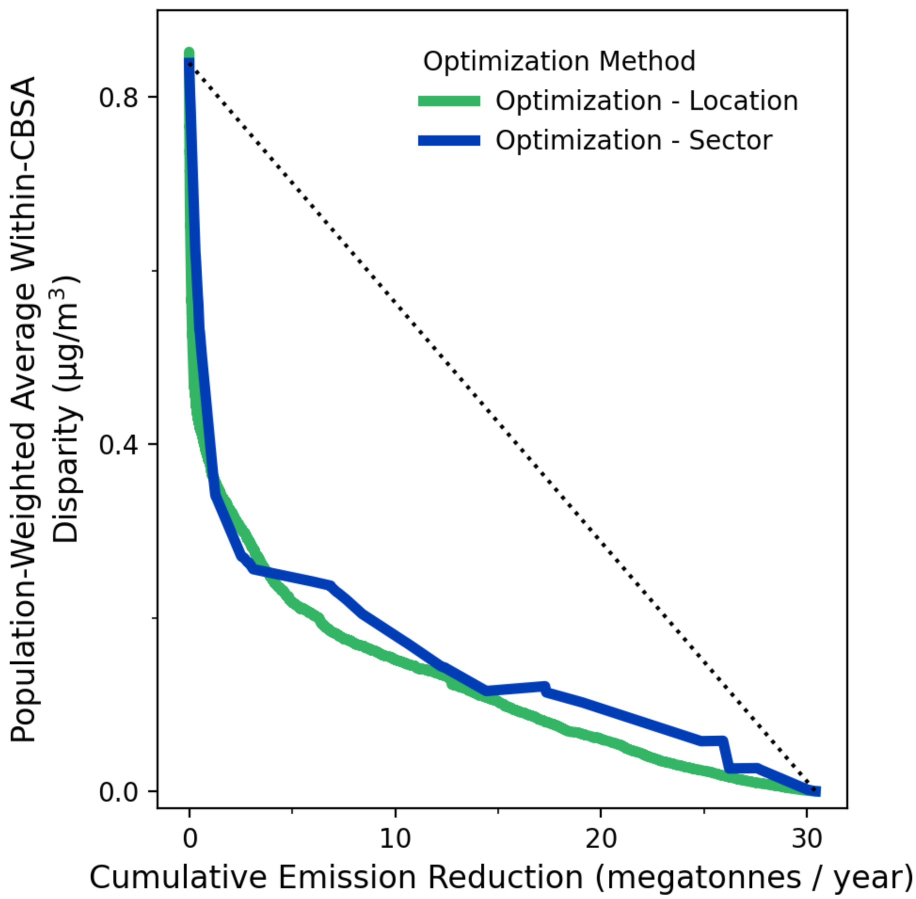

Fig. S5. Urban disparity reduction curves for the two optimization approaches.

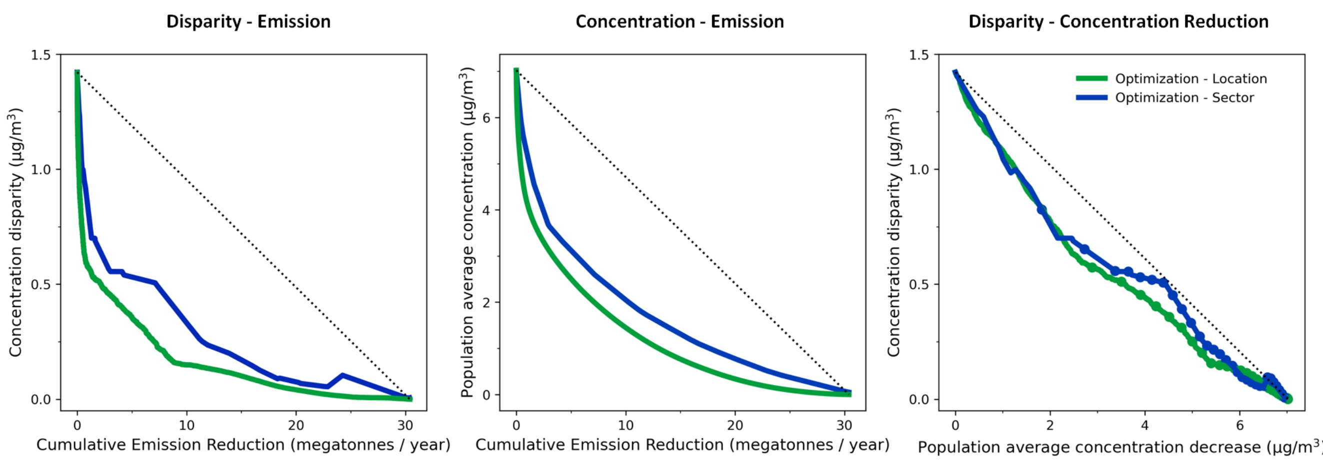

Fig. S6. $\mathrm{PM}_{2.5}$ exposure disparity and concentration reduction curves reflecting optimization to reduce average exposure concentration.

