FINAL REPORT: CONTRACT NUMBER # 11-336

Update and Statewide Expansion of the Environmental Justice Screening Method (EJSM)

Principal Investigators:

Rachel Morello-Frosch, Ph.D., M.P.H.
University of California, Berkeley
Department of Environmental Science Policy and Management &
School of Public Health

James Sadd, Ph.D.
Occidental College
Department of Geology and Environmental Science

Manuel Pastor Jr., Ph.D.
University of Southern California
Program on Environmental and Regional Equity
Departments of Geography and American Studies and Ethnicity

Prepared For the California Air Resources Board (CARB) by:

Rachel Morello-Frosch
With
James Sadd and Manuel Pastor

Disclaimer

The statements and conclusions in this report are those of the researchers on this contract and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

Acknowledgements

This study was supported by the California Air Resources Board (ARB Contract 11-336). We wish to thank the contracts and grants staff at the University of California Berkeley, Department of Environmental Science, Policy and Management (Rachel Morello-Frosch), the University of Southern California, Program for Environmental and Regional Equity (PERE) (Manuel Pastor), and Occidental College (James Sadd) for pre- and post-award support. We also thank Cynthia Garcia, our project manager, and Dr. Álvaro Alvarado at CARB, both of whom assisted us in acquiring critical data for this project; provided invaluable feedback on our work throughout the contract period; facilitated meetings with scientists and staff at CARB and Cal-EPA to enable us to present our preliminary results; attended our stakeholder meeting in Los Angeles; and hosted a second stakeholder meeting/webinar in Sacramento toward the end of this project. Bill Jesdale, Lara Cushing, and Carolina Balazs from UC Berkeley and UC Davis, along with Madeline Wander, Justin Scoggins, Pamela Stephens, Erika Petroy, and Vanessa Carter from USC PERE provided invaluable collaboration in the processing, cleaning and analysis of all of the data. Finally, we want to thank the California environmental justice advocates, too many to enumerate, who have worked with us on this project and helped make our scientific work better. They participated in phone calls and meetings where we previewed study results, provided excellent feedback to improve our screening method, and collaborated on “ground-truthing” efforts in Southern California and the Central Valley. Most importantly, they ensured that we leveraged our scientific knowledge in the realms of environmental health, epidemiology, social science, economics and GIS to develop a screening tool that spatially characterizes the cumulative impacts of multiple environmental hazards and social stressors, with an eye toward advancing environmental justice in regulatory decision-making and the policy arena.

Table of **Contents**

1 List of Tables and Figures 4
1.1 Main Tables and Figures 4
1.2 Supplemental Tables and Figures 5
2 Abstract 7
3 Introduction 8
4 Methods 9
4.1 Development of Land Use Base Map 9
4.2 Hazard Proximity and Sensitive Land Use Layer Updates 10
4.3 Health Risk and Exposure Layer Updates: 14
4.4 Social and Health Vulnerability Layer Updates: 15
4.5 New Climate Change Vulnerability Layer 16
4.6 New Drinking Water Layer 18
5 Results 21
5.1 Comparison with Other Screening Approaches 39
6 Additional Activities 40
7 Conclusion 40

1 List of Tables and Figures

1.1 Main Tables and Figures

Table 1. Environmental Hazard and Sensitive Land Use Indicators

Table 2. Exposure and Health Risk Indicators

Table 3. Health and Social Vulnerability Indicators

Table 4. Climate Change Vulnerability Indicators

Table 5: Derivation and Scoring of Drinking Water Quality Indicators

Table 6: Scoring Approaches in EJSM, CES and CEVA

Figure 1A: Application of Point Distance Tool to Develop Hazard Proximity Indicators

Figure 1: Different CI polygon sizes complicate application of Point Distance Tool

Figure 1C: Point Distance Tool Compared to Buffer Approach

Figure 1D: Aggregating from Block to Tract using Population Weights

Figure 2. Regions Analyzed for the EJSM

Figure 3A-F: SF Bay Area EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 4A-D: SF Bay Area Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 5A-F: Southern CA EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 6A-D: Southern California Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 7A-F: SANDAG EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 8A-D: SANDAG Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 9A-F: San Joaquin Valley EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 10A-D: San Joaquin Valley Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 11A-F: SACOG EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 12A-D: SACOG Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 13A-F: Central Coast EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 14A-D: Central Coast Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 15A-F: Eastern Sierra EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 16A-D: Eastern Sierra Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 17A-F: Northern California EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

Figure 18A-D: Northern California Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

Figure 19A-D: California Impact Score -- Statewide

1.2 Supplemental Tables and Figures

Table SB.1. CARB FOI facilities in the San Joaquin Valley found by ground-truthing to have been mis-located by at least 10 kilometers

Table SB.2. Locational error for California Department of Toxic Substances Control permitted hazardous waste handling facilities and generators located in the San Joaquin Valley

Figure SA.1. CI polygons derived from SCAG 2008 land use geospatial data layer for a portion of Riverside County

Figure SA 2: NLCD land use rasters corresponding to classes 22, 23, and 24

Figure SA.3: EJSM total score for Los Angeles County calculated using CI Polygons constructed using SCAG 2008 land use geospatial data

Figure SA4: EJSM total score for Los Angeles County calculated using CI Polygons constructed using NLCD 2011

Figure SA5: Net change EJSM total score calculated using NLCD land use classes

Figure SA6: EJSM total score calculated for Riverside County using CI Polygons constructed using SCAG 2008 land use geospatial data

Figure SA.7: EJSM total score calculated for Riverside County using CI Polygons constructed using NLCD 2011

Figure SA.8: Net change EJSM total score calculated using NLCD land use classes

Figure SB.1: CARB “Facilities of Interest” (FOI) locations in the San Joaquin Valley, showing location corrections

Figure SB.2: EJSM total score for San Joaquin Valley prior to ground truth correction

Figure SB3: EJSM total score for San Joaquin Valley after ground truth correction

2 Abstract

Environmental health researchers, practitioners, advocates, and policy-makers are increasingly concerned about the origins and persistence of health disparities in California. Scientific research indicates that the inequitable distribution of health is linked to environmental and social conditions combined with underlying vulnerability factors that put people at "risk of risks" (Phelan et al., 2010). This combination of environmental hazard exposures and socioeconomic stressors has been described as a form of double jeopardy (Institute of Medicine, 1999) that disproportionately impacts vulnerable groups, particularly communities of color and the poor, immigrants and linguistically isolated groups. Although the importance of cumulative impacts may be theoretically obvious, the task of measuring and quantifying these impacts is challenging because data on interactions among these exposures are unavailable, information on place- and population-specific exposures is lacking, and validated models relating exposure to health effects for multiple chemicals and nonchemical stressors do not exist (Sexton, 2012). However, spatial screening allows decision-makers to identify areas that are over-burdened with environmental hazards and that are socially vulnerable so that communities might be targeted for regulatory and policy action to improve environmental conditions and protect public health. The Environmental Justice Screening Method (EJSM) facilitates such mapping of cumulative impacts using multiple health, environmental and social vulnerability measures organized along diverse categories. This project extended the original EJSM to create additional metrics, including indicators of climate change vulnerability, and increased cumulative impacts screening coverage from its initial focus on Southern California and the San Francisco Bay Area, to all California regions. In addition, we integrated our work with OEHHA on drinking water into our final maps, which are presented here. Methodological improvements include data updates, corrections of facility and sensitive land use locations in data provided by CARB, a more streamlined method for developing the land use base maps, and an approach to assessing area-level proximity to environmental hazards that reduces data processing time and enhances flexibility for implementing different buffers and scoring approaches.

We present maps of the EJSM scores for eight California regions, including intermediate scores for each category of cumulative impact (e.g. hazard proximity and sensitive use; health risk and exposure; social and health vulnerability; climate change vulnerability; drinking water quality) and total cumulative impact. Comparisons of cumulative impact scores derived using regional versus state quintile distributions indicate that different geographic definitions for deriving scores affect screening results. Although the EJSM is flexible enough to allow for comparisons of cumulative impact scores across different study areas (e.g., within regions or across the state) we favor a regional application of scoring because generally land use planning, industrial and transportation development, and environmental regulation are regionally rooted and require regionally specific interventions to reduce hazard exposures or to address social and health vulnerability factors. In addition, statewide scoring can mask important within-region inequities. Nevertheless, there may be certain policy and decision-making contexts when statewide distributions for scoring may be appropriate for larger scale impacts such as climate change vulnerability.

Spatial screening methods such as the EJSM are critical tools that can help decision-makers advance environmental justice goals by more efficiently targeting efforts and resources to remediate cumulative impacts, environmental inequities, and focus regulatory action at the neighborhood level. As environmental health science develops a better understanding of cumulative impacts, standard approaches in risk assessment may need to change and be harmonized with cumulative impact screening methods to assure the protection of public health.

3 Introduction

Although environmental regulations for pollutants in air, water, soil, food, and other sources have been effective in controlling community exposures to some environmental hazards, they do adequately address multiple pollutants from diverse sources or incorporate nonchemical stressors and health vulnerabilities. Health disparities that disproportionately affect minority and low-income populations may enhance the effects of environmental chemicals. Cumulative exposures to environmental stressors, against a background of vulnerability, can result in heightened health risks and impacts across a population. Although the importance of cumulative impacts may be theoretically obvious, the task of measuring and quantifying these impacts is enormously challenging; indeed, quantifying cumulative risk from chemical and nonchemical stressors is impractical or impossible in most real-world situations because data on interactions among these exposures are unavailable, information on place- and population-specific exposures is lacking, and validated models relating exposure to effect for multiple chemicals and nonchemical stressors do not exist (Sexton, 2012). Therefore, a key challenge for decision-makers is to reconcile these data gaps with the pressing need to proactively address and reduce potential cumulative impacts and environmental health disparities among diverse communities.

Spatial screening allows decision-makers to identify areas that are over-burdened with environmental hazards and that are socially vulnerable so that communities might be targeted for regulatory and policy action to improve environmental conditions and protect public health. The key to this approach is the use of geographic information systems (GIS) mapping to integrate chemical and nonchemical stressors, vulnerability, and background risk factors in a semi-quantitative manner. The Environmental Justice Screening Method (EJSM) facilitates such mapping and spatial screening of cumulative impacts using multiple health, environmental and social vulnerability measures organized along diverse categories. During its first iteration, the EJSM included three main categories of metrics: (1) hazard proximity and land use; (2) estimated air pollution exposure and health risk; (3) social and health vulnerability (Sadd et al., 2014, 2011). At the request of the California Air Resources Board (CARB), this project sought to extend the original EJSM to create additional metrics, including indicators of climate change vulnerability, and to extend coverage from its initial focus on Southern California and the San Francisco Bay Area, to all California regions. CARB also requested that two community workshops be conducted to provide opportunities to review results with community residents and environmental justice advocacy and public health organizations and to collect feedback for inclusion in the final project report. It was also requested that the Project manager be trained concurrently with the researchers in EJSM data preparation, analysis and scoring, to monitor and help beta test the different modules of the EJSM as they were developed. CARB also requested that we conduct a final presentation/workshop at CARB in Sacramento for interested agency staff, similar to the one at the conclusion of the original 2010 project. Finally, CARB asked that transfer of the data and programming used in this project to CARB take place as the data become completed and finalized.

This report describes how we completed project deliverables, including how we addressed logistical and methodological challenges along the way; we also discuss additional products from our work that were beyond the scope of our contract, and present the final results of our EJSM analysis. We conclude with a short discussion of the implications this work, and potential synergies with emerging initiatives within CARB and Cal-EPA more broadly.

4 Methods

4.1 Development of Land Use Base Map

The EJSM involved a four-step process: (a) conduct an initial GIS spatial assessment to create a detailed land use base map to isolate the land use types (i.e. residential and "sensitive" uses) used in the analysis to follow; (b) apply GIS techniques to appropriately summarize the resulting environmental hazard proximity indicators for all census tracts; (c) integration of the resulting tract level scores with tract level data on other indicators of environmental and social stressors, including estimated air pollution exposure and/or health risk, climate change vulnerability, drinking water quality, and social and health vulnerability; (d) a cumulative ranking based on all the tract-level indicators that is then presented visually.

The first step in this process entailed acquiring parcel level spatial data files with accompanying land use attribute information for each county in California. This information was used to identify cumulative impact (CI) land uses as a step in constructing the CI polygons for EJSM scoring. The base map for the state is constructed by integrating specified residential and sensitive land use classes as classified by CARB (CARB, 2005) This approach focuses cumulative impact screening on areas with land uses where people reside or locations hosting schools, hospitals, day care centers, parks and other sensitive receptor locations. Areas that are, for example, strictly industrial or commercial or undeveloped open space are not included in the base map.

For most counties, land use code attribute information was acquired at the parcel level. However, for some counties, such as Mariposa, the data was not available or had inadequate use code information for many individual parcels. To address this problem, we developed an interim CI land use spatial data layer using the rough descriptions available, and then supplemented this information with NLCD data to further correct land use for these areas. Similarly, some counties used inconsistent definitions of use code. For example, in Merced and Stanislaus counties, parcels composed primarily of agricultural land but with a small area with residential buildings were classified as “rural residential” in the same way as small parcels in rural areas dominated by residential structures. In order to avoid over-representing farm fields as residential land, we use the NLCD to identify and filter out land that did not fit our residential criteria. Finally, some parcel data had use code detail available in which case we utilized NLCD and regional government land use information (e.g. from the Sacramento Area Council of Governments (SACOG)) to make these interpretations better and more accurate. The extra processing done for these data refinements in the development of the land use layer and CI polygons is codified in our ModelBuilder programming.

We also validated the use of NLCD to replace our more complex and integrative approach of using county parcel data for generating our land use base maps for the EJSM. The NLCD is the highest spatial resolution (30M grid cells) land cover data available and it is updated by the US Geological Survey every six years. The results of this validation process comparing approaches to derive the base land use maps are shown in **Supplement A**.

To geographically link the land use base map with tract-level metrics of environmental and social/health vulnerability, the residential and sensitive land use polygons were intersected using a GIS procedure with census block polygons from the 2010 Census, to create a base map composed of neighborhood-sized cumulative impact (CI) polygons, each with a known land use class and attribute key to attach census information.

4.2 Hazard Proximity and Sensitive Land Use Layer Updates

The first step required to score the proximity of hazards to analysis involved attaching to each of the

Table 1: Environmental Hazard and Sensitive Land Use Indicators

Indicator	Description	Data Source
Environmental Hazard		
Large industrial facilities	Includes facilities emitting >25,000 metric tons of CO2-equivalent (CO2e) and facilities emitting > 10 tons of particulate matter and/or toxic air contaminants per year.	Provided by CARB: Greenhouse Gas Inventory Database CA Emission Inventory Development and Reporting System (CEIDARS)
Small area source emitters	Includes auto body and paint shops, gas stations, permitted hazardous waste facilities	Provided by CARB: California Dept. of Toxic Substances Control (auto paint and body shops, and hazardous waste); County Departments of Weights and Measures (gas stations)
Large area emitters	Railroads and railyards, airports, intermodal distribution facilities	Facilities of Interest Database provided by CARB
Traffic Volume	Major roadways and freeways	CA DPH-EHIB
Sensitive Land Use		
Schools	Public and private pre-college schools and related eduational facilities	A Dept of Education 2011; SCAG 2008; ABAG 2005; SANDAG 2009 SACOG 2014; county tax parcel use codes 2010
Childcare facilities	Licensed childcare and daycare facilities	County tax parcel use codes 2010, 2011; State of California Licensing Division, 2011; Dun and Bradstreet by NAICS code 624410, 2011
Playgrounds, parks and recreation centers	Improved parks and playgrounds with grounds and facilities used in children's sports, and other outdoor play activities; does not include open space, hiking or biking paths, or similar unimproved parkland.	Regional planning agencies (SCAG 2008; Association of Bay Area Governments (ABAG) 2005; San Diego Association of Governments (SANDAG) 2009; Sacramento Area Council of Governments (SACOG) 2014); geocoded locations from addresses from Dun and Bradstreet by NAICS codes 621491 and 524114, 2009: California Spatial Information Library 2010; county tax parcel use codes 2010
Residential neighborhoods	Areas described as in use or zoned for residential uses, ranging from low density single family detached, to higher density multifamily, to mobile home parks, etc.	Regional planning agencies (SCAG 2008; Association of Bay Area Governments (ABAG) 2005; San Diego Association of Governments (SANDAG) 2009; Sacramento Area Council of Governments (SACOG) 2014); geocoded locations from addresses from Dun and Bradstreet by NAICS codes 621491 and 524114, 2009: California Spatial Information Library 2010; county tax parcel use codes 2010
Healthcare or senior housing facility	Residential care and living facilities dedicated or limited to serving the elderly (age >64).	Regional planning agencies (SCAG 2008; Association of Bay Area Governments (ABAG) 2005; San Diego Association of Governments (SANDAG) 2009; Sacramento Area Council of Governments (SACOG) 2014); geocoded locations from addresses from Dun and Bradstreet by NAICS codes 621491 and 524114, 2009: California Spatial Information Library 2010

CI polygons on our regional base map a set of hazard proximity indicators and then summarizing these to create scores at the tract level.

Table 1 shows the hazard and sensitive land use indicators and their data sources that we used to derive the environmental hazards and sensitive land use scores. Locational data were verified and corrected as needed using a tiered approach, which is described in detail in **Supplement B**.

Each CI polygon—consisting of either a residential or sensitive land use—was scored as follows: We applied the ArcGIS Point Distance procedure, which measures the distance between two points (in this case CI polygon centroids and the point hazard location), rather than a buffer distance approach, which measures the distance between a point and a polygon boundary (used in our first iteration of the EJSM and described in detail elsewhere) (Sadd et al., 2011). This newer approach has several advantages over the old one. The Point Distance method significantly reduce data processing time because it runs quickly and is done only once, and provides much greater flexibility to score using any conceivable buffer distance and strategy, which helps facilitate sensitivity analyses. For purposes of the EJSM, we currently use a 3000 foot threshold, which works well for smaller CI polygons. However, in less urbanized locations and in places with large areas of a single land use, CI polygons tend to be much larger, making the Point Distance tool less effective. To remedy this, we divided these larger CI polygons into smaller ones using a 1000 foot statewide grid (or “fishnet”) which is intersected with the large CI polygons. We then used the centroid to generate a table specifying the distance between the CI polygon centroid and point hazard. To address the different areas of each CI polygon we added the radius of a circle with the equivalent area to the CI polygon to each of the distance bands, which enables the identification of facilities or other emission sources within 1000, 2000, and 3000 feet of a CI polygon. **Figures 1A-C** provide a visual example of this process.

Figure 1A. Application of Point Distance Tool to Develop Hazard Proximity Indicators

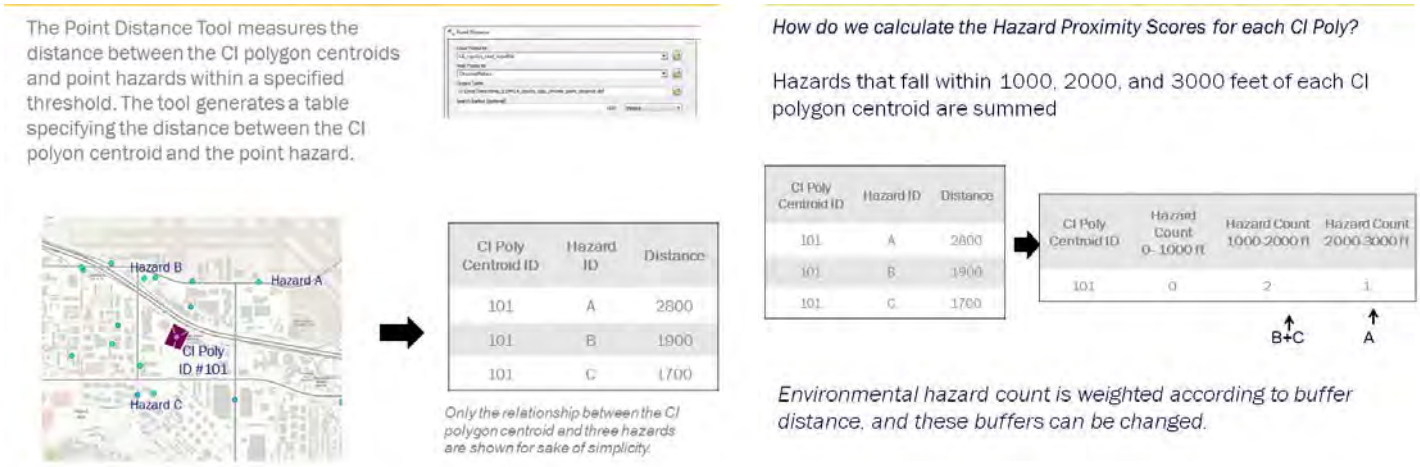


Figure 1B. Different CI polygon sizes complicate application of Point Distance Tool

