RESEARCH BRIEF

FEBRUARY 2022

USC Dornsife
Equity Research Institute

Up in the Air:

Revisiting Equity Dimensions of California's Cap-and-Trade System

by Manuel Pastor, Michael Ash, Lara Cushing, Rachel Morello-Frosch, Edward-Michael Muña, and James Sadd

Acknowledgements

Thank you to the participants in the January 22, 2021 environmental justice workshop for researchers and activists sponsored by USC's Equity Research Institute (ERI) and to participants of the June 3, 2021 Symposium on Climate Change Policy sponsored by the Stanford Energy Modeling Forum and USC's Schwarzenegger Institute for providing comments on earlier versions of this work. Thanks also to Kyle Meng and Danae Hernández-Cortés for various exchanges comparing their analytical strategy and ours and to researchers at California Office of Environmental Health Hazard Assessment (OEHHA) for both discussing approaches to assessing the impacts of cap-and-trade and helping us thoroughly clean the data. Thanks also to Madeline Wander of UCLA for her earlier assistance in assembling the data, to Jeffer Giang and Lance Hilderbrand of USC ERI for their assistance in programming, to Rhonda Ortiz, Dawy Rkasnuam and Sabrina Kim of USC ERI for assistance in copy-editing and final design and to Emma Yudelevitch for her work planning the January environmental justice workshop. Funding for this project was provided by the Energy Foundation.

Executive Summary

California's cap-and-trade program is a key instrument for achieving reductions in greenhouse gas (GHG) emissions under AB32, the California Global Warming Solutions Act. While there are many equity considerations with regard to cap-and-trade, including the ways in which funds are collected and disbursed, one point of controversy between proponents and critics of cap-and-trade systems is whether they can ameliorate or exacerbate disparities in exposure to co-pollutants.

Of course, cap-and-trade makes uneven GHG and co-pollutant reductions likely, given that emitters can choose between reducing their own pollution locally or paying for others to reduce theirs through allowance purchases or offsets. In previous work, we demonstrated that the first several years of California's cap-and-trade program (2013-2015) did exhibit a pattern of unequal changes in GHG and co-pollutant emissions that were consistent with the worries of environmental justice advocates: pollution actually increased in areas with a higher share of people of color or a higher score from a statewide environmental hazard and social vulnerability spatial screening tool called CalEnviroScreen (CES).

This report examines whether such disparities have persisted when we compare two years of recent data (2016-2017) to the two-year period immediately prior to the beginning of the cap-and-trade program (2011-2012). We look at the issue from both the facility and neighborhood perspective, comparing the changes in period averages for covered GHGs, PM2.5, PM10, NOX, and SOX. Dividing either the facilities or the neighborhoods into thirds by the change in pollutant emissions, we find that the facilities and neighborhoods that were "least improved" (which in all cases saw GHG and co-pollutant levels rise) tended to have higher proportions of people of color, people living below 200 percent of the federal poverty level, households that are linguistically isolated, and individuals who are less educated, as well as higher CES scores.

We also look at the change in pollution more directly. To do this, we take advantage of the fact that the state of California labels certain neighborhoods with high CES scores as "Disadvantaged Communities" or DACs. Because both the percentage change and the difference in pollution emissions over the two time periods are not normally distributed, we use non-parametric analysis to compare emissions changes in DACs that hosted facilities to other areas that hosted facilities. In nearly all cases, the DACs saw some improvements in terms of reduced pollutants from cap-and-trade facilities but these improvements were less than those in the non-DACs, with many of the contrasts being statistically significant.

Because a recent working paper by Hernandez-Cortes and Meng has been interpreted as suggesting that cap-and-trade actually reduced environmental disparities, we examine the approach, data, and methods of that paper. We suggest that Hernandez-Cortes and Meng address a different question: whether average local-pollutant emissions from a set of cap-and-trade facilities selected for comparability to non-cap-and-trade facilities declined relative to the non-cap-and-trade facilities. The extrapolation from the average assumes a common percentage reduction at all regulated facilities and necessarily finds more improvement in places that were more polluted to begin with. We highlight this and discuss additional data and methodological issues that may affect the results.

We close by noting that the issue of co-pollutants is only one of many equity considerations that arises from efforts to reduce GHGs. Nonetheless, it has raised concerns and we suggest ways to modify market mechanisms to better acknowledge and address distributional issues. We also call for additional research, including data assembly and modeling, and suggest that a robust discussion of the California experience can help other states and the nation as others also seek to integrate sustainability and environmental justice goals in climate change policy.

Table of Contents

Introduction

California's cap-and-trade program is a key strategy for achieving reductions in greenhouse gas (GHG) emissions under AB32, the California Global Warming Solutions Act. For residents living near large industrial facilities, AB32 offered the possibility that along with reductions in GHGs, emissions of other harmful pollutants would also decrease in their neighborhoods. While GHGs including carbon dioxide (CO2) may not directly impact local health, GHG emissions are usually accompanied by releases of other pollutants such as particulate matter (PM2.5 and PM10) that are harmful to health.

In earlier work (Cushing et al. 2016, 2018), we assessed the pattern of inequalities in the location of GHG-emitting facilities and in the amount of GHGs and particulate matter emitted by facilities regulated under cap-and-trade. To do this, we combined pollutant emissions data from the California Air Resources Board (ARB) mandatory GHG and criteria pollutant reporting systems as well as the U.S. EPA Toxic Release Inventory, data on neighborhood demographics and cumulative environmental health impacts from the American Community Survey and CalEPA's CalEnviroScreen (CES 3.0) tool, and information from ARB about how regulated companies fulfilled their obligations under the first several years (2013-2015) of the cap-and-trade program.

Our analyses focused on what are called "emitter covered emissions," or localized emissions from large point sources in the cap-and-trade system. We found:

1. Facilities regulated under California's cap-and-trade program were disproportionately located near communities of color and communities with socioeconomic disadvantages;
2. Local emissions of GHGs and co-pollutants (PM2.5, VOCs, SOX, NOX, and air toxics) were correlated, meaning that changes in GHG emissions, including changes caused by regulation, likely have implications for the co-pollutant burdens; and
3. During the initial three years of the cap-and-trade program (2013-2015), in-state GHG emissions actually increased relative to the pre-trade period (2011-2012), with variation by industrial sector, and average co-pollutant emissions rose most in neighborhoods with higher concentrations of people of color, residents with low educational attainment and lower socioeconomic status, and in "disadvantaged communities" as defined by CES.

Members of our research team had warned that carbon trading might result in unequal reduction in co-pollutants and exacerbate inequalities in co-pollutant exposure, with the potential to generate inequalities in the pace of improvement or even to potentially increase exposure in "hotspots" (Boyce and Pastor 2013; Shonkoff et al. 2009). However, the possibilities of disequalizing decarbonization did not necessarily mean that it would occur, and we have also suggested that more carefully structured trading programs could ameliorate harms and reduce disparities. In any case, our initial work validated the concerns of many environmental justice (EJ) advocates and has been used to characterize carbon trading as being problematic in terms of equity.

Since the release of that study, there have been some changes to the cap-and-trade program and we have had several more years of program operation. In addition, the State of California has become more sensitive to EJ concerns related to the cap-and-trade program, and changes in knowledge and rules could affect the behavior of both polluters and regulators. We were therefore interested in examining the effects of a more mature carbon-trading system on environmental quality and justice and assessing the extent to which the original patterns may have changed.

While we were preparing our analysis, a new working paper emerged that attracted attention on the part of the press and supporters of carbon trading systems.[1] That study, by Danae Hernandez-Cortes and Kyle Meng of UC Santa Barbara, combined an identification strategy focused on regulated and non-regulated facilities with air modeling of predicted air pollutant concentration. They concluded that cap-and-trade had, in fact, generated more improvement in EJ communities than in non-EJ communities (Hernandez-Cortes and Meng 2020, 2021). While the working paper provoked some early critiques of the authors' approach (for example, on the use of zip codes rather than census tracts to define disadvantaged communities), few critics dug as deeply as they might into the approach and the data (Cullenward and Valenzuela 2020).

In this report, we review the recent trends and show that concerns remain regarding emissions outcomes from cap-and-trade that are not optimal in terms of equity. Specifically, we show that there is a pattern where the deepest reductions in GHG and co-pollutant emissions are occurring in higher socioeconomic status neighborhoods, and that there is less improvement and often worsening of pollutant emissions in neighborhoods that have higher shares of residents of color and/or are defined by CES as "disadvantaged communities" (or DACs). We offer suggestions for research and policy to improve both tracking and performance.

How do we arrive at findings different than those of Hernandez-Cortes and Meng (HCM)? As it turns out, HCM's method may not be amenable to the question with which most EJ proponents are asking – whether there are distinct temporal patterns among those sectors regulated by the cap-and-trade program, and whether the pattern exhibits improvement or not in racial/ethnic or other forms of inequality in emissions burden. Instead, HCM's focus on estimating a "pure" cap-and-trade effect yields an estimate of the average change in pollution loads for cap-and-trade facilities relative to those facilities that were not regulated by the program; to assess equity, they apply that single estimated percent change to every facility in their sample.

However, the whole point of a market-based system is that some polluters will purchase more emission allowances and others will simply cut emissions, necessarily leading to percentage changes that vary across facilities and different industrial sectors. More important for equity considerations, applying an estimated common percentage effect to every regulated facility means that the predicted improvement in absolute terms will be largest where emissions levels were initially highest. For example, if the initial geographic pattern of pollution emissions is unequal by neighborhood and the common cap-and-trade effect is negative, the baseline HCM method is bound to show estimated improvements in equity.[2]

This report starts with a discussion of our data sources and what we believe to be the appropriate analytical approach to address the EJ question of whether there were social and demographic disparities in emissions reductions during the most recent years of California's cap-and-trade program. Then, we turn to the HCM working papers and discuss their analysis; aside from raising general concerns, we note how HCM's designation of facilities as being in or out of the cap and trade system and their actual regression specification can impact their findings. We close with a discussion of what our results mean for future research and policy.

1 See https://www.carbontax.org/issues/carbon-pricing-and-environmental-justice/ for a review of the debate from a pro-market perspective.

2 We say "baseline" because HCM do introduce size heterogeneity as one specification, a topic we discuss in the data appendix.

Analyzing the Equity Impacts of Cap-and-Trade

Setting Up the Data

To analyze the equity impacts of cap-and-trade, we draw on data from the California Air Resource Board's (ARB) Pollution Mapping Tool that includes all facilities in the Mandatory Reporting of Greenhouse Gas Emissions (MRR) system.[3] The facility-level data include information on GHGs, covered GHGs, and also a series of co-pollutants such as PM2.5, PM10, SOX, and NOX. Prior to the development of the Pollution Mapping Tool, GHG and co-pollutant emissions were maintained in separate emissions inventories, and so in our own previous work we were forced to painstakingly link the co-pollutants using information from the CEIDARS inventory available from the California Air Resources Board (Cushing et al. 2018). The co-pollutants are now provided in the newer integrated dataset, although there are still inconsistencies in reporting for different pollutants and issues with the oil and gas facilities which we discuss below.

The data are organized by reporting entity, which is indicated with a variable, ARBID. Each ARBID is also connected to a geocode permitting researchers to locate the reporting entity.[4] In most cases, the reporting entity (with the geocode) is the site of the emitting facility itself; in the case of about half of the oil and gas facilities, the reporting entity can be a headquarters or management office that owns multiple emitting facilities and is not necessarily itself a site of GHG or co-pollutant emissions. Analysts at the California Air Resources Board provided us with a list of geocodes for these oil and gas emitters; there were 24 ARB oil and gas reporting entities associated with 288 underlying facilities. Most of the reporting entities just have a small number of "satellite" emitting facilities but one entity has 141 underlying emitter locations.

3 See https://ww3.arb.ca.gov/ei/tools/pollution_map/ The specific pollution data we used was downloaded on April 17, 2020. This is important since there are occasional revisions to the data, and downloads on a different date may have slightly different numbers or even designations as to whether a facility is part of the cap-and-trade system or not. We discuss these issues both in the main text and in a data appendix.

4 The geocodes are not as accurate as they could be; the latitude and longitude coordinates are only available at the three decimal place which means that there could be locational imprecision of up to 110 meters. However, since we are using relatively large distance bands (we offer a 2.5 mile buffer analysis in the text but other buffer lengths yield similar results), the level of accuracy here is reasonable and a spot-check of geo-coded locations indicates that they are close to the stacks.

Figure 1. Oil and Gas Emitter Locations Relative to Reporter Location for One MRR Facility

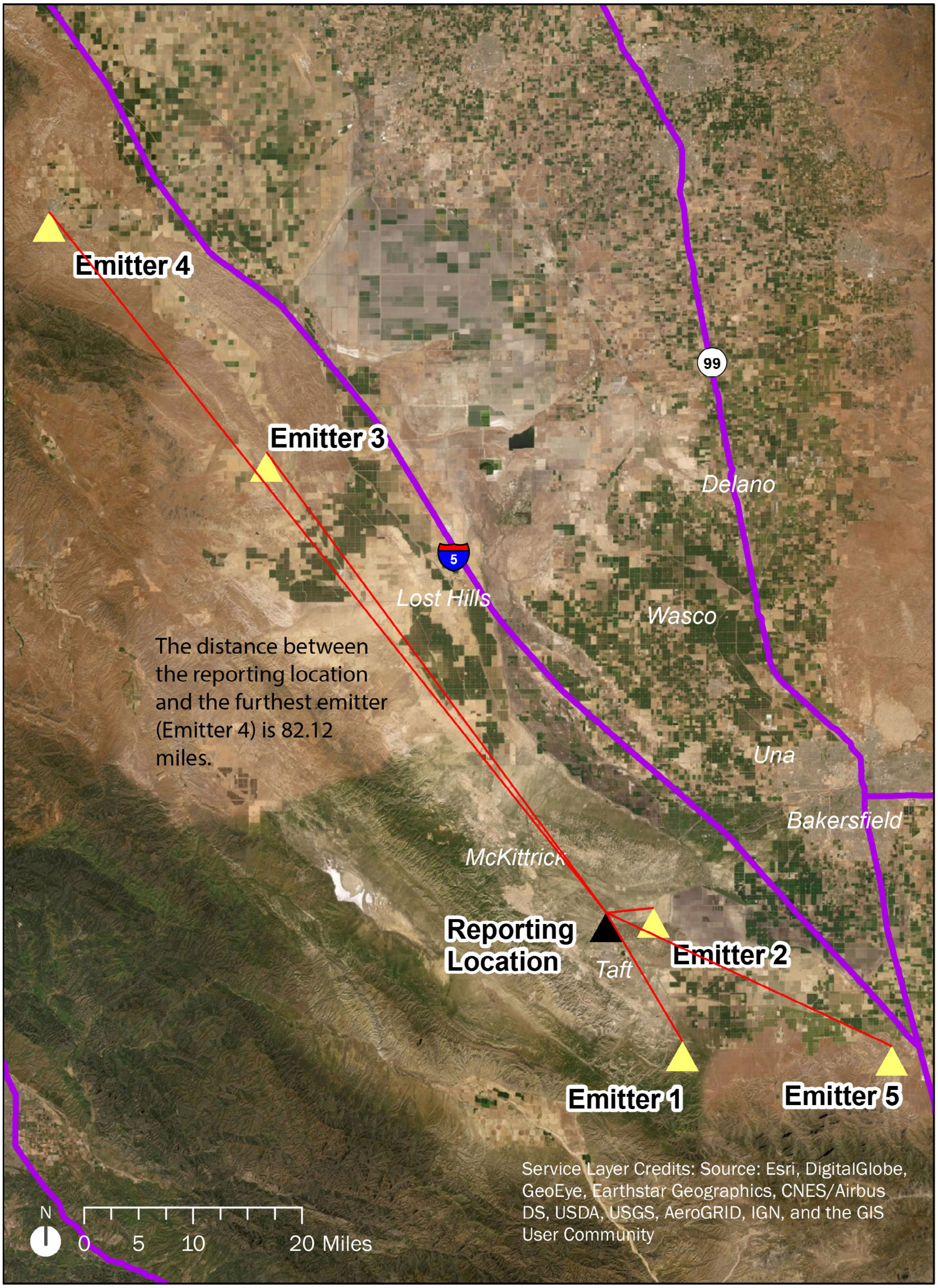


Figure 1 shows the pattern for one reporting facility that represents five emitters, only one of which is close to the reporting facility. For those unfamiliar with the scale, the northernmost emitter is near Coalinga, California, and is just over eighty miles as the crow flies from the reporting facility near Taft. Comparing the demographics around the reporting facility could seriously misstate the demographics around the emitter (or say, the combined demographics around emitters).

