SENSITIVITY ANALYSES OF THE CALENVIROSCREEN MODEL AND INDICATORS

CALIFORNIA ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF ENVIRONMENTAL HEALTH HAZARD ASSESSMENT
JUNE 2013

INTRODUCTION

The California Communities Environmental Health Screening Tool (CalEnviroScreen)[1] is designed by the Office of Environmental Health Hazard Assessment (OEHHA) and the California Environmental Protection Agency (Cal/EPA) to identify California's most pollution-burdened and vulnerable communities. It uses a practical methodology to evaluate multiple pollution sources and stressors in California's approximately 1800 ZIP codes using data from federal and state sources. Finalized in April 2013, the CalEnviroScreen 1.0 model is made up of four components in two broad groups. Exposure and Environmental Effects components comprise a Pollution Burden group, and the Sensitive Populations and Socioeconomic Factors components comprise a Population Characteristics group. The four components organize environmental, health and socioeconomic data from 18 indicators (see Figure 1). The CalEnviroScreen score is calculated by combining the individual indicator scores within each component, then multiplying the Pollution Burden and Population Characteristics scores to produce a final score. The indicators and general scoring method are shown in Figure 1, below. For more information, see the Method section of the CalEnviroScreen 1.0 report.[1]

During the development of CalEnviroScreen, OEHHA received comments from the public and from a panel of scientific experts convened in September 2012 recommending a sensitivity analysis to test how much the community rankings or categories would change if various factors in the model were altered, ranging from individual indicators to the overall mathematical approach. In response to these suggestions, we applied a series of statistical analyses, as well as simpler numerical and graphical comparisons, to answer questions about the behavior of the model and the confidence we place in the results.[2,3]

The analyses we conducted enabled us to systematically compare our draft model to alternative versions, and helped to inform the final decisions incorporated into the CalEnviroScreen Version 1.0. Because the primary purpose of CalEnviroScreen is to identify communities with high pollution burdens and vulnerabilities, we paid particular attention throughout this analysis to the top 10 percent of the 1769 ZIP codes evaluated (177 ZIP codes)

---

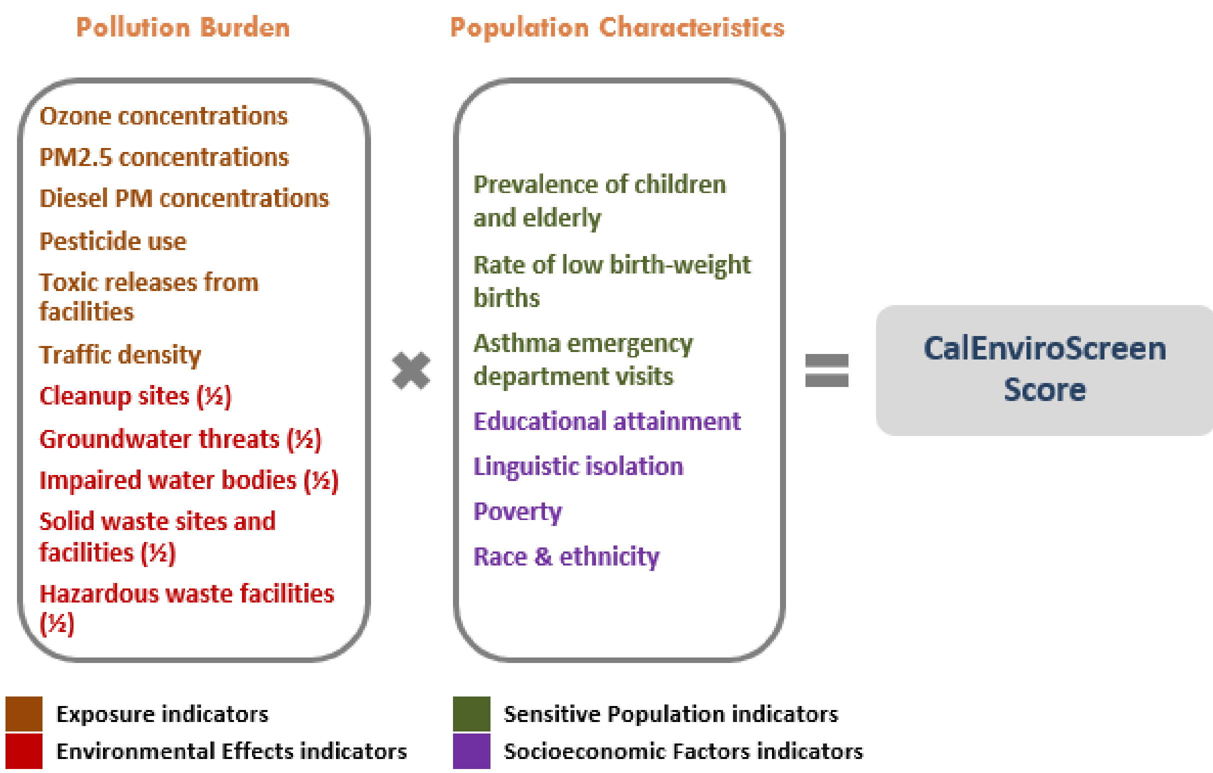
[1] California Communities Environmental Health Screening Tool, Version 1 (CalEnviroScreen 1.0). Guidance and Screening Tool. Office of Environmental Health Hazard Assessment and the California Environmental Protection Agency, Sacramento, CA http://www.oehha.ca.gov/ej/ces042313.html. Available in English and Spanish.

[2] Response to Major Comments on the CalEnviroScreen Public Review Drafts, http://www.oehha.ca.gov/ej/pdf/042313ResptoCom.pdf.

[3] Summary of Technical Comments from Panel of Academic Experts, http://www.oehha.ca.gov/ej/pdf/SummaryAcademicWorkshop090712.pdf.

with the highest CalEnviroScreen scores. The sensitivity analysis presented here is based on the January 2013 draft CalEnviroScreen data and results,[4] which contained 17 indicators, compared to 18 indicators in CalEnviroScreen 1.0. The results of the sensitivity analysis on the January 2013 draft that informed the finalization of Version 1.0 of the CalEnviroScreen are presented in this document. Here we respond to questions we received during the public comment periods and from the academic review panel. Descriptions of the specific methods we used are described in the Appendix.

Figure 1. CalEnviroScreen 1.0 indicators and scoring.



[4] California Communities Environmental Health Screening Tool. Second Public Review Draft. Available at http://www.oehha.ca.gov/ej/cipa011613.html

How can we be sure that the chosen indicators are not duplicative? Specifically, does the CalEnviroScreen model double- or triple-count air pollution or other indicators?