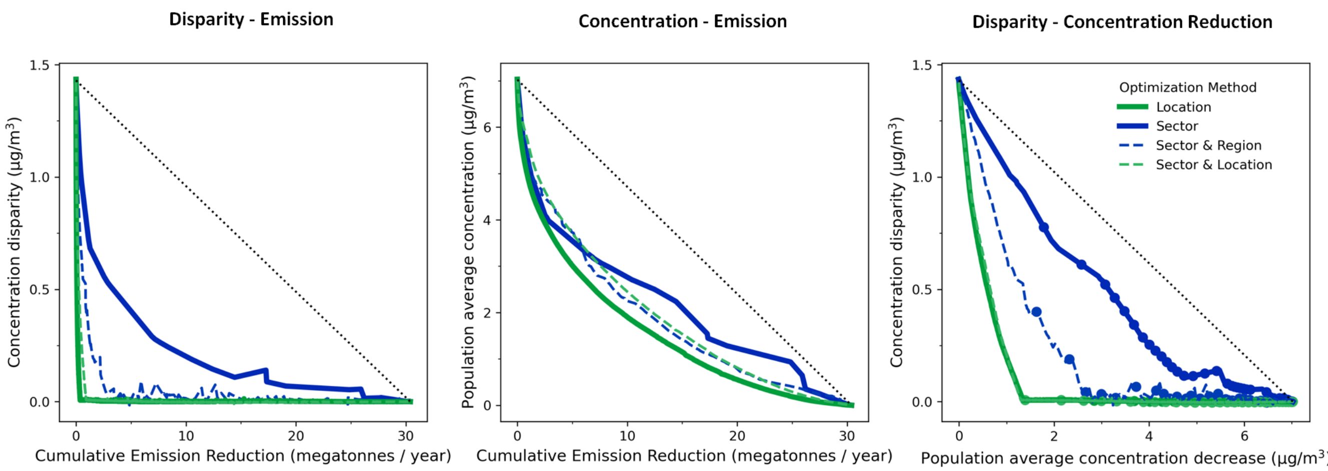

Fig. S7. $\mathrm{PM}_{2.5}$ exposure disparity and concentration reduction curves, comparing four approaches to emission-reduction: optimization by sector (blue line, same as Fig. 1), optimization by sector and geographic regions (blue dash line), optimization by location (green line, same as Fig. 1), and optimization by sector and location (green dash line).

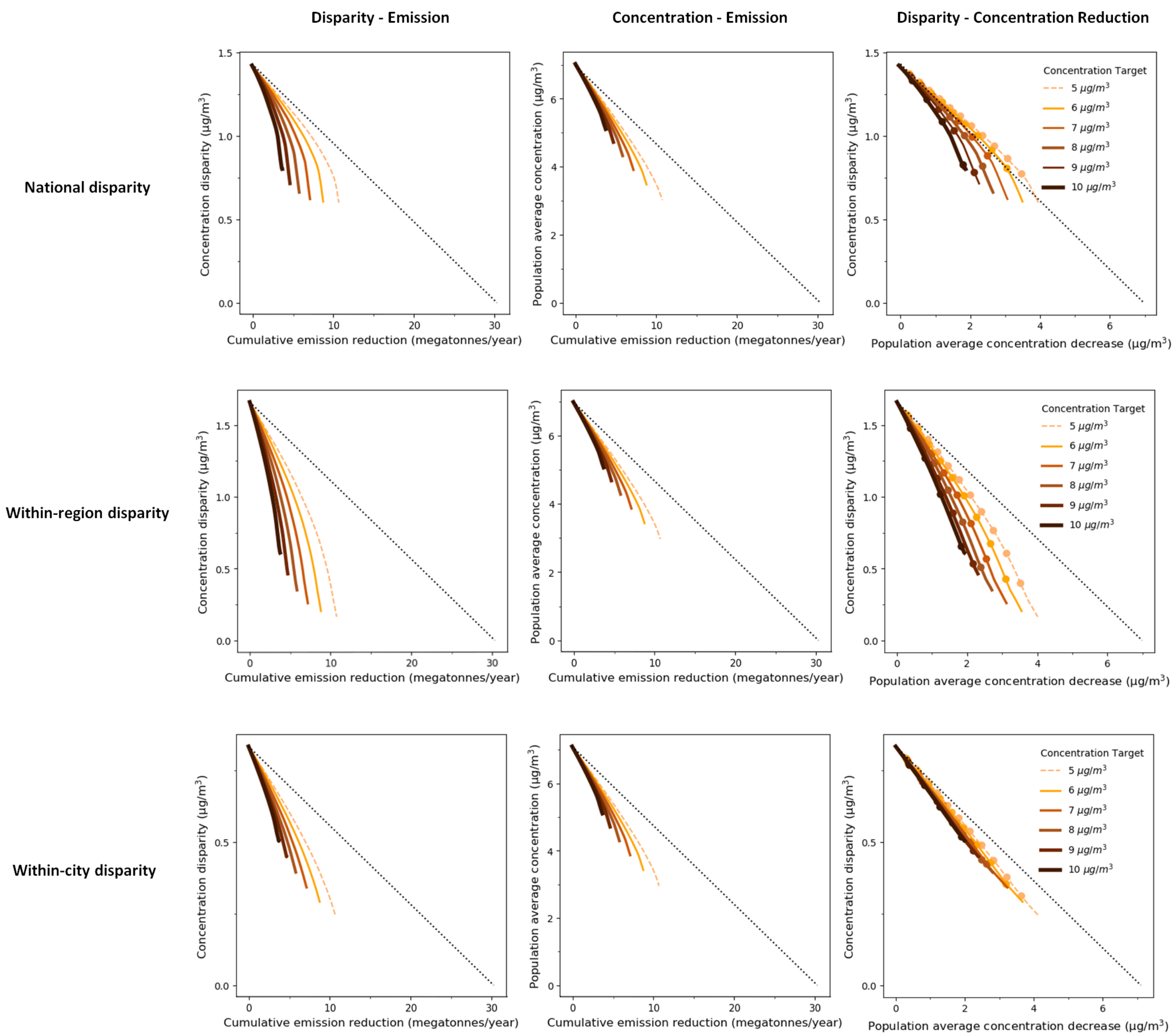

Fig. S8. $\mathrm{PM}_{2.5}$ exposure disparity and concentration reduction curves for "NAAQS-like" approach. Rows and columns are analogous to Fig. 1. Here, each CBSA reduces emissions inside that CBSA to meet the concentration target (5, 6, 7, 8, 9, or 10 $\mu\mathrm{g/m}^3$); the figure shows the resulting disparities and concentrations (top row: nationally; middle row: by regional; bottom row: by CBSA). None of the scenarios investigated here result in disparities reaching zero.

