Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

Vista-LA: Mapping methane emitting infrastructure in the Los Angeles megacity

Valerie Carranza[1,2], Talha Rafiq[1,3], Isis Frausto-Vicencio[1,4,*], Francesca Hopkins[1,*], Kristal R. Verhulst[1,3], Preeti Rao[1,**], Riley M. Duren[1], Charles E. Miller[1]

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109, U.S.A.

[2] Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, 90024, U.S.A.

[3] Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, 90024, U.S.A.

[4] Department of Chemistry, University of California, Los Angeles, Los Angeles, CA 90024, U.S.A.

* Now at: University of California, Riverside, Riverside, CA, 92521 U.S.A.

**Now at: University of Michigan, Ann Arbor, MI 48109 U.S.A.

Correspondence to: Kristal Verhulst (Kristal.R.Verhulst@jpl.nasa.gov)

Abstract

Methane is a potent greenhouse gas (GHG) and a critical target of climate mitigation efforts. However, actionable emission reduction efforts are complicated by large uncertainties in the methane budget at relevant scales. Here, we present Vista, a Geographic Information System (GIS)-based approach to map potential methane emissions sources in greater Los Angeles, an area with a dense, complex mixture of sources. The goal of this work is to provide a database that, together with atmospheric observations, improves methane emissions estimates in urban areas with complex infrastructure. We aggregated methane source location information into three sectors (energy, agriculture, and waste) following the frameworks used by the State of California GHG Inventory and the IPCC Guidelines for GHG Reporting. Geospatial modelling was applied to publicly available datasets to precisely geolocate facilities and infrastructure comprising major anthropogenic methane source sectors. The final database, Vista-Los Angeles (LA), is presented as maps of infrastructure known or expected to emit methane. Vista-LA

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

contains over 33,000 features concentrated on <1% of land area in the region. Currently, Vista-LA is used as a planning and analysis tool for atmospheric measurement surveys of methane sources, particularly for airborne remote sensing, and methane "hot-spot" detection using regional observations. This study represents a first step towards developing an accurate, spatially-resolved methane flux estimate for point sources in California's South Coast Air Basin (SoCAB), with the potential to address discrepancies between bottom-up and top-down methane emissions accounting. The final Vista-LA datasets and associated metadata have been submitted to the Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics (ORNL DAAC; https://doi.org/10.3334/ORNLDAAC/1525).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

1 Introduction

Methane (CH_4) is the second most important anthropogenic driver of climate change (Myhre et al., 2013). Recent studies have shown that mitigating CH_4 emissions yields large near-term climate benefits due to CH_4's relatively short atmospheric lifetime (Dlugokencky et al., 2011). Reducing CH_4 emissions is complicated by the incomplete understanding of the CH_4 budget at scales relevant to actionable emissions reduction efforts. Cities are important for GHG mitigation, since they represent high-density emissions regions with the appropriate scale to reduce GHG emissions (Duren and Miller, 2012; Kennedy et al., 2009). Additionally, political will and commitment is needed to implement mitigation efforts for reducing GHG emissions (Gurney et al., 2015). However, enacting emission controls is challenging in urban areas that are highly complex and heterogeneous, with various emission sources located in close proximity.

Understanding urban emissions requires knowledge of source sectors and their respective activities at scales that align with urban policy and planning (typically 10's to 100's of meters). Such information has been assembled for fossil fuel carbon dioxide (CO_2) emissions using the Hestia approach, which quantifies urban sources down to the building level (Gurney et al., 2012). To date, Hestia has generated detailed estimates of urban CO_2 emissions for several cities, including Los Angeles (LA) (Rao et al., *in review*), Indianapolis (Gurney et al., 2012), and Salt Lake City (Patarasuk et al., 2016). A CH_4 emissions product with spatial information equivalent to the scale of Hestia is needed for CH_4 emissions mitigation efforts. However, the sources of CH_4 differ significantly from those of CO_2, which are primarily driven by fossil fuel combustion. Therefore, the methods used to develop Hestia are not directly transferable to CH_4, which has distinct source processes and spatial patterns from CO_2.

Urban areas are globally significant sources of CH_4 emissions, primarily coming from energy use and waste management (Hopkins et al., 2016a; Marcotullio et al., 2013). However, knowledge of the location and relative contribution of these emission sources is highly uncertain, especially in urban areas where energy, waste treatment, and other CH_4 emission sources are located in close proximity to one another. Global emissions inventories based on nightlights and/or population scaling methods (e.g., EDGAR v4.2 European Commission Joint Research Centre, 2010; Olivier and Peters, 2005) are limited in their usefulness for estimating emissions at the scale of a city or air basin. Official CH_4 emission inventories

made using bottom-up approaches (e.g., IPCC, 2006) underestimate CH_4 emissions and are driven by a different mixture of sources compared to those inferred from atmospheric measurements, as observed in Los Angeles (Hopkins et al., 2016b; Hsu et al., 2009; Townsend-Small et al., 2012; Wennberg et al., 2012; Wong et al., 2016, 2015; Wunch et al., 2009), Boston (McKain et al., 2015), Indianapolis (Cambaliza et al., 2015), Florence (Gioli et al., 2012), London (Helfter et al., 2016), and San Francisco (Jeong et al., 2017)). Consequently, there is a need for a new approach of urban CH_4 assessment that overcomes these shortcomings by incorporating both top-down (observation-based) and bottom-up (activity-based) information.

Atmospheric CH_4 in the urban landscape is dominated by CH_4 hotspots that primarily come from fossil fuel-derived sources (e.g., Hopkins et al., 2016b). Many of these hotspots are associated with leaks—fugitive CH_4 emissions— in natural gas systems (e.g., Jackson et al., 2014; Phillips et al., 2013). Across the natural gas infrastructure, CH_4 emissions are disproportionately emitted by a small fraction of "super-emitters" (Brandt et al., 2014). Fugitive CH_4 emissions sources are more challenging to inventory than activity-related emissions, and contribute to uncertainty in the magnitude and spatial pattern of CH_4 emissions in urban areas (Hopkins et al., 2016b; Lamb et al., 2015). In recent studies, the locations of fugitive emissions have been identified using observational data, such as mobile surveys, airborne campaigns, and sustained monitoring (e.g., Cambaliza et al., 2015; Frankenberg et al., 2016; Hopkins et al., 2016b; Verhulst et al., 2017).

Inaccuracies and coarse information in city-scale inventories of CH_4 pose a direct obstacle to city mitigation plans. Shortcomings in bottom-up methods have been identified to be: inaccurate representation of fugitive CH_4 sources at fine spatial scales, the existence of unreported CH_4 sources in urban areas, and/or incomplete accounting of known CH_4 sources, such as from oil and gas activities (Hopkins et al., 2016a; Lyon et al., 2015; Zavala-Araiza et al., 2015). CH_4 emissions estimates for urban regions can be improved by more complete accounting of potential CH_4 sources at the facility scale, along with targeted observations that can detect fugitive emissions and super-emitter behavior.

89 Here, we present Vista, a GIS-based CH_4 emissions mapping database designed to address shortcomings
90 in current urban CH_4 inventories. Vista encompasses key CH_4 emissions categories from the
91 Intergovernmental Panel on Climate Change (IPCC) GHG Inventory methodology. The primary goal of
92 this research effort is to improve understanding of CH_4 emissions at urban scales with complex mixtures
93 of sources, exemplified by the LA Megacity. Emissions monitoring and verification efforts in LA are
94 highly relevant for California's statewide emissions control efforts. The LA Megacity emits a significant
95 fraction of California's GHG emissions, with 42% of the state's population concentrated in 4% of the
96 state's land area (CARB, 2014b).

97 In this study, we present the Vista-LA database for the spatial domain of California's South Coast Air
98 Basin (SoCAB), the air basin that contains the majority of LA Megacity GHG emissions. Vista-LA
99 consists of detailed spatial maps for facilities and infrastructure in the SoCAB that are known or expected
100 sources of CH_4 emissions. Vista-LA illustrates the spatial distribution of potential CH_4 sources,
101 representing a first step towards developing an urban-scale CH_4 emissions gridded inventory for the
102 SoCAB. The final Vista-LA database contains over 33,000 entries, which are presented as CH_4 emitting
103 infrastructure maps. SoCAB is an ideal testbed due to the density of sources and availability of
104 observations from the LA Megacity Carbon Project (https://megacities.jpl.nasa.gov/portal/) tower
105 network (Newman et al., 2016; Verhulst et al., 2017), the California Laboratory for Atmospheric Remote
106 Sensing (CLARS) (Wong et al., 2016, 2015), and a total column carbon observing network site (Wunch
107 et al., 2009). The Vista data product is a key tool for CH_4 emissions research and mitigation efforts; by
108 (1) mapping areas of CH_4 emitting infrastructure, (2) identifying targets for CH_4 surveys, and (3) enabling
109 interpretation of atmospheric observations, including source attribution, and comparison of measured
110 emissions to permitted or reported emissions. Combined with atmospheric observations, Vista enables
111 systematic study of urban CH_4 emission sources.

112

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

2 Methods Overview

2.1 Vista-LA Structure and Organization

The spatial domain for the Vista-LA database is SoCAB, the air-shed for the greater Los Angeles urban extent, including the urbanized parts of Los Angeles, Orange, Riverside, and San Bernardino Counties. Vista-LA follows the IPCC CH_4 emissions reporting framework (IPCC, 2006). Following IPCC methodology provides compatibility with the State of California CH_4 emissions inventory and allows the approach used in this study to be adapted to other regions globally (CARB, 2014a, 2015). For example, Vista-LA can be easily adapted to the Environmental Protection Agency (EPA) national inventory structure since it also follows the 2006 IPCC Guidelines for National GHG Inventories as shown in Figure A1 (EPA, 2016).

The Vista-LA structure enables sectoral tracking of emissions. We used the State of California GHG Inventory for 2015 (CARB, 2016), the most policy-relevant inventory that includes the SoCAB domain, to rank the top CH_4 emitting sources (Figure 1). According to the State of California GHG Inventory, ~99% of California's CH_4 emissions are expected to result from just three IPCC Level 1 categories – energy, agriculture, and waste, and eight IPCC Level 3 categories – fuel combustion activities related to energy industries and transport (IPCC − 1A1 & 1A3), fugitive emissions related to oil and natural gas (IPCC − 1B2), livestock emissions related to enteric fermentation (IPCC − 3A1) and manure management (IPCC − 3A2); and solid and liquid waste disposal and treatment, including managed waste disposal sites (4A1), and domestic and industrial wastewater treatment and discharge facilities (IPCC − 4D1 & 4D2, respectively) (see Figure 1). The other 27 Level 3 categories cumulatively contribute less than 1% of California's CH_4 emissions, and are hence assumed to have negligible impact on SoCAB CH_4 emissions. This approach greatly simplifies the database, allowing us to focus our attention on the top-emitting sources. By design, Vista-LA only includes Scope 1 emissions—that is, direct GHG emissions from sources that are owned or controlled by a company within the study domain, as defined by the GHG Protocol (http://www.ghgprotocol.org/corporate-standard). Therefore, sources that are not expected to result in significant direct emissions of CH_4 in the SoCAB were excluded, such as emissions from imported electricity, geothermal energy production, and from solid fuels such as coal.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

Vista-LA also includes two additional sources that are not explicitly accounted for in the California GHG Inventory but are potentially significant sources of fugitive CH_4 emissions in SoCAB: compressed natural gas (CNG) fueling stations and liquefied natural gas (LNG) fueling stations, which were categorized under IPCC Level 2 – 1B2. This case study of Vista-LA focuses on anthropogenic sources of CH_4, and every effort has been taken to make the Vista-LA dataset as complete, accurate, and timely as possible. Because Vista-LA is designed to incorporate solely anthropogenic sources of CH_4, natural CH_4 sources such as wetlands and geologic seeps are excluded this version. This is consistent with the most recent version of the California GHG Inventory, which categorizes petroleum gas seeps separately as "excluded" emissions, though they were previously categorized under IPCC – 1B2.

