Atmospheric Environment 247 (2021) 117665

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: http://www.elsevier.com/locate/atmosenv

Improving spatial surrogates for area source emissions inventories in California

Yiting Li [a], Caroline Rodier [b], Jeremy D. Lea [c], John Harvey [c], Michael J. Kleeman [c,*]

[a] *Department of Land, Air, and Water Resources, University of California, Davis, United States*
[b] *Institute of Transportation Studies, University of California, Davis, United States*
[c] *Department of Civil and Environmental Engineering, University of California, Davis, United States*

HIGHLIGHTS

- We updated spatial surrogates that describe location of area-source emissions across California.
- New spatial surrogates use public datasets and can be projected to future years.
- Changes to construction surrogates had the largest effect on emissions, followed by employment surrogates.
- New surrogates improved predicted PM concentrations by 1.5–10% in Sacramento, San Francisco, and Los Angeles in the year 2016.

ARTICLE INFO

Keywords:
Spatial surrogate
Area-source
Non-point source
Construction
Landuse

ABSTRACT

Ten spatial surrogates describing the detailed locations of air pollution emissions in regional air quality assessments for California were updated/created for the base year 2010 and future years from 2015 to 2040: (i) total population, (ii) total housing, (iii) single-family housing, (iv) total employment, (v) service & commercial employment, (vi) industrial employment, (vii) agricultural employment, (viii) industrial-related surrogate, (ix) off-road construction, and (x) on-road construction surrogates. The first seven surrogates were updated using the latest version of census-based datasets at finer resolution. New industrial-related, off/on-road construction surrogates were developed using realistic datasets to more accurately describe the location of construction projects and industrial facilities. Adoption of the new spatial surrogates caused changes to the spatial distribution of air pollution emissions in air quality calculations. The changes to the off-road construction surrogate resulted in the largest shift in PM emissions distribution for year 2015, followed by changes to the on-road construction surrogate. Industrial-related, service & commercial employment, and off-road construction surrogates all contributed to changes in NOx emissions. The changes to SED-derived surrogates were subtle and did not significantly influence emissions. Air quality simulations were carried out over the entire year 2016 to examine the impact of the new surrogate methodologies on simulated concentration fields. Changes to predicted pollutant concentrations followed the same pattern as changes in emissions, which indicates that proximity to sources is a dominant factor to determine the impact of spatial surrogates on model performance. The updated spatial surrogates generally improved predicted PM mass and EC concentrations in the Sacramento area (~10% for PM, ~3% for EC), the Bay Area (~3% for PM, ~1.5% for EC), and the region surrounding Los Angeles (~5% for PM, ~4% for EC). The updated spatial surrogates also improved predicted NOx concentrations in the core region of Los Angeles (~6%). These improvements demonstrate that development and adoption of new methodologies for emissions spatial surrogates can improve the accuracy of regional chemical transport models for criteria air pollutants.

1. Introduction

Chemical Transport Models (CTMs) are used to predict air pollutant concentrations over scales ranging from 10's of meters to 100's of kilometers (Eastham et al., 2018; González et al., 2018; Hu et al., 2017; Joe et al., 2014; Kuik et al., 2016; Li et al., 2016; Schaap et al., 2015;

* Corresponding author.
E-mail address: mjkleeman@ucdavis.edu (M.J. Kleeman).

https://doi.org/10.1016/j.atmosenv.2020.117665
Received 10 February 2020; Received in revised form 6 May 2020; Accepted 31 May 2020
Available online 4 June 2020

Table 1
Fraction of total area source emissions allocated by each surrogate.

Original surrogate		Updated surrogate		Fraction of area source criteria emissions					
				PM	TOG	NOx	CO	SOx	NH_3
585	Construction equipment	586	Construction	13.1%	0.77%	8.46%	1.50%	0.15%	
		587	Off-road construction						
		588	On-road construction						
730	Industrial-related			0.26%	6.85%	13.9%	4.36%	12.4%	3.01%
300	Industrial employment								
250	Total housing				0.41%				
440	Total population			0.10%	5.69%	0.26%	0.33%	0.15%	17.8%
620	Service & commercial employment				0.21%	2.8%	0.25%	2.35%	
650	Single-family housing				1.20%	0.23%	2.56%		

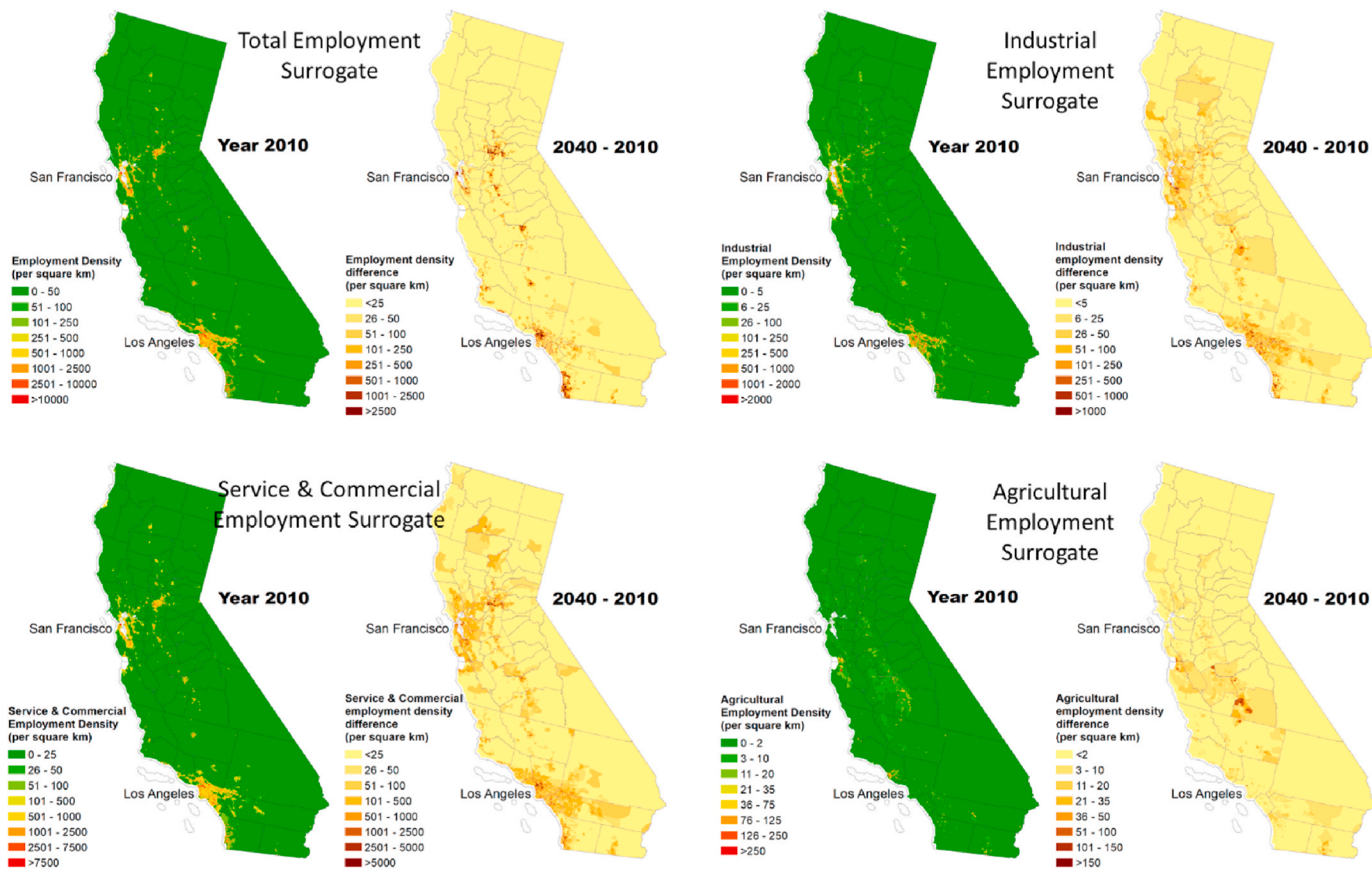

Fig. 1. Total employment, industrial employment, service & commercial employment, and agricultural employment in the years 2010 and spatial surrogate difference between 2040 and 2010.

Woody et al., 2016). One of the most important uses of CTMs is to assist in the design of emissions control programs that will achieve compliance with the National Ambient Air Quality Standards (NAAQS) (Herrera et al., 2010; Hogrefe and Rao, 2001; Kelly et al., 2019; Macpherson et al., 2017; Saylor et al., 1999; Zhang et al., 2011). Another important use of CTMs is to estimate population exposure to various air pollutants (Chen et al., 2014; Huang et al., 2018; Laurent et al., 2014, 2013; Ostro et al., 2015; Stieb et al., 2016; Van Donkelaar et al., 2015; Wang et al., 2016). Multiple studies have concluded that the errors introduced into CTMs by coarse spatial resolution could affect human health impact assessments (Fenech et al., 2018; Thompson and Selin, 2012) and so it is desirable to apply CTMs at the finest possible spatial resolution. Accurately describing the location of emission sources is often a critical factor that determines the fidelity of this overall process to protect public health across the United States (Cohan et al., 2006; Pan et al., 2017; Tan et al., 2015; Valari and Menut, 2008; Zheng et al., 2017).

A top-down approach is widely used to create spatial gridded emissions, and spatial surrogates play an important role in accurately mapping aggregated emissions to model grid cells (Bieser et al., 2011; Bun et al., 2010; Kuenen et al., 2014; US EPA, 2017). The effort needed to prepare spatially accurate emission inventories varies by source category that can be broadly summarized as (i) point sources, (ii) mobile sources, or (iii) area-sources. Major air pollution point sources in the United States have exact latitude and longitude recorded with their emissions permits making it easy to specify their exact location in emissions inventories. Likewise, mobile sources emit pollutants along well-defined roadways that often have monitors to measure traffic volume (Fameli and Assimakopoulos, 2015; Fu et al., 2017; McDonald et al., 2014). In contrast, the location of area-sources (or non-point sources) are difficult to describe accurately in emissions inventories (Dai and Rocke, 2000; Gkatzoflias et al., 2013; Trombetti et al., 2018). Hundreds or thousands of different types of area sources exist in a

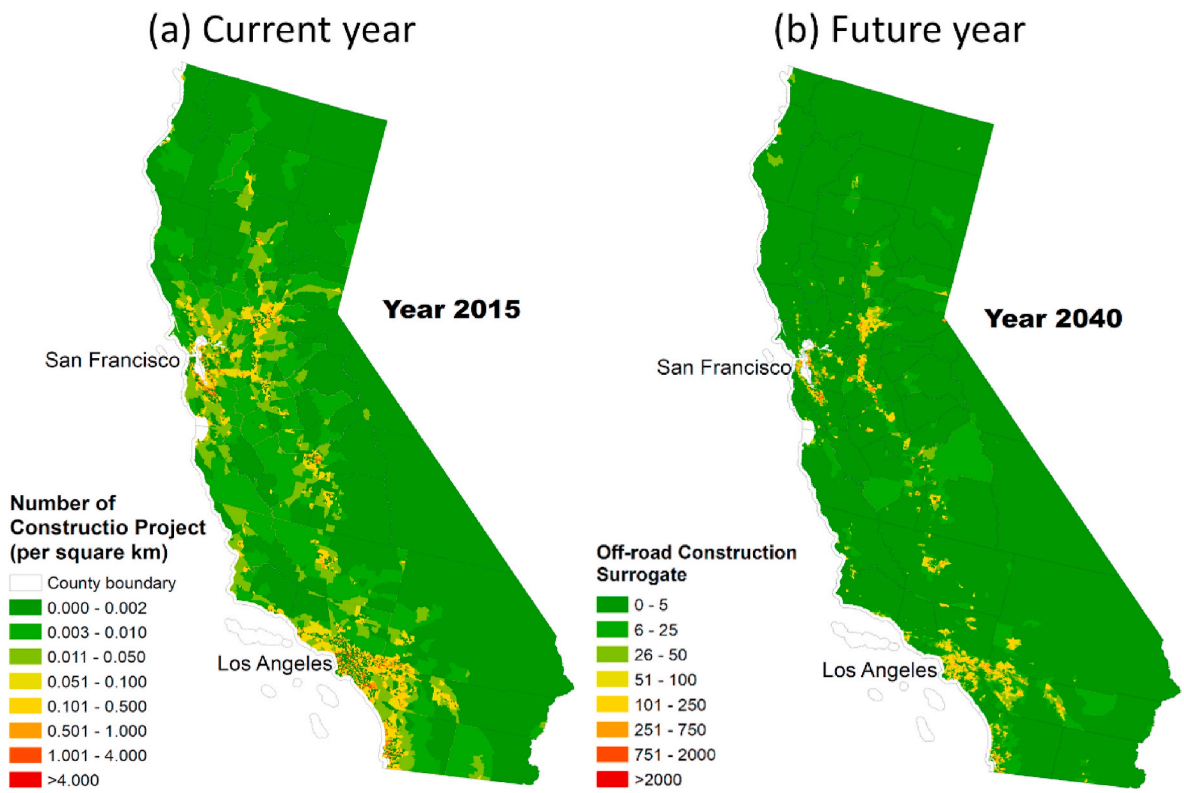

Fig. 2. Current and future off-road construction surrogate. Figure (a) shows year 2015 off-road construction surrogate created from a dataset of construction permits. Figure (b) shows an example of future year off-road construction surrogate, which is the population growth in the preceding 5 year period.

