Source Apportionment of Ambient Methane Enhancements in Los Angeles, California to Evaluate Emission Inventory Estimates

Toshihiro Kuwayama1*, Jessica G. Charrier-Klobas1, Yanju Chen1, Nicholas M. Vizenor2, Donald R. Blake2, Thomas Pongetti3, Stephen A. Conley4, Stanley P. Sander3, Bart Croes1, and Jorn D. Herner1

1 California Air Resources Board, 1001 I Street, Sacramento, CA 95812 USA
2 University of California at Irvine, 570 Rowland Hall, Irvine, CA 92697 USA
3 NASA Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109 USA
4 Scientific Aviation, Inc., 3335 Airport Road Suite B, Boulder, CO 80301 USA

*Corresponding Author: Tel +1.916.324.9287
E-mail: toshihiro.kuwayama@arb.ca.gov

Supplemental Information

Summary
Total number of pages (including Table of Contents and List of Tables and Figures): 26
Total number of Tables: 18
Total number of Figures: 8
Table of Contents

List of Tables and Figures ..................................................................................................................3
Tables and Figures Directly Referenced in the Manuscript..............................................................5
Key assumptions for inventory- and measurement-based CH₄ emission estimates.........................13
Inventory-Based CH₄ Emission Estimation Approach: IB1 ..............................................................13
  Transportation ..............................................................................................................................13
  Population ....................................................................................................................................13
  Land Use ......................................................................................................................................14
Inventory-Based CH₄ Emission Estimation Approach: IB2 ..............................................................14
  100 Landfills ..............................................................................................................................16
  102 Compost ..............................................................................................................................16
  150 Pipeline Fugitives ..............................................................................................................16
  151 Oil and Gas ..........................................................................................................................16
  152 Petroleum Refining Wastewater .........................................................................................17
  160 Electricity Generation (in state) .........................................................................................17
  162 Refinery and Hydrogen ........................................................................................................17
  Dairy (general description) .........................................................................................................17
  210 Non-Dairy Cattle Feedlot, 211 Non-Dairy Range Cattle, 212 Horses, and 213 Poultry ....18
    210 Non-Dairy Cattle Feedlot .................................................................................................20
    211 Pasture Non-Dairy Cow Spatial Surrogate .................................................................21
    212 Horses .............................................................................................................................21
    213 Poultry ............................................................................................................................21
  251 Rice ......................................................................................................................................22
  301 Wastewater, Domestic – Anaerobic Digesters ....................................................................22
  302 Septic ....................................................................................................................................22
  303 Non-AD Wastewater Treatment (including wastewater recycling plants) ....................23
  40X On-Road Transportation ....................................................................................................23
  980 Seeps ....................................................................................................................................24
  990 Wetlands .............................................................................................................................24
Assessing the robustness of linear regression models using repetitive randomized sub-
sampling technique .........................................................................................................................24
Reference .........................................................................................................................................26
List of Tables and Figures

Table S1. List of whole-air canister sampling dates and times, where M = 0:00 – 1:00, A = 11:00 – 12:00, B = 13:00 – 14:00, C = 15:00 – 16:00, and D = 17:00 – 18:00.............5
Table S2. Chemical compounds measured at MWO. Highlighted compounds were used in the PMF analysis. ................................................................................................................5
Table S3. PMF source contributions to the average monthly ΔCH4. Values are reported in ppb with standard errors of the monthly means.................................................................6
Table S4. IB1 surrogate scaling factors for SCLA from 2012 to 2016. .........................................................14
Table S5. SCLA-specific annual CH4 emissions by sectors from 2012 to 2016 (IB1). Units in MMT CO2e/year. ..................................................................................................................14
Table S6. IB2 surrogate information summary. ..........................................................................................15
Table S7. Distribution of pipeline fugitive CH4 emissions. ........................................................................16
Table S8. Emission types according to well types for Oil and Gas activities. .............................................17
Table S9. Dairy Manure Management Spatial Surrogate Descriptions..................................................18
Table S10. Summary of data sources used for surrogates 210, 211, 212, and 213. ..............................19
Table S11. Land use codes for surrogates 210, 211, 212, and 213. .......................................................19
Table S12. Animal types used for surrogates 210 and 211. .................................................................20
Table S13. Land use codes for surrogates 210 ..............................................................20
Table S14. Land use codes for surrogates 212. .....................................................................................21
Table S15. Land use codes for surrogates 213. .....................................................................................22
Table S16. Land use codes for surrogates 213. .....................................................................................23
Table S17. QC results for MB1; CH4:CO correlation. ............................................................................25
Table S18. QC results for MB2; CH4:CO correlation. ............................................................................25

Figure S1. CH4:CO orthogonal regression analysis using real-time analyzer data. Empty boxes indicate missing data or monthly data that did not pass the quality assessment. ........7
Figure S2. CH4:CO orthogonal regression analysis using canister data. Empty boxes indicate missing data or monthly data that did not pass the quality assessment. ...........8
Figure S3. Ethane:CH4 regression of natural gas and landfills factors compared to signatures from CH4 sources identified in SCLA1 .................................................................9
Figure S4. Comparison of actual (model input) and modeled (PMF output) ΔCH4. Left panel (A) shows the correlation based on individual data points. The right panel (B) shows the comparison based on monthly sum of ΔCH4..........................................................9
Figure S5. Atmospheric conditions observed at downtown Los Angeles, California from July 2014 to April 2016.................................................................10
Figure S6. Average monthly ΔCH4 from the analysis of the whole-air canister samples and the high-precision, fast-response CH4 analyzer data from MWO. ..............................10
Figure S7. Seasonal trend of CH$_4$:CO orthogonal regressions. Data between 2014 and 2016 were averaged each month with the exception of October 2015 (Aliso Canyon CH$_4$ leak incident$^2$). Monthly data had RSQ > 0.8 for real-time data and RSQ > 0.5 for canister data. Winter months are assumed to be between November and February. Summer months are assumed to be between June and September.