2.2 Overview of Data Sources

Within each of the three major CH_4 source sectors (IPCC 1 - Energy, IPCC 3 - Agriculture, Forestry, and other Land Use; and IPCC 4 - Waste), we defined the types of infrastructure associated with emissions. We sought out publicly available datasets that mapped their spatial locations (Table 1). Spatial datasets were compiled from reliable and verified public databases on government and federal/state research agency portals. The data are presented as shapefiles and kmz files that include point, line, and polygon data. Table 1 summarizes the spatial datasets by the year, source of data, data type (points, lines, or polygon data) and also indicates the corresponding IPCC Level 3 CH_4 emissions category. Sections 3-5 describe the data sources and information in Vista-LA in further detail and also describe the specific data processing techniques applied to the GIS dataset for each of the Level 3 emissions category.

Some of the spatial datasets we obtained (see e.g., Southern California Association of Governments, http://gisdata.scag.ca.gov/Pages/GIS-Library.aspx) and EPA Facility Registry Service (FRS) were useful for evaluating information from more than one type of CH_4-producing infrastructure (e.g., petroleum refineries and wastewater treatment plants). Due to the variety of data sources used to create Vista-LA, the same level of detail (e.g., spatial resolution, data completeness, available metadata) was not always available for every CH_4 emitting source. The level of completeness or detail for each spatial dataset will be discussed below under the data sources and limitations sections. We processed and standardized GIS

datasets through geo-referencing, spatial configuration, and verification using ArcGIS software packages. All spatial features and raster layers were geo-located using the World Geodetic System 1984 datum and the Universal Transverse Mercator Zone 11 North coordinate system. Considerations for privacy including restrictions and limitations on some of these datasets were taken into account for the final product. Consequently, Vista-LA datasets for natural gas compressor stations and natural gas pipelines are only included as static representations in Figures 2 and 3. Vista-LA does not include sub-facility level information. Timely data are critical for understanding methane dynamics in SoCAB, therefore we used the most current publicly available information in the development of the Vista-LA database.

3 Energy (IPCC Level 1 – Category 1)

The Energy (IPCC Level 1 − 1) sector includes CH_4 emitted by fuel combustion activities (IPCC Level 2 − 1A) and fugitive emissions from fuels (IPCC Level 2 − 1B). CH_4 emissions from fuel combustion are mainly produced by energy industries and transportation, with minor contributions from manufacturing, commercial, industrial, residential and agricultural fuel combustion (CARB, 2016). Fugitive emissions are defined by the IPCC as an intentional or unintentional release of gas from anthropogenic activities not including combustion (IPCC, 2001). Fugitive CH_4 emissions come from leaks or failures of equipment, off-gassing, or venting, arising from sources such as natural gas storage facilities, oil and gas wells, and pipelines. They occur mainly in the oil and gas sector (~95% of California's estimated fugitive CH_4 according to CARB, 2016), with a small contribution from industrial and manufacturing sources. Many facilities, including petroleum refineries and power plants, include both combustion and fugitive CH_4 emissions.

3.1 Fuel Combustion Activities (IPCC Level 2 - 1A)

Fuel combustion activities (IPCC Category 1A) includes CH_4 emissions from energy industries, which encompass petroleum refining and electricity generation via combustion of natural gas in power plants, and transportation. Other combustion sources have only a small expected CH_4 emission rate (totaling <0.1% of statewide CH_4), according to the California GHG Inventory, hence are not included in Vista-LA (CARB, 2016). The physical infrastructure associated with combustion and fugitive CH_4 emissions

192 from energy industries in SoCAB are natural gas-fired power plants and petroleum refineries.
193 Transportation comprises ~1.1% of inventoried statewide CH_4, primarily from on-road sources (e.g.,
194 conventionally fueled cars, light- and heavy-duty trucks), but is not included in this version of Vista-LA
195 (CARB, 2016).

196 **3.1.1 Energy Industries (IPCC Level 3 - 1A1)**

197 **3.1.1.1 Petroleum Refineries (Vista-LA layer)**

198 *Data sources:*

199 The Vista-LA petroleum refinery dataset provides location and extent data for 12 facilities in the domain.
200 The primary spatial datasets for petroleum refineries (IPCC – 1A1) were gathered from the U.S. Energy
201 Information Administration (EIA) for the year 2016. EIA reports information about all operable
202 petroleum refineries and electricity generation plants in the United States, including plants that are active,
203 on standby, and those short-term or long-term out of service (EIA, 2016). Additional information came
204 from Southern California Association of Governments (SCAG) land use data for the years 2005 and 2012
205 (see http://gisdata.scag.ca.gov/Pages/GIS-Library.aspx).

206 *Data processing and validation:*

207 Petroleum refinery locations were verified using multiple datasets, including EIA, SCAG, and the ESRI
208 Basemap aerial imagery, and Google Earth imagery. EIA was the primary source of information, as it
209 contains the most recent data. SCAG was used to verify that there were no missing petroleum refineries
210 from EIA. This process provided data quality assurance from the most updated publicly available spatial
211 database of petroleum refineries.

212 The original EIA data series on petroleum refineries includes geolocations as points, and information on
213 production capacity, current and projected capacity of crude oil separated by atmospheric distillation,
214 downstream charge, as well as fuel, electricity, and steam purchased and consumed by 141 refineries

across the United States (U.S. Energy Information Administration, 2015). This dataset contains information on nine refineries located in SoCAB, all of which are located in Los Angeles County.

To map the area of petroleum refinery and power plant facilities, we added data from SCAG for the years 2005 and 2012, which maps land use areas to a minimum two acre resolution (see http://gisdata.scag.ca.gov/Pages/GIS-Library.aspx). The SCAG database only contains land use classifications for the State of California, and lacks facility-level information. We performed feature identification using SCAG land use code 1322 "Petroleum Refining and Processing". This category includes major oil refineries, as well as associated petrochemical plants. This data was used to identify, extract, and define the spatial extent of each refinery and match the geolocations of the refineries listed in the EIA 2016 dataset.

The SCAG land use code was used to identify and extract 30 polygons in the SCAG 2005 dataset and 60 polygons in the SCAG 2012 dataset related to petroleum refineries in SoCAB. Because SCAG polygon features were fragmented and not assigned to an individual refinery, they had to be manually merged based on their geolocation and spatial relation to the EIA 2016 dataset. The polygon features dataset categorized as "Petroleum Refining and Processing" were merged together and then compared to the nine refineries identified in the raw EIA dataset. In some cases, these SCAG polygon features were geographically misplaced in residential locations or in the middle of streets and had to be manually adjusted to fit the actual extent of that facility. There were three facilities identified in the SCAG dataset that were not identified in the EIA dataset. The existence and operation of these three facilities identified in SCAG were further verified using refinery planning documents and environmental assessment reports and then were appended to the EIA dataset. The true spatial extent of all polygons was verified using aerial imagery. During validation of refinery spatial extents with Google Earth Imagery and Esri Basemap aerial imagery, focus was given on identifying acres of storage tanks situated in a matrix formation, large intake pipes, storage vats, and large industrial infrastructure.

Limitations:

The additional refineries identified in SCAG and validated through the Vista verification procedures do not contain facility level metrics that were provided in the EIA dataset. Obtaining detailed sub-facility-level information for each petroleum refinery will be crucial to developing accurate CH_4 emission factor estimates.

Results:

The final Vista-LA petroleum refinery dataset includes all 12 petroleum refineries operated by eight different companies within SoCAB. The final dataset includes operational data from EIA, recalculated locational data, and validation notes including any changes made and date of last update.

3.1.1.2 Power Plants (Vista-LA layer)

Data sources:

The Vista-LA layer for power plants (IPCC – 1A1) relies on data from EIA, SCAG 2005, SCAG 2012, Google Earth and Esri Basemap aerial imagery (EIA, 2016; see http://gisdata.scag.ca.gov/Pages/GIS-Library.aspx). The Vista-LA power plant dataset provides accurate location and extent data as well as facility level information on the type of power generation methods and energy production statistics.

Data processing, validation and limitations:

The EIA 2016 contains records for 7,995 power plants in the United States, including 385 power plants in SoCAB. For our analysis, we selected only the power plants that used the following primary fuels: biomass, natural gas, petroleum, or other—which matches the methods of the California GHG inventory (CARB, 2016). This excluded power plants with primary fuel categories such as wind, solar, hydroelectric, or pumped storage. After filtering by primary fuel type, the new dataset contained 110 power plants in SoCAB.

Polygon features for each of the 110 power plants were created based on Google Earth Imagery, Esri Basemap Aerial Imagery, SCAG 2005 and SCAG 2012 land use datasets. The SCAG land use code 1431

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

("Electrical Power Facilities"), was used to verify and determine the spatial extent of the EIA power plants. SCAG polygons were geolocated with the point data from the EIA dataset. In total, there were 1,490 individual polygon features related to land use code 1431 in SCAG 2005 and 6,932 in SCAG 2012. In addition to power plants, the SCAG land use code 1431 also includes land used for distribution of electricity and substations with power plants, hence visual inspection using high-resolution aerial imagery was required to validate each individual power plant and to generate accurate polygon representations.

When visually inspecting individual power plants, we looked for typical power plant infrastructure features such as smoke or steam stacks with towers, racks, piping, and vents, transformers and/or electrical equipment. Some power plant locations were more difficult to validate. In some cases, the power plant point data was placed on the street near the operating utility and sometimes it did not match the address that was listed in the metadata. Sometimes the point was located on the center of a site, which could be within another facility (e.g. a refinery) and thus had to be manually adjusted with appropriate understanding of the context of its location using Google Earth and Esri Basemap aerial imagery. Polygons were created using GIS methods including geoprocessing and digitizing with Google Earth and Esri Basemap aerial imagery as reference. Power plant latitude and longitude coordinates were recalculated appropriately for each power plant. Power plants whose geolocations were verified but their spatial extents could not be determined using this method were tagged with a circular placeholder and their EIA facility level metrics were maintained and marked in the metadata.

We used the 2014 Fossil Fuel Data Assimilation System (FFDAS) point dataset to validate our results (Asefi-Najafabady et al., 2014). The 105 power plant point locations in the 2014 FFDAS dataset match with 105/110 power plants in the final Vista-LA layer. The FFDAS dataset includes only those facilities registered through CAMD and EIA reporting, which explains the difference in the number of locations between the two datasets. One of the five plants is a landfill gas plant, so it is not tracked in FFDAS because it is not a fossil-based source of CO_2 emissions.

