Cumulative Impacts near California Hazardous Waste Operating Facilities: Data Analysis and Methods

Prepared for DTSC by Nicholas Depsky, Lara Cushing, and Rachel Morello-Frosch

Sustainability and Health Equity Lab, UC Berkeley

Updated: February 3, 2020

Table of Contents

SCOPE OF ANALYSIS ***2***

DATA SOURCES ***4***

Hazardous Waste Facility Locations **4**

CA Statewide Parcel Data **5**

Population Data **5**

Building Footprint Data **7**

Facility Operating Permits **8**

CalEnviroScreen 3.0 .... **9**

Domestic Well Data **9**

Oil and Gas Well Data **10**

Voter Turnout Data **12**

Sensitive Land Use Data **13**

METHODS ***16***

Defining Facility Boundaries **16**

Entire Facility Boundaries 16

Waste-Specific Boundaries 18

Estimating Community Characteristics Near Facilities **19**

Areas of Analysis 19

Populated Areas 20

Unweighted Metrics: Minimum and Maximum CES 3.0 Values, Oil & Gas Wells, Sensitive Land Use (SLU) Counts 22

Area-Weighted Metrics: Domestic Drinking Water Wells 24

Population-Weighted Metrics: Mean CES 3.0 Values, Racial Composition, Voter Turnout 27

RESULTS ***27***

APPENDIX 1 - DTSC Community Vulnerability Metrics Explanations and Justifications for Inclusion ***29***

CalEnviroScreen 3.0 (CES 3.0) **29**

Racial Composition **31**

Healthcare & Senior Care Facilities 31
Parks 31
Prisons 32
Schools and daycare 32
Oil and Gas Wells 33
References for Oil & Gas Wells: .... Error! Bookmark not defined.
APPENDIX 2 .... *39*
CES 3.0 Score vs. Percentile 39
Residential Parcel Classifications 41
Populated Area Layer Construction Example 43
APPENDIX 3 – Population Weighting vs. Area-Weighting *46*
Area Weighting 46
Population Weighting 47

SCOPE OF ANALYSIS

This document summarizes analysis conducted in support of the contract "Integrating a Community Cumulative Impacts Framework in the Implementation of AB 617 and SB 673". It was conducted with data, input and guidance from the California Department of Toxic Substances Control (DTSC) and the California Air Resources Board (CARB). The objective of this analysis was to characterize communities near currently operating hazardous waste facilities (HWFs) regulated by DTSC with respect to their proximity to multiple environmental hazards and vulnerability to the health impacts of pollution. This phase of analysis utilized CalEnviroScreen 3.0 (CES) scores and percentiles as relative metrics of cumulative environmental health impact and community disadvantage. It also includes a number of community metrics surrounding each facility that are not currently included in CES.

The full **list of metrics analyzed** in the areas of analysis surrounding each HWF:

- Mean CalEnviroScreen 3.0 Score & Percentile (See Appendix 1 for distinction between score and percentile)
- Max CalEnviroScreen 3.0 Score & Percentile
- Racial/ethnic composition (% people of color)

- Domestic drinking water well count
- Active oil and gas well count
- Average voter turnout in the 2012 and 2016 general elections (% of registered voters casting votes)
- Sensitive Land Use (SLU) Count – Parks
- SLU Count – Prisons
- SLU Count – Healthcare Facilities
- SLU Count – Senior Care Facilities
- SLU Count – Schools
- SLU Count – Childcare & Daycare Facilities
- SLU Count (All) – Parks, Prisons, Healthcare Facilities, Senior Care Facilities, Public Schools, Childcare and Daycare Facilities

Specifically, this analysis improves upon existing practices for assessing cumulative impacts near hazardous facilities in the following ways:

- ***Polygon boundaries:*** HWFs were defined spatially using polygons instead of a single point.
- ***Entire-facility and waste-specific boundary polygons***: HWF polygons were delineated in two ways: 1) around the entire property boundary, and 2) around the area within property boundaries that is permitted to process or store hazardous waste. The results using both methods are provided in separate spreadsheets.
- ***Population-weighted metrics:*** Community metrics (e.g. mean CES score/percentile, % people of color) were calculated using population-weighting rather than area-weighting to better reflect cumulative impacts experienced by populations near HWFs.
- ***High-resolution population distribution data:*** Populated areas were defined by combining information on the location of residential parcels (data provided by CARB) with block-level population estimates derived from the 2010 decennial US census and block-group-level estimates from the 2013 - 2017 American Community Survey, and building footprint data from a remotely-sensed national dataset produced by Microsoft in 2018. This approach better estimates conditions where people live by omitting places that are unlikely

to be inhabited, improving upon standard methods that assume a spatially uniform distribution of the population across the block group's entire area.

DATA SOURCES

Hazardous Waste Facility Locations

The names and locations of hazardous waste facilities (HWFs) currently permitted to operate in California were supplied by DTSC in the form of a geospatial point shapefile, with single points representing the approximate location of each HWF. This original shapefile contained coordinates for 82 sites. Five facilities that are no longer operating or are undergoing closure were omitted from the analysis after consultation with DTSC, leaving a final list of 77 active HWFs across the state (**Figure 1)**.

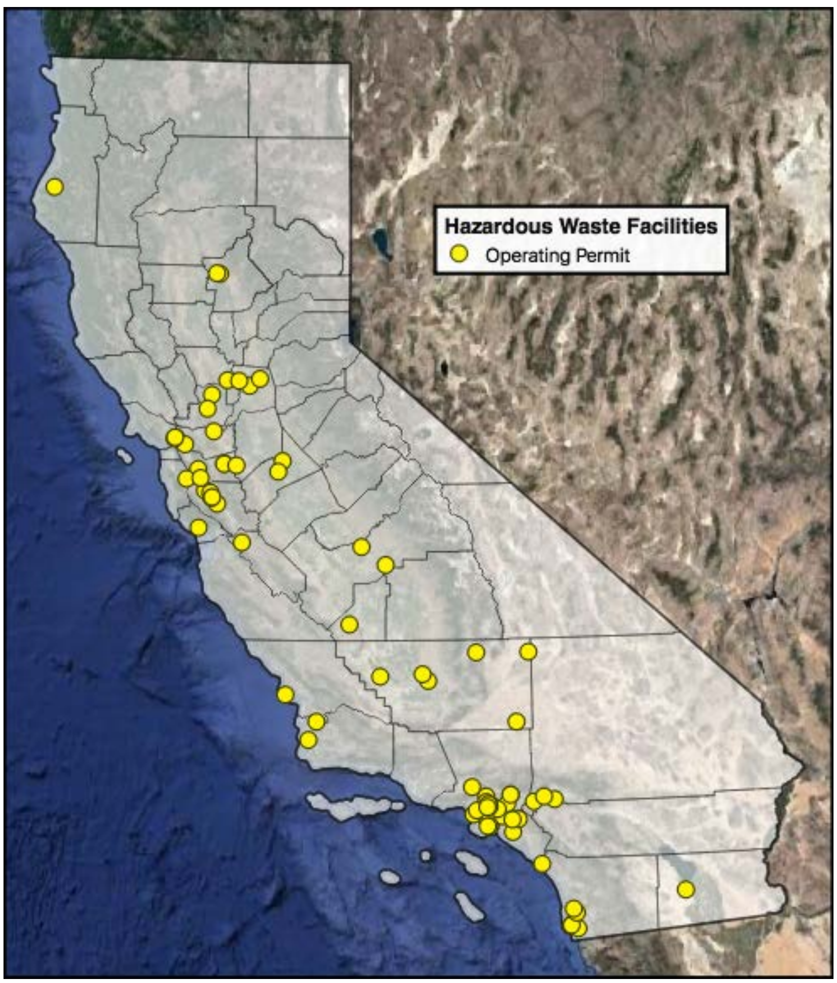


Figure 1. Location of active HWFs

CA Statewide Parcel Data

We utilized a comprehensive, statewide shapefile of all California parcels obtained from CARB was utilized to: 1) construct facility polygons around the point locations provided by DTSC and 2) classify residential regions within census block groups for the purpose of calculating population-weighted metrics of cumulative impact. Each parcel in this dataset has a number of attributes pertaining to various use code classifications which were used to distinguish between residential and non-residential parcels.

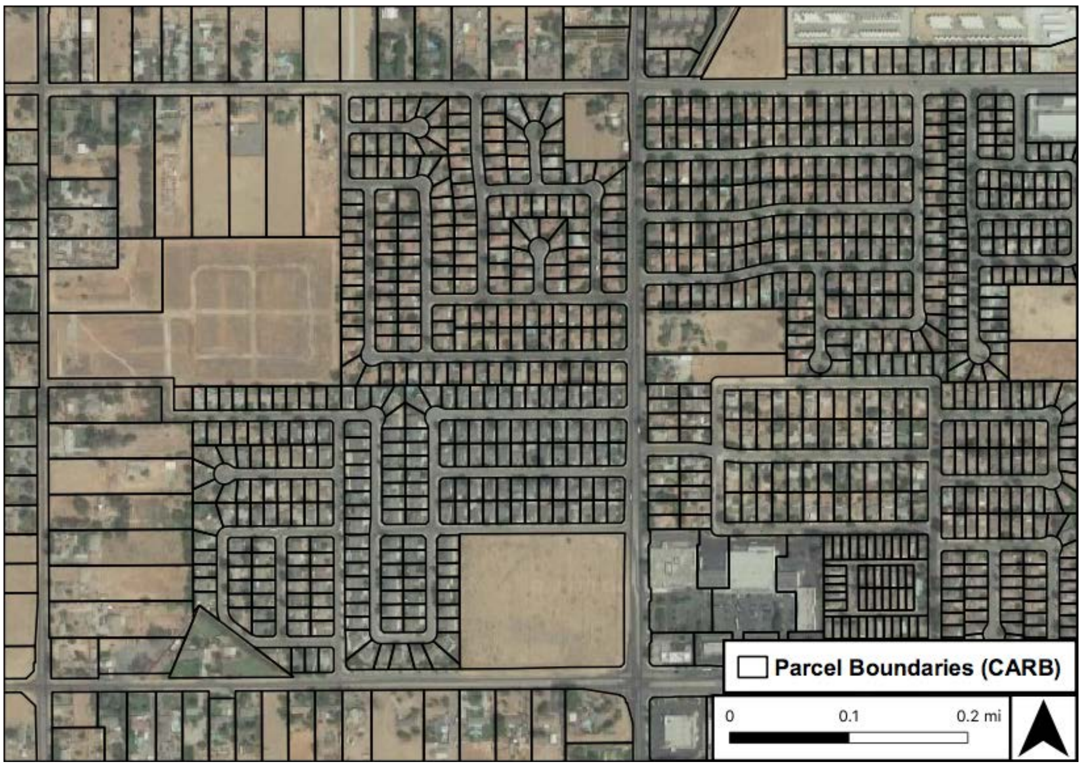


Figure 2. Example parcel boundaries near Fresno, CA.

Population Data

Block-level estimates of population – the highest spatial-resolution of analysis recorded by the US Census Bureau (USCB) – were utilized. However, estimates at the block-level were only last enumerated as part of the 2010 decennial census. Estimates of population at coarser units of analysis, however, are updated continuously as part of the ongoing 5-year American Community Survey (ACS) projects by the USCB. Therefore, population estimates at the block group-level – the finest spatial unit of analysis available in ACS datasets – were taken from the 2013 – 2017 ACS dataset in order to utilize more contemporary estimates of population. This was done by utilizing the block-level population distribution patterns within block groups represented in the 2010 decennial census and applying these same distributional patterns with the updated population totals from the 2013 – 2017 ACS. These population estimates were then disaggregated to the block level using the block-level distributional patterns of population within each individual block-group seen during the 2010 decennial census.

For example, say a block group is comprised of two blocks (Block 1 & Block 2), and, according to the 2010 census, there are 40 people living in one block 1 and 60 people living in Block 2. This means that 40% of the block group's total population lives in Block 1 and 60% in Block 2. Perhaps the new population total for that same block group according to the 2013 – 2017 ACS estimates has increased to 120 people. In order to roughly approximate how these 120 people are distributed between Block 1 & Block 2, it was assumed that the relative distributions of people is equal to those observed in 2010 (40%, 60%). Therefore, the new block-level population estimates for the 2013 – 2017 period was assumed to be 40% of 120, or 48 people for Block 1, and 60% of 120, or 72 people for Block 2. This assumption is obviously not valid in all cases, but is based on the belief that relative population distributions within block groups likely does not dramatically change over a ~5-year timeframe. For block groups with non-zero populations in the 2013 – 2017 ACS that had no population in the 2010 census, the ACS population was assumed to be uniformly distributed across all blocks in that block group given the lack of antecedent knowledge.

Tabulated decennial census and ACS data were downloaded from the NHGIS data server[1] for the state of California along with geospatial polygon shapefiles of the 2010 block and 2017 block group boundaries. The percentage of residents of color was defined as the percentage of residents who identified as Hispanic or as being of any non-White race (including multiracial) and was estimated using ACS data from 2013 – 2017. These values were assumed to be uniformly distributed across the populated areas (as identified using parcels, block populations and building footprints – see Methods) within each block group.

[1] https://www.nhgis.org/

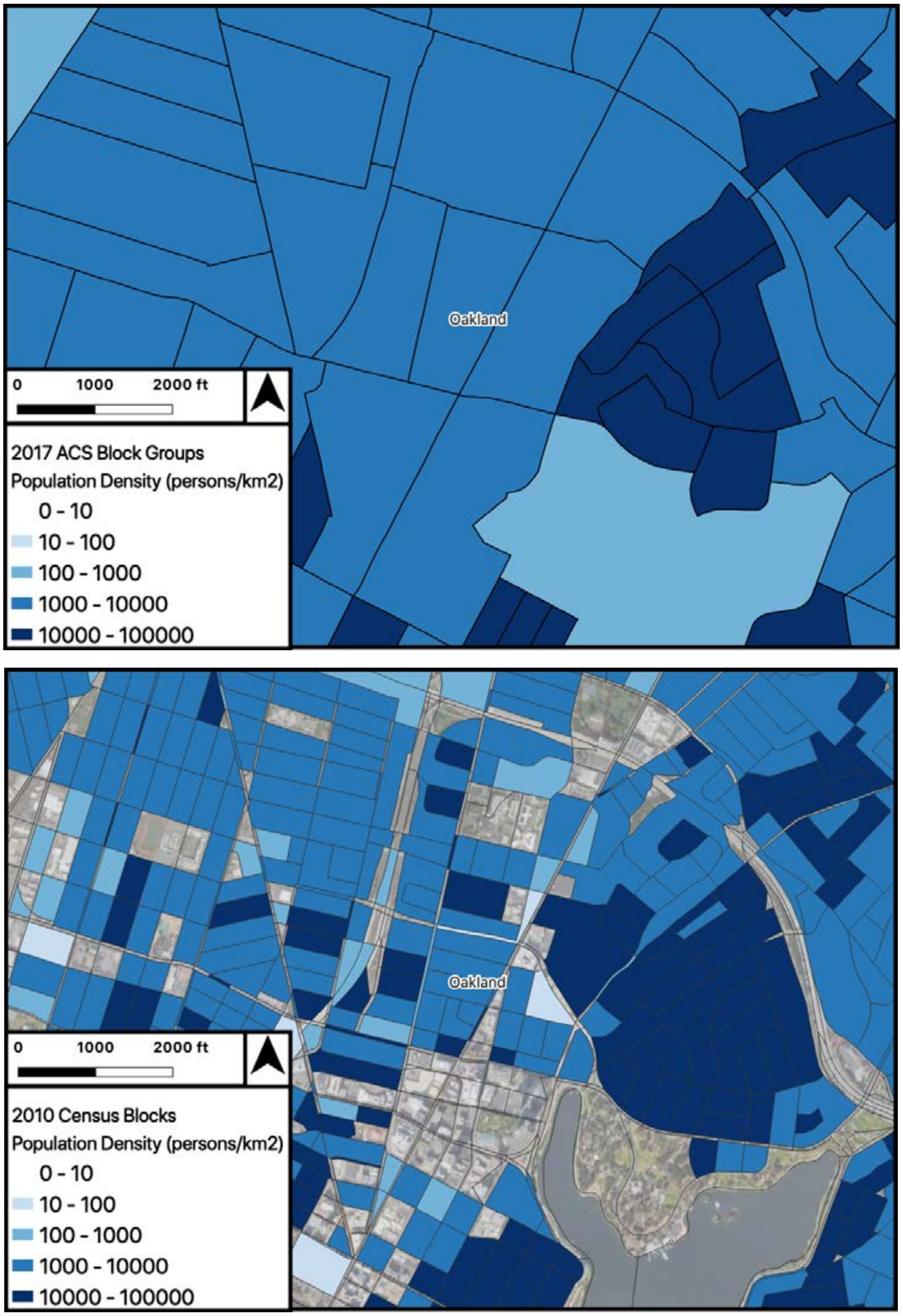


Figure 3. Example block group (top) and block-level (bottom) population data in Oakland.