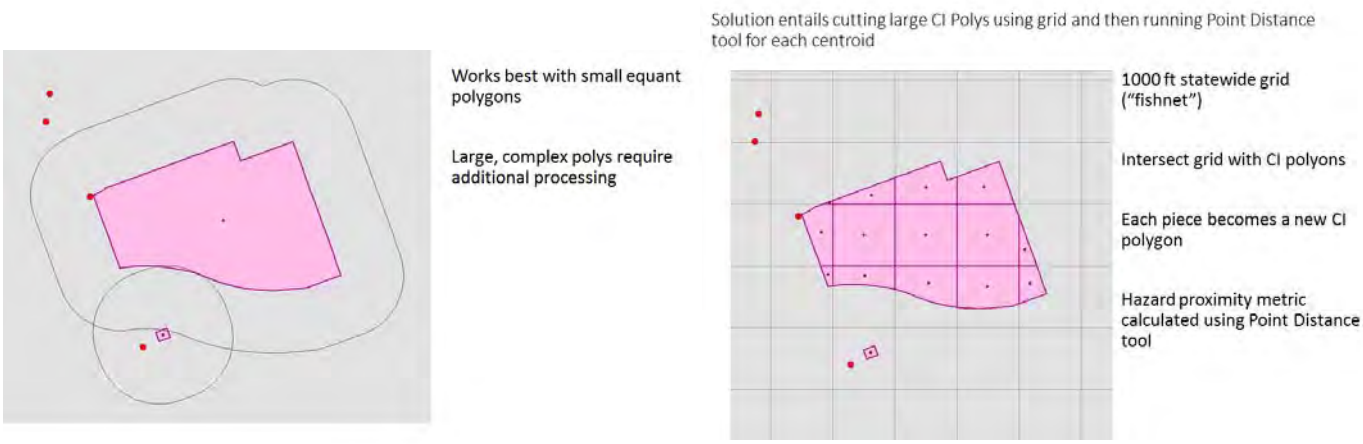


Figure 1C. Point Distance Tool Compared to Buffer Approach

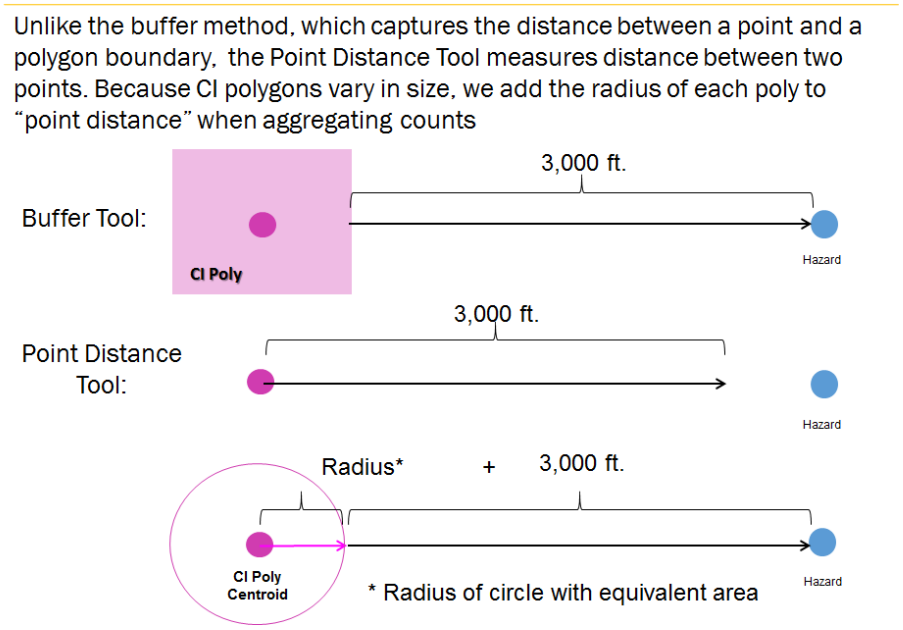
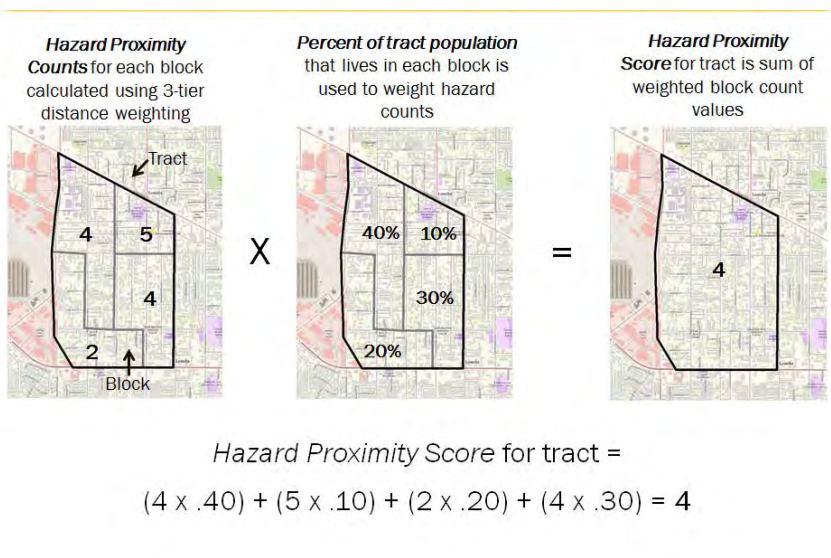


Figure 1D. Aggregating from Block to Tract using Population Weights



We evaluated this method in several counties and decided to also add the step of eliminating any new polygons (after the intersect) with areas < 100 square feet, because the spatial precision of the hazard and sensitive land use data are not adequate enough to warrant keeping these small polygons, and such a small area is not relevant from a land use perspective. Overall, the combination of these point-distance and fishnet techniques ran very quickly (about 3 hours to process) for the statewide data layer and generated approximately 4 million CI polygons. This GIS approach is also advantageously nimble in terms of facilitating sensitivity analyses that allow for changes in distance bands and integration of new data, where appropriate.

The number and type of hazardous land uses (represented as point or area features, such as airports and railroad tracks) were calculated for every CI polygon; we then applied a distance-weighted scoring procedure where the influence of the hazards on the sum attached to the CI polygon diminishes with distance as those areas with closer proximity to numerous air quality hazards are assumed to be more highly impacted. The distance-weighted hazard count for each CI polygon was derived with the following formula:

$$(W_{<1000ft} * H_{<1000ft}) + (W_{1000\text{-}2000ft} * H_{1000\text{-}2000ft}) + (W_{>2000\text{-}3000ft} * H_{>2000ft\text{-}3000ft})$$

Where H is the hazard, and W is the distance weight (weights = 1 for <1000 ft; 0.5 for 1000-2000 ft; and 0.1 for >2000-3000 ft).

We added to the distance-weighted hazard proximity counts a binary dummy variable indicating whether the CI polygon was residential land or a non-residential sensitive land use CI polygon. Sensitive land uses included, CI polygons with a school, playground childcare center, park or health care facility. A tract-level hazard proximity score was then calculated by combining the hazard proximity and sensitive land use measures and then attaching to each CI polygon a population weight derived from assigning population using the underlying intersection of census block data and polygon land area. We then weighted the scores to derive a census tract average count for hazard proximity/sensitive land uses. (**Figure 1D**). Finally, a quintile ranking from 1 (low) to 5 (high) was applied to derive a tract-level score which integrates the presence of both sensitive and hazardous land uses.

4.3 Health Risk and Exposure Layer Updates:

Similar to the first iteration of the EJSM, this category includes updated metrics of the ambient air pollution concentrations and health risk indicators associated with modeled TRI emissions and air toxics exposures all calculated at the census tract level. Specifically we integrated the following indicators (**Table 2**); 1) toxicity weighted hazard scores for air pollutant emissions averaged for 2007-2010 from Toxic Release Inventory facilities included in the U.S. EPA's Risk Screening Environmental Indicators (RSEI), estimated at the census tract level using a Gaussian-plume fate-and-transport model (US EPA, NO); 2) estimated cumulative estimated lifetime cancer risk associated with ambient air toxics exposures from mobile and stationary sources for 2005 derived by integrating US EPA's National Air Toxics Assessment (NATA) data and cancer potency values from California's Office of Environmental Health Hazard Assessment (Cal-EPA, ND; US EPA, ND); tract-level estimates of cumulative respiratory hazard derived from the 2005 National Air Toxics Assessment (NATA) and OEHHA Reference Exposure Levels (RELs)(Cal-EPA, ND); tract-level ambient concentration estimates interpolated from the CARB statewide criteria air pollutant monitoring network for PM2.5 averaged for 2009–2011 and ozone exceedances, defined as sum of the portion of the daily maximum 8 hour concentration over the California standard of 0.070 ppm averaged over 2009-11.

For this new EJSM iteration, we also added an indicator for agricultural pesticide use in pounds per square meter, averaged for 2009-2011.

Table 2: Exposure and Health Risk Indicators

Indicator	Description	Data Source	Year
Toxicity-weighted air pollution concentrations from large industrial facilities (TRI)	Includes facilities required to report industrial emissions to EPA's Toxic Release inventory. Estimates are averaged across 4 years.	US EPA's Risk Screening Environmental Indicators (RSEI)	2007-10
$PM_{2.5}$ concentrations	Annual average concentration estimates interpolated to tract level from CARB air monitors. Averaged over three years	CARB	2009-11
Ozone exceedances	Annual sum of the portion of the daily max 8 hour concentration over CA standard of 0.070 ppm averaged over three years	CARB	2009-11
Lifetime cancer risk from air toxics	Estimates cancer risks associated with modeled estimates of ambient air toxics categorized as known, probable or possible carcinogens	US EPA's National Air Toxics Assessment Cancer Potencies from OEHHA	NATA 2005 OEHHA 2012
Chronic respiratory hazard from air toxics	Estimates respiratory hazard associated with modeled estimates of ambient air toxics based on a hazard ratio derived using RELs	US EPA's National Air Toxics Assessment RELs from OEHHA	NATA 2005 OEHHA 2012
Pesticide Use	Average annual number of pounds of pesticide applied in pounds/m^2	Pesticide Use Reporting System, compiled by the Environmental Health Investigations Branch of CDPH	2009-11

Intermediate scores for each health risk and exposure metric were calculated based on their quintile distribution rankings (with scores ranging from 1–5) for all tracts in a study area (state or region). As these health risk and exposure metrics are at the tract level, each CI polygon receives the metric score for its host census tract and the ranking is done at the tract level. For example, a tract in the least impacted 20% for each of the six exposure and health risk metrics (PM2.5 concentration, ozone exceedance, estimated cumulative cancer risk and respiratory hazard for air toxics, toxicity-weighted pollutant emissions from RSEI, and pesticide use) would receive a total health risk and exposure score of 6 (6 metric scores of 1), whereas a tract that ranked in the highest quintile for all six metrics would have a total exposure and health risk score of 30 (6 metric scores of 5). These total intermediate scores are then re-ranked into quintiles by tract to derive the final score for this air pollution exposure/health risk category, which ranges from 1 to 5.

4.4 Social and Health Vulnerability Layer Updates:

We updated all of the tract level demographic metrics using five year estimate of the American Community for the years 2008-2012. We also expanded the voter turnout metric to include an average percent of the votes cast among all registered voters averaged for the 2004, 2006, 2008, and 2010 general elections (**Table 3**).

Intermediate social and health vulnerability indicator scores were calculated using the same quintile distribution and normalization technique employed for the health risk and exposure indicators, above, with scores ranging from 1 to 5. To ensure that social and health vulnerability scores were not distorted by missing data or based upon anomalously small populations, tracts with fewer than 50 people and those with fewer than six indicator values were not scored. Some of these tracts had already been eliminated in the hazard proximity scoring phase owing to having no residential land. To insure comparability between tracts with all metrics and those tracts missing 1 to 4 metrics, we summarized the ranks for the individual metrics and then calculated a score based on dividing that sum by the number of non-missing metrics.

Table 3: Health and Social Vulnerability Indicators

Indicator	Description	Data Source	Year
Social Vulnerability Indicators			
Racial/ethnic make-up	% residents of color	American Community Survey (ACS)	2008-12
Poverty rate	% residents living below twice the national poverty level	ACS	2008-12
Homeownership	% of residents living in rented households	ACS	2008-12
Housing Value	median housing value	ACS	2008-12
Potential Biological Vulnerability Indicators			
Age of residents	% residents <5 years old	ACS	2008-12
	% of residents > 65 years old	ACS	2008-12
Birth outcomes	% of preterm and small-for-gestational age infants	California Dept of Public Health	2006-2012
Civic Engagement Capacity Indicators			
Linguistic isolation	% pop. >age 4 in households where no one >age 15 speaks English well	ACS	2008-12
Voter Turnout	% votes cast among all registered voters averaged for 2004, 2006, 2008, 2010 general elections	California Secretary of State (CA Secretary of State, ND)	2004, 2006, 2008, 2010

4.5 New Climate Change Vulnerability Layer

For this new iteration of the EJSM, we developed a novel category of metrics to address the issue of climate vulnerability. Research, including our own, indicate that climate change is having a disproportionate impact on the health of poor communities and communities of color in the US (Shonkoff et al., 2011; Jesdale et al., 2013; English et al., 2013). At the same time, industrial and vehicular sources of the greenhouse gases (GHGs) and other climate-forcing pollutants are also disproportionately located in low income communities of color. Mitigation efforts to combat climate change in California have the potential to deliver substantial co-benefits to the health of disadvantaged communities by reducing the hazardous air pollutants that are emitted from these sources during the combustion of fossil fuels. Other types of mitigation projects, some of which are funded through the Greenhouse Gas Reduction Fund, and that include initiatives related to urban planning and forestry, public transportation, household energy efficiency or renewable energy

generation, can also offer co-benefits to low income households in the forms of neighborhood greenspace, employment and savings on energy expenditures. However, very few analytic frameworks exist for integrating equity into the design and assessment of climate policy in order to capture the co-benefits of better environmental quality in disproportionately impacted communities. The integration of climate change vulnerability metrics into the EJSM can support efforts to systematically highlight those communities in the state that are likely to benefit most from adaptation and mitigation efforts going forward. Moreover, adding the climate change vulnerability category to the EJSM provides a foundational analytic framework for understanding of the scope and scale of changes in health equity that result from climate change mitigation polices and track the extent to which mitigation efforts maximize health co-benefits for the most vulnerable populations. Accordingly we integrated indicators of heat island risk, projected changes in temperature, and additional measures of social isolation and lack of mobility into the climate change vulnerability category (**Table 4**). Intermediate climate change vulnerability indicator scores were calculated using the same quintile distribution and normalization technique described above, with scores ranging from 1 to 5.

Given the health risks associated with wildfires and their increasing frequency and intensity in California (Holstius et al., 2012), we also examined the viability of integrating using CalFire data to

Table 4: Climate Change Vulnerability Indicators

Indicator	Description	Data Source	Year
Heat Island Risk Indicators			
Tree Canopy	% of area covered by tree canopy	NLCD	2012
Impervious Surface	% of area covered by impervious surface	NLCD	2012
Temperature change indicators			
Projected maximum temperature	Projected max monthly temperature average 2050-2059	National Center for Atmospheric Research, downscaled Community Climate System Model, scenario B1, ensemble average & Cal ADAPT	2011
Projected change in maximum monthly temperature	Change in projected max monthly temperature (2050-2059) – (2000-2009)	National Center for Atmospheric Research, downscaled Community Climate System Model, scenario B1, ensemble average & Cal ADAPT	2011
Projected change in warm nights	Change in degree-days of warm nights (19°C)	National Center for Atmospheric Research, downscaled Community Climate System Model, scenario B1, ensemble average & Cal ADAPT	2011
Social Isolation Indicators			
Isolated elderly residents	% of elderly residents living alone	ACS	2008-12
Mobility	% of residents owning a car	ACS	2008-12

develop a metric of wildfire risk. However, the CalFire dataset had severe limitations, including low spatial resolution (1/8°), little variation from one spatial data point to another, and no definition of risk

for urban areas or cultivated lands. Moreover, local risk of fire is a poor proxy for the actual effects on populations due to widespread exposures to air pollution from wildfire smoke. Therefore, we chose not to add this metric into the climate vulnerability category. If we are able to locate improved modeled estimates that better integrate air quality impacts at a finer geographic resolution and with a broader geographic scope that includes urban areas, we will likely add a wildfire risk metric in the future.

4.6 New Drinking Water Layer

Communities struggling with access to safe drinking water in the US typically face a composite burden related to both exposure to contamination and the ability to cope adequately. This burden is based, among other things, on vulnerability of the physical infrastructure and weak managerial and financial capacity (Balazs and Ray, 2014). Assessing the cumulative impacts of these burdens is critical from a public health perspective. Based on our previous empirical research (Balazs et al., 2011, 2012), an assessment of what data sources are readily available, and consultation with environmental justice groups working on drinking water, we created a composite drinking water quality metric that includes: Potential exposure to high levels of contaminants, and compliance with monitoring standards, and/or degree of missing data. Exposure to contaminants captures the potential health risks from drinking contaminated water, particularly at levels that exceed the maximum contaminant levels (MCLs) for 17 compounds. Monitoring compliance captures the extent to which the system meets minimum monitoring requirements of the Safe Drinking Water Act and/or lacks data on specific contaminants. We also integrated visual indicators in our water quality maps (using hatch marks) of Technical, Managerial and Financial (TMF) Capacity of Community Water Systems (CWS) and Physical Vulnerability of a CWS's water sources, although these were not folded into the drinking water quality scores. TMF capacity is one measure of system-level vulnerability, indicating overall system sustainability. Given data availability, we measured it with a proxy variable based on the estimated population served by the system. Physical vulnerability captures how vulnerable the water system is to shortages based on the source of water and how many total sources of supply the system has (if on groundwater).

To derive a tract-level rank of cumulative impacts of drinking water at a regional and statewide level, we followed three main steps: 1) using existing estimated water system geographies in order to estimate system-and tract-level measures, 2) calculating system-level measures of contamination and compliance, 3) estimating tract-level measures based on the system-level measures. The first step involved three main components. First, we used estimated boundaries of community water systems (CWSs) derived from OEHHA's compilation of the Environmental Health Investigation Branch's Tracking Program and a method of estimating boundaries developed by OEHHA. In essence, two types of public community water system geographies are scored: CWSs with known boundaries (1,562 systems covering 33.7 million people), and CWSs with estimated boundaries (1429 systems covering 1.3 million people). Second, we used estimated boundaries of areas not served by CWSs, also developed by OEHHA. Here, township grids were used as the geographic boundary and unit of analysis to derive our drinking water for these populations. For areas of the state not covered by CWSs, a 6x6 mile grid of townships is used to define areas where people are likely to be drinking well water (approximately 1.5 million people). (Note that ~.6 million people are not assigned water quality because they are not within a township that has a groundwater sample.) Together, we used OEHHA's aerial weights for the contribution of a water system or township to a tract. Here, both sets of boundaries were intersected with 2010 US census block boundaries. The area for each system-block portion was calculated. The proportion of that system contained within the block (the aerial weight) was then used to calculate an aerially-weighted population contribution for that tract. As an example, if 100% of one system is contained in one block, the block's entire

population is counted. If 50% of a system covers a block, then 50% of the population gets counted when deriving weights.

System-level (i.e. at CWS or township level) metrics are summarized in **Table 5**. Two types of data were used to determine if a water system had a monitoring and reporting burden. First, we used monitoring and reporting violations as tracked in the PICME database, from 2005-2013. For each system, we determined whether the system had received: any "monitoring and reporting" (M&R) violations for 4 key contaminants: arsenic, nitrate, perchlorate and TCR. These contaminants were included because of the nature of their monitoring/waiver requirements. Inorganic contaminants have to be monitored at least once every 9 years, even if they have a waiver. All other contaminants—organic VOCs, organic non-VOCs, radiological and DBPs can be given continual waivers. If a system has a waiver, it doesn't have to keep monitoring for the contaminant. Without knowing which systems have which waivers, it is not possible to indicate which systems should have been monitoring. Therefore, the 4 inorganic contaminants are used as a general proxy for how well the system is in compliance with monitoring requirements. The fact that arsenic, nitrate and TCR are among the top contaminants across the state, is a second reason why we focused on these contaminants.

Potential exposure to contaminants was determined for the following analytes (Arsenic, Barium, Benzene, Cadmium, Carbon Tetrachloride, Lead, Mercury, MTBE, Nitrate, Perchlorate, PCE, Radium 226, TCE, THMs, TCR, Toluene and Xylene) using a time-weighted average for each contaminant of interest for each water system. Water quality data was obtained from CDPH's Water Quality Monitoring database (WQM) for the years of 2005-2013. A similar time-weighted average is used to calculate water quality in townships, using water quality characterizing groundwater in the local basin. Inclusion criteria for contaminants was based on the percentage of water systems that had water quality information for specific contaminants. If more than 80% of systems had water quality data for that contaminant, the contaminant was included. This resulted in the inclusion of 17 key contaminants. Since Total Coliform is only assessed for presence or absence, a system was given a binary value of 0 or 1 if it had received a maximum contaminant level (MCL) violation for Total Coliform during the study period. If a block has no contaminant information for a particular contaminant, it was not used to assess the population-weighted average for that contaminant. The tract-level contaminant average is divided by its corresponding MCL. If data was missing for the tract, a regional-average was used for that contaminant's tract-level average; for TCR a 1 or 0 is added whether the tract had at least one MCL violation. The sum of each of these "MCL ratios" is then taken to create a total sum of ratios. Intermediate scores for drinking water quality were calculated using a quintile distribution and normalization technique with scores ranging from 1 to 5. In addition, we used hatch marks on the water quality maps to also indicate systems that were considered vulnerable due to a lack technical, managerial and financial capacity, physical vulnerability due to few water sources and compliance with testing and reporting requirements.