We also considered using the 2010 Open-source Data Inventory for Anthropogenic CO_2 (ODIAC) for validation of power plants (Oda and Maksyutov, 2011). Cross-validation with ODIAC was not straightforward because the online data product is gridded and is at lower resolution than the EIA and

290 SCAG datasets. The publicly available version of ODIAC also had significant latency compared to the
291 EIA and SCAG datasets used in this study.

292 *Results:*

293 The Vista-LA power plant dataset merged polygon extent data with the EIA metadata. The final dataset
294 includes the facility level statistics from EIA and along with data validation information using Google
295 Earth, SCAG 2005, and SCAG 2012 in the metadata for all 110 power plants originally identified in the
296 EIA dataset. Based on the 2016 EIA electrical output data, there are only 17 power plants with greater
297 than 100 Megawatts/hour of electrical output in SoCAB. For this reason, we include the production
298 metrics in the Vista-LA database, as they may be useful for generating CH_4 emission estimates in the
299 future. The largest producing power plants in SoCAB might be expected to have significant emissions of
300 CH_4 compared to smaller power plants.

301 **3.2 Fugitive Emissions from Fuels (IPCC Level 2 - 1B)**

302 Fugitive emissions from fuels (IPCC Category 1B) include CH_4 emissions from the lifecycle (production,
303 processing, storage, transportation) of oil, natural gas, solid fuels, and geothermal energy production
304 occurring in SoCAB. We omit the latter two sources from consideration since California air quality
305 restrictions do not permit coal-burning (Perata, 2006), and there are no active coal mining or geothermal
306 energy sites in the SoCAB. In the California GHG inventory, fugitive emissions are primarily from oil
307 and gas extraction (30%) and natural gas pipelines (65%) (CARB, 2016). Vista-LA includes spatial
308 information for oil & gas wells, natural gas pipelines, natural gas storage fields, natural gas processing
309 plants, and natural gas compressor stations. Petroleum refineries emit fugitive CH_4 (IPCC 1B2), but
310 because of the spatial overlap with refinery combustion emissions at the facility level, we do not treat
311 them separately in Vista-LA (see refinery layer in IPCC Category 1A1). We also include two more
312 potentially significant sources of fugitive CH_4 emissions in SoCAB that have no assignment in the
313 California GHG inventory or IPCC categories: compressed natural gas (CNG) fueling stations and
314 liquefied natural gas (LNG) fueling stations. Vista-LA does not yet include the locations of petroleum
315 storage tanks due to a lack of publicly available information for these elements. Data for energy-related

316 sources is also available for purchase from consulting companies; however, one of the objectives of this
317 work is to generate a product that is openly accessible to the public. Therefore, we did not utilize
318 proprietary or "for-purchase" information in the development of the Vista-LA database.

319 **3.2.1 Oil and Natural Gas (IPCC Level 3 - 1B2)**

320 **3.2.1.1 Compressed Natural Gas (CNG) Fueling Stations (Vista-LA layer)**

321 *Data sources:*

322 Geospatial data of active compressed natural gas (CNG) fueling stations was obtained from the U.S.
323 Department of Energy's (DOE) Alternative Fuels Data Center (AFDC) for the year 2017 (DOE, 2017).
324 Currently, CNG fueling stations are not included in a separate IPCC category, so for the purposes of this
325 study we have classified these data under IPCC Level 3 – 1B2)

326 *Data processing, validation and limitations:*

327 The raw file was downloaded from DOE/AFDC through a series of queries for compressed natural gas
328 stations. Next, the CNG dataset was geocoded using latitude/longitude coordinates. Coordinates for this
329 dataset generated points for 1,792 stations across the United States including 336 in the state of California.
330 163 of the 336 compressed natural gas stations were in SoCAB, further reduced to 109 after removing
331 duplicate entries. The geolocations of these 109 CNG fueling stations were verified by comparing the
332 reported street address to Google aerial imagery (e.g. Google Earth, Google Maps, and Google Earth
333 Street View) and Esri Basemaps. Out of the 109 points, 88 polygon extent features were created based on
334 aerial imagery. For the remaining 21 CNG stations, placeholder polygons were created for stations whose
335 natural gas infrastructure could not be visually identified using aerial imagery, but whose location was
336 otherwise verified. During visual validation of the CNG stations, we focused on identifying pumps, gas
337 station infrastructure, and piping/compressed gas storage cylinders near parking lots and salvage yards.

338 A major portion of the DOE/AFDC dataset placed the locations of CNG fueling stations adjacent to the
339 fueling station, making it challenging to discern the exact location of station infrastructure on the map.

This dataset counts the entire station as one polygon, despite multiple fuel dispensers. Sub-facility-level information about individual fueling dispensers is not currently identified in Vista-LA.

Results:

The final Vista CNG station layer contains geolocations for 109 polygons. The metadata contains information about the station name, pressures (units: pounds per square inch or psi), types of dispensing capability, maximum vehicle size accommodation, and the recalculated latitude and longitude coordinates, along with validation notes including any changes made and date of last update.

3.2.1.2 Liquefied Natural Gas (LNG) Fueling Stations (Vista-LA layer)

Data sources:

Similar to CNG stations, fugitive emissions from LNG fueling stations were not inventoried in the 2015 California GHG Inventory. Thus, we assigned LNG stations under IPCC – 1B2. Geospatial data of active LNG fueling stations was obtained from the U.S. Department of Energy's (DOE) Alternative Fuels Data Center for the year 2017 (DOE, 2017).

Data processing, validation and limitations:

DOE data was originally geocoded using the coordinates listed. The raw DOE dataset contained 187 records of LNG stations across the U.S., 47 of those stations were located in the state of California with 27 currently open and operational in SoCAB. 15 of the 27 LNG stations shared the same location with CNG stations. The geolocations of the 12 LNG-only stations were verified by comparing the reported street address to Google aerial imagery and Esri Basemap aerial imagery. Similar to the CNG stations, extent polygons were generated of the remaining 12 stations using aerial imagery. During visual validation of the LNG stations, focus was given on identifying gas station infrastructure, and piping/compressed gas storage cylinders near parking lots and salvage yards.

This dataset is updated annually by the DOE, meaning additional validation will need to be completed in the future as more LNG fueling stations come online. Two stations were already listed as being planned in SoCAB and will be operational in less than a year and will need to be added to the dataset in the future. Similar to the CNG dataset, the LNG dataset assigns the entire station as one polygon, regardless of the number of fuel dispensers.

Results:

The final Vista-LA LNG fueling stations dataset contains polygons for 27 stations in SoCAB. The LNG dataset also includes metadata describing the station name, pressures (units: psi), types of dispensing capability, maximum vehicle size accommodation, recalculated GPS coordinates, and validation notes including any changes made and date of last update.

3.2.1.3 Natural Gas Compressor Stations (Vista-LA layer)

The natural gas compressor station (IPCC – 1B2) dataset was obtained using the U.S. Environmental Protection Agency's Facility Level Information on GHG online reporting Tool (EPA FLIGHT). We approximated their locations of non-reporting facilities using the EIA compressor station database, which provides postal-code-level information on compressor stations (Maasakkers et al., 2016). We identified two natural gas compressor stations in SoCAB. The locations of both natural gas compressor stations were validated using Google Earth and Esri Basemap aerial imagery (EPA, 2015). Due to restrictions, we only show the location of the reporting compressor station facility in Figures 2 and 3. The EPA data can be complemented with the EIA database for non-reporting facilities for potential future development of gridded emissions products.

3.2.1.4 Natural Gas Pipelines (Vista-LA layer)

Information for natural gas pipelines (IPCC – 1B2) was collected from the California Energy Commission (CEC) and the EIA 2017 dataset (Maasakkers et al., 2016). The CEC dataset provides infrastructure information of major gas transmission and hazardous liquid transmission pipelines in the United States for the year 2012 (CEC, 2012). For California, the raw CEC dataset was georeferenced and clipped to fit

387 the spatial extent of SoCAB. We also obtained high-resolution natural gas transmission pipeline maps
388 from the National Pipeline Mapping System (NPMS) to validate the CEC pipeline layers. The NPMS
389 dataset includes a level of detail similar to that of the CEC dataset, but can be obtained for the entire U.S.
390 There were very minor differences between the CEC and NPMS layers; However, because the NPMS
391 restricts distribution or visualization of this data, they were retained for internal use only. Due to security
392 concerns, the CEC dataset is only shown as static representations in Figures 2 & 3.

393 Unlike the CEC and NPMS data, the EIA dataset is available publicly. The EIA dataset has a lower level
394 of accuracy compared to the CEC dataset. The positional accuracy of the EIA dataset is ± 3,000 meters
395 while the positional accuracy of CEC ± 150 meters. The EIA dataset also contains less information on the
396 disaggregation of pipeline segments. The final Vista-LA natural gas pipeline dataset includes a
397 georeferenced and processed version of the EIA dataset and contains 111 polyline segments, however due
398 to the uncertainties noted, it would be ideal to use more spatially resolved datasets for future work.

399 **3.2.1.6 Natural Gas Storage Fields (Vista-LA layer)**

400 *Data sources:*

401 Under IPCC – 1B2, natural gas storage facility point data for the United States was obtained from the
402 U.S. Energy Information Administration (EIA) online database for 2016 (EIA, 2016). Natural gas storage
403 geolocations and spatial extent data were obtained using oil field extents from California Department of
404 Conservation's Division of Oil, Gas, & Geothermal Resources (DOGGR) for 2016 (DOGGR, 2016). The
405 EPA FLIGHT tool was also used for data quality assurance (EPA, 2015).

406 *Data processing, validation and limitations:*

407 Point locations for natural gas storage from EIA in 2016 included 415 points for the entire United States,
408 three of which are in SoCAB. Since underground natural gas storage in California is done in depleted oil
409 fields (EIA, 2008), we determined natural gas storage field extents using the 2016 DOGGR dataset, which
410 contained 516 polygon features for oil field extents in the state of California. Both datasets were first
411 georeferenced. The EIA metadata contained operation metrics for each storage field, which we appended

to the DOGGR polygon shapefiles for these three extracted entries. The EPA FLIGHT online GHG reporting tool was used to validate the geolocations of the Aliso Canyon and Honor Rancho storage facilities. Southern California Gas Company's online information on natural gas storage facilities validated the geolocation and extent of the Playa Del Rey storage facility.

We did not include former gas storage fields that are no longer used, such as the Montebello Oilfield. Nevertheless, it is possible these former storage facilities are still leaking, as it takes many years to deplete the gas to pre-storage conditions (Chilingar and Endres, 2005).

Results:

The Vista-LA natural gas storage field polygon layer contains the spatial information and attribute information of the three natural gas storage facilities located within SoCAB: Aliso Canyon, Honor Rancho, and Playa Del Rey. The final Vista-LA layer contains metadata relating to field type, company name, amount of base gas, working capacity, field capacity, maximum delivery, and recalculated locational coordinates along with validation notes including any changes made and date of last update.

3.2.1.5 Natural Gas Processing Plants (Vista-LA layer)

Data sources:

Natural gas processing plant (IPCC 1B2) geospatial data was obtained from the U.S. Energy Information Administration (EIA) online database for the year 2014 (EIA, 2016).

Data processing, validation and limitations:

The raw 2014 EIA dataset contained point geolocations for 551 processing plants across the United States, with six of these in SoCAB, which were georeferenced. Because the raw EIA dataset was limited to a scale of 1:1,000,000, extensive manual geolocation and validation had to be completed for each plant. EIA's geolocation was manually augmented using aerial imagery, planning documents and environmental assessment reports related to the operating companies of each processing plant. Exact spatial extents for

435 the processing plants were created by identifying infrastructure features such as electrical equipment,
436 piping, vents, smoke or steam stacks with towers, racks, and transformers using both Google Earth and
437 Esri Basemap imagery.