<u>Correlations between indicators</u>

One way to consider whether the indicators are duplicative is to look at correlations between them. We used Spearman's correlations coefficients to examine the relationship between the 17 indicators in the January 2013 CalEnviroScreen model (See Appendix, Method I).

There are multiple reasons that two indicators may be correlated. These include the possibility that the measure of one indicator reflects something that is caused by the other (for example, frequency of lightning and that of thunder). Another is that there are some similar underlying factors that explain the measure of both indicators (for example, a common source for a pair of pollutants). The role of chance in contributing to correlations cannot always be excluded, either. For these reasons, calculating correlation coefficients is an early step in examining possible relationships between indicators.

Table 1 shows the correlation coefficients between the raw scores for individual indicators with indicators that are highly or moderately correlated shown in bold. Indicators with correlation coefficients between -0.5 and +0.5 are weakly correlated. Correlation coefficients between 0.5 and 0.8 (either positive or negative) are moderately correlated, and those with coefficients below -0.8 or above 0.8 are highly correlated. Linguistic isolation and race/ethnicity were the only two indicators that were highly correlated ($\rho = 0.82$). Diesel particulate matter (PM) was moderately correlated with ambient PM2.5 ($\rho = 0.71$) and traffic ($\rho = 0.62$). However, traffic was only weakly correlated with PM2.5 ($\rho = 0.39$), which indicates that although there is some interrelationship between diesel PM, PM2.5, and traffic, these indicators do not duplicate each other. Ozone was only weakly correlated with PM2.5 and was not at all correlated with diesel or traffic. Cleanup sites and groundwater threats were moderately correlated with each other and with traffic ($\rho = 0.62$ to 0.66). The education indicator was also moderately correlated with the other socioeconomic factors ($\rho = 0.61$ to 0.67). Age was the indicator that was most consistently, although weakly, negatively correlated with other indicators.

Our analysis using Spearman's correlation coefficients found that only one pair out of the total of 17 indicators were highly correlated with each other. A larger number of individual indicators were moderately or weakly correlated. These findings are reassuring in that they do not indicate significant double- or triple- counting in the choice of indicators. Within the exposure component, the moderate correlations between diesel PM, PM2.5 and traffic were not surprising since traffic is a major source of diesel PM and diesel PM contributes to ambient PM2.5. However, the lack of a strong correlation indicates that the indicators are not duplicative. Instead it is likely that certain non-road locations (for example, ports and rail yards) are hot spots of diesel exhaust emissions that are not necessarily captured by either the PM2.5 or traffic indicators. In addition, diesel exposure and total PM2.5 exposure are known to have distinct effects on health due to differences in sources and composition (Fujita et al., 2007a; Fujita et al., 2007b; Rogge et al., 1993).

We found that the linguistic isolation indicator was highly correlated with the race/ethnicity indicator ($\rho = 0.82$). ZIP codes with higher proportions of immigrants are likely to score high for both indicators, but some communities of color, notably those with a high proportion of African American or Native American residents, are less likely to be linguistically isolated. These communities are captured by the race/ethnicity indicator but missed by the linguistic isolation indicator. Lack of facility in English, on the other hand, brings with it some unique challenges, including difficulty getting medical treatment, understanding care instructions, or affecting policy decisions through civic engagement. This sensitivity analysis indicates that ZIP codes that are high in both linguistic isolation and race/ethnicity are given somewhat higher weight in the CalEnviroScreen. Further research could examine this relationship in greater depth.

Age is the only indicator that was generally negatively correlated with other indicators, although only weakly. This indicator, which includes both the proportion of children under 10 and the proportion of adults over 65 years of age, warrants further examination in future versions of CalEnviroScreen. There may be other population attributes or features that are surrogates for age that could better capture vulnerability related to age. In particular, we will consider whether disparities in life expectancy, which can vary by over a decade across racial and socioeconomic groups and neighborhoods (Burd-Sharps & Lewis, 2011), may bias this indicator towards identifying more advantaged communities, and whether this indicator may be unduly influenced by the existence of some popular wealthy retirement communities. We will continue to review the age indicator and consider changes in the future.

Table 1. Spearman's correlation coefficients (ρ) between indicator raw scores.*

			Ozone	PM2.5	Diesel PM	Pesticides	TRI	Traffic	Cleanup sites	Groundwater threats	Impaired water bodies	Waste sites	Age	Asthma	Low Birth Weight	Education	Linguistic isolation	Poverty	Race/ ethnicity
Pollution Burden	Exposures	Ozone	1																
		PM2.5	0.44	1															
		Diesel PM	0.10	**0.71**	1														
		Pesticides	0.07	0.05	-0.07	1													
		TRI	-0.01	0.16	0.17	0.00	1												
		Traffic	0.04	0.39	**0.62**	0.12	0.23	1											
	Environmental Effects	Cleanup sites	-0.03	0.34	0.48	0.10	0.37	**0.62**	1										
		Groundwater threats	-0.15	0.21	0.36	0.15	0.24	**0.62**	**0.66**	1									
		Impaired water	-0.31	-0.07	0.06	0.26	0.05	0.24	0.22	0.34	1								
		Waste sites	0.07	0.11	0.04	0.15	0.27	0.25	0.37	0.42	0.16	1							
Population Characteristics	Sensitive Populations	Age	0.08	-0.22	-0.38	0.11	-0.06	-0.24	-0.22	-0.15	-0.05	-0.01	1						
		Asthma	0.11	0.03	-0.07	-0.06	0.20	-0.12	0.07	0.06	-0.09	0.09	0.05	1					
		Low Birth Weight	0.11	0.28	0.32	-0.16	0.02	0.22	0.15	0.08	-0.06	0.01	-0.12	0.06	1				
	Socioeconomic Factors	Education	0.21	0.30	0.16	0.27	0.20	0.16	0.25	0.23	0.01	0.23	-0.01	0.40	0.03	1			
		Linguistic isolation	0.06	0.46	**0.56**	0.20	0.18	**0.50**	0.47	0.43	0.11	0.14	-0.21	0.10	0.17	**0.61**	1		
		Poverty	0.17	0.14	-0.04	0.08	0.14	-0.08	0.10	0.11	-0.09	0.16	0.05	**0.50**	-0.03	**0.67**	0.37	1	
		Race /ethnicity	0.12	**0.53**	**0.57**	0.21	0.28	**0.50**	0.49	0.41	0.07	0.18	-0.28	0.22	0.24	**0.65**	**0.82**	0.38	1

*Spearman's correlation coefficient measures the degree to which two indicators tend to vary together. Values near -1 mean the indicators are strongly inversely related. Values of 1 mean the indicators are positively correlated. Values of 0 mean there is no clear relationship between the indicators. Strong and moderate correlations are shown in bold. Pairs with missing values were omitted from the analysis.