Table 2
NAICS codes and LEHD fields name and definitions.

Surrogate	NAICS code	LEHD field	Definitions
Agricultural Employment	11	cns01	Agricultural, forestry, fishing and hunting
Industrial Employment	21	cns02	Mining
	22	cns03	Utilities
	23	cns04	Construction
	31–33	cns05	Manufacturing
	42	cns06	Wholesale trade
	48–49	cns07	Transportation and warehousing
Service & Commercial Employment	44–45	cns08	Retail trade
	51	cns09	Information
	52	cns10	Finance and insurance
	53	cns11	Real estate rental and leasing
	54	cns12	Professional, scientific, and technical services
	55	cns13	Management of companies and enterprise
	56	cns14	Administrative and support and waste management and remediation services
	61	cns15	Educational service
	62	cns16	Health care and social assistance
	71	cns17	Arts, entertainment, and recreation
	72	cns18	Accommodation and food services
	81	cns19	Other services (except public administration)
	92	cns20	Public administration

typical urban region. As emissions from major point sources and mobile sources are reduced, these numerous area sources have emerged as a very important category for continued progress towards improved air quality (McDonald et al., 2018). Accurately describing the location of area source emissions is of paramount importance for the next generation of regulatory programs to address current and future conditions in the United States.

Area-source emission rates in the United States are often estimated using a formula such as

$$Total\ Emissions = Activity \times (Emissions\ /\ Activity)$$

where *"activity"* is a measure such as fuel consumed or production rate and *"(emissions/activity)"* describes the release of air pollutants for each unit of fuel consumed or product created. Emissions totals are calculated for broad spatial zones such as counties or Geographic Area Index (GAI) regions. For example, the California Air Resources Board (CARB) divides the state into 69 GAI regions based on the intersection of county, air basin, and air district political boundaries. The detailed location of area-source emissions within each GAI is then described using spatial surrogates that are assumed to be proportional to the target emission rates. Statistical activity data such as census data, industry registration, traffic information and fuel consumption are commonly used to disaggregate various county-level emissions, but the availability of these datasets varies by location. Population is the most common dataset used to create spatial surrogate (Bieser et al., 2011; Kuenen et al., 2014; Zasina and Zawadzki, 2017; Zhao et al., 2012; Zhou et al., 2017) but CARB currently uses over 100 additional spatial surrogates depending on the exact source-type as summarized in Table S1.

The purpose of this paper is to update ten important spatial surrogates used to specify the location of area-source emissions in California: (i) total population, (ii) total housing, (iii) single-family housing, (iv) total employment, (v) service & commercial employment, (vi) industrial employment, (vii) agricultural employment, (viii) industrial-related surrogate, (ix) off-road construction, and (x) on-road construction surrogates. These ten spatial surrogates allocate emissions locations for sources ranging from lawn and garden equipment to construction and mining to residential/commercial/industrial natural gas combustion. The updated surrogates take advantage of publicly available datasets

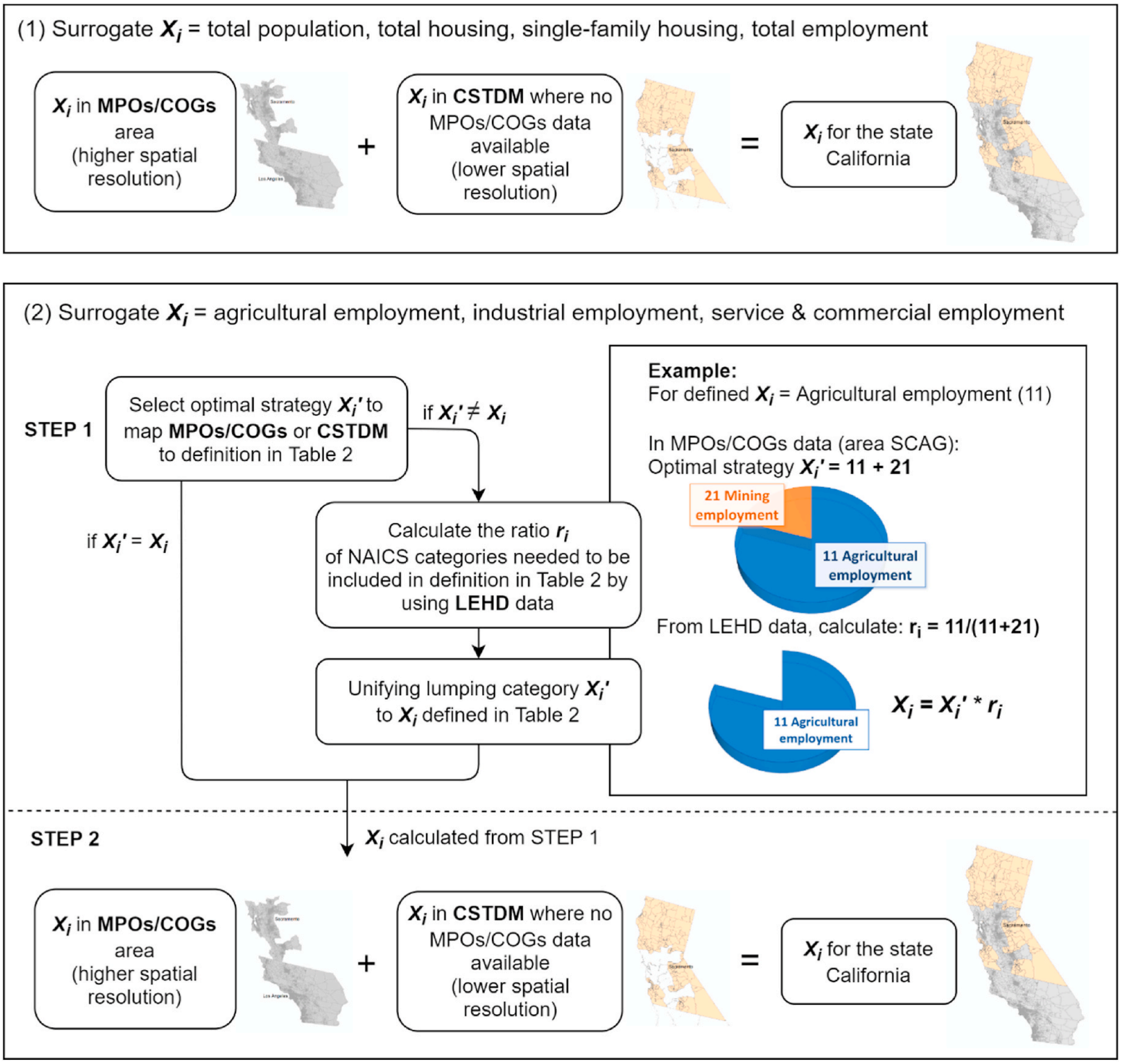

Fig. 3. Flow chart of methodology for SED spatial surrogates, including total population, total housing, single-family housing, total employment, agricultural employment, industrial employment and service & commercial employment. Subscript i is the geographic unit of SED dataset.

that can be extrapolated to future years. In the current study, new spatial surrogates were created for the years 2010, 2015 and extrapolated to the years 2020 through 2040. Updated emission spatial patterns were compared to previous emission spatial patterns and a regional air quality model was used to predict differences in ground-level pollutant concentrations resulting from adoption of the updated emissions. The findings from this study help to improve the spatial accuracy of emissions inventories in California, which can be used as a model for other locations in the United States.

2. Methods

The ten spatial surrogates updated in the current paper can be divided into three major categories based on their data source and emission source related to: socio-economic data (SED), industrial-related surrogates, and construction equipment surrogates. The construction equipment surrogate allocates 13.1% of particle matter and 8.46% of NOx within total area source emissions (Table 1). The industrial employment and industrial-related surrogates are used primarily to allocate gaseous emissions, including 13.9% of NOx, 12.4% of SOx, and 6.85% of TOG (Total Organic Gas). The industrial employment surrogate usually serves as a secondary surrogate for industrial emission that will be used when the primary surrogate is not available within the target geographical region. The SED surrogates accounts for a smaller fraction of the emissions across California compared to the construction equipment and industrial surrogates, but SED surrogates can still have non-negligible impacts on populated areas. Updates to each of the eight spatial surrogates are described in the sections below. Two new SED surrogates – total employment and agricultural employment were created in Section 2.3.

2.1. Industrial-related surrogate

The industrial-related surrogate (730) is used to describe the location of manufacturing processes and industrial fuel combustion (including natural gas not associated with major point sources). The original industrial-source spatial surrogate was created from the 2016 Dun and Bradstreet Financial Database (DUNS database, http://www.dnb.com/). The DUNS database often lists the address for company headquarters rather than actual industrial facility locations where emissions are released. Moreover, DUNS employment types are classified by the Standard Industrial Classification (SIC) system, which groups industries based on demand or production of goods. As a result, DUNS employment totals include industrial occupations combined with office/managerial

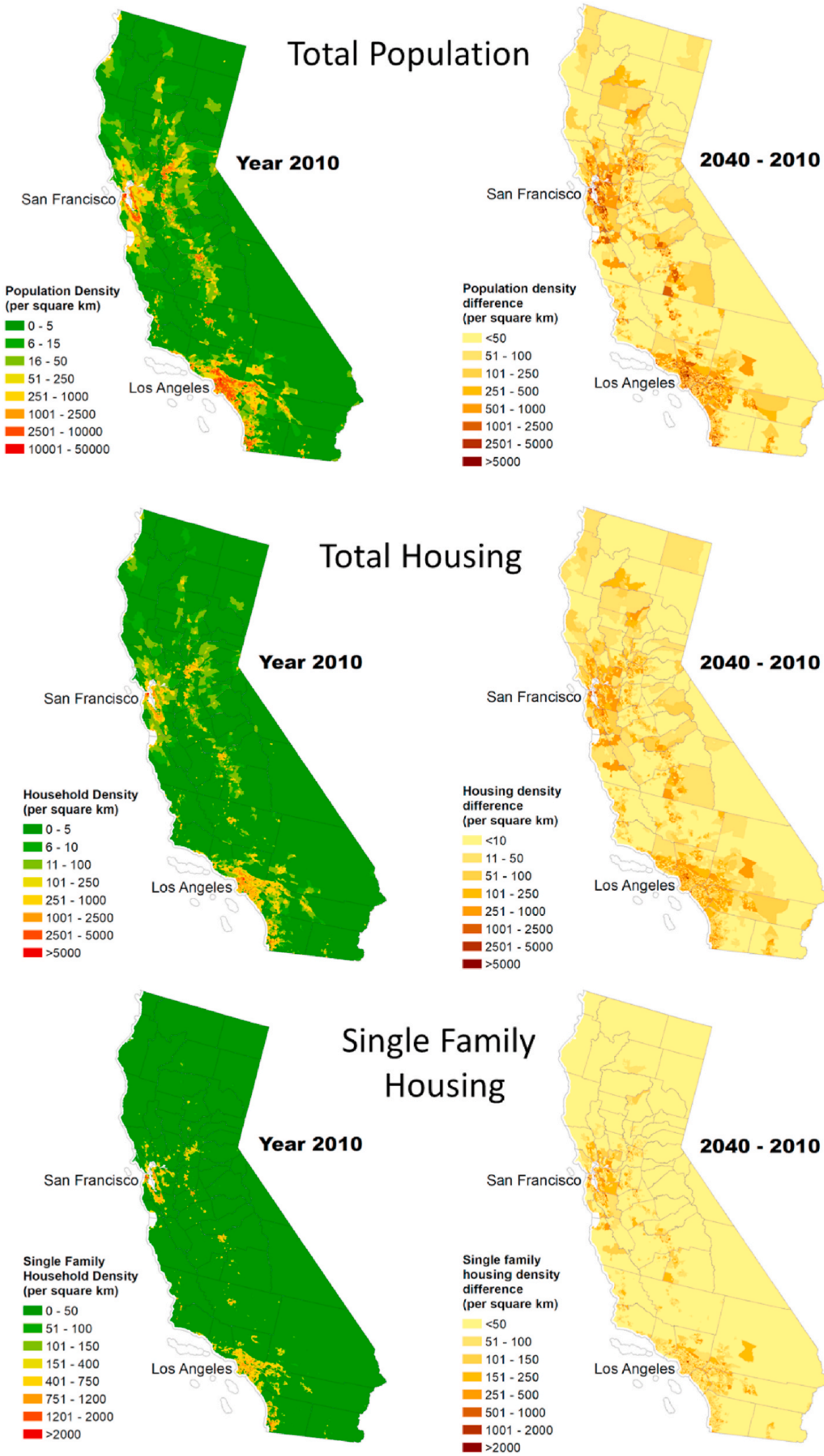

Fig. 4. Total population, total housing, and single-family housing in the years 2010 and spatial surrogate difference between 2040 and 2010.

occupations that may not be correlated with emissions. Approximately 11% of organizations in the SIC-based DUNS database in California are actually not related to industrial process (Fig. S1). In contrast, North American Industry Classification System (NAICS) groups are organized based on the likeness of the process used to generate goods or services. NAICS codes for industrial occupations used in the Longitudinal Employer-Household Dynamics (LEHD) "OnTheMap" dataset (United State Census Bureau, 2019) are distinct from NAICS codes for