438 *Results:*

439 The Vista-LA natural gas processing plant layer contains verified geolocated polygons of six facilities
440 located in SoCAB. The associated metadata includes information on facility and operator name, plant
441 flow in million cubic feet per day, dry storage in million metric standard cubic feet, energy content of
442 natural gas in British thermal units, barrels of liquid natural gas stored at each facility, recalculated
443 locational coordinates of each polygon, along with validation notes including any changes made and date
444 of last update.

445 **3.2.1.7 Oil and Gas Wells (Vista-LA Layer)**

446 *Data sources:*

447 Data on oil and gas wells was collected from DOGGR for the year 2016 (DOGGR, 2016). The oil and
448 gas well dataset contains information on well status, type, coordinates, and whether it has undergone
449 hydraulic stimulation treatment. Another dataset from DOGGR also includes historical production and
450 injection statistics, owner and operators of the well, and the state of the well.

451 *Data processing, validation and limitations:*

452 In SoCAB, the DOGGR dataset includes 32,537 oil and gas wells, associated with activities such as gas
453 storage, pressure maintenance, water disposal, and other (DOGGR, 2016). Due to the sheer size of the
454 datasets, we assumed the locations of the wells in the DOGGR dataset to be valid for the purposes of this
455 study. The dataset includes 5,804 abandoned wells, some of which may be located underneath buildings
456 and other structures which hinder validation of their locations (Chilingar and Endres, 2005). Validation
457 of this dataset is beyond the scope of this work, even in cases where manual visual inspection methods

and/or automated feature extraction from aerial imagery might be useful. We discuss possible methods for automated feature extraction using aerial imagery further in Section 6.

According to DOGGR, well information varies in accuracy, scale, origin and completeness (DOGGR, 2016). DOGGR uses a variety of sources to establish well locations. These sources include handheld measurements using GPS units derived from DOGGR Division staff, coordinates provided by operators, well summary reports, official notices regarding the intent to drill, coordinates derived from aerial imagery, coordinates generated from a tool in MapInfo based on corner call locations, and coordinates from digitized maps. However, we note that some wells in LA were drilled before accurate records were kept by DOGGR (Chilingar and Endres, 2005).

4 Agriculture, Forestry, and Other Land Use (IPCC Level 1 - Category 3)

In the California GHG Inventory, emissions from Livestock (IPCC Category 3A) are the largest of the IPCC Level 3 categories (Figure 1). Emissions from Biomass Burning (IPCC − 3C1) contribute at the ~0.1% level (Figure 1), and are therefore considered negligible for the purposes of this study. Emissions from Wetlands (IPCC − 3B4) and all other emissions from the Land (IPCC Category 3B), Aggregate Sources and Non-CO_2 Emissions (IPCC Category 3C) and Other (3D) source types of the Agriculture, Forestry and Other Land Use category are also considered insignificant in the domain, and were not included as part of this study. Within the Livestock category, dairies and cattle farms are the major contributors in the SoCAB region (Viatte et al., 2017). Below we describe our methods for collecting GIS data related to these activities within the SoCAB.

4.1 Livestock (IPCC Level 2 - 3A)

The livestock category (IPCC − 3A) includes emissions from enteric fermentation (IPCC − 3A1) and manure management (IPCC − 3A2). Manure management systems vary from facility to facility and broadly fall into dry and wet management practices (Kaffka et al., 2016). Manure can be handled and stored using dry lots, deep pits, solid manure storage, daily spread, digesters (CARB, 2016). In slurry/liquid systems, waste from feedlots and other livestock areas are washed and is collected in ponds, which are commonly referred to as anaerobic lagoons (Kaffka et al., 2016). Emissions from manure

depend the type of management practices employed by the farm or facility (Kaffka et al., 2016). Wet manure management involves washing of feedlots and other livestock areas, and the waste runoff is typically collected in lagoons where CH_4 is produced due to anaerobic conditions. By contrast, dry manure management practices do not wash waste with water, thus reduce anaerobic conditions. Dairies in SoCAB primarily use dry manure management practices due to the copious amounts of water needed in wet manure management practices for the collection, movement and storage of animal wastes. However, recent mobile measurement campaigns verified CH_4 emissions from a small number of dairies with anaerobic lagoons as recently as 2015 (Hopkins et al., 2016b; Viatte et al., 2017). In addition to dairy locations, the locations of some anaerobic lagoons were identified as part of the Vista-LA database as described below.

4.1.1 Enteric Fermentation (IPCC Level 3 - 3A1)

4.1.1.1 Dairies (Vista-LA Layer)

Data sources:

Dairy and cattle farm facility data were collected from the California Regional Water Quality Control Board (RWQCB), Santa Ana Region. The data was drawn from annual reports, which contain information on the location of each dairy, the number and type of the herd (i.e., milking cow, dry cow, heifer or calf) and other livestock located at each facility for the year 2015 (Kashak, 2016). We defined cattle farms as facilities that did not contain any milking cows. Overall, dairy facilities were primarily found to be located in the Chino and in the San Jacinto Basins.

Data processing, validation and limitations:

First, the raw RWQCB dataset was georeferenced to match the spatial information of the datasets in Vista-LA. Next, all 110 locations of dairies and cattle farms were validated with Google Earth's historical imagery tool for the year 2015. Facility addresses and coordinates were used to validate the true locations of the farms. When verifying with aerial imagery, focus was given to dairy/cattle facility infrastructure such as: feedlots, manure lagoons, animal housing structures, and open pastures. The dairy and cattle farm

509 locations were deemed accurate if the geographic location in the RWQCB dataset was confirmed with
510 aerial imagery and the coordinates did not overlap another facility. During processing and validation, we
511 identified and manually corrected the locations for two farms with incorrect addresses based on aerial
512 imagery. Additionally, we corrected geolocations for twelve farms that were located incorrectly in the
513 original RWQCB dataset using Google Earth aerial imagery and the facility address information given in
514 the RWQCB dataset.

515 The RWQCB did not report the quantity of feedlots or manure lagoons per dairy, but did include several
516 other types of information which will be useful for estimating emissions from manure management,
517 including: annual manure produced, manure hauled, manure spread to cropland and amount of wastewater
518 produced. The RWQCB dataset also contained several facilities that were neither dairy nor cattle farm,
519 such as a livestock market and beef packing facilities. We removed these from the final Vista-LA layer
520 because we do not expect significant CH_4 emissions from these facilities.

521 It was difficult to obtain spatial extent for each facility because they were difficult to differentiate, as
522 many facilities displayed common features. For this reason, all final dairy/cattle facility locations are
523 point-based in the Vista-LA database. The RWQCB database did not report the number of milking cows,
524 dry cows, heifers or calves for the seven of the farms in Chino. While it does not affect the geolocation
525 in our facility maps, information on cattle populations specific to these farms will be helpful for estimating
526 CH_4 emissions from these facilities.

527 *Results:*

528 The final Vista-LA dairy layer contains a total of 110 livestock facilities in SoCAB: 22 dairies and one
529 cattle farm in the San Jacinto Basin; and 56 dairies, 26 cattle farms and five other livestock farms in
530 Chino. Locations for all facilities were validated using Google Earth imagery. Validation notes have been
531 appended to the spatial dataset for further information on which facilities were corrected for location.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

532 **4.1.2 Manure Management (IPCC Level 3 - 3A2)**

533 **4.1.2.1 Anaerobic lagoons (Vista-LA layer)**

534 *Data sources:*

535 In terms of manure management practices, Vista-LA focused on the collection of GIS structures for wet
536 manure management. Specifically, anaerobic lagoons (IPCC – 3A2) were identified using visual
537 inspection of aerial imagery since publicly available GIS datasets on anaerobic lagoons were not
538 available. Therefore, we created a preliminary geospatial dataset of SoCAB anaerobic lagoons starting
539 from the dairies and cattle farms GIS data. Anaerobic lagoons, also commonly called manure lagoons,
540 are considered sub-facility infrastructures within dairy/cattle farms.

541 *Data processing, validation and limitations:*

542 Point locations of manure lagoons at each dairy farm were visually determined using aerial imagery from
543 the National Agriculture Imagery Program (NAIP) and Google Earth's Time Tool for the year 2015.
544 Infrastructure of anaerobic lagoons were identified near a dairy/cattle farm facility by their distinct
545 rectangular shape and brown to dark blue color associated with the color of wash water from manure
546 waste. Often, aerial imagery showed evidence of cattle, further confirming the facility location. Once a
547 lagoon structure was identified, GIS processing tools were used to create point data for the geolocation.

548 Most anaerobic lagoons in SoCAB were found in the Chino/Ontario region, the area with the densest
549 clusters of dairy farms. However, the identified geolocation of lagoons for year 2015 is likely to change
550 in the near future due to rapid land use developments in the region and fluctuating manure management
551 practices (Hirsch, 2006). Further work could be done with automated feature extraction with
552 contemporaneous imagery, as discussed in Section 6.3.

553 *Results*

The Vista-LA layer for anaerobic lagoons contains 228 point locations in the Chino, Ontario, and Riverside regions. The final layer includes geolocations given by latitude and longitude with validation notes including any changes made and date of last update.

5 Waste (IPCC Level 1 – Category 4)

5.1 Solid Waste Disposal (IPCC Level 2 - 4A)

Solid waste disposal includes both managed and unmanaged waste disposal sites, as well as uncategorized disposal sites. The largest CH_4 emissions are expected from managed waste disposal sites, hence the site category we included in the Vista-LA data product.

5.1.1 Managed Waste Disposal (IPCC Level 3 - 4A1)

Managed Waste Disposal (IPCC − 4A1) is the third largest IPCC Level 3 emissions source in the California GHG inventory, trailing only the Enteric Fermentation (IPCC − 3A1) and Manure Management (IPCC − 3A2) categories. Landfills (solid waste disposal sites) constitute the source of Vista-LA data for this source. Vista-LA includes both active and inactive landfills, with the status recorded in the metadata.

5.1.1.1 Landfills (Vista-LA Layer)

Data sources:

The Vista-LA layer for landfills (IPCC – 4A1) was created using the California Air Resource Board's 2014 landfill data and the 2015 California's Department of Resources Recycling and Recovery's (CalRecycle) Solid Waste Information System (SWIS) dataset (CARB, 2014; CalRecycle, 2015). Geolocation and spatial extent for each individual landfill facility was generated and verified using the 2005 and the 2012 SCAG land use dataset, Google Earth, and Esri Basemap aerial imagery (see http://gisdata.scag.ca.gov/Pages/GIS-Library.aspx).

Data processing, validation and limitations:

The CARB 2014 landfill dataset contained locational records for 372 potential methane-producing landfills for the state of California with 73 of those landfills located in SoCAB. The CalRecycle/SWIS point dataset contained information on 3,087 landfills for all of California, with 759 in SoCAB. CalRecycle landfills located in SoCAB were queried to isolate only the 353 tagged as "solid waste disposal facilities" or "solid waste landfills". Finally, 19 duplicate entries were identified and removed.