Table 5: **Derivation and Scoring of Drinking Water Quality Indicators**		
Drinking water contaminants (potential health risks from drinking contaminated drinking water)	*At the system level:* Captures potential *average* exposure given time-weighted average during a 9-year compliance cycle. Calculated for 17 contaminants Arsenic, Barium, Benzene, Cadmium, Carbon Tetrachloride, Lead, Mercury, MTBE, Nitrate, Perchlorate, PCE, Radium 226, TCE, TTHMs, Toluene, Xylene and Total Coliform (this is a binary—whether system had TCR MCL violation or not) *At the tract level:* For each contaminant, a population-weighted aerial average of the 9-year average is calculated for each tract. If a block has no contaminant information for a said contaminant, it is not used to assess the population-weighted average for that contaminant. *The tract-level average is divided by the contaminant's MCL. If data is missing for the tract, a regional-average is inserted for that contaminant's tract-level average; for TCR a 1 or 0 is added whether the tract had at least one MCL violation. The sum of each of these "MCL ratios" is then taken to create a total sum of ratios.	CDPH's WQM Water Quality Data, calculated as time-weighted average for each contaminant for 9-year compliance period: 2005-2013 OEHHA's compiled "Water Boundary Layer"+ townships calculations (see CalEnviroScreen for more information) Using EHIB Tracking Program's data, Version 1 included only: Calculated for nitrate, arsenic, uranium, Atrazine, TCE, PCE and Total Coliform, TTHMs
TMF Capacity (measure of system-level vulnerability, indicator of overall system sustainability)	*At the system level:* TMF is calculated using a categorical variable based on population served of water system. Townships are given the highest TMF score (=4), since populations not served by CWSs are assumed to be very vulnerable in this regard. *If data is missing for the system, it is left blank, different from TPC variable below. *At the tract level:* Aerial-weighted average TMF is calculated across systems in the tract. *If data is missing for the tract, a regional-average is inserted for that tract.	WQM dataset (# of people served)
*Combined Vulnerability (*TMF+Physical+ Compliance Burden)	*Step 1: At the system-level*: TMF + Physical Vulnerability + Compliance Burden are added together to produce a range from 2-13. *Systems that had blanks for TMF or Physical Vulnerability are filled in by the regional system-level average. This allows all systems to receive a vulnerability score that then gets aggregated to the tract-level. Note the regional average is not population-weighted. *Step 2 at the tract level:* a) % of population drinking from a high vulnerability (score of 8-10) or very high vulnerability (score of 11-13) CWS divided by total CWS population in a tract is calculated, b) % of population drinking from a township is calculated in a tract *Step 3 at the tract level:* 90th percentile for parts a & b above is calculated. Binary variable is created if tract is in 90th percentile for either a or b (in Step 2)	PICME dataset (# of Monitoring & Reporting violations) OEHHA's water quality data (whether the system had any missing water quality data for a specific contaminant)

5 Results

We present maps of the EJSM scores for eight regions which are demarcated in **Figure 2** (San Francisco Bay Area, Southern California, San Diego, San Joaquin, Sacramento, Central Coast, Northern California). We also showed cumulative impact scores for (**Figures 3-18**). Maps display both intermediate scores (e.g. mapped distributions for each of hazard proximity and sensitive use; health risk and exposure; social and health vulnerability; climate change vulnerability; drinking water quality) and total cumulative impact scores across all and subsets of the five dimensions. We also provide a comparison of cumulative impact scores derived using regional versus state quintile distributions to derive intermediate and final scores. Although the EJSM is flexible enough to allow for comparisons across different study areas (e.g., within regions or across the state) we tend to favor a regional application of scoring because generally land use planning, industrial and transportation development, and environmental regulation are regionally rooted and require regionally specific interventions to reduce hazard exposures or to address social and health vulnerability factors. In addition, statewide scoring can mask important within-region inequities (See **Figures 3A-D** for SF Bay Area and **Figures 8A-D** for San Diego, for example) which can make these areas fall below the regulatory radar screen. At minimum, we recommend examining cumulative impact scores within regions using both regional and statewide distributions to derive quintile scores for each EJSM dimension. There may be certain policy and decision-making contexts when statewide distributions for scoring may be appropriate for larger scale impacts such as climate change vulnerability, while other environmental hazards that are more locally driven require a regional scoring approach to effectively elucidate areas that may require targeted regulatory attention.

Figure 2: Regions Analyzed for the EJSM

Figure 3A-F: SF Bay Area EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

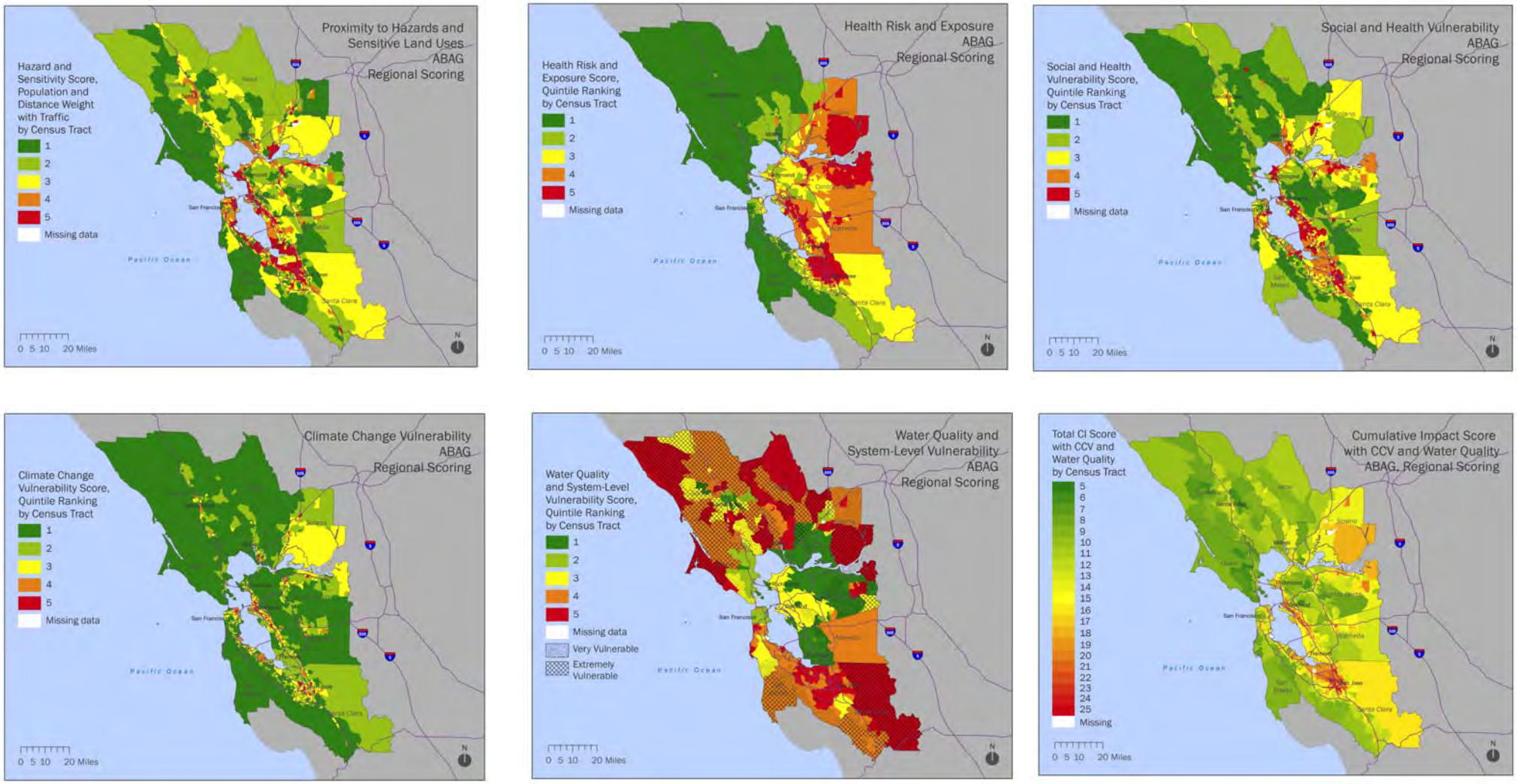


Figure 4A-D: SF Bay Area Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

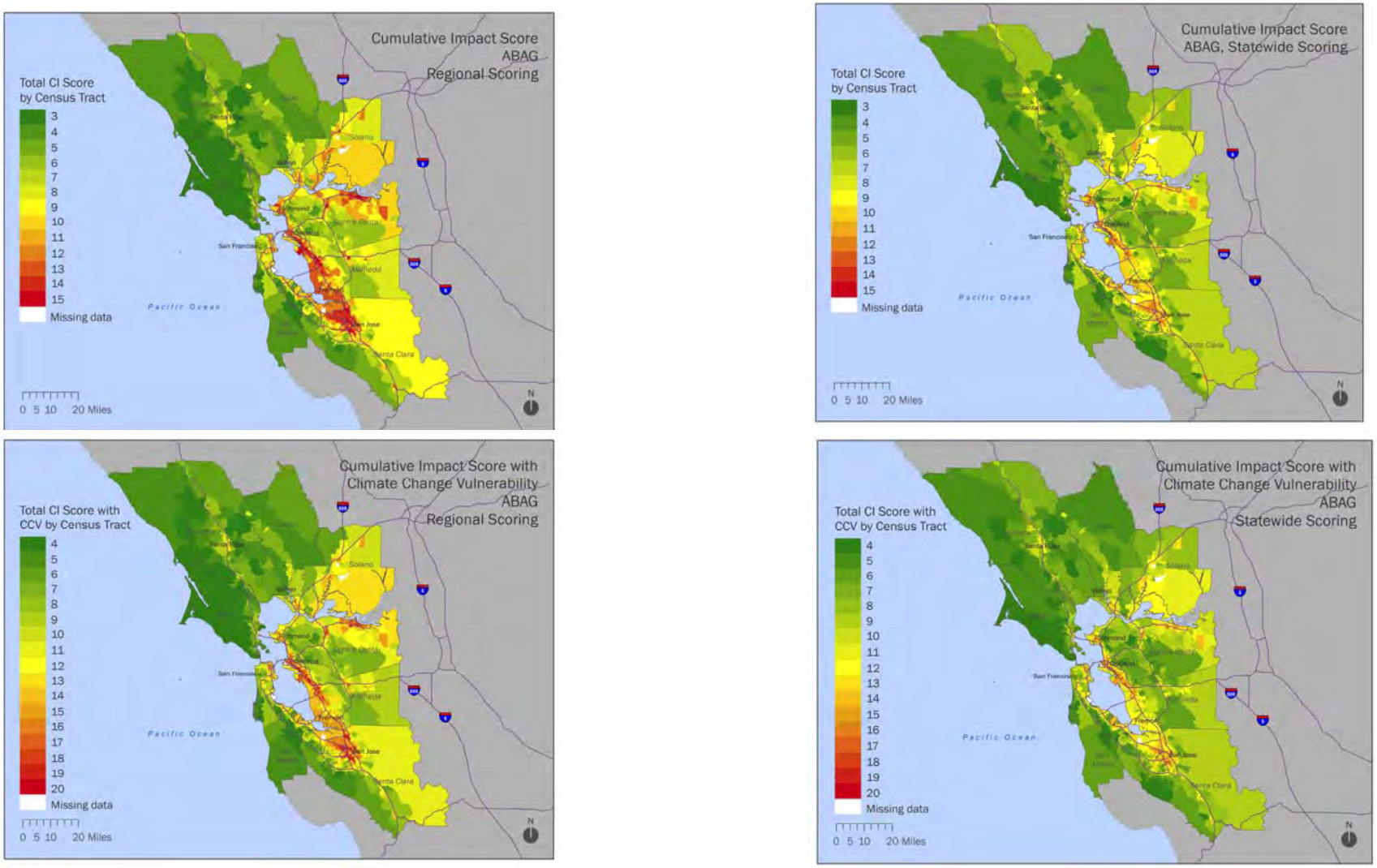
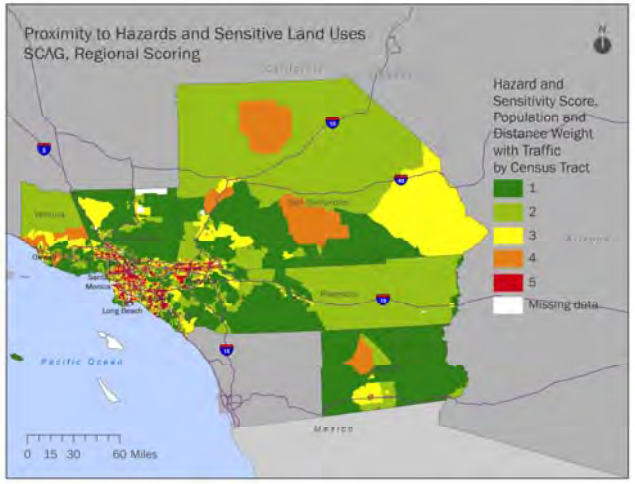
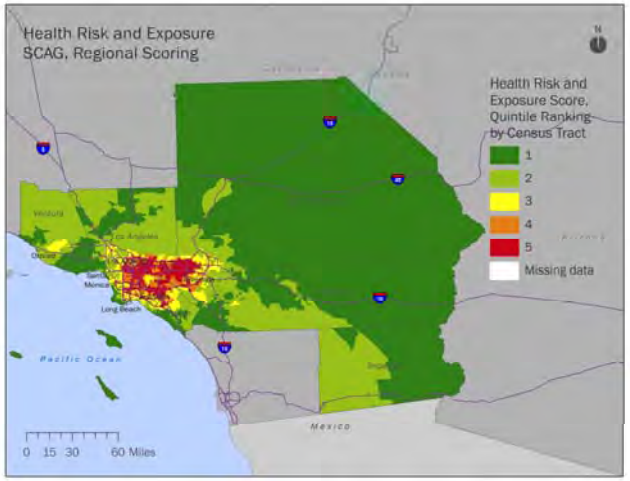
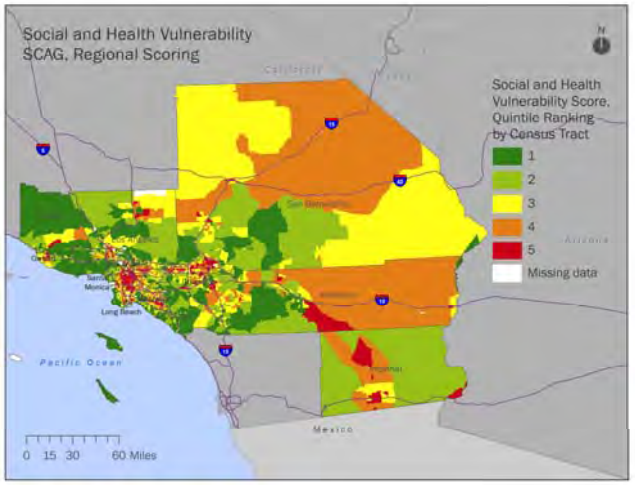
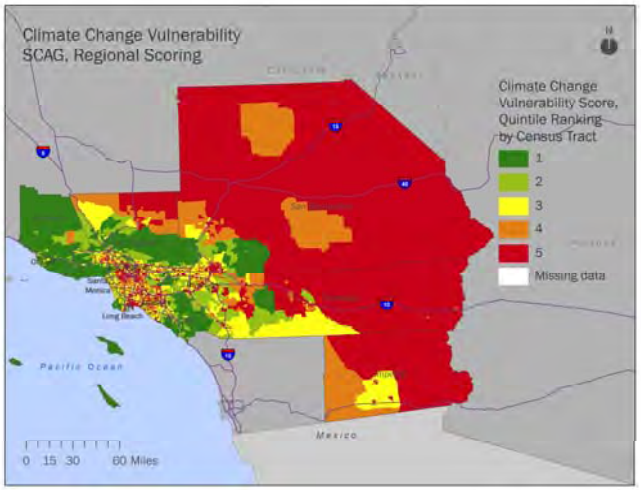


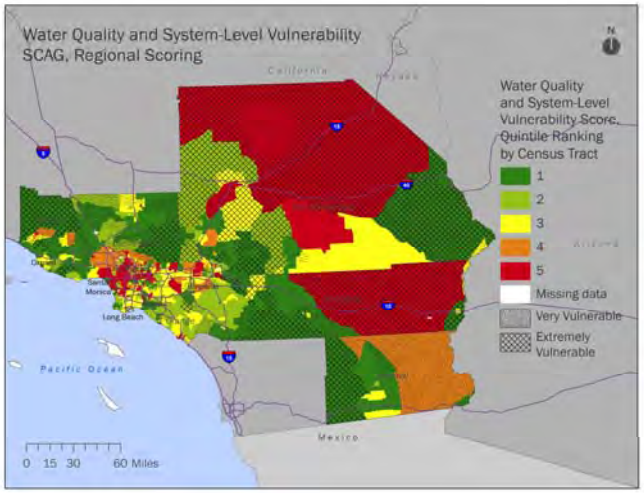
Figure 5A-F: Southern CA EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)











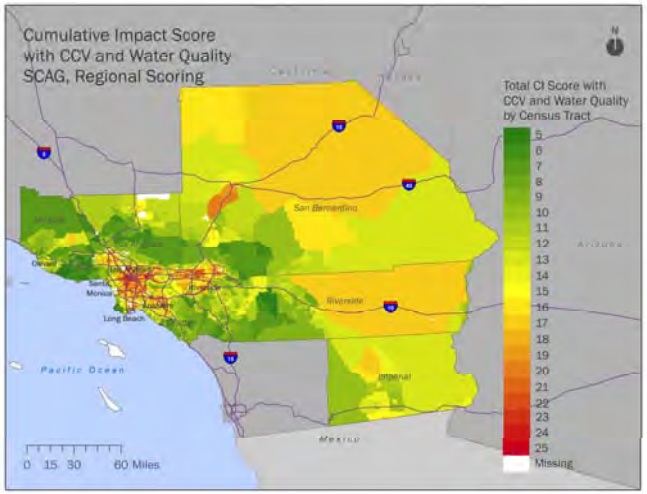


Figure 6A-D: Southern California Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

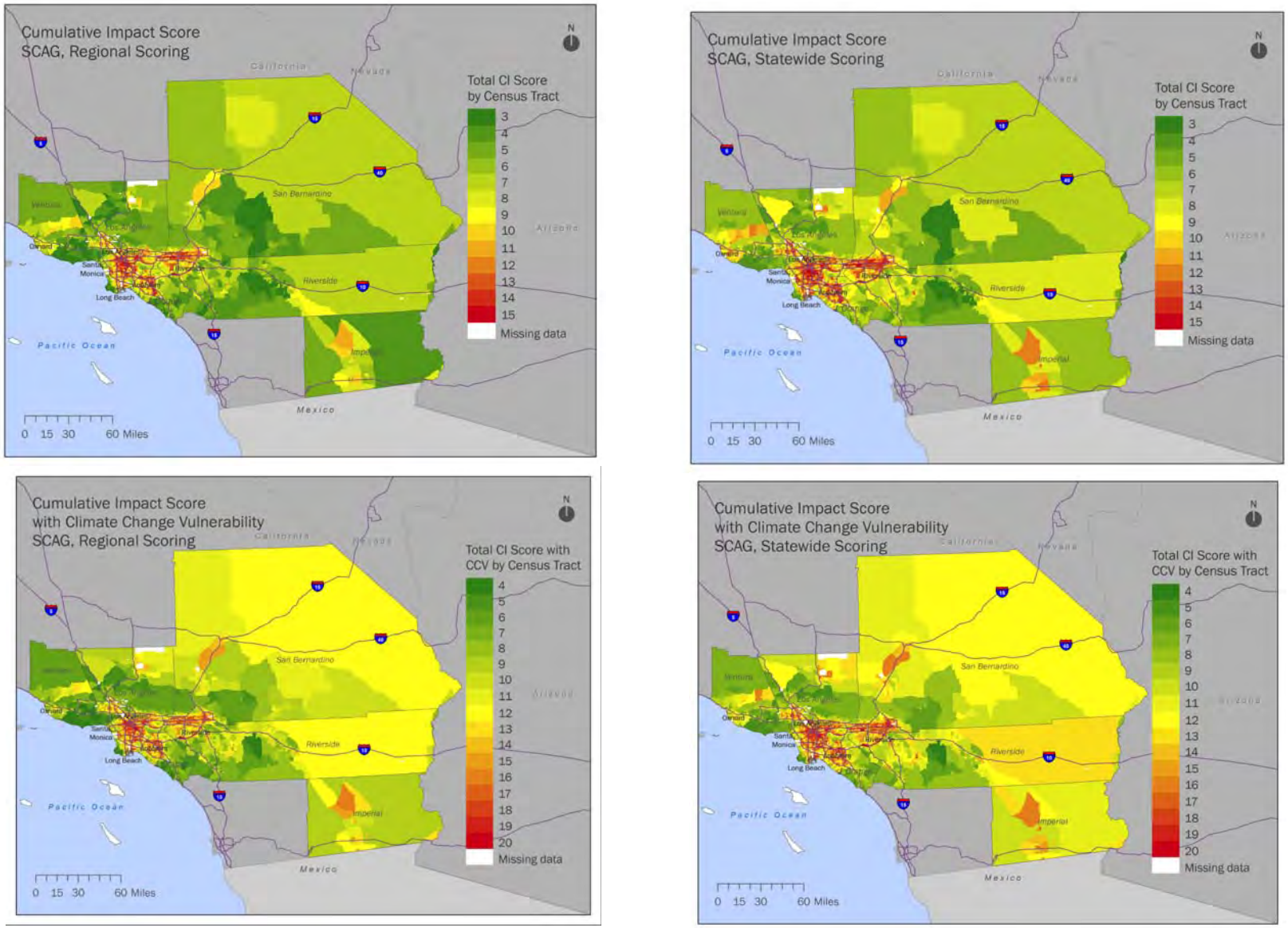
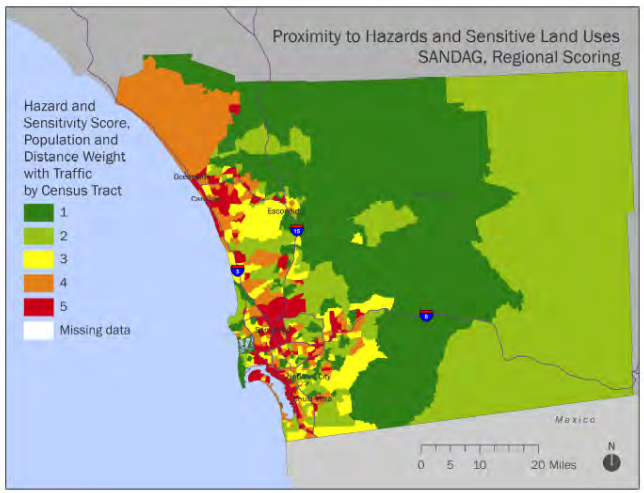
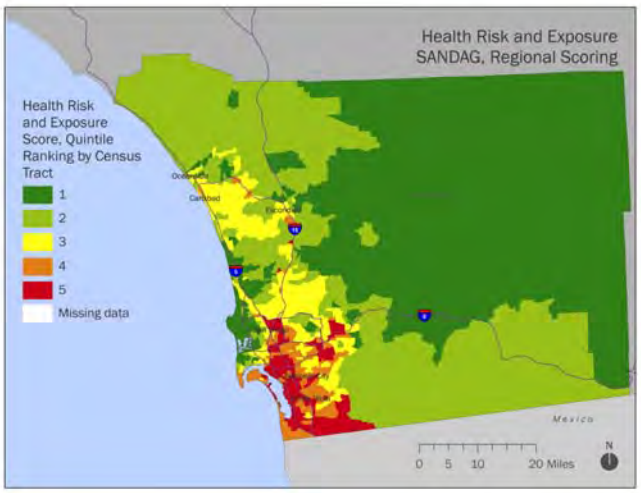
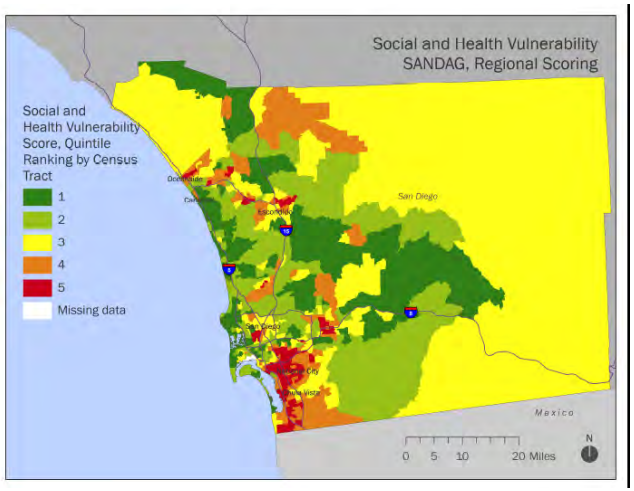
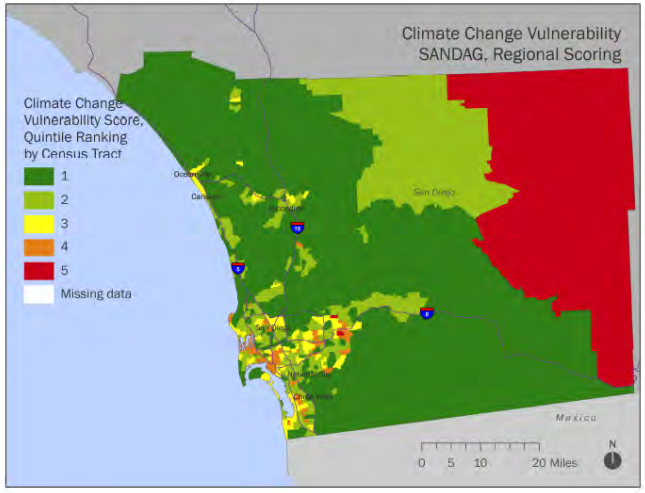