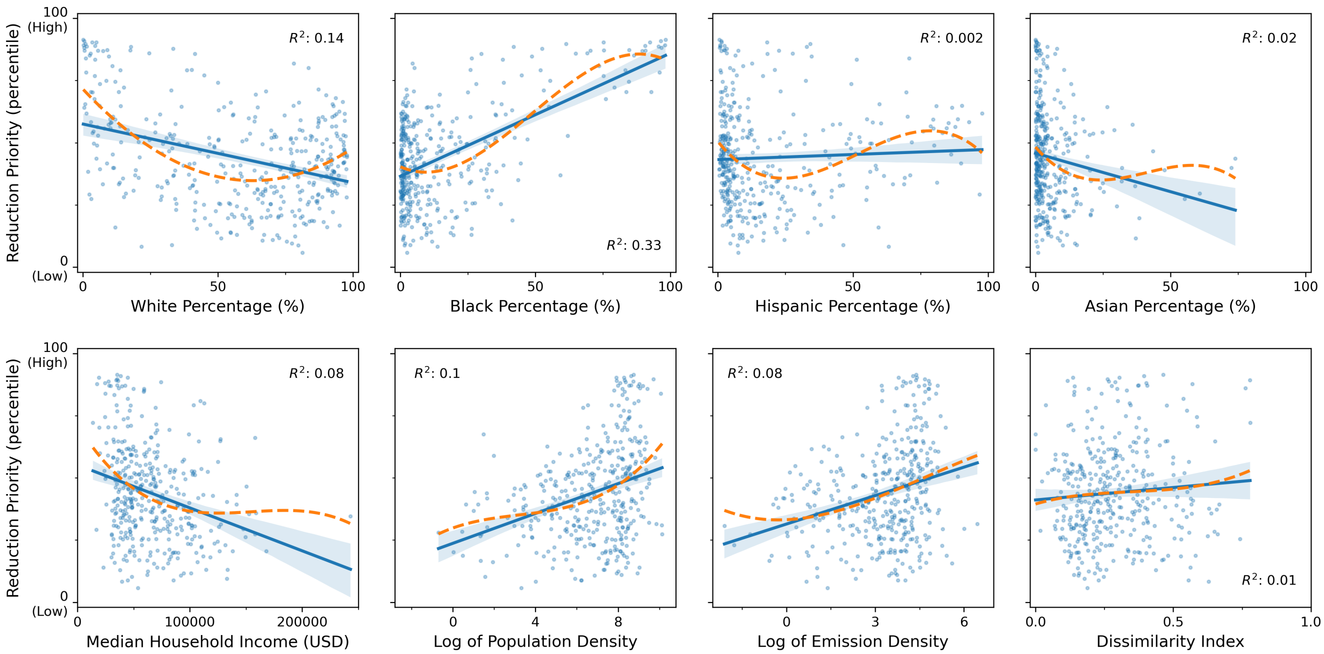

Fig. S9. Scatter plots with best fit line (blue lines) and spline smoothing line (orange lines; order = 3) of reduction priority versus racial-ethnic composition, household income, population density, emission density, and racial segregation index in the location (grid cell). Reduction priority is converted to 0–100 scale: 100 represents highest priority, 0 represents lowest priority. Points, best-fit lines, and regression R-squares are for the 1% random sub-sample of all the locations with none missing value, non-zero emissions, and non-zero populations (n = 398). Population density and emission density are at log-scale. The unit of population density is log of persons per square kilometer; the unit of emission density is log of tonnes per square kilometer.