The geolocation and spatial extent of the 334 unique landfills in the CalRecycle/SWIS dataset were verified using SCAG 2005 and SCAG 2012 land use datasets. We used the SCAG land use code 1432 (solid waste disposal facilities) to identify the polygon features associated with active dumps and sanitary landfill operations. We used both SCAG 2005 and SCAG 2012 for maximum amount of information on landfill extent information because neither SCAG dataset contained all landfill locations. In the raw SCAG 2005 dataset, there were 247 individual polygons associated with the land use code 1432 in SoCAB. The SCAG 2005 dataset showed multiple polygon features for the same facility in some cases, so this dataset was further refined by merging multiple polygons that comprise a known facility location based on the refined CalRecycle/SWIS dataset. Simplified polygon features were created for all the distinguishable solid waste disposal sites based on cross-checking the SCAG dataset with the point data from CalRecycle and World Imagery. In Los Angeles County, the 86 total polygons associated with land use code 1432 were aggregated into 18 individual landfills; in Orange County, 60 polygons were aggregated into 7 individual landfills; in Riverside County, 61 polygons were aggregated into 13 individual landfills; and in San Bernardino County, 40 polygons were aggregated into 10 individual landfills. This process was again repeated using the SCAG 2012 dataset that had 211 polygons designated as land use code 1432. Overall, 48 landfills of the total 334 were identified from the intersection of SCAG 2005/2012 solid waste disposal facilities and landfills from the CalRecycle/SWIS dataset.

The location and spatial extent of the remaining 286 landfills from the CalRecycle/SWIS dataset had to be manually validated and/or generated using Google Earth Imagery along with other GIS methods including geoprocessing and digitizing. We were able to verify the location of 188 out of the 286 designated landfills. Unfortunately, their extent and shape could not be determined using imagery since they had long been closed and thus modified or built upon significantly. A placeholder polygon was

generated indicating the historical location with extent and shape determined and digitized using Google Earth imagery (for example canyons, excavated pits, and barren land). However, land use changes that occur on surfaces of former landfills vary from site to site. As seen with verification procedures for power plants, time sensitive imagery is critical when evaluating the existence and geolocations of landfills. Using the CalRecycle/SWIS metadata for address location for enhanced verification proved to be difficult because often there were no landfill related features at these addresses, requiring manual geolocation and correction.

The verified 334 polygons were subset by matching the SWIS number in the CalRecycle metadata to the SWIS number of the 73 potential methane-producing landfills in the CARB dataset in order to produce the final Vista-LA dataset.

Results:

In total, 73 potential methane-producing landfills were identified, all locations verified, and polygons were generated in the final dataset using the actual extent or placeholder if extent could not be verified. The metadata from the CalRecycle/SWIS dataset was appended to the final Vista-LA landfill polygon dataset. This includes site-specific information, such as throughput, capacity, and waste types (CalRecycle, 2015). Validation notes include whether facility extent was derived from SCAG 2005, SCAG 2012, or Google Earth imagery. The final polygon data for landfill extents can be separately categorized by landfill status: active, closed, clean-closed, closing, absorbed, and inactive. The operation status breakdown is as follows: 2 were absorbed, 17 were active, 3 were clean-closed (site is considered to cease to exist as a solid waste disposal site, but records are kept to document the status of the site), 308 were closed, 1 was closing, and 3 were inactive.

5.2 Wastewater Treatment and Discharge (IPCC Level 2 – 4D)

Wastewater treatment and discharge includes both domestic and industrial wastewater treatment facilities (Table 1).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

627 **5.2.1 Domestic and Industrial Wastewater Treatment and Discharge (IPCC Level 3 – 4D1 and 4D2)**

628 Wastewater treatment in SoCAB is primarily done through aerobic sludge digestion, which has no
629 associated CH_4 emissions in the California GHG inventory (CARB, 2016). However, in low oxygen
630 conditions, CH_4 may be emitted as a by-product of enhanced denitrification present in water recycling
631 systems (e.g., open tanks in treatment facilities; Townsend-Small et al., 2012). Many wastewater
632 treatment plants also use anaerobic digesters which collect CH_4 for eventual combustion, but may have
633 fugitive CH_4 emissions. Of the urban sources of CH_4, this source is perhaps the most uncertain (Hopkins
634 et al., 2016a). For the purposes of this study, we assume CH_4 emissions are most likely to arise from the
635 plants with the largest daily flow capacity; however, emissions could also potentially arise from various
636 points of collection and/or drainage of wastewater and sewage.

637 **5.2.1.1 Wastewater Treatment Plants (Vista-LA Layer)**

638 *Data sources:*

639 The final Vista-LA wastewater treatment plant layer (IPCC – 4D1 and 4D2) relies on data from the State
640 Water Resources Control Board Facility Report Tool (SWRCB FRS), SCAG 2012 land use dataset,
641 Google Earth and Esri Basemap aerial imagery. The Vista-LA wastewater treatment plant dataset
642 provides accurate location, extent, and facility level metrics for the largest domestic wastewater treatment
643 plants in SoCAB.

644 *Data processing, validation and limitations:*

645 From the SWRCB FRS, we obtained information for wastewater treatment plants in the LA Basin for
646 2016, including facility names, addresses, coordinates, and the design flow rate in million gallons per day.
647 The raw data, which contained information on 152 plants for the state of California, 36 of which were
648 located within SoCAB, 26 which contained design flow metrics. We geocoded the CARB data as points,
649 and found many uncertainties in geolocation of wastewater treatment plants. Because of the relatively
650 small number of plants with metrics in SoCAB, we were able to successfully resolve this uncertainty.

We generated polygons and validated plant geolocation for the 26 SoCAB wastewater treatment plants using Google Earth and Esri Basemap aerial imagery and SCAG 2012 land use data. SCAG data was used to first obtain facility spatial extents. The SCAG land use code 1433, ("liquid waste disposal facilities"), was used to verify locations and determine the spatial extent of the original list of wastewater treatment plants. In total, 189 polygons were classified with this land use code. 44 of those polygons were directly matched with point locations from the SWRCB FRS dataset. Since some facilities had multiple polygons associated to them, they had to be manually merged in order to associate one polygon with one wastewater treatment facility. After this merging procedure, 44 polygons were merged to 11 polygons that directly matched the location of 11 SWRCB FRS wastewater treatment plants. Polygons for the other 15 plants were digitized using both Google Earth and Esri Basemap imagery as reference for spatial extent. Each of the 26 plants was successfully validated using aerial imagery. During the manual validation procedure, attention was given to identifying features such as spreading grounds, aeration fields, water injection plants, and circular tanks. EPA FRS data for the year 2013 was used as a verification source and contains information about wastewater treatment plants operating within petroleum refineries and power plants (U.S. Environmental Protection Agency, 2016).

The ten additional plants from the SWRCB FRS dataset were not included in the final Vista-LA dataset because they did not contain facility level metrics and design flow rates, only location information. Additionally, it is difficult to identify sub-facility plant infrastructure since aerial imagery can only provide a certain degree of context.

Results:

The final Vista-LA wastewater treatment plant layer contains geolocations for 26 wastewater treatment plant facilities. Five of these wastewater treatment plants were found to be co-located with power plants. The metadata contains information about the station name, design flow rate metrics, recalculated locational coordinates of each facility, and validation notes including any changes made and date of last update.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

6 Discussion

6.1 Vista-LA Data Summary

The Vista-LA database consists of 33,353 individual features as points, lines and polygons among thirteen spatial layers, providing a spatial representation of major CH_4 production sources in SoCAB (Figure 2). For the nine polygon layers, Vista-LA depicts the true spatial extent of each facility, a major advance over the original source data. Pipelines are represented as lines, and oil and gas wells are represented as points, which we consider to be the most accurate representation of these sources. The remaining two sources currently represented by points— dairies and anaerobic lagoons— require future work to accurately describe their spatial extents.

The maps in Figures 3-5 show the spatial distributions of potential sources of CH_4 in the IPCC Level 1 categories: Energy (1), Agriculture (3), and Waste (4), respectively. The highest density of CH_4 emitting infrastructure is located in the western portion of SoCAB in Los Angeles and Orange Counties (Figure 2). The Energy sector, specifically oil and gas wells, account for the majority of the spatial inventory (32,537 features) and are primarily located in southern and northwestern Los Angeles County and northern Orange County (Figure 3). The Agriculture sector is dominated by dairies and cattle farms located in the Chino and San Jacinto Basins of San Bernardino and Riverside Counties (Figure 4). By contrast, landfills and wastewater treatment plants are relatively evenly distributed throughout SoCAB (Figure 5).

In total, Vista-LA polygons cover 117 km^2, or 0.68% of the 17,108km^2 extent of SoCAB, substantially narrowing the area over which surveys for fugitive and not-inventoried CH_4 sources should be carried out. This spatial structure more closely matches the "hotspot" nature of atmospheric CH_4 that has been observed in SoCAB at the scale of meters to kilometers (Hopkins et al., 2016b) than is represented by existing gridded products such as CALGEM (10 km x10 km; Jeong et al., 2013) and EPA/Harvard (0.1° x 0.1°; Maasakkers et al., 2016).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

700 **6.2 Vista-LA Data Completeness and Uncertainty**

701 The goal of Vista-LA is to provide a complete representation of potential anthropogenic CH_4 emission
702 sources in SoCAB; however, ensuring complete spatial coverage of important CH_4 sources is challenging.
703 We made the simplifying assumption that including the eight IPCC Level 3 sources which constitute
704 ~99% of the expected California GHG inventory emissions would capture the most important CH_4
705 sources in SoCAB, omitting rice cultivation, imported electricity, and coal mining which are not present
706 in the domain (Figure 1). We also added two new source types that are not included in the California
707 GHG inventory: compressed natural gas (CNG) fueling stations (total: 109 locations) and liquefied
708 natural gas (LNG) fueling stations (total: 27 locations). While these sources are not presently included in
709 the inventory, there was sufficient evidence for fugitive CH_4 emissions for inclusion in Vista (e.g., Figure
710 8; Clark et al., 2017; Hopkins et al., 2016b), particularly given that SoCAB contains 32% and 57% of the
711 state's CNG and LNG fueling stations, respectively.
712
713 We verified our list of included source categories against the key source categories at the national level
714 from the U.S. EPA inventory (EPA, 2016), and found that rice cultivation and coal mining were the only
715 source types contributing >1% of total emissions that were not included. We also compared Vista's source
716 categories to observations of CH_4 hotspots in Los Angeles. The known sources of enhanced CH_4 levels—
717 landfills, cattle, water treatment, power plants, CNG fueling, natural gas pipelines, oil refineries, and oil
718 fields— corresponded to Vista layers with the exception of geologic seeps. This correspondence suggests
719 that Vista-LA is well suited for source attribution of anthropogenic CH_4 hotspots in SoCAB.
720
721 We omitted several categories that might have important contributions to CH_4 emissions in SoCAB, such
722 as transportation. Although transportation produces ~1% of California inventoried CH_4 emissions (and
723 <0.3% of national emissions; EPA, 2016), it likely comprises a greater fraction of SoCAB emissions
724 given the greater density of traffic in the region. We have chosen not to include a spatial layer for
725 transportation in this version of Vista—we view Vista as primarily a tool for attribution of large fugitive
726 CH_4 emission sources, and there is no evidence for this type of emission from conventionally fueled
727 vehicles. Fugitive CH_4 emissions detected along roadways are more likely to originate from leaks in

natural gas pipelines that lie underneath the road surface (Chamberlain et al., 2016). Adding a spatial layer for transportation would be simple to achieve, such as by including a map of the road network or from an existing high resolution inventory such as Hestia (Rao et al., *in review*), and needs to be included in an emissions inventory.