Figure S8. Waste from Los Angeles County being disposed into respective counties (in tons).
### Tables and Figures Directly Referenced in the Manuscript

#### Table S1. List of whole-air canister sampling dates and times, where M = 0:00 – 1:00, A = 11:00 – 12:00, B = 13:00 – 14:00, C = 15:00 – 16:00, and D = 17:00 – 18:00.

| Week | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Year | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |

#### Table S2. Chemical compounds measured at MWO. Highlighted compounds were used in the PMF analysis.

| Method    | Compound     | Formula | LOD  | Accuracy | % Precision |
|-----------|--------------|---------|------|----------|-------------|
| Off-axis ICOS | Carbon monoxide | CO      | -    | -        | 0.1 ppb     |
| CRDS      | Methane (12C) | CH4     | -    | -        | 50 ppb +0.05% of reading |
| GC-FID    | Carbon monoxide | CO      | 5 ppb| 7%       | 3%          |
| GC-FID    | Carbon dioxide | CO2     | 1 ppb| 1 ppm    | 1%          |
| GC-FID    | Methane       | CH4     | 1 ppb| 3 ppb    | 1%          |
| GC-FID    | Ethane        | C2H6    | 3 ppt| 5%       | 1%          |
| GC-FID    | Ethene        | C2H4    | 3 ppt| 5%       | 2%          |
| GC-FID    | Ethyne        | C3H2    | 3 ppt| 5%       | 1%          |
| GC-FID    | Propene       | C3H6    | 3 ppt| 5%       | 5%          |
| GC-FID    | Propane       | C3H8    | 3 ppt| 5%       | 3%          |
| GC-FID    | i-Butane      | C4H10   | 3 ppt| 5%       | 4%          |
| GC-FID    | n-butane      | C4H10   | 3 ppt| 5%       | 1%          |
| GC-FID    | i-pentane     | C5H12   | 3 ppt| 5%       | 1%          |
| GC-FID    | n-pentane     | C5H12   | 3 ppt| 5%       | 6%          |
| GC-FID    | Isoprene      | C5H8    | 3 ppt| 5%       | 3%          |
| GC-FID    | n-hexane      | C6H14   | 3 ppt| 5%       | 6%          |
| GC-FID    | 2-methylpentane | C6H14  | 3 ppt| 5%       | 2%          |
| GC-FID    | 3-methylpentane | C6H14  | 3 ppt| 5%       | 6%          |
| GC-FID    | n-heptane     | C7H16   | 3 ppt| 5%       | 9%          |
| GC-FID    | Benzene       | C8H8    | 3 ppt| 5%       | 7%          |
| GC-FID    | Ethylbenzene  | C9H10   | 3 ppt| 5%       | 5%          |
| GC-FID    | m/p-xylene    | C9H10   | 3 ppt| 5%       | 4%          |
| GC-FID    | o-xylene      | C9H10   | 3 ppt| 5%       | 6%          |
Table S3. PMF source contributions to the average monthly ΔCH\textsubscript{4}. Values are reported in ppb with standard errors of the monthly means.

| Month-Year | Factor 1 | Factor 2 | Factor 3 | Factor 5 | Factor 6 | Factor 7 |
|------------|----------|----------|----------|----------|----------|----------|
|            | Landfills | Petroleum Gas and Industrial Processes | FF\textsubscript{CS} | FF\textsubscript{HR} | Natural Gas | Biogenic |
| Jul-14     | 12.7±1.7  | 0.2±0.04  | 5.7±0.7  | 1.7±0.3  | 46.4±5.3  | 9.7±0.8  |
| Aug-14     | 18.8±2.5  | 0.4±0.06  | 7.0±0.5  | 2.8±0.2  | 65.4±3.8  | 4.2±0.2  |
| Sep-14     | 17.5±2.8  | 0.2±0.04  | 4.6±0.8  | 1.8±0.3  | 48.9±6.0  | 2.0±0.2  |
| Oct-14     | 16.6±2.9  | 0.3±0.04  | 5.0±0.7  | 2.2±0.3  | 62.8±6.8  | 1.3±0.1  |
| Nov-14     | 19.7±2.5  | 0.5±0.05  | 3.6±0.5  | 2.8±0.4  | 48.4±6.5  | 2.1±0.1  |
| Dec-14     | 14.5±1.9  | 0.3±0.03  | 2.3±0.3  | 2.7±0.2  | 36.1±3.2  | 2.0±0.1  |
| Jan-15     | 16.3±2.3  | 0.3±0.04  | 1.9±0.2  | 2.0±0.5  | 30.1±1.9  | 2.0±0.3  |
| Feb-15     | 19.2±3.2  | 0.5±0.11  | 4.1±1.0  | 3.1±0.4  | 55.0±10.7 | 3.2±0.3  |
| Mar-15     | 12.7±2.0  | 0.3±0.05  | 3.4±0.6  | 2.4±0.4  | 38.5±3.6  | 3.6±0.3  |
| Apr-15     | -         | -         | -        | -        | -         | -        |
| May-15     | 8.8±1.5   | 0.2±0.03  | 2.6±0.3  | 1.9±0.2  | 36.9±2.7  | 2.6±0.4  |
| Jun-15     | 25.5±3.3  | 0.8±0.20  | 6.2±1.1  | 2.9±0.5  | 58.3±8.3  | 9.9±0.6  |
| Jul-15     | 8.5±0.9   | 0.4±0.11  | 3.7±0.6  | 2.4±0.3  | 45.0±5.9  | 6.6±0.5  |
| Aug-15     | 5.8±1.1   | 0.2±0.03  | 5.5±0.5  | 2.5±0.3  | 65.3±5.2  | 7.3±0.6  |
| Sep-15     | 8.1±1.2   | 0.3±0.06  | 4.5±1.0  | 2.9±0.4  | 55.1±8.5  | 5.3±0.6  |
| Oct-15\*   | 7.1±1.0   | 0.2±0.03  | 3.3±0.4  | 2.9±0.3  | 65.2±19.6 | 3.7±0.4  |
| Nov-15\*   | 9.0±1.5   | 0.4±0.10  | 1.9±0.4  | 4.8±0.5  | 38.8±13.1 | 2.6±0.4  |
| Dec-15\*   | 9.4±1.2   | 0.5±0.08  | 1.9±0.3  | 3.8±0.5  | 65.3±7.9  | 1.5±0.1  |
| Jan-16\*   | 11.2±2.4  | 0.4±0.07  | 2.6±0.4  | 4.2±0.6  | 53.6±10.5 | 1.4±0.2  |
| Feb-16\*   | 6.9±0.8   | 0.4±0.04  | 2.5±0.4  | 2.6±0.4  | 45.1±5.4  | 2.5±0.2  |
| Mar-16     | 8.3±1.2   | 0.4±0.05  | 4.5±0.5  | 3.1±0.3  | 46.3±6.0  | 2.1±0.2  |
| Apr-16     | 11.2±1.7  | 0.4±0.10  | 2.9±1.0  | 2.7±0.7  | 29.0±6.8  | 3.8±0.6  |