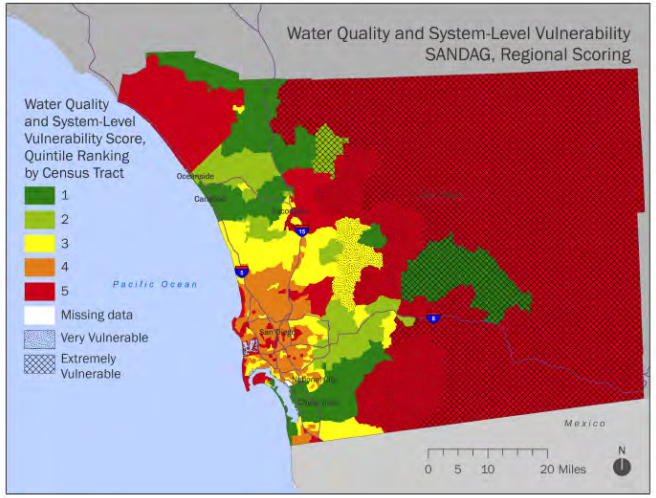
Figure 7A-F: SANDAG EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)











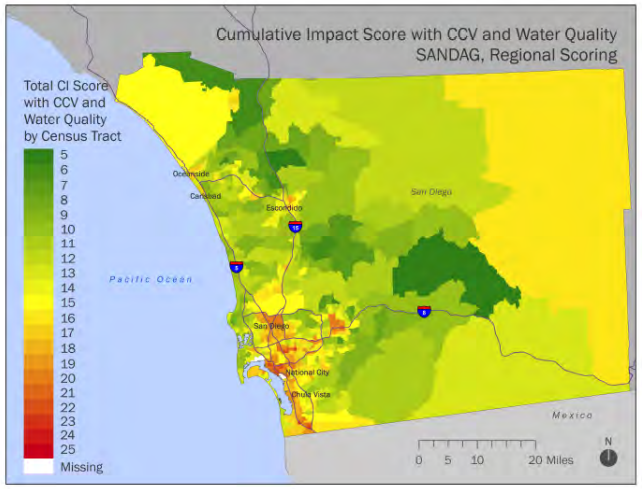


Figure 8A-D: SANDAG Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

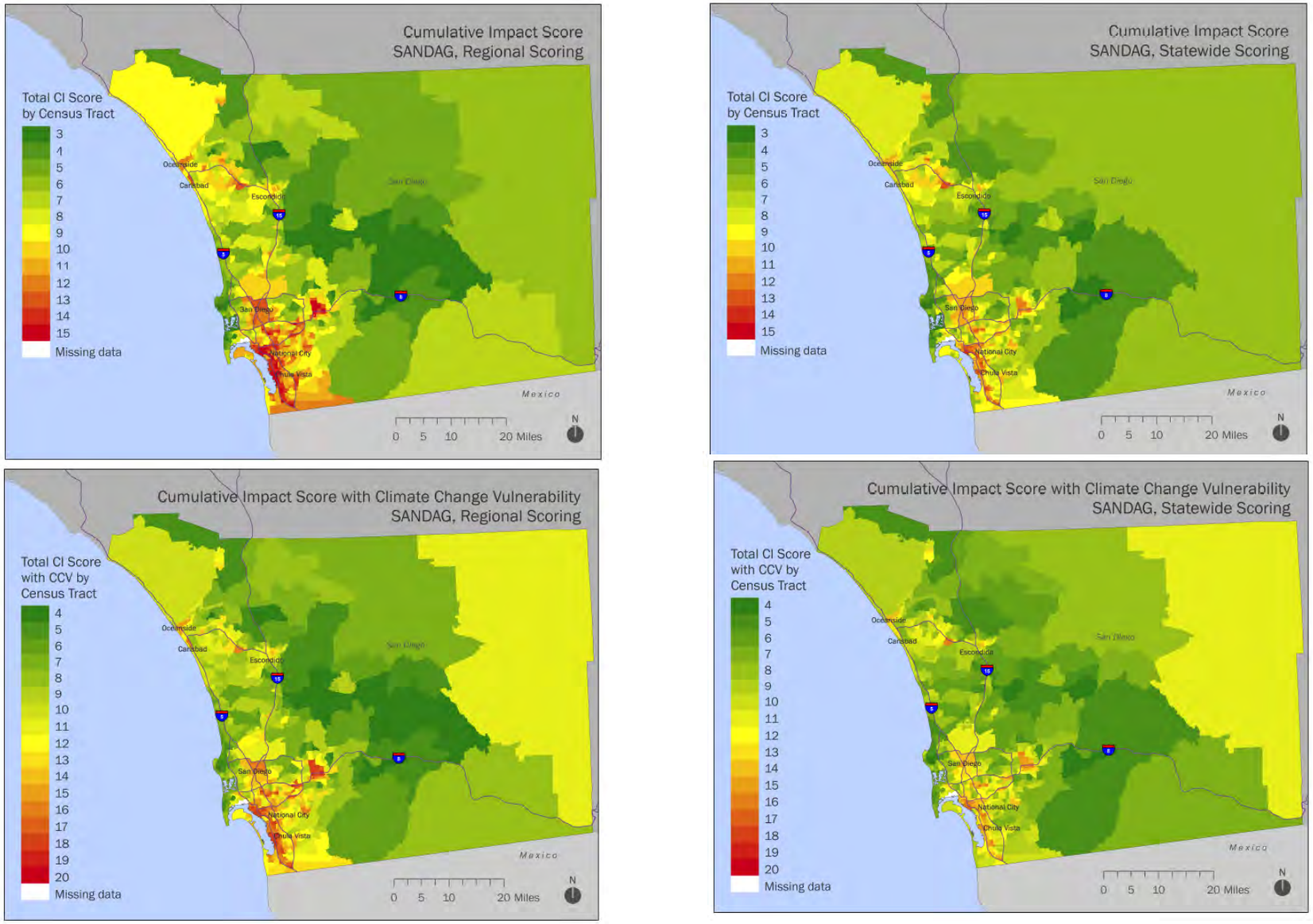


Figure 9A-F: San Joaquin Valley EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

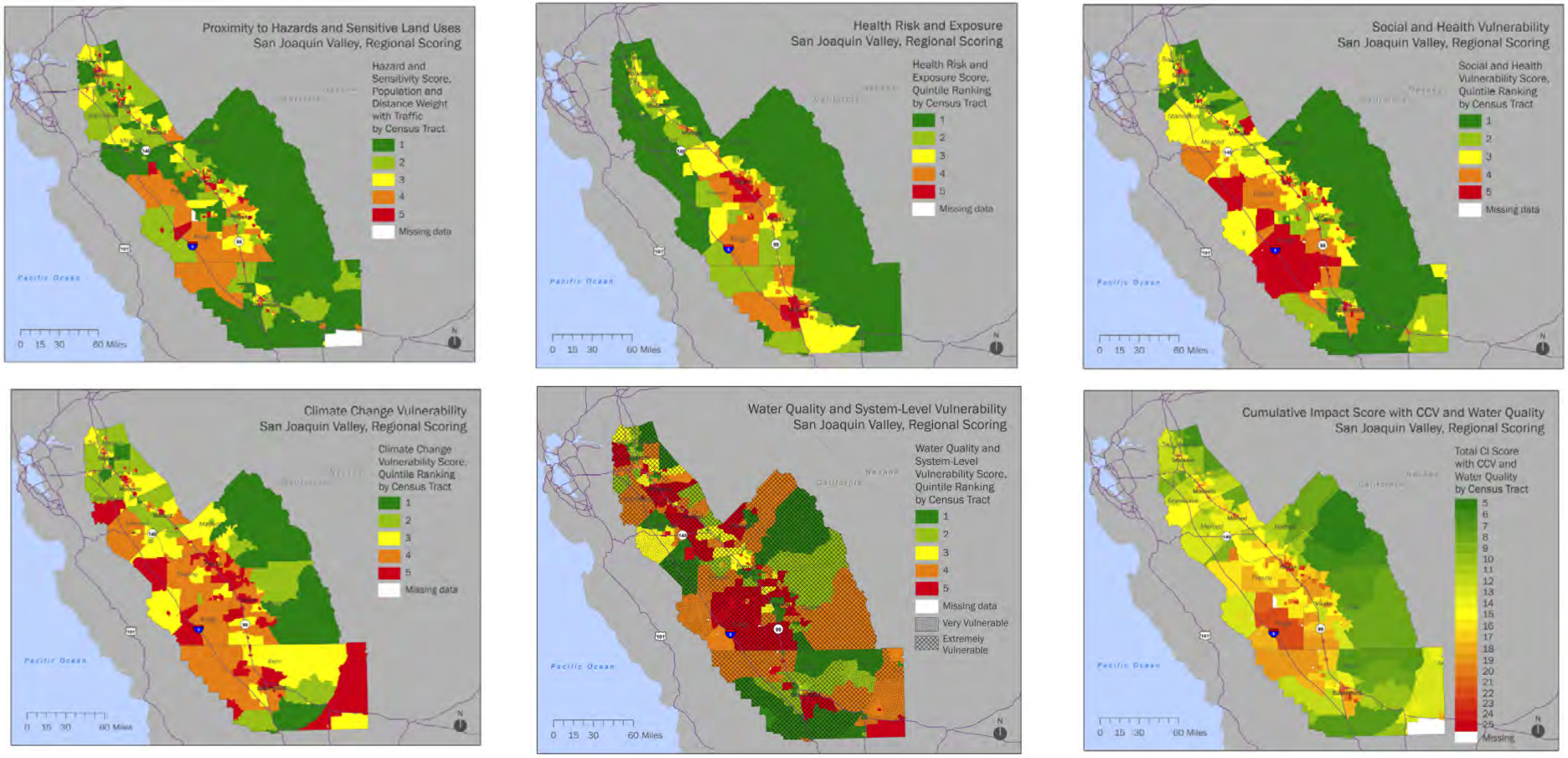


Figure 10A-D: San Joaquin Valley Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

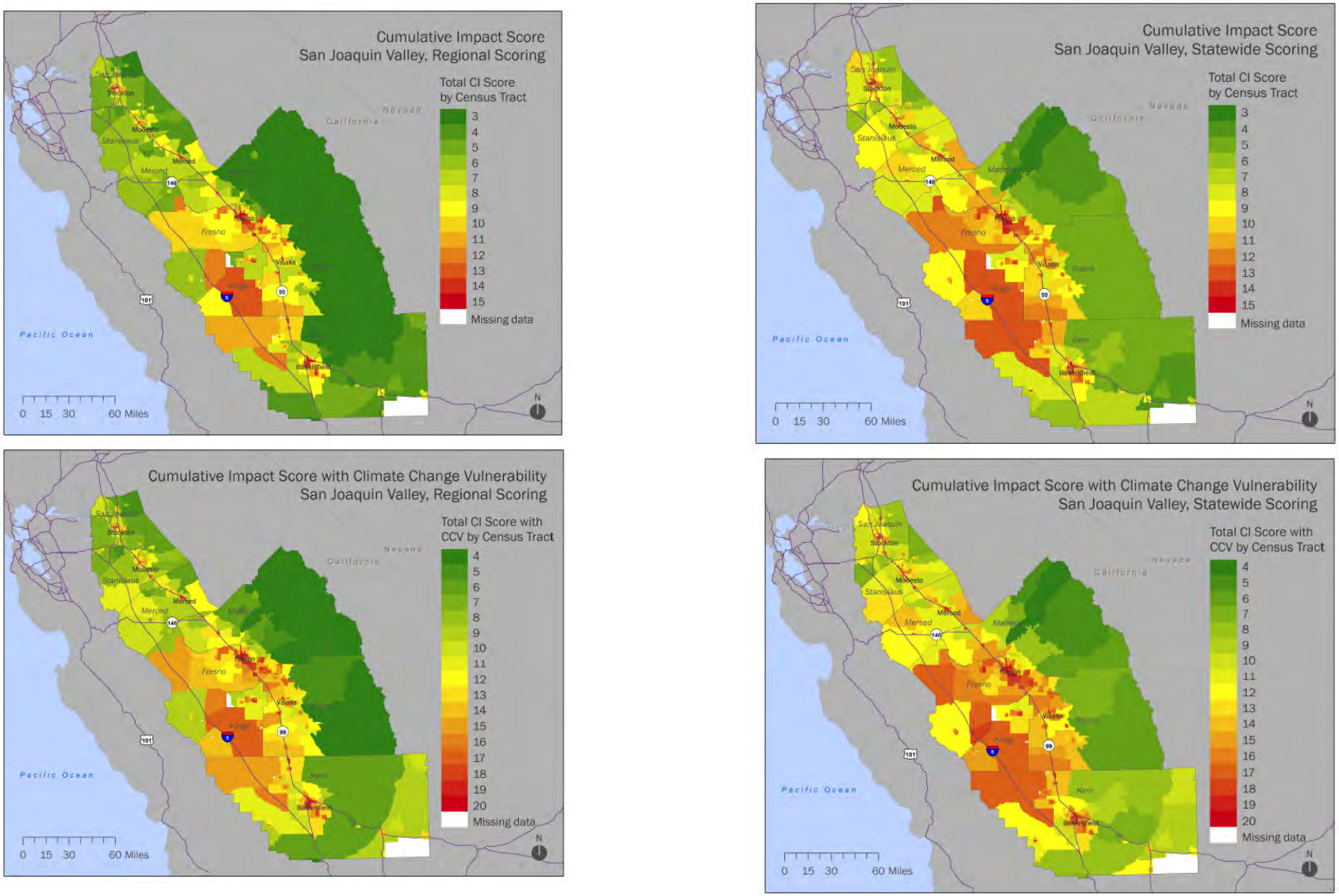
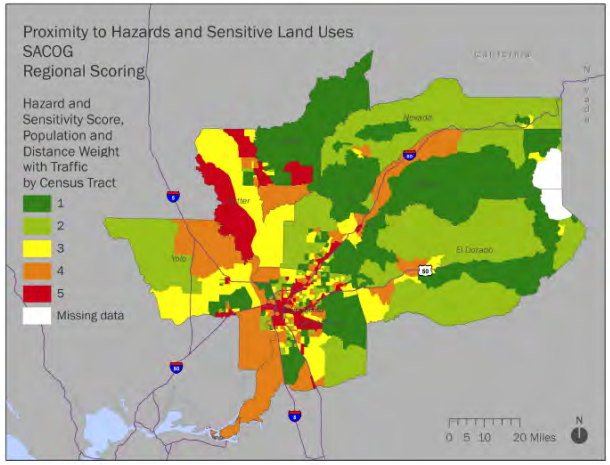
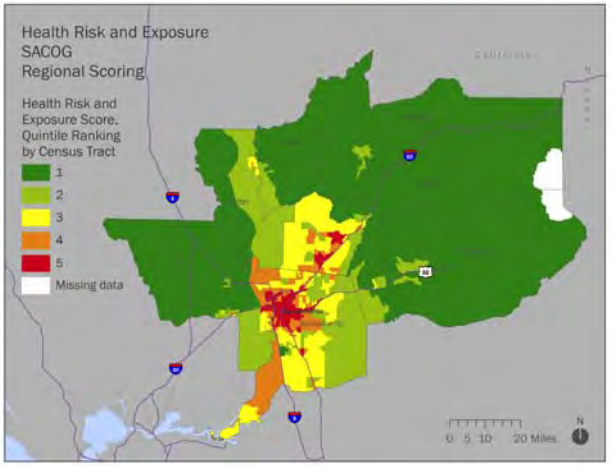
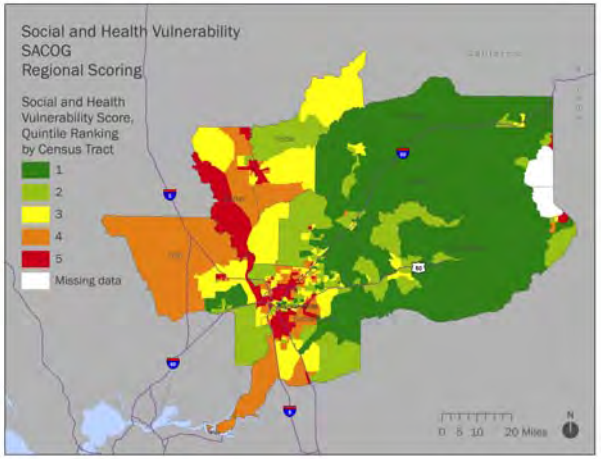
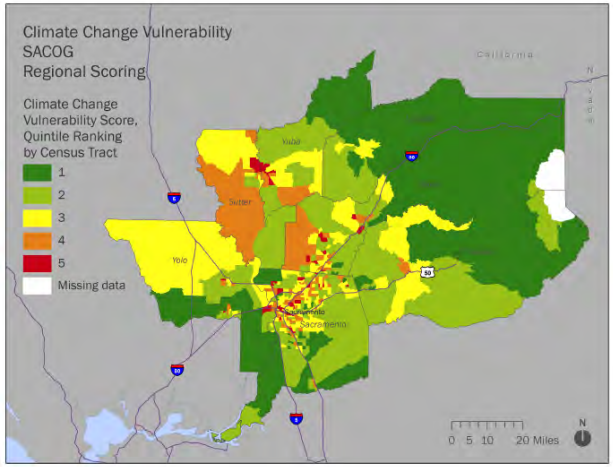


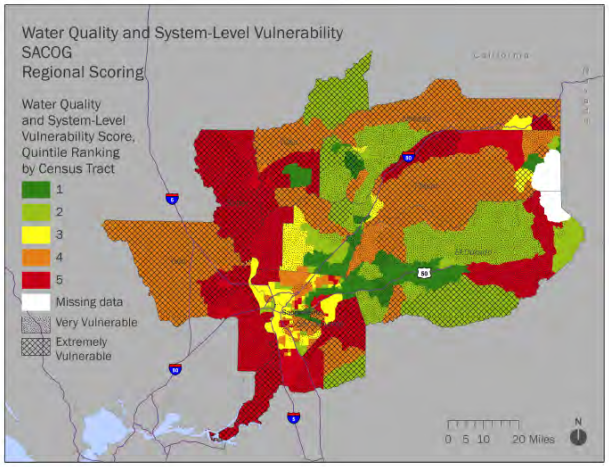
Figure 11A-F: SACOG EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)