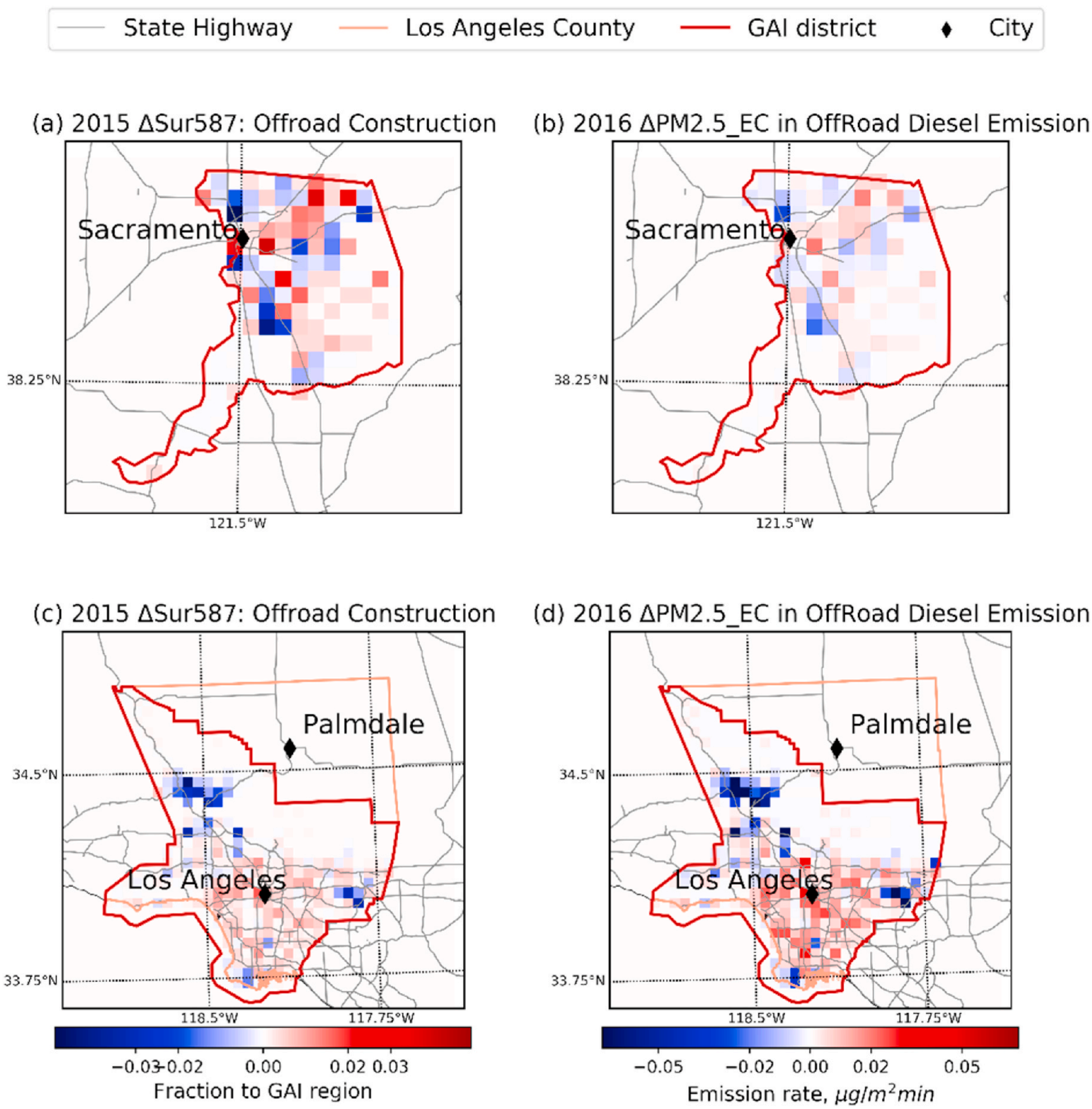

Fig. 5. Relationship between off-road construction surrogate and PM2.5 EC in off-road diesel emission. Figure (a) and (c) are surrogate difference between original 585 and updated 587 at Sacramento county and Los Angeles GAI 6059. Figure (b) (d) are PM2.5 EC difference in off-road diesel emission at the same corresponding area. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

office/managerial occupations yielding more accurate industrial employment totals. Furthermore, LEHD employment locations are specified at the actual industrial facility, which more accurately represents industrial activities than the headquarters location. Adoption of the LEHD manufacturing employment to replace the SIC manufacturing employment yields a spatial distribution with greater variability in Southern California (see Fig. S2). LEHD manufacturing employment data is also consistent with major industrial activities permitted under the Stormwater Multiple Application and Report Tracking system (SMART) database maintained by the California Water Resources Board (see Fig. S3). The agreement between these independent indicators of industrial activity builds confidence in the accuracy of the LEHD manufacturing distribution.

Future-year industrial-related surrogates were adopted from the future-year SED data related to industrial employment generated in Section 2.3. Fig. S4 plots the industrial employment distribution alongside the location of industrial activities locations from the SMART database. The results indicate that the future-year industrial employment distribution also captures the spatial pattern of real industrial activities, building confidence in the approach to use future industrial employment as a spatial surrogate for future industrial emissions.

Fig. 1 shows industrial employment surrogates for the years 2010 and spatial surrogate difference between 2040 and 2010. Total employment, industrial employment, and service & commercial employment are clustered in urban areas in both current and future years. Agricultural employment is significantly lower with most activity focused on the San Joaquin Valley in central California.

2.2. Construction equipment surrogate

The construction equipment surrogate (585) is used to describe the spatial location of equipment burning gasoline and diesel fuel for the purpose of creating buildings and roads, and dust from construction activities. The current CARB construction equipment spatial surrogate blends information from two sources: (i) the change in “impervious

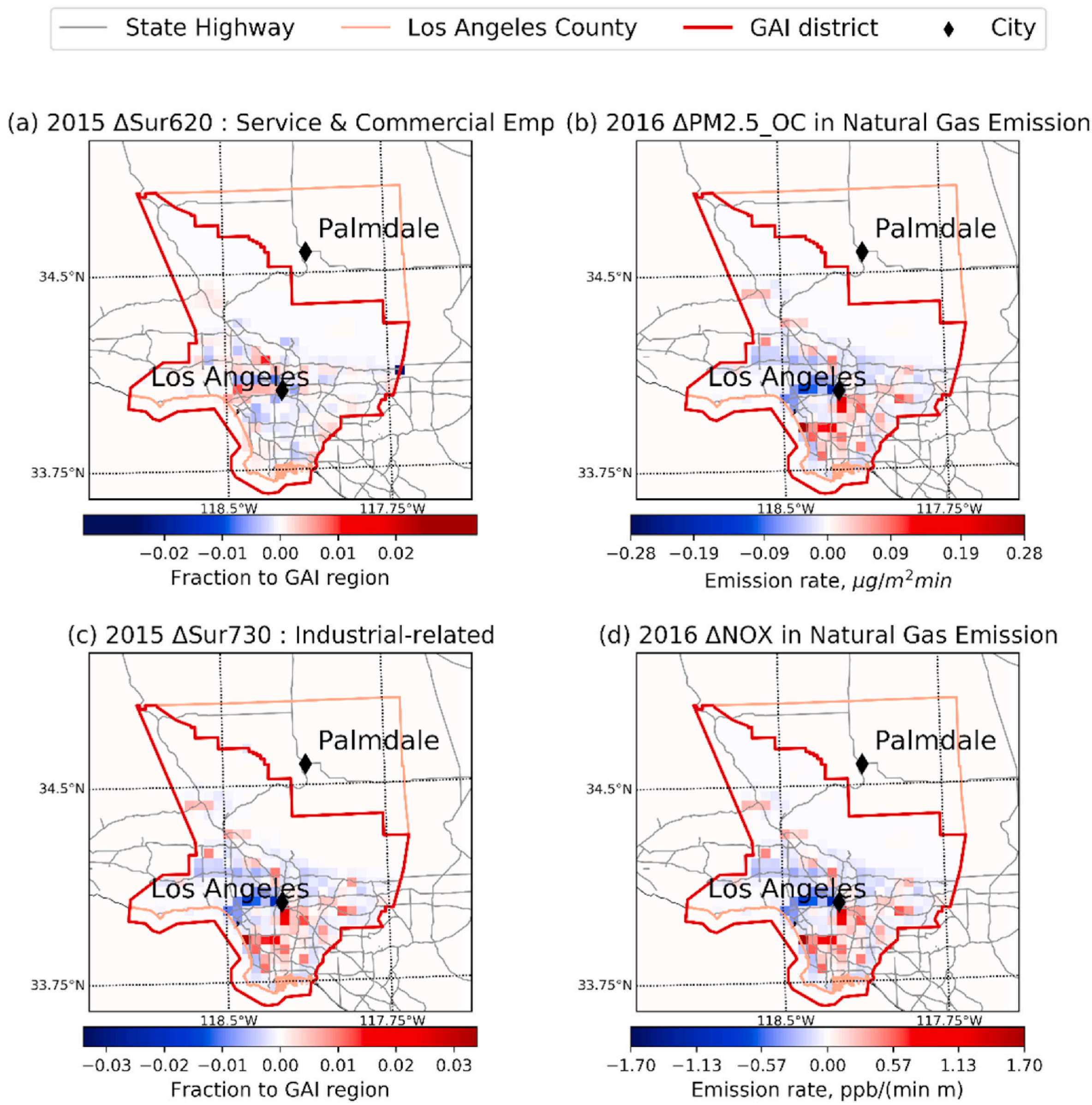

Fig. 6. Relationship between service & commercial employment, industrial-related surrogates and $PM_{2.5}$ OC, NOx in natural gas emission. Figure (a) and (c) are surrogate difference between original and updated service & commercial/industrial-related surrogates at Los Angeles GAI 6059. Figure (b) (d) are $PM_{2.5}$ OC and NOx difference in natural gas emission at the same corresponding area. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

surfaces (imperviousness)" between 2006 and 2011 from the National Land Cover Database (NLCD) and (ii) the California Department of Transportation (Caltrans) on-road Truck Network (482) surrogate. The combined construction surrogate weights the input impervious surface factor by 90% and the Truck Network factor by 10%. Impervious surfaces are mainly artificial structures or impenetrable materials such as pavements. Changes in imperviousness can be used to identify the pattern, nature and magnitude of the change in urban land cover. From an emissions perspective, an increase in impervious surfaces is almost always associated with construction equipment. However, some construction activities do not change impervious landscape, such as demolition and reconstruction activities for existing buildings. The NLCD impervious surface data is only available in 5-yr increments which does not identify the exact date of the construction activity. The Truck Network represents the location of commercial truck routes including terminal access to ports and national network routes. The construction surrogate assumes that all roads are subject to repair over their lifecycle, and so construction emissions are distributed uniformly on this network. This approach is reasonable over an averaging time of ~15 years but not realistic within in any given year. The surrogate created from the combination of NLCD imperviousness and truck network data only approximately represents construction activities and it is difficult to apply over all potential years of interest.

The construction equipment surrogate created in the current project was separated into three individual surrogates to better represent the different types of construction activity: (i) off-road construction surrogate (587) represents off-road construction recorded in the SMART database; (ii) on-road construction surrogate (588) represents projects from Caltrans highway records; and (iii) construction surrogate (586) is a combination of 50% surrogate 587 and 50% surrogate 588 as recommended by staff at the California Air Resources Board based on their testing of a range of on-road and off-road weighting factors. Surrogate 586 is hereafter reserved as a backup or secondary surrogate if the primary surrogate is not available in certain areas.