* Asterisked months (October 2015 to February 2016) include data collected during the Aliso Canyon CH\textsubscript{4} leak incident. Although Aliso Canyon influences at Mt. Wilson were infrequent, the asterisked data should be used with caution.
Figure S1. CH$_4$:CO orthogonal regression analysis using real-time analyzer data. Empty boxes indicate missing data or monthly data that did not pass the quality assessment.
Figure S2. CH$_4$-CO orthogonal regression analysis using canister data. Empty boxes indicate missing data or monthly data that did not pass the quality assessment.
Figure S3. Ethane:CH$_4$ regression of natural gas and landfills factors compared to signatures from CH$_4$ sources identified in SCLA$^1$.

Figure S4. Comparison of actual (model input) and modeled (PMF output) ΔCH$_4$. Left panel (A) shows the correlation based on individual data points. The right panel (B) shows the comparison based on monthly sum of ΔCH$_4$. 
Figure S5. Atmospheric conditions observed at downtown Los Angeles, California from July 2014 to April 2016.

Figure S6. Average monthly $\Delta$CH$_4$ from the analysis of the whole-air canister samples and the high-precision, fast-response CH$_4$ analyzer data from MWO.
Figure S7. Seasonal trend of CH$_4$:CO orthogonal regressions. Data between 2014 and 2016 were averaged each month with the exception of October 2015 (Aliso Canyon CH$_4$ leak incident$^2$). Monthly data had RSQ > 0.8 for real-time data and RSQ > 0.5 for canister data. Winter months are assumed to be between November and February. Summer months are assumed to be between June and September.
Figure S8. Waste from Los Angeles County being disposed into respective counties (in tons).
Key assumptions for inventory- and measurement-based CH₄ emission estimates

1) CO is a representative surrogate that captures the CH₄ emissions behavior in well-mixed air mass that travels from SCLA.
2) Aggregated monthly orthogonal regression slopes can be used to calculate representative CH₄ emissions.
3) Scaling factors such as population, VMT, and land-use can be used to apportion regional emission contributions from various source sectors.
4) Regionalization of Statewide CH₄ and CO Emission Inventory to SCLA level leads to a representative inventory-based emission estimate.
5) CH₄ emissions are well-mixed and the distribution of CH₄ sources within SCLA do not strongly bias ambient CH₄ measured at MWO.
6) IB2 and MB2 comparison presented in Comparing Source Contributions to CH₄ Emissions section assumes proportionality of PMF-derived source distribution in describing CH₄ emission sources located within SCLA as a result of well-mixed condition.
7) IB2 and MB2 comparison presented in Comparing Source Contributions to CH₄ Emissions section assumes representativeness of PMF-derived source distribution in describing CH₄ emission sources in SCLA that extend toward central Los Angeles, with increased uncertainties and reduced influences from sources located near the coast.

Inventory-Based CH₄ Emission Estimation Approach: IB1

The IB1 approach utilized CARB’s 2017 Edition Statewide GHG Emission Inventory as the basis for calculating annual SCLA-specific CH₄ emission estimates. Below are information regarding the source of the surrogate data used to develop the scaling factors:

Transportation – EMFAC2014 (with EMFAC2011 vehicle categories) was used to query the annual VMT data for SCLA with aggregated travel speeds and by fuel types. In general, the annual VMT was primarily dominated by LDA, which accounting for over 50% of the activities. Since most of the CH₄ emissions in the transportation sector are attributed to gasoline sources in the inventory, the annual VMT for gasoline sources were used for further comparison. The ratios between the SCLA total and the state total were used as the scaling factor for CH₄ sources related to transportation sector.

Population – 2014 Population data from U.S. Census Bureau was used to calculate the population in SCLA. Since information specific to SCLA was not readily available, population from cities outside of the physical boundaries of SCLA were subtracted from the population data for Los Angeles County. The excluded cities accounted for less than 10% of the Los Angeles County population. The ratios between the SCLA total and the state total were used as the scaling factor for CH₄ sources related to commercial and residential, electric power generation, industrial, and recycling and waste sectors.
**Land Use** – 2014 land use data was obtained from the Department of Conservation, specifically to understand land use for farmland and grazing land (i.e., agriculture). The primary data was downloaded in GIS format to extract SCLA-specific information. The cumulative average of the farmland and grazing land data was used to calculate the ratio between the SCLA total and the state total. This was used as the scaling factor for CH$_4$ sources related to agricultural sector. In general, agricultural activities in SCLA accounted for less than 1% of the statewide total.