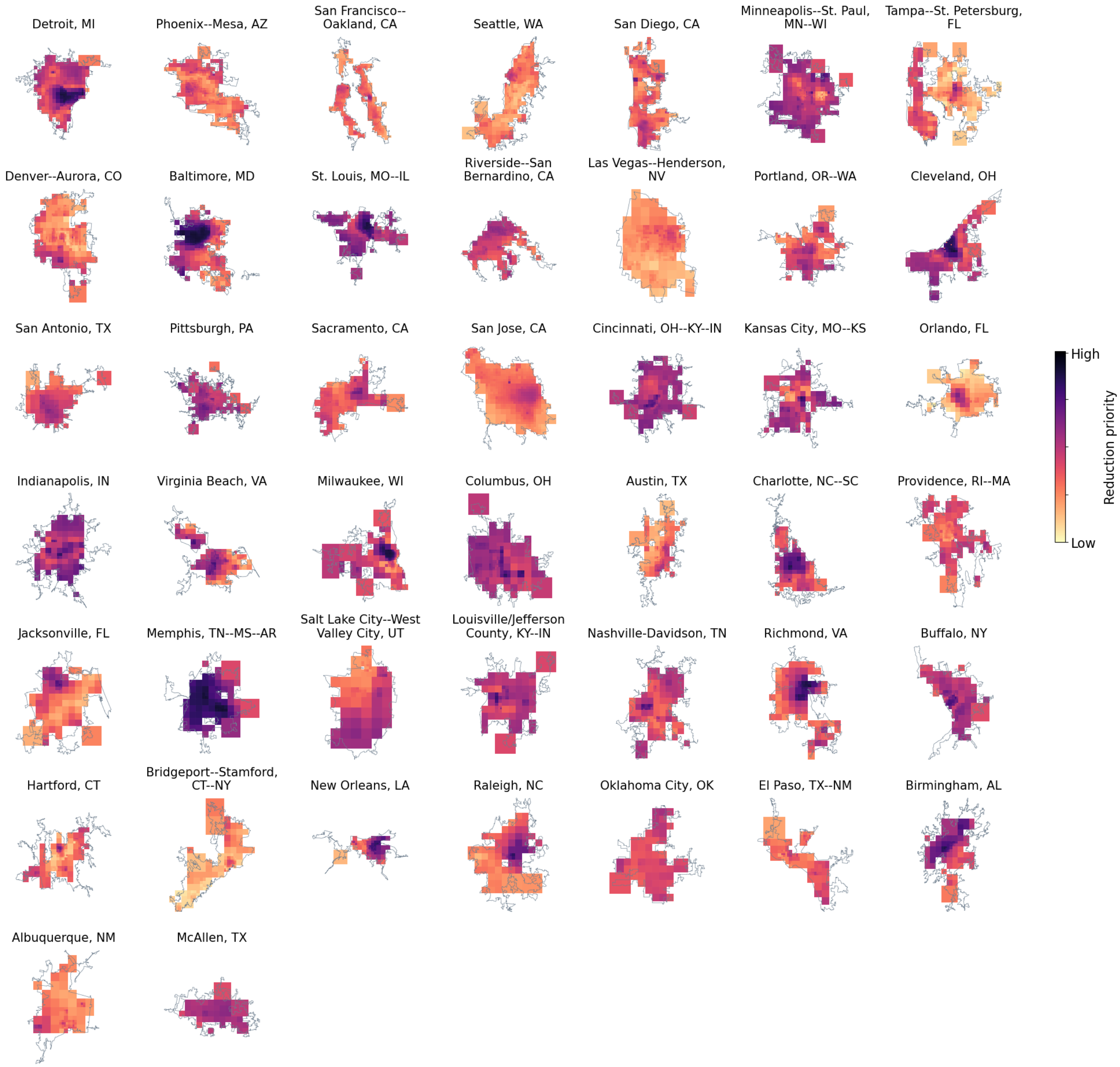

Fig. S10. Reduction priority maps for optimization by location methods for 44 medium Urban Areas.

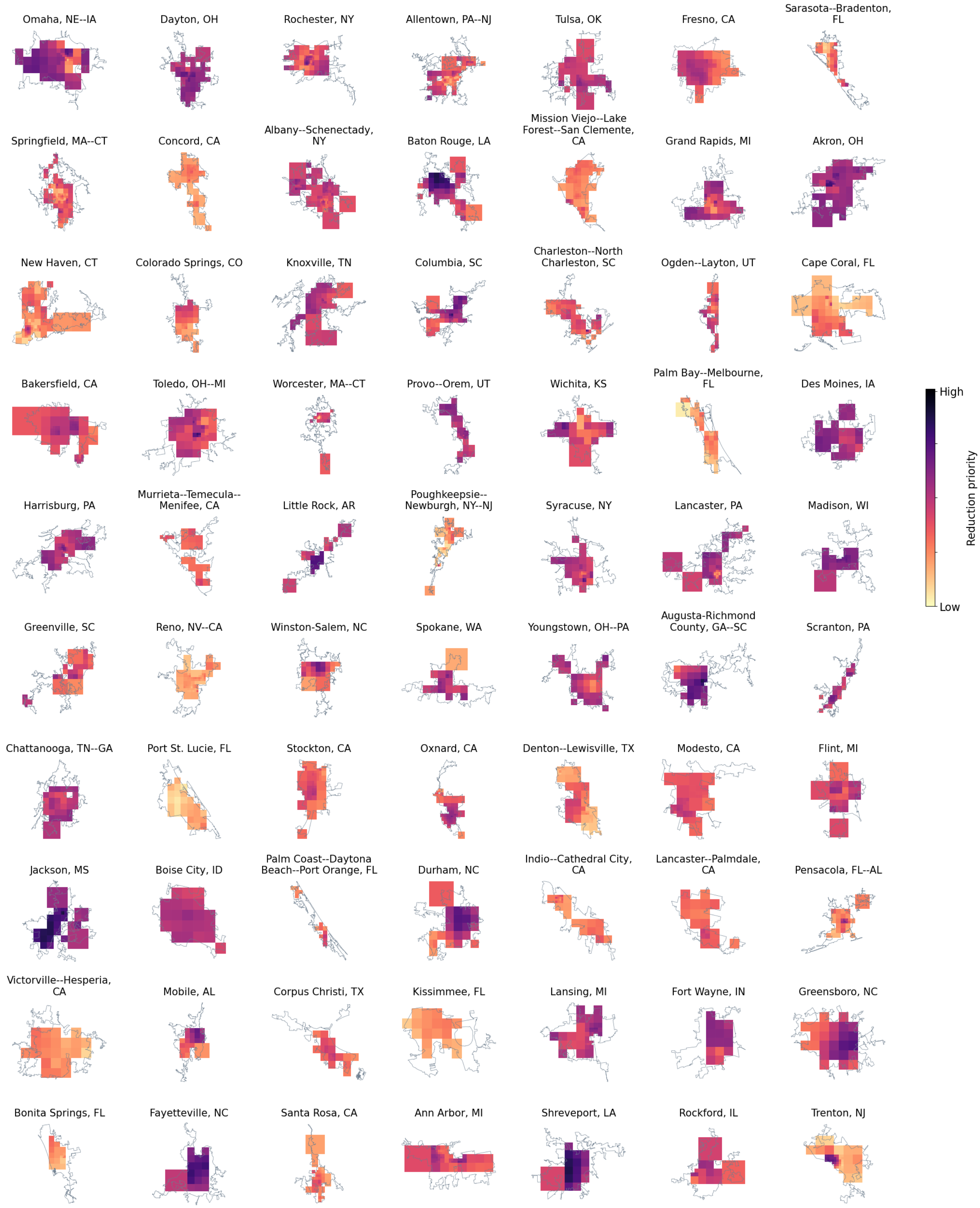

Fig. S11. Reduction priority maps for optimization by location methods for 70 (out of 381) small Urban Areas.