Are any individual indicators unduly "driving" the results of the model by more substantially affecting the overall score? Conversely are there any indicators that don't really contribute much to the overall score?

We used the degree of change in the top 10 percent of ZIP codes to compare the CalEnviroScreen results to results when individual indicators were removed (See Appendix, Method II). Fagin's G Measure and Inverse Rank Measure (IRM) M were used to compare the top 10 percent results of the CalEnviroScreen model to results when individual indicators were removed (see Appendix, Method IV).

Table 2 shows the number and percent of ZIP codes that fall in or out of the top 10 percent in the CalEnviroScreen when individual indicators were removed from the model. Removing an exposure indicator generally resulted in more changes than removing an environmental effects indicator or one of the population characteristics indicators. Other statistical measures were also lower when an exposure indicator was removed, indicating that the rankings produced without exposure indicators were less similar to those of the full model.

When we removed various individual indicators from the model, between about 4 and 16 percent of the ZIP codes making up the top ranks changed. The indicators that caused the largest changes to the top-ranked ZIP codes were TRI and diesel PM, and those that caused the least change were education and poverty. This analysis suggests that each of the indicators makes a contribution to the model. On the other hand, when individual indicators were removed, the values of Fagin's G and IRM M remained close to 1, providing evidence that the ordering of ZIP codes at the top was not overly affected by any one indicator. That is, no single indicator is exceptionally influential in determining the model's results, yet all indicators contribute something to the overall model.

As a result of comments received during the public participation process for developing CalEnviroScreen, diesel PM and linguistic isolation indicators were added to the model. From the results in Table 2, we observed that leaving the diesel PM indicator out of the analysis results in a 14.7 percent change in the top 10% of ZIP codes identified, among the highest level of change by omitting a single indicator. This finding supports the concept that the diesel PM indicator represents additional contributions that are not encompassed by the other indicators.

Table 2. Influence of individual indicators on top 10 percent of ZIP codes.

Indicator Removed	Degree of change		Statistical Measures	
	No. of ZIP codes	%	Fagin's G*	IRM M*
TRI	28	15.8	0.85	0.79
Diesel PM	26	14.7	0.88	0.89
Ozone	23	13	0.84	0.77
PM 2.5	23	13	0.89	0.91
Pesticides	23	13	0.83	0.77
Asthma	21	11.9	0.90	0.90
Low Birth Weight	20	11.3	0.88	0.84
Traffic	19	10.7	0.90	0.92
Cleanup sites	14	7.9	0.94	0.96
Impaired water bodies	12	6.8	0.91	0.90
Waste sites	11	6.2	0.92	0.93
Linguistic isolation	11	6.2	0.94	0.94
Age	10	5.6	0.90	0.89
Race/ethnicity	10	5.6	0.94	0.95
Groundwater threats	9	5.1	0.94	0.95
Poverty	9	5.1	0.94	0.93
Education	8	4.5	0.94	0.96

*Fagin's G and IRM M are measures that compare rankings within two sets, here, a comparison of the top 10% using two different models. Values near 0 indicate that ZIP codes are not ranked similarly in the model with the specified indicator removed compared to the full CalEnviroScreen model, while values closer to 1 indicate that ZIP codes are similarly ranked in both models. See text for further details.

The CalEnviroScreen model includes groups of indicators representing Pollution Burden and Population Characteristics (including socioeconomic factors). Does one of these groups unduly influence the CalEnviroScreen score? Specifically, does the CalEnviroScreen essentially reflect poverty and other socioeconomic factors?

To compare the CalEnviroScreen model to alternative models in which groups of indicators were removed, we again used a measure of the degree of change in the top 10 percent of ZIP codes to compare the CalEnviroScreen results to the results of the alternative models (see Appendix, Method II). The CalEnviroScreen Model is expressed below:

CalEnviroScreen Model (multiplicative model)
$CalEnviroScreen\ score_{0-100} = Pollution\ Burden_{0-10} \times Population\ Characteristics_{0-10}$

Several other models in which groups of indicators are removed are expressed below as Alternative Models 1 through 3:

Alternative Model 1: Pollution Burden alone
$Model\ Score_{0-10} = Pollution\ Burden_{0-10}$
Alternative Model 2: Population Characteristics alone
$Model\ Score_{0-10} = Population\ Characteristics_{0-10}$
Alternative Model 3: Socioeconomic Factors alone
$Model\ Score_{0-10} = Socioeconomic\ Factors_{0-10}$

Fagin's G and IRM M were used to compare the top 10 percent results of Alternative Models 1 through 3 to the CalEnviroScreen model (see Appendix, Method IV).

Leaving out all Pollution Burden, Population Characteristics, or Socioeconomic Factors indicators led to more significant changes than removing single indicators (38 to 51 percent change in the top 10 percent of ZIP codes), with G and M scores considerably less than 1.0 (see Table 3). Ranking ZIP codes by Pollution Burden alone resulted in a smaller change (higher G and M) than did ranking ZIP codes by Population Characteristics or the Socioeconomic Factors indicators alone, indicating that pollution is overall more important in the final score than are the various population and socioeconomic factors. However, the removal of all Population Characteristics indicators still resulted in a substantial difference among the highest-ranked ZIP codes (G = 0.59, M = 0.36), showing that these factors are also important in the final score. This analysis supports a view that no single group of measures or components exerts the primary influence on the results and that none is likely to be an adequate substitute for the full model.

Table 3. Influence of different models on the top 10 percent of ZIP codes.

		Degree of Change		Statistical Measures	
	Alternative Model	No. of ZIP codes	%	Fagin's G*	IRM M*
CalEnviroScreen score compared to …	Pollution Burden model	67	38	0.59	0.36
	Population Characteristics model	87	49	0.43	0.27
	Socioeconomic Factors model	90	51	0.35	0.11

*Fagin's G and IRM M are measures that compare rankings within two sets. Values near 0 indicate that ZIP codes are not similarly ranked in the model with the specified indicator removed, while values closer to 1 indicate that ZIP codes are similarly ranked.