Other data inconsistencies in the MRR and Pollution Mapping Tool include address or zip code changes over time, but these make little difference to our analysis because we rely on the actual geocodes. A bigger issue is that the same CARB facility may report different NAICS codes (industry classifications) from year-to-year. Beyond straightforward changes, such as shifting from a four-digit code to a more precise six-digit code, there are nine facilities under the AB32 cap-and-trade system that change NAICS codes over time. These account for less than one percent of the total covered GHG emissions in 2017, and none shift in terms of their classification by what ARB designates as its primary sector.[5]

A bigger data issue is that there were some inconsistencies in terms of which entities were designated as being part of the cap-and-trade system. We utilized a CARB database downloaded in April 2020 which indicated that 310 entities were regulated by cap-and-trade. However, a careful check of the data, including the use of the online Pollution Mapping Tool, suggested that some entities were reporting covered emissions in 2016 and 2017 and should have been designated as being regulated. Working with colleagues at the Office of Environmental Health Hazard Assessment (OEHHA), we identified ten additional entities that should have been tagged as cap-and-trade.[6]

As a result, we wound up with a database covering up to the year 2017 that had 784 reporting entities, of which 320 were regulated by cap-and-trade, 336 were required to report to MRR but were not emitting a sufficient level of GHGs to actually be part of the cap-and-trade program, and 128 were initially in the MRR reporting but dropped out before policy implementation and are not assigned a “Yes” or a “No” as cap-and-trade tags. Many of these drop-out facilities were involved in biomass combustion and were originally required to report to MRR but then had no compliance obligation under cap-and-trade, and so stopped being recorded in that dataset.[7] Because we were interested in the question of whether there were relative shifts in the burden of pollution from cap-and-trade facilities – that is, what was the pattern of increases and decreases over time in those facilities regulated under cap-and-trade –we focused on the 320 in our own work.

A series of considerations reduced that sample of cap-and-trade entities further. We were particularly interested in the difference between the two years pre-trade (2011-2012) and the most recent two years (2016-2017) of the last compliance period. Our early work analyzed the pattern in the program’s first several years (2013-2015) and noted that the pattern in those years might have been anomalous because of the increase in in-state electrical generation due to shifts toward cleaner energy sources. Moreover, there has been sufficient time for learning and adjustment which may have also changed

5 While it is not generally relevant for this analysis, we assigned the most recent NAICS codes for all facilities. We did the same for the address and zip code which occasionally shift over time as well, usually because a street is spelled differently in one year.

6 They also provided information that led us to follow them in dropping several years of data where a facility had been shut but was still reporting either zero GHGs or a negligible amount.

7 There are also a few cement plants that reported emissions in 2008 but were in the process of shutting down, and so they also enter the dataset with observations for several years but had closed and hence spent no time under cap-and-trade regulations. These observations are coded with neither a “yes” nor a “no” but rather a blank. Nearly all observations with the “blank” on the cap-and-trade variable drop out of the database in 2010, with the remaining 4 dropping out in 2011.

firm decisions and helped to ameliorate disparities.

To capture both before and after and to ensure some continuity along the way, we initially required that any entity considered had at least one observation for total GHGs in each of three time periods – the pre-trade period of 2011-2012, the previously studied period of 2013-2015, and the most recent two years, 2016-2017. We should note that while GHG and particulate emissions estimates prior to 2011 are available, they are not comparable due to changes in reporting requirements and emissions accounting effective in 2011, an issue discussed in the data appendix.

Using that criteria, we wound up with 289 entities of the 320 cap-and-trade entities in the database; the entities dropping out either started after cap-and-trade (so were missing data for the first period) or closed before 2016-2017.[8] After some experimentation with different inclusion criteria, we focused on facilities that reported positive GHG emissions for every year of the study period (2011-2017), which brought us to 262 facilities.[9] Finally, we excluded three more facilities that had data issues, including, for example, the Exide Technologies facility which had a large decline which was due to a shutdown not related to cap-and-trade but rather to extensive lead emissions in its local neighborhood.[10] If summed up over the whole period, the included facilities (N=259) represent 95% of the covered GHGs. This loss in pollution coverage – which is largely due to the requirement of having all years of GHG data – is, we think, a reasonable tradeoff so that we are looking at a consistent series.

This process created the basic set of facilities for analysis but there were further considerations for each of the co-pollutants. While we required covered GHGs for all years, we had a looser standard for the co-pollutant measures because far more had missing observations for some years, likely due to different and less stringent and less frequent reporting requirements on those pollutants which are integrated into the Pollution Mapping Tool from the CEIDARS data. For those, we required that they be in the overall GHG set but only required at least one observation for each of the time periods so we could calculate an average for the pre-compliance and the other two periods.[11] After that is done, we have slightly fewer facilities for each co-pollutant than we do for the GHG set but the loss in overall pollution is quite minor: within the dataset that reliably includes GHGs for all years, we retain between

8 Some facilities that were shut in the last period do report data although sometimes they report a total GHG level of zero. We carefully did searches to document whether facilities had shut and also collaborated with research at OEHHA to clean the data of these cases.

9 We initially hoped to have a looser constraint, requiring just one observation for each of the time periods, which is what we do for co-pollutants. However, after doing Google searches on the 27 cases where an occasional year was missing, it seemed that the vast majority were from facilities starting up, shutting down, and some retrofitting in the middle of the time period for reasons largely unrelated to the cap-and-trade policy. The few cases we could not document looked like they had a similar pattern. Requiring positive GHG emissions was imposed because every case where total GHGs are reported as zero seems to be a shutdown year for the facility in question.

10 We also excluded a peaker plant that had observations for all years but reports on its website that it began operations in 2012 and had a pattern that indicated this was so. Finally, we excluded one oil and gas entity where there were multiple changes in ownership and reporting jumped up and down dramatically as the entity reported the emissions from its on-again, off-again business partner; we verified this was a problem case with colleagues at OEHHA.

11 Unlike with the total GHG series, we do not automatically exclude observations of zero since an examination of the underlying CEIDARS database that is the source for the co-pollutants suggested that these might be due to rounding very low numbers. However, we did strike cases where the first two years (2011-2012) were both zero or had one missing year and the other was zero. We did this because inspection of those cases (all of which had positive GHGs) suggested that that several of these could be due to CEIDARS reporting lags since there was a subsequent sharp increase in co-pollutants while the GHG series remained steady over time; moreover, it would have been impossible to calculate a rate of change over the two periods with a zero starting point. In the case of PM10 and PM2.5, this dropped six reporting entities; the process dropped no facilities due to NOX and 16 due to SOX. As might be surmised, dropping facilities with very low levels of pollution to start with has extraordinarily modest impacts on coverage.

94-96% of reported emissions for PM2.5, PM10, NOX, and SOX.

We should stress again that "facilities" may be a bit of a misnomer here; these are really reporting entities which are generally emitter locations as well. However, this is not the case for all oil and gas reporters, some of which have actual emission sites that are not the same as the geo-coded addresses in the main ARB database. Using the information provided by ARB, we created a crosswalk between the oil and gas reporters and the actual oil and gas emitter locations. As in Cushing et al. (2018), we assigned the reported pollution in equal shares to each of the individual emitters; a more complex approach, in which we attributed a share of various pollutants based on reporting by the oil and gas facilities into the CEIDARS database, produced little difference in the outcomes.[12]

Basic Patterns

As with our previous work, we report results at the facility level or the neighborhood (census block group) level. We do this by creating a point-distance matrix that links every facility location to every census block group centroid in the state, taking care to ensure that we are using the right location for the underlying oil and gas facilities. We can then either aggregate certain neighborhood characteristics to the facility or allocate facility-level pollution to the neighborhood. In both cases, we use a simple buffer rule for defining the facility-neighborhood relationship. Throughout the statistics presented in this report, that buffer is 2.5 miles; wider buffers, or a more complex distance-weighted approach, yield similar results.

It is useful to start with a quick look at the facility locations and changing emissions. We start with maps detailing the facilities that reported positive covered GHGs over the period 2016-2017.[13] Figure 2 shows the full state; Figure 3 shows the pattern in Los Angeles County, with the background layer for both maps being census tracts ranked into quartiles by the percentile score assigned to each tract by CES 3.0. One can visually see, particularly in Los Angeles County, that many (but certainly not all) of the facilities are in or near the top quarter of the CES scores; these, as we note, are generally considered by the State of California to be "disadvantaged communities" (or DACs).

12 In that exercise, we tried to attribute the potential share of pollution to each individual oil and gas emitter by taking the pollutant with the highest correlation with total GHGs, PM10, and extracting that from the CEIDARS inventory. We then used the share of total PM10 coming from each reporting entity to allocate a share of other pollutants as well. Because the facility identification numbers can repeat across counties and air districts, we created a unique identifier based on facility, air district, air basin, and county, and then matched in PM10 by year. While potentially more accurate, it is also significantly more complex and can generate its own data problems; since the equal allocation strategy was our previous approach and one that passed peer review in Cushing, et al., (2018), we stick with it in the analysis.

13 For the purposes of this visualization, we are using a different break of facilities than we do for the facility-level analysis below. As we explain there, we do the analytical ranking at the level of the reporting facility and aggregate neighborhood demographics when an oil and gas reporter has multiple emitting locations to the reporter; in this simple visualization, we simply rank the emitter, implying that the lowest third includes a significant share of the oil and gas multiple emitters (since the overall reported GHG is spread among them as we discuss below).