Building Footprint Data

In some cases, distinguishing between open space and potentially populated areas within blocks was done with the help of a dataset of remotely-sensed building extents, or footprints, produced by Microsoft for the entire country in 2018. This dataset used publicly-available satellite imagery of the

US and employed a series of machine learning classification algorithms to identify likely building rooftops, converting these footprints to a polygon shapefile for each state. More information on the production of this dataset can be found on its download page[2]. Further explanation as to how these data were used for this analysis is provided in Methods.

Figure 4. Example of building footprint data in Oakland.

Facility Operating Permits

We used the final operating permit documents for each HWF in combination with the parcel data to better delineate facility boundaries and to determine the specific locations of waste stored within each facility boundary. Permits were reviewed for each facility both by DTSC staff and the UC Berkeley project team in order to identify property lines and waste locations from maps and figures

[2] https://github.com/Microsoft/USBuildingFootprints/

in the permitting documentation. The permits for operating facilities can be found on DTSC's EnviroStor web platform, or directly using this link[3].

CalEnviroScreen 3.0

The third version of the California Communities Environmental Health Screening Tool (CalEnviroScreen 3.0, or CES 3.0) was downloaded from the Office of Environmental Health Hazard Assessment website[4] and used to assess cumulative impacts surrounding each HWF. CES 3.0 is an aggregate index combining 20 indicators of pollution burden and population vulnerability into a relative cumulative impact score for each census tract in the state. The final scores are also expressed on a percentile scale from 1 to 100, with higher scores/percentile indicating higher levels of cumulative impact. Both raw CES scores and percentiles were calculated and included in the results.

Domestic Well Data

The location of domestic drinking water wells was estimated using the Online System for Well Completion Reports (OSWCR) maintained by the California Department of Water Resources (DWR) and downloaded September 1, 2018[5]. This dynamic dataset includes information on the approximate location of domestic drinking water wells across the state. Exact locations (<50 ft of precision error) are known for some wells but the vast majority of wells are currently represented as part of a generalized well count for ~2-3 km^2 "well sections". These sections form a somewhat uniform grid of rectangular land areas across the state with each section containing a certain number of active drinking water wells whose precise location within the section is unknown (**Figure 5)**.

---

3 https://tinyurl.com/y2md3mrw (List of Operating Facilities with Permits)
4 https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30
5 https://water.ca.gov/Programs/Groundwater-Management/Wells/Well-Completion-Reports

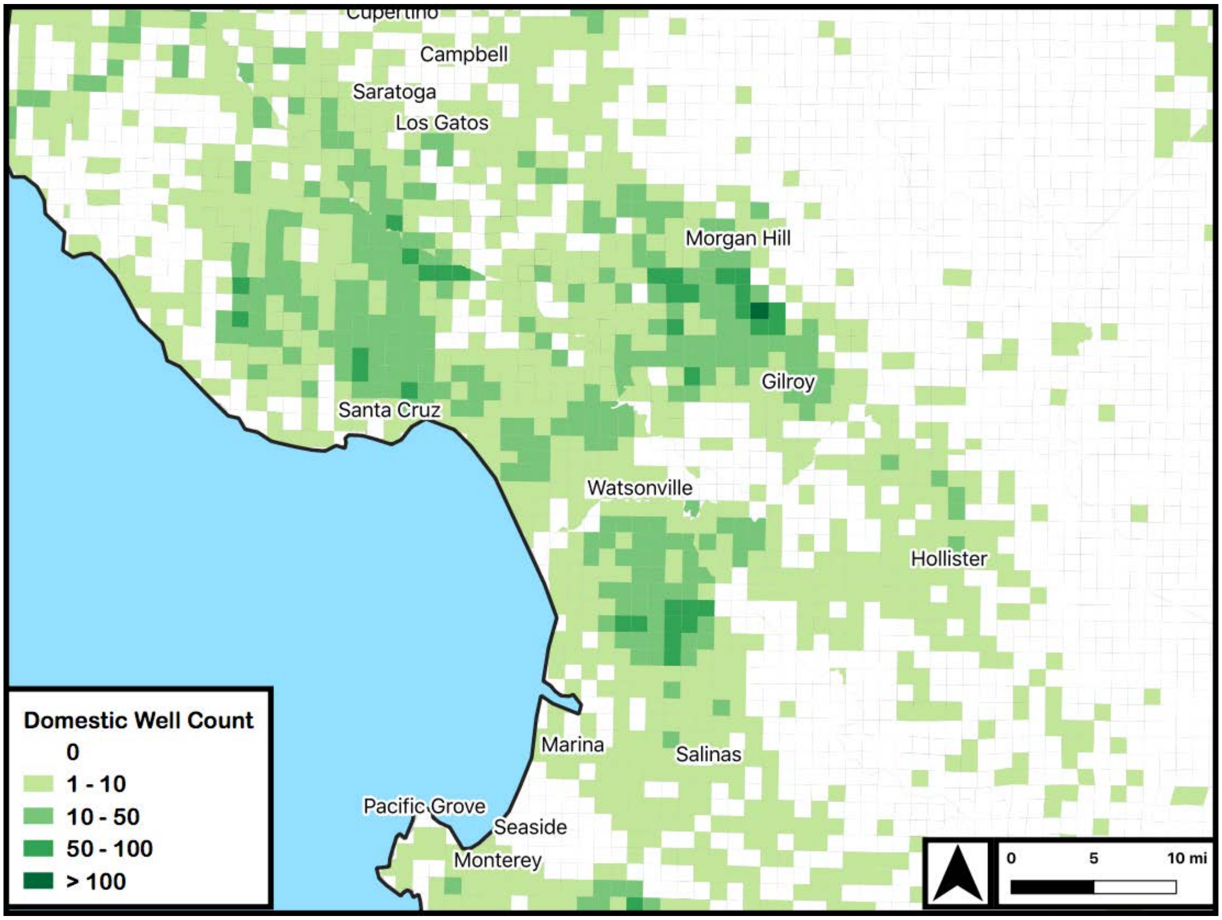


Figure 5. Domestic drinking water well counts by section in the Monterey Bay area of California.

Oil and Gas Well Data

A dynamic database of the location and characteristics of oil and gas wells across the state is maintained by the California Department of Conservation's Division of Oil, Gas, and Geothermal Resources (DOGGR)[6]. We utilized the "All Wells" shapefile for this study, which contains point locations of oil and gas wells throughout the state as well as their current operating status (**Table 1**). Only wells that were classified as "New" or "Active" were included in the analysis (69,531 total wells). The vintage of the dataset used is July 10, 2019.

[6] https://www.conservation.ca.gov/dog/maps/Pages/GISMapping2.aspx

Table 1. Oil and Gas wells counts by operation status

Well Operating Status	Count
Abeyance	2
Active	**65450**
Buried	2
Canceled	7453
Idle	36185
New	**4081**
Plugged	122851
Unknown	1966

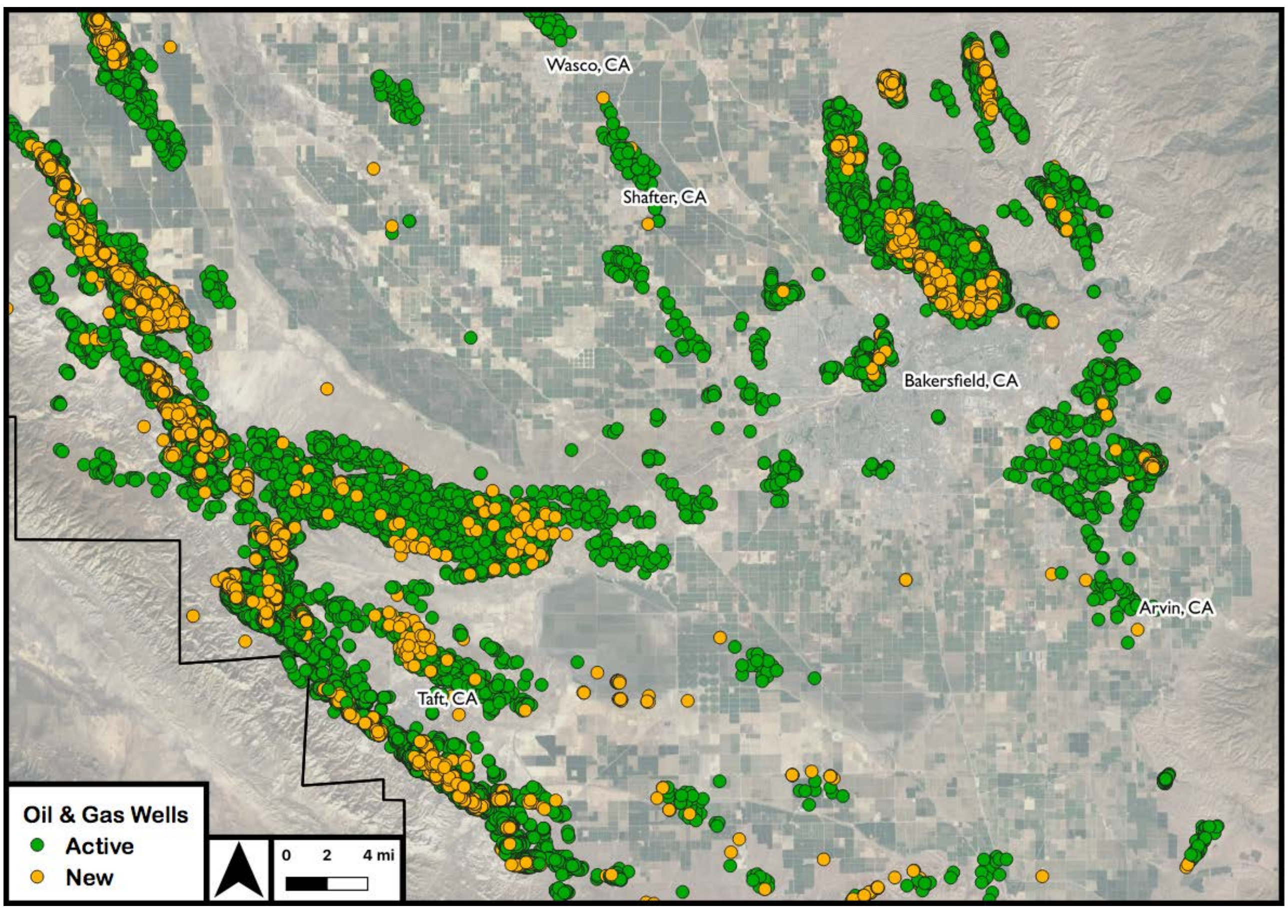


Figure 6. Active or new oil and gas well locations in Kern County, CA.

Voter Turnout Data

The number of people who vote in elections provides a measure of civic engagement capacity and the degree to which communities are involved in local decision-making, which may have implications for community engagement in permitting and regulatory decisions. We utilized voter data from the UC Berkeley Statewide Database 2016 and 2012 General Election Precinct Data at the registration precinct level (RGPREC)[7] to derive the average (mean) percent of registered voters who participated in the 2012 and 2016 elections.

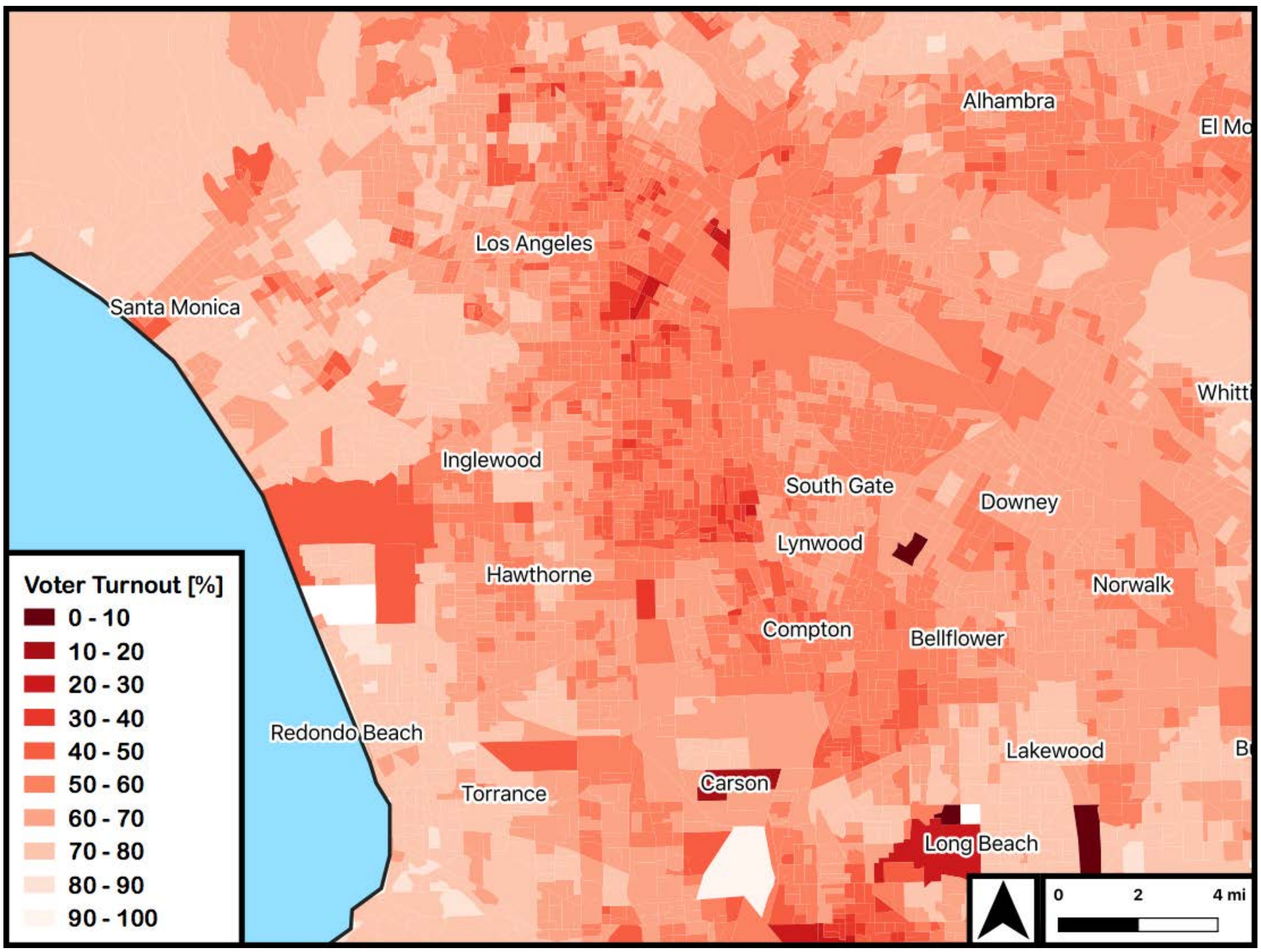


Figure 7. Average registered voter turnout by census block group in the 2012 and 2016 general elections in the Los Angeles metro area.