Table S4. IB1 surrogate scaling factors for SCLA from 2012 to 2016.

| Category            | Sub-Category                  | 2012   | 2013   | 2014   | 2015   | 2016   |
|---------------------|-------------------------------|--------|--------|--------|--------|--------|
| Transportation      | VMT                           | 22.1%  | 21.9%  | 21.7%  | 21.6%  | 21.4%  |
| Population          | Population                    | 25.4%  | 25.4%  | 25.3%  | 25.3%  | 25.3%  |
| Land Use            | Farmland + Grazing Land       | 0.4%   | 0.4%   | 0.4%   | 0.4%   | 0.3%   |

Table S5. SCLA-specific annual CH$_4$ emissions by sectors from 2012 to 2016 (IB1). Units in MMT CO$_2$e/year.

| Source Sector       | Surrogate                      | 2012   | 2013   | 2014   | 2015   | 2016   |
|---------------------|-------------------------------|--------|--------|--------|--------|--------|
| Agriculture         | Land Use                      | 0.09   | 0.08   | 0.08   | 0.08   | 0.08   |
| Commercial and Residential | Population       | 0.01   | 0.02   | 0.01   | 0.01   | 0.01   |
| Electric Power      | Population                    | 0.03   | 0.06   | 0.06   | 0.07   | 0.07   |
| Industrial          | Population                    | 1.80   | 1.80   | 1.84   | 1.80   | 1.79   |
| Recycling and Waste | Population                    | 2.19   | 2.19   | 2.21   | 2.22   | 2.24   |
| Transportation      | VMT                           | 0.05   | 0.05   | 0.05   | 0.04   | 0.04   |

**Inventory-Based CH$_4$ Emission Estimation Approach: IB2**

Information about the spatial surrogates used for IB2 is provided below. Each ID in the table is linked to information that was used to develop the regional surrogate(s) for corresponding sectors/sources. In general, the surrogates were used to regionalize the Statewide CH$_4$ Emission Inventory. A large amount of information is needed to spatially distribute state-level emissions with high level of confidence. In many cases, detailed data is not available for this purpose and relatively coarse assumptions must be made. This process necessarily adds complexity to inventory calculations and increases the uncertainty of any emission estimates. The variability among sources in different regions may not always be captured, thus accurate statewide estimates may become inaccurate when distributed to a specific region. Every effort has been made to identify the most accurate and useful surrogates and data sources in this spatial distribution, however, the data limitations must be acknowledged.
Table S6. IB2 surrogate information summary.

| ID   | Surrogate Name                  | Spatial Data Source                                                                 | Activity Data Source                                                                 |
|------|---------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| 100  | Landfills                       | CARB Landfill Model                                                                 | CARB Landfill Model facility-specific emissions                                      |
| 102  | Compost                         | CalRecycle                                                                          | CalRecycle facility-specific data                                                   |
| 150  | Pipeline Fugitives               | EIA Transmission Pipeline Map, Estimates of Distribution Pipeline Extent, Housing Units | Distributed by mile of pipeline, housing density, or among housing units |
| 151a-151c | Oil & Gas Extraction              | DOGGR                                                                               | Apportioned among wells by type                                                    |
| 152  | Petroleum Refining Wastewater    | State Water Board CIWQS                                                              | State Water Board CIWQS                                                             |
| 162  | Refinery and Hydrogen            | CARB MRR Data                                                                        | CARB MRR Data                                                                        |
| 201  | Dairy Enteric/Manure Distributed by Population | CARB-developed shapefile                                                           | Regional Water Board General Order dairy reports (for largest) and USDA county-level population weighted by farm geographic extent (for smaller) |
| 202  | Dairy Anaerobic Digester         | UC Davis Biomass Collaborative                                                       | UC Davis Biomass Collaborative                                                      |
| 203  | Dairy Lagoon                     | CARB-developed shapefile                                                             | Regional Water Board General Order dairy reports (for largest) and USDA county-level population with farm footprint (for smaller) |
| 210  | Non-dairy Feedlot Cattle         | Parcel Data                                                                          | USDA county-level feedlot estimates distributed by area                              |
| 211  | Non-dairy Range Cattle (not complete) | Bureau of land management grazing lands and NLCD pasture | County-level range cattle distributed by area                                         |
| 212  | Horses                           | Parcel Data                                                                          | Distributed by Area                                                                 |
| 213  | Poultry                          | Parcel Data                                                                          | USDA county-level population distributed by area                                     |
| 214  | Sheep, Swine, Goats Enteric      | Parcel Data                                                                          | USDA county-level population distributed by area                                     |
| 251  | Rice                             | CropScape (USDA)                                                                     | Distributed by area                                                                 |
| 301  | Wastewater Domestic - AD         | UC Davis Biomass Collaborative                                                       | UC Davis Biomass Collaborative                                                      |
| 302  | Septic                           | Housing Density                                                                      | Distributed evenly among houses thought to have septic                              |
| 303  | Non-AD Wastewater Treatment      | State Water Board CIWQS                                                              | State Water Board CIWQS                                                             |
| 40X  | All on-road transportation surrogates | Existing CARB air quality modeling surrogates                                    | Existing CARB air quality modeling surrogates                                       |
| 990  | Wetlands                         | NLCD Land Use Data, using Method of Potter 2010                                      | Potter 2010 Emissions distributed evenly across Wetlands                              |
100 Landfills – Landfill location and facility-specific emissions were obtained from the CARB Landfill Model.

102 Compost – The spatial surrogate for commercial compost operations was provided by CalRecycle in November 2015. The location and SWIS throughput permitted maximums of compost facilities are publically available, but permitted maximums do not necessarily correspond to the relative emissions at each facility. Confidential estimates of excess capacity provided by CalRecycle was used in conjunction with permitted maximums to approximate actual throughput. This estimated throughput was assumed to be proportional to emissions.