Why not simply add up the scores instead of multiplying the Pollution Burden scores by the Population Characteristics scores? (CalEnviroScreen compared to an additive model)

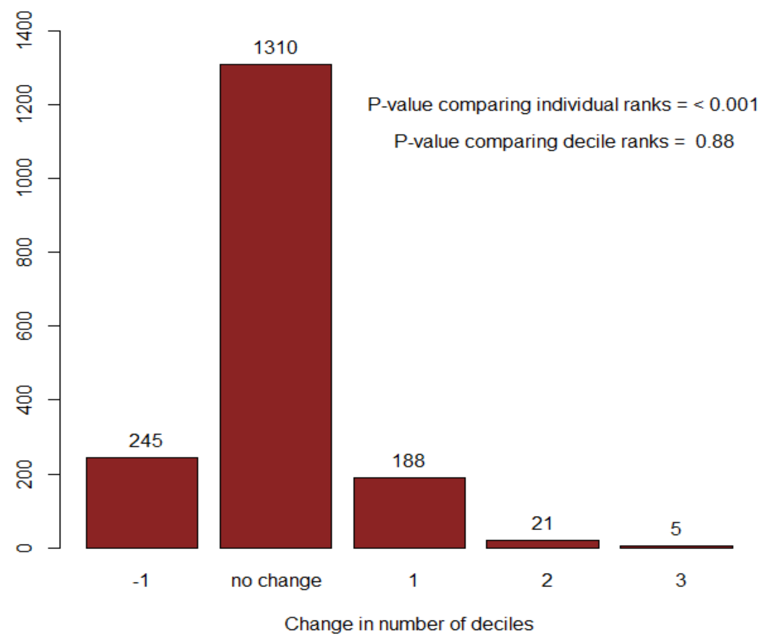
Following the initial release of the July 2012 CalEnviroScreen draft, we received a number of public comments suggesting an "additive model" in which the Pollution Burden and Population Characteristic scores are added together, rather than multiplied. Here, we compared the results of the January 2013 CalEnviroScreen multiplicative model with an additive model based on the equation below.

Alternative Model 4: Additive model

$$\boldsymbol{Model\ Score_{0-20} = Pollution\ Burden_{0-10} + Population\ Characteristics_{0-10}}$$

We compared the CalEnviroScreen model to an additive model using decile ranks for the ZIP codes because we are less concerned with small incremental changes in ZIP code rankings than we are with large changes in rankings. A decile refers to a 10 percent group, where the ZIP codes are divided into ten equal groups based on their model score. In other words, the highest scoring 10 percent of ZIP codes are in the highest decile, the next highest scoring 10 percent of ZIP codes are in the next-highest decile, and so forth. Figure 2 shows the degree of reordering of ZIP codes among deciles that resulted from adding the Pollution Burden and Population Characteristic scores, rather than multiplying them.

Figure 2. Degree of change in the decile rank of all California ZIP codes for the additive model compared to CalEnviroScreen (January 2013).



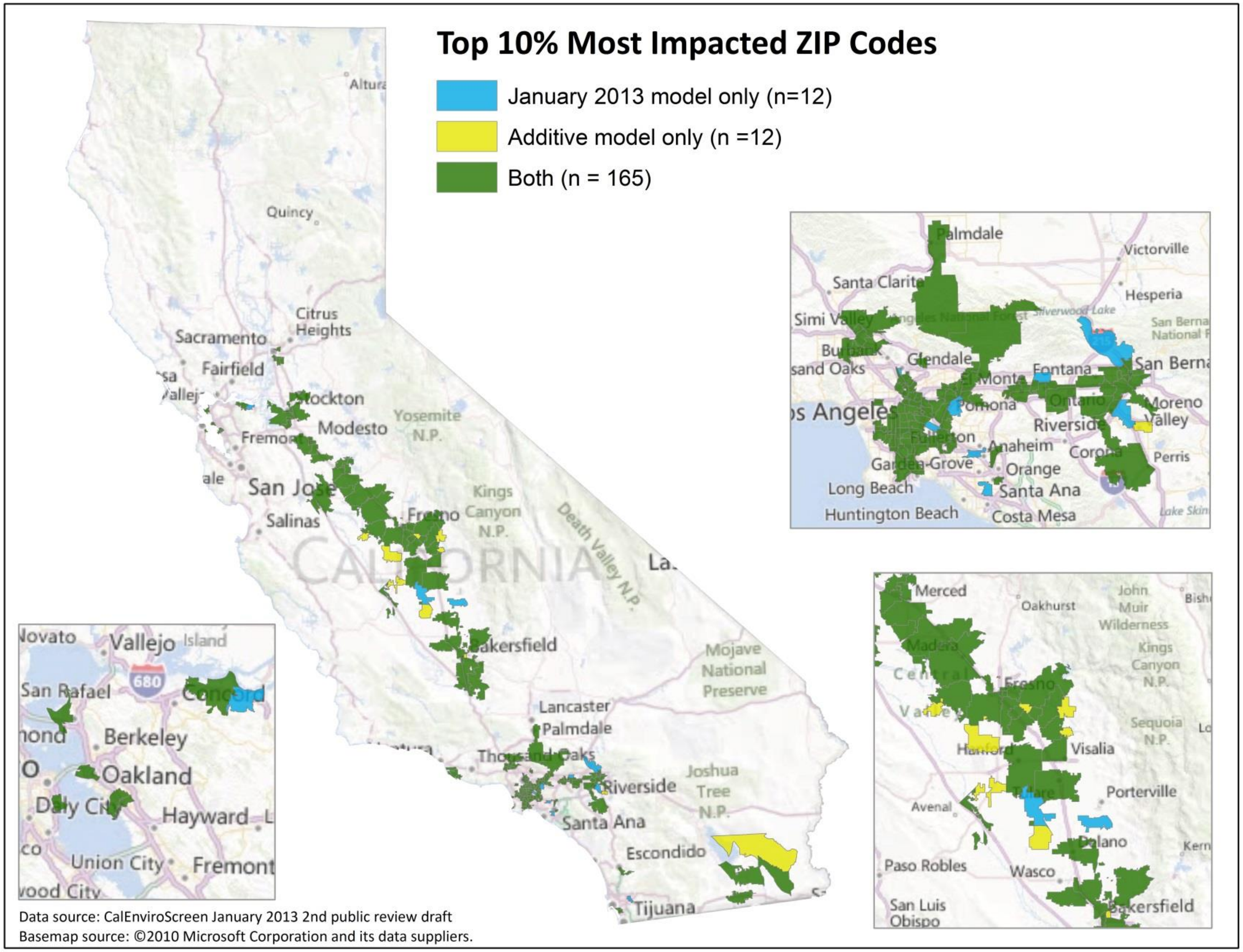
We performed the Wilcoxon Signed Rank test[5] for the decile-ranked ZIP codes for the CalEnviroScreen and additive models, and found the two models were not significantly different (p = 0.88, indicating the models are not significantly different). When the Wilcoxon Signed Rank test was performed on all ZIP codes (not grouped into deciles), the CalEnviroScreen model results were different from the additive model (p < 0.001). Figure 2 indicates that 75 percent of ZIP codes (1310 ZIP codes) remained in the same decile when the additive model was used rather than CalEnviroScreen. The vast majority of the remaining ZIP codes changed by only one decile rank. This indicates that use of an additive model does not produce significantly different results than CalEnviroScreen.