[7] https://statewidedatabase.org/d10/g16.html

Sensitive Land Use Data

Locational data for six different types of sensitive land uses (SLUs) were included in this analysis to assess the proximity of such land uses to the HWFs studied. SLUs were defined in this context as areas in which vulnerable populations (children, the elderly, people with respiratory illness) either reside or spend a substantial amount of time. The six SLU classes were chosen to be consistent with the "sensitive uses" used by CARB in its Air Quality and Land Use Handbook[8], and are defined for this project as: parks and playgrounds, pre-college level schools, childcare/daycare facilities, healthcare facilities of the type that house vulnerable populations, senior care and residential facilities, and prisons/correctional facilities. The shapefiles used in our analysis were built using the most recent geospatial information available for each use that is consistent statewide, combining data (and removing duplicates) when more than one data source is available. We also compared these locations to earlier versions of this type of data used in our research as a means of validating locations and understanding changes in the location and number of these uses statewide. These locations were located either as geocoded points (healthcare facilities, senior care facilities) or polygons (parks, prisons, schools).

The statewide parcel data provided by CARB was of limited utility in this process because of the poor match between the parcel use codes and the land use designation of most sensitive land use parcels. Queries using use code to identify sensitive uses returned far fewer locations than other data sources that also identify these uses. Parcels identified in this way were included in the final datasets for each sensitive use. However, the parcel dataset cannot be used alone to comprehensively identify any of the sensitive uses recognized in this project, and use of the parcel data generally should be cautious as it is primarily constructed to reflect land ownership for tax purposes. Where appropriate, geocoding was performed using two different address locater street data layers[9,10]. A minimum geocoding score of 0.8 was be required for each location, and features located by the two street layers had to agree with 10 meters. Geocoded locations were validated when appropriate using comparison with aerial imagery in Google Earth Pro. These point location shapefiles were also cross-checked with older datasets of each facility type, and with the parcels identified using use codes to ensure consistency.

---

[8] https://ww3.arb.ca.gov/ch/handbook.pdf

[9] https://www.openstreetmap.org/#map=5/38.007/-95.844

[10] https://www.tomtommaps.com/mapdata/

Healthcare Facilities: This point shapefile was generated using the 2019 "Licenses and Certified Health Facility Listing" from the Dept of Public Health[11], which lists descriptive information, including addresses, for over 30 types of healthcare facilities. We selected the subset of these facilities that are consistent with the SLU definition for this analysis, and geocoded the addresses. The final shapefile consists of 2375 point-locations.

Senior Care Facilities: This point shapefile was generated using the 2019 "Community Care Licensing for Residential Eldercare" data from CA Dept of Social Services[12]. Similar to the healthcare listings, there were duplicate names, addresses, and addresses that limited the records that could be geocoded. The final shapefile consists of 7470 point-locations.

Childcare/Daycare: The most recent data is the list of "community care facilities" licensed as of Nov 2018 by the CA Dept of Social Services[13]. This includes both facility locations and "family daycare" homes with capacity of 8 or more. There are nearly 20,000 licensed facilities, but once duplicate locations, head-start preschools, closed or inactive facilities and addresses that cannot be geocoded were removed, the final shapefile became 7952 point-locations. We compared these locations with a dataset from 2015 Dun & Bradstreet showing businesses describing themselves as childcare or daycare[14]. Unfortunately, the parcel data from CARB was very incomplete, containing only 248 parcels statewide that are listed as a childcare use.

Schools: We have high confidence in this dataset as there are several high-quality datasets available. We relied on the two most authoritative sources available. The California Department of Education online data portal ("Schools and Districts Datafiles") provides school addresses of both public and private schools[15], which we geocoded to produce point locations. These were combined with school polygons from the recently-updated 'California School Campus Database' from GreenInfo Network[16], which comes from an ongoing project with the Stanford Prevention Research Center.

---

11 https://healthdata.gov/dataset/licensed-and-certified-healthcare-facility-listing
12 https://data.chhs.ca.gov/dataset/community-care-licensing-residential-elder-care-facility-locations
13 https://data.chhs.ca.gov/dataset/community-care-licensing-child-care-center-locations
14 https://www.dnb.com/products/marketing-sales/dnb-hoovers.html
15 https://www.cde.ca.gov/ds/si/ds/pubschls.asp
16 https://www.greeninfo.org/work/project/cscd

The final shapefiles are comprised of 8688 public schools as polygons, and 3076 private schools as points.

Parks and Playgrounds: The parks SLU layer was derived by combining land use polygons from four different statewide datasets:

- Real estate tax parcels provided by CARB for this project; park parcels were identified by "use code"
- The California Protected Areas Database a dataset maintained and updated by the Greeninfo Network[17] that captures open space lands, parks, conservation easements, and preserves statewide, mapped using assessor ownership parcels with more extensive attribute information than the parcel data provided by CARB
- USA Parks – a geospatial dataset produced by ESRI in partnership with TomTom, a private company specializing in location technologies and digital geodatabase products and services. This layer, which ESRI considers its "authoritative" data on parks, gardens and forests, combined with boundary information for national, state and local parks[18].

When compared to current (2018) aerial imagery it is apparent that some parks are represented by polygons in two or more of these data layers. It is also apparent that no one dataset is sufficiently comprehensive to be used alone to represent parks and sensitive land uses for this project. From these three data layers, a single composite and validated dataset was produced by using aerial imagery to identify each candidate SJU ("parks and playgrounds" as defined by CARB in their Air Quality and Land Use Handbook[19] and selecting from each layer the polygon(s) that best represent that SLU visible in the aerial imagery.. The aerial imagery was also used to determine which of these parks qualify as an SLU, using the presence of improvements such as athletic facilities, play structures etc.

Prisons: The polygon data of prison boundaries statewide were collected via ESRI's OpenData site. The 'Prison Boundaries' layer[20] was constructed by the Homeland Infrastructure Foundation - an "online community" of the federal Department of Homeland Security. This is part of the Homeland Infrastructure Foundation-Level Data (HIFLD) Subcommittee, which is responsible for

---

17 https://www.greeninfo.org/work/project/cpad-the-california-protected-areas-database
18 https://www.arcgis.com/home/item.html?id=578968f975774d3fab79fe56c8c90941
19 https://ww3.arb.ca.gov/ch/landuse.htm
20 https://hifld-geoplatform.opendata.arcgis.com/datasets/2d6109d4127d458eaf0958e4c5296b67_0

improvements in data collection, processing, sharing and protection of National geospatial information across multiple levels of the federal government to provide common data sources to multiple agencies.

Because each of these data layers contains some of the uses that fit our criteria, features from these three datasets were combined to produce the final SLU layer for parks and playgrounds used in this project.

METHODS

Defining Facility Boundaries

Entire Facility Boundaries

We created a set of polygons delineating each facility's property boundary using the following process:

- **Step 1** – We reviewed the current operating permit document for the HWF for relevant maps and figures showing the facility location and boundary.
- **Step 2** – We validated the coordinates of DTSC's point location for the site based on the facility address and the permitting documents. For points that appeared to be incorrectly located, we adjusted the location using the permit and provided site address information. For a number of sites, the existing DTSC point appeared to be located at a different address than that listed for the facility. These locations on Google Maps were cross-checked with the permit documents before correcting the point location.
- **Step 3** – We intersected the resulting HWF point locations with the statewide parcel dataset in order to identify the parcels within which each point is located. If this parcel looked to agree with the facility boundaries depicted in the permit, we used this parcel as the final site boundary polygon.
- **Step 4** – When parcel boundaries identified in Step 3 did not appear to match facility boundaries depicted in the permit, we selected different or additional parcels to match the facility boundaries depicted in the permit.
- **Step 5** – If there was no clear depiction of the facility's property boundary in the permit document, we conducted additional online searches regarding the facility and reviewed

satellite imagery from Google Earth, made an educated guess as to its approximate property boundaries, and manually drew the final boundary polygon.

Roughly a third of the site boundaries agreed nearly perfectly with a single intersected parcel and only required Steps 1-3. The majority of the sites required some form of manual alteration described in Step 4. An example of one of each type of site is given in **Figure 8**. Four sites (Edwards, Travis and Vandenberg Air Force Bases and Naval Air Weapons Station China Lake) required rough estimation of facility boundaries described in Step 5 due to their large areas, irregular borders, and lack of access to official property boundary maps or shapefiles.
Additionally, there were four sites for which the entire-facility boundaries were limited to the specific region of waste within them due to their unusual size and the fact that their exact boundaries would be difficult to construct. These sites included the Lawrence Berkeley National Lab, whose facility boundaries are dispersed across the eastern Berkeley hills, and three sites within the San Diego Naval Station/Naval Air Station, which is a massive facility the stretches across islands up the coastline of downtown San Diego. For these four sites, we felt a single polygon for the entire boundary would be large and potentially misleading. Therefore, the "entire-facility" polygons for these four sites either exactly correspond to the "waste-specific" polygons or represent the sub-region of the facility which encompasses the waste storage area.

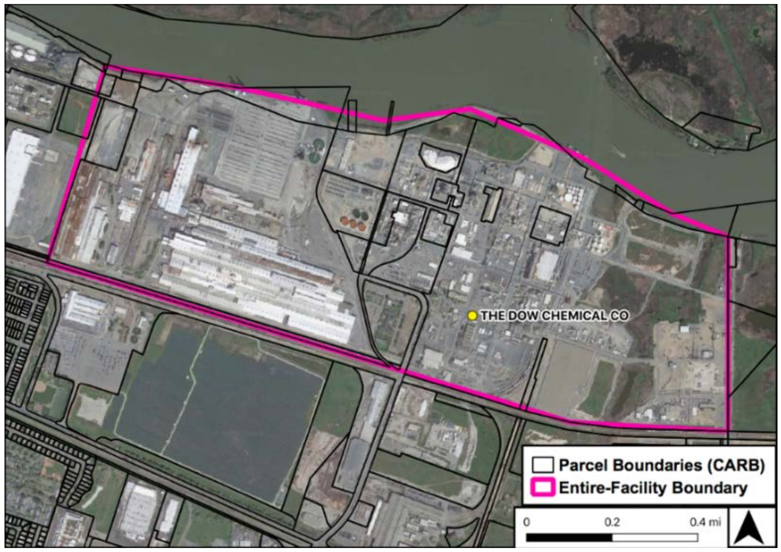


Figure 8. Example of a site whose facility boundary exactly matches a single parcel (top). Example of a site whose facility boundaries spans multiple parcels and required manual drawing of its boundary (bottom).

Waste-Specific Boundaries

A second set of polygons representing the specific locations permitted to process or house hazardous waste within each facility was constructed via additional manual processing. We constructed each of these polygons site by site following a two-step process:

- **Step 1** – We reviewed the current operating permit document for the 77 HWFs for relevant maps and figures showing the specific permitted location of waste within the facility.
- **Step 2** – We manually drew polygons around the waste sites within the facility. This frequently entailed delineating single buildings, tank arrays or storage facilities within the greater property boundary according to permit maps and figures in conjunction with Google Earth satellite imagery. These locations were available in all 77 operating permits.

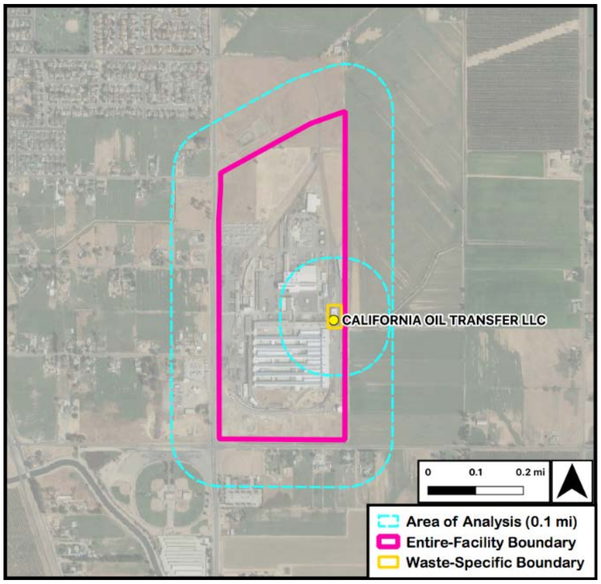
Relatively few of the sites have waste permits for their entire-facility boundaries. Therefore, for the majority of sites the "entire-facility" and "waste-specific" boundaries differ, with the waste-specific being smaller (**Figure 9**).

Figure 9. Entire-facility and waste-specific polygon boundaries for sites within the Chevron refinery complex in Richmond, CA.

Estimating Community Characteristics Near Facilities

Areas of Analysis

In order to assess the characteristics of communities surrounding each HWF, we considered 13 different buffer distances from 0.1 to 7 miles: 0.1, 0.3, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7 miles. The areas delineated by these various buffer distances are referred to as "Areas of Analysis (AoAs)". Each of the 13 AoAs were constructed using both sets of HWF polygon boundaries discussed above (**Figure 10)**.



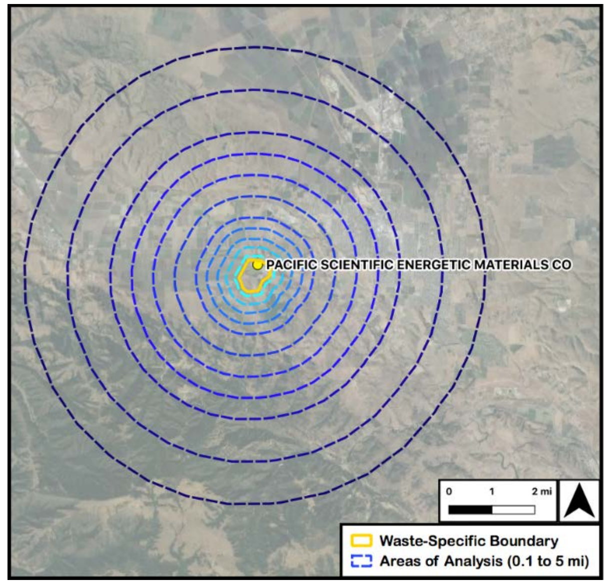


Figure 10. Example of an AoA (0.1 mi) around both the entire-facility boundary and the waste-specific boundary at a facility in Stanislaus County (left). Example of AoAs from 0.1 to 5 mi drawn around a site's waste-specific boundary in San Benito County (right).

Populated Areas

Using the census block-level population estimates from the ACS provides a fairly high-resolution map of population characteristics across the state. However, block groups contain a lot of land area that is non-residential, such as open space, water, vacant land, retail, industrial, or other non-residentially-zoned areas. As a result, it is inaccurate to assume that the population of a given block group is evenly distributed across its land area. Filtering out non-residential areas from each block group yields a more spatially accurate representation of where people live.

This was done using residential parcels from the CARB parcel data, census population at the block group and block-levels (from 2013-2017 and 2010, respectively) and building footprints from Microsoft's US buildings dataset. The final map of population using these data was made in the following steps and illustrated graphically in Appendix 1:

1. **Extrapolate block-level populations** from the 2010 decennial census forward in time using population estimates from the 2013-2017 ACS for parent block-groups. Proportional distribution of population amongst the blocks within each block group was kept constant according to patterns observed in 2010, but with their totals updated to reflect values in the ACS dataset.

2. **Identify residential parcels** from the CARB parcel data using the "USE_CODE_2" classification, which has some 278 unique land use types, of which 30 were identified (see Appendix 1) as being residential (e.g. 'Single Family Residential', 'Apartment House (5+ units)'). We also included "planned residential unit developments" because many of these parcels have already been developed, as evidenced by recent satellite imagery.
3. Create a spatial polygon **layer of only residential parcels.**
4. Of this parcel subset**, identify those residential parcels that likely contain a large amount of open, unpopulated space**. This was defined as individual parcels with an area of more than 1-acre for low-density residential classes (e.g. 'single-family residential') or with more than 50-acres for high-density residence classes (e.g. 'apartment house (100+ units)'). The distinction in thresholds between low and high-density residence types was made due to the observation that for most low-density uses, parcels may be large but only contain a small portion where a home is located and for which people likely are present., leading to the 1-acre cutoff. However, in densely-populated regions, it is common to see single parcels encompass large apartment or condominium developments that can span large areas of urban space, leading to the 50-acre area cutoff for these parcels.
5. Assume that all parcels not excluded in step 4 (< 1-acre or < 50-acre areas), are populated areas, with population distribution assumed to be uniform within each individual parcel. These parcel areas account for roughly 91.8% of the state's total population.
6. For those parcels excluded in step 4 (> 1-acre or > 50-acres), identify the buildings within these parcels using the Microsoft US buildings layer, and make the assumption that the population within these large parcels is distributed only amongst the building areas within it. These areas account for roughly 4.9% of the population.
7. For any blocks with a non-zero population but containing no residential parcels, identify buildings within them and assume population is distributed in these buildings. These areas represent roughly 3.0% of the population.
8. Finally, for any blocks with non-zero population but which contain neither residential parcels nor buildings, simply assume that its population is uniformly distributed across the entire block area. This pertains to blocks containing only roughly 0.3% of the population.
9. Using a combination of these four polygon geometries, (i) small residential parcels, ii) buildings within large residential parcels, iii) buildings within populated blocks with no residential parcels, and iv) boundaries of populated blocks with no residential parcels or

buildings), create a polygon layer representing the union of all of them and assign the block-level population totals only to these areas within each block, assuming uniform population density throughout the block.