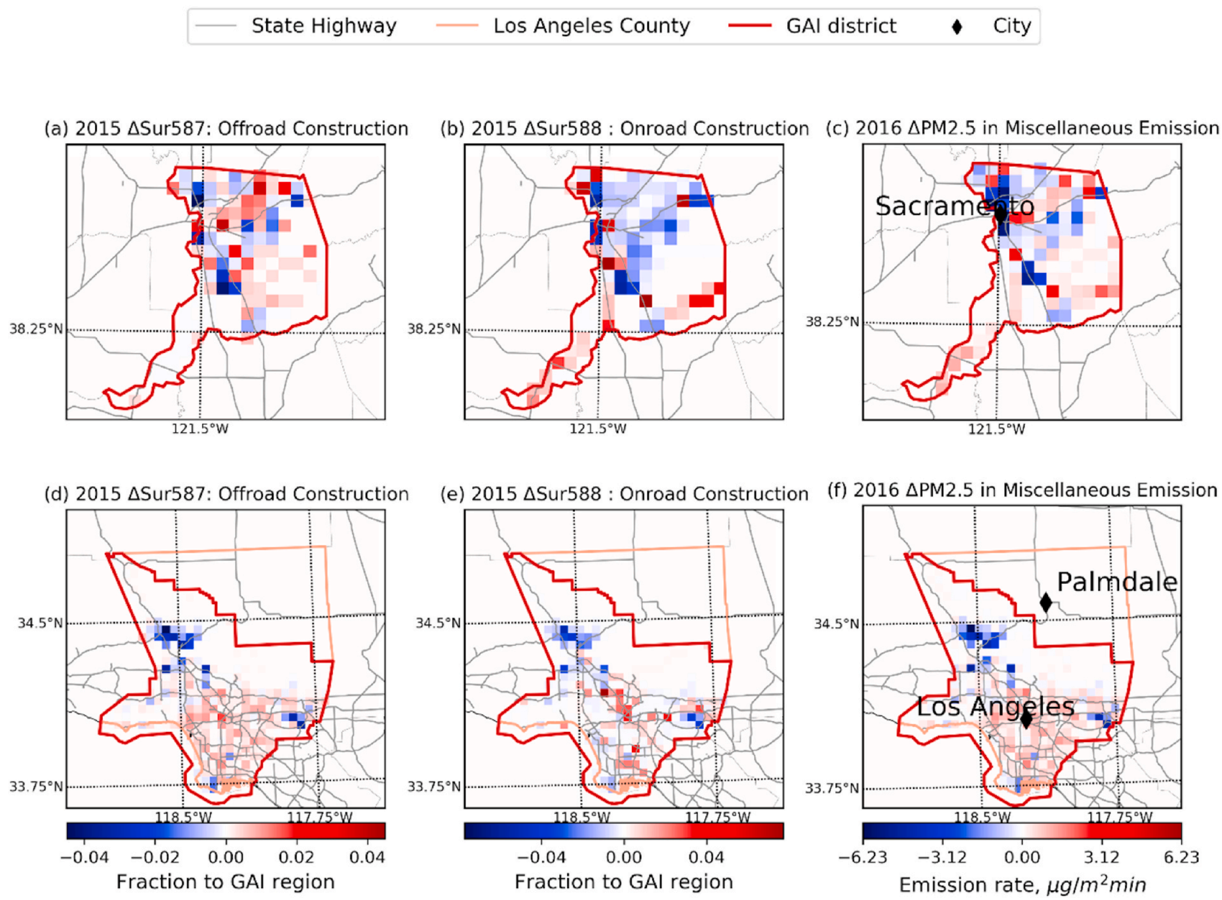

Fig. 7. Relationship between off-road/on-road construction surrogates and PM2.5 total mass emissions from miscellaneous sources at Sacramento county and Los Angeles GAI 6059. Figure (a) (b) (d) and (e) are surrogate differences between original 585 and updated 587/588. Figure (c) (f) are differences in PM2.5 total mass emissions from miscellaneous sources as defined in Section 2.4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Diesel engine exhaust is one of the major sources of NOx and PM emissions from construction activities (Millstein and Harley, 2009). Heavy diesel equipment including tractors, loaders, backhoes, and skid steer loaders account for 80% of all construction equipment and 78% of annual operating hours in California (California Air Resource Board, 2020, 2010). Tractors, loaders, backhoes, and skid steer loaders are estimated to contribute 67% of total NOx emissions and 70% of total PM emissions from statewide construction and mining activities (California Air Resource Board, 2020, 2010). Heavy diesel-powered equipment is used primarily in large-scale new construction projects or large-scale reconstruction/renovation projects as opposed to small scale construction projects. It is desirable for the off-road construction surrogate to represent these major construction/renovation projects as accurately as possible. Construction surrogates are also used to spatially allocate PM dust emissions during building/road construction. Construction dust comes from material, equipment and transportation (Sandanayake et al., 2016) which are also expected to be highest around large scale construction sites.

Two updates were made to increase the accuracy of the current-year construction equipment spatial surrogate. First, the NLCD impervious surface surrogate was replaced with off-road construction project permits from the SMART database that describe the project location (lat/lon), construction type, imperviousness change, and distributed activity area. It should be noted that permits in SMART are only required for projects larger than 1 acre, meaning SMART data captures large-scale construction projects that account for the majority of the heavy-duty diesel equipment and dust construction emissions. As a second update, the generic Truck Network surrogate was replaced with the actual location of highway construction projects described in records publicly filed with Caltrans including highway number, the start and stop mile along that highway, and the number of active working days in the project. These actual off-road and on-road construction locations were converted to the standard map projections used for CARB spatial surrogates, leading to increased accuracy in the location of construction emissions.

The off-road construction surrogate is mainly used to distribute building construction emissions. Most new buildings are associated with an increase in population (either residences or commercial services). A statistical analysis of the correlation between the county-level population growth between 2010 and 2015 and the number of new construction projects yields a high correlation ($R^2 = 0.89$), which builds confidence in the strength of the association between changes in population and changes in construction activity (see Fig. S5). Therefore, future-year off-road construction surrogates are based on population increase calculated from the changes in population surrogates generated in Section 2.3. Fig. 2 shows a current year and a future year off-road construction surrogate. Most of the current-year building construction activity occurs in urban areas (Fig. 2(a)), which is consistent with the spatial pattern of population difference between years (Fig. 2(b)).

The 10-year future on-road construction surrogate was created from the publicly available State Highway Operation and Protection Program (SHOPP) 10-year plan that lists all possible construction projects in the coming decade with accurate location and road treatment type. SHOPP is a product of information from the Caltrans pavement network analysis tool – PaveM system (Caltrans, 2015) and local Caltrans decisions made using that information. PaveM uses databases describing pavement type,

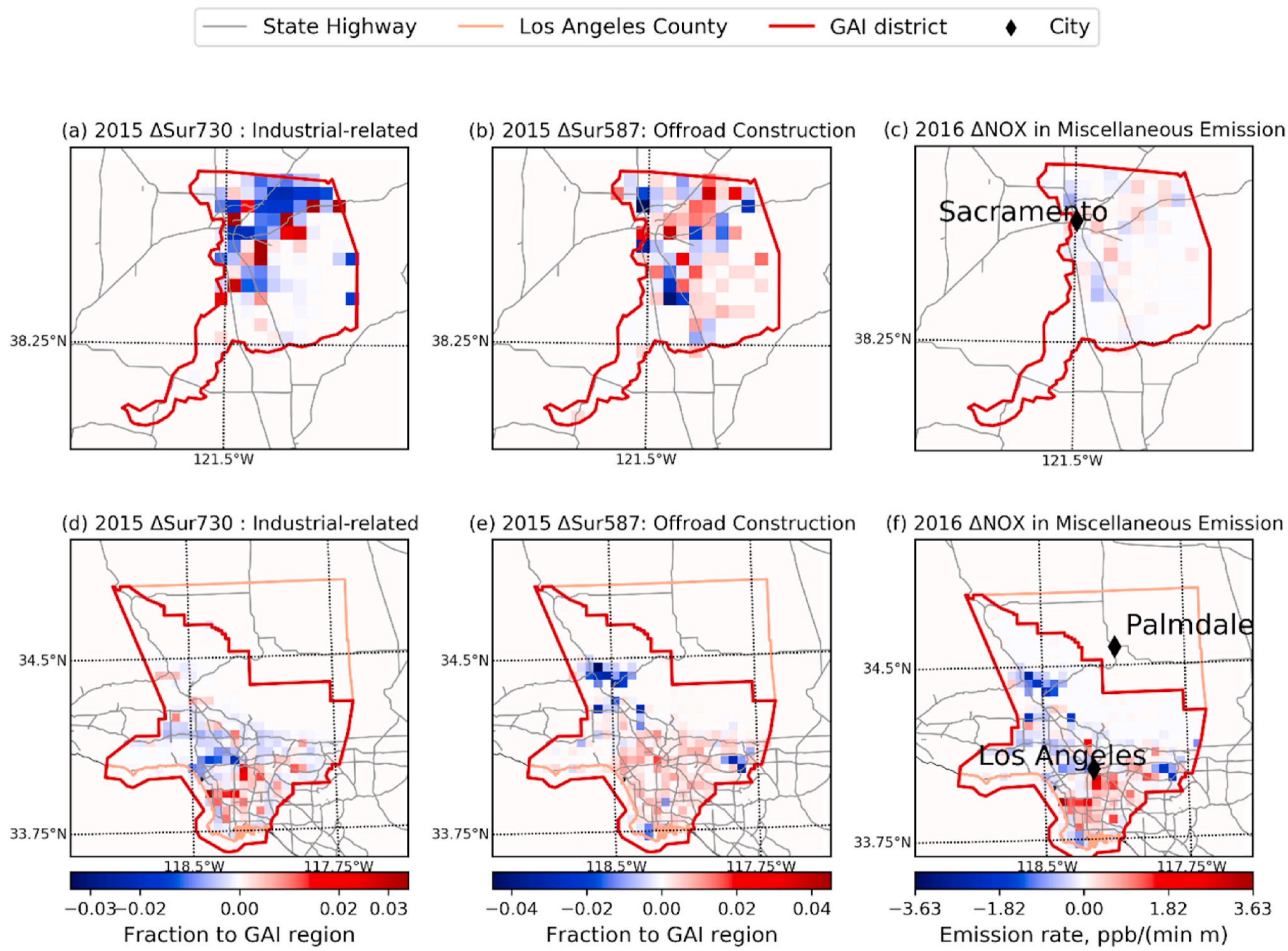

Fig. 8. Relationship between off-road/on-road construction surrogates and NOx emissions from miscellaneous sources at Sacramento county and Los Angeles GAI 6059. Figure (a) (b) (d) and (e) are surrogate differences between original 585 and updated 587/588. Figure (c) (f) are differences in NOx emissions from miscellaneous sources as defined in Section 2.4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

pavement condition, project history, climate, and anticipated future load to predict future construction projects for each mile of pavement in California, and can be used to plan for the rehabilitation and reconstruction of all state highways. PaveM is directly integrated into the decision making and optimization tools used by Caltrans to allocate future funding. For on-road construction projects more than 10 years in the future, we adapted the PaveM methodology to create a simplified projection of long-term future highway maintenance needs. A statistical forecast model was created to calculate the probability of replacing each mile of roadway in California based on pavement type, pavement condition, project history (age), climate zone, and anticipated future Equivalent Single Axle Loads (ESAL). Each segment of replaced roadway can select between different pavement types with probabilities determined by the same underlying factors that determine likelihood of replacement. A series of randomized "Monte Carlo" simulations were conducted to predict on-road construction projects by year in California through the year 2050. The average emissions from the Monte Carlo simulations were adopted as the future year on-road construction surrogate (see section S2.2.2 for additional details). The randomized "Monte Carlo" simulations to a certain degree were able to capture that the road rehabilitations will likely happen on highways where there are higher annual ESALs, such as Los Angeles, the Bay Area, and several main highway in California (I-80, I-5, US101, and State Route 99), and more severe climate conditions (see Fig. S8).

2.3. SED surrogates

SED-derived surrogates distribute emissions related to human activities. Total population surrogate 440 serves as a default surrogate if no other surrogate is assigned. Seven SED-derived surrogates were updated in the current study: total households (250), industrial employment (300), total population (440), service & commercial employment (620), single-family households (650), total employment (744, new), and agricultural employment (745, new). Three datasets served as the basis for the new surrogates: (i) data from Metropolitan Planning Organizations (MPOs)/local Council of Governments (COGs), (ii) data from the Caltrans Statewide Travel Demand Model (CSTDM), and (iii) the Longitudinal Employer-Household Dynamics (LEHD) "OnTheMap" data (United State Census Bureau, 2019). MPOs/COGs are agencies created by federal law to provide regional planning and implementation of federal transportation funds to urbanized areas with more than 50,000 people. Eighteen MPOs/COGs are designated in California (see Table S2), accounting for approximately 98% of the state's population (CALCOG, 2019). CSTDM is a tool used to forecast all personal travel made by every California resident, plus all commercial vehicle travel (Cambridge Systematics Inc, 2014a, 2014b). CSTDM socioeconomic data includes population, housing, and employment within specific sectors. MPOs/COGs and CSTDM datasets have total population and housing data from the US census data. All three datasets have employment census data classified by the North American Industry Classification System (NAICS). NAICS is used by Federal statistical

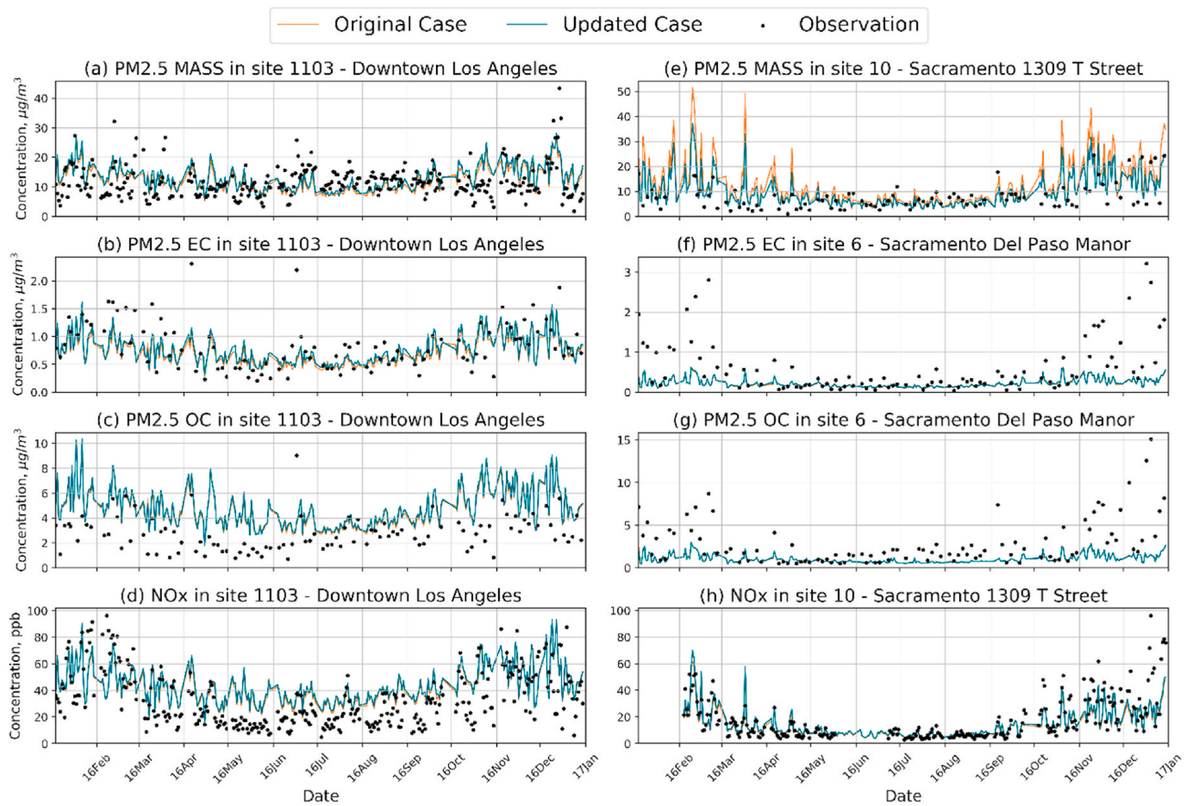

Fig. 9. Time series for predicted (original case in green line and updated case in orange line) and observed (black dot) PM2.5 mass, PM2.5 EC, PM2.5 OC and NOx concentrations at Los Angeles (shown in figure (a) (b) (c) (d)), Sacramento (shown in figure (e) (f) (g) (h))) during year 2016. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Annual PM2.5 mass mean fractional bias (MFB) and mean fractional error (MFE) for South California Sites. **Bold** Indicates improved performance.