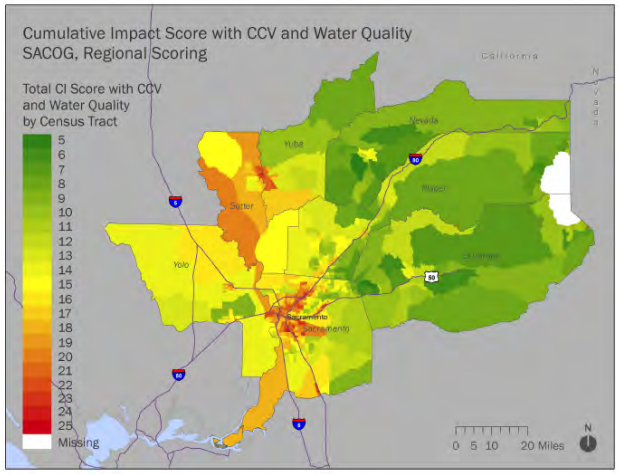
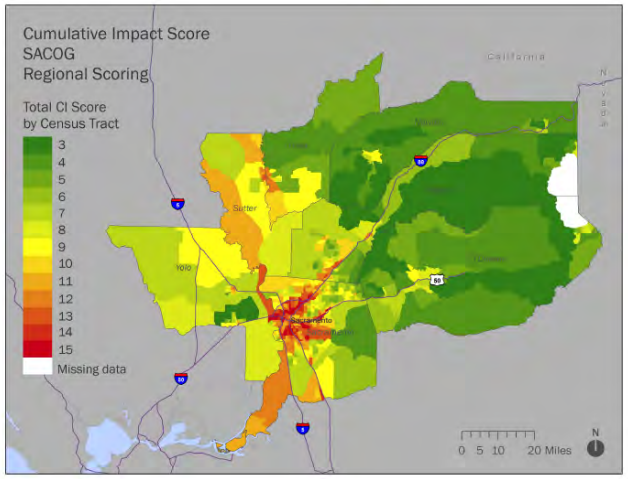
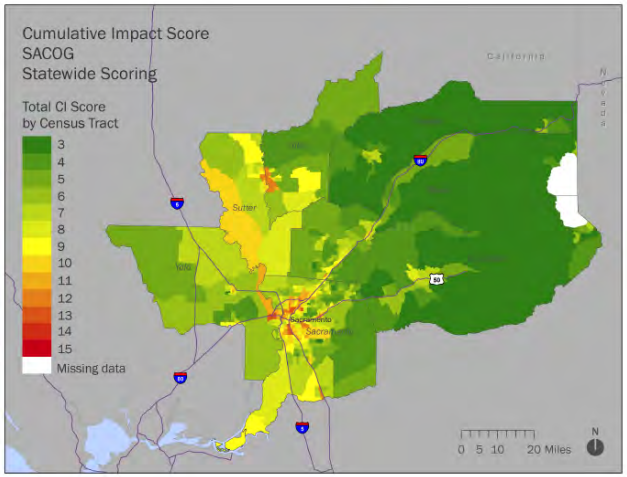
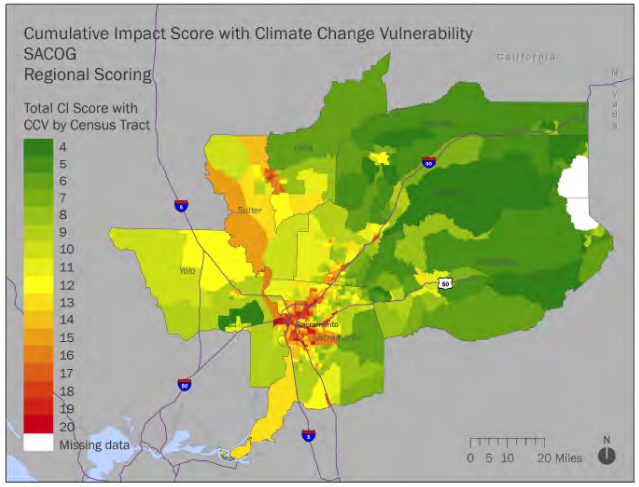


Figure 12A-D: SACOG Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)







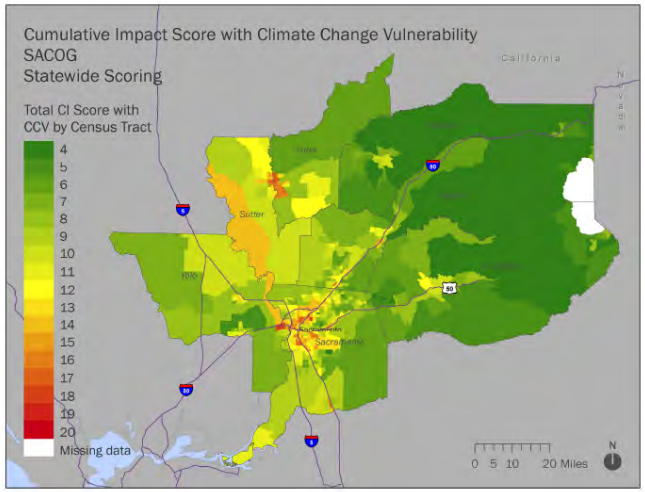
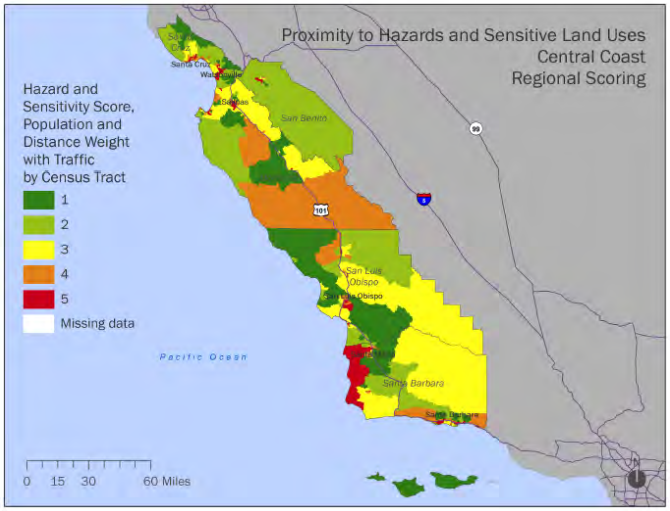
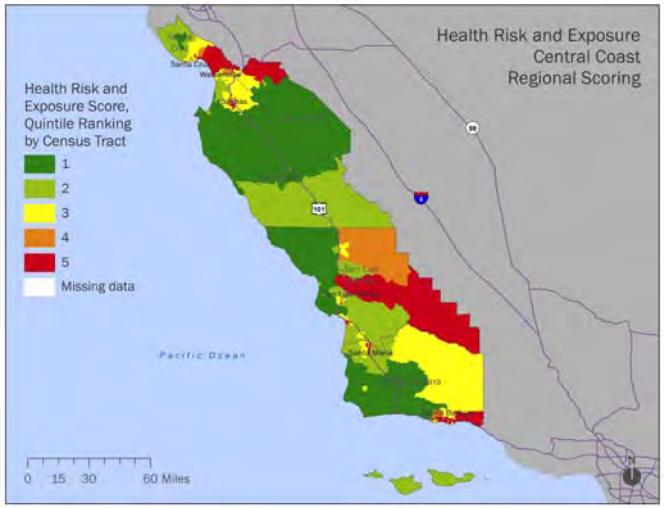
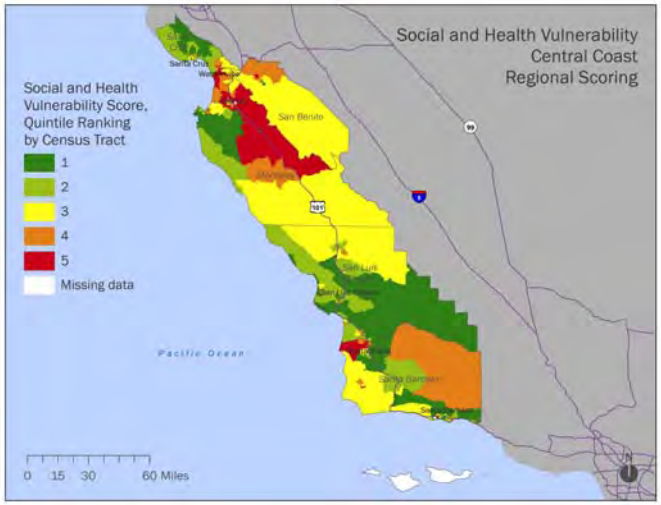
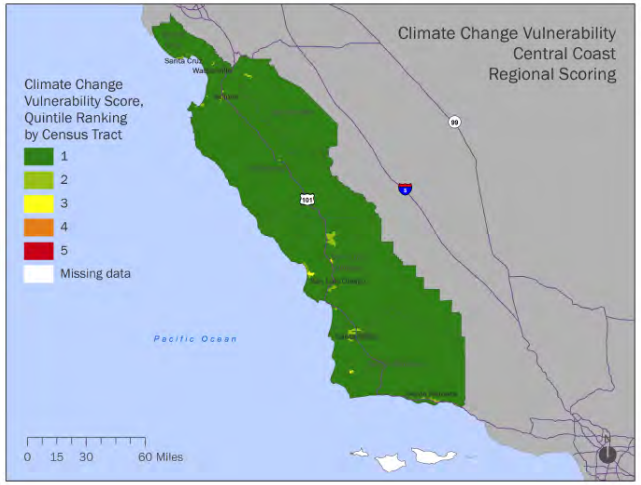


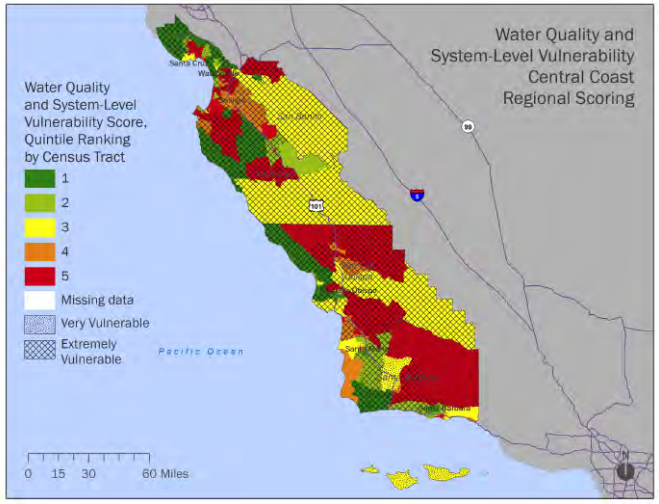
Figure 13A-F: Central Coast EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)











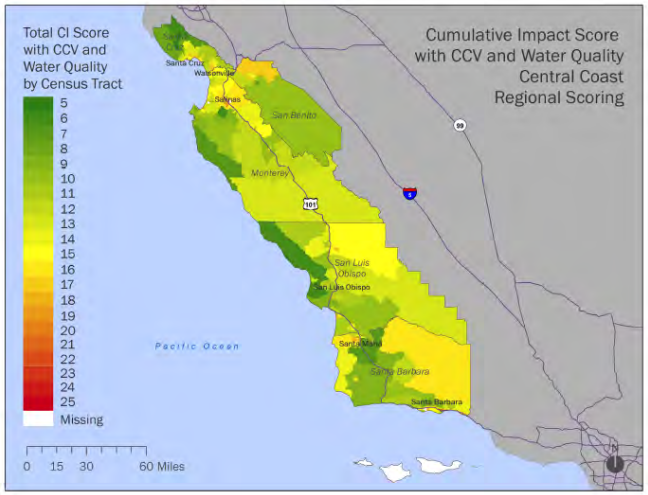
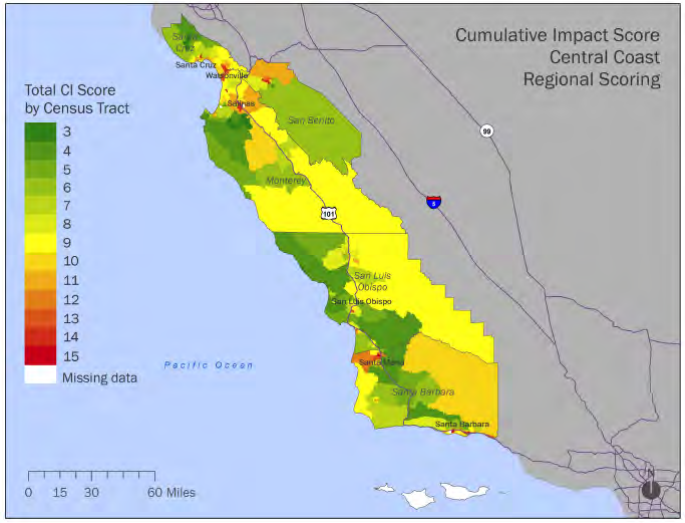
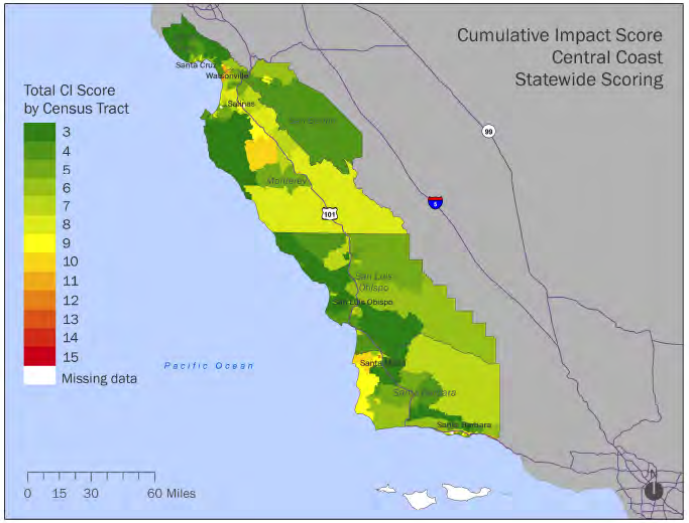
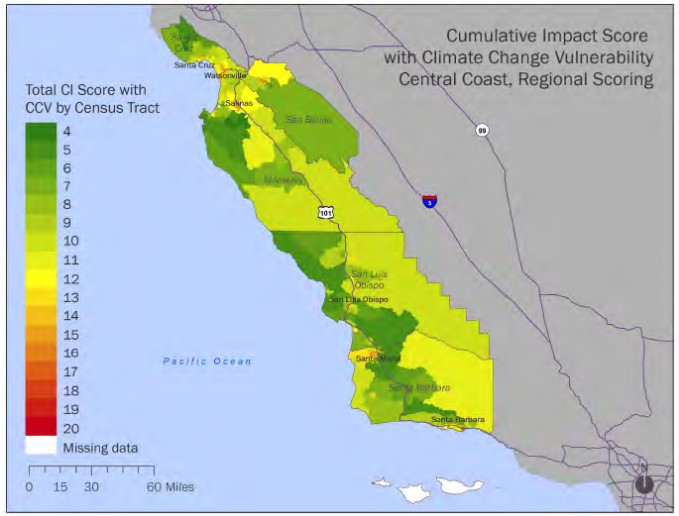


Figure 14A-D: Central Coast Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)







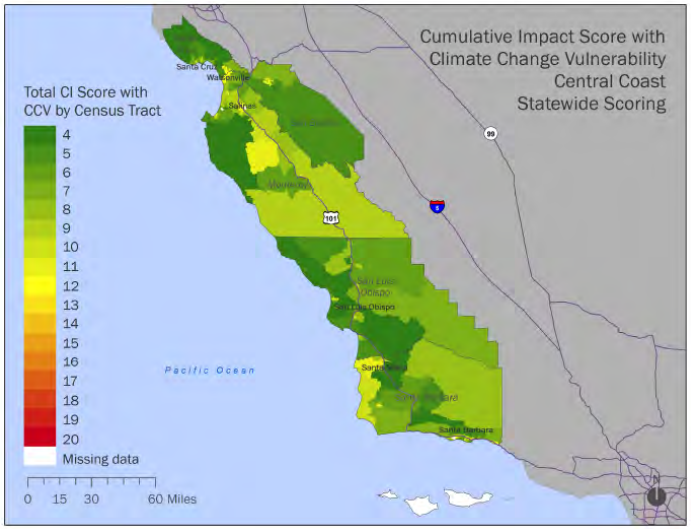


Figure 15A-F: Eastern Sierra EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)

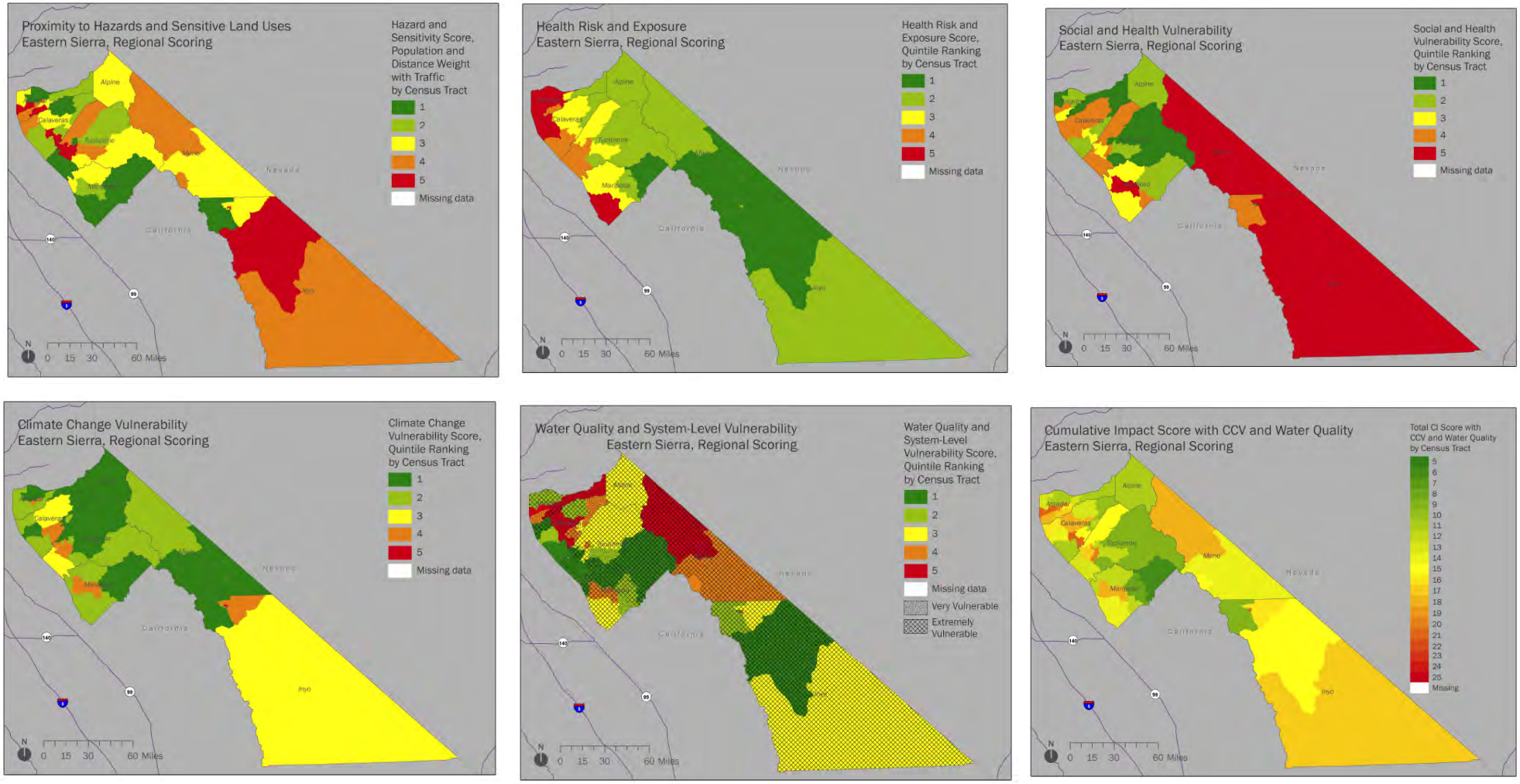
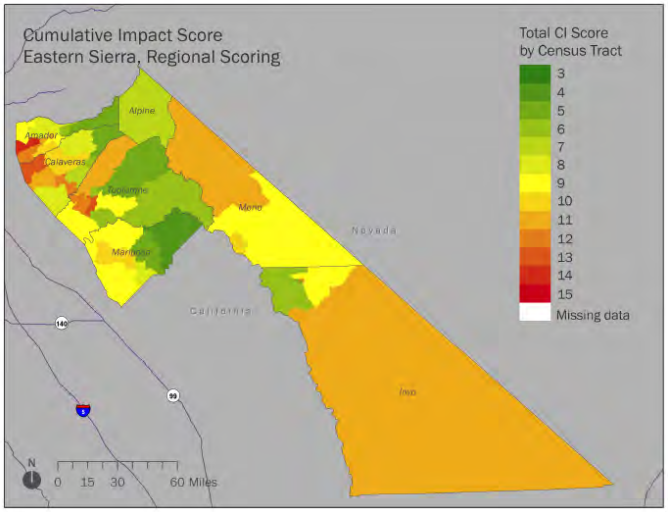
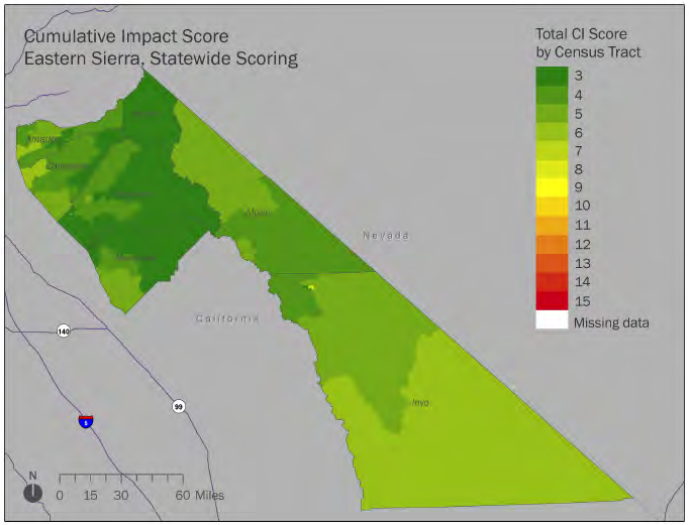
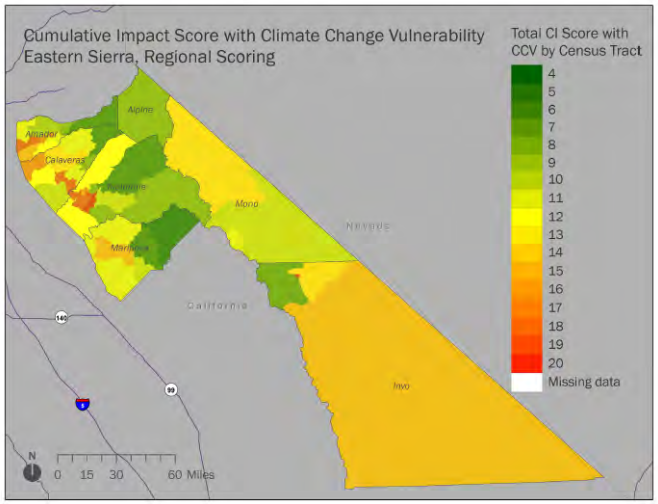