Figure 2. Cap-and-Trade Facilities and CES Scores in California, 2016-2017

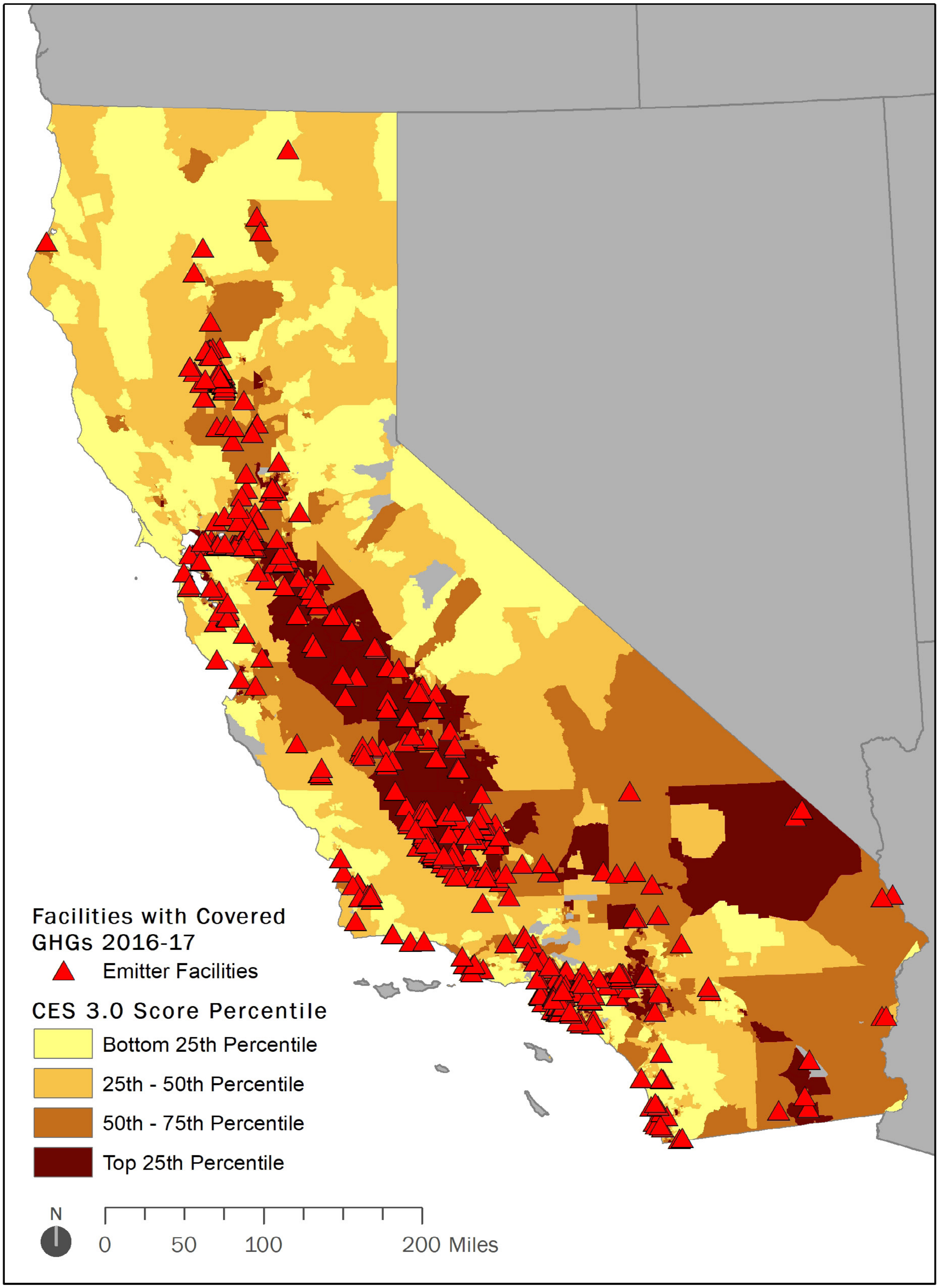
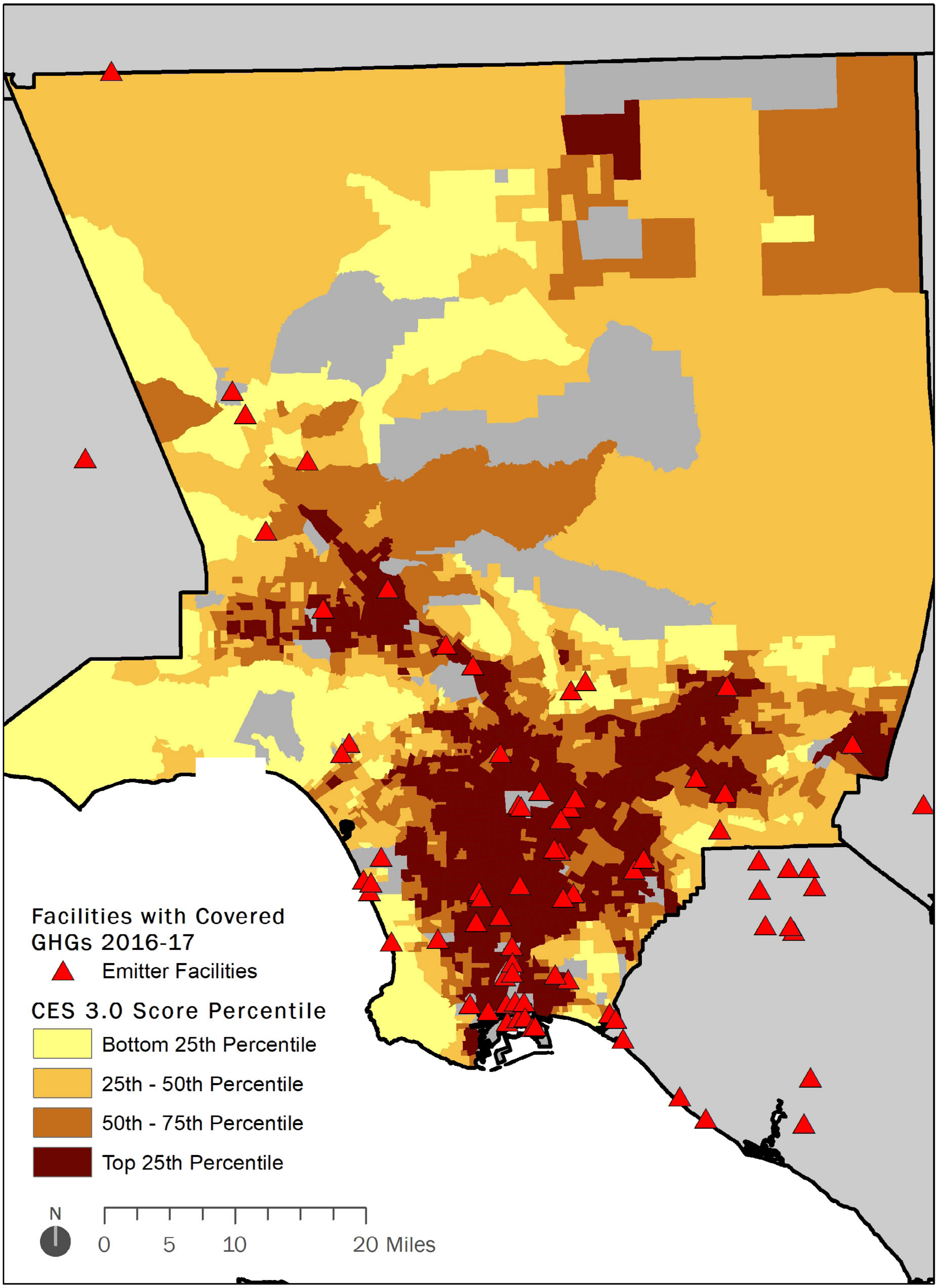
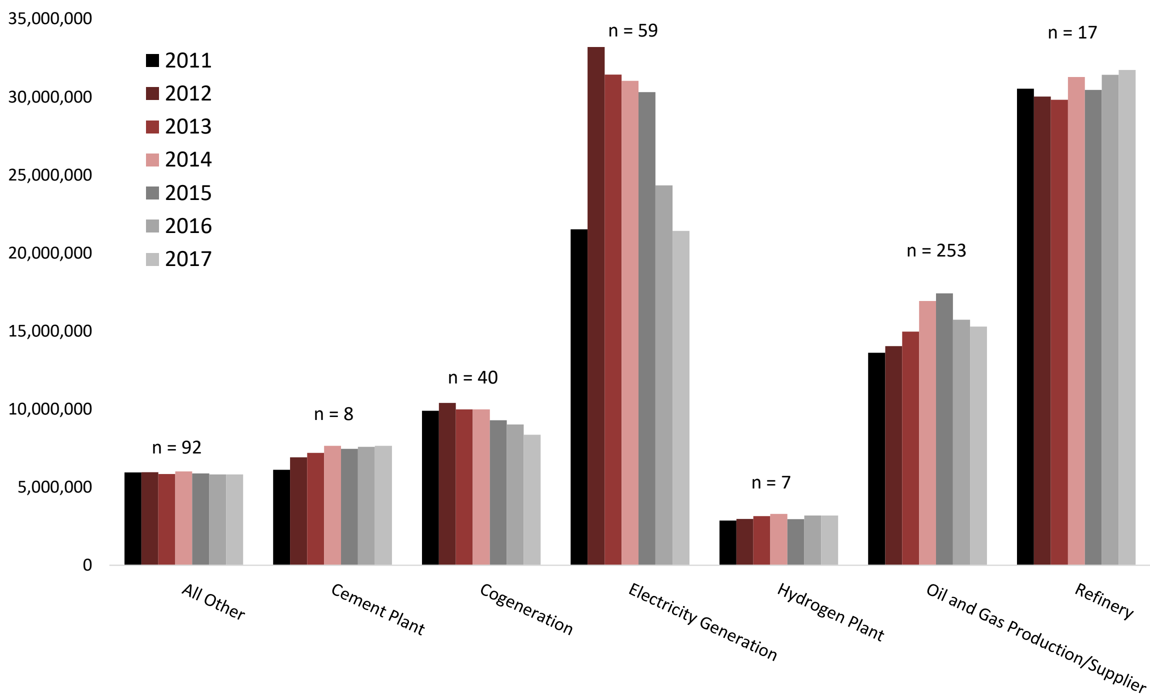


Figure 3. Cap-and-Trade Facilities and CES Scores in Los Angeles County, 2016-2017



Another basic pattern to examine is the change in GHG emissions over time by industry sector (measured in metric tons of CO2 equivalent, or MTCO2e).[14] In Figure 4, we do this for "covered" GHGs, where covered means regulated under the cap-and-trade program and, in the case of the data coming from the mapping tool, locally emitted.[15] As one can see, the substantial improvements in covered GHGs over time are in electrical generation (after an initial increase), cogeneration, and in the latter years, oil and gas production. The pattern for electrical generation is particularly important since the high levels in 2013-2015 compared to the mean for 2011-2012, which seems to result from a shift in electricity sourcing from out-of-state coal generation to in-state natural gas generation, played a role in the dynamics examined in our earlier work. Also of note, is the lack of improvement in the refinery sector, a pattern of special concern for the low-income communities of color that often live near the fence lines of refineries.[16]

Figure 4. Total Covered GHGs (MTCO2e) by Industry Sector, California, 2011-2017



14 Note that the dataset examined here requires that a facility have data on covered GHG for all years so that no shifts in an annual basis are due to a facility closing or opening for reasons unrelated to the cap-and-trade system. As noted, in the analysis below, we also exclude other facilities based on their documented history; adopting those exclusions yields a similar pattern.

15 In nearly all cases, covered GHGs and total GHGs are identical. In some cases, they diverge slightly, often because of some small amount of biomass combustion. We should add that we are only considering "emitter covered" emissions in the Pollution Mapping Tool. As ARB notes on the Pollution Mapping Tool website, "Since 2011, the new reporting regulation requires total GHG emissions that include not only emitter's GHG emissions, but also GHG emissions from fuel suppliers. In order for emission numbers to be consistent and comparable throughout the years in the tool, fuel suppliers' GHG emissions were deducted from the total facility GHG emissions. In the reporting website, on the other hand, the total emissions for a facility include emissions from all sources. These differences between the tool and the reporting website are most likely to occur in 2011 and subsequent years, specifically in refineries and oil and gas production facilities." Finally, there were a few cases where covered GHGs were missing but total GHGs were available; to fill in those seeming data gaps, we created a variable which was the average of covered GHGs to nonbiomass GHGs for each of those facilities and applied that ratio to estimate what would have been reported as covered.

16 The reported number of facilities also provides a sense of which sectors might be most important in terms of focusing on individual emitters; certainly, refineries fit that bill. The number reported for oil and gas producers includes other combustion sources and is the number of emitter sites rather than reporting entities; the number of reporting oil and gas entities with data for all years is 38, of which 15 are reporters which map to multiple locales.

Facility or Source Analysis

So what was the nature of the facility-level change relative to neighborhood characteristics? To get at this question, we constructed a "neighborhood" or catchment area for each facility by calculating, for example, the share of people of color or the share of households that are linguistically isolated for all the block groups with centroids that fall within 2.5 miles of the emitter location. We also constructed a neighborhood CES percentile score for the facility by taking the average of the CES percentile associated with the block groups within 2.5 miles of the emitter location.

While this is completely straightforward for most of the facilities, some of the oil and gas data are a bit more complex since the data can come from one reporting entity that can have multiple emitter locations. For these, we used the correct emitter location to collect the demographics but summarized them to a single reporter (so for the oil and gas reporter depicted in Figure 1, we collected demographics from each of the emitter locations and calculated various measures as though this was one catchment area).[17, 18]

We then calculated the facility-level change between 2011-2012 and 2016-2017 for covered GHGs, PM2.5, PM10, NOX, and SOX, using the mean for each two-year period for each of the pollutants. For each of the pollutants, we ranked the facilities into thirds by the change in emissions, with the largest negative change or reduction designated as "most improved." As we note below, the third of the facilities that we mark as "least improved" actually meant an increase in GHGs and co-pollutants, suggesting that we could actually term these "worsening" facilities or "worst performers;" we chose to use the language of relative improvement since there was improvement on average for the cap-and-trade facilities as a whole.[19] We used this categorical approach to distinguish between those facilities (and later those neighborhoods) that experienced the most dramatic, rather than the most modest, shifts since a slight move up (or down) might be of less concern (or celebration) than a larger swing in either direction.[20]

In doing the analysis, we examined the most significant outliers in terms of increases and decreases and noticed that many of the outlier entities exhibited a pattern similar to those entities with a documented history of shutdowns or reboots unrelated to the cap-and-trade program. We thus trimmed each sample of outliers, eliminating five percent of the reporters (accounting for less than 1 percent of the covered GHGs), with half of that five percent coming from those that reported the largest percentage increases and half from those reporting the largest percentage decreases.[21] In

17 We took this approach for several reasons. First, in some cases, the oil and gas emitters under a single reporting facility actually share the same geo-coded location but are listed as separate emitters and so should be considered together. Second, this makes the oil and gas "facility" more comparable, in size and impact, to the other entries in the ARB database while still retaining the particular demographics of who is impacted. If instead you "spread" the pollution loads from the parent reporter to all the child emitters, those emitters are then very likely to show small increases or decreases and dominate the middle of any distribution of facilities. An alternative approach of calculating demographics by emitter location yields similar but less clear results as does an approach that simply eliminates the oil and gas facilities to test for sensitivity. We thank a colleague for suggesting this approach.

18 The CES scores for the reporting entity were calculated as the average of all the underlying average emitter scores.

19 The exception from the facility view was on SOX where the median rate of change for the facilities examined was exactly 0; for all other pollutants, the median rate of change was negative, indicating improvement.

20 We should note an interesting pattern on the source side: using the initial 2011-2012 level of covered GHGs as the baseline, the larger facilities are actually more likely to be in neighborhoods that have a smaller share of people of color or people who live in households making less than 200 percent of the poverty line. We discuss this more in the data appendix.

21 Expanding the buffer zone to five miles to retain some of these facilities does not change the pattern of results.

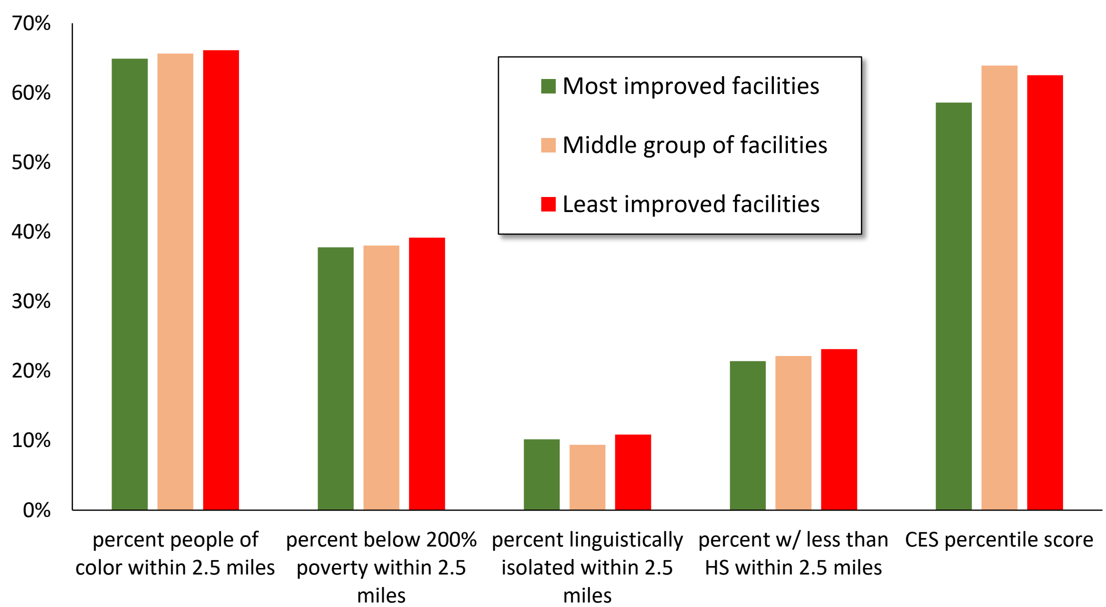
addition, in the analysis below, if there was no block group centroid within 2.5 miles, the facility in question drops out of consideration.

So what do we find for the facilities? To get a sense, we ranked the facility change in two ways for the five pollutants, first by percentage change and second by absolute change, assigning each to thirds by the change in pollution. Since the patterns are generally the same and our subsequent presentation from a neighborhood view is more sensible using a percentage change approach, we present percentage change in the charts below (although we do include the absolute difference in nonparametric tests in which we examine what happened to pollution levels in disadvantaged communities).

To give a sense of how the facilities break into thirds, when ranked by percent changes in covered GHGs, the third of the facilities that were most improved facilities saw GHG decreases of 11 percent or more (i.e., less than -11 percent) while the worst third of the facilities saw increases in covered GHGs of 4.5 percent or more, with the middle range being a mix of increases and decreases. There is a similar range of improvements and worsening by facility terciles for the other pollutants. For PM2.5, the most improved third saw decreases of 20.4 percent or more and the worst third saw increases of 9.4 percent or more; similar markers for PM10 were -19.3 percent and 6.9 percent, for NOX, -18 percent and 12 percent, and for SOX, -14.8 percent and 11.1 percent.