Using an additive rather than the multiplicative CalEnviroScreen model only resulted in 12 changes among the top 10 percent of ZIP codes. The values of Fagin's G (0.97) and IRM M (0.95) are both very close to 1, indicating a high degree of similarity between the ways CalEnviroScreen and the additive model ranked the top 10 percent of ZIP codes. Additionally, none of the ZIP codes within the top 10 percent in CalEnviroScreen fell below the top 15 percent in the additive model.

The state map below (Figure 3) illustrates the differences in location of the 12 changes in the top 10 percent of ZIP codes between CalEnviroScreen and the additive model. The green areas are ZIP codes that ranked in the top 10% by both models. Blue areas were ranked in the top 10 percent only by CalEnviroScreen; yellow areas were ranked high only by the additive model.

[5] The Wilcoxon Signed Rank test measures the degree of change in the rank of two sets and is appropriate for non-normally distributed data.

Figure 3. Map of California showing top 10 percent of ZIP codes generated by CalEnviroScreen and an additive model.

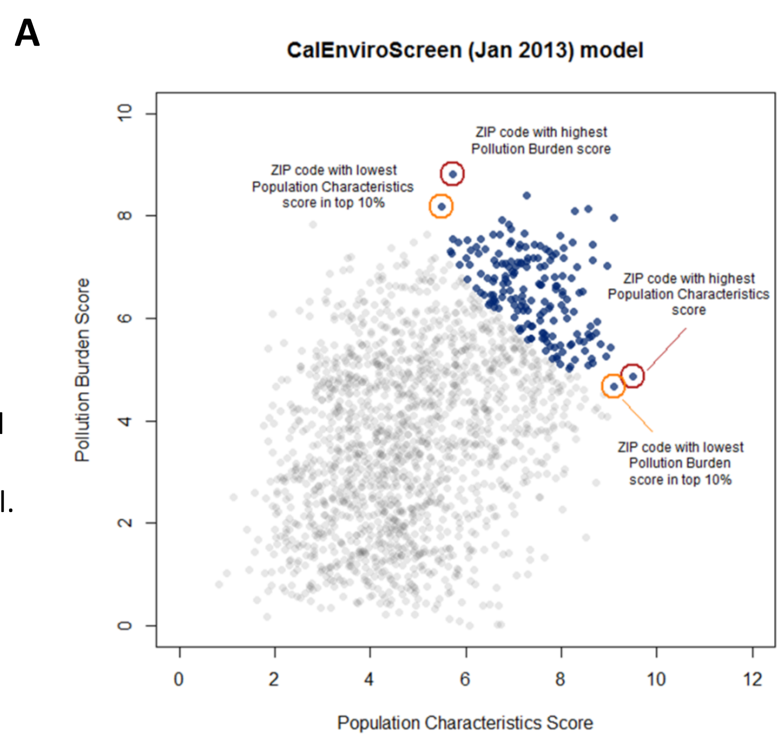


We also examined the differences between the multiplicative and additive models by using scatterplots of ZIP code scores produced by the two models. Figure 4 displays scatterplots of each ranked ZIP code on x-y coordinates of Population Characteristics Score and Pollution Burden Score. The scatterplots are shown for the CalEnviroScreen model and for an additive model (Figure 4A and Figure 4B, respectively). In each scatterplot, the top 10 percent of ZIP codes are highlighted in blue and the top scoring ZIP code for Pollution Burden and for Population Characteristics is circled in red. (These circle the same points in all scatterplots.) The two points circled in orange in each scatterplot are the two ZIP codes that have either the lowest Pollution Burden or the lowest Population Characteristics score within the top 10 percent for each model.

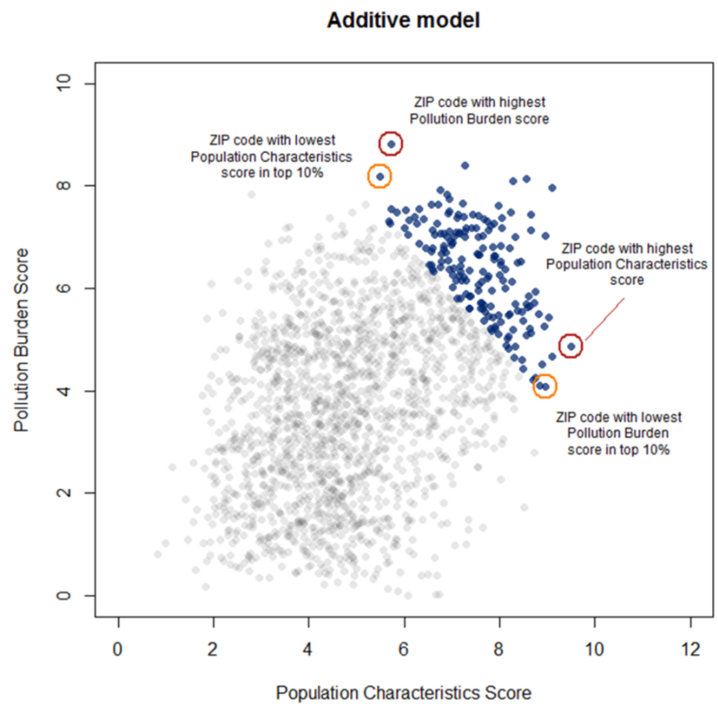
Visual inspection of a scatterplot of the Pollution Burden and Population Characteristics scores shows that both the additive model and CalEnviroScreen include the highest scoring ZIP code in each indicator group and both cover similar range of scores in the top 10 percent. This visually supports the conclusion above that both the multiplicative and the additive model appear to be approximately equally reasonable choices for identifying communities with the highest pollution burdens and vulnerabilities.

Figure 4. Scatterplots of scores produced by (A) CalEnviroScreen (January 2013) model, (B) an additive model for all California ZIP codes. Blue points represent the top 10 percent of ZIP codes identified by each model.

Top 10% scoring ZIP codes (blue) identified using scores calculated using the CalEnviroScreen model.