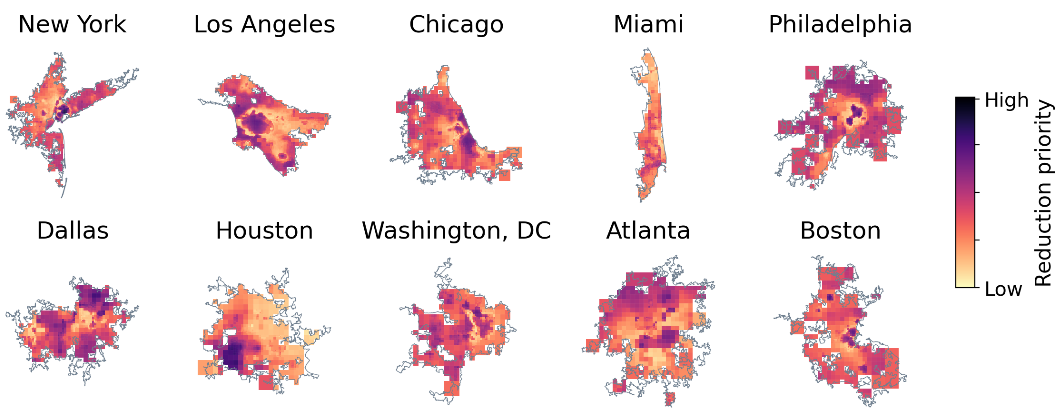

Fig. S12. Reduction priority maps for urban-level optimization by location methods for 10 large Urban Areas.

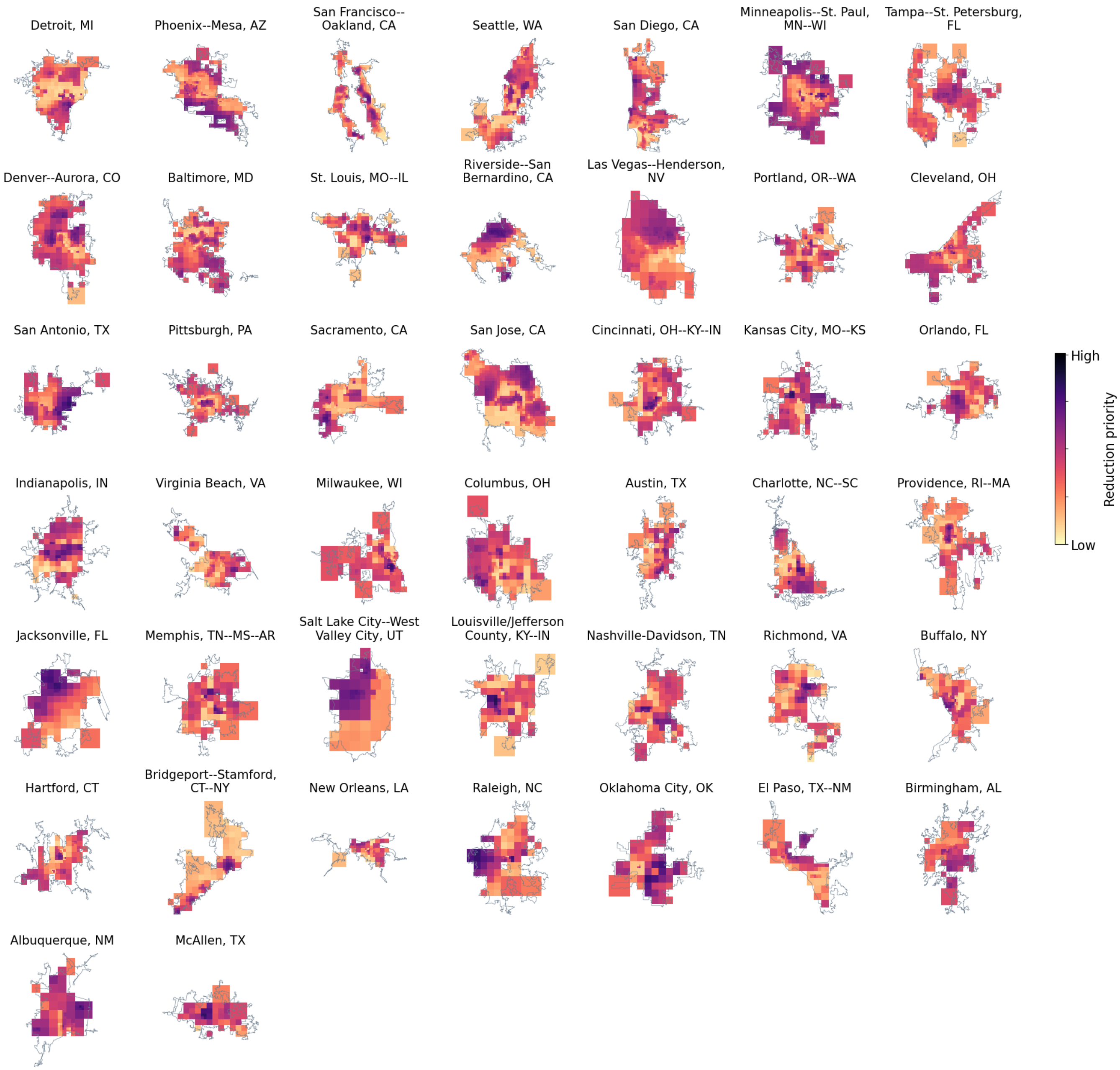

Fig. S13. Reduction priority maps for urban-level optimization by location methods for 44 medium Urban Areas.