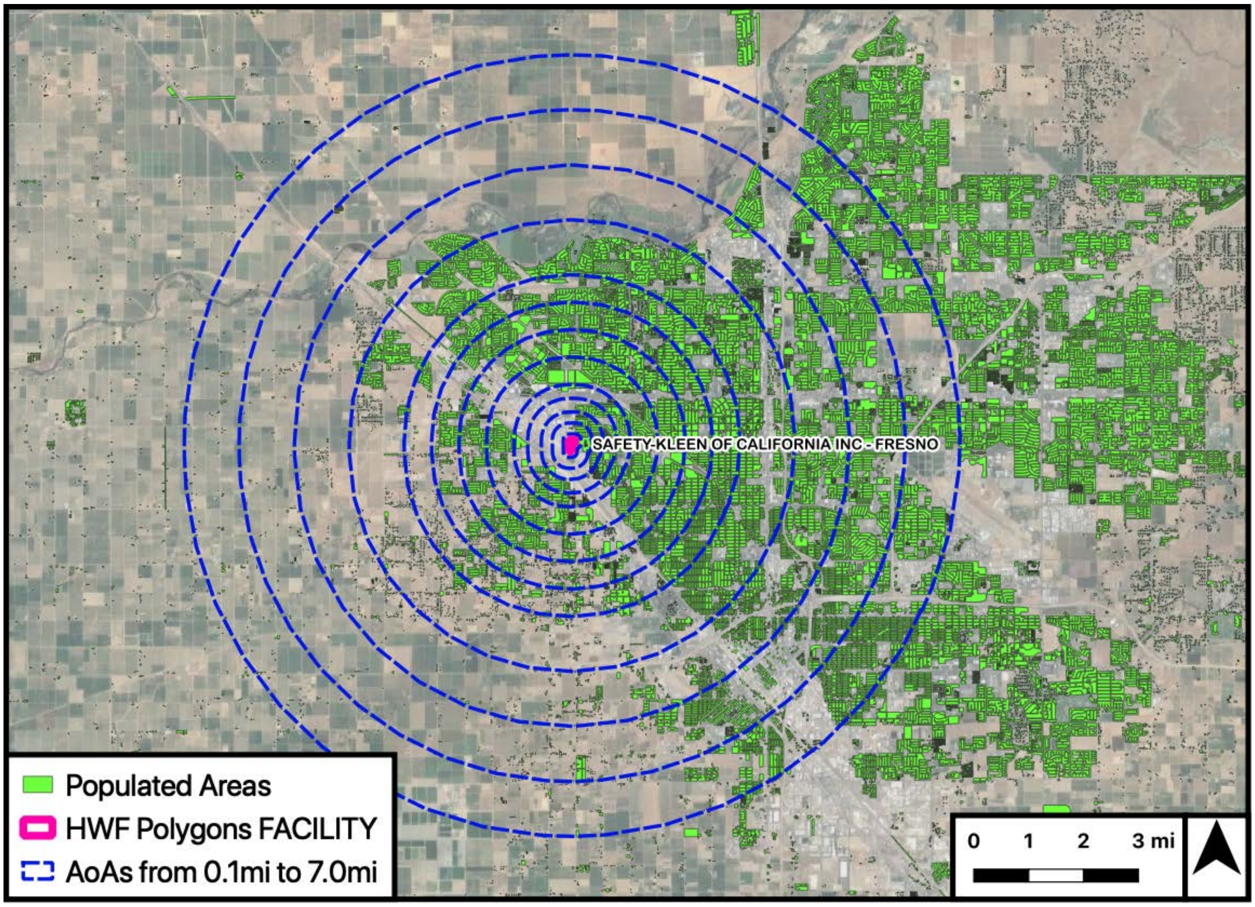


Figure 11. Example of AoAs from 0.1 to 7mi drawn around the waste-specific polygon at a facility in Fresno. Only the green area represents populated areas were included in the analysis.

Unweighted Metrics: Minimum and Maximum CES 3.0 Values, Oil & Gas Wells, Sensitive Land Use (SLU) Counts

The **minimum and maximum CES values (scores and percentiles)** were calculated by simply identifying the smallest or largest CES 3.0 value amongst the populated areas encountered within the given AoA. For example, in **Figure 12**, the maximum raw CES 3.0 score for the 0.5 mi AoA is determined by the populated area encountered in the eastern half of the AoA (value between 40-60). For the 1 and 2-mile AoAs, the maximum score is found in more southern populated areas and is between 60-80.

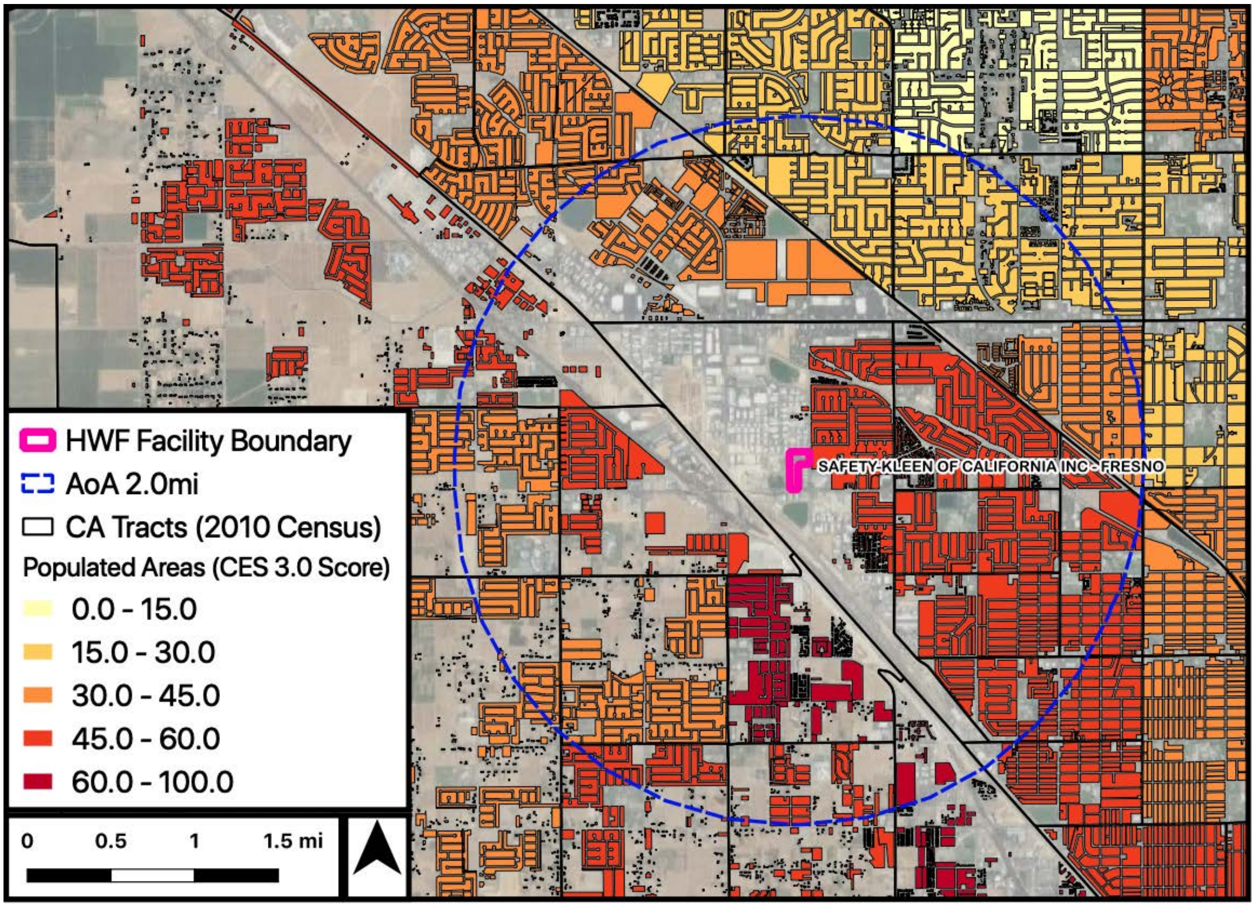


Figure 12. Map of tract-level CES 3.0 scores near the Fresno Safety-Kleen facility. Populated areas were assigned the CES 3.0 score of the tract that contained them.

Counts of sensitive land use (SLU) zones within each AoA were estimated using the point or polygon geometries of each SLU type. If a point or *any part* of a SLU boundary polygon intersected with an AoA, it was counted as being in the AoA. Therefore, all SLUs are each summarized as simple counts, with a total count for all six SLU types reported as well.

Counts of **new or active oil and gas wells** in each AoA surrounding HWFs were calculated using the point-location data of the wells from the DOGGR dataset, with the final counts representing the total number of well points (active or new) that fall within a given AoA.

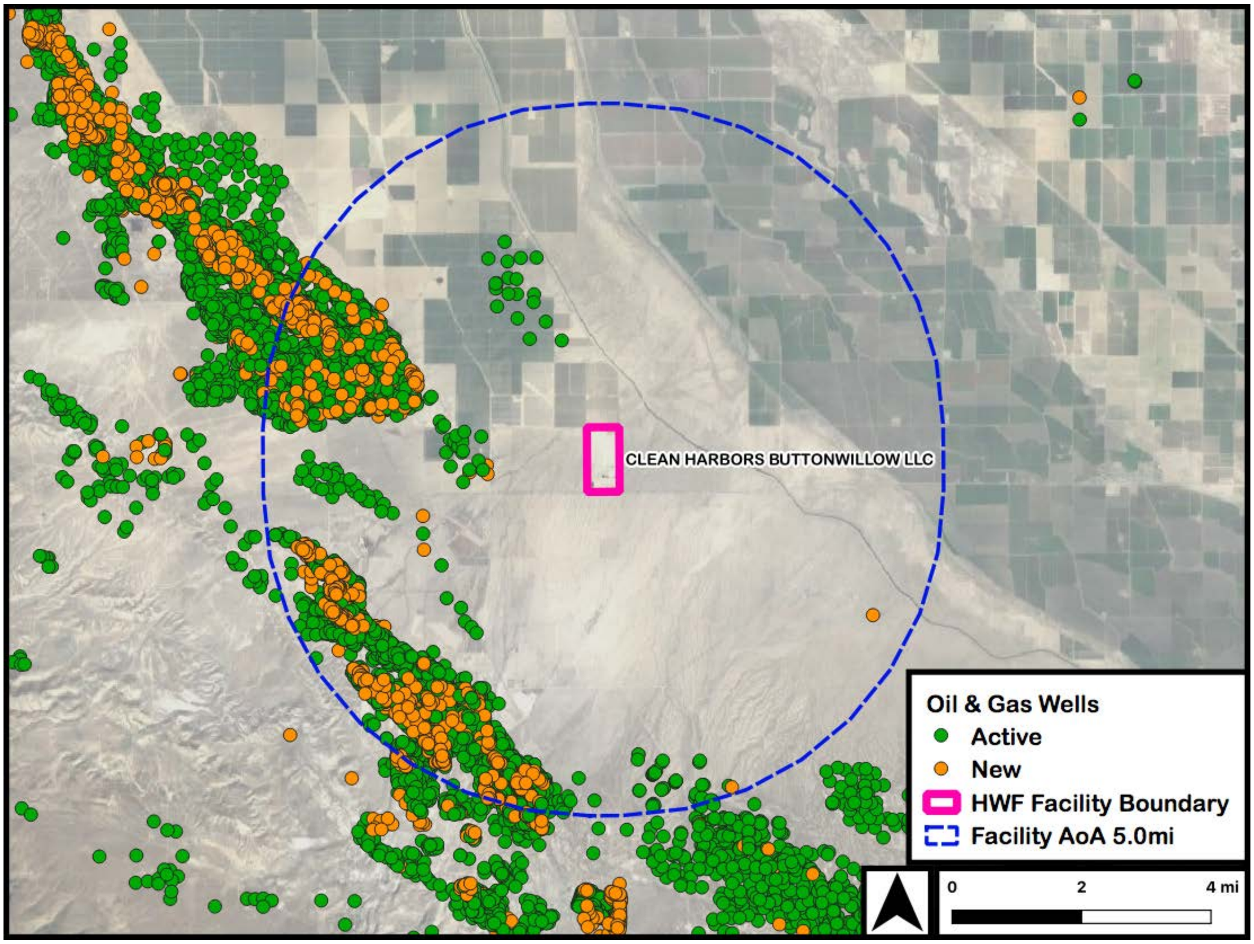


Figure 13. There are 5174 active or new oil and gas wells within the 5.0mi AoA surrounding the Clean Harbors facility in Buttonwillow.

Area-Weighted Metrics: Domestic Drinking Water Wells

In order to estimate the number of domestic drinking water wells within AoAs of each HWF, we utilized a simple area-weighted averaging approach using the sectional well totals provided from the OSWCR dataset. This was done in the following steps:

- **Step 1** – We filtered out all well section geometries that have a well count of 0.
- **Step 2** – Assuming that domestic drinking water wells predominantly occur within populated areas, we intersected the well section geometries with the populated area geometries and assigned the well totals for each section to the populated areas within each section, excluding non-populated areas from analysis. For sections containing wells that did not intersect with any populated areas, we assumed that their wells are uniformly distributed across the section area. Roughly 5.4% of all registered domestic drinking water wells fell

into this category, suggesting that some domestic wells may no longer be used, and/or that the populated area data being used does not fully capture all residences.