Figure 16A-D: Eastern Sierra Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)







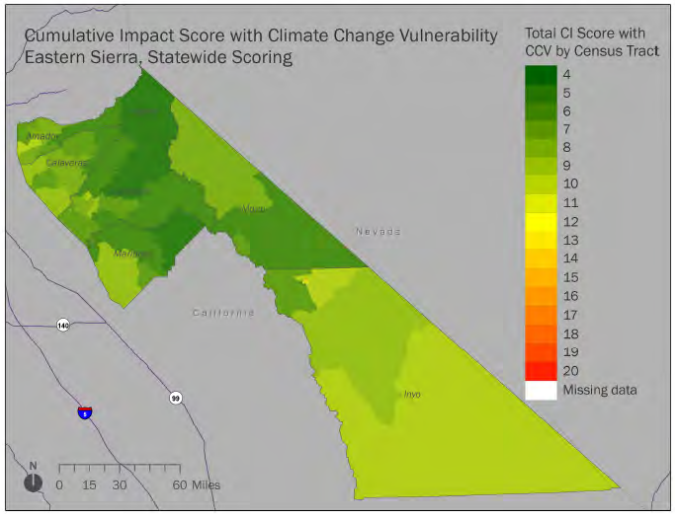
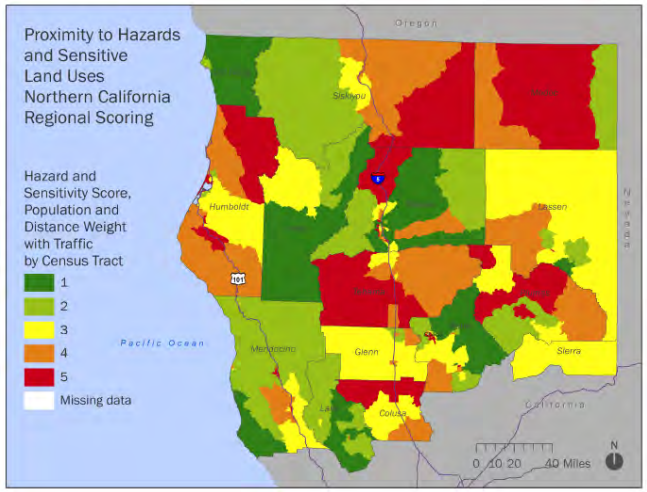
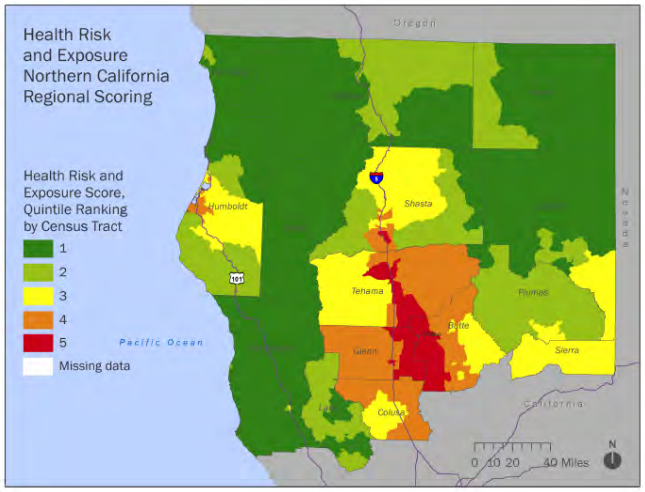
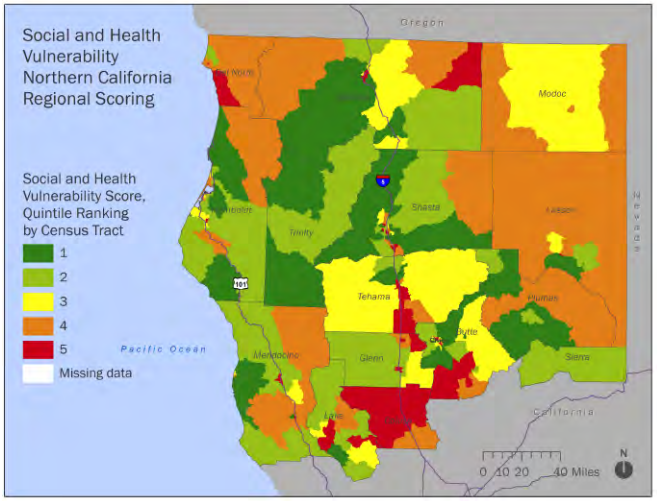
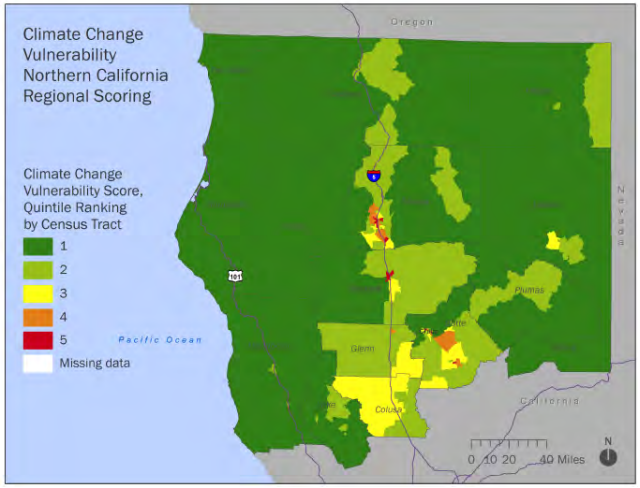


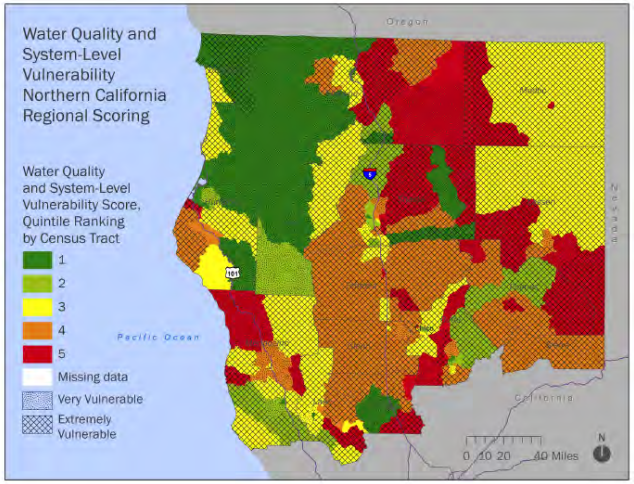
Figure 17A-F: Northern California EJSM Intermediate and Cumulative Impact Scores (Regional Scoring)











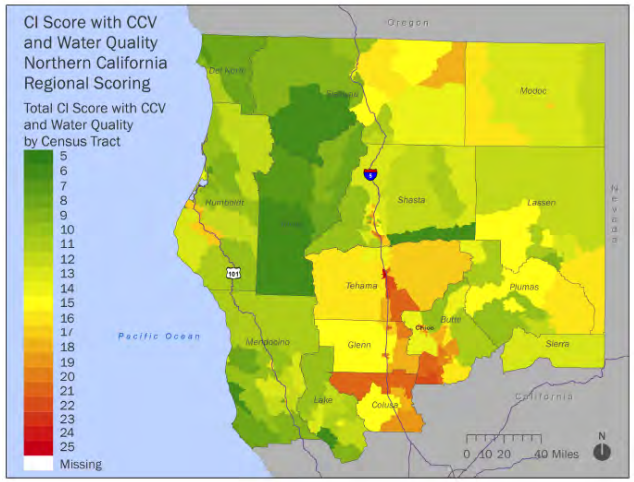


Figure 18A-D: Northern California Cumulative Impact Score Comparing Regional versus State Scoring (with and without Climate Change Vulnerability)

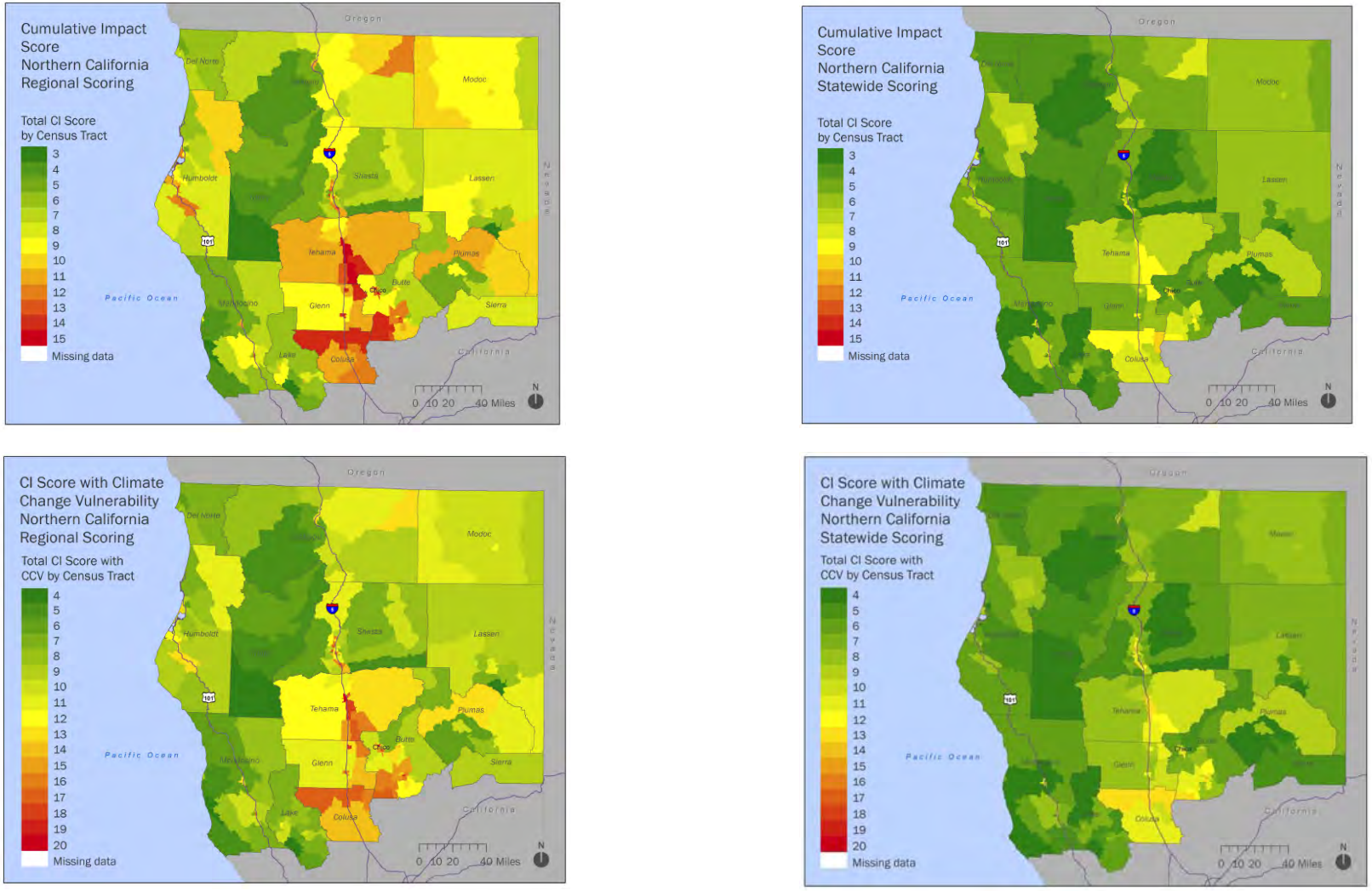
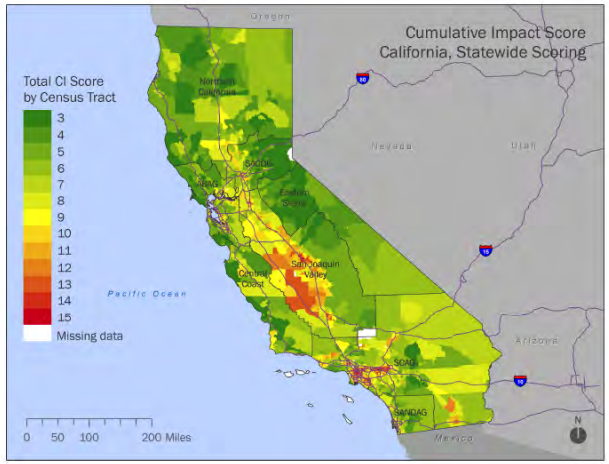
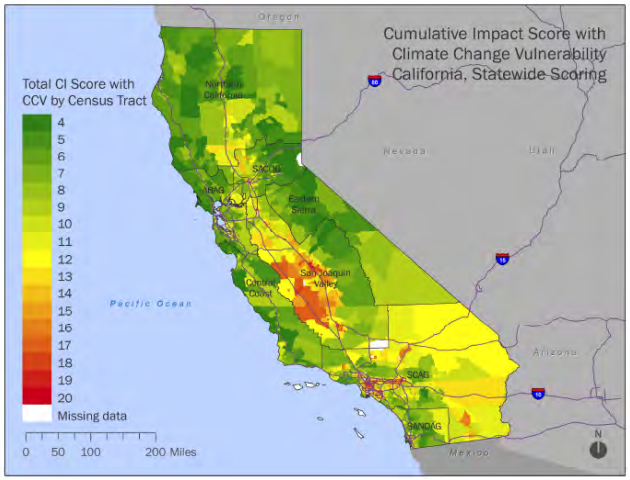
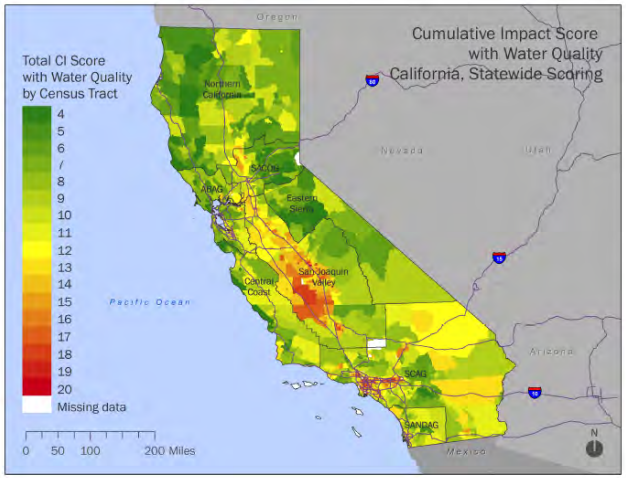
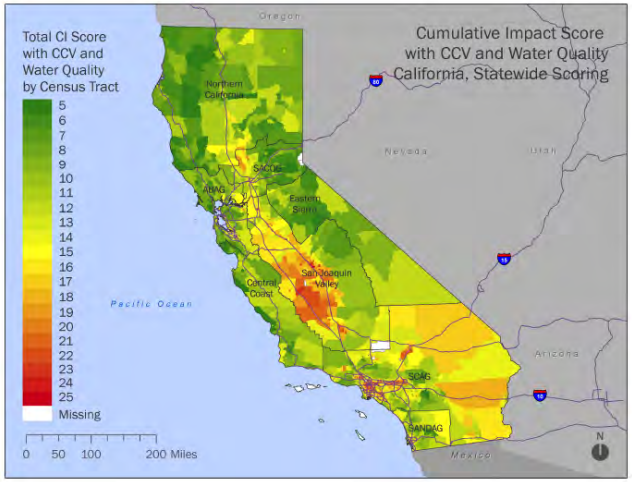


Figure 19A-D: California Impact Score -- Statewide









5.1 Comparison with Other Screening Approaches

Although systematic comparisons between the EJSM and other screening tools, such as OEHHA's CalEnviroScreen (CES) and Cumulative Environmental Assessment (CEVA) developed by London and colleagues at UC Davis, show many similarities in both approach and data metrics used, some notable differences and similarities are worth mentioning. First, although there is considerable overlap in some of the data inputs and metrics in all three screening methods, only the EJSM systematically incorporates sensitive and residential land use for the development of its base maps and its approach to exposure (by proximity) estimation. This orients regulatory attention more toward populated areas and less toward areas that are dominated by industrial and commercial land uses and enables a more granular (higher spatial resolution) evaluation of hazard proximity. This method also allows for areas and uses not scored to be "masked off" in final thematic maps, improving the interpretation of spatial patterns of scores which can be complicated by the presence of

Table 6: Scoring Approaches in EJSM, CES and CEVA

EJSM	
	1. Scoring procedure • Linear quintile ranking of each metric within the 5 cumulative categories • Each quintile score is summed and re-ranked to derive a total quintile score for each cumulative impact category • Final cumulative impact score is derived by summing all category scores 2. Approach is open-ended to accommodate additional indicators and allow for new categories 3. Preferred scoring strategy is regional, although method accommodates statewide scoring as well
CES	
	1. Indicator categories are multiplied to yield a continuous, open-ended score 2. Scores are grouped into percentiles (1-20) with same number of tracts in each value category 3. Statewide scoring only
CEVA	
	1. 3x3 scoring matrix (1-9) with separate axes for impact and vulnerability 2. Scores applied to specific regions.

large census polygons with low scores dominating non-urban parts of the State. Second, EJSM uses a greater number of indicator metrics than either of the other two approaches, which both increases its potential utility and reduces the influence of any single metric on the final score. Third, EJSM is the only screening approach that includes a layer with metrics on climate change vulnerability. This element could be potentially useful to guide decision-making regarding future investments in climate change mitigation strategies so that funds are more targeted toward neighborhoods that have higher risks of heat islands or more dramatic changes in projected temperature, for example.

The three screening approaches use different scoring methods, shown in **Table 6**. EJSM is distinctive in that its scoring approach results in the frequency distribution of scores following a bell shaped distribution, which means that outlier tracts are likely to be fewer than in CES.

6 Additional Activities

We finalized the programming of our scoring approach in ArcGIS Model Builder and SPSS to automate the processing and scoring of metrics. These scripts were provided to CARB staff. We also finalized the translation of SPSS programming into SAS, a task that was not one of the deliverables for this contract, but which we agreed to provide to CARB staff to enhance the use of our EJSM scoring routines by SAS programmers at the Agency.

The research team held one in-person stakeholder meeting on January 16th 2015 and a webinar on May 13th 2015 to share the final results of our EJSM analysis. The January meeting was hosted by the Program on Environmental and Regional Equity at the University of Southern California, and was attended by 32 people from agencies, such as US EPA, SCQAMD, and the Los Angeles Department of Health, staff from non-governmental organizations, and university researchers. CARB staff attended as well. In addition to presenting the final version of the EJSM, we also outlined improvements in the methods and final steps before completion of the project. We also discussed scoring comparisons using statewide versus regional quantiles and provided comparisons with results for CES. Feedback on the work was positive and some stakeholders had suggestions for additions to the hazard proximity layer, including locations of oil wells, and sites of non-conventional oil exploration, dairies and other sites. While adding these metrics is beyond the purview of this contract, our team is considering the viability of folding these data sources into the EJSM, including assessing the resources required to verify the locational accuracy of new datasets. Our final stakeholder meeting on May 13th was a webinar, hosted by CARB in Sacramento, to accommodate participants who were unable to travel. The meeting was well attended, with over 109 people in attendance remotely and 25 of whom were in the audience.

In addition to this report, we are preparing a manuscript for publication to highlight the key results from our new statewide analysis of the EJSM. Potential journal outlets include: *Environmental Health Perspectives, Environmental Health* or *International Journal for Environmental and Public Health Research*. We are prioritizing open-access journals for this manuscript to enable this work to be available to diverse audiences. In addition, we are collaborating with CARB staff to co-author another paper, currently in preparation, to apply the EJSM to analyze spatial patterns of temporal changes in air quality in the state.

7 Conclusion

The National Environmental Justice Advisory Committee, EJ advocates, and community organizations have long argued that scientists and regulatory agencies should incorporate the cumulative impacts of environmental and psychosocial stressors when ranking the priorities for regulatory enforcement activities instead of using the traditional chemical-by-chemical and source-specific assessments of potential health risks of environmental hazards, which do not reflect the multiple environmental and psychosocial stressors faced by vulnerable communities. These stakeholders have voiced their concern and have called for additional methods to consider and include cumulative impacts in developing regulatory and enforcement priorities. Regulatory agencies have responded to this need by embracing the National Research Council's call for the development of "cumulative risk frameworks" within their scientific programs and enforcement activities.

Spatial screening methods such as the EJSM have become key tools that can help decision-makers advance environmental justice goals by more efficiently targeting efforts and resources to remediate cumulative impacts, environmental inequities, and focus regulatory action at the neighborhood level. All too often, the burden of proof is placed on communities to demonstrate the cumulative impacts of environmental and social stressors and push for action. Cumulative impact screening such as the

EJSM provides environmental policy and programs with a more proactive approach that removes this burden from vulnerable communities so that those without an active environmental justice movement or capacity for civic engagement can also receive regulatory attention and protection. CARB's ongoing support of the development of the EJSM has facilitated the leveraging of our method to develop other key regulatory tools, such as CalEnviroScreen (OEHHA, 2016) and now US EPA's EJSCREEN (US EPA). During this contract period, all members of our research team have collaborated in different ways with both Cal-EPA and US EPA to provide advice, feedback, data, and methodological input into development these other regulatory screening tools which now being applied in innovate ways address cumulative impacts, for example to guide the implementation of SB535 and direct investments of Greenhouse Gas Reduction Fund monies generated through California's cap-and-trade program. Ultimately, as environmental health science develops a better understanding of cumulative impacts, standard approaches in risk assessment may need to change and be harmonized with cumulative impact screening methods to assure the protection of public health. Environmental and social stressors clearly converge in disadvantaged communities, and tools to measure these impacts are needed for improved decision-making to advance environmental equity. The use of cumulative impact screening approaches such as the EJSM increases the likelihood that disadvantaged communities may receive critical attention, improving existing conditions and reducing future harm.