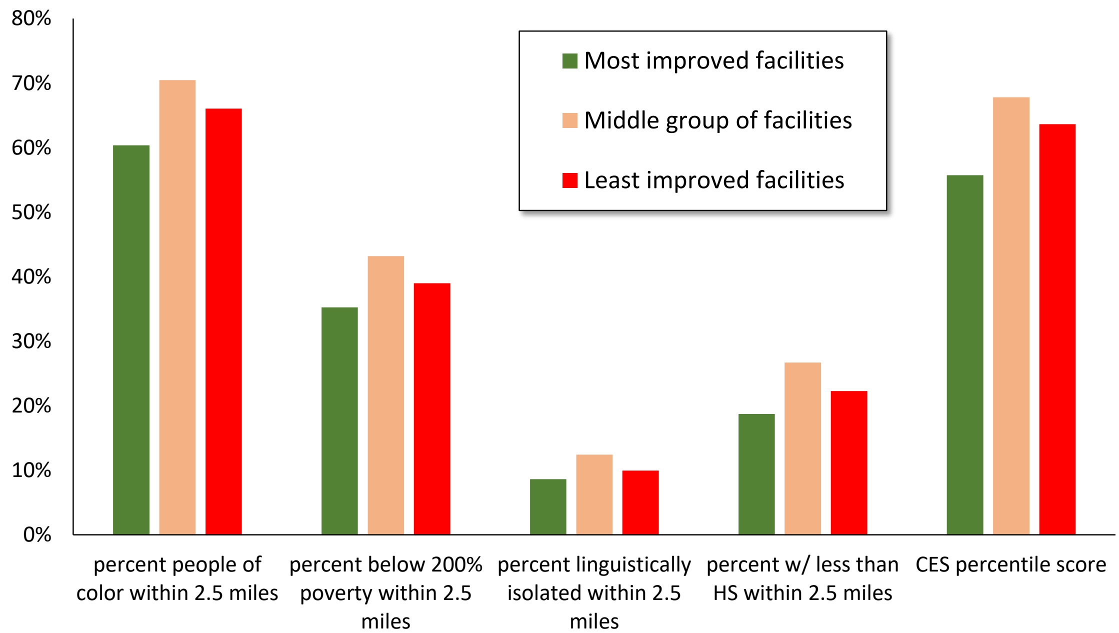
As can be seen in Figure 5, the least improved (or "worst") facilities in terms of the percentage change in covered GHGs tend to have a higher share of people of color, people living below 200 percent of the poverty line, households that are linguistically isolated, adults with less than a high school education, and somewhat higher CES scores. However, the pattern is not perfectly monotonic: facilities in the middle group (between the most improved and least improved) occasionally have values that are not consistent with a straight-line pattern. On the other hand, when comparing the best and worst thirds of facilities ranked by pollution changes, we see patterns consistent with environmental justice concerns.

Figure 5. Facilities Ranked by % Change in Covered GHG, 2011-2012 to 2016-2017 (measures summarized to the ARBID reporting entity)



Since many of the concerns about cap-and-trade are about co-pollutants, in Figure 6, we look at facilities grouped in thirds by their change in PM2.5. While the pattern is non-monotonic, the figure shows a pattern with the least improvements (that is, actual worsening in the levels of PM2.5) occurring at facilities where the surrounding community is more likely to have a high share of people of color, poorer families, linguistically isolated households, or lesser-educated adults, and a higher CES scores.

Figure 6. Facilities Ranked by % Change in PM2.5, 2011-2012 to 2016-2017 (measures summarized to the ARBID reporting entity)



In order to save space on charts, in Figure 7, we list the patterns for the three other co-pollutants of concern. The pattern of PM10 is much like that depicted above for PM2.5, although a bit more muted for the middle group of facilities. The pattern for NOX and SOX is monotonic, with environmental justice measures (such as percent people of color, percent below 200 percent of the poverty, and CES scores) rising consistently as we move from most improved to least improved facilities. Overall, the general finding is that the facilities falling into the worst category – which in this case all produced increases in the various pollutants – are more likely to have communities that exhibit environmental justice characteristics, including generally higher CES percentile scores.

Figure 7. Facilities Ranked by Difference in Three Pollutants, 2011-2012 to 2016-2017 (summarized to the ARBID reporting entity)

		percent people of color within 2.5 miles	percent below 200% poverty within 2.5 miles	percent linguistically isolated within 2.5 miles	percent w/ less than HS within 2.5 miles	CES percentile score
PM10	Most improved facilities	61%	35%	9%	19%	56%
	Middle group of facilities	70%	41%	11%	25%	66%
	Least improved facilities	66%	41%	11%	24%	65%
NOX	Most improved facilities	63%	37%	9%	21%	59%
	Middle group of facilities	64%	38%	9%	21%	62%
	Least improved facilities	68%	42%	12%	26%	65%
SOX	Most improved facilities	60%	34%	8%	18%	55%
	Middle group of facilities	66%	39%	10%	23%	63%
	Least improved facilities	67%	43%	12%	26%	66%

Another approach to the facility analysis – and one more parallel to what we do with the neighborhoods – is to classify the facilities not by their changes in pollutants but rather by an EJ marker, and then use that to consider shifts in pollution load. One handy marker is whether or not the surrounding community – defined in our case by the character of the block groups within 2.5 miles of the facilities – meets the California benchmark for being a "Disadvantaged Community" or DAC. This is a meaningful measure (and one emphasized by HCM in their work) because the DAC designation takes into account not only population vulnerability but pre-existing pollution burdens from multiple sources, underlining the significance of cumulative impact.

We designate facilities as DAC-affected when the average CES percentile score for block groups in the entity's buffer is equal to or above the 75th percentile, the cutoff the California Air Resources Board uses to designate a census tract as a "Disadvantaged Community;" roughly 35 percent of the facilities in our analysis are thus tagged as DAC-affecting. Since the percentage changes and the absolute differences in the pollutants are not normally distributed, in Figure 8, we present medians and the interquartile range, as well as the results from nonparametric median and Mann-Whitney ranked sum tests to determine whether there are significant differences in the changes in pollution for facilities located in DAC versus non-DAC neighborhoods.

Figure 8. Changes in Emissions for Facilities in DAC and non-DAC Neighborhoods

	Non-DAC Facilities (Median, IQR)	*DAC Facilities (Median, IQR)*	*Median Test*	*Highest Rank*	*Mann-Whitney*
% Change in Covered GHG	-3.1% (-20.5%, 8.1%)	1.5% (-10.8%, 8.8%)		DAC	#
Diff in Covered GHG (MTCO2e)	-3,799 (-26,077, 6,739)	818 (-7,832, 5,727)		DAC	#
% Change in PM2.5	-9.0% (-37.4%, 20.0%)	1.8% (-16.7%, 29.3%)	#	DAC	*
Difference in PM2.5 (tons)	-0.40 (-3.05, 1.00)	0.10 (-1.35, 0.95)	**	DAC	#
% Change in PM10	-6.9% (-36.9%, 13.6%)	3.1% (-13.0%, 23.8%)	**	DAC	*
Difference in PM10 (tons)	-0.50 (-3.85, 0.85)	0.15 (-1.40, 1.35)	*	DAC	#
% Change in NOX	-3.8% (-28.5%, 19.9%)	-1.6% (-18.9%, 29.1%)		DAC	
Difference in NOX (tons)	-0.85 (-9.65, 2.90)	-0.05 (-9.65, 1.80)		DAC	
% Change in SOX	-5.1% (-33.3%, 19.1%)	3.9% (-12.5%, 37.5%)	**	DAC	***
Difference in SOX (tons)	-0.05 (-0.70, 0.10)	0.10 (-0.20, 0.45)	**	DAC	***

*** significant at the .01 level
** significant at the .05 level
* significant at the .10 level
# significant at the .20 level

Figure 8 shows that DAC-affecting facilities consistently underperform in terms of emissions reductions compared to the non-DAC facilities. In fact, the medians for each group nearly always indicate absolute increases for the facilities in DAC neighborhoods and absolute decreases for the non-DAC facilities; the one exception is the pattern for NOX where there is a decline for both groups, with a steeper decline for the non-DAC group, although these differences were not statistically significant. However, one can see from the interquartile ranges (IQRs) that there is a wide range for the variables in question and so we need to dive deeper to understand the significance (or not) of this result.

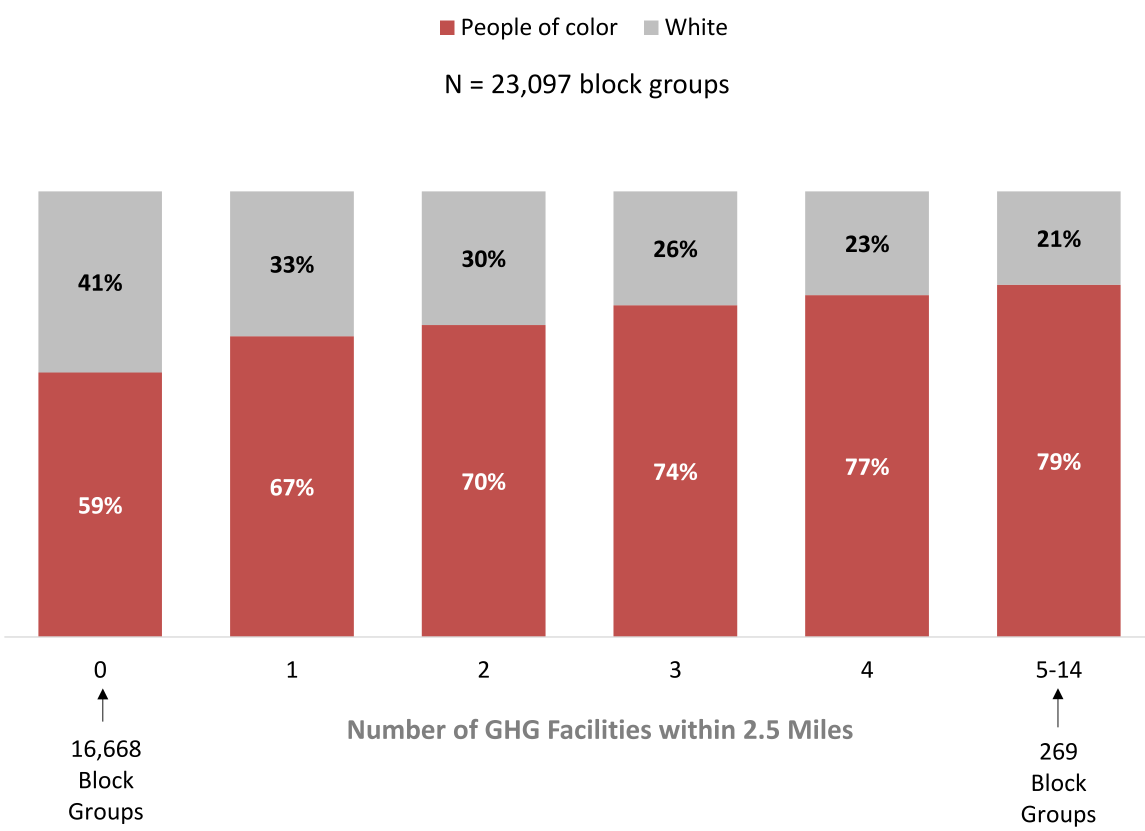
For this, we rely on two non-parametric tests. The median test compares the shares of DAC and non-DAC facilities that are above and below the overall median for the various emissions; by this measure, we find higher shares of above-median polluters in DAC neighborhoods in all cases with statistically significant differences for PM2.5, PM10, and SOX. The non-parametric Mann-Whitney test ranks all cases and then compares the mean ranks of DAC and non-DAC facilities. By that measure, there are statistically significant differences in the cases of covered GHG and SOX, and for the percentage change measures for the two PM measures, with DAC-adjacent facilities always performing worse than the non-DAC facilities. For NOX, we find that DAC facilities always do worse (they have a higher mean rank of pollution change) than non-DAC facilities but the difference is not statistically significant.

Neighborhood or Receptor Analysis

We now turn to a neighborhood or receptor analysis, looking at the potential burden from GHG facilities on block group populations. Note that from this lens, we are taking full advantage of the locational accuracy gained from properly geocoding oil and gas facilities. While for the subset of the regulated facilities, we were previously summarizing the demographics to the parent facilities to consider the "surrounding neighborhood" (which might really be an amalgam of several neighborhoods), we now instead distribution the various calculated pollutant emissions to emitter sites and consider the block group to be the neighborhood. This approach accounts for the potential cumulative impact of clusters of facilities in particular neighborhoods.

Figure 9 looks at all California block groups for which there is population data and examines the characteristics of those neighborhoods by the number of facilities within 2.5 miles.[22] As can be discerned, there is a large racial disparity between those block groups with no nearby GHG facility and block groups that have one or more. Moreover, as the number of facilities near a block group rises, so does the percent people of color living there.

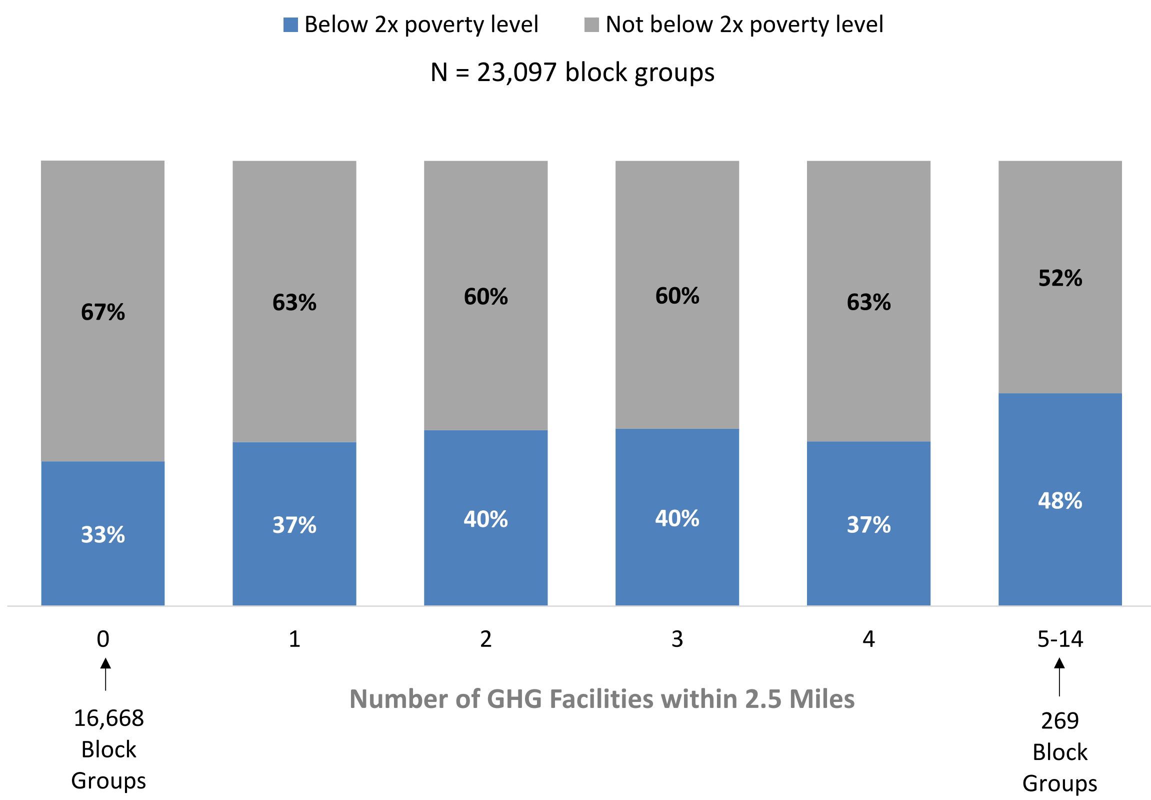
Figure 9. Racial Composition of Neighborhoods by Number of Nearby Facilities



22 ; all groupings are population-weighted but the pattern is the same if we choose to eschew population-weighting and instead derive simple means.