- **Step 3** – We then intersected these populated-area-only well sections with the AoAs and, assuming that each section's wells are uniformly distributed across its populated area, calculated an area-weighted well count for each AoA. For example, if an AoA intersects two sections, encompassing 50% of each section's populated areas, and the well counts of the sections are 6 and 10, respectively, then the estimated total number of wells within the AoA will be:

$$(0.5*6) + (0.5*10) = 8 \text{ domestic drinking water wells in AoA}$$

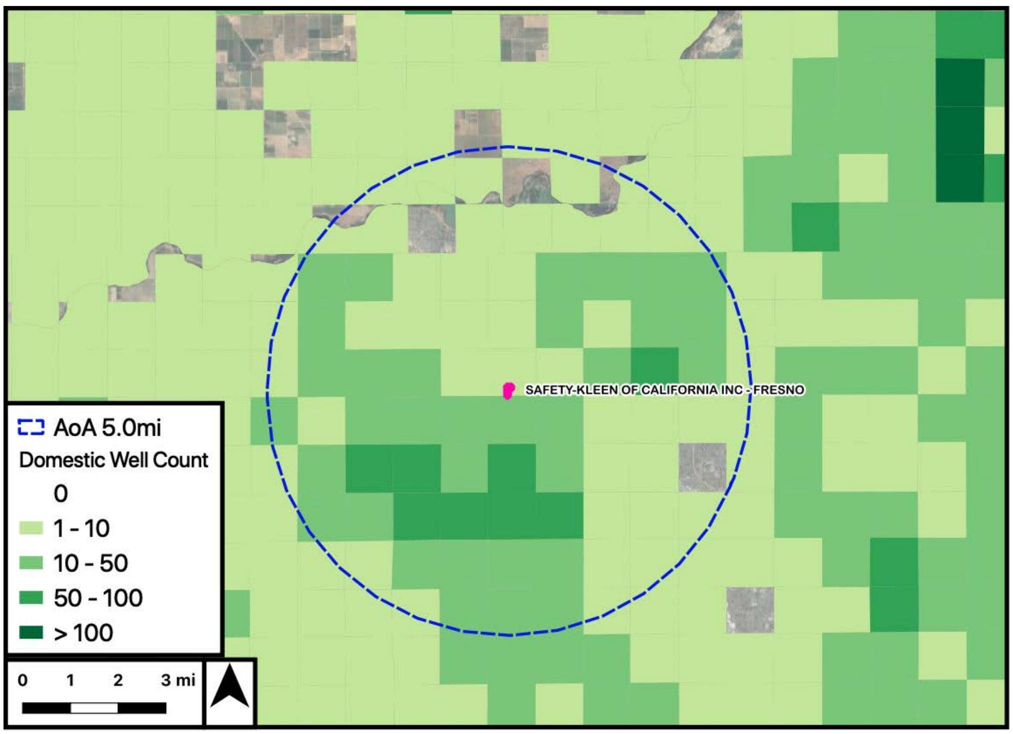


Figure 14a. Domestic drinking water well sections surrounding the Safety-Kleen facility in Fresno with well sections with 0 wells removed

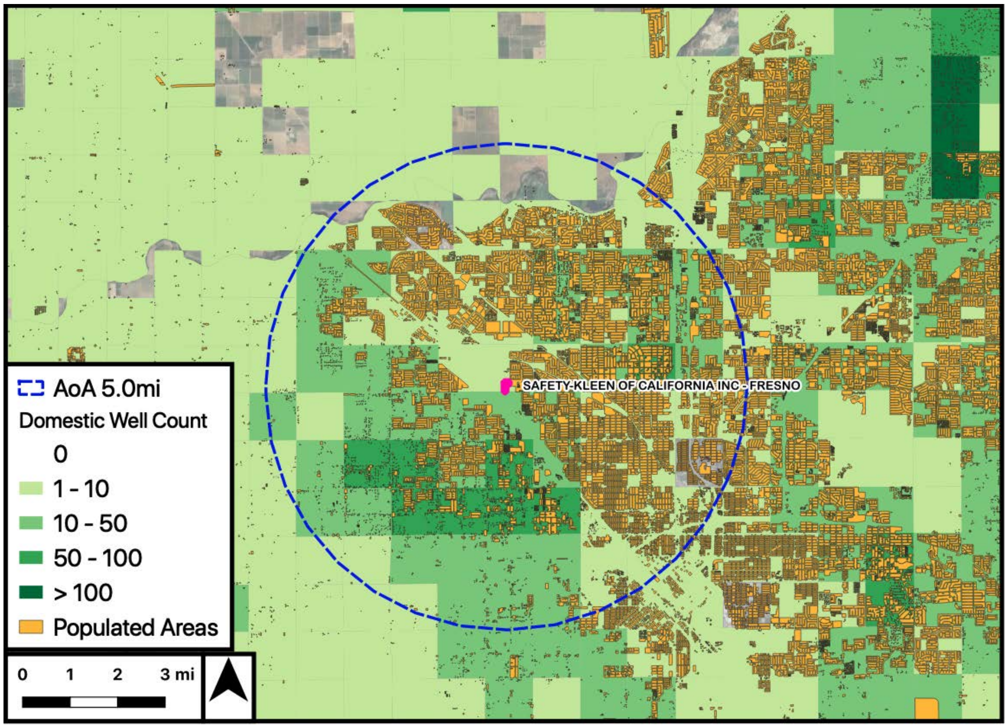


Figure 14b. Populated areas intersected with well sections

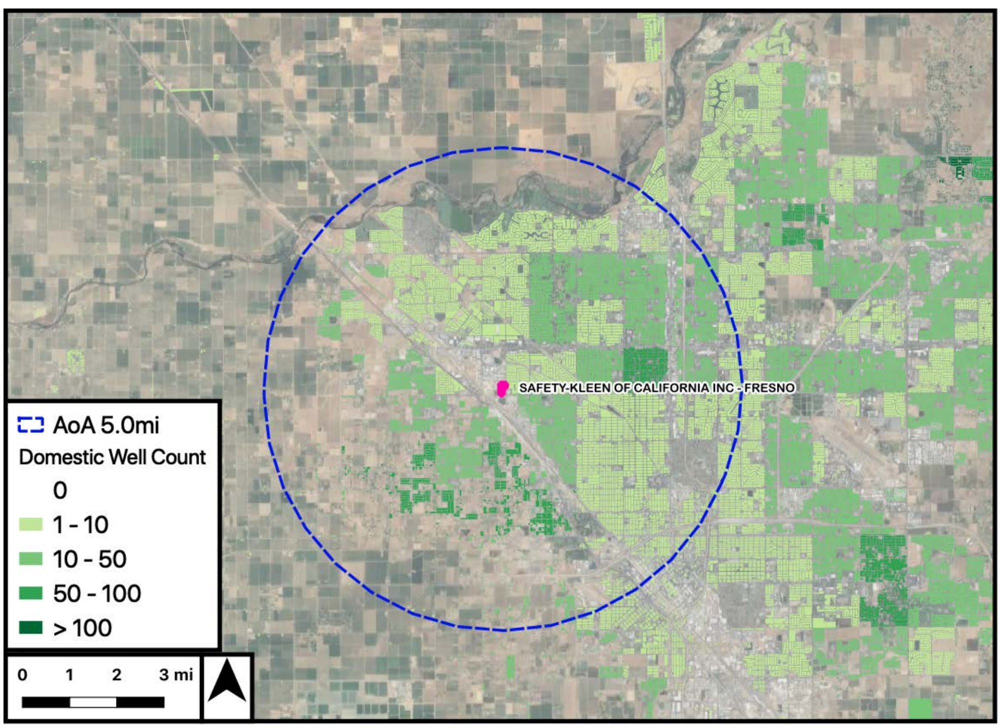


Figure 14c. Well counts assigned to populated areas within sections and intersected with the AoA to find the area-weighted mean well count. Sections with non-zero well counts and no intersecting populated areas are left intact.

Population-Weighted Metrics: Mean CES 3.0 Values, Racial Composition, Voter Turnout

Mean CES 3.0 scores and percentiles, racial composition, and average voter turnout are all metrics that were weighted by population and then averaged to generate a population-weighted mean as follows:

- **Step 1** – We assigned the populations of each block group to the populated area polygons within it assuming a uniform population distribution across the populated areas in each block group.
- **Step 2** – We intersected these populated-area-only block group geometries with each AoA and calculated the percentages of each area that fell within the AoA, and in turn the percentage of their populations within the AoA. This provided an estimation of the total number of people living in each populated area polygon or portion of a populated area polygon in the AoA.
- **Step 3** – Using the population estimates derived in Step 2 as weights, the average metric was calculated by summing the product of the weights and metric values (e.g. CES 3.0 score) from each populated area polygon or portion of a polygon within the AoA. We repeated steps 1-3 for each facility AoA to calculate population-weighted CES scores, percentiles, racial composition and average voter turnout.

Voter turnout data from the UC Berkeley Statewide Database (UCBSD) was downloaded for the 2012 and 2016 general elections at the level of voter registration precinct (RGPREC), which was then re-mapped to populated areas within the 2017 census block groups using the RGPREC to census blocks crosswalk protocol available in the same UCBSD data repository. This average percentage voter turnout by block group was the metric used in the population weighting scheme described above.

RESULTS

All results are tabulated in two excel workbooks, one for community characteristics within AoAs based upon the entire-facility polygons and one for those based upon the waste-specific polygons. Separate sheets are included for each metric (11 sheets total: *Mean CES 3.0 Score, Min CES 3.0 Score, Max CES 3.0 Score, Mean CES 3.0 Percentile, Min CES 3.0 Percentile, Max CES 3.0 Percentile, Non-White %, Domestic Drinking Wells Count, Oil & Gas Wells Count, Sensitive Land Use Counts, Voter Turnout %*).

The facilities are presented as separate rows in alphabetical order. The “ID” column represents a unique identifier for each facility that can be used to merge data between sheets or workbooks. These identifiers were originally created when working with the original set of 82 HWFs, which is why some of the IDs are higher than 77.

Metric values by AoA are presented in separate columns for each HWF and may contain NA values. NA values entail that the given AoA did not intersect any populated areas and therefore has no values of cumulative impact to evaluate.

APPENDIX 1 - DTSC Community Vulnerability Metrics Explanations and Justifications for Inclusion

CalEnviroScreen 3.0 (CES 3.0)

This statewide tool provides information regarding environmental health indicators at the census-tract levels across the entire state. Commissioned and maintained by the California Environmental Protection Agency (CalEPA) and, more specifically, the Office of Environmental Health Hazard Assessment (OEHHA), this database serves as a tool for information transfer and environmental screening at the community level. The newest iteration of this product, version 3.0, incorporates a wide array of pollution, demographic and socioeconomic metrics to estimate cumulative environmental burdens facing communities. This product is widely used both by policy-makers, practitioners, academics and community organizations in order to identify and implement policies that are sensitive and responsive to environmental inequities[21,22,23,24].

Cumulative burdens are reported in terms of raw scores (ranging from roughly 0 to 95.0), which are calculated via a multi-step algorithm that incorporates the multiple factors considered, as well as in percentile terms (ranging from 0 - 100), which provides a relative measure of burden experienced by a given community compared to the rest of the state. Both the raw scores and percentiles were provided in this analysis, and may each be appropriate for use in assessing community vulnerability, depending on the context of the research being done or questions being asked. Using the raw scores will provide a true reflection of the actual cumulative burden experienced by each census tract, while using percentiles will only provide a relative measure.

Using a simplified example, suppose there are only ten tracts in the state, three of which have a score of 30.0, one of which has a raw score of 80.0, and the remaining six with scores of 95.0.

---

[21] Padula, Amy M et al. "Environmental pollution and social factors as contributors to preterm birth in Fresno County." *Environmental health : a global access science source* vol. 17,1 70. 29 Aug. 2018, doi:10.1186/s12940-018-0414-x

[22] Cushing, L., Faust, J., August, L. M., Cendak, R., Wieland, W., & Alexeeff, G. (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), 2341-2348.

[23] Meehan August, L., Faust, J. B., Cushing, L., Zeise, L., & Alexeeff, G. V. (2012). Methodological considerations in screening for cumulative environmental health impacts: Lessons learned from a pilot study in California. International journal of environmental research and public health, 9(9), 3069-3084.

[24] Mataka, A., & Galaviz, V. (2016, October). CalEnviroScreen: A Pathway to Address Environmental Justice Issues in California. In APHA 2016 Annual Meeting & Expo (Oct. 29-Nov. 2, 2016). American Public Health Association.

Analyzing these raw scores will tell the observer that most of the tracts have a very high level of burden, with 7 out of 10 experiencing a score of 80 or higher. However, using the percentile analysis, could distort this understanding to some extent. In our simple example above, given the high proportion of scores equal to 95.0, the tract with the score of 80.0 would be placed in the 40th percentile. In other words, the percentile value of 40% for the tract with a score of 80.0 would indicate that 60% of the state has a higher score than this tract, which may make it seem like the tract has a low level of burden, but in reality is only saying that its level of burden is lower *relative* to the remainder of the state's tracts. However, if the analysis at hand is specifically oriented towards identifying the *relative* level of burden experienced by each tract relative to the rest of the state, then using percentiles would be appropriate. It is up to the investigator to decide the most appropriate metric to utilize.

When studying the CES 3.0 scores and percentile values to assess the level of environmental health burden in a given area of analysis (AoA) that encompasses multiple tracts, it is also prudent to consider whether the tract-averaged values are the best metric to consider, or simply the maximum score or percentile present within the AoA. Using a simple maximum will highlight the *most* burdened tract in the AoA, a value that is probabilistically expected to increase if the AoA grows in size and more tracts are included. This is valuable if the analysis at hand is aimed at identifying the presence of *any* particularly high-burdened tracts rather than assessing the average level of burden across the AoA. However, if multiple AoAs are being assessed and compared, using a simple maximum score/percentile metric could be inadequate to truly assess the relative differences in burdens experienced between different AoAs as a whole.

For example, it is possible that one AoA could have a low-level of burden overall, with most of its tracts having low CES scores, but perhaps has one small tract with a high CES score. Perhaps a neighboring AoA has a much higher level of burden overall, with all of its tracts with higher CES scores. However, suppose that none of the tracts in the more-burdened AoA individually have a score equal to or higher than that of the single high-score tract in the first AoA. Using a simple maximum CES score as the metric of analysis would identify the first AoA as being more highly-burdened as compared to the second AoA, even though on average, the level of burden across the second AoA as a whole is much higher than in the first. Using instead an average CES score or percentile metric would identify the second AoA as more burdened than the first, though it would mask the presence of the single high value in the first. Therefore, it is likely always appropriate to consider both the mean and maximum metrics when conducting analyses of multiple AoAs and is again up to the investigator to choose the priorities of their analysis in order to inform the way in which they interpret these metrics.

Racial Composition

Analysis of racial and ethnicity-based metrics is commonly done when assessing issues of community vulnerability and environmental equity/justice more broadly. Given the legacy of segregation, inequality and marginalization of communities of color in the United States, they are often disproportionately exposed to hazards, environmental and otherwise. There is a very strong precedent for including such metrics in environmental health and community vulnerability studies, especially in the last three to four decades[25,26,27,28,29].

Healthcare & Senior Care Facilities

Senior centers and medical facilities such as hospitals, health clinics, and nursing homes, are all considered sensitive land uses, as individuals within these types of facilities are the most vulnerable to health risks from exposure to poor air quality. Individuals older than 65 years of age are more susceptible to air pollution-related illnesses such as stroke, asthma, heart disease, lung cancer, and other respiratory diseases. Similarly, those individuals with pre-existing medical conditions, such as those people admitted in hospitals and other healthcare facilities, are more prone to developing air pollution-related illnesses[30].

Parks

Park are sensitive land uses in which populations uniquely susceptible to environmental hazard exposures, including children and older adults, are likely to spend time[31]. While parks bring health

---

[25] Bullard, R. D. (1993). Race and environmental justice in the United States. Yale J. Int'l L., 18, 319.
[26] Maantay, J., & Maroko, A. (2009). Mapping urban risk: Flood hazards, race, & environmental justice in New York. Applied Geography, 29(1), 111-124.
[27] Bullard, R. D., Mohai, P., Saha, R., & Wright, B. (2008). Toxic wastes and race at twenty: Why race still matters after all of these years. Envtl. L., 38, 371.
[28] Morello-Frosch, R., Pastor, M., & Sadd, J. (2001). Environmental justice and Southern California's "riskscape" the distribution of air toxics exposures and health risks among diverse communities. Urban Affairs Review, 36(4), 551-578.
[29] Pastor, M., Sadd, J., & Hipp, J. (2001). Which came first? Toxic facilities, minority move-in, and environmental justice. Journal of urban affairs, 23(1), 1-21.
[30] CARB Land Use Handbook, available at: https://ww3.arb.ca.gov/ch/handbook.pdf
[31] Air Quality and Land Use Handbook: A Community Health Perspective; California Environmental Protection Agency; California Air Resources Board, 2005.

benefits through facilitating outdoor physical activities, performing physical activities in polluted environments also has adverse health effects[32]. Therefore, reducing potentially hazardous exposures to pollution in parks can ensure their net health benefits.

Prisons

Compared with the general population, prisoners tend to have higher rates of underlying health conditions, including higher odds of chronic (e.g. asthma, cardiovascular disease, arthritis, and cancer)[33] and infectious diseases (e.g. HIV , hepatitis, and tuberculosis), and mental disorders[34]. By virtue of being incarcerated, prisoners have little to no control over their living conditions and are also likely to have inadequate access to health care[35]. Furthermore, prisoners are faced with worse living conditions such as overcrowding, which in turn leads to the prevalence of infectious diseases and mental disorders[36]. These conditions can make this community uniquely susceptible to the adverse health effects of environmental hazard exposures.