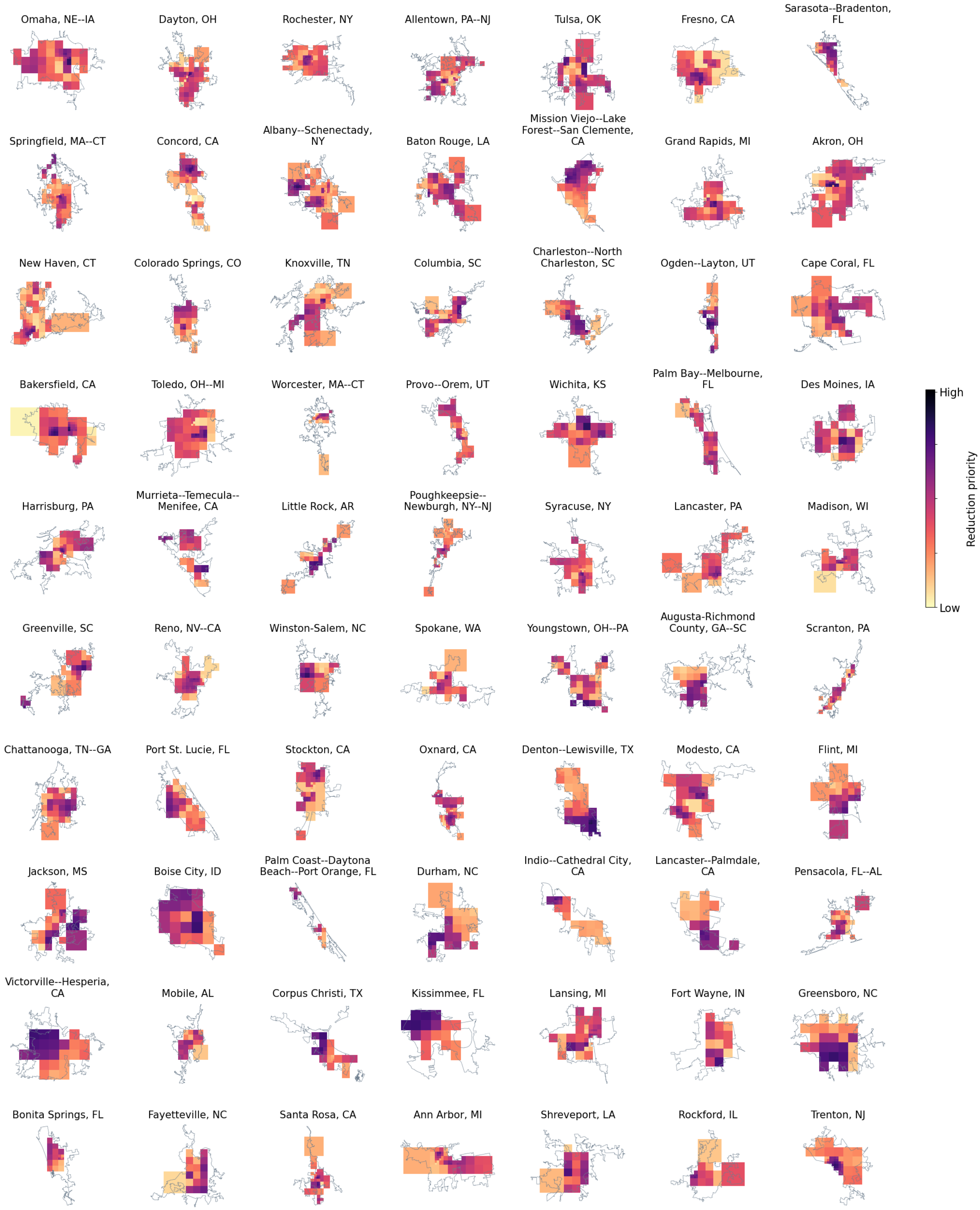

Fig. S14. Reduction priority maps for urban-level optimization by location methods for 70 (out of 381) small Urban Areas.

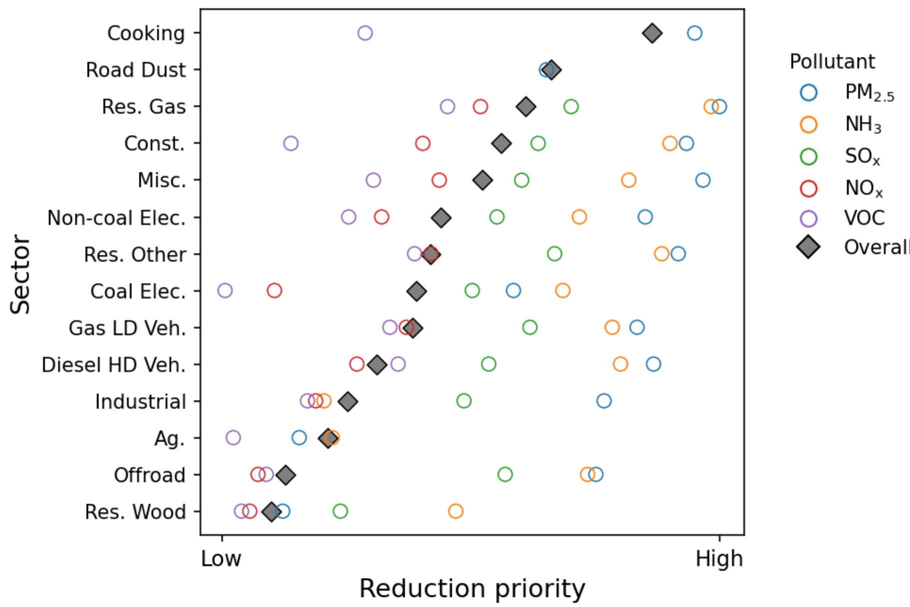

Fig. S15. Emission reduction priority for optimization by sector method. The plot is an alternative version of Fig. 2b-left, where the icons are equally size (so they are more easily visible) instead of sized proportionately to emissions.