150 Pipeline Fugitives – Pipeline natural gas fugitive emissions include transmission, distribution, and residential/commercial meters. Anything beyond the meter is not included in the CARB statewide inventory, such as leaks from appliances within a home. The distribution of fugitive CH$_4$ emissions between transmission, distribution, and residential units was based on a 2007 CARB survey (https://www.arb.ca.gov/cc/oil-gas/finalreport.pdf).

| Source        | % of Pipeline Fugitives |
|---------------|-------------------------|
| Distribution  | 69.3%                   |
| Transmission  | 10.7%                   |
| Residential Meters | 20.0%               |

To spatially distribute fugitive CH$_4$ emissions from distribution, it was first assumed that the transmission pipelines and natural gas service area data generally covered core information that can be used to estimate natural gas distribution contribution in California. South Coast and San Diego urban areas were assumed to have a complete spatial coverage of natural gas systems. Natural gas availability in other urban regions away from the main transmission pipelines were verified using housing sales data. Specifically, houses that use propane/butane are excluded from the inventory calculations. Based on the data availability, buffers were drawn around the natural gas transmission pipelines to represent distribution pipelines and housing units connected to natural gas. 89.3% of pipeline fugitives were distributed in this area according to housing units, and the remaining emissions were distributed along the transmission pipeline by length.

151 Oil and Gas – The division of Oil, Gas, and Geothermal Resources (DOGGR) shapefile was used to identify oil and gas wells in California. CH$_4$ missions were distributed evenly by well types, with assumption that all wells of the same type have equal emissions. There are multiple emission categories, and not all well types will emit in all categories. The table below defines which emission types are attributed to each well type. For emissions data, refer to specific surrogates identified in the table.
Table S8. Emission types according to well types for Oil and Gas activities.

| Well Type | Petroleum Fugitives (151a) | Process Loss Fugitives (151b) | Wastewater Fugitives (151b) | Fuel Combustion (151c) |
|-----------|-----------------------------|-----------------------------|-----------------------------|------------------------|
| N = new   | Yes                         | Yes                         | Yes                         | Yes                    |
| A = active| Yes                         | Yes                         | Yes                         | Yes                    |
| I = idle  | Yes                         | Yes                         | Yes                         | No                     |
| P = plugged| Yes                        | No                          | No                          | No                     |
| U = unknown| No                         | No                          | No                          | No                     |
| C = cancelled| No                        | No                          | No                          | No                     |
| B = Buried-Idle| No                        | No                          | No                          | No                     |
| Total Wells Included | 223,349                      | 104,226                      | 90,558                      |

152 Petroleum Refining Wastewater – Permitted wastewater treatment flow for petroleum refineries were obtained from the State Water Board California Integrated Water Quality System (CIWQS) using facility type: petroleum refinery. Locations were determined by geocoding the facility address. Missing permitted flows were filled in using refinery production obtained by matching to Energy Information Administration (EIA) GIS data.

160 Electricity Generation (in state) – Electricity generation emissions were distributed using Mandatory Reporting Regulation (MRR) data from the GHG mapping tool, using vintage 2015 mapping data.

162 Refinery and Hydrogen – MRR emissions and spatial data were used to apportion CH₄ emissions.

Dairy (general description) – Dairies emit CH₄ from enteric fermentation as well as manure management. The emission levels depend on multiple attributes of a dairy including, but not limited to, the age and type of animal and manure management practice used. Various manure management practices and animal types are differentiated in the CARB statewide GHG emission inventory requiring multiple surrogates to spatially distribute the emissions. Dairy emissions are very low in the urban Los Angeles region within the South Coast Air Basin (< 1% of total), therefore these surrogates are not heavily relied on in this analysis.

The location, population, and manure management practices on California dairies were combined from multiple data sources in order to provide the most accurate possible spatial distribution of these emissions. CARB dairy GIS layers were improved using satellite imagery to confirm dairy locations, and changes in locations over time. As part of the General Order, most dairies are required to submit Annual Reports, which include detailed dairy population data and water usage. Population data included five categories: milking head, dry cows, heifers split by two age groups, and calves split by two age groups. Populations were compiled for the largest 300 dairies, representing over 50% of the milking head in the state, then were geocoded.
to provide detailed farm-specific distribution of animal head. The cow populations on the remaining farms were calculated using the 2012 U.S. Department of Agriculture (USDA) Census county animal populations minus the population that was already matched, weighted by the dairy footprint size. Enteric emissions were based on this population data. Assumptions about manure management practices were based on information obtained from regional water boards, knowledge of the primary management practices in different regions, and satellite imagery. Some dairies will use multiple manure management methods. Error! Reference source not found. summarizes the data and methods used to distribute emissions for each manure management practice.

Table S9. Dairy Manure Management Spatial Surrogate Descriptions

| Management Practice | Description |
|---------------------|-------------|
| Anaerobic digester (202) | UC Davis Dairy California Biomass Collaborative (https://biomass.ucdavis.edu/tools/california-biomass-facilities-reporting-system/) |
| Anaerobic lagoon (203) | Population for those dairies likely to use anaerobic lagoon |
| Daily spread (201) | Population, all dairies |
| Deep pit (204) | Population for those dairies not likely to use anaerobic lagoon |
| Dry lot (201) | Population, all dairies |
| Liquid/slurry (204) | Population for those dairies not likely to use anaerobic lagoon |
| Pasture (205) | Assumed North Coast dairies only, by population |
| Solid storage (201) | Population, all dairies |

210 Non-Dairy Cattle Feedlot, 211 Non-Dairy Range Cattle, 212 Horses, and 213 Poultry – There are minimal emissions (< 1% of total) from non-dairy livestock in the urban Los Angeles region within the South Coast Air Basin. Still, multiple land use datasets were combined to spatially distribute non-dairy livestock emissions. These data have different vintages and land use categorization methodologies. The Southern California Association of Governments (SCAG) data covers 9 counties in Southern California (2008 vintage), the San Diego Association of Governments (SANDAG) data covers the San Diego region (2013 vintage), and other regions were covered, to extent possible using the Department of Water Resources Land Use Survey (various vintages from 1993 through 2015). Santa Clara and Inyo Counties were missing from these datasets, therefore parcel data was used to establish land use. The parcel maps were a part of the CARB GIS library which has been developed over time. These data were originally likely from the individual counties but the vintage is unknown. A summary of the data sources by county is shown below:
Table S10. Summary of data sources used for surrogates 210, 211, 212, and 213.