Figure 10 shows the same groupings by number of facilities but with a look at the percent of people living below and above 200 percent of the poverty level. There are also differences between those block groups with no facility nearby and those with a facility nearby. The share of those in the higher poverty category also generally rises with the number of facilities. However, in both cases, the pattern is a bit less pronounced for poverty than it is for race, a finding in keeping with much of the environmental literature in which the exposure gradient is steeper by race than by class (Cushing, Faust, et al. 2015; Cushing, Morello-Frosch, et al. 2015; Ringquist 2005).[23]

Figure 10. Neighborhood Poverty Prevalence by Number of Nearby Facilities



So what were the relative improvements in neighborhoods? To consider this, we take a relatively simple approach that is parallel to the facilities analysis. We attribute to each block group the pollution of any facility within a 2.5 mile radius that made the initial screens discussed above. These included: (1) requiring that facilities have all years of GHG data; (2) eliminating three facilities that we could document had very large increases or decreases for idiosyncratic reasons other than cap-and-trade; and (3) taking a slightly less strict approach with the facilities that had all years of GHGs but were missing co-pollutant emission data for a particular year. We then added all the emissions impacting any block centroid from all nearby facilities, and calculated the percentage change in covered GHGs,

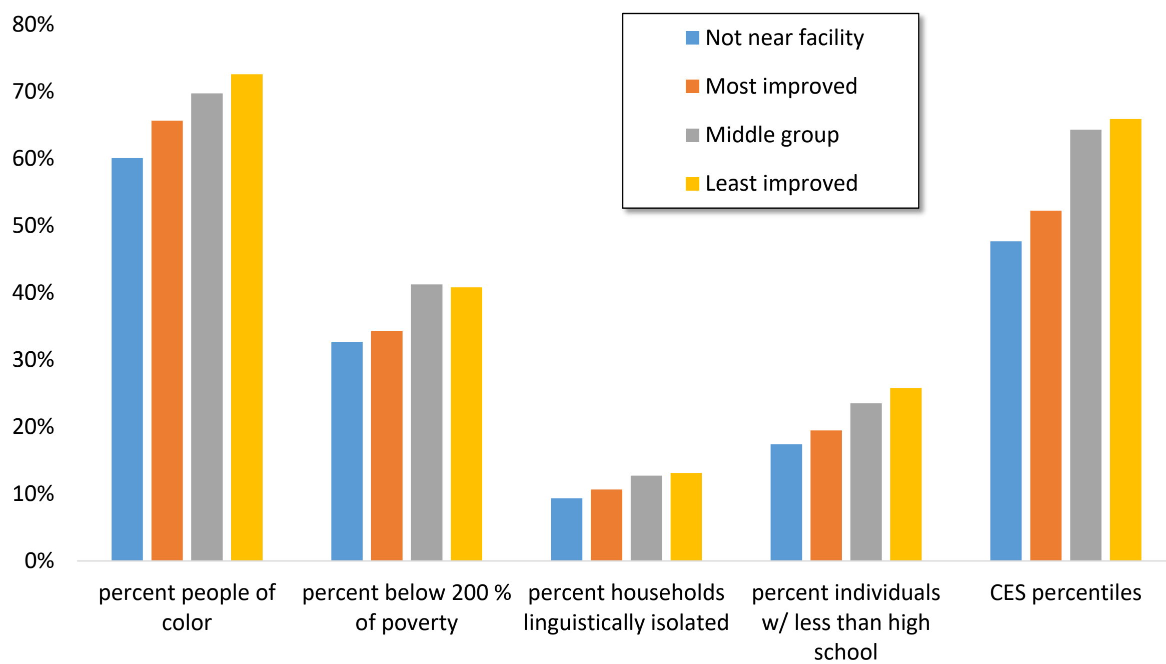
23 In these initial figures, we are adding up at a neighborhood level all facilities regardless of whether they had sufficient observations in all years to make it into the dataset used to analyze within-unit change over time. Restricting our attention to that tighter set yields a very similar pattern and the relative flatness of the income gradient compared to the steepness of the race gradient is more pronounced.

PM2.5, PM10, NOX, and SOX; as with the facility analysis, we examined and then excluded the outliers with extreme increases or decreases.[24]

We once again broke the sample into terciles, except this time by the third of neighborhoods based on pollutant emissions from nearby facilities between 2011-2012 and 2016-2017. To get a sense of the bands of effect, for covered GHGs, the most improved neighborhoods saw reductions of at least 15.1 percent in the tons of GHG emitted in total by all facilities located within 2.5 miles of the block group centroid; the "worst" block groups actually saw GHGs go up by more than 2.2 percent. The breaks for the other pollutants were not dissimilar – and for all (PM2.5, PM10, NOX, and SOX), "least improved" neighborhoods actually saw higher levels of pollution between the two time periods and so could be termed "worsened" or labeled as "worse performers" relative to "best performers." Once again, we use the language of relative improvement here because the neighborhoods on average saw decline in the cap-and-trade related pollutants, as evidenced by a median rate of change for each pollutant that was negative across the whole sample.

Figure 11 shows the demographic pattern by the percentage change in Covered GHG; in this case, we can also include the population not near any facilities as a baseline comparison. As can be seen from that, there is a general environmental justice issue with regard to who lives near a facility or not. Focusing on the demographics by change in pollution, we find that improvements were concentrated in in whiter and wealthier areas than in areas with larger shares of people of color, individuals living below 200 percent of the poverty line, linguistically isolated households, less educated individuals, and a higher CES percentile.

Figure 11. Demographic pattern by % change in covered GHG (2011-2012 versus 2016-2017) among block groups with and without facilities



24 We did this after noticing that these outliers generally tracked with the facilities we had eliminated when trimming the sample in the facility analysis. We followed the same strategy, reducing the block groups under consideration by five percent overall, with half from the extreme high in terms of percentage change and half from the extreme lows in percentage change. We also tried tracking the facilities eliminated in that earlier analysis and eliminating their impact on the pollution landscape. The results were similar but that approach tended to leave in far more extreme outliers in the case of SOX.

Figure 12 offers a table showing the pattern for PM2.5, PM10, NOX and SOX. The patterns for several of the co-pollutants are less monotonic than we see for covered GHGs but the neighborhoods in the tercile with the least percentage change improvement nearly always have a higher share of people of color, poorer families, and less educated individuals than those in the tercile considered most improved; they also always have higher average CES percentiles.[25] We also calculated what the pattern would look like if we excluded the pollution from the oil and gas entities that report from multiple emitters; the pattern for just those facilities points more strongly and monotonically in the direction of environmental inequity.[26]

Figure 12. Block groups ranked by % change in emissions for four pollutants, between 2011-2012 and 2016-2017 (population-weighted)

		percent people of color in community	percent below 200% poverty in community	percent linguistically isolated in community	percent w/ less than HS in community	average CES percentile score of communities
PM2.5	Not Near a Facility	60%	32%	9%	17%	47%
	Most improved	71%	37%	12%	22%	62%
	Middle group	73%	45%	15%	27%	67%
	Least improved	71%	39%	11%	24%	64%
PM10	Not Near a Facility	60%	32%	9%	17%	47%
	Most improved	70%	36%	11%	22%	62%
	Middle group	73%	44%	15%	27%	65%
	Least improved	71%	40%	11%	24%	65%
NOX	Not Near a Facility	60%	33%	9%	17%	47%
	Most improved	68%	35%	11%	20%	58%
	Middle group	69%	39%	11%	23%	63%
	Least improved	73%	43%	14%	27%	64%
SOX	Not Near a Facility	60%	33%	9%	17%	47%
	Most improved	68%	37%	13%	22%	63%
	Middle group	68%	39%	11%	23%	61%
	Least improved	76%	42%	14%	27%	66%

We now repeat our non-parametric analysis of pollutant changes, this time with a focus on what occurred in DAC and non-DAC neighborhoods, with a block group deemed a DAC if the CES percentile associated with its census tract was equal to or exceeded the DAC threshold (75%). We present those results in Figure 13, looking only at neighborhoods that have some pollution from cap-and-trade facilities. In general, median improvements in pollution levels occurred in the DAC neighborhoods that fall into this broad group but they were not as strong as in non-DAC neighborhoods impacted by cap-and-trade.

25 The patterns shown are weighted by block group population; the pattern is similar if we do not population weight.

26 This was an appropriate robustness test because we are making assumptions about pollutant distribution to those multiple emitters and also because including them brings in a large number of neighborhoods often with quite small pollution loads. When we exclude the oil and gas facilities with multiple emitters, we are still capturing 89 percent of facility-level covered GHGs and PM10, 86 percent of PM2.5, and 95 and 96 percent of NOX and SOX respectively relative to the samples considered in the facility analysis above.

The Mann-Whitney test indicates that the difference in both percent change and the absolute change in pollution levels were statistically significant for Covered GHG and PM10, with worse performance in the DACs. For PM2.5, the percent change was statistically significant with sharper reductions in the non-DAC neighborhoods; using the Mann-Whitney ranked sums procedure, the absolute difference was larger (less improvement in pollution) in the DAC neighborhood and also statistically significant at a lesser level although the median test was not significant. For SOX, the difference between the absolute difference in DAC neighborhoods versus non-DAC neighborhoods was statistically significant but the percentage change was not; for both the change and the difference, the mean rank and median tests pointed in the direction of the DAC neighborhoods experiencing less improvement, something that can also be seen in the median pollution changes presented in the table.[27]

Figure 13. Changes in Emissions from Facilities in DAC and non-DAC Block Groups

	Non-DAC Neighborhoods (Median, IQR)	*DAC Neighborhoods (Median, IQR)*	*Median Test*	*Highest Rank*	*Mann-Whitney*
% Change in Covered GHG	-8.8% (-17.5%, 5.5%)	-0.8% (-13.5%, 7.5%)	***	DAC	***
Diff in Covered GHG (MTCO2e)	-5,662 (-26,131, 4,537)	-450 (-7,310, 11,719)	***	DAC	***
% Change in PM2.5	-19.6% (-46.5%, 6.9%)	-4.6% (-29.0%, 8.3%)	***	DAC	***
Difference in PM2.5 (tons)	-0.80 (-4.00, 0.40)	-0.65 (-2.95, 0.85)		DAC	*
% Change in PM10	-16.5% (-39.0%, 6.9%)	-1.2% (-26.7%, 6.9%)	***	DAC	***
Difference in PM10 (tons)	-0.65 (-4.05, 0.45)	-0.18 (-4.05, 1.00)	***	DAC	***
% Change in NOX	-3.1% (-23.9%, 31.2%)	-2.2% (-15.7%, 38.2%)		DAC	***
Difference in NOX (tons)	-0.45 (-5.40, 2.90)	-0.05 (-6.65, 6.05)	***	DAC	***
% Change in SOX	-5.6% (-25.0%, 19.5%)	-1.0% (-25.0%, 9.9%)	***	DAC	
Difference in SOX (tons)	-0.05 (-0.45, 0.10)	-0.01 (-1.26, 0.20)	#	DAC	#

*** significant at the .01 level
** significant at the .05 level
* significant at the .10 level
# significant at the .20 level

While the pattern is sometimes uneven, both the facility-level and neighborhood-level analysis above generally suggest that equity concerns about changes in GHG and co-pollutant emissions between 2011-2012 and 2016-2017, from facilities regulated under cap-and-trade may be warranted. We discuss in the conclusion some research and policy efforts that might address these concerns. First, however, we explore and explain some recent research that suggested another pattern of results.

27 The pattern for SOX seems to be sensitive to extreme outliers that remain in the data even after trimming. For this analysis too, we reproduced the results looking at just emissions from facilities that were not the reporters with multiple emitters; the basic pattern of direction of effect and statistical significance is similar and, in fact, stronger for covered GHGs, PM2.5, PM10, and NOX; the results are muddled for SOX with median tests pointing to less reduction in DAC neighborhoods but the ranked sum test pointing in the other direction.

Understanding Another Approach

As indicated in the introduction, both the popular press and California environmental officials were intrigued by a 2020 analysis by Danae Hernandez-Cortes and Kyle Meng that indicated that California's carbon market has not widened but rather closed disparities in terms of environmental exposure. It should be stressed that HCM are essentially answering a different question than the one we explored above; in their analysis, HCM seek to identify the average effect of cap-and-trade on equity by comparing the average change in performance of cap-and-trade facilities to that of generally smaller facilities not regulated under the cap-and-trade program. This estimate – rather than any actual observed change – is then applied to an air model to generate geographic changes in exposure.

HCM's analytical approach is not necessarily well-suited to assessing the temporal patterns of emissions and equity performance of regulated facilities only—the question generally posed by EJ advocates. Their baseline method requires each facility to share the common percentage increase or decrease depending on what occurs in the overall regulated sector – and in that baseline model, they obtain heterogeneity in the size of the reduction only by applying a common percent change to different starting levels. While they do model differential reductions by the level of initial emissions as a robustness check, the bulk of their analysis relies on the common percentage effect and we discuss in the appendix why size heterogeneity may not capture EJ heterogeneity.

The key analytical point is that if EJ communities are initially the most burdened by high levels of emissions, an imposed common percent reduction (if that is what the data suggest) will inevitably yield a larger predicted improvement in pollutant tons in those communities with the highest emissions. This reflects the starting point and the nature of the common-percent reduction model rather than any actual change in any community's pollution burden.[28]

HCM actually have two basic versions of their paper, one which was published as an NBER Working Paper in 2020 (Hernandez-Cortes and Meng 2020) and attracted significant attention, and a recent NBER update in 2021 that employs much of the same methodology but significantly reduces the number and the range of characteristics of facilities included in statistical modeling (Hernandez-Cortes and Meng 2021). Both the early and most recent versions raise some methodological and data issues which we discuss below.

The HCM Method

HCM use a regression model to compare regulated and non-regulated facilities. Formally, they obtain a distinct trend for cap-and-trade facilities relative to non-regulated facilities after the launch of the cap-and-trade system. In general form, their estimating equation is:

Pollution = f(facility, year, CnT x time, CnT x time_post2012)

Which could be written in perhaps more accessible terms as:

28 Again, they do introduce size heterogeneity later in their analysis and we discuss this below.

Pollution = f[baseline facility fixed effect, year-specific effect for each year shared by every facility in that year, a pre-policy trend (shared by all Cap-and-Trade facilities, and a post-policy change in trend (shared by all Cap-and-Trade facilities]

The baseline for each facility, captured with a so-called fixed effect, controls for the size, vintage, and underlying technology of any emitter.[29] Overall trends in emissions that potentially affect all facilities are controlled for with year fixed effects; these year fixed effects essentially strip away common patterns in increases or decreases in pollution because of, for example, an expansion or contraction in the economy.[30] The penultimate term captures the trend while the last term, the key explanatory variable for HCM, expresses the altered trajectory, in percent change in pollution per year, shared by all cap-and-trade facilities after the policy went into effect in 2013.

HCM designates the last term, the post-policy change in trend shared by all cap-and-trade facilities, as the effect of cap-and-trade; all this is relative to a control group of facilities that are also in the regression but are not under cap-and-trade and so have their pattern fully captured by just the facility and year effects. After running the regression, HCM obtain a set of predicted pollution levels for each facility, strip them of their common year effect, and this estimated level of pollution simulated as the outcome of the policy provides the inputs for the next stage of their analysis.