Site	Downtown LA[a]		Reseda[a]		Compton[a]		Pico Rivera[a]	
Site Number	6037 1103		6037 1201		**6037 1302**		**6037 1602**	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	9.89%	39.70%	−34.81%	49.74%	−15.03%	42.35%	−27.58%	42.85%
Updated Case	16.40%	40.53%	−37.88%	51.84%	−8.05%	40.88%	−21.86%	39.78%
Site	Long Beach (South)[a]		Long Beach near Route 710[a]		Pasadena[a]		Long Beach[a]	
Site Number	6037 4004		**6037 4008**		**6037 2005**		**6037 4002**	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−0.50%	39.88%	−35.18%	45.70%	−10.46%	36.67%	−7.61%	42.73%
Updated Case	4.46%	40.35%	−29.56%	42.90%	−4.60%	35.87%	−2.06%	42.58%
Site	Anaheim[b]		Mission Viejo[b]		Rubidoux[c]		Mira Loma[c]	
Site Number	**6059 0007**		**6059 2022**		**6065 8001**		**6065 8005**	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−3.95%	39.11%	−8.53%	39.88%	−48.99%	59.19%	−56.07%	64.23%
Updated Case	0.45%	38.98%	−6.64%	39.42%	−47.66%	58.48%	−54.32%	62.87%
Site	Downtown San Diego[d]		San Diego military[d]		El Cajon[d]		Pala[d]	
Site Number	6073 1010		6073 1016		6073 1018		6073 1201	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	10.69%	38.59%	14.00%	41.77%	0.65%	31.24%	−25.26%	47.06%
Updated Case	41.91%	53.92%	16.51%	41.88%	4.24%	31.77%	−25.08%	46.88%

[a] Los Angeles county.
[b] Orange county.
[c] Riverside county.
[d] San Diego county.

Table 4
Annual PM2.5 mass mean fractional bias (MFB) and mean fractional error (MFE) for North California Sites – Sacramento and Santa Clara counties. **Bold** Indicates improved performance.

Site	Sacramento - Del Paso Manor[a]		Sacramento - 1309 T Street[a]		Folsom[a]		Sacramento Health Department[a]	
Site Number	6067 0006		**6067 0010**		6067 0012		**6067 4001**	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−17.99%	43.23%	38.24%	57.34%	−0.74%	53.45%	22.48%	47.26%
Updated Case	−18.76%	43.41%	22.18%	50.49%	−1.08%	53.60%	15.03%	45.25%
Site	San Jose - Konx Avenue[b]		San Jose – Jackson[b]		Gilory[b]			
Site Number	**6085 0005**		**6085 0006**		6085 0002			
Statistics	MFB	MFE	MFB	MFE	MFB	MFE		
Original Case	−6.40%	38.52%	−20.66%	37.53%	15.70%	52.06%		
Updated Case	−3.56%	38.04%	−17.95%	36.68%	−16.55%	54.01%		

[a] Sacramento county.
[b] Santa Clara county.

Table 5
Annual PM2.5 elemental carbon (EC) and PM2.5 organic carbon (OC) mean fractional bias (MFB) and mean fractional error (MFE) for all available California sites. **Bold** Indicates improved performance.

Site	Downtown LA[a]				Lebec (rural)[a]			
Site Number	6037 1103				6037 9034			
Species	**PM2.5 EC**		PM2.5 OC		PM2.5 EC		PM2.5 OC	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−5.49%	33.38%	57.13%	59.55%	−46.88%	86.60%	−72.51%	96.25%
Updated Case	0.64%	32.09%	59.30%	61.41%	−48.32%	87.73%	−73.17%	96.84%
Site	El Cajon[b]				El Cajon 2[b]			
Site Number	6073 1018				60731022			
Species	**PM2.5 EC**		**PM2.5 OC**		PM2.5 EC		PM2.5 OC	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−67.75%	76.55%	−12.86%	45.51%	−82.37%	83.22%	−67.05%	68.24%
Updated Case	−62.56%	72.31%	−9.30%	44.42%	−80.92%	81.60%	−67.09%	68.27%
Site	Sacramento - Del Paso Manor[c]				San Jose - Knox Avenue[d]			
Site Number	6067 0006				6085 0005			
Species	**PM2.5 EC**		PM2.5 OC		**PM2.5 EC**		PM2.5 OC	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−52.90%	70.77%	−64.07%	70.26%	−4.02%	44.73%	−12.31%	31.98%
Updated Case	−49.82%	68.97%	−63.84%	70.11%	−2.70%	44.48%	−13.23%	31.70%

[a] Los Angeles county.
[b] San Diego county.
[c] Sacramento county.
[d] Santa Clara county.

agencies to classify businesses for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy (https://www.census.gov/eos/www/naics/). LEHD provides annual employment statistics linking home and work locations (latitude/longitude) at the 2010 census block-level for individual NAICS categories, but the LEHD data is only available for historical years starting from 2002.

Updated SED surrogates were created for the base year 2010 and future years from 2015 to 2040, in 5 year increments. MPOs/COGs and CSTDM forecast future SED to multiple years up to 2050 (available years listed in Table S4) based on anticipated growth rate in each area in order to plan future infrastructure needs. Each of these local projections uses accepted practices for forecasting future trends. MPO/COG forecasts were interpolated in time as needed to produce uniform projections across California in target future years. Table S8 tests the accuracy of this time interpolation procedure for population and housing in the years 2010 and 2015 by comparing interpolated values to data from the United States Census for 2010 and 2015. Most of MPO/COG regions agree with the real census data within 5%. KERN, KINGS, SACOG, and SLOCOG had relatively minor errors (<10%), but the overall uncertainty introduced by time interpolation is still minor compared to the other uncertainty inherent in the surrogate projections.

The definitions of total population, total housing, single-family housing, and total employment for each MPO/COG and CSTDM are relatively consistent and so these spatial surrogates were derived directly from the variables provided by each new data source. Data from individual MPOs/COGs typically has better spatial resolution and has undergone more rigorous quality control than data from CSTDM. MPO/COG data was therefore used wherever possible, with CSTDM data filling in locations where MPO data was not available (see Section S2.3.1 for additional details). The approach used in the current project retains the fine-grain detail of the original MPO data wherever possible to increase the accuracy of the final spatial surrogate fields. MPO/COG data greatly enhances spatial resolution for moderately urbanized areas, such as Sacramento and central CA, but has little impact in highly urbanized areas, such as the Bay Area and Southern California.

Spatial surrogates describing employment are more complicated than the SED categories discussed above. The definitions for each employment surrogate are shown in Table 2. Each MPO/COG creates its own specialized grouped employment categories and/or modifies the definition of the standard NAICS employment categories to suit their own needs. The LEHD dataset is used in the current study to unify these heterogeneous fields into a standard set of spatial surrogates for employment in subcategories of agriculture, industry and service &

Table 6
Annual NOx mean fractional bias (MFB) and mean fractional error (MFE) for sites with changes in model performances. **Bold** Indicates improved performance.

Site	Compton[a]		Pico Rivera[a]		Long Beach (Hudson) [a]		Long Beach near Route 710[a]	
Site Number	6037 1302		**6037 1602**		**6037 4006**		**6037 4008**	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	26.70%	56.27%	−13.89%	43.17%	−6.14%	44.45%	−34.31%	62.46%
Updated Case	32.50%	58.27%	−7.41%	42.64%	−1.55%	43.46%	−29.29%	59.71%
Site	Santa Clarita[a]		Lancaster - Division Street[a]		Glendora[a]		Anaheim[b]	
Site Number	6037 6012		6037 9033		**6037 0016**		6059 0007	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−21.61%	43.19%	−60.96%	93.20%	−24.34%	41.79%	21.87%	53.14%
Updated Case	−40.17%	55.89%	−65.05%	97.37%	−15.42%	37.41%	25.79%	54.41%
Site	Anaheim Near-road[b]		Mira Loma[c]		Rubidoux[c]		Lake Elsinore[c]	
Site Number	**6059 0008**		**6065 8005**		**6065 8001**		6065 9001	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−73.97%	74.89%	−19.64%	44.63%	−16.37%	56.38%	−31.59%	50.69%
Updated Case	−70.38%	71.39%	−17.97%	44.10%	−14.96%	56.04%	−36.32%	54.19%
Site	Alpine[d]		Donovan[d]		El Cajon[d]		Sacramento-Goldenland Ct. [e]	
Site Number	**6073 1006**		6073 1014		**6073 1018**		**6067 0014**	
Statistics	MFB	MFE	MFB	MFE	MFB	MFE	MFB	MFE
Original Case	−43.25%	60.27%	−5.02%	55.20%	−26.48%	47.06%	49.31%	57.73%
Updated Case	−40.65%	58.13%	−14.79%	56.22%	−20.86%	43.72%	45.14%	55.06%

[a] Los Angeles county.
[b] Orange county.
[c] Riverside county.
[d] San Diego county.
[e] Sacramento county.

commercial within each county. The ratio of each individual LEHD sector to the LEHD lumped category total provides a profile that can be multiplied into the MPO/COG or CSTDM lumped categories to estimate employment which satisfy definitions in Table 2. An example given in Fig. 3 shows how to calculate agricultural employment category (NAICS code 11) in MPOs/COGs area Southern California Association of Governments (SCAG). SCAG combined mining employment (NAICS code 21) and agricultural employment in a single native field, which is the original variable (X_i') without invoking other datasets. Subscript i represents each geographic unit in MPOs/COGs or CSTDM, census block in MPOs/COGs and Traffic Area Zone in CSTDM. The ratio of agricultural employment (11) in SCAG data (r_i) can be estimated as LEHD agricultural sector (11) divided by LEHD lumped sector (11 + 21). Thus, the agricultural employment (11) in SCAG (X_i) can be calculated by multiplying the ratio of agricultural employment to the original variable ($X_i = X_i' * r_i$). These NAICS employment totals are then recombined into the categories defined in Table 2 (see Section S2.3.2 for additional details).

Figs. 1 and 4 illustrates SED surrogates in the years 2010 and spatial surrogate difference between 2040 and 2010. All current and future year SED surrogates are concentrated in urban areas including the Bay Area, Sacramento, Los Angeles, Fresno, and Bakersfield along Highway 99. The spatial patterns for socioeconomic surrogates (other than agricultural employment) expand outward from urban centers through year 2040.

2.4. Surrogate evaluation methods

Gridded surrogates were generated using the spatial allocator tool from United States Environmental Protection Agency (EPA) (CMAS, 2019). The spatial allocator is a set of programs that allows users to generate data files related to emissions and air quality modeling without requiring the use of a commercial Geographic Information System (GIS) package. Given a map projection, domain boundaries, and weight shapefiles (developed in Section 2), the spatial allocator can generate gridded spatial surrogates that can be directly used in emission models, such as the Sparse Matrix Operator Kernel Emissions (SMOKE) (CMAS, 2016). In this project, the spatial allocator was applied with the Lambert Conformal map projection (−120.5° longitude, 37° latitude, standard parallels 30° and 60°) and 4 km spatial resolution. Domain boundaries were specified based on the 69 Geographic Area Index (GAI) region shapefile (https://www.arb.ca.gov/ei/gislib/gislib.htm) which is consistent with previous statewide emissions inventories from ARB.

The raw spatial surrogates produced in the current project were combined with emissions inventories to analyze how the updates would influence the spatial distribution of actual emissions compared to the original CARB surrogates. These two cases are hereafter referred to as "updated" and "original".