| Counties                      | Data Source                           | Vintage          |
|-------------------------------|---------------------------------------|------------------|
| Santa Clara                   | Parcel Data                           | Unknown          |
| Inyo                          | Parcel Data                           | Unknown          |
| Los Angeles, Riverside, Imperial, Orange, San Bernardino | Southern California Association of Governments (SCAG), land use survey | 2008             |
| San Diego                     | San Diego Association of Governments (SANDAG) land use data | 2013             |
| All Others                    | Department of Water Resources (DWR)    | 1993 - 2015      |

The land use categorization methodology varied by county. For most non-dairy livestock, county-level populations were available, therefore different categorization methods will not unduly affect the statewide spatial distribution. A summary of land use codes selected for non-dairy livestock analyses are summarized below (bolded are categories used):

Table S11. Land use codes for surrogates 210, 211, 212, and 213.

| Counties                      | Data Source | Categories Included               | ID   |
|-------------------------------|-------------|-----------------------------------|------|
| Santa Clara                   | Parcel Data | Agriculture Intensive             | n/a  |
| Inyo                          | Parcel Data | 540 Misc Agriculture              | 540  |
| Los Angeles, Riverside, Imperial, Orange, San Bernardino | SCAG | Irrigated Cropland and Pasture Non-Irrigated Cropland and Pasture Orchards and Vineyards Nurseries Dairy and Intensive Livestock Poultry Other Agriculture Horse Ranches | 2110 2120 2200 2300 2400 2500 2600 2700 |
| San Diego                     | SANDAG      | Field Crops Orchard or Vineyard Intensive Ag | 8003 8001 8002 |
| All Others                    | DWR         | Various Crop Categories Available Farmstead (potentially with residence) Livestock Feeding Dairy Poultry Farmstead (without residence) | n/a 1 2 3 4 5 |

There are two distinct spatial patterns of non-dairy cows that were developed for this process: cows at feedlots (surrogate 210) and range cows (surrogate 211). The GHG emission data was allocated into these two categories according to internal CARB estimate:
Table S12. Animal types used for surrogates 210 and 211.

| Emission Type | Animal Type                         | Spatial Category |
|---------------|-------------------------------------|------------------|
| Enteric       | Beef cows                           | Range            |
| Enteric       | Beef replacements 12-24 months      | Range            |
| Enteric       | Beef replacements 0-12 months       | Range            |
| Enteric       | Bulls                               | Feedlot          |
| Enteric       | Heifer feedlot                      | Feedlot          |
| Enteric       | Heifer stockers                     | Feedlot          |
| Enteric       | Steer feedlot                       | Feedlot          |
| Enteric       | Steer stockers                      | Feedlot          |
| Enteric       | Beef calves                         | Range            |
| Manure        | Feedlot - heifers 500+ lbs          | Feedlot          |
| Manure        | Feedlot - steers 500+ lbs           | Feedlot          |
| Manure        | Feedlot - heifers 500+ lbs          | Feedlot          |
| Manure        | Feedlot - steers 500+ lbs           | Feedlot          |
| Manure        | Not on feed - beef cows             | Range            |
| Manure        | Not on feed - bulls 500+ lbs        | Range            |
| Manure        | Not on feed - calves <500 lbs       | Range            |
| Manure        | Not on feed - heifers 500+ lbs      | Range            |
| Manure        | Not on feed - steers 500+ lbs       | Range            |

210 Non-Dairy Cattle Feedlot – Land use data was compiled to estimate the location of beef feedlot operations. These data have various temporal scales, accuracies, and categorization methodologies as described above. The following land use datasets were included:

Table S13. Land use codes for surrogates 210.

| Counties                              | Data Source | Categories Included                                           | ID     |
|---------------------------------------|-------------|---------------------------------------------------------------|--------|
| Santa Clara                           | Parcel Data | Agriculture Intensive                                         | n/a    |
| Inyo                                  | Parcel Data | 540 Mise Agriculture                                          | 540    |
| Los Angeles, Riverside, Imperial, Orange, San Bernardino | SCAG  | Dairy and Intensive Livestock and Associated Facilities | 2400   |
| San Diego                             | SANDAG      | Intensive Ag                                                  | 8002   |
| All Others                            | DWR         | Livestock Feeding                                             | class S and subclass 2 |

Features that intersected with known dairies (based on the CARB dairy GIS layer) were excluded. CH₄ emissions at each parcel were determined by weighting both the county-level animal populations and the surface area of the parcel. Discrepancies in the number of feedlot cows between various sources were used to update the weighting.
211 Pasture Non-Dairy Cow Spatial Surrogate – Cattle can be confined to pasture areas, and are also grazing on public lands. The Bureau of Land Management provides a shapefile of grazing allotment on public lands. This was combined with the National Land Cover Database (NLDC 2006) shapefile, which identifies the location of pasture and hay. By incorporating hay, which cannot be separated from pasture, this dataset will overestimate the area of range cattle, but was the best available pasture surrogate. Once their locations were identified, the county-level animal populations were used to distribute the CH$_4$ emissions. These CH$_4$ emissions are distributed among a large area and the exact location of the livestock at any given time is not known.