B



Top 10% scoring ZIP codes (blue) identified using scores calculated using an additive model (Alternative Model 4).

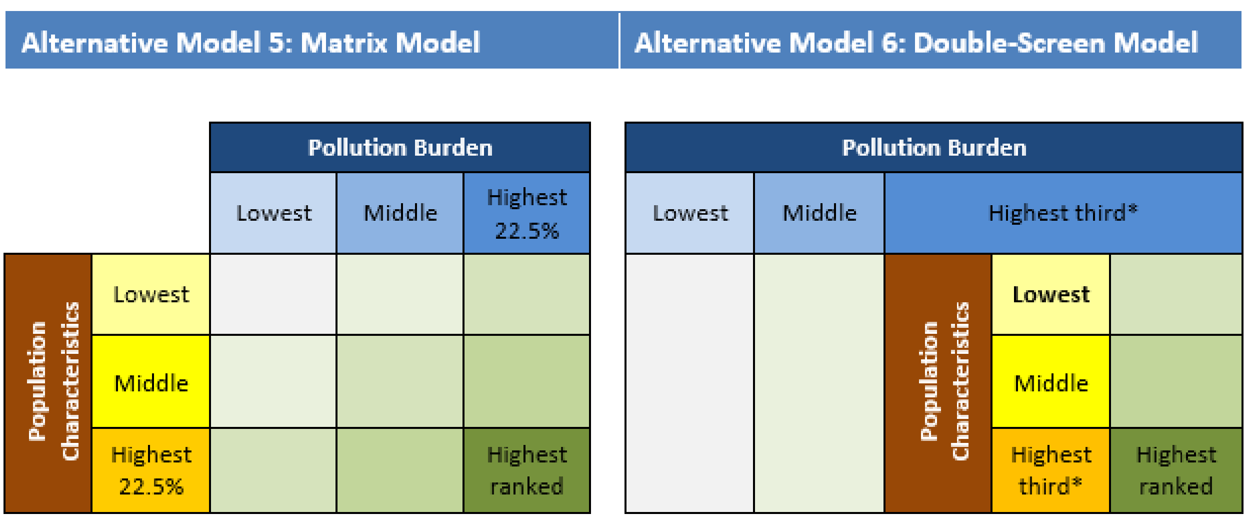
Although some of the comments we received concerning the January 2013 CalEnviroScreen model suggested that addition would be a better calculation method than multiplication, we found that there was relatively little difference between the additive model and CalEnviroScreen. When we looked at the entire state and compared the additive model to the multiplicative CalEnviroScreen model using the Wilcoxon Signed-Rank test, most of the changes were confined within decile groups, which would not appreciably alter the ZIP codes identified in the top 10 percent. And none of top 10 percent ZIP codes in the CalEnviroScreen model fell below the top 15 percent in the additive model. Research has shown that population characteristics are effect modifiers that can increase the risk of pollution exposure and increase vulnerability to pollution's adverse effects. For this reason, we felt that the use of the multiplier in the model is scientifically supported (Samet & White, 2004, Brody et al., 2012). In our comparison, the additive model brings in a number of less-polluted ZIP codes into the highest scoring places.

The academic workgroup and other commenters suggested various other possible approaches for identifying impacted communities, including a matrix method. How do these compare?

Based on public comments, we considered two further alternative models, a "matrix model" and a "double-screen model". The matrix model identifies the ZIP codes with Pollution Burden and Population Characteristics scores that fall above a common threshold. The number of ZIP codes identified with the highest pollution burdens and vulnerabilities using this model will depend on 1) the thresholds chosen and 2) the underlying distributions of the data. Here we sought a threshold that would identify roughly the same number of ZIP codes in the top 10 percent using the CalEnviroScreen model (177 ZIP codes). For this reason, in the matrix model we selected ZIP codes that met a threshold of having the top 22.5% of scores for both Pollution Burden and Population Characteristics (See Figure 5 below).

The double-screen model first identifies the top third of all ZIP codes in terms of Pollution Burden. It then identifies within this group approximately the one-third of ZIP codes with the highest Population Characteristics score. The matrix and double-screen models are illustrated below.

Figure 5. Illustration of the matrix and double-screen models for identifying the highest ranked ZIP codes based upon Pollution Burden and Population Characteristics scores.



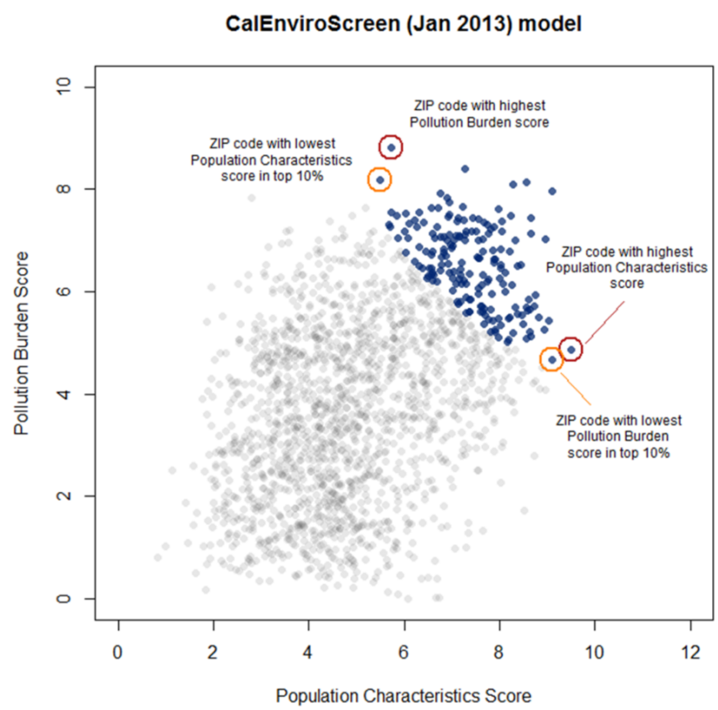
* In order to select the same number of highest ranked ZIP codes, we set the threshold at slightly less than one-third for both Pollution Burden and Population Characteristics

Figure 6 displays scatterplots of each ranked ZIP code on x-y coordinates of Population Characteristics Score and Pollution Burden Score. The scatterplots are shown for the CalEnviroScreen model and for two other models suggested by various commenters – the matrix and double-screen models. In each scatterplot, the top 10 percent of ZIP codes are highlighted in blue and the top scoring ZIP code for Pollution Burden and for Population Characteristics is circled in red. (These circle the same points in all scatterplots.) The two points circled in orange in each scatterplot are the two ZIP codes that have either the lowest Pollution Burden or the lowest Population Characteristics score within the top 10 percent for each model.