In that stage, HCM apply an air dispersion model to gauge where the predicted pollution from their regression drifted before regulation and where predicted pollution drifted after regulation. Seeking to avoid an arbitrary definition of environmental justice, the authors instead characterize neighborhoods based on the disadvantaged community designation (DAC) that is available under state environmental legislation, an approach we used as well. HCM examines the "EJ gap" in total exposure between DACs and other communities before and after.

They conclude that the total gap has closed and so, contrary to the EJ critique of California's carbon markets, a rising environmental tide may have lifted all boats. They rightly stress that their analysis does not exhaust the range of distributional concerns or environmental justice issues and highlight the need for specific EJ policies to address issues of disparity, a conclusion we applaud and an assessment we share.

The introduction of air modeling is a significant contribution to the literature since proximity-based approaches like ours do not account for wind patterns and other conditions affecting the fate and transport of emissions. HCM take an exhaustive approach to the air modeling, running the model over the entire time period for which they have data, 2008-2017, at four-hour intervals and tracking pollution for the subsequent 24 hours, with care taken to alter the parameters and distance decay functions for various co-pollutants. Such air modeling is certainly a methodological improvement and it is one we seek to deploy in the future. However, the data inputs into the model are important – which are the result of the regression exercise described above – and here there are a few issues.

29 Such a fixed effect does not indicate whether the issue is size, vintage, or technology but simply controls for that.

30 In the regression we implemented, we chose as the base or default year 2012 so that the fitted lines would tend to most closely match the actual data at that turning point rather than at the starting year, 2008. This does not affect the key slope variables but it makes the graphic presentation of the difference between the actual data and the fitted lines a bit clearer for the reader.

The HCM Data

In constructing their dataset, HCM started with a pollution download from 2008-2016, and then appended 2017 data. They noted that 39 facilities "switched" regulatory status in 2017 (i.e., flipped from being coded as "No" with respect to cap-and-trade regulation to being coded as "Yes" and vice versa). In an initial version of their research, they dropped these 39 facilities from the main sample, later noting that that exclusion did not alter the results (Hernandez-Cortes and Meng 2020). In their most recent work, they include the 39 facilities in the broader universe from which they draw a subsample, using a facility's status during 2015-2016 as its time-consistent and definitive designation (Hernandez-Cortes and Meng 2021:10).[31]

Splicing two series in such a fashion can create problems, particularly when the agency itself provides a time-consistent and updated dataset. For example, the dataset HCM use has 128 facilities that never entered cap-and-trade; these are tagged as a "No" in the HCM dataset when they are tagged in current data, including the Pollution Mapping Tool, with a blank indicating that they were never part of the system and really should not enter any comparative analysis.[32] Of those entities that remain, HCM has 30 facilities tagged as "Yes" that according to our download (and the extensive checking with the Pollution Mapping Tool and OEHHA detailed above) should be tagged "No" and 21 facilities (apart from the 128 that should be tagged with a blank) that are tagged in their data as "No" that should be tagged as a "Yes."[33]

As we note below, HCM actually prefer to run their regression on a reduced sample in which they exclude facilities in the refinery or electrical generation sector and also exclude entities reporting GHG emissions below a certain threshold; that yields a total of 333 facilities of which 41 actually dropped out before cap-and-trade and 26 are characterized differently than in our cleaned and cross-checked dataset.[34] As we also note below, the HCM regression also does not require facilities to be in the sample both before and after the imposition of cap-and-trade; if we drop all the blank tags, and require what is necessary for a within-unit temporal analysis, the sample reduces to 135 reporting entities, with 13 percent of the cap-and-trade tags differing from their designation according to OEHHA.

We demonstrate the impacts of the classification and regressions specifications below. Before we do, we should note that here are other issues in the HCM data, including the fact that the HCM air model utilizes the reporting location for many of the oil and gas reporting entities rather than the actual emitting locations, which can lead to inaccuracy when determining which communities are affected by pollution.[35] Another issue which appears to have an impact is that reporting requirements for both

31 As noted above, our basic standard is whether the entity was tagged as having covered emissions in 2016-2017 and whether if was considered to be cap-and-trade by OEHHA.

32 This was likely a problem in the early version of the data available from the Pollution Mapping Tool that HCM use.

33 The HCM data also has one more facility than our data; this is because one facility changed its ARB identification after its first year, something that has been cleaned up in our version of the dataset. We should note that of the 51 entities with a reporting issue, 34 of them are part of the 39 that HCM switched back to their previous reporting status.

34 HCM obtain a total of 332 with those cuts; we are not sure why there is a difference but this seems to be a minor issue.

35 HCM also define "disadvantaged communities" or DACs using version 1.1 of a state tool called CalEnviroScreen. That version was never actually used to designate DACs, partly because it relied on zip codes as the geographic unit, an approach that was widely rejected by researchers, community organizers, and eventually policy makers and state decision-makers. Newer versions of CalEnviroScreen use a different method and indicators to calculate DACs based on census tracts. Relying on zip codes from CES 1.1 in the way they do misclassifies about 30 percent of the census tracts. This is a relatively minor issue as they demonstrate that their results are robust to

GHGs and particulate matter shifted between the 2008-2010 period and the 2011-2017 period, which suggests that regression analysis over the whole period should be done with caution; this is also one reason why we started our own analysis in 2011 and we return to this when discussing estimates of the pre-policy trend in the data appendix.

The HCM Estimates

HCM have had two basic iterations of their paper, one in which their estimating equation was performed over the entire sample and a second in which the estimating equation was applied to a reduced sample in which the authors excluded refineries and electrical generators, including co-generation facilities, and also restricted the size of the facilities considered.[36] The logic for the sector exclusion was that refineries and electric generators were also governed by other regulations, such as California's Renewable Portfolio Standard and Low Carbon Fuel Standard; the logic for the size exclusion was that one should not compare smaller and larger cap-and-trade facilities with much smaller unregulated facilities.[37]

While the sectoral and size constraints may generate more similar comparison facilities and help to isolate a "pure" cap-and-trade effect, the sample restrictions eliminate 87 to 95 percent of the total pollution, depending on pollutant, emitted by facilities under cap-and-trade regulation. It is, however, this second approach that is preferred by HCM and that is the restricted sample that we work with below.

We conduct three basic regressions: a first estimate that utilizes the cap-and-trade tags that HCM use; a second estimate that utilizes the cap-and-trade tags we use (and have been verified by OEHHA and the Pollution Mapping Tool); and a third estimate in which we utilize the verified end-of-period cap-and-trade tags and also require that the analysis be within-unit, meaning that to remain in the sample, entities must have observations before and after the cap-and-trade policy is imposed.

What do we find? With our first regression (in which we stick with the HCM tags for cap-and-trade), we can perfectly match the results in Table 1 of Hernandez-Cortes and Meng (2021) in terms of coefficient estimates, significance levels, size of sample, other key measures.[38] To see what the estimated cap-and-trade trends looks like, we show the predicted values for Total GHGs for four randomly selected

---

the use of CES 3.0. There is also an issue of repeated observations in the co-pollutants that we discuss in the data appendix.

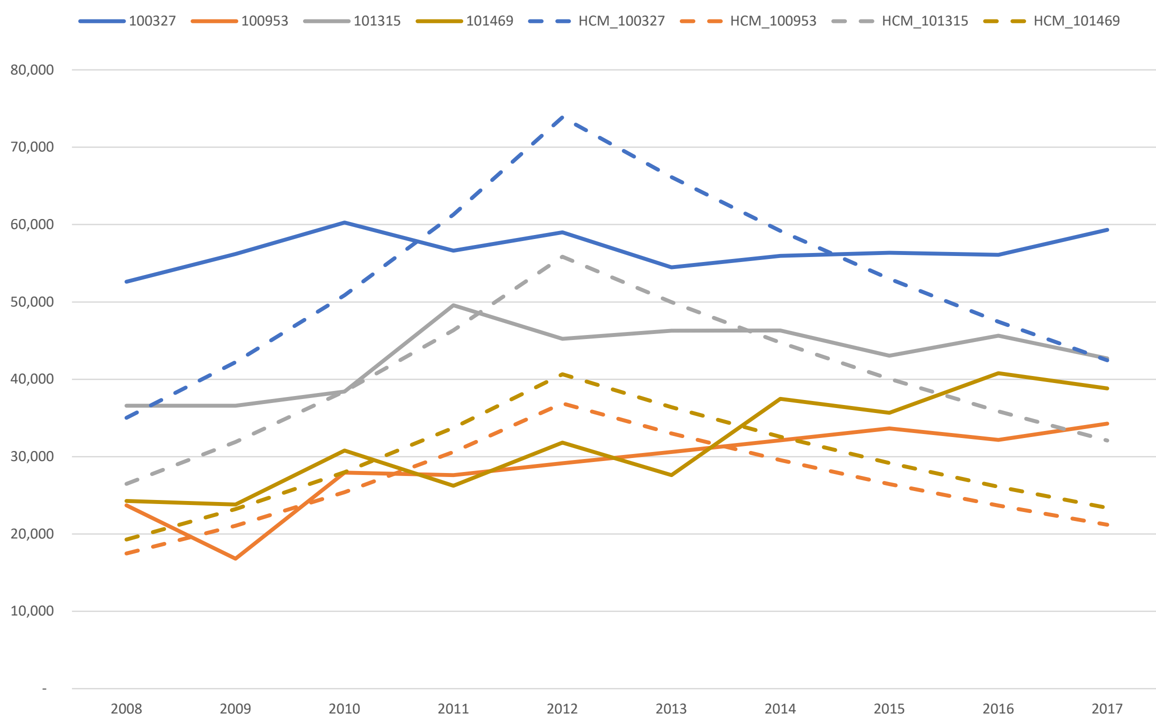
36 In the first version, HCM used the natural log of the pollutant as the dependent variable, something which conveniently eliminated the zero observations, something we think might be a better approach given the nature of this data. However, as we discuss in the appendix, we conduct the regression with the dependent variable being the inverse hyperbolic sine of the dependent, a strategy that allows zeros to persist in the data.

37 HCM indicate that they set the size threshold at 75 percent of average annual total GHGs in their sector-constrained sample. This is actually a bit misleading for two reasons. First, they include in their sample facilities that never entered cap-and-trade. Second, rather than choosing the entity at the 75th percentile, they choose the observation at the 75th percentile; since many entities do not have data for every year of the sample, these will differ.

38 We utilize the actual HCM database so we can more perfectly match their results; as noted above, our own data was also cleaned more thoroughly of extreme observations where a facility was, for example, shut down or where the pattern of reduction (such as with the Excide plant) or increase had little to do with cap-and-trade. We thank HCM for sharing their data and, of course, we returned the favor. The sample size in the GHG regression is less than the count of facilities because "singletons" drop out of the regression; we match the HCM sample size.

representative cap-and-trade facilities in Figure 14.[39,40]

Figure 14. Total GHG Patterns and Baseline HCM Estimates



As can be seen, the estimates suggest a sharp increase in GHGs (relative to the non-cap-and-trade sector) prior to 2012 and then a sharp decrease in GHGs after 2013; this sharp decline is the primary reason that HCM can obtain an improvement in the so-called EJ gap (imposing a common-percentage reduction in pollution to an already unequal "riskscape" necessarily shrinks absolute gaps between communities). Of course, the estimated pattern for these four facilities does not bear a strong resemblance to the pattern of the actual data for the four cap-and-trade entities we selected randomly. While this could be an anomaly, one could also note that the predicted relative decline for GHGs from this first estimate is on the order of 42 percent between 2012 and 2017 (and 36 percent between 2013 and 2017), far outpacing the reported 5.3 percent decline in state GHGs between 2013 and 2017.[41]

To examine the impact of different regression specifications, we took the average of the facility estimates for all three approaches specified above and compared them to the actual patterns in Figure 15. As one can see, the first averaged estimate mirrors what we showed for the four estimates in the

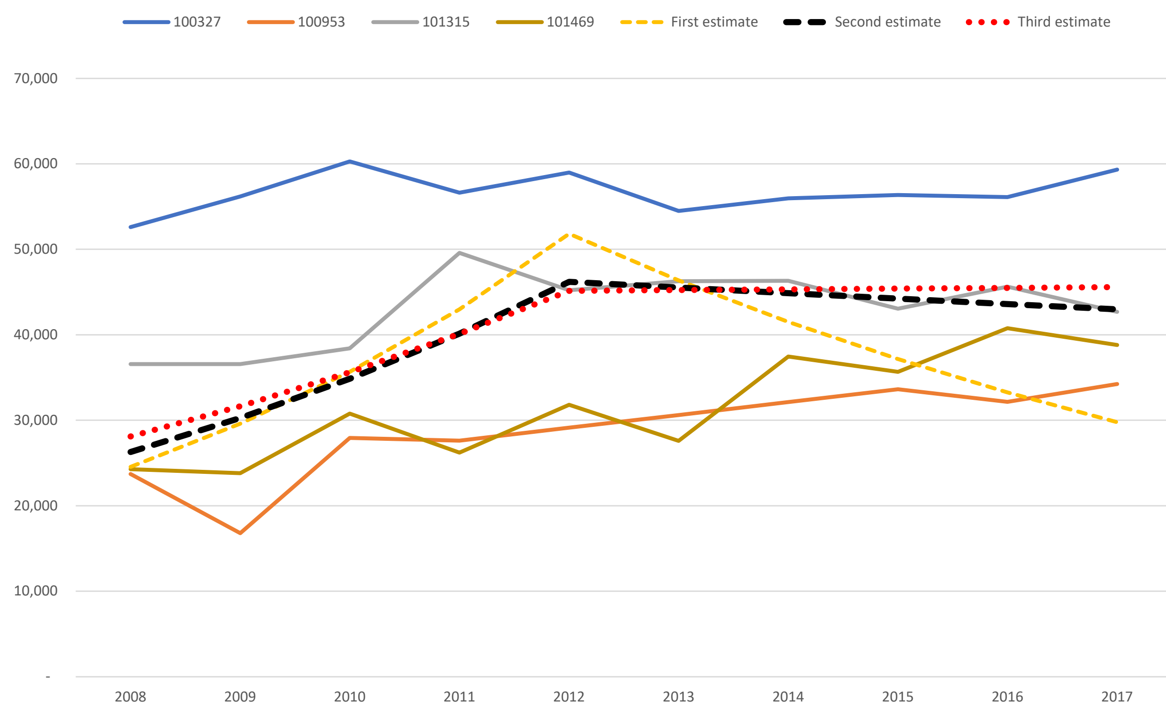
39 We should also note that here we are analyzing total GHGs as HCM do rather than covered GHGs which we use in our analysis of the cap-and-trade entities; as discussed in an earlier note, these are frequently close for the covered sector but can diverge significantly for the uncovered sector, particularly when GHGs include biomass.

40 We wanted to show just a few facilities to make the needed points but avoid clutter. To avoid bias in the choice, we chose the four facilities by first selecting nine facilities that were near the median for PM10. We then ran a random number generator on the nine facilities and chose the four with the lowest values, and then looked at the predicted series for TotalGHG.

41 https://www.c2es.org/content/california-cap-and-trade/

previous figure: a sharp increase followed by a sharp decrease. In our second estimate, we use the cap-and-trade tags that are in the Pollution Mapping Tool and are accurate according to OEHHA; aside from correcting the "Yes" and "No" tags, we are also dropping the facilities that are blanks and dropped out of the sample but are tagged in HCM as "No." This reduces the number of entities in the regression from 316 to 282. As can be seen, there is a less steep increase prior to cap-and-trade followed by a slight decline over time; interestingly, the estimate decline between 2013 and 2017 is 5.6 percent, quite close to what actually happened for state GHG emissions.