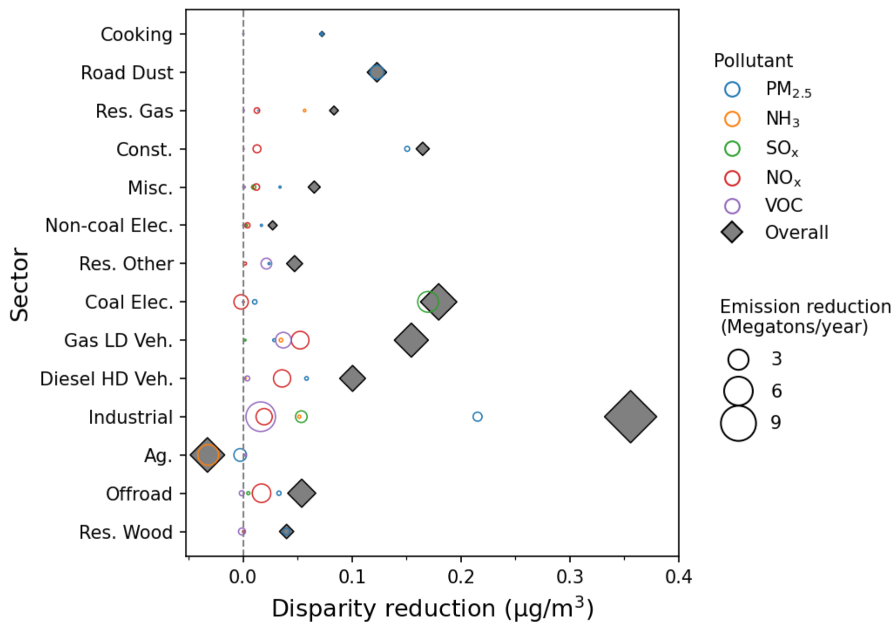

Fig. S16. Disparity reduction for optimization by sector method. The icons sizes are proportionately to emissions.

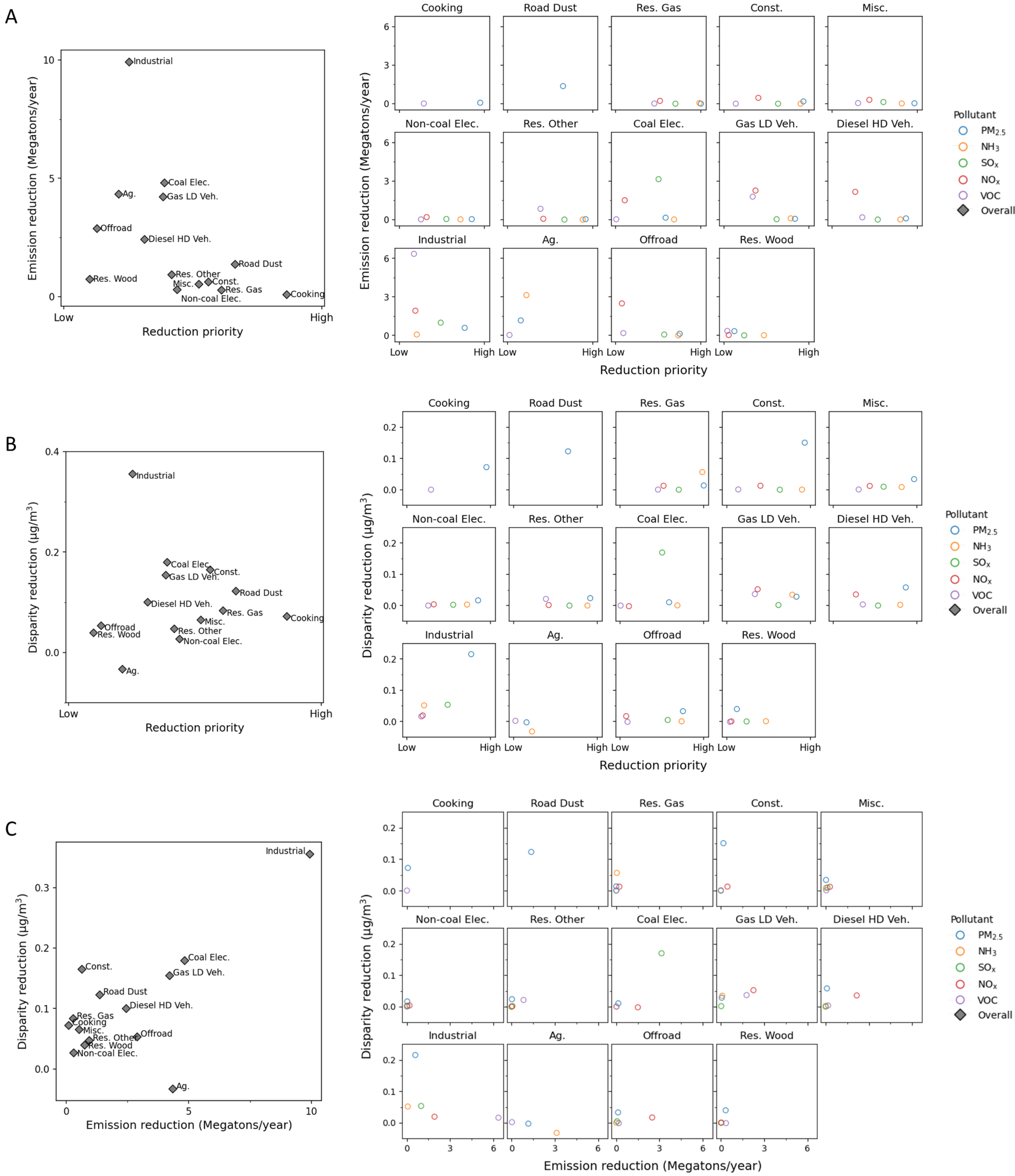

Fig. S17. Emission reduction priority, emission reduced and disparity reduced for optimization by sector method. This figure is an alternative version of Fig. 2b.