Schools and daycare centers

Children are sensitive to pollution given their small size, high metabolic rates, and developing lung structure and immune systems. In addition to health consequences, air pollution may cause some students to be absent from school, leading to other social cost (e.g. school dropout, parents missing work, and cut in attendance-based school funding). For children with respiratory issues, not going to school on a heavily polluted day is either a result of respiratory problems triggered by air pollution or a preventive measure. Since children spend more time indoors, their exposures are strongly

---

32 Li, F.; Liu, Y.; Lü, J.; Liang, L.; Harmer, P. Ambient Air Pollution in China Poses a Multifaceted Health Threat to Outdoor Physical Activity. J Epidemiol Community Health 2015, 69 (3), 201–204. https://doi.org/10.1136/jech-2014-203892.

33 Binswanger, I. A.; Krueger, P. M.; Steiner, J. F. Prevalence of Chronic Medical Conditions among Jail and Prison Inmates in the USA Compared with the General Population. J. Epidemiol. Community Health 2009, 63 (11), 912–919. https://doi.org/10.1136/jech.2009.090662.

34 Fazel, S.; Baillargeon, J. The Health of Prisoners. The Lancet 2011, 377 (9769), 956–965. https://doi.org/10.1016/S0140-6736(10)61053-7.

35 Wilper, A. P.; Woolhandler, S.; Boyd, J. W.; Lasser, K. E.; McCormick, D.; Bor, D. H.; Himmelstein, D. U. The Health and Health Care of US Prisoners: Results of a Nationwide Survey. Am. J. Public Health 2009, 99 (4), 666–672. https://doi.org/10.2105/AJPH.2008.144279.

36 García-Guerrero, J.; Marco, A. Overcrowding in Prisons and Its Impact on Health. Rev. Esp. Sanid. Penit. 2012, 14 (3), 106–113.

correlated with pollution concentration in schools and home environments and during transportation[37,38].

Oil and Gas Wells

Oil and gas well development (OGD) involves the development of oil/gas sites and wells (production and injection for enhanced recovery), transport of materials to and from well sites, drilling, operation of equipment to recover oil/gas, and collection and disposal of chemicals and waste separated from the raw oil and gas[39,40]. These activities are associated with diverse environmental hazards including air and water pollutants, noise, odors, excessive and inappropriate lighting, and undesired land use changes[41,39]. As of 2017, California (CA) was one of the top five producers of crude oil in the country[42]. Four of the ten largest US oil fields are in CA's San Joaquin and Los Angeles Basins[39,40] and unlike newer shale gas plays, most of CA's natural gas is extracted from reservoirs also producing oil[39,40]. Stimulation techniques, such as water and steam injection and hydraulic fracturing (HF), are o used at established sites rather than newly drilled wells. Oil recovered via water flooding and steam injection (conventional enhanced oil recovery methods) accounted for 76% of the state's oil production in 2009 while HF accounted for 20% of CA's oil production in the last decade[39,40]. The application of unconventional techniques can enhance

---

[37] Currie, J.; Hanushek, E. A.; Kahn, E. M.; Neidell, M.; Rivkin, S. G. Does Pollution Increase School Absences? Rev. Econ. Stat. 2009, 91 (4), 682–694. https://doi.org/10.1162/rest.91.4.682.
[38] Ashmore, M. R.; Dimitroulopoulou, C. Personal Exposure of Children to Air Pollution. Atmos. Environ. 2009, 43 (1), 128–141. https://doi.org/10.1016/j.atmosenv.2008.09.024.
[39] Long JCS, Feinstein LC, Bachmann CE, Birkholzer JT, Camarillo MK, Domen JK, et al. 2015a. An Independent Scientific Assessment of Well Stimulation in California Volume II: Potential Environmental Impacts of Hydraulic Fracturing and Acid Stimulations.
[40] Long JCS, Feinstein LC, Dirkholzer J, Jordan PD, Houseworth JE, Dobson PF, et al. 2015b. An Independent Scientific Assessment of Well Stimulation in California Volume I: Well Stimulation Technologies and their Past, Present, and Potential Future Use in California.
[41] Adgate JL, Goldstein BD, McKenzie LM. 2014a. Potential Public Health Hazards, Exposures and Health Effects from Unconventional Natural Gas Development. Environ Sci Technol 48:8307–8320; doi:10.1021/es404621d.
[42] US EIA. 2018a. CA - State Profile and Energy Estimates. Available: https://www.eia.gov/state/analysis.cfm?sid=CA. US EIA. 2018b. U.S. Crude Oil, Natural Gas, and Natural Gas Liquids Proved Reserves. Available: http://www.eia.gov/naturalgas/crudeoilreserves/

environmental burdens as additional toxic chemicals are used that can potentially be released into air, water, and soil[41,39,40,43,44,45].

Air pollutants associated with OGD include particulate matter with an aerodynamic diameter of < 2.5μm (PM2.5), diesel PM, nitrogen oxides (NOx), secondary ozone formation, mercury, and volatile organic compounds (VOCs) like benzene, toluene, ethylbenzene and xylene (BTEX) from

---

[43] Macey GP, Breech R, Chernaik M, Cox C, Larson D, Thomas D, et al. 2014. Air concentrations of volatile compounds near oil and gas production: a community-based exploratory study. Environmental Health 13:82; doi:10.1186/1476-069X-13-82.

[44] Roy AA, Adams PJ, Robinson AL. 2014a. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. Journal of the Air & Waste Management Association 64:19–37; doi:10.1080/10962247.2013.826151.

[45] Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. 2014a. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. Environmental Science & Technology 48:8334–8348; doi:10.1021/es405118y.

truck traffic, drilling, hydraulic fracturing, production and flaring[46,47,48,49,50,51,52,53,54,42,55,56,57,43,58]. Additionally, fugitive toxic air contaminants can escape at the wellhead[59,57] that might impact health of communities living near points of release. Water contaminants associated with OGD include gas-phase hydrocarbons, chemicals mixed in drilling fluids, and naturally occurring salts, and metals and radioactive elements within shale that surface with wastewater along with recovered oil and gas and

---

46 Allshouse WB, McKenzie LM, Barton K, Brindley S, Adgate JL. 2019. Community Noise and Air Pollution Exposure During the Development of a Multi-Well Oil and Gas Pad. Environ Sci Technol 53:7126–7135; doi:10.1021/acs.est.9b00052.

47 Brantley HL, Thoma ED, Eisele AP. 2015. Assessment of volatile organic compound and hazardous air pollutant emissions from oil and natural gas well pads using mobile remote and on-site direct measurements. Journal of the Air & Waste Management Association 65:1072–1082; doi:10.1080/10962247.2015.1056888.

48 Colborn T, Schultz K, Herrick L, Kwiatkowski C. 2014. An Exploratory Study of Air Quality Near Natural Gas Operations. Human and Ecological Risk Assessment: An International Journal 20:86–105; doi:10.1080/10807039.2012.749447.

49 Eapi GR, Sabnis MS, Sattler ML. 2014. Mobile measurement of methane and hydrogen sulfide at natural gas production site fence lines in the Texas Barnett Shale. Journal of the Air & Waste Management Association 64:927–944; doi:10.1080/10962247.2014.907098.

50 Esswein EJ, Snawder J, King B, Breitenstein M, Alexander-Scott M, Kiefer M. 2014. Evaluation of Some Potential Chemical Exposure Risks During Flowback Operations in Unconventional Oil and Gas Extraction: Preliminary Results. Journal of Occupational and Environmental Hygiene 11:D174–D184; doi:10.1080/15459624.2014.933960.

51 Franklin M, Chau K, Cushing LJ, Johnston JE. 2019. Characterizing Flaring from Unconventional Oil and Gas Operations in South Texas Using Satellite Observations. Environ Sci Technol 53:2220–2228; doi:10.1021/acs.est.8b05355.

52 Goetz JD, Floerchinger C, Fortner EC, Wormhoudt J, Massoli P, Knighton WB, et al. 2015. Atmospheric Emission Characterization of Marcellus Shale Natural Gas Development Sites. Environ Sci Technol 49:7012–7020; doi:10.1021/acs.est.5b00452.

53 Koss AR, Yuan B, Warneke C, Gilman JB, Lerner BM, Veres PR, et al. Observations of VOC emissions and photochemical products over US oil- and gas-producing regions using high-resolution H3O+ CIMS (PTR-ToF-MS). Atmos. Meas. Tech., 10, 2941–2968, 2017 https://doi.org/10.5194/amt-10-2941-2017.

54 Lan X, Talbot R, Laine P, Torres A, Lefer B, Flynn J. 2015. Atmospheric Mercury in the Barnett Shale Area, Texas: Implications for Emissions from Oil and Gas Processing. Environ Sci Technol 49:10692–10700; doi:10.1021/acs.est.5b02287.

55 Marrero JE, Townsend-Small A, Lyon DR, Tsai TR, Meinardi S, Blake DR. 2016. Estimating Emissions of Toxic Hydrocarbons from Natural Gas Production Sites in the Barnett Shale Region of Northern Texas. Environ Sci Technol 50:10756–10764; doi:10.1021/acs.est.6b02827.

56 Maskrey JR, Insley AL, Hynds ES, Panko JM. 2016. Air monitoring of volatile organic compounds at relevant receptors during hydraulic fracturing operations in Washington County, Pennsylvania. Environmental Monitoring and Assessment 188; doi:10.1007/s10661-016-5410-4.

57 Mellqvist J, Samuelsson J, Andersson P, Brohede S, Isoz O, Ericsson M. 2017. Using Solar Occultation Flux and other Optical Remote Sensing Methodsto measureVOC emissions from a variety of stationary sourcesin the South Coast Air Basin.

58 Warneke C, Geiger F, Edwards PM, Dube W, Pétron G, Kofler J, et al. 2014. Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah: oil and gas well pad emissions compared to ambient air composition. Atmospheric Chemistry and Physics 14:10977–10988; doi:10.5194/acp-14-10977-2014.

59 Garcia-Gonzales DA, Shonkoff SBC, Hays J, Jerrett M. 2019. Hazardous Air Pollutants Associated with Upstream Oil and Natural Gas Development: A Critical Synthesis of Current Peer-Reviewed Literature. Annu Rev Public Health 40:283–304; doi:10.1146/annurev-publhealth-040218-043715.

can contaminate potable water via leaks and spills or evaporation[41,60,39,40]. Noise pollution is associated with well pad construction, truck traffic, drilling, pumps, flaring of gases, and other processes[45,61]. Drilling and production activities occur both during the daytime and nighttime, and light pollution has been previously reported as a nuisance in communities undergoing OGD[39,40], suggesting OGD may impact the health of nearby communities via increased psychosocial stress.

To date, most epidemiological studies on the impacts of OGD have focused on populations in Pennsylvania, Colorado, and Texas. For example, several recent studies have found associations between OGD and various adverse birth outcomes, including reductions in term birth weight[62,63] and increased odds or incidence of low birth weight[64,61], preterm birth[65,66,67] and small for gestational age birth[61,62]. One study indicates that asthma exacerbation is also of concern in relation to OGD[68].

Drinking Water Wells

Communities served by water with elevated contaminant levels are disproportionately poor and Latino, raising environmental justice concerns[69,70]. In 2012, California passed Assembly Bill (AB)

---

[60] Hildenbrand ZL, Carlton DD, Fontenot BE, Meik JM, Walton JL, Taylor JT, et al. 2015. A Comprehensive Analysis of Groundwater Quality in The Barnett Shale Region. Environ Sci Technol 49:8254–8262; doi:10.1021/acs.est.5b01526.

[61] Blair BD, Brindley S, Dinkeloo E, McKenzie LM, Adgate JL. 2018. Residential noise from nearby oil and gas well construction and drilling. J Expo Sci Environ Epidemiol 28:538–547; doi:10.1038/s41370-018-0039-8.

[62] Hill EL. 2018. Shale gas development and infant health: Evidence from Pennsylvania. Journal of Health Economics 61:134–150; doi:10.1016/j.jhealeco.2018.07.004.

[63] Stacy SL, Brink LL, Larkin JC, Sadovsky Y, Goldstein BD, Pitt BR, et al. 2015. Perinatal Outcomes and Unconventional Natural Gas Operations in Southwest Pennsylvania. J. Meliker, ed PLOS ONE 10:e0126425; doi:10.1371/journal.pone.0126425.

[64] Currie J, Greenstone M, Meckel K. 2017. Hydraulic fracturing and infant health: New evidence from Pennsylvania. Science Advances 3:e1603021; doi:10.1126/sciadv.1603021.

[65] Casey JA, Savitz DA, Rasmussen SG, Ogburn EL, Pollak J, Mercer DG, et al. 2015. Unconventional Natural Gas Development and Birth Outcomes in Pennsylvania, USA: Epidemiology 1; doi:10.1097/EDE.0000000000000387.

[66] Walker Whitworth K, Kaye Marshall A, Symanski E. 2018. Drilling and Production Activity Related to Unconventional Gas Development and Severity of Preterm Birth. Environmental Health Perspectives 126; doi:10.1289/EHP2622.

[67] Whitworth KW, Marshall AK, Symanski E. 2017. Maternal residential proximity to unconventional gas development and perinatal outcomes among a diverse urban population in Texas. PLOS ONE 12:e0180966; doi:10.1371/journal.pone.0180966.

[68] Rasmussen SG, Ogburn EL, McCormack M, et al. Association between unconventional natural gas development in the Marcellus Shale and asthma exacerbations. JAMA Intern Med. 2016;176:1334–1343.

[69] Balazs CL, Morello-Frosch R, Hubbard AE, Ray I. Environmental justice implications of arsenic contamination in California's San Joaquin Valley: a cross-sectional, cluster-design examining exposure and compliance in community drinking water systems. *Environ Health*. 2012;11(1):84. doi:10.1186/1476-069X-11-84

[70] Balazs C, Morello-Frosch R, Hubbard A, Ray I. Social Disparities in Nitrate-Contaminated Drinking Water in California's San Joaquin Valley. *Environ Health Perspect*. 2011;119(9):1272-1278. doi:10.1289/ehp.1002878.

685[71], known as the Human Right to Water law, which recognizes the universal right to clean, safe, affordable water among all Californians including disadvantaged communities in rural and urban areas served by community water systems (CWS -- with at least 15 service connections or serving at least 25 year-round residents), small water systems (i.e. <15 service connections) and private domestic wells. Several state and regional agencies tasked with implementing California's Human Right to Water law include the State Regional Water Boards, the Department of Water Resources, and Cal EPA's Office of Environmental Health Hazard Assessment. A major barrier to achieving universal access to clean drinking water is a lack of regulatory oversight and data on untreated drinking water sources, including small water systems and private wells. Little water quality information about these water sources exists because they fall outside the purview of state and federal drinking water regulations. Nevertheless, it is estimated that as many as 1.5 – 2.5 million Californians[72],[73] rely on small water systems or private wells (referred to herein as "domestic wells"), which may face even more significant water quality challenges compared to regulated CWS. Previous studies have sought to characterize the extent to which Californians rely on domestic wells and estimate their water quality and suggest that domestic well users are uniquely vulnerable to potential contamination from diverse agricultural, industrial and other sources with significant EJ concerns[74,75,76,77,78,79].

Voter Turnout

---

71 AB-685 State water policy. The Human Right to Water. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201120120AB685. Accessed November 8, 2019.

72 Johnson TD, Belitz K. Identifying the location and population served by domestic wells in California. *J Hydrol Reg Stud*. 2015;3:31-86. doi:10.1016/j.ejrh.2014.09.002

73 Dieter CA, Maupin MA, Caldwell RR, et al. *Estimated Use of Water in the United States in 2015*. U.S. Geological Survey; 2018. doi:10.3133/cir1441

74 Balazs CL, Ray I. The Drinking Water Disparities Framework: On the Origins and Persistence of Inequities in Exposure. *Am J Public Health*. 2014;104(4):603-611. doi:10.2105/AJPH.2013.301664

75 Anning, David, Paul, Angela P., McKinney, Tim, Huntington, Jena, Bexfield, Laura, Thiros, Susuan. *Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwestern United States*. U.S. Geological Survey; 2012.