212 Horses – A spatial surrogate for horses was developed using the land use data as described above. The following land use data were included in surrogate 212:

Table S14. Land use codes for surrogates 212.

| Counties                        | Data Source | Categories Included                  | ID    |
|--------------------------------|-------------|--------------------------------------|-------|
| Santa Clara                    | Parcel Data | Agriculture Intensive                | n/a   |
| Inyo                           | Parcel Data | 540 Misc Agriculture                 | 540   |
| Los Angeles, Riverside, Imperial, Orange, San Bernardino | SCAG         | Horse Ranches                        | 2700  |
| San Diego                      | SANDAG      | Intensive Ag                         | 8002  |
| All Others                     | DWR         | Farmstead (with residence) / Farmstead (without residence) | class S and subclass 1 or 5 |

Since county-level horse populations are not available, CH$_4$ emissions (representing 1% of statewide 2012 CH$_4$ emissions) were distributed by land area of the categories above, across the whole state. A non-normalized area-weighted statewide approach over weighted contributions from Santa Clara County, as large intensive agriculture occupied large area of land and could not be differentiated from regions that contain horses. Based on further analyses, it was determined that Santa Clara was 5 times more heavily weighted than neighboring counties. Therefore, the distribution was normalized by devaluing the Santa Clara data by a factor of 5.

213 Poultry – County-level poultry populations were obtained from the USDA 2012 Census of agriculture. USDA excludes some county-level data for confidentiality reasons. The number of operations in the county is still reported without the population information. The missing poultry and operations data were calculated by comparing the USDA data to the statewide total. The statewide number of missing poultry was divided by the statewide number of missing operations. This value was multiplied by the number of missing operations in each county to provide an estimate of the total population in each county.

A spatial surrogate for poultry was developed using the land use data as described above. The following land use data were included in surrogate 213:
Table S15. Land use codes for surrogates 213.

| Counties                        | Data Source | Categories Included       | ID     |
|--------------------------------|-------------|---------------------------|--------|
| Santa Clara                     | Parcel Data | Agriculture Intensive     | n/a    |
| Inyo                            | Parcel Data | 540 Misc Agriculture      | 540    |
| Los Angeles, Riverside, Imperial, Orange, San Bernardino | SCAG       | Poultry operations        | 2500   |
| San Diego                       | SANDAG      | Intensive Ag              | 8002   |
| All Others                      | DWR         | Poultry                   | class S and subclass 4 |

Approximately 5% of the poultry population reported in the USDA 2012 survey did not have corresponding land use information. These counties were excluded from the spatial distribution as they represent only 5% of the GHG emissions, and no land use data is available to distribute those emissions at the sub-county-level. In total, data from 32 counties were used to distribute the poultry CH$_4$ emissions.

**251 Rice** – 2012 CropScape data was used to determine the total land area used for rice. This data was cross-compared to USDA NASS Quickstats to verify the counties that had rice activities. Those counties that were not in the USDA NASS Quickstats were removed from further analysis. After the removal, the total rice area according to CropScape was 2,360,525,026.9 sq. meters or approximately 583,298.4 acres. USDA reports 561,968 ± 22,000 acres harvested in 2012, indicating reasonable agreement. This county-level rice area was used to distribute the Statewide CH$_4$ emissions from rice.

**301 Wastewater, Domestic – Anaerobic Digesters** – Data for wastewater with anaerobic digesters was acquired from UC Davis ([http://biomass.ucdavis.edu/tools/california-biomass-facilities-reporting-system/](http://biomass.ucdavis.edu/tools/california-biomass-facilities-reporting-system/)). Average dry weather flow was used as the activity metric to distribute the CH$_4$ emissions. Best estimate of flow based on the population of the region and the surrounding facility capacity was used to fill in missing facility data.

**302 Septic** – The septic tank CH$_4$ emissions were distributed by housing units, using statistics differentiated by urban and non-urban areas and housing density. The number of homes in California with septic tank was consistently 10%, and approximately 10% of all new homes are expected to have septic tanks according to the Water Board. Using this information and data form 2010 Census, approximately 1,368,000 housing units are expected to have septic tanks. 2013 American Housing Survey data was used to determine that approximately 1% of urban homes and 65% of homes outside of urban areas used septic tanks in the Pacific region (including CA, OR, and WA). Housing density was incorporated into the analysis to estimate the fraction of total homes with septic tanks before distributing the Statewide CH$_4$ emissions:
Table S16. Land use codes for surrogates 213.

| Category                  | Urban/Rural | Housing Density (per sq. mile) | Number of Housing Units | % of Statewide Housing Units | % of housing units assumed to have septic |
|---------------------------|-------------|--------------------------------|-------------------------|-----------------------------|----------------------------------------|
| Very Rural                | Rural       | <1,000                         | 569,439                 | 4.2%                        | 100%                                   |
| Rural                     | Rural       | >=1,000                        | 64,131                  | 0.7%                        | 65%                                    |
| Low Density Urban         | Urban       | <500                           | 450,348                 | 5.1%                        | 65%                                    |
| Medium Density Urban      | Urban       | 500 – 1,000                    | 248,052                 | 5.2%                        | 35%                                    |
| High Density Urban        | Urban       | >=1,000                        | 11,610                  | 85%                         | 0.1%                                   |
| Total                     |             |                                | 1,343,582               | 9.8%                        |                                        |

303 Non-AD Wastewater Treatment (including wastewater recycling plants) – The wastewater treatment plants (WWTP) and wastewater recycling plants (WRP) with active permits were identified using CIWQS under “waste water treatment facilities” with permit type: “active”, which includes all regulation categories. Total of 1165 facilities were identified out of the 1216 on record. Facilities with missing geospatial information and significant design flow were tracked and confirmed using satellite maps. Two facilities were not identified due to lack of information. WWTP and WRP with missing design flow were excluded since there were no accurate methods to estimate the throughput. The CH$_4$ emissions will depend on the specific management practices employed at the facility, as well as the total waste treated. Data on actual volume treated or specific management details were not accessible for the purpose of regionalization. Therefore, the permitted design flows were used as a surrogate to apportion Statewide CH$_4$ emissions. The actual CH$_4$ emissions may deviate significantly from this estimate since the design flow do not correspond directly to daily throughput and CH$_4$ emissions.