Figure 15. Comparing GHG Patterns and Alternative HCM Estimates



In our third estimate, we introduce an additional constraint that we think may be appropriate: requiring that the trend break be identified within-unit. In early communication with HCM about the first draft of their paper, the authors indicated that their regressions were run with the requirement that each entity include at least two years before the introduction of cap-and-trade and at least two years after, on the grounds that that was necessary to accurately estimate the pre- and post-trends. But that is not, in fact, what they seem to have done: their regression sample includes facilities that stopped reporting before the cap-and-trade program began, facilities that started reporting after cap-and-trade, and a large number of facilities that started reporting in 2012 and so really have no pre-cap-and-trade trend.

While this might seem appropriate for comparing, say, one sample of students' performance to another sample of students' performance after a shift in teaching strategies, such sampling strategies occur when you for some reason cannot take a look at the full universe of cases. In this case, we have the full universe of facilities in the cap-and-trade system and the analogy would be more like wanting to look at the longitudinal change in the performance of students after a change in teaching strategy in a particular school district but including students who dropped out before teaching practices changed and others who arrived later; it's hard to know if they are the same sort of students and one would at least prefer to look at a within-unit analysis. Once we require that entities have at least one observation before cap-and-trade, one after, and one at the pre-change peak in 2012, we wind up with 135 entities, 90 of which were under cap-and-trade and 45 of which were not.[42]

As can be seen in Figure 15, that third estimate shows a relative rise in GHGs followed by essentially a flat performance after the imposition of cap-and-trade. While we have illustrated this pattern in the charts only for total GHGs, Figure 16 shows the pattern for Total GHGs and the five co-pollutants for all three regression specifications. The pre-policy trend refers to the estimated upward slope in Figure 15, the post-policy break refers to the change in that slope, and the post-policy trend captures the slope of the line after the imposition of the policy. What can be seen is that using the corrected tags for cap-and-trade does most of the work of flattening the estimated post-policy trend; when we also require that the observations actually be within-unit or longitudinal, we see virtually no estimated change in this sector (with the exception of NOX).[43]

Figure 16. Coefficient Estimates from Three Different Regression Specifications

		Total GHGs	PM2.5	PM10	NOX	SOX
First Estimate (HCM, Table 1)	*pre-policy trend*	18.7%	5.8%	8.3%	7.5%	0.6%
	post-policy break	-29.7%	-9.7%	-11.7%	-10.4%	-3.7%
	post-policy trend	-11.1%	-3.9%	-3.4%	-2.9%	-3.1%
Second Estimate (verified, cap-and-trade tags)	*pre-policy trend*	14.1%	5.7%	8.2%	6.4%	-2.8%
	post-policy break	-15.5%	-6.5%	-8.0%	-5.4%	1.0%
	post-policy trend	-1.4%	-0.7%	0.3%	1.0%	-1.8%
Third Estimate (verified & within-unit)	*pre-policy trend*	11.8%	4.8%	7.6%	2.5%	-3.5%
	post-policy break	-11.6%	-4.4%	-6.6%	3.7%	2.4%
	post-policy trend	0.2%	0.4%	1.0%	6.2%	-1.1%

While it may seem odd that there was little substantive change in this sector- and size-constrained sample, note that these are smaller facilities and may have been more likely to forgo expensive investments to reduce local GHG emissions and instead purchase allowances; by focusing on a particular (and very small) part of the whole cap-and-trade system, this may be less a "pure" cap-and-trade effect than might be initially surmised.[44]

42 58 of the entities eliminated started in the HCM data in 2013 or later and 1 had a single observation in the year 2012 so we really have no pre-policy trend for them; 98 had their start year in 2012. We believe that having about 20 percent of the sample entities start reporting after the imposition of cap-and-trade and roughly 40 percent of the sample start reporting in the last year of the pre-compliance period makes estimating a reliable pre-trend challenging.

43 The exception is the NOX series but by the time we get to the third estimate, we have very much limited the sample and unusual results are possible, particularly given the nature of the co-pollutant data.

44 Indirect evidence that this might be the case comes when we run a similar regression on the whole sample on our own data, using the verified cap-and-trade tags, requiring a within-unit analysis, and implementing a method to drop both cap-and-trade and non-cap-and-trade observations that are likely shut-down years. We obtain a post-policy trend estimate for total GHGs of -.01, which would be somewhat consistent with the actual reduction in GHGs in the state. Estimates for the other pollutants in such a specification generally hover around zero.

We discuss several other issues in the appendix, including the reliability of the estimate of an upward sloping pre-policy trend. The main point of this analysis is that to obtain a relative EJ improvement in the baseline HCM model, one likely needs to see the estimates of pollution actually trend downward; while this is the case for the estimate HCM present in their paper, using the cap-and-trade tags in the Pollution Mapping Tool and verified by OEHHA and requiring that the temporal trends include the same facilities over time seems to largely change that result. We discuss in the data appendix how introducing variation by size impacts this analysis to some degree.

It should also be stressed that even if a negative post-policy trend is found, this does not necessarily get at the question usually at stake in these debates: whether or not the actual changes are equitable in terms of their impacts on different communities. To see this, consider what a DAC neighborhood analysis of the type above would be likely to look like if we assume a common negative percentage effect from cap-and-trade: the median percentage change between 2011-2012 and 2016-2017 would be the same (or nearly the same) in DAC and non-DAC areas, while the absolute change would be larger in the DAC neighborhoods because they generally start off with higher levels of initial pollution. One can compare that with the actual pattern illustrated with a simple buffer approach in Figure 13.

Conclusion

This paper has compared patterns of pollutant emissions in 2011-2012 and 2016-2017 from facilities regulated under California's cap-and-trade system. We find that communities with higher shares of people of color, people living below 200 percent of the poverty level, households that are linguistically isolated, individuals with less education, and higher CES percentiles are more likely to live near cap-and-trade facilities and, while the patterns are complex and not entirely consistent, are also less likely to have seen improvement in pollution emissions. When we make use of the CES percentiles to specifically consider "Disadvantaged Communities" (or DACs) – by breaking either facilities into DAC-adjacent and non-adjacent, or breaking the communities themselves by their DAC status – we also find a general pattern that leans in the direction of less improvements for communities often considered to be of prime EJ concern.

These results are consistent with our earlier analysis (Cushing et al. 2018). That published study has been contrasted with recent work by Hernadez-Cortes and Meng (2020, 2021). However, it is important to note that HCM are asking a very different question: whether or not pollution went down among a select set of facilities regulated under cap-and-trade versus those facilities that were not regulated under the program. In essence, they are describing the average behavior of the regulated sector compared to the unregulated sector. While they report a relative reduction in GHGs and co-pollutants from regulation, those results are quite sensitive to the tagging of cap-and-trade facilities and somewhat sensitive to requiring that that analysis be of the same entities over time.

More importantly, the HCM approach does not necessarily shed light on what might be happening among those facilities that are included in the cap-and-trade system. In their baseline model, HCM are estimating a common percentage effect from cap-and-trade, applying that to all communities, and then finding that communities that were most burdened initially are now (theoretically) most relieved. Yet the point of a market system is to allow companies to pursue differential increases or decreases in pollution. Most important for this report, because applying a common negative percentage effect will likely drive down estimated pollution in the most polluted areas, a closing of environmental justice gaps is more or less a forgone conclusion of the HCM modeling strategy even when the reported data on individual emitters does not fit that pattern.

Future research in this area will benefit from better data and methods. For example, HCM have provided a real step forward by integrating air modeling and future efforts should follow suit to the extent possible. On the data side, the harmonization of GHG and co-pollutant emissions inventories in the Pollution Mapping Tool has been a welcome development. Nonetheless, CEIDARS data remains a problem in terms of consistency and frequently repeated observations due to reporting lags; more frequent and regular reporting to CEIDARS could be useful for future work. It would also be helpful if researchers could gain access to digital allowance and offset data to gain a finer sense of who is buying allowances, and what types of projects are being supported through offsets, either within or outside of California. Moreover, use of air modeling would be a significant improvement over the buffer approach used here and will be part of our future work on this topic.

Standing back from this work, we should acknowledge that environmental justice issues arising from cap-and-trade per se may be a limited part of the overall challenge of exposure disparities (Anderson et al. 2018). For example, although we examined offsets, in terms of project types and their locations in our prior study (Cushing et al 2018), offset data are not amendable for the level of spatial granularity required to assess equity concerns within California. For this updated analysis, we did not examine the equity implications of offsets, a policy that has been criticized for potentially overselling pollution reduction and forgoing key local benefits with investments in projects that address GHG in distant locations (Halper 2021).

Moreover, equity concerns from cap-and-trade are not limited to the co-pollutant issue, and there are many other issues raised by EJ advocates and many positive actions have been taken by the state to respond. For example, the allocation of a substantial share of California's Climate Investments funds raised by cap-and-trade (formerly known as the Greenhouse Gas Reduction Fund) to DACs is a welcome move in the direction of fairness. The Transformative Climate Communities program is a good example of how those resources are being targeted to address the most distressed communities. Both of these programs have provided a model for federal climate justice initiatives including Justice40.

There are also bigger equity issues ahead as we address climate change, including how adopting the principle of "just transition" can impact who will obtain jobs and which communities will see a decline in pollution (Cha et al. 2019). And future research can help by temporally tracking the performance and equity implications of other GHG reduction programs, including those that focus on mobile GHG sources, such as California's Clean Vehicle Rebate Programs (Ju, Cushing, and Morello-Frosch 2020).

With regard to the issues raised in this paper, we remain convinced that there are a series of reasonable fixes to alleviate environmental justice concerns from carbon markets. Declaring "no-trade" zones in places that are highly polluted could accelerate actual local emissions improvements in those neighborhoods. Creating price incentives to deepen GHG and co-pollutant reductions in DAC communities is also an option. Paying more regulatory attention to those facilities or sectors with high correlations of GHGs and co-pollutants could also address equity concerns.

Modeling of these scenarios could be useful, particularly since a number of other states and jurisdictions are considering cap-and-trade programs or carbon fee programs and are looking to California as they try to design their systems. Comparing "pure" cap-and-trade with market modifications and other hypothetical emissions reductions approaches would allow for a better understanding of what regulatory tools might optimize climate protection, air quality improvements, and environmental equity. We hope that this paper contributes to a discussion of such alternatives that intentionally integrate sustainability and environmental justice goals.

Data Appendix

This appendix discusses data issues raised in both our own analysis and our attempt to replicate the HCM analysis.

Mix and Match

The Pollution Mapping Tool combines MRR records on GHG with CEIDARS inventory data on the co-pollutants. The MRR mandates annual verified reports. CEIDARS, in contrast, involves self-reporting by companies and only requires updating every few years, depending on the air district authority, leading the inventory to repopulate with the previous year's data when there has been no update.

Specifically, the CARB CEIDARS emissions inventory is a database of estimated toxic pollutants discharged into the atmosphere, with reporting requirements specified by various State and Federal mandates, including the California Health and Safety Code, the California Clean Air Act of 1988, the Federal Clean Air Act Amendments of 1990, and the 1987 Air Toxics "Hot Spots" Act (AB2588). Reported emissions for each source include basic data on the facilities, stacks, devices, and processes that emit criteria pollutants into the air, and specify chemical, geographic area, time span (e.g., in a calendar year) and source category (e.g. point and area). The ID for each reported source, the county ID, facility ID, air basin code and district code are combined to uniquely identify each reported emission.

There are important differences in reporting requirements for GHG, Criteria, and Toxic pollutant emissions. Regulated facilities report GHG emissions using CARB-designated quantification methods. Criteria and Toxics emissions data are reported by local air districts - in some cases, these estimates use methodologies prescribed by the district and may vary from district to district; they may differ from the emissions estimated by the facility. Only facilities meeting specific criteria are required to report. Facilities emitting 10,000 tons or more of CO2-equivalent emissions report annually. Toxics emissions are generally required to report every four years, but some air districts require more frequent updates. Criteria emissions reports are required annually from high emitting facilities (250 tons per year by legislation, but CARB guidelines set a reporting threshold of 10 tons per year); smaller facilities report every third year, and local air districts have the flexibility to set their own reporting thresholds.

For oil and gas facilities, GHG emissions are reported as an aggregate of a company's operations in a geologic basin (large areas that often cover one or more counties). Criteria and Toxics emissions from oil and gas production are reported at the sub-facility level - either a point-source or a collection of smaller sources within a contiguous lease. In many cases, the individual sub-facilities may be exempt from the local air district's toxic pollutant reporting requirements. In order to appropriately compare GHG emissions with Criteria and Toxics emissions that can be, the sub-facility values are aggregated to match the MRR facility definitions, hence the issue with oil and gas facilities noted above.

In any case, consideration of these rules suggest why we see something unusual in the data: a much higher prevalence of non-variation for the co-pollutants overall – and, in particular, a much higher prevalence of repeat observations for the entities that did not enter cap-and-trade. This is likely because they are smaller facilities and are not required to update their data as frequently. While we avoid some of the problems this might introduce in our own analysis because we focus on the cap-and-trade facilities which have more variation and because we are comparing period average at two tail ends of a time period, it is an issue for considering whether the “control” and “treatment” groups in a regression exercise are strictly comparable.

There is another CEIDARS-related issue worth mentioning. While the Pollution Mapping Tool integrates in the CEIDARS data on PM, NOX, and SOX, the data is rounded to the one-digit level when the raw data includes much more precision. We work with what is reported in the Pollution Mapping Tool, partly because it is the public database and partly because it includes PM2.5, a pollutant that is not in the CEIDARS annual data that is readily available to download. This rounding helps to explain some of the percentage change outliers we eventually trim in the analyses above.

In and Out

Our own exercise excludes the data from 2008-2010 because the GHG and PM series are not strictly comparable. For example, the MRR website (https://ww2.arb.ca.gov/mrr-data) notes that “GHG emissions data reported between 2008 and 2010 were subject to slightly different applicability, calculation, methodology, and verification requirements.”

As well, on the ARB website where the data can be downloaded (https://ww3.arb.ca.gov/ei/tools/pollution_map/), there is a caveat indicating that:

> Oversight agencies (U.S. EPA, CARB, and local air districts) may periodically update the emission reporting requirements, which may result in an expansion of the types of emissions that facilities must report. One of the more recent updates was the U.S. EPA requirement for facilities that were not reporting condensable particulate matter (emissions that are released as a gas but condense into a semi-solid particle upon cooling) to start reporting it in 2011. The new reporting of condensable PM caused an increase in PM, PM10, and PM2.5 reported emissions.

Of course, the regression analysis in the HCM paper and in our work above includes 2008-2010 data. According to correspondence with the authors, HCM are not concerned with this potential discontinuity issue because a sudden year jump in pollutant levels would be absorbed by year fixed effects for all facilities. However, this assumes that the jump is the same percentage wise for every facility and for every sector, something that seems unlikely, particularly when one examines the data. As we see below, this shift in reporting requirement may also help to explain part of the relative pre-policy trend.