The new spatial surrogates for the year 2015 were processed using SMOKE along with the raw emissions for the year 2016 to generate gridded emissions for six pollutants: CO, NOx, TOG, NH3, SOx, and PM. These emissions were then processed with the UCD emissions processing system (EMINV) to create model-ready inputs including speciated VOCs and size- and composition-resolved PM. Emissions were segregated into nine source categories for easier interpretation of the results: (i) on-road gasoline, (ii) off-road gasoline, (iii) on-road diesel, (iv) off-road diesel, (v) woodsmoke, (vi) food cooking, (vii) aircraft emissions, (viii) natural gas, and (ix) miscellaneous, which are emissions not included in the categories listed above. Emissions were used by the UCD/CIT air quality model applied to the entire state of California.

3. Results and discussion

3.1. Relationships between surrogates and emissions

The relationship between the spatial pattern of surrogates and emissions was analyzed in two detailed case studies for Sacramento county and Los Angeles GAI 6059. Sacramento has the largest shift of predicted PM mass concentration in response to the adoption of updated spatial surrogates. Los Angeles is one of the most populated areas in California. Each emissions source within these two study regions is

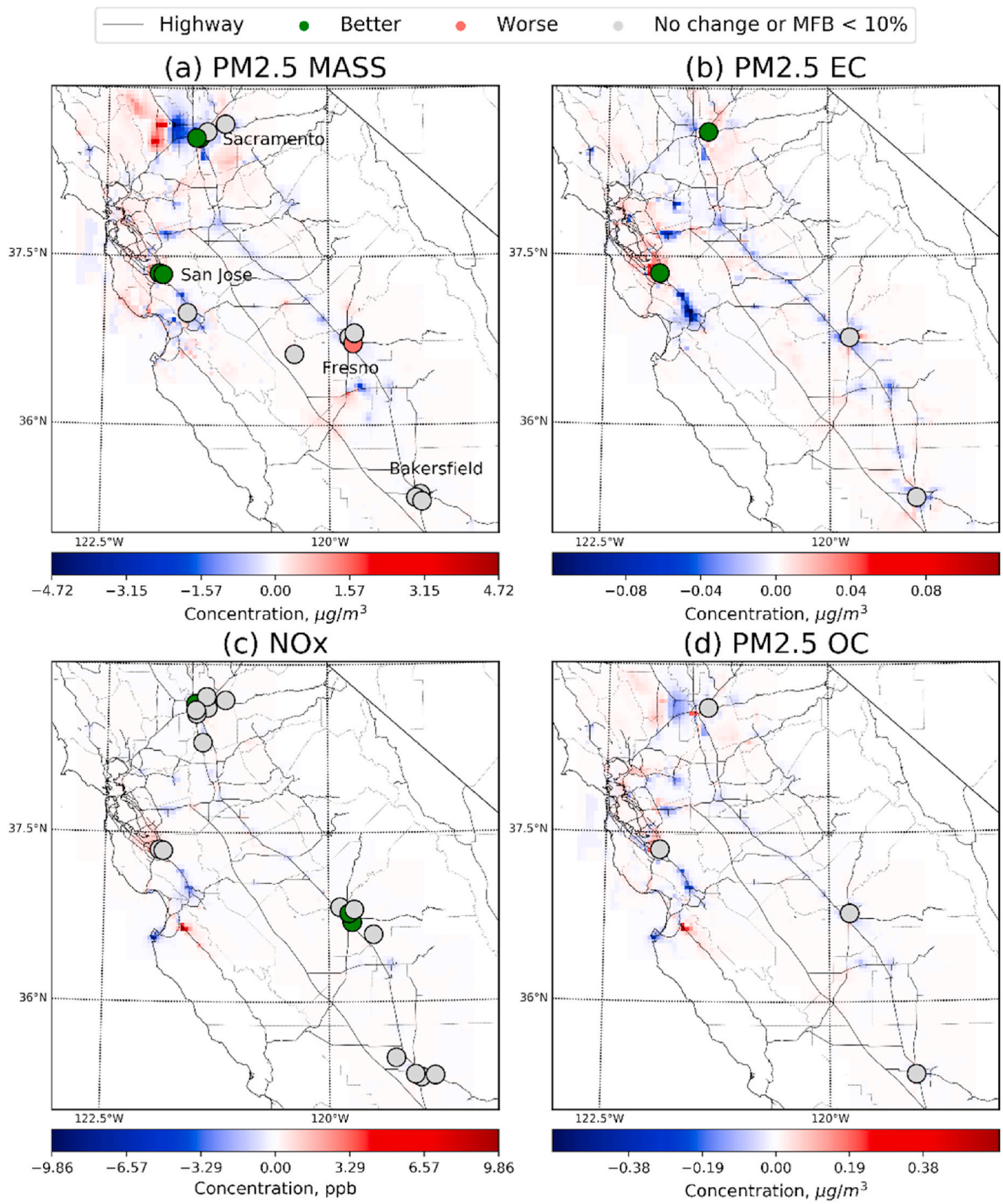

Fig. 10. Change in predicted ground-level concentrations due to the adoption of new spatial surrogates in norther California. Blue indicates concentrations decrease while red indicates concentration increase as shown by the key below each panel. Circles quantify change in model performance when compared to measurements at monitoring locations. Green circles indicate improved performance, red circles indicate degraded performance relative to the original case.

typically affected by a combination of spatial surrogates as described below.

Fig. 5 illustrates how off-road diesel emissions are mainly affected by updated off-road construction equipment surrogates 585 and 587 (see Fig. S9 and Fig. S12). Since surrogates only allocate emissions within each GAI region, Fig. 5(a) and (c) show the fractional change in each GAI total between original construction surrogate 585 and revised off-road construction surrogate 587. Fig. 5(b) and (d) show the change in absolute $PM_{2.5}$ EC emissions. A strong spatial correlation is apparent between changes to off-road construction equipment surrogate 587 and changes to $PM_{2.5}$ EC emissions from off-road diesel vehicles. Fig. 5(a) and (b) both reflect a major increase in the area northeast of Sacramento between two major highways. Fig. 5(c) and (d) also illustrate increasing emissions around major highways in the urban LA area, and major decrease northwest of LA. $PM_{2.5}$ EC from off-road diesel engines contributes strongly to total $PM_{2.5}$ EC concentrations. Changes to spatial surrogate 587 are therefore expected to significantly influence predicted overall $PM_{2.5}$ EC spatial patterns in cities.

California's emissions inventory treats a subset of the industrial and commercial natural gas combustion as area sources allocated using

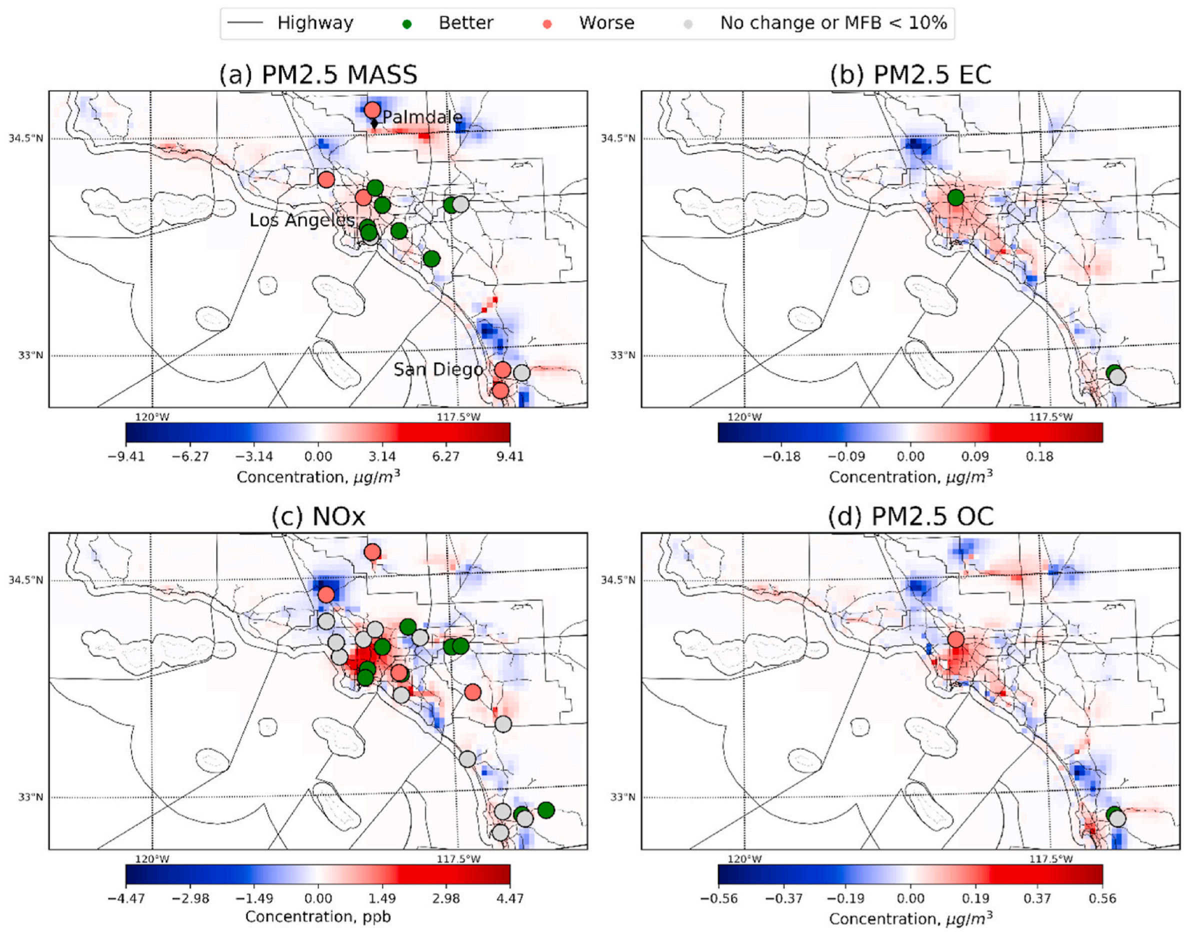

Fig. 11. Change in predicted ground-level concentrations due to the adoption of new spatial surrogates in southern California. Blue indicates concentrations decrease while red indicates concentration increase as shown by the key below each panel. Circles quantify change in model performance when compared to measurements at monitoring locations. Green circles indicate improved performance, red circles indicate degraded performance relative to the original case.

spatial surrogates 730 (industrial-related) and 620 (service & commercial employment), respectively. Fig. 6 and Fig. S15 illustrate how changes to updated surrogates 730 and 620 work together to change the pattern of natural gas combustion emissions in Los Angeles and Sacramento. Fig. 6(c) and (b) (d) show a strong spatial correlation between changes in surrogate 730 and natural gas combustion $PM_{2.5}$ OC and NOx emissions in Los Angeles. Note that increases in surrogate 730 are balanced by decreases in surrogate 620 in some locations (see Fig. 6(a) (c)). Changes to surrogate 730 result in significant changes to $PM_{2.5}$ OC and NOx in natural gas combustion emissions in LA area, ~0.35 μg/m2/min for $PM_{2.5}$ OC and ~1 ppb/min·m for NOx. The updated surrogate 730 also alters the spatial pattern of natural gas combustion emission in the San Francisco Bay Area (see Fig. S10). Updates to surrogate 620 modify the spatial pattern of natural gas combustion emissions in Sacramento (Fig. S15), but changes are modest because past census data already produced accurate spatial patterns for surrogate 620 in the original inventory.

Fig. 7 illustrates how changes to surrogate 587 and 588 influence the spatial pattern of $PM_{2.5}$ total mass emissions from miscellaneous sources. In general, off-road/on-road construction equipment surrogates 587/588 work together on changes in $PM_{2.5}$ total mass, and off-road surrogate 587 has a relatively larger impact than on-road surrogate 588. In Sacramento county (see Fig. 7(a) (b) and (c)), surrogates 587/588 have opposite change patterns northeast of Sacramento. The impact from changes to off-road surrogate 587 are weakened by changes to on-road surrogate 588. This causes a significant decrease for $PM_{2.5}$ total mass emissions in downtown Sacramento (~6 $\mu g/m^2$/min). In Los Angeles (see Fig. 7(d) (e) and (f)), off-road/on-road construction surrogates both increase PM mass emissions around the LA urban area, and decrease PM mass emissions to the northwest of downtown LA. The identical shift from surrogate 587/588 increased PM mass emissions in the LA urban area (~2.5 $\mu g/m^2$/min), although surrogate 587/588 has less change compared to other regions across the state. $PM_{2.5}$ total mass emissions from construction contributes strongly to total $PM_{2.5}$ mass emissions. Changes to construction spatial surrogate 587/588 are therefore expected to significantly influence the spatial pattern of predicted overall $PM_{2.5}$ total mass concentrations.

Fig. 8 shows how changes to off-road construction surrogate 587 and industrial-related 730 affect NOx emissions from miscellaneous sources differently across the state. In Sacramento county, changes in NOx follow the pattern of surrogate 587. However the absolute value of changes in NOx emissions is minor, because decreases in surrogate 587 are typically balanced by increases in surrogate 730 (see Fig. 8(a) (b) and (c)). Changes to the spatial pattern of miscellaneous NOx emissions stem from a combination of changes in surrogate 730 and surrogate 587 in the region surrounding Los Angeles, with a maximum shift of ~1.5 ppb/min·m. These patterns add to the shifts in the spatial pattern of NOx emissions associated with natural gas combustion (Fig. 6).