We also omitted two possible natural sources of CH_4 emissions in SoCAB—geologic seeps and wetlands—because they are not included in the most recent version of the California GHG Inventory (geologic seeps are inventoried as "excluded" sources). These sources have not been included in this version of Vista despite the large observed emissions from geologic seeps in SoCAB (Farrell et al., 2013; Hopkins et al., 2016b). We have identified possible spatial datasets to include in future versions of Vista (USGS maps of Natural Oil and Gas Seeps in California, https://walrus.wr.usgs.gov/seeps/ and U.S. Fish and Wildlife Service National Wetlands Inventory, https://www.fws.gov/wetlands/). There is a potential that either sources may contribute significantly to the SoCAB emissions budget—we anticipate including these in future versions.

There is also uncertainty in the spatial representation of sources in Vista-LA. We assumed that the spatial location of sources by linking facility-level (or pipeline-, or oil well-level) datasets to IPCC source categories (Table 1), although there may not be a perfect correspondence between mappable infrastructure and these sources. An alternative approach would have been to map out the lifecycle of each IPCC Level 1 sector (e.g., Energy, Agriculture, Waste), and determine spatial locations of each lifecycle phase in the domain. We chose the IPCC approach because is more standardized, and hence applicable to other cities. The lifecycle approach requires local knowledge, and is likely to differ among cities and regions (Hopkins et al., 2016a). We also assumed that most CH_4 emissions come from the main facility associated with an emitting activity, such as wastewater treatment plants for wastewater emissions, rather than the sewer network. This assumption may not hold true, for example when manure is exported from dairies and treated elsewhere. In contrast to waste and agriculture, we included a higher level of detail for the natural gas system because there is more evidence of quantitatively significant CH_4 emissions from distributed parts of the network (e.g., pipelines).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Finally, we recognize error inherent in the availability of data, and in the original data sources themselves. Vista-LA relies on publicly available datasets. Consequently, we are constrained by (a) lack of data on some infrastructure types that may be transient or has never been collected, such as the locations of manure piles; (b) unavailability of proprietary data, such as the locations of petroleum storage tanks or gathering pipelines; and (c) data that poses a security risk, and hence cannot be distributed, such as specific locations of natural gas compressor stations and high pressure transmission pipelines. Within the datasets themselves, we found errors in geolocation and missing facilities. As described in the text, we took steps to control these errors inherent to the source datasets by performing visual validation and using multiple datasets for the same source category. In most original datasets, geolocation of some facilities was incorrect by up to several kilometers for various reasons, and some geolocations corresponded to the offices or street address of facilities rather than actual facility location (e.g., for dairies). Raw datasets for landfills, natural gas processing plants, and wastewater treatment plants also had coarse spatial resolution resulting in uncertain geolocations, which we corrected. However, we were unable to visually validate all sources, including oil wells and landfills. With respect to omission of CH_4 emitting facilities, in many cases, the total number of facilities varied among the raw datasets that were used to construct the Vista-LA layers (e.g., landfills, natural gas processing plants, natural gas storage fields, power plants, and wastewater treatment plants). For power plants, there are many small facilities in California (around 4,020 plants) that are not included in the EIA dataset (110 plants). Overall, these small power plant facilities only represent about 3% of the total statewide electricity production related fossil fuel CO_2 emissions, and therefore most are not tracked. Some of these facilities are also not grid-tied, but are facilities that generate electricity for an industrial process (K. Gurney, *personal communication*). We found that at least nine power plants were located within the bounds of a petroleum refinery facility in SoCAB. We have not considered potential emissions from smaller facilities or those contained within other facilities, but this may be important to consider as part of future work.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

781 **6.3 Vista-LA Applications**

782 At present, Vista-LA does not contain the bottom-up empirical measurements required for the creation of
783 an accurate fine-scale CH_4 inventory (as in Lyon et al., 2015). However, making and interpreting
784 atmospheric CH_4 measurements is easier for the dense, heterogeneous landscape of SoCAB with the
785 guidance of the Vista-LA spatial layers. Vista-LA represents a much-needed first step towards the
786 development of a fine-scale urban CH_4 emissions inventory that can be used for design and interpretation
787 of CH_4 hotspot surveys. Importantly, Vista-LA compliments the tiered remote sensing observation
788 strategies for regional top-down CH_4 emission measurements. Figure 6 summarizes potential applications
789 of Vista-LA, which are described in the following sections.

790 **6.3.1 Research Planning**

791 Vista-LA is already in use as a planning tool for research aimed toward better understanding CH_4
792 emissions in SoCAB, through guiding design of CH_4 super-emitters surveys and appropriate selection of
793 locations for stationary CH_4 monitoring. Figure 7 shows how Vista-LA has been used for planning
794 airborne remote sensing campaigns to survey CH_4 point sources. Guidance from Vista-LA allows airborne
795 campaigns to be designed for maximum coverage of key infrastructure identified in SoCAB. In Figure 7,
796 aircraft flight lines shown in green illustrate a path optimized for coverage of key CH_4 infrastructure in
797 SoCAB.

798 **6.3.2 CH_4 Hotspot Detection**

799 In addition, Vista-LA can be used to interpret observations of atmospheric CH_4 measurements, addressing
800 the challenge of source attribution of CH_4 hotspots detected in airborne and ground-based surveys.
801 Pinpointing the source of fugitive CH_4 emissions sources in dense mixed-land use urban areas has been
802 an ongoing challenge (Cambaliza et al., 2015; Hopkins et al., 2016b). In many urban areas, many potential
803 CH_4 sources are located in close proximity, such as in the Ports of Los Angeles and Long Beach, which
804 contain extensive fossil fuel use and transportation, active oil drilling and refining, and wastewater
805 treatment. Figure 8 shows an overlay of CH_4 observations from a mobile CH_4 survey in June 2013, where

high frequency in situ CH_4 observations were made from a moving van, as described in Hopkins et al. (2016b). Atmospheric CH_4 levels are shown as colored boxes with an "x" in the center, with blue colors representing near background levels, and warmer colors (red) representing elevated CH_4. The same region was flown by the Hyperspectral Thermal Emission Spectrometer (HyTES) in July 2014. Plumes of CH_4 were retrieved from HyTES radiance data, and are shown in green (Hulley et al., 2016). In the scene, Vista-LA shows roughly a dozen oil and gas wells and a CNG fueling station near the elevated CH_4 observations, narrowing down the number of potential sources greatly. Together with wind direction, these observations plus Vista-LA enable attribution to the facility level.

6.3.3 Stakeholder and Public Engagement Tool

Vista-LA also has the potential to guide CH_4 mitigation efforts by identifying persistent CH_4-emitting infrastructure when combined with atmospheric measurements. In order to control CH_4 emissions effectively in cities, it is essential to understand the fine-scale spatial distribution of stationary sources from a broad range of industries in the energy, agriculture, and waste sectors. Local regulators and city planners may be able to use the location information presented in Vista-LA, combined with targeted surveys or observations, to develop enhanced CH_4 monitoring and mitigation strategies for their cities. Vista-LA may also be extensible to air-quality mitigation efforts. For example, some processes that emit CH_4 also result in co-emission of other gases that are important for climate and air quality. By incorporating new information in Vista, such as information on permitting and/or sub-facility infrastructure information, users may be able to evaluate the air quality co-benefits associated with fugitive CH_4 mitigation strategies.

6.4 Future Directions: Vista CH_4 Database

Vista-LA was primarily developed to identify CH_4 emitting infrastructure in SoCAB. However, we anticipate our approach could be scaled to other regions and over larger spatial scales, including the state of California, the contiguous U.S., and possibly internationally. Expanding the Vista-LA database across the state of California is highly feasible given that our framework is consistent with how the State of California reports CH_4 emissions. Additionally, many of the raw data sources used in the development of

Vista-LA already encompass state-level or national-level spatial extents (Table 1). In theory, the approach could also be expanded to any region on Earth, as long as an IPCC (or similar) inventory and geolocation data for the top-emitting sources are publicly available. Our framework is also dependent on the availability of timely, reliable public datasets. In this regard, Los Angeles and the state of California are perfect testbeds for development. By contrast, many regions may not have such information available publicly, especially in developing nations.

Efforts to expand the database could be enhanced by the use of automated feature extraction techniques. For example, the use of automated feature extraction techniques could expedite the process of identifying, extracting, and updating relevant infrastructure features. Automated feature extraction involves machine-learning algorithms used to recognize patterns through image processing (see e.g., Yuan, 2016; Castelluccio et al., 2015). In this way, aerial imagery in the SoCAB could essentially be used to parse through and precisely locate features such as converted landfills, oil and gas wellheads, anaerobic lagoons, and/or wastewater treatment plants. Future work may benefit from the use of automated feature recognition algorithms using software such as eCognition, ERDAS IMAGINE, GeoMedia, InterIMAGE, RemoteView and SOCET GXP in order to identify and update spatial information for facilities and sources not yet known or housed in current publicly available databases.

In the future, the information in the Vista-LA could be used generate a high-resolution bottom-up gridded CH_4 emissions product for SoCAB. While the initial goal of the Vista-LA database was to provide facility location information, some of the facility- and sub-facility-level activity and operation information contained within the metadata may also be useful for assigning emission factors to each source (see Metadata in the Supplementary Information). For example, the Vista-LA dairy layer contains information on herd population by type, which could be used to estimate emissions factors for enteric fermentation and manure management. Manure management practices are known to vary widely by region, even within the state of California. Potentially, the information in the Vista-LA database could also be combined with top-down observations and provide independent validation of bottom-up CH_4 flux estimates, in a similar approach to that shown in Section 6.2.2. In general, utilizing the spatial information in Vista-LA to assign CH_4 emissions estimates could significantly minimize errors in the spatial representation of sources

compared to previous estimates for this region. Updates such as changes to locations, spatial extents, and removal or addition of facilities will be required on at least an annual basis to provide timely and accurate emissions information.

7 Data availability

The final Vista-LA datasets and associated metadata are open access and are available in the Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics (ORNL DAAC) (Carranza et al., 2017; https://doi.org/10.3334/ORNLDAAC/1525).

8 Conclusions

Vista-LA adopts a GIS-based approach to map CH_4 emissions in dense-mixed-land use areas like the South Coast Air Basin. Characterizing CH_4 emissions at the urban scale is incredibly complex, as there exist thousands of structures known to be associated with CH_4 emissions. Vista-LA successfully identifies 33,503 potential CH_4 emitters from three IPCC sectors: Energy, Agriculture, and Waste. Vista-LA contains accurate and validated spatial extent information for nine sources including compressed natural gas fueling stations, liquefied natural gas fueling stations, landfills, natural gas compressor stations, natural gas storage fields, natural gas processing plants, petroleum refineries, power plants, and wastewater treatment plants. It also includes point location of anaerobic lagoons, dairies, and oil and gas wells as well as a natural gas pipeline network for SoCAB. Vista-LA has been assisted in flight planning for CH_4 airborne campaigns, can be used in potential CH_4 hotspot detection, and can essentially be merged with top-down flux estimates for the identification of individual point sources. In this way, Vista-LA represents a first step towards developing a gridded emissions spatial product that can illustrate spatial distribution of CH_4 emissions at a fine-scale. By fusing Vista-LA, automated surface feature recognition, and other various remote sensing point source data products, we could dramatically improve the attribution of methane emissions.