FINAL REPORT: CONTRACT NUMBER # 11-336

Update and Statewide Expansion of the Environmental Justice Screening Method (EJSM)

SUPPLEMENTAL MATERIALS

Principal Investigators:

Rachel Morello-Frosch, Ph.D., M.P.H.
University of California, Berkeley
Department of Environmental Science Policy and Management &
School of Public Health

James Sadd, Ph.D.
Occidental College
Department of Geology and Environmental Science

Manuel Pastor Jr., Ph.D.
University of Southern California
Program on Environmental and Regional Equity
Departments of Geography and American Studies and Ethnicity

Supplement A: Using the National Land Cover Dataset (NLCD) with the EJSM.

One of the primary challenges of this project was producing geospatial data files on land use statewide, for construction of the CI Polygon layer, that are consistent and comparable. Land use information varies in quality, availability, and classification on a variety of geographic scales. To accomplish this, tax parcel information was used for many counties and portions of counties lacking quality land use information. Parcel records include use codes, which describe the current and allowed (by zoning) uses for each parcel, and parcel data is commonly used to automate or impute land use. However, among these various counties, there is an inconsistent definition of use code that made it difficult to produce an equivalent dataset for each county. The most obvious example is the classification of parcels composed primarily of agricultural land but with a small area with residential buildings as "rural residential", the same as small parcels in rural areas dominated by the residential structures. This was a particular problem in Merced and Stanislaus counties. In order not to over-represent farm fields as residential land, we use the 2011 National Land Cover Dataset (NLCD http://www.mrlc.gov/nlcd2011.php) to correct this type of problem identify and filter out land that did not fit our criteria. This extra processing is codified in the ModelBuilder programming that we used for preparing CI polygons, and was provided to CARB in the data transfer at the end of the project. In addition, the Project Manager has seen how this process works.

In spite of the best efforts of both CARB staff and our research team, we were unable to acquire tax parcel data with use codes for four counties – Trinity and Plumas (Northern California Region), and Amador and Alpine (Eastern Sierra Region). These are very sparsely populated counties with very little exposure to environmental hazards. For these counties we used the NLCD 2011 land cover data, classes 21-24, to create CI Polygon layers, which were subsequently validated by checking a random sample against the aerial imagery in Google Earth Pro. Sensitive land uses were identified by address geocoding and confirmation in Google Earth Pro. We also used this technique to correct large rural or agricultural parcels in other counties and have confidence in its utility and accuracy. It also represents an excellent way to update land use data in the future.

Our contract also included a test of using NLCD land cover data alone in the EJSM process. To do this, we completed processing the NLCD 2011 land cover data for the SCAG region, an area with very high quality and reliable land use information that we understand well. By converting rasters to vector layers (both polygons representing grid cells, and grid cell centers as points) for land cover classes 22, 23 and 24 (see **Figures SA1 and 2**, below). This land use layer was compared the SCAG 2008 land use geospatial data, using Google Earth Pro imagery. As can be seen in the two maps below for a portion of the Moreno Valley in Riverside County, the NLCD land use classes 22 and 23 do a very good job replicating the SCAG land use data.

This comparison demonstrated that in urban and suburban areas, land cover class 21 does not replicate residential or sensitive land uses well. We also performed a crosswalk using the grid cell centers (points) to assign a census block identifier to each grid cell, and identified grid cells

that represent sensitive, rather than residential, land uses by spatially joining our various sensitive land use data layers with the grid cells. The grid cell centers were then used to calculate hazard proximity metrics as with CI polygons, and the SCAG region EJSM metrics and scores were recalculated, to determine the impact of using NLCD land use data on scores derived.

This test revealed that there is little variation in EJSM scores calculated using the NLCD data vs. those calculated using the more standard SCAG 2008 land use information, as can be seen in the EJSM maps for Los Angeles and Riverside Counties, below (Figures C through H). Note that in areas of high EJSM scores, the variation is particularly low, indicating that using NLCD to update land use information in future iterations of the EJSM is both viable and would deliver significant savings in time and effort updating the EJSM going forward.

Testing the grid-data format was completed for the SCAG region as demonstration of its success and utility. The results were reported to CARB as part of the October 24, 2014 meeting and research presentation, and the data provided as part of the deliverables.

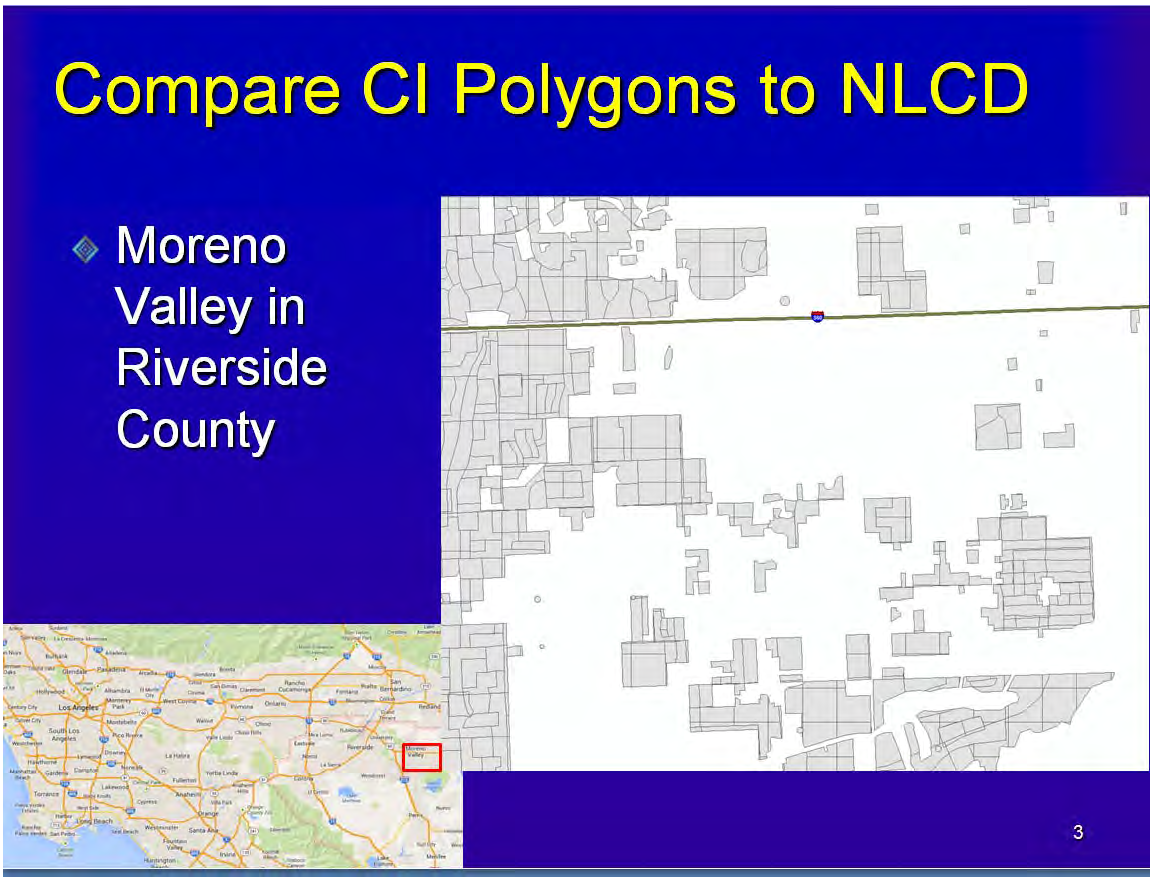


Figure SA.1. CI polygons derived from SCAG 2008 land use geospatial data layer for a portion of Riverside County.

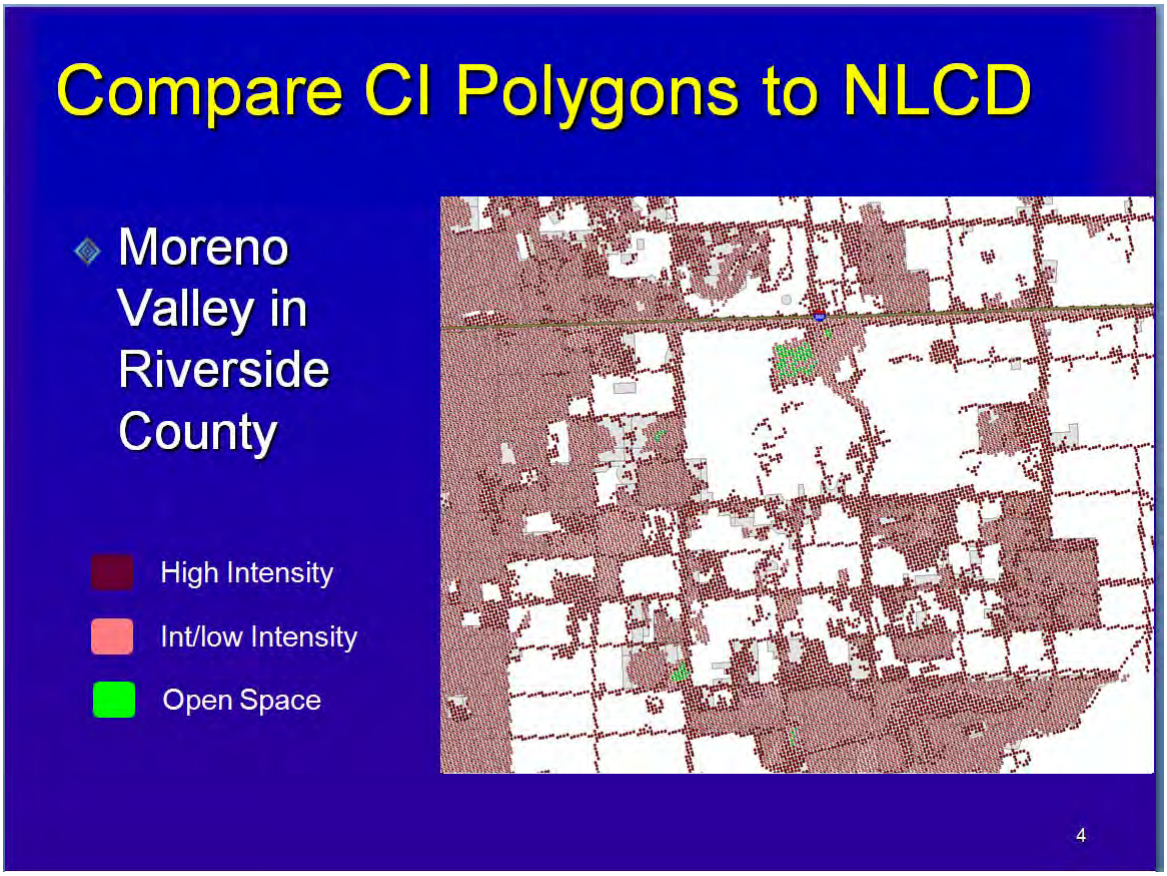


Figure SA 2: NLCD land use rasters corresponding to classes 22, 23, and 24.

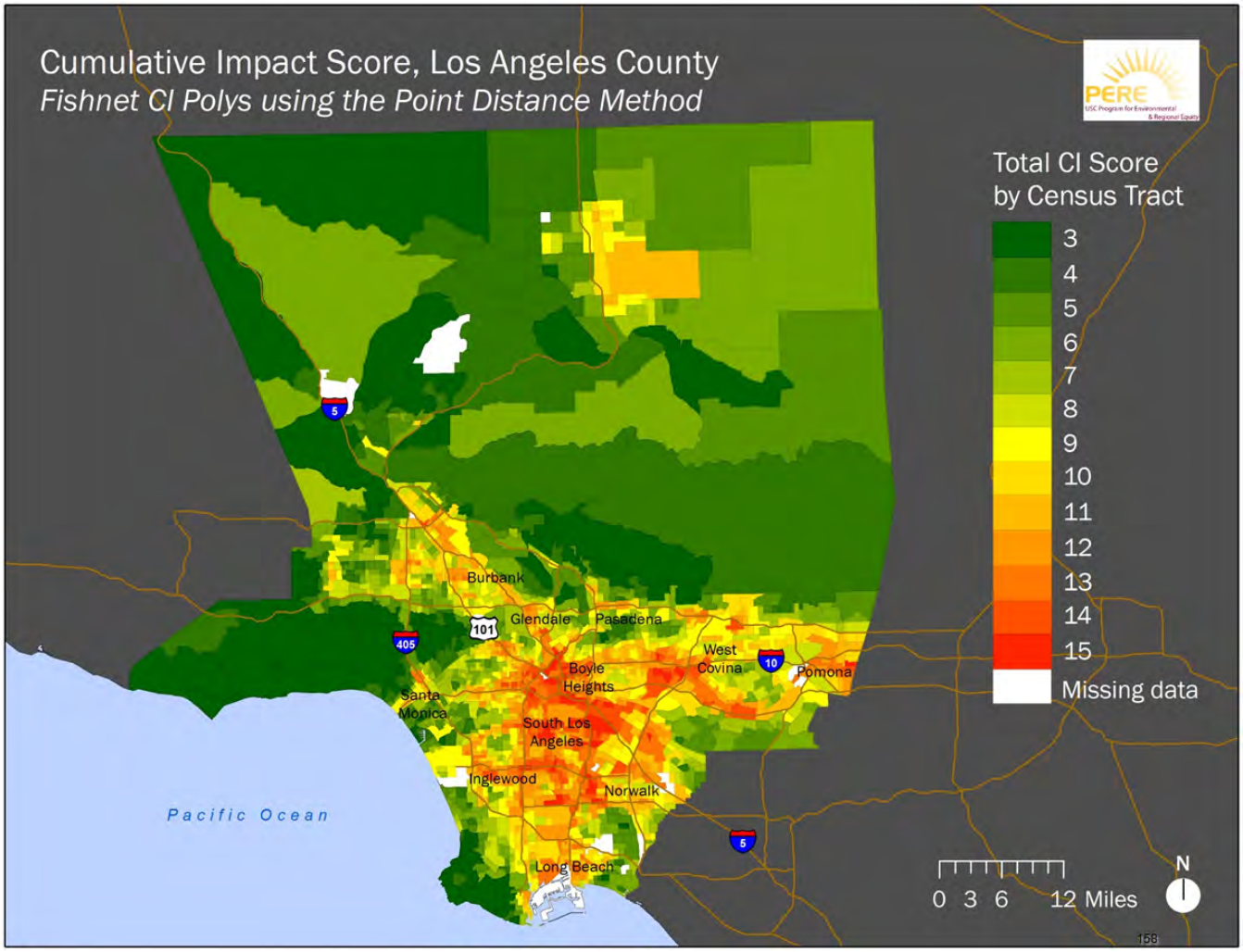


Figure SA.3. EJSM total score for Los Angeles County calculated using CI Polygons constructed using SCAG 2008 land use geospatial data.

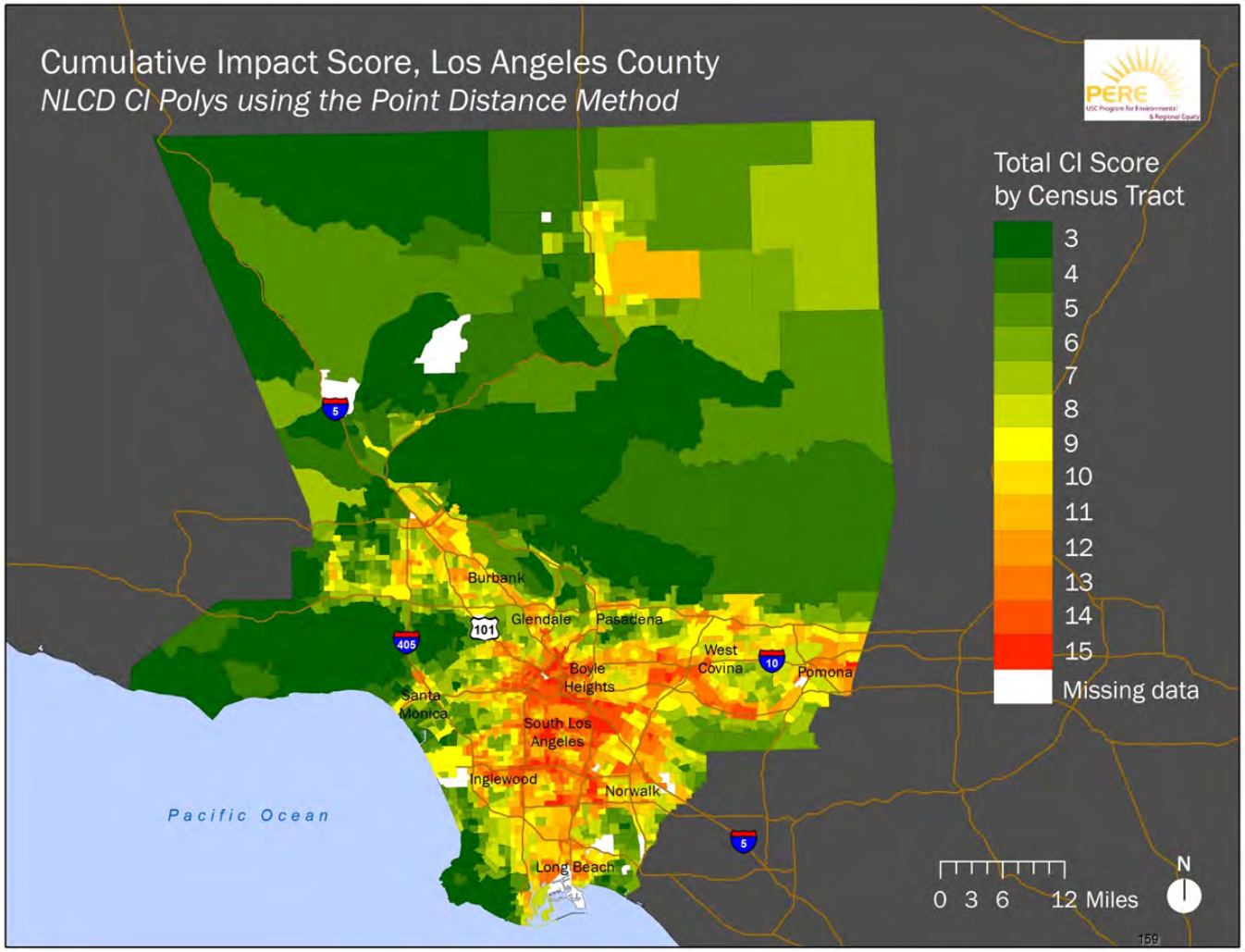


Figure SA4. EJSM total score for Los Angeles County calculated using CI Polygons constructed using NLCD 2011.