The specific trends illustrated in Figs. 5–8 are generally apparent throughout California. Changes for SED spatial surrogates are concentrated in cities or along highways where population is highest. In contrast, most of the changes in the spatial distribution of the industrial surrogates are found outside of the major urban areas. The off-road construction spatial surrogate 587 reflects the shift to project-based records as opposed to changes in impervious surfaces and is more concentrated within cities associated with urban renovation projects (see Fig. S12). The on-road construction surrogate 588 also reflects the shift to describing emissions from individual road construction projects instead of evenly distributing road construction emissions along the entire Truck Network. The intensity of the construction spatial surrogate generally decreases slightly along major highways throughout California but increases in isolated locations along those highways (Fig. S13). The updated industrial-related surrogate 730 generally concentrates industrial activity from a diffuse region into a concentrated source, perhaps associated with a central facility (see Fig. S14). The updated surrogates mainly change the location of off-road diesel emissions, natural gas emissions, and miscellaneous emissions. In general, "miscellaneous emissions" of $PM_{2.5}$ total mass, and NOx have the strongest response to the adoption of updated spatial surrogates (Fig. S11).

3.2. Surrogates performance: air quality model predictions

The updated spatial surrogates created in the current study were tested using the source-oriented UCD-CIT air quality model (Kleeman and Cass, 2001; Ying et al., 2008) for California during the year 2016. A single 24 km domain covering the entire state and two nested 4 km domains covering major population centers in northern California and southern California were selected for the analysis. Model simulations were carried out using both the original spatial surrogates and the updated spatial surrogates. Measurement data was downloaded from EPA website: https://aqs.epa.gov/aqsweb/airdata/download_files.html. There are 17 measurement sites in the southern California domain and 13 measurement sites in the northern California domain to evaluate $PM_{2.5}$ model performance; six sites across the state are available to evaluate $PM_{2.5}$ EC and OC predictions; 31 sites in southern California and 18 measurement sites in northern California are available to evaluate NOx predictions.

Fig. 9 shows the time series of $PM_{2.5}$ mass, EC, OC and NOx daily average concentration at central Los Angeles and Sacramento during the year 2016 using the original spatial surrogates (orange line) and the updated spatial surrogates (green line). Observed values are illustrated as black dots. Predicted concentrations based on the original and updated spatial surrogates are similar. Both cases capture the routine $PM_{2.5}$ mass, EC, OC and NOx concentrations with reasonable accuracy but they fail to capture the peak $PM_{2.5}$ mass concentration events which mostly occur in wintertime. Total $PM_{2.5}$ mass has relatively greater differences between original and updated cases compared to species such as $PM_{2.5}$ EC. Changes to spatial surrogates have minor effect on predicted $PM_{2.5}$ OC and NOx concentrations at the 2 measurement sites.

Statistical analysis was carried out for $PM_{2.5}$ mass, $PM_{2.5}$ EC, $PM_{2.5}$ OC, NOx, $PM_{2.5}$ nitrate, $PM_{2.5}$ sulfate, and two metals – Cu and Fe at all available measurement sites. $PM_{2.5}$ nitrate, sulfate, Fe and Cu do not respond strongly to the updated spatial surrogates and will not be discussed further in the present analysis. $PM_{2.5}$ mass Mean Fractional Bias (MFB) and Mean Fractional Error (MFE) are shown in Table 3 for southern California and Table 4 for northern California. $PM_{2.5}$ EC and OC MFB/MFE for all available California sites are shown in Table 5. NOx MFB/MFE for measurement sites with changed model performance are listed in Table 6. In general, air quality simulations carried out over the entire year 2016 determined that the effects of the updated spatial surrogates on predicted PM and NOx concentrations at measurement sites across the state are positive in most populated area, including the South Coast Air Basin (SoCAB) (including Los Angeles, Orange and Riverside counties), the region surrounding Sacramento, and the region south of San Francisco. $PM_{2,5}$ total mass has ~5% of improvement at most locations sites in the SoCAB (see Table 3); ~10% of improvement at downtown Sacramento (see site 6067 0010 and 6067 4001 in Table 4); and ~3% of improvement at San Jose (south of San Francisco). $PM_{2,5}$ EC improves by ~4% at two sites in southern CA, and ~2% at two sites in northern CA (see Table 5). NOx model performance improves (~6%) southeast of LA county, where industrial-related surrogate 730 has the largest changes (see Table 6). Updated spatial surrogates have minor impact on predicted concentrations of NOx in central and northern California. Likewise, updated spatial surrogates have minor impact on predicted concentrations of $PM_{2,5}$ mass, OC, and EC in central California (including Fresno and Bakersfield).

The spatial pattern of change caused by the adoption of the updated spatial surrogates can be understood more clearly by plotting the results over the entire region. Fig. 10 shows the change in the predicted annual-average $PM_{2.5}$ mass, EC, OC and NOx concentrations in the northern California domain due to the adoption of the updated spatial surrogates. Red grid squares represent areas of increased concentrations while blue grid squares represent areas of decreased concentration. The measurement sites in the model domain are represented as circles, with green coloring indicating improved performance and red coloring indicating degraded performance due to the adoption of the new surrogates. In Fig. 10(a), $PM_{2.5}$ mass concentrations decrease by 1–2 μg m^{-3} in the region west of Sacramento in response to the updated surrogates, leading to a 8–14% improvement at two sites (60670010, 60674001) in Sacramento. $PM_{2.5}$ EC predictions in Sacramento also improve by ~3% (Fig. 10(b)). Only one site (60670014) out of six in Sacramento has ~4% improvement for NOx (Fig. 10(c)). $PM_{2.5}$ concentrations in San Jose are also influenced by the updated spatial surrogates, with generally higher $PM_{2.5}$ mass concentrations throughout the urban region. The two measurement sites in San Jose both have ~3% improvement. $PM_{2.5}$ EC in San Jose slightly increased and improved ~1.5% in response to the updated spatial surrogates.

Fig. 11 displays the change in predicted annual-average $PM_{2.5}$ mass, EC and OC concentrations attributed to the updated spatial surrogates in southern California. Updated spatial surrogates have relatively larger impact in southern California compared to northern California. Fig. 11 (a) shows that $PM_{2.5}$ mass concentrations in the central region of Los Angeles increase by approximately 0.5 μg m^{-3} when the updated spatial surrogates are adopted, bringing the predictions into closer agreement with measured values at stations throughout this area. The overall spatial trends shown in Fig. 11(b) and (c) are similar to the trends shown in Fig. 11(a), but the performance of $PM_{2.5}$ EC improves (or does not change) at all available measurement sites when the updated spatial surrogates are adopted. $PM_{2.5}$ OC has little response to updated surrogates. EC is a primary PM component and OC is dominated by primary emissions in the current simulations. Both primary emissions and secondary formation contribute to total $PM_{2.5}$ mass. This suggests that the complex pattern of increasing and decreasing performance illustrated in Fig. 11(a) may be related to secondary PM formation rates and offsetting model errors. Fig. 11(c) shows that NOx concentrations increase by approximately 2.5 ppb in Los Angeles area in updated case, bring the prediction into closer agreement with measured values in this area.

The PM total mass, EC, OC and gas-phase NOx concentration results summarized in Figs. 10 and 11 are consistent with the changes in off-road diesel, natural gas combustion, and miscellaneous emissions discussed in Section 2. These patterns indicate that proximity to sources is a dominant factor that determines the impact of spatial surrogates on model performance. Off-road construction surrogate 587 and on-road construction surrogate 588 induce the largest change in predicted PM concentrations, followed by more modest changes associated with industrial-related surrogate 730, service & commercial employment surrogate 620 and single-family housing surrogate 650. Off-road construction surrogate 587 and industrial-related surrogate 730 induce the largest change in predicted NOx concentrations in southern California. Altered concentrations are associated with emissions from construction equipment, natural gas combustion, and industrial processes.

4. Conclusions

Spatial surrogates including total population, total housing, single family housing, total employment, service & commercial employment, industrial employment, agricultural employment, industrial-related, off-road construction, and on-road construction were updated for use with California emissions inventories. SED surrogates were updated using the latest version of census-based datasets at finer resolution. Off-road construction, on-road construction and industrial-related surrogates were developed using new methods to more accurately describe the location of construction projects and industrial facilities. All surrogates were created for the past years 2010 and 2015 and projected in 5- year increments to the year 2040.

The changes to the off-road construction spatial surrogate caused the largest shift in the distribution of PM emissions in the year 2015, followed by changes to the on-road construction spatial surrogate. These changes logically manifested as altered emissions patterns associated with construction sources. The changes to NOx emissions varied with location. In southern California, the changes to the industrial-related surrogate resulted in the largest shift in the distribution of NOx emissions in the year 2015. In northern California, changes to service & commercial employment, off-road construction equipment, and industrial-related surrogates are balanced leading to little impact on the spatial pattern of NOx emissions. The redistribution of industrial emissions based on a more exact description of industrial employment resulted in some isolated shifts in industrial emissions but no systematic pattern was observed. Changes in the spatial distribution of SED-derived surrogates, caused a slight reduction of emissions in the core of small cities and an increase in emissions in surrounding areas. SED changes were subtle and did not significantly influence emissions.

Air quality simulations carried out over the entire year 2016 determined that the updated spatial surrogates generally improve predicted PM mass and EC concentrations in Sacramento area (~10%), the Bay Area (~3%), and the region surrounding Los Angeles (~5%). Adoption of the updated spatial surrogates also improved predicted NOx concentrations in the core region of Los Angeles (~6%). These improvements demonstrate that adoption of new methodologies to estimate the location of construction equipment related surrogates (separate to off-road and on-road) and industrial-related surrogates are feasible at 4 km spatial resolution. Moreover, the updated construction-related and industrial-related surrogates may be suitable for even greater spatial resolution in future studies. The methods used to increase the accuracy of emissions locations in the current years may be extended to study emissions predictions and public health effects in future years.

It should be noted that spatial surrogates are an approximate approach for distributing area-source emissions. Even perfectly accurate spatial surrogates may not be perfectly correlated with emissions rates, and so this method has inherent uncertainty that varies depending on the exact emissions sources. Future applications of image recognition and GPS data may enable more accurate tracking of detailed activities that generate area-source emissions, but appropriate safeguards must be used to balance privacy vs. utility before widespread adoption of these techniques. Future studies should investigate these issues in order to improve the accuracy of area-source emissions inventories.

CRediT authorship contribution statement

Yiting Li: Writing - original draft. **Caroline Rodier:** Methodology. **Jeremy D. Lea:** Methodology. **John Harvey:** Methodology. **Michael J. Kleeman:** Conceptualization, Software, Formal analysis, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the California Air Resources Board under contract 15AQP009. The statements and conclusions in this manuscript are those of the authors and not necessarily those of the California Air Resources Board.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2020.117665.

References

Bieser, J., Aulinger, A., Matthias, V., Quante, M., Builtjes, P., 2011. Geoscientific Model Development SMOKE for Europe-adaptation, modification and evaluation of a comprehensive emission model for Europe. Geosci. Model Dev. (GMD) 4, 47–68. https://doi.org/10.5194/gmd-4-47-2011.

Bun, R., Hamal, K., Gusti, · M., Bun, · A., 2010. Spatial GHG inventory at the regional level: accounting for uncertainty. Climatic Change 103, 227–244. https://doi.org/10.1007/s10584-010-9907-5.

CALCOG, 2019. CALGOG main webpage. WWW Document. https://www.calcog.org/index.php?submenu=TypeofRegions&src=gendocs&ref=TypeofRegions&category=RegionsLead.

California Air Resource Board, 2020. OFFROAD2017 (v1.0.1). Web Database [WWW Document]. URL. https://www.arb.ca.gov/orion/?_ga=2.241139715.1886892688.1581021011-2036378939.1578081391. accessed 2.6.20.

California Air Resource Board, 2010. 2010 OFFROAD Emission Inventory Document - Construction and Mining Sector.

Caltrans, 2015. 2015 State of the Pavement Report.

Cambridge Systematics Inc, 2014a. California statewide travel demand model version 2.0 - model overview final report.

Cambridge Systematics Inc, 2014b. California Statewide Travel Demand Model, Version 2.0 - Population, Employment, and School Enrollment Final Report.

Chen, G., Li, J., Ying, Q., Sherman, S., Perkins, N., Sundaram, R., Mendola, P., 2014. Evaluation of observation-fused regional air quality model results for population air pollution exposure estimation. Sci. Total Environ. 485–486, 563–574. https://doi.org/10.1016/j.scitotenv.2014.03.107.

CMAS, 2019. Spatial allocator user's manual. WWW Document. https://github.com/CMASCenter/Spatial-Allocator/blob/master/docs/User_Manual/README.md. accessed 8.9.19.

CMAS, 2016. SMOKE v4.0 User's Manual.

Cohan, D.S., Hu, Y., Russell, A.G., 2006. Dependence of ozone sensitivity analysis on grid resolution. Atmos. Environ. 40, 126–135. https://doi.org/10.1016/j.atmosenv.2005.09.031.

Dai, J., Rocke, D.M., 2000. A GIS-based approach to spatial allocation of area source solvent emissions. Environ. Model. Software 15, 293–302. https://doi.org/10.1016/S1364-8152(00)00004-9.