As illustrated in Figure 6, the main difference in results between CalEnviroScreen and these two models is that the January 2013 CalEnviroScreen model identifies more ZIP codes that have either a relatively high Pollution Burden or Population Characteristic score, but do not necessarily rank high in the other group. The matrix and double-screen models identify more ZIP codes that have more moderate scores in both categories.

Figure 6. Scatterplots of scores produced by (A) CalEnviroScreen (January 2013) model, (B) a matrix model, and (C) a double-screen model for all California ZIP codes. Blue points represent the top 10 percent of ZIP codes identified by each model.

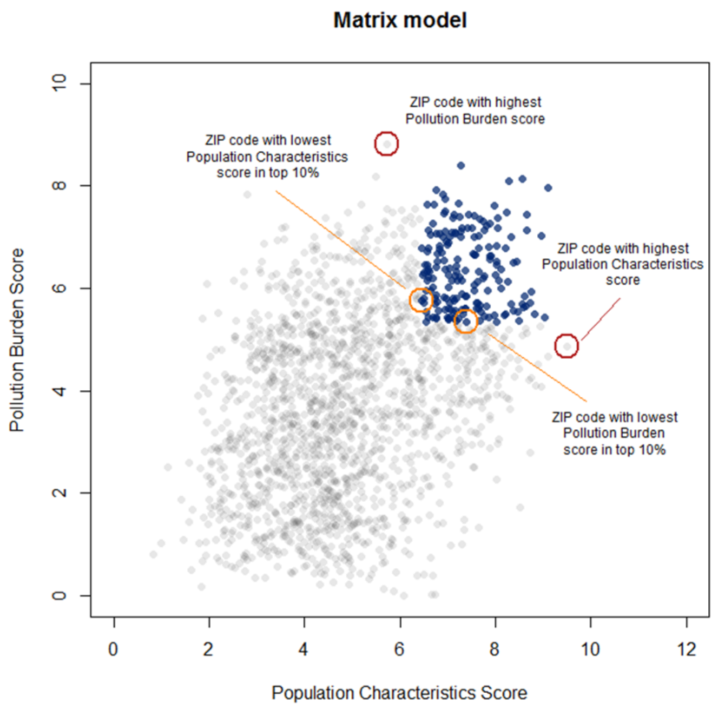
A



Top 10% scoring ZIP codes (blue) identified using scores calculated using the CalEnviroScreen model.

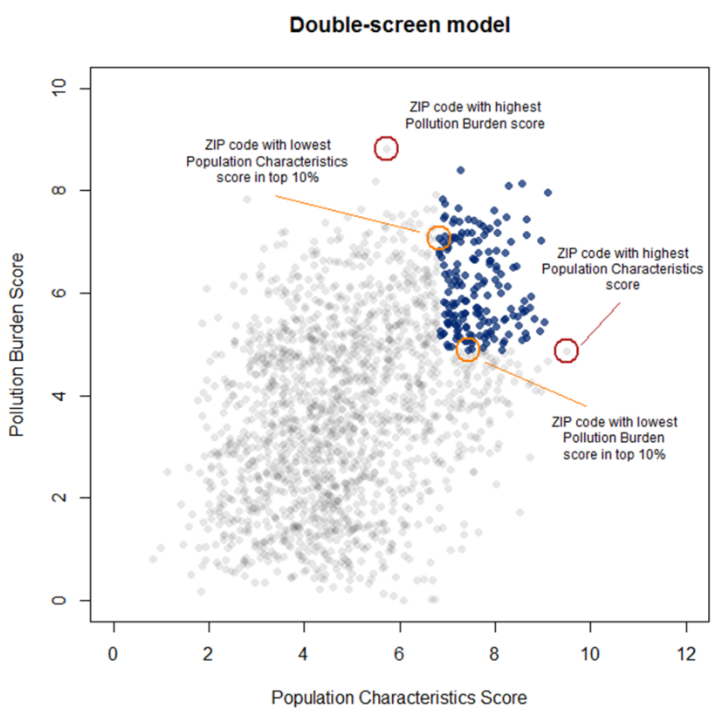
B

Top 10% (blue) identified by the matrix model described in Figure 5.



C

Top 10% (blue) identified by double-screen model described in Figure 5.



The scatterplots in Figure 6 highlight the Pollution Burden and Population Characteristics scores for ZIP codes identified as the most highly ranked by CalEnviroScreen and the matrix (Figure 6B) and double-screen (Figure 6C) models. In the CalEnviroScreen model, a ZIP code with an extremely high score in one group could have an intermediate score in the other group and still score in the top 10 percent of ZIP codes. In the matrix and double-screen models, ZIP codes would have to score fairly high in both components to be identified in the blue area.

All of the models we examined will capture the ZIP codes that rank high in both categories. However, neither the matrix nor the double-screen model selected the ZIP code with the highest Pollution Burden score or the highest Population Characteristics score (see Figure 4). In our opinion, this appears to be a limitation in the matrix and double-screen models. The double-screen model also involves a judgment as to which aspect – pollution or population vulnerability – should be considered first when characterizing overall community impact. In this example, we first screened for Pollution Burden. ZIP codes with extreme population vulnerability but only moderately high Pollution Burden would not be among the highest ranked under this scenario. The model could alternatively screen first for Population Characteristics, and thereby exclude places with extremely high Pollution Burden but only moderately high population vulnerability.

The matrix and double-screen models set thresholds below which no score will be considered. CalEnviroScreen, unlike the other two models, is supported by the academic literature that suggests that population characteristics can modify the impact of pollution. Our research suggests that race and poverty are often independently correlated with residential proximity to polluted sites (Meehan August *et al.*, 2012).

CONCLUSIONS

Sensitivity analysis here offered insights into differences in outcomes when the approach to evaluating community impacts from pollution and vulnerability is varied. Among the overarching findings is that the CalEnviroScreen model, because it is comprised of a relatively large number of indicators, is relatively robust to changes from removal of indicators and modest changes in model structure. Nonetheless, changes in results are substantial enough to suggest that each indicator makes a unique contribution to our measure of overall pollution burdens and vulnerabilities under this model. Changing the model structure more considerably by removing groups of indicators in structure does not offer preferable alternatives to the CalEnviroScreen model.

The design of CalEnviroScreen owes a great deal to the public participation process that we have followed. The comments we received helped us make improvements to the model, and led to the testing of the model with this sensitivity analysis. Nevertheless, CalEnviroScreen will continue to evolve. Indicators may be revised and new indicators will be included in the future if they add to the value and suitability of the model.