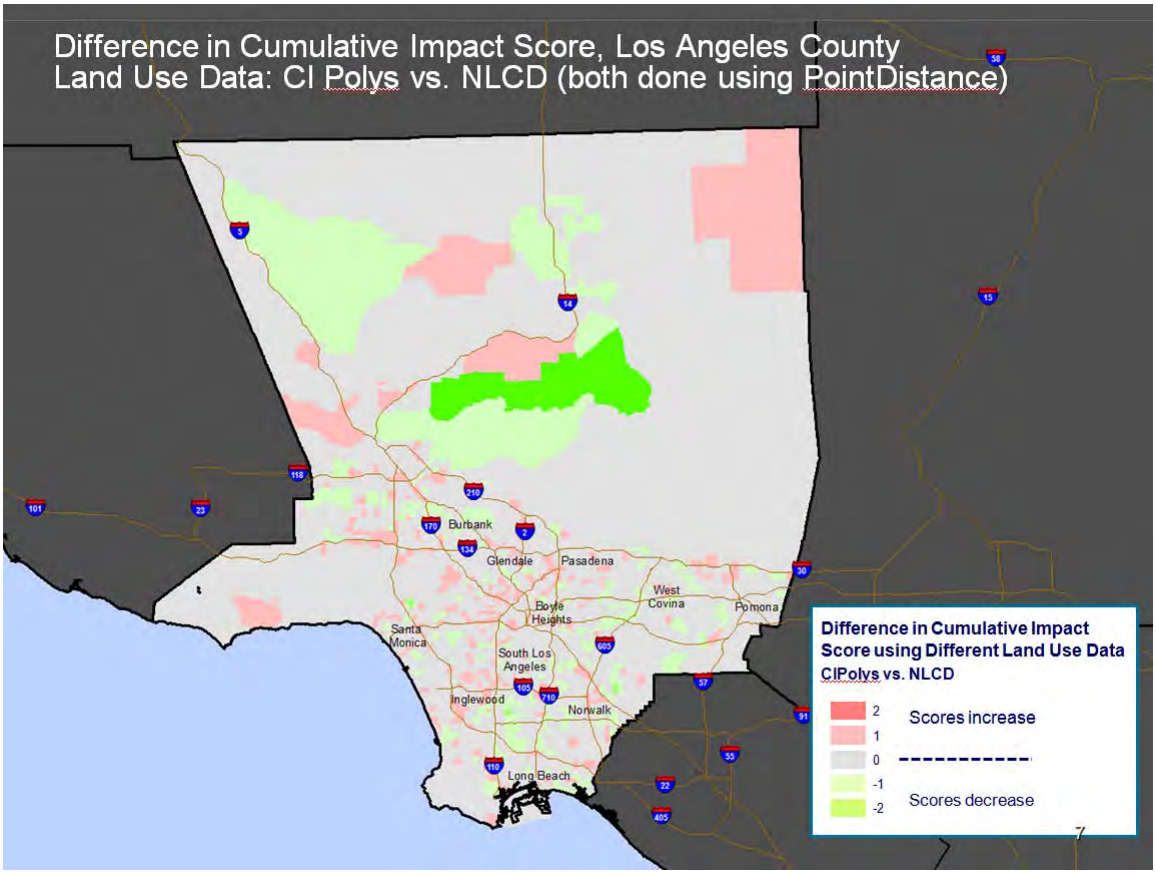


Figure SA5. Net change EJSM total score calculated using NLCD land use classes

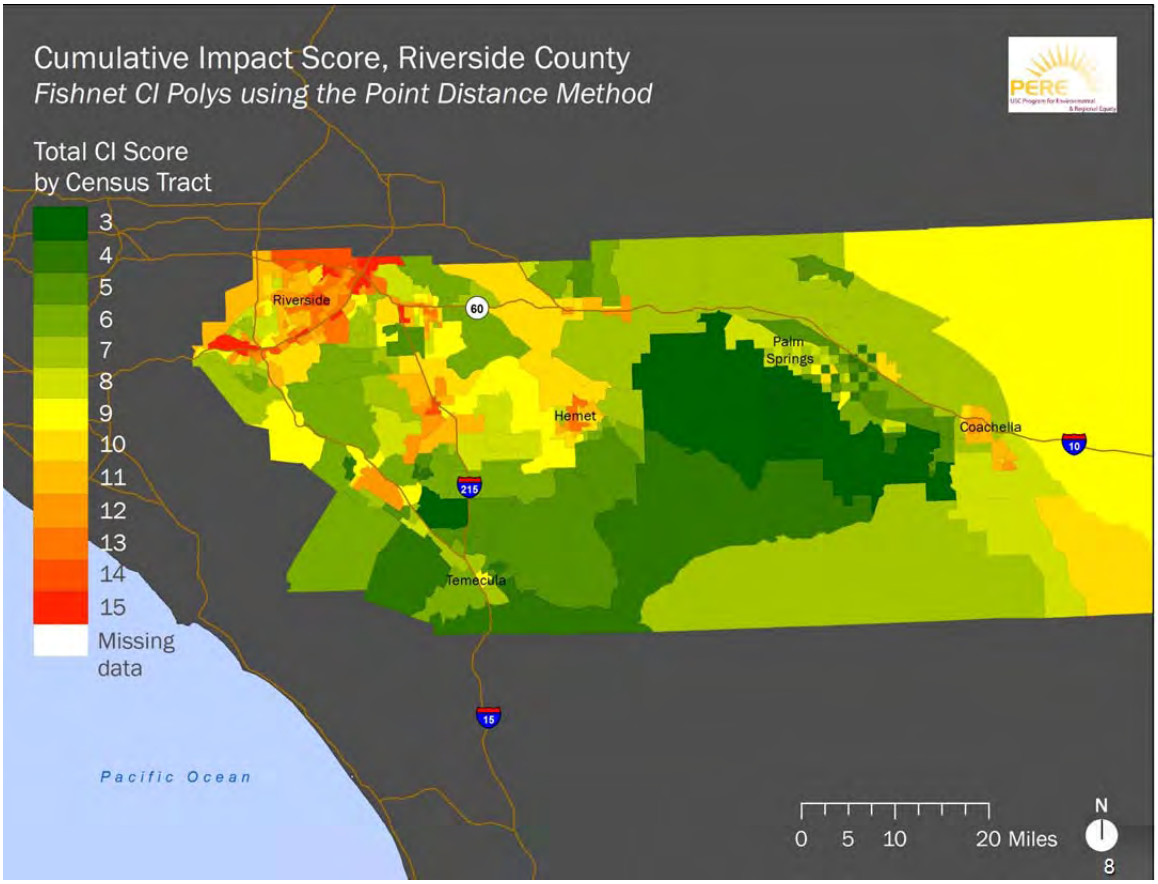


Figure SA6. EJSM total score calculated for Riverside County using CI Polygons constructed using SCAG 2008 land use geospatial data.

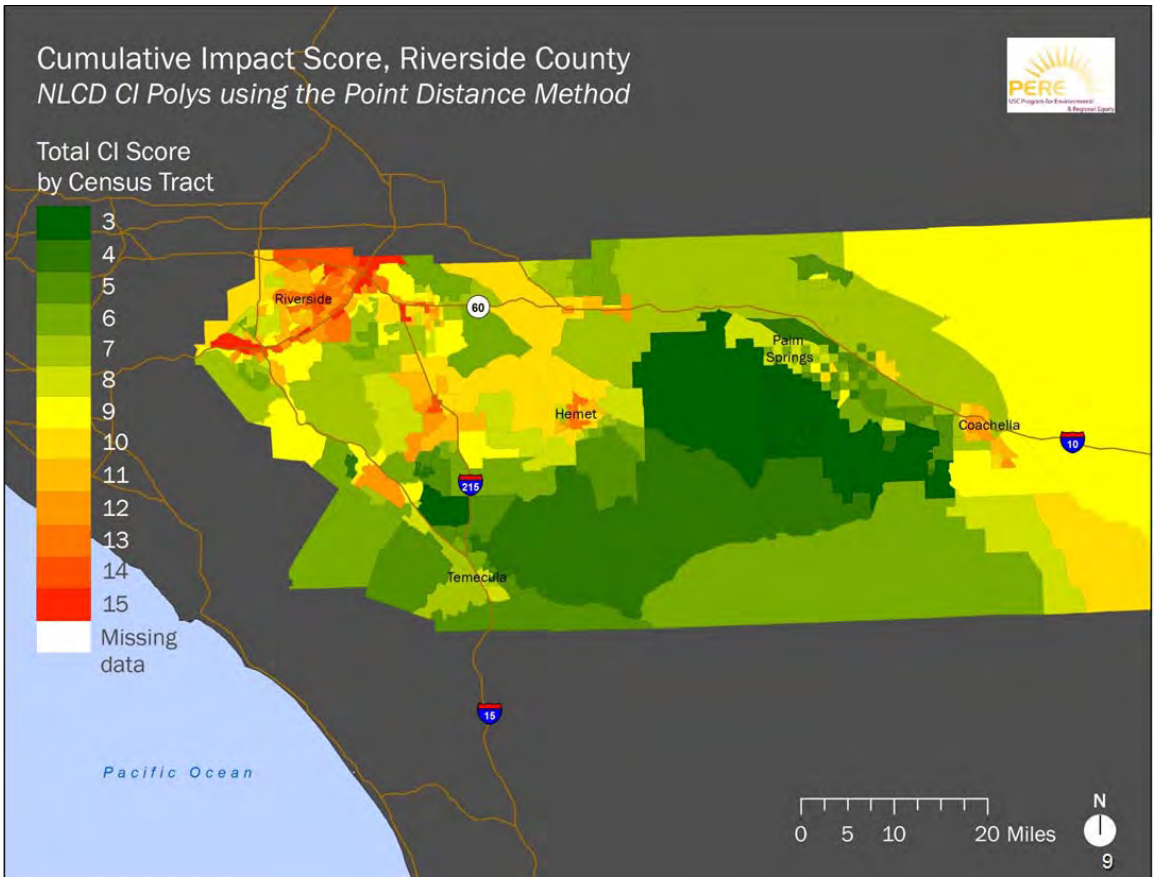


Figure SA.7. EJSM total score calculated for Riverside County using CI Polygons constructed using NLCD 2011

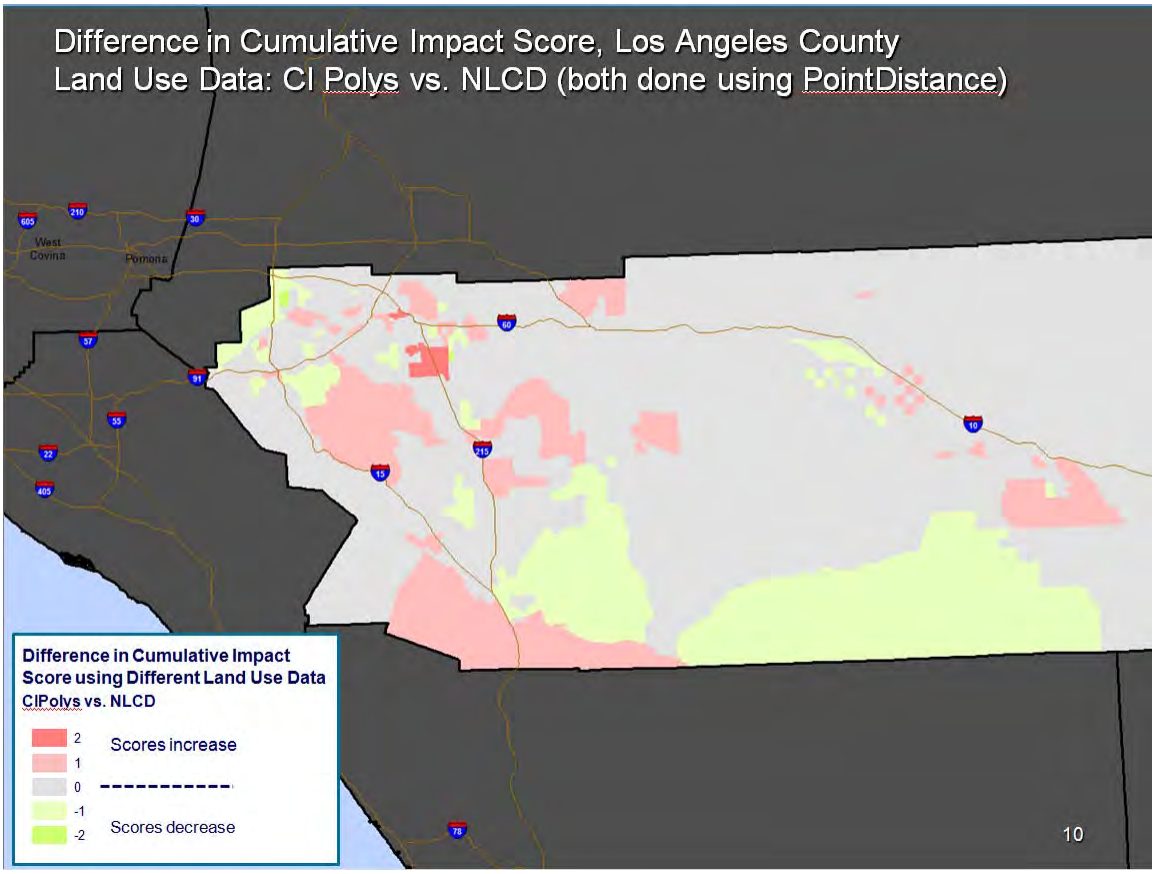


Figure SA.8. Net change EJSM total score calculated using NLCD land use classes

Supplement B: Locational accuracy of point source facility datasets

The original contract included the task of geocoding and error checking the CARB dataset of facilities required to report under the AB2588 “toxic hot spots: program, and using these facilities in the EJSM hazard proximity calculation. After completing this task in August 2013, we were asked to instead substitute it with a new GIS data layer - the “Facilities of Interest (FOI)” that was compiled by CARB staff because the Agency considers these facilities emitting criteria air pollutants and air toxics to better reflect significant threats to health from air pollution exposure than does AB2588, which has quite broad reporting requirements resulting in a large percentage of its facilities emitting very low air toxics loads and, thereby, complicating the inherent assumption of the EJSM method that considers all of these facilities equal in terms of exposure impact on nearby communities. AB2588 also contains a very large number of facility locations, with numerous challenges to locating them accurately.

The FOI is a subset of the facility emission inventory data extracted from CARB’s two primary emission inventory databases: the Greenhouse Gas (GHG) Mandatory Reporting database, and the California Emission Inventory Development and Reporting Systems (CEIDARS) database. CEIDARS contains the emission estimates for both the criteria and toxic air pollutants and includes the AB2588 facilities. The GHG database includes large facilities that must report under mandates from the AB32 Climate Change program emit significant quantities of air toxics as co-pollutants. The FOI layer and facilities is a work in progress, and is based on a draft set of parameters used to identify the facilities that account for the majority of the reported point-source emissions of GHGs, criteria pollutants, and air toxics. In keeping with the design philosophy requirements of the EJSM, all the data included in the FOI layer are publicly available.

A working version of the FOI was delivered to us with point locations provided by CARB. We conducted a preliminary evaluation of the locational accuracy and breadth of FOI, and compared it to the original AB2588 dataset to assess the extent to which they captured different facility types and any systematic differences. Details of this facility comparison evaluation were discussed extensively in our Q3 report. After more consultation with CARB staff regarding the results of our evaluation, they agreed to provide the research team with industry-wide data layers that included auto-body shops and gas stations to extend the FOI data layer to include the types of facilities identified as requiring distance buffers for separation from sensitive land uses in the CARB guidance on land use and air pollution sources (CARB, 2005). We believed that these additional facility categories were important to integrate, based on our prior ground-truthing work in the Bay Area and Southern California. We received these industry-wide data layers in late March 2014 and integrated them along with the FOI data into the EJSM as a preliminary draft to visualize the revised results in our hazard proximity layer with these new facility datasets.

After extensive evaluation of the locational accuracy of FOI and the industry-wide datasets, the research team discovered problems with the locational accuracy of a significant portion of the facilities included in the FOI dataset, with errors ranging from hundreds of feet to several miles. CARB staff worked further on the FOI dataset to check and correct facility locations, and sent a revised version of FOI for use in this project.

We checked all FOI locations using a variety of validation methods, including internet searches to verify facility information, and Google Earth Pro to visually inspect each reported location, and its geocoding capability for address matching. This allowed us to verify (and, in some cases, correct) the locations of some facilities. There are surely many reasons for these inaccuracies, among these are the practice of expressing facility locations using geographic coordinates with only three significant digits of precision, rather six or seven digits required for accurate positioning, and obtainable using industry standard address matching GIS layers. The layer also included a considerable number of sites that could not be address geocoded due to missing addresses, address information that is not geographically specific enough for mapping, or address errors.

Because the location error rate appeared to be significantly high, and the amount of apparent error often equaled or exceeded the distance buffers used in determining hazard proximity, the decision was made jointly with CARB for the research team to error-check the entire FOI layer and make appropriate corrections. This was accomplished using Google Earth Pro on duplicate datasets. The FOI shapefile was converted to .kml Placemarks, which were checked individually (one by research staff at USC, the other by staff at Occidental College) using the facility name and/or address search capability, tax parcel information overlays, and aerial imagery provided within Google Earth Pro. In this way, each point location was validated and relocated as needed; using either actual facilities or parcel centers as the corrected location. The resulting location-checked shapefile was sent to CARB and was used to develop the EJSM scores reported in this report.

Based on the results of FOI location correction, we also checked and corrected the DTSC hazardous waste facility point shapefile provided by the OEHHA CalEnviroScreen team for incorporation into the EJSM hazard proximity score. This file also had a high error rate, with 16% of sites located inaccurately within the 1000 to 3000 foot range, and 4.2% inaccurate by more than 3000 feet (see **Figure SB.1** and **Table SB.1**, below). It also contained two duplicate sites, and four facilities that could not be found or verified using Google Earth Pro and other internet searches. It is also important to note that many of these facilities are very large, and representing them as points in calculating proximity metrics will result in error and misclassification. It would be best to measure proximity using distance from the boundary of these large facilities, and our testing found this had a significant effect on the hazard proximity score. We automated these facilities as polygons, and included them in the data provided to CARB. These polygons and location corrections were also sent to OEEHA.

The locational accuracy of the industry-wide facilities data provided by CARB (auto paint and body shops, gas stations) were also checked using a 10% random sample of each dataset to examine error rate and degree. The >1000 ft error rates for gas stations (3%) and auto paint and body shops (5.5%) were much lower than hazardous waste and FOI facilities. We elected not to correct these entire datasets because (a) of the very large number of sites (9682 and 3701); (b) the fact that these are mostly urban businesses with geocodable street addresses gave us more confidence in the accuracy of the dataset overall, and (c) we lacked the resources under this contract to correct these large datasets.

The California Department of Toxic Substances Control lists 17 permitted hazardous waste handling facilities and generators located in the San Joaquin Valley. Ground truth validation

demonstrated significant locational error for most of these sites, with most locations off by well over 100 meters (see **Table SB1 and Figures SB-1**).

Figure SB.1. California Department of Toxic Substances Control permitted hazardous waste handling facilities and generators located in the San Joaquin Valley showing selected location corrections.

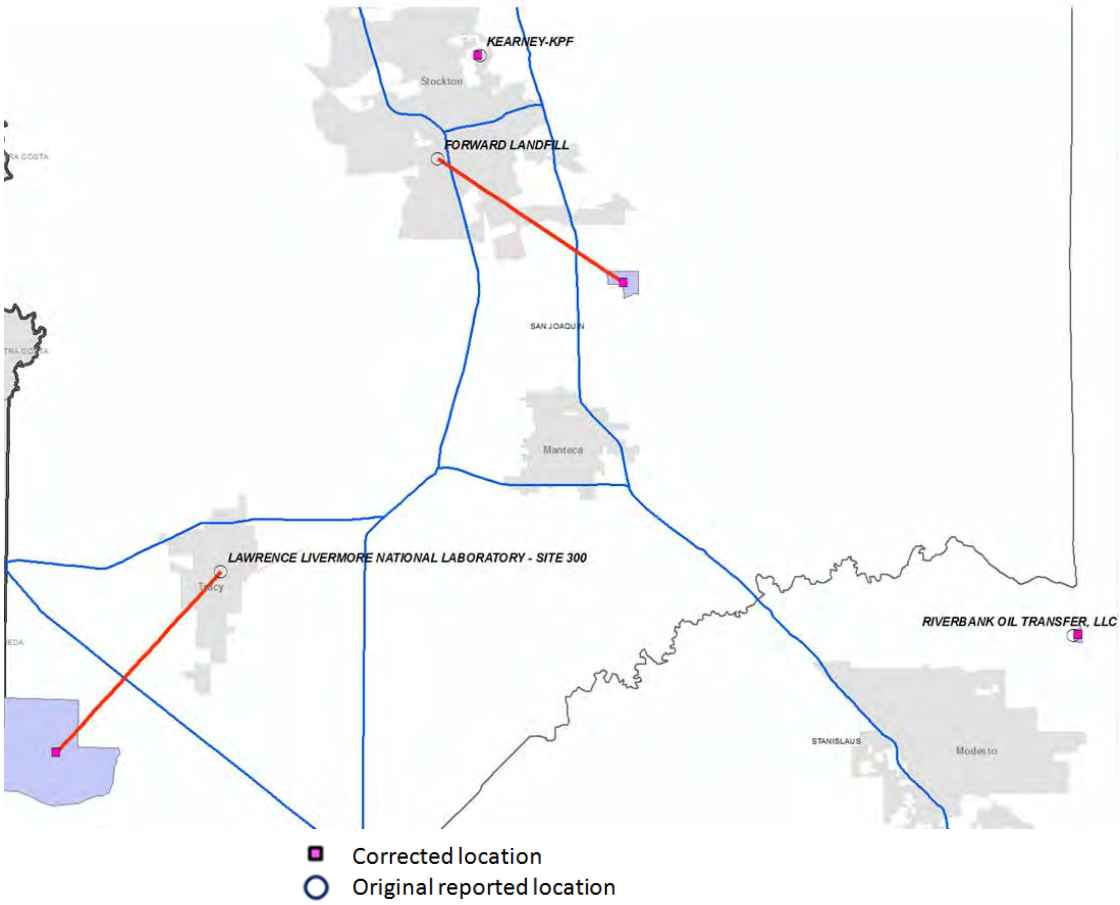


Table SB.1. Summary of locational error for California Department of Toxic Substances Control permitted hazardous waste handling facilities and generators located in the San Joaquin Valley.

EPA_ID	PROJECT NAME	ADDRESS	CITY	Error (m)
CA2890090002	LAWRENCE LIVERMORE NATIONAL LAB - SITE 300	CORRAL HOLLOW RD	TRACY	12,764
CAD990794133	FORWARD LANDFILL	9999 S AUSTIN RD	STOCKTON	11,705
CA1570024504	EDWARDS AIR FORCE BASE	5 E POPSON AVE	EDWARDS	1,519
CA4170024414	OCCIDENTAL OF ELK HILLS INC	28590 HIGHWAY 119	TUPMAN	1,500
CAD980813950	CRANE'S WASTE OIL INC	16095 HIGHWAY 178	WELDON	614
CAT000646117	CHEMICAL WASTE MANAGEMENT INC KETTLEMAN	KETTLEMAN HILLS LDFL HWY 41	KETTLEMAN CITY	478
CAL000190816	RIVERBANK OIL TRANSFER, LLC	5300 CLAUS RD	RIVERBANK	238
CAL000282598	BAKERSFIELD TRANSFER INC	1620 E BRUNDAGE LN	BAKERSFIELD	231
CA2170023152	NAVAL AIR WEAPONS STATION CHINA LAKE	1 ADMINISTRATION CIR	RIDGECREST	188
CAD982446882	EVERGREEN OIL INC FRESNO	4139 N VALENTINE AVE	FRESNO	144
CAD066113465	SAFETY-KLEEN	3561 S MAPLE AVE	FRESNO	115
CAD981429715	KEARNEY-KPF	1624 E ALPINE AVE	STOCKTON	107
CAL000102751	WORLD OIL - SAN JOAQUIN LLC	14287 E MANNING AVE	PARLIER	99
CAT080010606	BIG BLUE HILLS PESTICIDE CONT DISPOSAL	10 MILES NORTH OF COALINGA	COALINGA	76
CAD982435026	KW PLASTICS OF CALIFORNIA	1861 SUNNYSIDE CT	BAKERSFIELD	34
CAT080010283	EPC WESTSIDE DISPOSAL FACILITY	26251 HIGHWAY 33	FELLOWS	33
CAD980675276	CLEAN HARBORS BUTTONWILLOW LLC	2500 WEST LOKERN RD	BUTTONWILLOW	21