Yuzhou Wang, Josh S. Apte, Jason D. Hill, Cesunica E. Ivey, Regan F. Patterson, Allen L. Robinson, Christopher W. Tessum and Julian D. Marshall

Table S1. Multiple linear regression results for reduction priority in relation to racial-ethnic composition, household income, population density, emission density, racial segregation index, and interaction term between racial-ethnic composition and racial segregation in the location (grid cell).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Percentage of non-White population		2.6 (2.3, 2.9)				2.9 (2.6, 3.2)				1.3 (0.6, 2.0)			
Percentage of Black population			4.2 (4.1, 4.3)				4.4 (4.3, 4.5)				6.9 (6.6, 7.2)		
Percentage of Hispanic population				-2.5 (-2.6, -2.3)				-1.4 (-1.6, -1.3)				3.5 (3.1, 3.9)	
Percentage of Asian population					-1.5 (-1.6, -1.4)				-1.0 (-1.1, -0.9)				0.5 (0.3, 0.7)
Median household income	-4.5 (-4.7, -4.3)	-3.9 (-4.1, -3.7)	-2.1 (-2.3, -1.9)	-5.2 (-5.4, -5.0)	-3.6 (-3.8, -3.4)	-2.3 (-2.5, -2.1)	-1.8 (-2.0, -1.7)	-3.8 (-4.0, -3.6)	-2.3 (-2.5, -2.1)	-2.4 (-2.6, -2.2)	-1.8 (-2.0, -1.6)	-3.5 (-3.7, -3.3)	-2.2 (-2.4, -1.9)
Log of population density	7.5 (7.2, 7.8)	7.0 (6.7, 7.3)	6.5 (6.2, 6.8)	8.1 (7.8, 8.4)	8.0 (7.7, 8.3)	8.7 (8.4, 9.0)	8.4 (8.1, 8.7)	10.1 (9.8, 10.4)	10.2 (9.9, 10.6)	8.6 (8.3, 8.9)	8.6 (8.3, 9.0)	10.3 (10.0, 10.6)	10.5 (10.1, 10.8)
Log of emission density	2.8 (2.4, 3.1)	2.3 (2.0, 2.6)	1.4 (1.2, 1.7)	3.2 (2.9, 3.5)	3.2 (2.9, 3.5)	2.9 (2.6, 3.2)	1.5 (1.2, 1.8)	3.6 (3.3, 3.9)	3.4 (3.1, 3.7)	2.8 (2.5, 3.2)	1.6 (1.3, 2.0)	3.8 (3.5, 4.1)	3.6 (3.2, 3.9)
Racial segregation index (non-White/White)						5.4 (5.2, 5.7)				6.5 (6.0, 6.9)			
Racial segregation index (Black/White)							3.6 (3.4, 3.9)				4.9 (4.6, 5.2)		
Racial segregation index (Hispanic/White)								5.6 (5.4, 5.9)				8.1 (7.7, 8.4)	
Racial segregation index (Asian/White)									5.3 (4.9, 5.7)				6.2 (5.8, 6.6)
Percentage of non-White population * Racial segregation index (non-White/White)										1.9 (1.2, 2.7)			
Percentage of Black population * Racial segregation index (Black/White)											-2.0 (-2.2, -1.7)		
Percentage of Hispanic population * Racial segregation index (Hispanic/White)												-4.0 (-4.3, -3.7)	
Percentage of Asian population * Racial segregation index (Asian/White)													-1.7 (-1.9, -1.5)
Number of observations	41,439	41,439	41,439	41,439	41,439	41,070	39,427	40,757	39,354	41,070	39,427	40,757	39,354
Adjusted R-square	0.21	0.21	0.32	0.22	0.22	0.24	0.32	0.25	0.22	0.24	0.32	0.26	0.22

Notes: The table reports the multiple linear regression results for reduction priority in relation to the location's characteristics. Model 1 only includes income, population density, and emission density (three "baseline" variables); Models 2-5 include all the three "baseline" variables as well as racial-ethnic composition; Models 6-9 include three "baseline" variables, racial-ethnic composition, and segregation index; Models 10-13 include three "baseline" variables, racial-ethnic composition, segregation index, and interaction of racial-ethnic composition and segregation index. Models 2, 6, and 10 are for the total non-White population; Models 3-5, 7-9, and 11-13 are for specified racial-ethnic groups (Black, Hispanic, or Asian). Reduction priority at each location is calculated by the emission weighted average priorities for the five pollutants, and is normalized to 0–100 scale (percentile; 100 represents highest priority, 0 represents lowest priority). All the predictor variables are normalized to the interquartile range (IQR), so the slopes represent the increase in disparity reduction priority (percentile) in relation to per IQR increase in each characteristics. For all the models, the p-values for all the slopes are less than 0.05.

References

1. Y Wang, et al., Spatial decomposition analysis of NO_2 and $PM_{2.5}$ air pollution in the united states. *Atmos. Environ.* **241**, 117470 (2020).