40X On-Road Transportation – Four on-road transportation spatial surrogates were used to distribute emissions:

- 401 Heavy-Duty Diesel
- 402 Light-Duty and Passenger Diesel
- 403 Heavy-Duty Gasoline and Ethanol
- 404 Light-Duty and Passenger Gasoline and Ethanol

Gridded EMFAC2014 maps were developed using the default database setting for a typical summer weekday in 2012. The resulting NO$_x$ emissions were normalized to produce a statewide spatial surrogate. These gridded NO$_x$ emissions were used as a proxy for CH$_4$ emissions.
980 **Seeps** – Seeps are in the excluded category of the GHG inventory, and are based on incomplete information. Statewide CH$_4$ emissions from seeps are not well known, however, emissions from some specific seeps are available in the literature. Continued work to understand emissions from seep is warranted. La Brea Tar Pits emissions in the excluded GHG emission inventory was estimated to be 1.5 MMT CO$_2$e/year and Coal Oil Point was estimated to be 0.657 MMT CO$_2$e/year$^3$.

990 **Wetlands** – Wetlands are a natural source of CH$_4$ that is not included in the CARB anthropogenic GHG inventory. Ambient measurements include both natural and anthropogenic sources, therefore this source was estimated and added to the analysis to maintain consistent scope between estimation methods. The magnitude and location of CH$_4$ emissions from wetlands were estimated using the method described in Potter et al. (2006, 2010). NLCD 2011 wetland data were extracted for two types: woody wetlands and herbaceous wetlands. The total CH$_4$ from wetlands were estimated to be 0.467 MMT CO$_2$e/year using AR4 GWP and the CH$_4$ emissions were distributed evenly according to wetland area per grid cell.

**Assessing the robustness of linear regression models using repetitive randomized sub-sampling technique**

Sub-sampling technique utilized to determine the robustness of the linear regression model assumes that sub-sampling approximately 50% of $n$ can generally describe the variability of the resulting data such that $n_s$ is $\text{int}[0.5n]+1$ when the dataset size, $N$, is equal to the analysis population of $n$, and when the new sub-sampled population can act as an estimator for the original sample population at a specific efficiency$^7$. The number of sub-sampling events, $r^*$, used to describe the efficiency or the representativeness of the sub-samples must be determined using the desired efficiency, $E_f(S_g)$, $n$, and $n_s$.

$$r^* = \text{int} \left[ \frac{\log(1 - E_f(S_g))}{\log(n - n_s) - \log(n)} \right] + 1$$

(Eq. 1)

The total number of sub-samples required to understand the variability, $k$, however, must accommodate for the outliers, $m$, that exist in the dataset. It can be assumed that $m = 0$ as a result of the data constraining using time and meteorological parameters. Therefore, the probability that the random sub-sample is representative of a non-biased sample set, $p_{g^*}$, is 1. The probability of $i$ representative sub-samples in $k$ sub-sampling events is described by:

$$p(k, i) = 1 - \sum_{j=0}^{i-1} \binom{k}{j} p_{g^*}^j (1 - p_{g^*})^{k-j}$$

(Eq. 2)
To determine $k$, the above equation must be solved for the minimum value in which the probability is satisfied for $E_i(S_g) = 99\%$ efficiency.

$$k = \arg \min \{ p(l, r^*) \geq 0.99 \}$$  \hspace{1cm} (Eq. 3)

However, the effective probability, $p(k,i)$, when $m = 0$ is $1$ (since $p_g = 1$). Therefore, $k$ is equivalent to the total number of randomized-sub-sampling events required to capture the variability of the linear regression model without outliers, $r^*$, which is described by the following equation:

$$r^* = \text{int} \left[ \frac{\log(1 - 0.99)}{\log(n - n_*) - \log(n)} \right] + 1$$  \hspace{1cm} (Eq. 4)

Sub-sampling, therefore was performed generally $k = 7$ times to understand the variability of the linear regression slopes obtained for each month.

Below are the monthly aggregated data that passed this and other quality control checks:

Table S17. QC results for MB1; CH$_4$:CO correlation.

| MB1 | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2012 | | | | | | | | | | | | |
| 2013 | | | | | | | | | | | | |
| 2014 | | | | | | | | | | | | |
| 2015 | | | | | | | | | | | | |
| 2016 | | | | | | | | | | | | |

Note: Green = pass, yellow = not pass, gray = no data.

Table S18. QC results for MB2; CH$_4$:CO correlation.

| MB2 | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2014 | | | | | | | | | | | | |
| 2015 | | | | | | | | | | | | |
| 2016 | | | | | | | | | | | | |

Note: Green = pass, yellow = not pass, gray = no data.
Reference

1. Hopkins, F. M.; Kort, E. A.; Bush, S. E.; Ehleringer, J. R.; Lai, C. T.; Blake, D. R.; Randerson, J. T. Spatial patterns and source attribution of urban methane in the Los Angeles Basin. *J. Geophys. Res.: Atmos.* **2016**, 121, (5), 2490-2507. DOI: 10.1002/2015JD024429.

2. Aliso Canyon Natural Gas Leak. *California Air Resources Board 2016*. https://www.arb.ca.gov/research/aliso_canyon_natural_gas_leak.htm (Accessed: October 21, 2016).

3. CARB. California Greenhouse Gas Emission Inventory (2017). *California Air Resources Board 2017*. http://www.arb.ca.gov/cc/inventory/data/data.htm (Accessed: July 2017).

4. CARB. EMFAC Web Database. 2014.

5. Census. Population Estimates. United States Census Bureau: 2014.

6. DLRP. Farmland Mapping and Monitoring Program. California Department of Conservation, Division of Land Resource Protection: 2012.

7. Tsao, M.; Ling, X. Subsampling method for robust estimation of regression models. *Open Journal of Statistics* **2012**, 2, (03), 281. DOI: 10.4236/ojs.2012.23034.