Eastham, S.D., Long, M.S., Keller, C.A., Lundgren, E., Yantosca, R.M., Zhuang, J., Li, C., Lee, C.J., Yannetti, M., Auer, B.M., Clune, T.L., Kouatchou, J., Putman, W.M., Thompson, M.A., Trayanov, A.L., Molod, A.M., Martin, R.V., Jacob, D.J., 2018. GEOS-Chem High Performance (GCHP v11-02c): a next-generation implementation of the GEOS-Chem chemical transport model for massively parallel applications. Geosci. Model Dev. (GMD) 11, 2941–2953. https://doi.org/10.5194/gmd-11-2941-2018.

Fameli, K.M., Assimakopoulos, V.D., 2015. Development of a road transport emission inventory for Greece and the Greater Athens Area: effects of important parameters. Sci. Total Environ. 505, 770–786. https://doi.org/10.1016/j.scitotenv.2014.10.015.

Fenech, S., Doherty, R.M., Heaviside, C., Vardoulakis, S., Macintyre, H.L., O'Connor, F. M., 2018. The influence of model spatial resolution on simulated ozone and fine particulate matter for Europe: implications for health impact assessments. Atmos. Chem. Phys. 18, 5765–5784. https://doi.org/10.5194/acp-18-5765-2018.

Fu, M., Kelly, J.A., Clinch, J.P., 2017. Estimating annual average daily traffic and transport emissions for a national road network: a bottom-up methodology for both nationally-aggregated and spatially-disaggregated results. J. Transport Geogr. 58, 186–195. https://doi.org/10.1016/j.jtrangeo.2016.12.002.

Gkatzoflias, D., Mellios, G., Samaras, Z., 2013. Development of a web GIS application for emissions inventory spatial allocation based on open source software tools. Comput. Geosci. 52, 21–33. https://doi.org/10.1016/J.CAGEO.2012.10.011.

González, C.M., Ynoue, R.Y., Vara-Vela, A., Rojas, N.Y., Aristizábal, B.H., 2018. High-resolution air quality modeling in a medium-sized city in the tropical Andes: assessment of local and global emissions in understanding ozone and PM 10 dynamics. Atmos. Pollut. Res. 9, 934–948. https://doi.org/10.1016/j.apr.2018.03.003.

Herrera, J.C., Work, D.B., Herring, R., Ban, X., Jacobson, Q., Bayen, A.M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment. Transport. Res. C Emerg. Technol. 18, 568–583. https://doi.org/10.1016/j.trc.2009.10.006 (Jeff).

Hogrefe, C., Rao, S.T., 2001. Demonstrating attainment of the air quality standards: integration of observations and model predictions into the probabilistic framework. J. Air Waste Manag. Assoc. 51, 1060–1072. https://doi.org/10.1080/10473289.2001.10464332.

Hu, J., Jathar, S., Zhang, H., Ying, Q., Chen, S.-H., Cappa, C.D., Kleeman, M.J., 2017. Long-term particulate matter modeling for health effect studies in California – Part 2: concentrations and sources of ultrafine organic aerosols. Atmos. Chem. Phys. 17, 5379–5391. https://doi.org/10.5194/acp-17-5379-2017.

Huang, R., Zhai, X., Ivey, C.E., Friberg, M.D., Hu, X., Liu, Y., Mulholland, J.A., Russell, A. G., 2018. Using air quality model-data fusion methods for developing air pollutant exposure fields and comparison with satellite AOD-derived fields: application over North Carolina, USA. In: Mensink, C., Kallos, G. (Eds.), Air Pollution Modeling and its Application XXV. Springer International Publishing, Cham, pp. 207–212.
Joe, D.K., Zhang, H., DeNero, S.P., Lee, H.H., Chen, S.H., McDonald, B.C., Harley, R.A., Kleeman, M.J., 2014. Implementation of a high-resolution source-oriented WRF/chem model at the port of oakland. Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2013.09.055.
Kelly, J.T., Jang, C.J., Timin, B., Gantt, B., Reff, A., Zhu, Y., Long, S., Hanna, A., 2019. A system for developing and projecting PM2.5 spatial fields to correspond to just meeting national ambient air quality standards. Atmos. Environ. X 2. https://doi.org/10.1016/j.aeaoa.2019.100019.
Kleeman, M.J., Cass, G.R., 2001. A 3D eulerian source-oriented model for an externally mixed aerosol. Environ. Sci. Technol. 35, 4834–4848. https://doi.org/10.1021/es010886m.
Kuenen, J.J.P., Visschedijk, A.J.H., Jozwicka, M., Denier Van Der Gon, H.A.C., 2014. TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling. Atmos. Chem. Phys. 14, 10963–10976. https://doi.org/10.5194/acp-14-10963-2014.
Kuik, F., Lauer, A., Churkina, G., Denier Van Der Gon, H.A.C., Fenner, D., Mar, K.A., Butler, T.M., 2016. Air quality modelling in the Berlin-Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data. Geosci. Model Dev. (GMD) 9, 4339–4363. https://doi.org/10.5194/gmd-9-4339-2016.
Laurent, O., Hu, J., Li, L., Cockburn, M., Escobedo, L., Kleeman, M.J., Wu, J., 2014. Sources and contents of air pollution affecting term low birth weight in Los Angeles County, California, 2001–2008. Environ. Res. 134, 488–495. https://doi.org/10.1016/j.envres.2014.05.003.
Laurent, O., Wu, J., Li, L., Chung, J., Bartell, S., 2013. Investigating the association between birth weight and complementary air pollution metrics: a cohort study. Environ. Health (Nagpur) 12, 1–15. https://doi.org/10.1186/1476-069x-12-18.
Li, Y., Henze, D.K., Jack, D., Kinney, P.L., 2016. The influence of air quality model resolution on health impact assessment for fine particulate matter and its components. Air Qual. Atmos. Heal. 9, 51–68. https://doi.org/10.1007/s11869-015-0321-z.
Macpherson, A.J., Simon, H., Langdon, R., Misenheimer, D., 2017. A mixed integer programming model for National Ambient Air Quality Standards (NAAQS) attainment strategy analysis. Environ. Model. Software 91, 13–27. https://doi.org/10.1016/j.envsoft.2017.01.008.
McDonald, B.C., Gouw, J.A. de, Gilman, J.B., Jathar, S.H., Akherati, A., Cappa, C.D., Jimenez, J.L., Lee-Taylor, J., Hayes, P.L., McKeen, S.A., Cui, Y., Kim, S.-W., Gentner, D.R., Isaacman-VanWertz, G., Goldstein, A.H., Harley, R.A., Frost, G.J., Roberts, J.M., Ryerson, T.B., Trainer, M., 2018. Volatile chemical products emerging as largest petrochemical source of urban organic emissions. Am. Assoc. Adv. Sci. 359 https://doi.org/10.1126/science.aaq0524.
McDonald, B.C., McBride, Z.C., Martin, E.W., Harley, R.A., 2014. High-resolution mapping of motor vehicle carbon dioxide emissions. J. Geophys. Res. Atmos. 119, 5283–5298. https://doi.org/10.1002/2013jd021219.
Millstein, D.E., Harley, R.A., 2009. Revised estimates of construction activity and emissions: effects on ozone and elemental carbon concentrations in southern California. Atmos. Environ. 43, 6328–6335. https://doi.org/10.1016/j.atmosenv.2009.09.028.
Ostro, B., Hu, J., Goldberg, D., Reynolds, P., Hertz, A., Bernstein, L., Kleeman, M.J., 2015. Associations of mortality with long-term exposures to fine and ultrafine particles, species and sources: results from the California teachers study cohort. Environ. Health Perspect. 123, 549–556. https://doi.org/10.1289/ehp.1408565.
Pan, S., Choi, Y., Roy, A., Jeon, W., 2017. Allocating emissions to 4 km and 1 km horizontal spatial resolutions and its impact on simulated NOx and O3 in Houston, TX. Atmos. Environ. 164, 398–415. https://doi.org/10.1016/j.atmosenv.2017.06.026.
Sandanayake, M., Zhang, G., Setunge, S., Li, C.Q., Fang, J., 2016. Models and Method for Estimation and Comparison of Direct Emissions in Building Construction in Australia and a Case Study. Energy Build. https://doi.org/10.1016/j.enbuild.2016.05.007.
Saylor, R.D., Chameides, W.L., Chang, M.E., 1999. Demonstrating Attainment in Atlanta Using Urban Airshed Model Simulations: Impact of Boundary Conditions and Alternative Forms of the NAAQS. Atmospheric Environment.
Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J.M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M.T., Rouïl, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., Wind, P., 2015. Performance of European chemistry transport models as function of horizontal resolution. Atmos. Environ. 112, 90–105. https://doi.org/10.1016/j.atmosenv.2015.04.003.
Stieb, D.M., Chen, L., Hystad, P., Beckerman, B.S., Jerrett, M., Tjepkema, M., Crouse, D. L., Omariba, W.D., Peters, P.A., Donkelaar, A. van, Martin, R.V., Burnett, R.T., Liu, S., Smith-Doiron, M., Dugandzic, R.M., 2016. A national study of the association between traffic-related air pollution and adverse pregnancy outcomes in Canada, 1999–2008. Environ. Res. 148, 513–526. https://doi.org/10.1016/j.envres.2016.04.025.
Tan, J., Zhang, Y., Ma, W., Yu, Q., Wang, J., Chen, L., 2015. Impact of spatial resolution on air quality simulation: a case study in a highly industrialized area in Shanghai, China. Atmos. Pollut. Res. 6, 322–333. https://doi.org/10.5094/apr.2015.036.
Thompson, T.M., Selin, N.E., 2012. Atmospheric Chemistry and Physics Influence of air quality model resolution on uncertainty associated with health impacts. Atmos. Chem. Phys. 12, 9753–9762. https://doi.org/10.5194/acp-12-9753-2012.
Trombetti, M., Thunis, P., Bessagnet, B., Clappier, A., Couvidat, F., Guevara, M., Kuenen, J., López-Aparicio, S., 2018. Spatial inter-comparison of Top-down emission inventories in European urban areas. Atmos. Environ. 173, 142–156. https://doi.org/10.1016/J.ATMOSENV.2017.10.032.
United State Census Bureau, 2019. OnTheMap: Data Overview (LODES Version 7).
US EPA, 2017. Emissions Inventory Guidance for Implementation of Ozone and Particulate Matter National Ambient Air Quality Standards (NAAQS) and Regional Haze Regulations.
Valari, M., Menut, L., 2008. Does an increase in air quality models' resolution bring surface ozone concentrations closer to reality? J. Atmos. Ocean. Technol. 25, 1955–1968. https://doi.org/10.1175/2008jtecha1123.1.
Van Donkelaar, A., Martin, R.V., Spurr, R.J.D., Burnett, R.T., 2015. High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America. Environ. Sci. Technol. 49, 10482–10491. https://doi.org/10.1021/acs.est.5b02076.
Wang, M., Sampson, P.D., Hu, J., Kleeman, M., Keller, J.P., Olives, C., Szpiro, A.A., Vedal, S., Kaufman, J.D., 2016. Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and PM2.5. Environ. Sci. Technol. 50, 5111–5118. https://doi.org/10.1021/acs.est.5b06001.
Woody, M.C., Baker, K.R., Hayes, P.L., Jimenez, J.L., Koo, B., Pye, H.O.T., 2016. Understanding sources of organic aerosol during CalNex-2010 using the CMAQ-VBS. Atmos. Chem. Phys. 16, 4081–4100. https://doi.org/10.5194/acp-16-4081-2016.
Ying, Q., Lu, J., Allen, P., Livingstone, P., Kaduwela, A., Kleeman, M., 2008. Modeling air quality during the California Regional PM10/PM2.5 Air Quality Study (CRPAQS) using the UCD/CIT source-oriented air quality model – Part I. Base case model results. Atmos. Environ. 42, 8954–8966. https://doi.org/10.1016/J.ATMOSENV.2008.05.064.
Zasina, D., Zawadzki, J., 2017. Spatial surrogate for domestic combustion's air emissions: a case study from Silesian Metropolis, Poland. J. Air Waste Manag. Assoc. 67, 1012–1019. https://doi.org/10.1080/10962247.2017.1316327.
Zhang, L., Jacob, D.J., Downey, N.V., Wood, D.A., Blewitt, D., Carouge, C.C., van Donkelaar, A., Jones, D.B.A., Murray, L.T., Wang, Y., 2011. Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with $1/2^{\circ} \times 2/3^{\circ}$ horizontal resolution over North America. Atmos. Environ. 45, 6769–6776. https://doi.org/10.1016/j.atmosenv.2011.07.054.
Zhao, B., Wang, P., Ma, J.Z., Zhu, S., Pozzer, A., Li, W., 2012. A high-resolution emission inventory of primary pollutants for the Huabei region, China. Atmos. Chem. Phys. 12, 481–501. https://doi.org/10.5194/acp-12-481-2012.
Zheng, B., Zhang, Q., Tong, D., Chen, C., Hong, C., Li, M., Geng, G., Lei, Y., Huo, H., He, K., 2017. Resolution dependence of uncertainties in gridded emission inventories: a case study in Hebei, China. Atmos. Chem. Phys. 17, 921–933. https://doi.org/10.5194/acp-17-921-2017.
Zhou, Y., Zhao, Y., Mao, P., Zhang, Q., Zhang, J., Qiu, L., Yang, Y., 2017. Development of a high-resolution emission inventory and its evaluation and application through air quality modeling for Jiangsu Province, China. Atmos. Chem. Phys. 17, 211–233. https://doi.org/10.5194/acp-17-211-2017.