Vista-LA serves as a prototype resource to aide in the development of high-resolution bottom-up gridded models of GHG emissions in densely populated urban areas with a complex variety of sources and can be adapted to larger scales in accordance to characteristics innate to each respective region. The

development of spatially-resolved carbon emission datasets can offer significant advances in understanding, managing, and mitigating carbon emissions from cities. Generally, uncertainties in emission sources and their locations in inventories hinder the implementation of mitigation policies. To be useful, CH_4 emissions information is needed at the scale of individual sources. In addition to accurate and timely spatial information, urban CH_4 emissions inventories should also be flexible enough to incorporate new information, while remaining relevant for observation-based research efforts such as surveys, hotspot detection, and inversion modeling. Finally, the information should be communicated in an open-source, transparent, and well-documented manner.

Acknowledgements

A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). The NASA DEVELOP program (http:// develop.larc.nasa.gov/) provided the primary support for this project for VC and TR during 2015. IFV also participated in the NASA DEVELOP program and received support from the NASA Scholars Program and the Minority University Research and Education (MUREP) Project during 2014-2016. FMH was supported by an appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, California Institute of Technology, administered by Universities Space Research Association under contract with NASA. KRV was supported by the National Institutes of Standards and Technologies (NIST) Greenhouse Gas and Climate Science Measurements Program. The authors are very thankful to Dr. Ben Holt, G. Miller, and C. Rains from the NASA/JPL DEVELOP program for helpful discussions regarding data analysis, and to M. Limb for comments on the manuscript. We are also thankful for valuable advice and guidance from the California Air Resources Board (ARB) Greenhouse Gas Emission Inventory Branch, including: A. Huang, L. Hunsaker, K. Eslinger, W. Widger, J. Charrier, and G. Ruiz, and T. Rose, U. Prins, F. Thong and G. Bemis for assistance with data from the California Energy Commission, and Edward Kashak from the Regional Water Quality Control Board, Santa Ana Region. This writing has used information provided by the California Energy Commission. This writing does not necessarily represent the views of the Energy Commission, its employees, or the State of California. The

911 Energy Commission and the State of California make no express or implied warranties, and assume no
912 legal liability for the information contained in this writing. © 2017. All Rights Reserved.

913 **References**

914 Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., Huang, J.,
915 Elvidge, C. and Baugh, K.: A multiyear, global gridded fossil fuel CO_2 emission data product: Evaluation
916 and analysis of results, J. Geophys. Res. Atmos., 119(17), doi:10.1002/2013JD021296, 2014.

917 Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Pétron, G., Joordan, S. M., Tans, P., Wilcox, J.,
918 Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardly, D. and Harriss,
919 R.: Methane Leaks from North American Natural Gas Systems, Science (80-.)., 343(6172), 733–735,
920 doi:10.1126/science.1247045, 2014.

921 CalRecycle: SWIS Facility/Site Search, [online] Available from:
922 http://www.calrecycle.ca.gov/SWFacilities/Directory/Search.aspx, 2015.

923 Cambaliza, M. O. L., Shepson, P. B., Bogner, J., Caulton, D. R., Stirm, B., Sweeney, C., Montzka, S. a.,
924 Gurney, K. R., Spokas, K., Salmon, O. E., Lavoie, T. N., Hendricks, A., Mays, K., Turnbull, J., Miller,
925 B. R., Lauvaux, T., Davis, K., Karion, A., Moser, B., Miller, C., Obermeyer, C., Whetstone, J., Prasad,
926 K., Miles, N. and Richardson, S.: Quantification and source apportionment of the methane emission flux
927 from the city of Indianapolis, Elementa, 3, doi:10.12952/journal.elementa.000037, 2015.

928 CARB: California's 2000-2012 Greenhouse Gas Emissions Inventory Technical Support Document.,
929 2014a.

930 CARB: The California Almanac of Emissions and Air Quality - 2013 Edition Rep., California Air
931 Resources Board, Sacramento, CA., 2014b.

932 CARB: California Greenhouse Gas Inventory for 2000-2013 — by Sector and Activity Electricity
933 Generation (In State) California Greenhouse Gas Inventory for 2000-2013 — by Sector and Activity.,
934 2015.

CARB: Documentation of California's 2000-2015 GHG Inventory — Index, [online] Available from: http://www.arb.ca.gov/cc/inventory/doc/doc_index.php (Accessed 10 August 2016), 2016.

Carranza, V., Rafiq, T., Frausto-Vicencio, I., Hopkins, F., Verhulst, K. R., Rao, P., Duren, R. M. and Miller, C. E. 2017. NACP Sources of Methane Emissions (Vista-LA), South Coast Air Basin, California, USA. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1525

Castelluccio, M., Poggi, G., Sansone, C. and Verdoliva, L.: Land Use Classification in Remote Sensing Images by Convolutional Neural Networks, arxiv.org, cs.CV, 2015.

Chamberlain, S. D., Ingraffea, A. R. and Sparks, J. P.: Sourcing methane and carbon dioxide emissions from a small city: Influence of natural gas leakage and combustion, Environ. Pollut., 218, 102–110, doi:10.1016/j.envpol.2016.08.036, 2016.

Chilingar, G. and Endres, B.: Environmental hazards posed by the Los Angeles Basin urban oilfields: an historical perspective of lessons learned, Environ. Geol., 47(2), 302–317, doi:10.1007/s00254-004-1159-0, 2005.

Clark, N. N., McKain, D. L., Johnson, D. R., Wayne, W. S., Li, H., Akkerman, V., Sandoval, C., Covington, A. N., Mongold, R. A., Hailer, J. T. and Ugarte, O. J.: Pump-to-Wheels Methane Emissions from the Heavy-Duty Transportation Sector, Environ. Sci. Technol., 51(2), 968–976, doi:10.1021/acs.est.5b06059, 2017.

Conley, S., Franco, G., Faloona, I., Blake, D., Peischl, J. and Ryerson, T.: Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA, Science (80-.)., 351(6279), 1317–1320, doi:10.1126/science.aaf2348, 2016.

Dlugokencky, E. J., Nisbet, E. G., Fisher, R. and Lowry, D.: Global atmospheric methane: budget, changes and dangers, Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci., 369(1943), 2058–2072, doi:10.1098/rsta.2010.0341, 2011.

959 DOE: Alternative Fuels Data Center, [online] Available from: http://www.afdc.energy.gov/, 2017.

960 DOGGR: GIS Mapping, [online] Available from:
961 http://www.conservation.ca.gov/dog/maps/Pages/GISMapping2.aspx (Accessed 18 June 2016), 2016.

962 Duren, R. and Miller, C.: Measuring the carbon emissions of megacities, Nat. Clim. Chang., 2(8), 560–
963 562, 2012.

964 EIA: Office of Oil & Gas, Natural Gas Division Gas, Gas Transportation Information System, [online]
965 Available from:
966 https://www.eia.gov/pub/oil_gas/natural_gas/analysis_publications/ngpipeline/undrgrndstor_map.html,
967 2008.

968 EIA: Maps: Layer Information for Interactive State Maps, [online] Available from:
969 https://www.eia.gov/maps/layer_info-m.cfm, 2016.

970 EPA: FLIGHT: 2015 Greenhouse Gas Emissions from Large Facilities, [online] Available from:
971 https://ghgdata.epa.gov/ghgp/main.do (Accessed 18 April 2017), 2015.

972 EPA: Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2014., 2016.

973 European Commission Joint Research Centre: Netherlands Environmental Assessment Agency (2010)
974 Emission Database for Global Atmospheric Research (EDGAR), Release Version 4.2., [online] Available
975 from: http://edgar.jrc.ec.europa.eu (Accessed 4 November 2013), 2010.

976 Farrell, P., Culling, D. and Leifer, I.: Transcontinental methane measurements: Part 1. A mobile surface
977 platform for source investigations, Atmos. Environ., 74, 422–431, doi:10.1016/j.atmosenv.2013.02.014,
978 2013.

979 Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J.,
980 Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S. and Green, R.
981 O.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region.,

982 Proc. Natl. Acad. Sci. , 113(35), 9734–9739, doi:10.1073/pnas.1605617113, 2016.

983 Gioli, B., Toscano, P., Lugato, E., Matese, A., Miglietta, F., Zaldei, A. and Vaccari, F. P.: Methane and
984 carbon dioxide fluxes and source partitioning in urban areas: The case study of Florence, Italy, Environ.
985 Pollut., 164, 125–131, doi:10.1016/j.envpol.2012.01.019, 2012.

986 Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B. and Abdul-Massih, M.: Quantification of
987 fossil fuel CO_2 emissions on the building/street scale for a large US city, Environ. Sci. Technol., 46(21),
988 12194–12202, doi:10.1021/es3011282, 2012.

989 Gurney, K. R., Romero-Lankao, P., Seto, K. C., Hutyra, L. R., Duren, R., Kennedy, C., Grimm, N. B.,
990 Ehleringer, J. R., Marcotullio, P., Hughes, S., Pincetl, S., Chester, M. V., Runfola, D. M., Feddema, J. J.
991 and Sperling, J.: Climate change: Track urban emissions on a human scale, Nature, 525(7568), 179–181,
992 doi:10.1038/525179a, 2015.

993 Helfter, C., Tremper, A. H., Halios, C. H., Kotthaus, S., Bjorkegren, A., Grimmond, C. S. B., Barlow, J.
994 F. and Nemitz, E.: Spatial and temporal variability of urban fluxes of methane, carbon monoxide and
995 carbon dioxide above London, UK, Atmos. Chem. Phys., 16(16), 10543–10557, doi:10.5194/acp-16-
996 10543-2016, 2016.

997 Hirsch, J.: Dairies Moving Out of Inland Empire, Los Angeles Times, 2006.

998 Hopkins, F. M., Ehleringer, J. R., Bush, S. E., Duren, R. M., Miller, C. E., Lai, C., Hsu, Y., Carranza, V.
999 and Randerson, J. T.: Mitigation of methane emissions in cities: How new measurements and partnerships
1000 can contribute to emissions reduction strategies, Earth's Futur., 4(9), 408–425,
1001 doi:10.1002/2016EF000381, 2016a.

1002 Hopkins, F. M., Kort, E. A., Bush, S. E., Ehleringer, J., Lai, C., Blake, D. and Randerson, J. T.: Spatial
1003 patterns and source attribution of urban methane in the Los Angeles Basin, J. Geophys. Res. Atmos.,
1004 121(5), 2490–2507, doi:10.1002/2015JD024429, 2016b.

Hsu, Y. K., VanCuren, T., Park, S., Jakober, C., Herner, J., FitzGibbon, M., Blake, D. R. and Parrish, D. D.: Methane emissions inventory verification in southern California, Atmos. Environ., 44(1), 1–7, doi:10.1016/j.atmosenv.2009.10.002, 2009.