76 Ayotte JD, Medalie L, Qi SL, Backer LC, Nolan BT. Estimating the High-Arsenic Domestic-Well Population in the Conterminous United States. *Environ Sci Technol*. 2017;51(21):12443-12454. doi:10.1021/acs.est.7b02881

77 Ayotte JD, Nolan BT, Gronberg JA. Predicting Arsenic in Drinking Water Wells of the Central Valley, California. *Environ Sci Technol*. 2016;50(14):7555-7563. doi:10.1021/acs.est.6b01914

78 Ransom KM, Nolan BT, A. Traum J, et al. A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. *Sci Total Environ*. 2017;601-602:1160-1172. doi:10.1016/j.scitotenv.2017.05.192

79 CalEnviroScreen 3.0 | OEHHA. https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30. Accessed October 7, 2019.

Studies in the economic, social science and environmental health literature suggest key linkages between voter turnout, as an indicator of community and local civic engagement capacity and environmental quality indicators[80,81]. Boyce et al. (1994, 1999) examined variations among US states using a composite index of environmental stress that incorporated 167 indicators of air and water pollution, toxic chemical releases, pesticide use, and other measures, as well as an index of state-level environmental policy related to these aspects of environmental quality[82,83]. Utilizing a cross-sectional study design, the authors found that an index of power equality that combined voter turnout, educational attainment, tax fairness, and access to Medicaid was associated with stronger environmental policies, which were, in turn, associated with less environmental stress. In separate models, greater environmental stress and power inequality were also associated with a higher infant mortality rate and a premature death rate.

---

80 Cushing L, Morello-Frosch R et al. et al. Annu. Rev. Public Health 2015. 36:193–209

81 Press, D. (1998). Local environmental policy capacity: framework for research. Nat Resources Journal, 38(1), 29-52.

82 Boyce JK. 1994. Inequality as a cause of environmental degradation. Ecol. Econ. 11:169–78

83 Boyce JK, Klemer AR, Templet PH, Willis CE. 1999. Power distribution, the environment, and public health: a state-level analysis. Ecol. Econ. 29:127–40.

APPENDIX 2

CES 3.0 Score vs. Percentile

The raw CalEnviroScreen 3.0 "scores" are simple the point total between 0-100 for each census tract computed as the sum of all the various exposure factors calculated as part of the CES 3.0 dataset. In contract, the percentile values represent percentiles (also from 0-100) of those scores. Therefore, if there are not many scores of given value range, it requires a large change in score to result in a difference in percentile. Indeed, we see this with the CES 3.0 score and percentile data, where due to the lack of high score instances, there is a very large range of scores which result in very high percentile values.

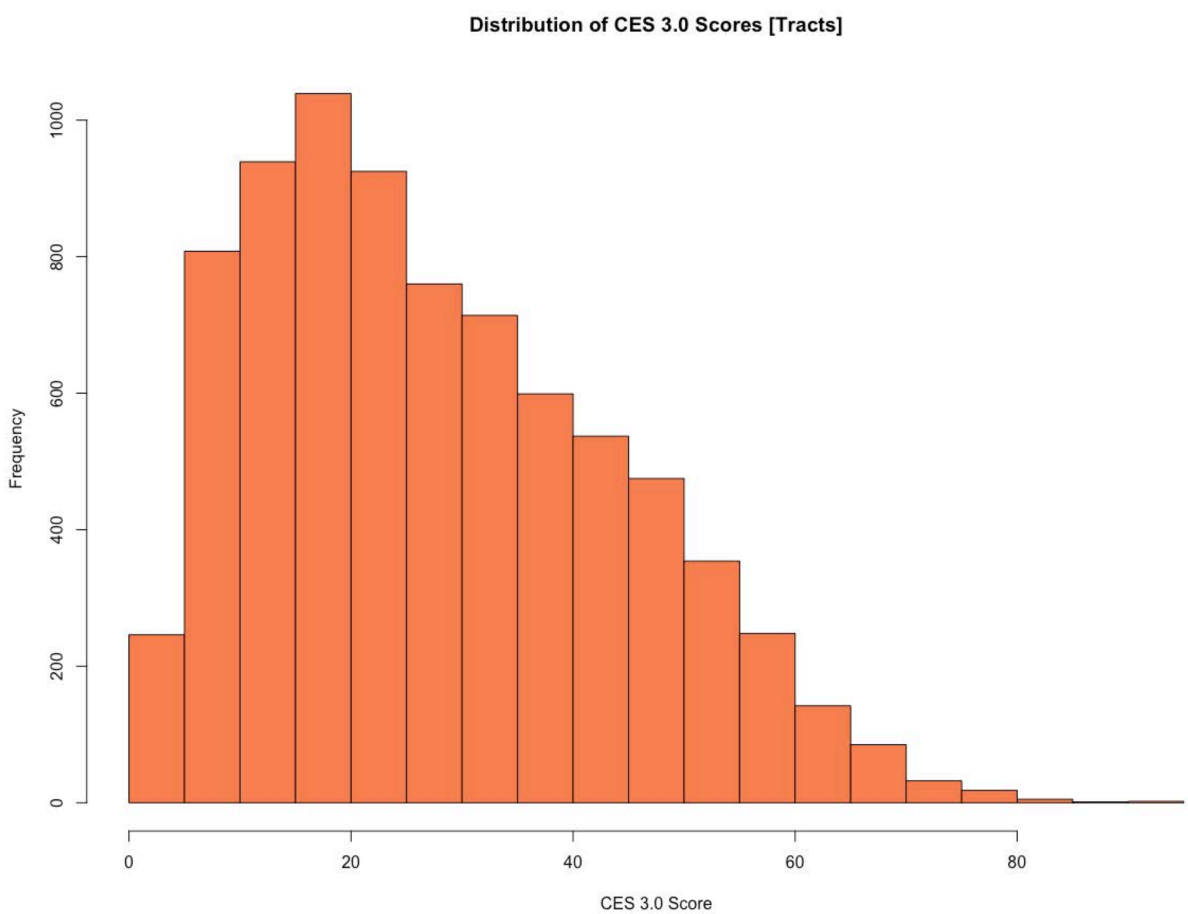


Figure A1.1. Distribution of CES 3.0 Scores

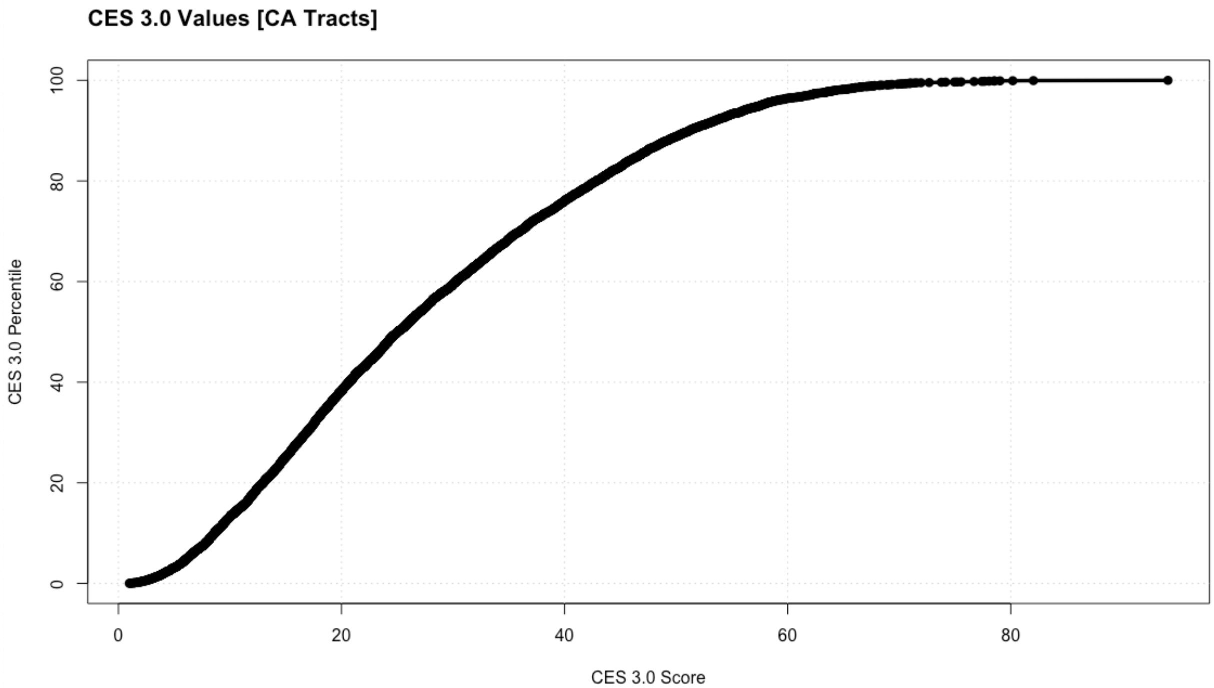


Figure A1.2. CES 3.0 Percentiles (y-axis) vs. CES 3.0 Scores (x-axis).

Residential Parcel Classifications

Classes identified as residential using the "USE_CODE_2" land classification field.

USE_CODE_2
APARTMENT HOUSE (100+ UNITS)
APARTMENT HOUSE (5+ UNITS)
APARTMENTS (GENERIC)
CLUSTER HOME (RESIDENTIAL)
COMM/OFC/RES MIXED USE
CONDOMINIUM (RESIDENTIAL)
COOPERATIVE (RESIDENTIAL)
DORMITORY, GROUP QUARTERS (RESIDENTIAL)
DUPLEX (2 UNITS, ANY COMBINATION)
FRATERNITY HOUSE, SORORITY HOUSE
GARDEN APT, COURT APT (5+ UNITS)
HIGHRISE APARTMENTS
HOMES (RETIRED; HANDICAP, REST; CONVALESCENT; NURSING)
MANUFACTURED, MODULAR, PRE-FABRICATED HOMES
MISC RESIDENTIAL IMPROVEMENT

MOBILE HOME

MOBILE HOME PARK, TRAILER PARK

MULTI-FAMILY DWELLINGS (GENERIC, ANY COMBINATION 2+)

PLANNED UNIT DEVELOPMENT (PUD) (RESIDENTIAL)

QUADRUPLEX (4 UNITS, ANY COMBINATION)

RESIDENTIAL (GENERAL) (SINGLE)

RESIDENTIAL COMMON AREA (CONDO/PUD/ETC.)

RESIDENTIAL INCOME (GENERAL) (MULTI-FAMILY)

RURAL RESIDENCE (AGRICULTURAL)

SINGLE FAMILY RESIDENTIAL

STORES & APARTMENTS

TIMESHARE (RESIDENTIAL)

TOWNHOUSE (RESIDENTIAL)

TRIPLEX (3 UNITS, ANY COMBINATION)

ZERO LOT LINE (RESIDENTIAL)

Populated Area Layer Construction Example

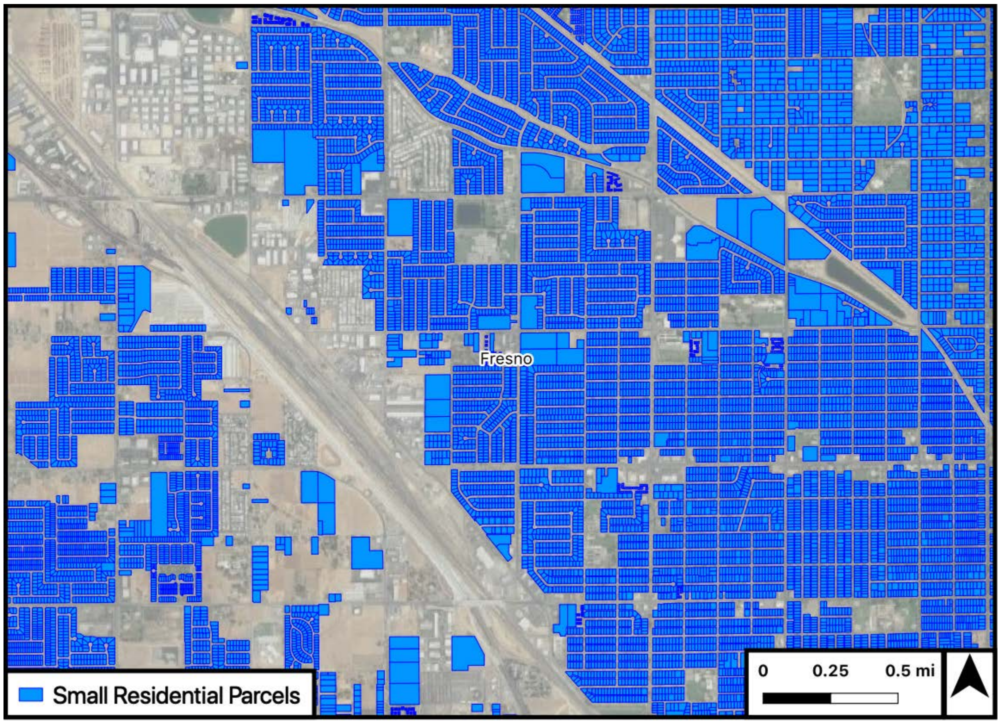


Figure A1.3. "Small" residential parcels (area < 1-acre for low-density, < 50-acre for high-density) for area in Fresno. These parcels were used in final populated area layer.

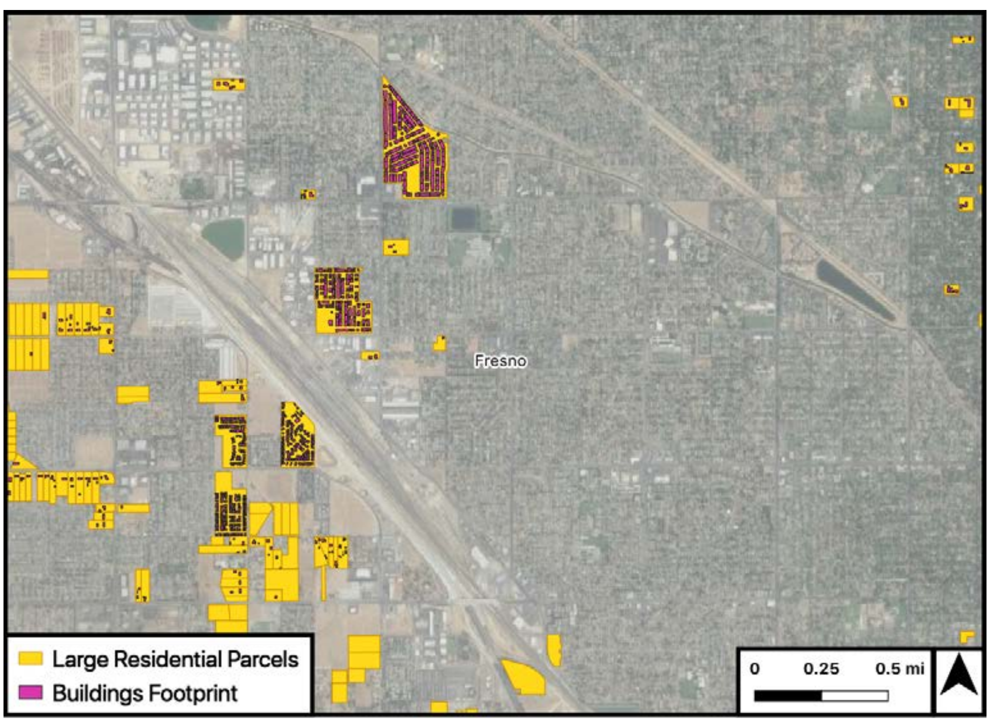


Figure A1.4. "Large" residential parcels. Intersecting buildings used in final populated areas layer.