General and Formal

To keep matters simple in the main text, we presented a general estimation equation for the HCM analysis. The more formal version of the HCM estimating equation is:

$$asinh\, Y_{jt}^{p} = \kappa_1^{p}[C_j \text{ x } t] + \kappa_2^{p}[C_j \text{ x } 1(t \geq 2013) \text{ x } t] + \Phi_j^{p} + \gamma_t^{p} + \mu_{jt}^{p}$$

where μ_{jt}^{p} is the error term, *p* refers to the pollutant, *j* refers to facility, C is the cap-and-trade dummy, and *t* as a subscript is the year. Φ_j^{p} is the facility effect, γ_t^{p} is the year effect, and κ_1^{p} is the pre-trend for the cap-and-trade facilities, and κ_2^{p} is the key estimate of the break in the trend after 2012 and the initiation of cap-and-trade regulations.[45] The *asinh* refers to an inverse hyperbolic sine, a function that performs much like taking a natural log but allows for the retention of observations where the value is zero.

The facility fixed effect allows each facility to have its own starting point or baseline level of pollution; this also helps to remove any particular issue related to the size, vintage, sector, or fixed technology of the facility. The year effect is meant to eliminate any impact on the pollution level that might be economy-wide and year-specific, such as an economic expansion or contraction. Calling the result of the regression estimate, P^*, what HMC then do is to subtract away the year effects, and obtain an estimated level of pollution, P^{**}, which is based exclusively on individual facility effect and the common differential percentage trends between cap-and-trade and non-cap-and-trade facilities applied to each facility's baseline quantity.[46] This policy-induced level of pollution is then entered into the air model for the facility.

We should stress something indicated in a footnote in the main text: when replicating the regression results of HCM, we utilized their data, first with their cap-and-trade tags and then with corrected tags. There are a number of minor issues with the HCM data; our more recent version of the data had some data corrections and additions, and we also thoroughly cleaned the data of shut-downs, including an effort to clean the non-cap-and-trade series. While we think ours is a better dataset, we wanted to be able to exactly reproduce the HCM results to more fairly show the importance of the cap-and-trade tags and within-unit analysis and so utilized their dataset.

Up and Down

As we noted in the main text, if the policy-induced pollutants do not shift much after cap-and-trade, then the preexisting disparities will persist. However, it is possible to argue that at least cap-and-trade moderated the relative increases that were going on prior to cap-and-trade, as evidenced by the negative trend break coefficients for Total GHGs and the two PM measures. This is not what is emphasized in HCM's work; their abstract concludes with the statement: ". . . we find that while EJ gaps across California for criteria air pollutants were widening prior to 2013, they have since fallen as a consequence of the carbon market."

Still, the fact that GHGs and PM in the cap-and-trade sector rose relative to GHGs and PM in the non-cap-and-trade sector in the pre-policy period could be due to early increases in that sector – or it could be due to decreases in those facilities that were not regulated by cap-and-trade. While teasing this apart would require more analysis, one simple way to examine this is to break down emissions in the HCM sample by whether a facility was in cap-and-trade (using the correct tags) over four periods: 2008-2010, 2011-2012, 2013-2014, and 2015-2017.

The last three periods are the official pre-compliance or benchmarking period, and the two official compliance periods (2013-2014 and 2015-2017). The 2008-2010 period is of interest because it is, strictly speaking, not comparable to the rest of the years: both GHG and PM reporting requirements shifted, which is why we started our own analysis in 2011.

45 The error terms are clustered at the county level, a specification we follow.

46 As can be gleaned from the formal regression as well as the discussion, what is being estimated is the natural log of the pollutant in question; one then exponentiates the estimates after subtracting the shared year effects.

To avoid bias due to differential patterns of dropping in and out of the sample in different years, we require that any included entity have observations for all years from 2008 to 2010. The pattern for the medians of emissions of the three pollutants in the HCM sample that shifted reporting requirements are shown in Figure 17; the pattern is similar if we look at means. As can be seen, there is a much sharper drop for the non-cap-and-trade facilities when the requirements shift in 2011 than for the cap-and-trade facilities. While this could be due to actual changes in emissions, a detailed look at the data suggests that one plausible explanation is that the requirements affected the entities differently and the sharp fall in pollutants in the non-cap-and-trade sector is at least partly responsible for the upward slope of relative emissions for the cap-and-trade sector.[47]

Figure 17. Median Emissions by Time Period for Three Pollutants That Shifted Reporting Requirement in 2011 (regression sample only, all years)

		Total GHGs (CO_2e tons)	PM2.5 (tons)	PM10 (tons)
Not Cap-and-Trade	2008-2010	29,215	2.70	2.80
	2011-2012	22,488	2.10	2.10
	2013-2014	21,066	2.10	2.20
	2015-2017	20,339	1.55	1.65
Cap-and-Trade	2008-2010	40,542	3.00	4.30
	2011-2012	39,227	3.30	4.25
	2013-2014	39,457	3.45	4.70
	2015-2017	38,240	3.30	4.30

Big and Small

Because the baseline HCM method assumes a common percentage effect and that forces a larger absolute change in more polluting facilities, the authors also introduce size heterogeneity in which larger facilities (within their sector- and size-constrained sample) are allowed to have a different post-policy trend depending on size. To us, this seems to simply transpose the problem: rather than assuming that all facilities exhibit the same trend, it assumes that all facilities of the same size will exhibit the same trend, regardless of whether that occurs for any particular facility.

Nonetheless, we followed their specification and were able to exactly reproduce the results indicated in Table S3 of their most recent paper (Hernandez-Cortes and Meng 2021). We then followed the process above and first introduced the corrected cap and trade tags, and then restricted the analysis to within-unit change.

The results are somewhat similar to their estimates on the size effects, albeit with a reversal of

47 For example, detailed inspection of the data reveals that at least part of the issue is the inclusion in non-cap-and-trade set of entities with very high shares of their GHGs due to biomass. When reporting shifts in 2011, the trend for non-biomass GHGs looks fairly stable but the total GHGs fall sharply as do the PM measures, suggesting something was occurring with the biomass calculations. When we run the within-unit regression and include a dummy variable for the years 2008-2010 interacted with the cap-and-trade tags, the initial upward slope drops from .118 to .040, with the post-policy trend remaining similar to that in the third within-unit estimate.

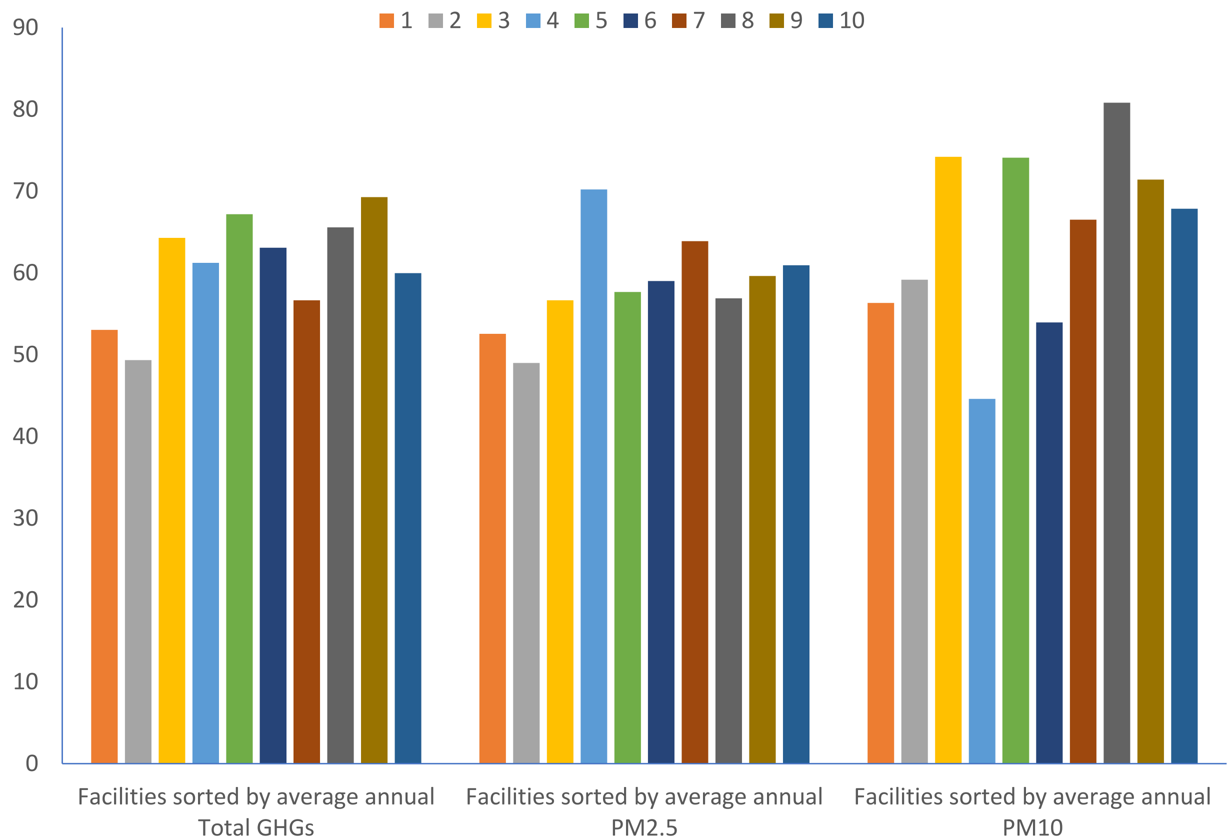
the sign on the GHG results. This may be important because HCM note that "a positive interaction coefficient would imply that larger emitting facilities are abating less, contradicting our assumption" (Hernandez-Cortes and Meng 2021:19). However, the other coefficients on co-pollutants are negative (and significant for PM10, NOX, and SOX), implying somewhat larger decreases in the larger facilities, although this is now from the flatter baseline noted in the figures above.

In any case, the basic issue for an EJ analysis is not the size of the facility but the nature of its surrounding neighborhood: size might be coincident with the EJ character of an entity's impact zone but it might not be. To explore this for the whole sample, we computed the correlations between a facility covered emissions in 2011-2012 and various neighborhood measures, utilizing the natural log of the facility's covered GHGs in order to have a more normal distribution. The coefficients are negative and significant for percent people of color, percent below 200 percent of the poverty line, percent of households that are linguistically isolated, and percent of the adult population with less than a high school degree. The relationship is negative but not significant for the CES percentile we attach to the facility; this is unsurprising since this is based on both demographics and pollution levels and the large facility is likely contributing to the pollution side of that measure.

While this relationship between size and the EJ characteristics of the neighborhood communities may seem unexpected at first, research suggests that the cumulative impact of clusters of polluting facilities, rather than the presence of a single facility, may be the central environmental challenge for neighborhoods. Indeed, the disparities are more sharply revealed in our subsequent neighborhood-based analysis that registers cumulative impacts.

In any case, how does the size-EJ relationship look in the sector- and size-constrained sample of cap-and-trade facilities where we have also corrected for the cap-and-trade tags and required a within-unit analysis? To look at this issue, we selected on that group, ranked into tenths by pollutant emissions the entities to which we could attach CES scores, requiring that the entities have observations for all years so that we could insure that the comparison was not due to entities dropping in and out of the reporting. The result is presented in Figure 18. As can be seen, there is a slight drift up in the mean CES values for the bigger facilities (although this is less clear for median values) and so HCM might be getting indirectly at EJ heterogeneity through the introduction of a size dimension. Of note, however, is that size works in the opposite or unexpected direction for TotalGHGs in the regression (recall the discussion of the sign above) which is the object of the policy. While this is worth further exploration, it is beyond the scope of this paper.

Figure 18. Mean CES Percentiles for Cap-and-trade Facilities in Sector- and Size-constrained Within-unit Regression Sample Ranks by Facility Emissions



References

Anderson, Christa M., Kendall A. Kissel, Christopher B. Field, and Katharine J. Mach. 2018. “Climate Change Mitigation, Air Pollution, and Environmental Justice in California.” *Environmental Science & Technology* 52(18):10829–38. doi: 10.1021/acs.est.8b00908.

Boyce, James K., and Manuel Pastor. 2013. “Clearing the Air: Incorporating Air Quality and Environmental Justice into Climate Policy.” *Climatic Change* 1–14. doi: 10.1007/s10584-013-0832-2.

Cha, J. Mijin, Manuel Pastor, Madeline Wander, James Sadd, and Rachel Morello-Frosch. 2019. *A Roadmap to an Equitable Low-Carbon Future: Four Pillars for a Just Transition.* The Program for Environmental and Regional Equity, at The University of Southern California.

Cullenward, Danny, and Katie Valenzuela. 2020. “A Critique of ‘Do Environmental Markets Cause Environmental Injustice? Evidence from California’s Carbon Market,’ a 2020 NBER Working Paper by Danae Hernández-Cortés and Kyle C. Meng.” *Climate Policy*. Retrieved May 31, 2021 (https://www.ghgpolicy.org/writing/hcm-error).

Cushing, Lara, Dan Blaustein-Rejto, Madeline Wander, Manuel Pastor, James Sadd, Allen Zhu, and Rachel Morello-Frosch. 2018. “Carbon Trading, Co-Pollutants, and Environmental Equity: Evidence from California’s Cap-and-Trade Program (2011–2015).” *PLOS Medicine* 15(7):e1002604. doi: 10.1371/journal.pmed.1002604.

Cushing, Lara, John Faust, Laura Meehan August, Rose Cendak, Walker Wieland, and George Alexeeff. 2015. “Racial/Ethnic Disparities in Cumulative Environmental Health Impacts in California: Evidence From a Statewide Environmental Justice Screening Tool (CalEnviroScreen 1.1).” *American Journal of Public Health* 105(11):2341–48. doi: 10.2105/AJPH.2015.302643.

Cushing, Lara, Rachel Morello-Frosch, Madeline Wander, and Manuel Pastor. 2015. “The Haves, the Have-Nots, and the Health of Everyone: The Relationship between Social Inequality and Environmental Quality.” *Annual Review of Public Health* 36:193–209.

Cushing, Lara, Madeline Wander, Rachel Morello-Frosch, Manuel Pastor Jr, Allen Zhu, and James Sadd. 2016. *A Preliminary Environmental Equity Assessment of California’s Cap-and-Trade Program*.

Halper, Evan. 2021. “Burned Trees and Billions in Cash: How a California Climate Program Lets Companies Keep Polluting.” *Los Angeles Times.* Retrieved September 19, 2021 (https://www.latimes.com/politics/story/2021-09-08/what-is-the-california-climate-credit-does-it-cut-pollution).

Hernandez-Cortes, Danae, and Kyle Meng. 2020. “Do Environmental Markets Cause Environmental Injustice? Evidence from California’s Carbon Market.” *NBER Working Paper 27205*. UC Santa Barbara: National Bureau of Economic Research.

Hernandez-Cortes, Danae, and Kyle Meng. 2021. “Do Environmental Markets Cause Environmental Injustice? Evidence from California’s Carbon Market.” *NBER Working Paper 27205.* Natonal Bureau of Economic Research.

Ju, Yang, Lara J. Cushing, and Rachel Morello-Frosch. 2020. “An Equity Analysis of Clean Vehicle Rebate Programs in California.” *Climatic Change* 162(4):2087–2105. doi: 10.1007/s10584-020-02836-w.

Ringquist, Evan J. 2005. “Assessing Evidence of Environmental Inequities: A Meta-Analysis.” *Journal of Policy Analysis and Management* 24(2):223–47.

Shonkoff, Seth B., Rachel Morello-Frosch, Manuel Pastor, and James Sadd. 2009. “Minding the Climate Gap: Environmental Health and Equity Implications of Climate Change Mitigation Policies in California.” *Environmental Justice* 2(4):173–77.