APPENDIX

METHODS

We used a range of methods to evaluate the CalEnviroScreen tool, including descriptive statistics and other statistical measures to inform the evaluation. We were primarily interested in evaluating the relationships between indicators, the effects of removing individual indicators from the model, and the CalEnviroScreen model compared to alternative models. Because the primary purpose of CalEnviroScreen is to identify communities with high pollution burdens and vulnerabilities, we paid particular attention throughout this analysis to the top 10 percent of the 1769 ZIP codes evaluated (177 ZIP codes) with the highest CalEnviroScreen scores.[6] Analyses were conducted using the statistical software, R (version 2.15.2).[7]

Method I. Spearman's Correlation Coefficients

Correlation is the degree to which two or more attributes or measurements of the same group of elements tend to vary together. Because the data we use in CalEnviroScreen are not normally distributed, we used Spearman's correlation coefficient to measure the strength of association between individual indicators. We calculated Spearman's correlation coefficients ρ (rho), which range from -1 (perfectly inversely correlated), to +1 (perfectly positively correlated), with a value of zero indicating no relationship in either direction.

Method II. Degree of change in the top 10 percent

We evaluated which of the top 10 percent of ZIP codes identified by the January 2013 CalEnviroScreen model were also identified using different models, including alternative CalEnviroScreen models suggested by stakeholders in their public comments. The degree of change within the top 10 percent of ZIP codes was calculated as followed:

$$\frac{Number\ of\ changes\ in\ ZIP\ codes_{top\ 10\%}}{Number\ of\ total\ ZIP\ codes_{top\ 10\%}} \times 100$$

Method III. Wilcoxon Signed-Rank Test

The Wilcoxon Signed-Ranked test looks at all ZIP codes' ranks in paired results from two models, for example, the CalEnviroScreen and a different model (Wilcoxon, 1945). If there is no difference in the way the two models rank all ZIP codes, then the differences between the ranks both models produce for each ZIP code is zero. The resulting test statistic W is used to calculate a p-value, which offers an indication of the statistical likelihood that the other model ranks individual ZIP codes differently.

Method IV. Fagin's G Measure and the Inverse-Rank Measure (IRM) M

We used Fagin's G Measure and the Inverse-Rank Measure (IRM) M to study changes in only the top 10 percent of ZIP codes. When comparing the rankings generated by the CalEnviroScreen model and those generated by different models, one model's subset of the top

[6] ZIP codes with no resident population as assessed by the 2010 U.S. Census were excluded from all analyses.

[7] http://www.R-project.org/

10 percent will likely not contain all the same ZIP codes as another model's subset. That is, some ZIP codes are likely to fall out of the top 10 percent and others will move into the top 10 percent. Bar-Ilan *et al.* (2006) describe the use of Fagin's G measure and the IRM M to compare rankings of lists that do not necessarily share all elements. G and M allow us to quantify the degree of similarity of the rankings between the subsets of ZIP codes identified as the highest-ranked 10 percent by the CalEnviroScreen model and an alternate model. They consider both the ZIP codes that comprise the subset and how they are ordered.

M assigns weights to identical or similar rankings near the top of the subset (i.e., the top of the top 10 percent); G does not assign weights. For the M measure, this means that if a ZIP Code is ranked 3rd by model A and 6th by model B, that comparison would be given more weight than a ZIP Code that is ranked 173rd by model A and 176th by model B. On the other hand, the G measure considers changes in lower-ranked ZIP codes equivalent to changes in the highest ranked communities in the subset. Values of G and M can range from zero (no ZIP codes identified in the CalEnviroScreen model were identified by the alternate model) to 1 (the two models identified exactly the same ZIP codes and ranked them in the same order). Values that are close to 1 indicate that the two models generated very similar rankings among the highest-ranked ZIP codes.

Method V. Scatterplots

Finally, scatterplots displaying all ranked ZIP codes on x-y coordinates of Population Characteristics Score and Pollution Burden Score were used to visually compare the highest ranked ZIP codes using the CalEnviroScreen model to alternative models.

REFERENCES

Bar-Ilan J, Mat-Hassan M, Levene M (2006). Methods for comparing rankings of search engine results. *Computer Networks* **50**(10):1448-63.

Brody TM, Di Bianca P, Krysa J (2012). Analysis of inland crude oil spill threats, vulnerabilities, and emergency response in the midwest United States. *Risk Analysis* **32**(10):1741-9.

Burd-Sharps S, Lewis K (2011). *A Portrait of California: California human development report*: Social Science Research Council. http://www.measureofamerica.org/california. Accessed March 2013.

Fujita EM, Campbell DE, Arnott WP, Chow JC, Zielinska B (2007). Evaluations of the chemical mass balance method for determining contributions of gasoline and diesel exhaust to ambient carbonaceous aerosols. *Journal of the Air & Waste Management Association* **57**(6):721-40.

Fujita EM, Zielinska B, Campbell DE, Arnott WP, Sagebiel JC, Mazzoleni L*, et al.* (2007). Variations in speciated emissions from spark-ignition and compression-ignition motor vehicles in California's South Coast Air Basin. *Journal of the Air & Waste Management Association* **57**(6):705-20.

Meehan August L, Faust JB, Cushing L, Zeise L, Alexeeff GV (2012). Methodological Considerations in Screening for Cumulative Environmental Health Impacts: Lessons Learned from a Pilot Study in California. *International Journal of Environmental Research and Public Health* **9**(9):3069-84.

Pastor M, Morello-Frosch R, Sadd J (2010). *Air pollution and environmental justice: integrating indicators of cumulative impact and socio-economic vulnerability into regulatory decision-making;* ed.: California Environmental Protection Agency, Air Resources Board, Research Division. Final Report. Contract No. 04-308 with the Air Resources Board.

Rogge WF, Hildemann LM, Mazurek MA, Cass GR, Simoneit BR (1993). Sources of fine organic aerosol. 2. Noncatalyst and catalyst-equipped automobiles and heavy-duty diesel trucks. *Environmental Science & Technology* **27**(4):636-51.

Samet JM, White RH (2004). Urban air pollution, health, and equity. *J Epidemiol Community Health* **58**(1):3-5.

Taylor M (2009). What is sensitivity analysis?, *What is … ? series* (pp. 8). London: Hayward Medical Communications. http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_sens_analy.pdf. Accessed March 2013.

Wilcoxon F (1945). Individual comparisons by ranking methods. *Biometrics bulletin* **1**(6):80-3.