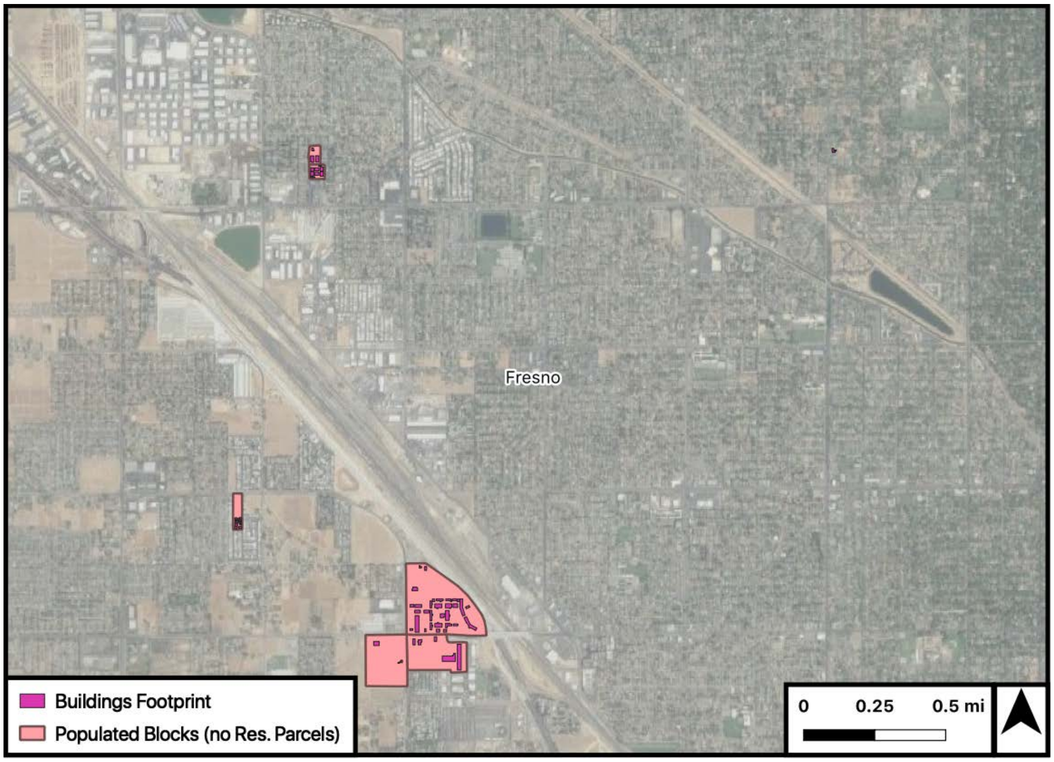


Figure A1.5. Populated blocks with no residential parcels within them. Intersecting buildings used in final populated areas layer.

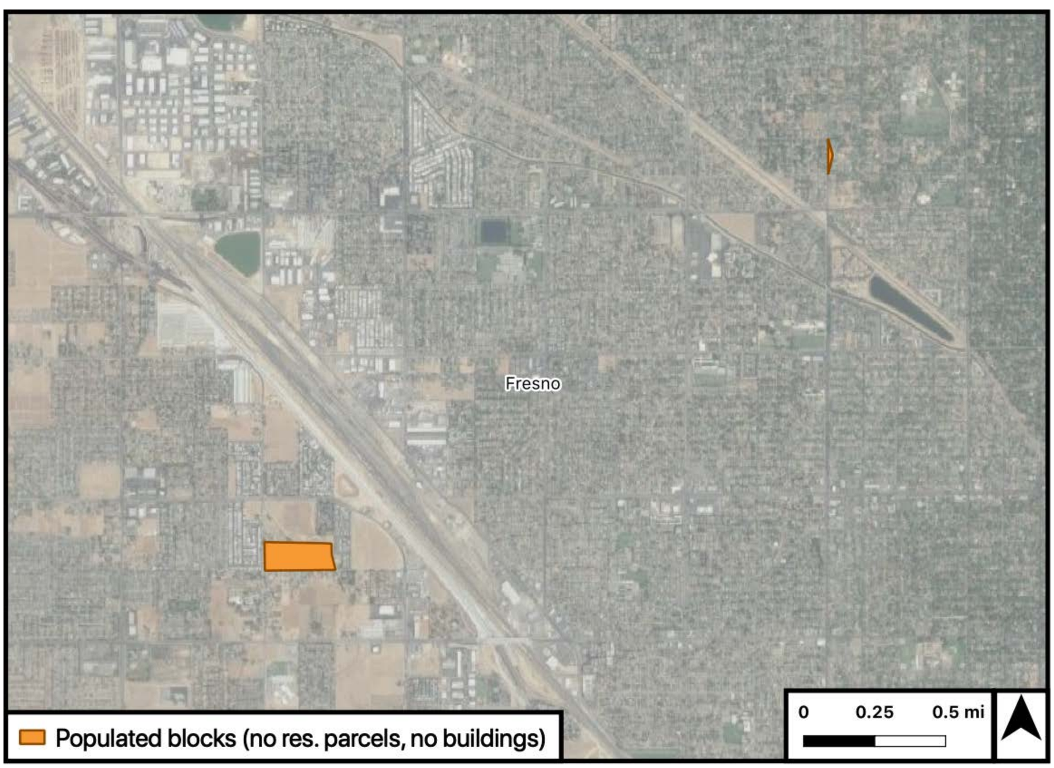


Figure A1.6. Populated blocks with no residential parcels or buildings within them. Block boundaries used in final populated areas layer.

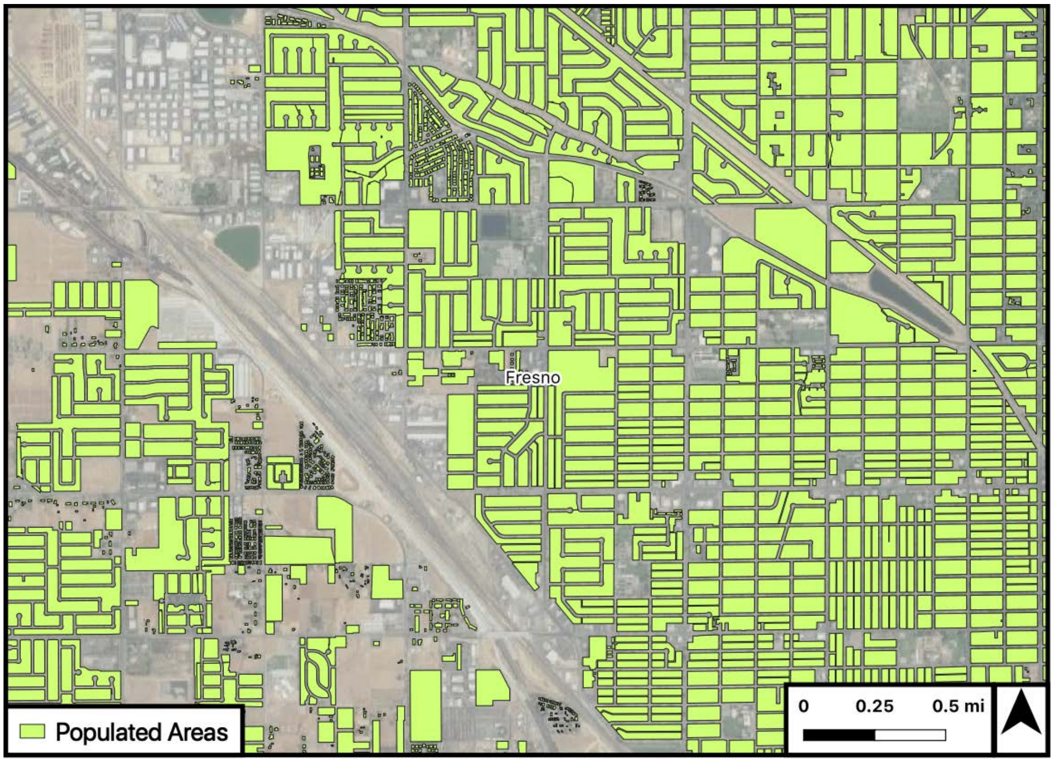


Figure A1.7. Composite of the four geometries highlighted in figures A1.3-A1.6, which makes up the final populated areas layer in this example region in Fresno.

APPENDIX 3 – Population Weighting vs. Area-Weighting

The use of population weighting for the analysis of many of the community metrics was done due to the fact that is accounts for the relative distribution of people within a given area of interest (AoA) and weights their respective characteristics accordingly. This is relevant when an AoA intersects residential areas of more than one block group (or tract when considering CES metrics).

Let's explore an example where we want to calculate the average CES 3.0 score within a given AoA. Say our AoA intersects residential portions of three different census tracts, let's call them A, B, and C. Let's pretend their characteristics are as follows:

Tract	Residential area within AoA [km^2]	Tract Residential Area [km^2]	Tract Population	CES 3.0 Score
A	0.4	1.6	30	65
B	0.4	1.2	200	40
C	0.2	1.0	750	85

Area Weighting

If we were to calculate the average CES score using a simple, area-weighted approach, we would just calculate the fraction of each tract's total residential area that is within the AoA and then use those values as our weights for the weighted-averaging process. This essentially assumes that all residential area, regardless of its parent tract's population, is treated equally when calculating average CES scores. The formula for each tract would look like this:

(Area in AoA/Total area)*CES + … + …

Putting this all together for all tracts using our hypothetical values:

(0.4/1.6)*65 + (0.4/1.2)*40 + (0.2/1.0)*85 = <u>46.6</u>

This approach is simple and does not require knowledge about each tract's population. However, area-weighting is most appropriate when the metric of concern is related to land area. For example, say you are surveying three different wheat fields, each of which has a different, uniform rate of yield. If you wanted to calculate the average yield across the area of these three fields, you would want to employ an area-weighted approach, with the weights corresponding to the total area of each field. However, when calculating metrics of vulnerability or exposure that inherently pertain to populations, it may not be sufficient to simply employ an area-weighted approach.

Population Weighting

For any metrics that relate to the residents of an area, using a simple area-weighted approach will still be using the area of land as its weights, rather than the presence or absence of people themselves, which can be misleading. Say we want to calculate the average, median income across two different census tracts, each with equal area, but one of which has 10 people and the other 100. An area-weighted approach would simply be the average of the two tracts' median incomes, which would mean that the incomes of the 10 people in the sparsely-populated tract would have equal influence as the incomes of the 100 people in the more densely-populated tract. However, if we employ a population-weighted approach, we would instead account for the populations of each tract and therefore find the true average income of all 110 people in the study area.

Using the numbers from our hypothetical example, let's recalculate the average CES score using population-weighting. The steps are as follows:

1. Assuming population is evenly distributed *within* each tract, estimate the total population within the AoA. This intermediate step is done using the area fractions of each tract within the AoA as follows:

(Area in AoA/Total area)* (Tract Population) + … + …

Total Population in AoA = (0.4/1.6)*30 + (0.4/1.2)*200 + (0.2/1.0)*750 = 224 people

2. Now that we have our total estimated population within our AoA, we can use that value as the denominator in our calculation of population weights for each tract:

(Area in AoA/Total area)* (Tract Population/Total Population in AoA)*CES + …

(0.4/1.6)*(30/224)*65 + (0.4/1.2)*(200/224)*40 + (0.2/1.0)*(750/224)*85 = <u>71.0</u>

Notice how different our population-weighted result (71.0) is from the simple, area-weighted one (46.6). This is due to the fact that the three tracts each have different population densities, and even though tract C has a smaller area of intersection with our AoA, it has far more people total, meaning that there are more people within this area of intersection, who have a higher CES score than tracts A and B. These people were being underrepresented in the simple, area-weighting scheme because it simply weighted tract C's CES score based on its area and not population.

Note above how in this population-weighting scheme we still use the areas to calculate how many people are presumed to be within each tract's area within the AoA. This is because we do not have detailed population distribution information regarding changing population densities within each tract and therefore must assume that the tract's population is spread uniformly across its residential area.

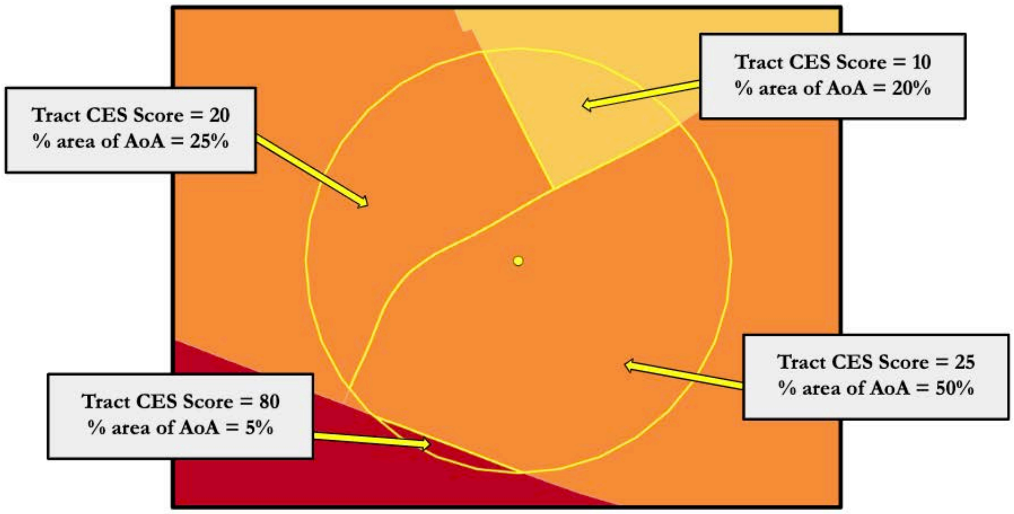
Other important clarifications:

- Population-weighting only accounts for populations *within* a given area of interest and not beyond. Therefore, this approach does not inherently under-weight AoAs in which few people reside compared to a different AoA with a higher total population. If 10 people reside in one AoA and all have a CES score of 60, the average CES score for that AoA will be 60, even if there exist other AoAs that have much higher populations within them around the state.

- It is true that population-weighting does weight more populous tracts within a given AoA, and therefore may dilute certain values of sparsely-populated tracts within the same AoA. However, this is an issue inherent to any form of averaging. If one would like to capture the highest or lowest CES in an AoA, we suggest using the Max/Min CES 3.0 metrics provided.

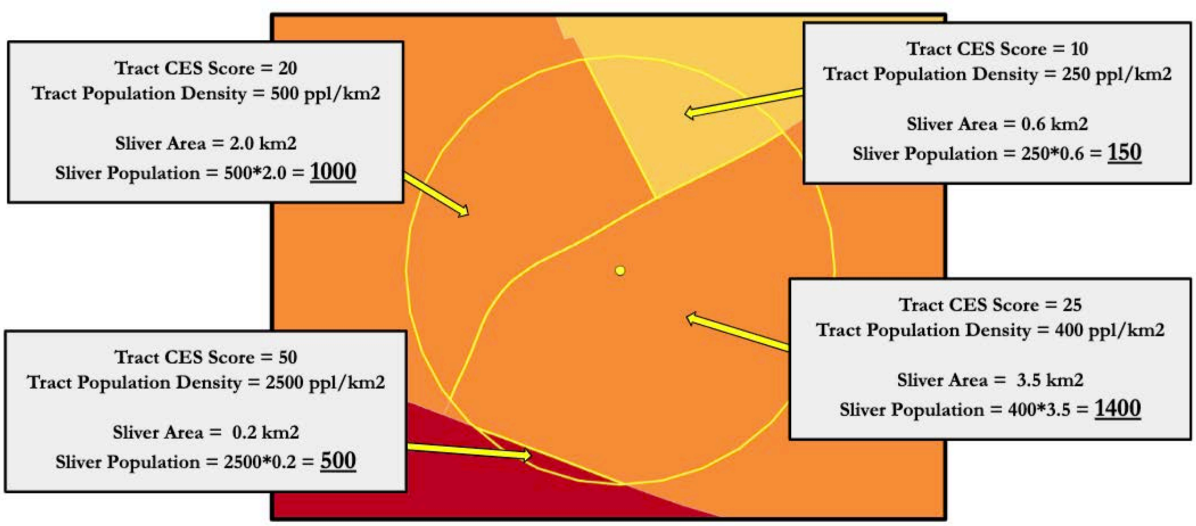
Another visual example for calculating the CES score showing a hypothetical AoA intersecting multiple census tracts is shown below:

Area-Weighting of Variables



Area-weighted CES Score = 20*(0.25) + 10*(0.20) + 25*(0.50) + 80*(0.05) = **23.5**

Population-Weighting of Variables



Total AoA Population = 1000 + 150 + 1400 + 500 **= 3050 people**

Pop.-Weighted CES Score = 20*(1000/2830) + 10*(150/2830) + 25*(1400/2830) + 80*(500/2830) = **34.1**