Hulley, G. C., Duren, R. M., Hopkins, F. M., Hook, S. J., Vance, N., Guillevic, P., Johnson, W. R., Eng, B. T., Mihaly, J. M., Jovanovic, V. M., Chazanoff, S. L., Staniszewski, Z. K., Kuai, L., Worden, J., Frankenberg, C., Rivera, G., Aubrey, A. D., Miller, C. E., Malakar, N. K., Tomás, J. M. S. and Holmes, K. T.: High spatial resolution imaging of methane and other trace gases with the airborne Hyperspectral Thermal Emission Spectrometer (HyTES), Atmos. Meas. Tech., 9(5), 2393–2408, doi:10.5194/amt-9-2393-2016, 2016.

IPCC: Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories., 2001.

IPCC: 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Ch. 4, Table 4.1., 2006.

Jackson, R. B., Down, A., Phillips, N. G., Ackley, R. C., Cook, C. W., Plata, D. L. and Zhao, K.: Natural Gas Pipeline Leaks Across Washington, DC, Environ. Sci. Technol., 48(3), 2051–2058, doi:10.1021/es404474x, 2014.

Jeong, S., Cui, X., Blake, D. R., Miller, B., Montzka, S. A., Andrews, A., Guha, A., Martien, P., Bambha, R. P., LaFranchi, B., Michelsen, H. A., Clements, C. B., Glaize, P. and Fischer, M. L.: Estimating methane emissions from biological and fossil-fuel sources in the San Francisco Bay Area, Geophys. Res. Lett., 44(1), 486–495, doi:10.1002/2016GL071794, 2017.

Kaffka, S., Barzee, T., El-Mashad, H., Williams, R., Zicari, S. and Zhang, R.: Evaluation of Dairy Manure Management Practices for Greenhouse Gas Emissions Mitigation in California. [online] Available from: http://biomass.ucdavis.edu/wp-content/uploads/2016/06/ARB-Report-Final-Draft-Transmittal-Feb-26-2016.pdf (Accessed 18 April 2017), 2016.

Kashak, E.: Edward Kashak from California Regional Water Quality Board, Santa Ana Region personal

communication with Francesca Hopkins, 2016.

Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., Havránek, M., Pataki, D., Phdungsilp, A., Ramaswami, A. and Villalba Mendez, G.: Greenhouse gas emissions from global cities, Environ. Sci. Technol., 43, 7297–7302, doi:10.1021/es900213p, 2009.

Lamb, B. K., Edburg, S. L., Ferrara, T. W., Harrison, M. R., Kolb, C. E., Townsend-Small, A., Dyck, W., Possolo, A. and Whetstone, J. R.: Direct Measurements Show Decreasing Methane Emissions from Natural Gas Local Distribution Systems in the United States, Environ. Sci. Technol., 49(8), 5161–5169, doi:10.1021/es505116p, 2015.

Lyon, D. R., Zavala-Araiza, D., Alvarez, R. A., Harriss, R., Palacios, V., Lan, X., Talbot, R., Lavoie, T., Shepson, P., Yacovitch, T. I., Herndon, S. C., Marchese, A. J., Zimmerle, D., Robinson, A. L. and Hamburg, S. P.: Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region, Environ. Sci. Technol., 49(13), 8147–8157, doi:10.1021/es506359c, 2015.

Maasakkers, J. D., Jacob, D. J., Sulprizio, M. P., Turner, A. J., Weitz, M., Wirth, T., Hight, C., DeFigueiredo, M., Desai, M., Schmeltz, R., Hockstad, L., Bloom, A. A., Bowman, K. W., Jeong, S. and Fischer, M. L.: Gridded National Inventory of U.S. Methane Emissions, Environ. Sci. Technol., 50(23), 13123–13133, doi:10.1021/acs.est.6b02878, 2016.

Marcotullio, P. J., Sarzynski, A., Albrecht, J., Schulz, N. and Garcia, J.: The geography of global urban greenhouse gas emissions: an exploratory analysis, Clim. Change, 121(4), 621–634, doi:10.1007/s10584-013-0977-z, 2013.

McKain, K., Down, A., Raciti, S. M., Budney, J., Hutyra, L. R., Floerchinger, C., Herndon, S. C., Nehrkorn, T., Zahniser, M. S., Jackson, R. B., Phillips, N. and Wofsy, S. C.: Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts., 2015.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T. and Zhan, H.:

1053 Anthropogenic and Natural Radiative Forcing: In Climate Change 2013: The Physical Science Basis.
1054 Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on
1055 Climate Change, in Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,
1056 pp. 659–740., 2013.

1057 Newman, S., Xu, X., Gurney, K. R., Hsu, Y. K., Li, K. F., Jiang, X., Keeling, R., Feng, S., O'Keefe, D.,
1058 Patarasuk, R., Wong, K. W., Rao, P., Fischer, M. L. and Yung, Y. L.: Toward consistency between trends
1059 in bottom-up CO_2 emissions and top-down atmospheric measurements in the Los Angeles megacity,
1060 Atmos. Chem. Phys., 16(6), 3843–3863, doi:10.5194/acp-16-3843-2016, 2016.

1061 Oda, T. and Maksyutov, S.: A very high-resolution (1km×1km) global fossil fuel CO_2 emission inventory
1062 derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys.,
1063 11(2), 543–556, doi:10.5194/acp-11-543-2011, 2011.

1064 Olivier, J. and Peters, J.: CO_2 from non-energy use of fuels: A global, regional and national perspective
1065 based on the IPCC Tier 1 approach, Resour. Conserv. Recycl., 45(3), 210–225, 2005.

1066 Patarasuk, R., Gurney, K. R., O'Keeffe, D., Song, Y., Huang, J., Rao, P., Buchert, M., Lin, J. C., Mendoza,
1067 D. and Ehleringer, J. R.: Urban high-resolution fossil fuel CO_2 emissions quantification and exploration
1068 of emission drivers for potential policy applications, Urban Ecosyst., 19(3), 1013–1039,
1069 doi:10.1007/s11252-016-0553-1, 2016.

1070 Perata: Senate Bill No. 1368, Chapter 598. [online] Available from:
1071 http://www.energy.ca.gov/emission_standards/documents/sb_1368_bill_20060929_chaptered.pdf, 2006.

1072 Phillips, N. G., Ackley, R., Crosson, E. R., Down, A., Hutyra, L. R., Brondfield, M., Karr, J. D., Zhao,
1073 K. and Jackson, R. B.: Mapping urban pipeline leaks: Methane leaks across Boston, Environ. Pollut., 173,
1074 1–4, doi:10.1016/j.envpol.2012.11.003, 2013.

1075 Pipeline and Hazardous Materials Safety Administration; U.S. Department of Transportation: National
1076 Pipeline Mapping System, [online] Available from: https://www.npms.phmsa.dot.gov/, 2013.

1077 Rao, P., Gurney, K. R., Patarasuk, R., Yang, S., Miller, C. E., Duren, R. M. and Eldering, A.: Spatio-
1078 temporal variations in on-road CO_2 emissions in the Los Angeles Megacity, AIMS Geosci., *(in review)*.

1079 State Water Resources Control Board; California Environmental Protection Agency: Facility Reporting
1080 Service, [online] Available from:
1081 https://ciwqs.waterboards.ca.gov/ciwqs/readOnly/CiwqsReportServlet?inCommand=reset&reportName
1082 =RegulatedFacility/, 2016.

1083 Townsend-Small, A., Tyler, S. C., Pataki, D. E., Xu, X. and Christensen, L. E.: Isotopic measurements of
1084 atmospheric methane in Los Angeles, California, USA: Influence of "fugitive" fossil fuel emissions, J.
1085 Geophys. Res. Atmos., 117(7), 1–11, doi:10.1029/2011JD016826, 2012.

1086 U.S. Energy Information Administration: Refinery Capacity Report., 2015.

1087 U.S. Environmental Protection Agency: EPA Facility Registry Service (FRS): Wastewater Treatment
1088 Plants, Data.Gov [online] Available from: https://catalog.data.gov/dataset/epa-facility-registry-service-
1089 frs-wastewater-treatment-plants, 2016.

1090 Verhulst, K., Karion, A., Kim, J., Salameh, P., Keeling, R., Newman, S., Miller, J., Sloop, C., Pongetti,
1091 T., Rao, P., Wong1, C., Hopkins, F. M., Yadav, V., Weiss, R. F., Duren, R. M. and Miller, C. E.: Carbon
1092 Dioxide and Methane Measurements from the Los Angeles Megacity Carbon Project: 1. Calibration,
1093 Urban Enhancements, and Uncertainty Estimates, Atmos. Chem. Phys., 17, 8313-8341,
1094 https://doi.org/10.5194/acp-17-8313-2017, 2017.

1095 Viatte, C., Lauvaux, T., Hedelius, J. K., Parker, H., Chen, J., Jones, T., Franklin, J. E., Deng, A. J., Gaudet,
1096 B., Verhulst, K. and Duren, R.: Methane emissions from dairies in the Los Angeles Basin, Atmos. Chem.
1097 Phys., 17, 7509-7528, https://doi.org/10.5194/acp-17-7509-2017, 2017.

1098 Wennberg, P. O., Mui, W., Wunch, D., Kort, E. A., Blake, D. R., Atlas, E. L., Santoni, G. W., Wofsy, S.
1099 C., Diskin, G. S., Jeong, S. and Fischer, M. L.: On the sources of methane to the Los Angeles atmosphere,
1100 Environ. Sci. Technol., 46(17), 9282–9289, doi:10.1021/es301138y, 2012.

1101 Wong, C. K., Pongetti, T. J., Oda, T., Rao, P., Gurney, K. R., Newman, S., Duren, R. M., Miller, C. E.,
1102 Yung, Y. L. and Sander, S. P.: Monthly trends of methane emissions in Los Angeles from 2011 to 2015
1103 inferred by CLARS-FTS observations, Atmos. Chem. Phys, 16(20), 13121–13130, doi:10.5194/acp-16-
1104 13121-2016, 2016.

1105 Wong, K. W., Fu, D., Pongetti, T. J., Newman, S., Kort, E. A., Duren, R., Hsu, Y.-K., Miller, C. E., Yung,
1106 Y. L. and Sander, S. P.: Mapping CH_4 : CO_2 ratios in Los Angeles with CLARS-FTS from Mount Wilson,
1107 California, Atmos. Chem. Phys., 15(1), 241–252, doi:10.5194/acp-15-241-2015, 2015.

1108 Wunch, D., Wennberg, P. O., Toon, G. C., Keppel-Aleks, G. and Yavin, Y. G.: Emissions of greenhouse
1109 gases from a North American megacity, Geophys. Res. Lett., 36(15), 1–5, doi:10.1029/2009GL039825,
1110 2009.

1111 Yuan, J.: Automatic Building Extraction in Aerial Scenes Using Convolutional Networks, arXiv.org,
1112 cs.CV, 2016.

1113 Zavala-Araiza, D., Lyon, D. R., Alvarez, R. A., Davis, K. J., Harriss, R., Herndon, S. C., Karion, A., Kort,
1114 E. A., Lamb, B. K., Lan, X., Marchese, A. J., Pacala, S. W., Robinson, A. L., Shepson, P. B., Sweeney,
1115 C., Talbot, R., Townsend-Small, A., Yacovitch, T. I., Zimmerle, D. J. and Hamburg, S. P.: Reconciling
1116 divergent estimates of oil and gas methane emissions, Proc. Natl. Acad. Sci., 112(51), 15597–15602,
1117 doi:10.1073/pnas.1522126112, 2015.

1118

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

Figures

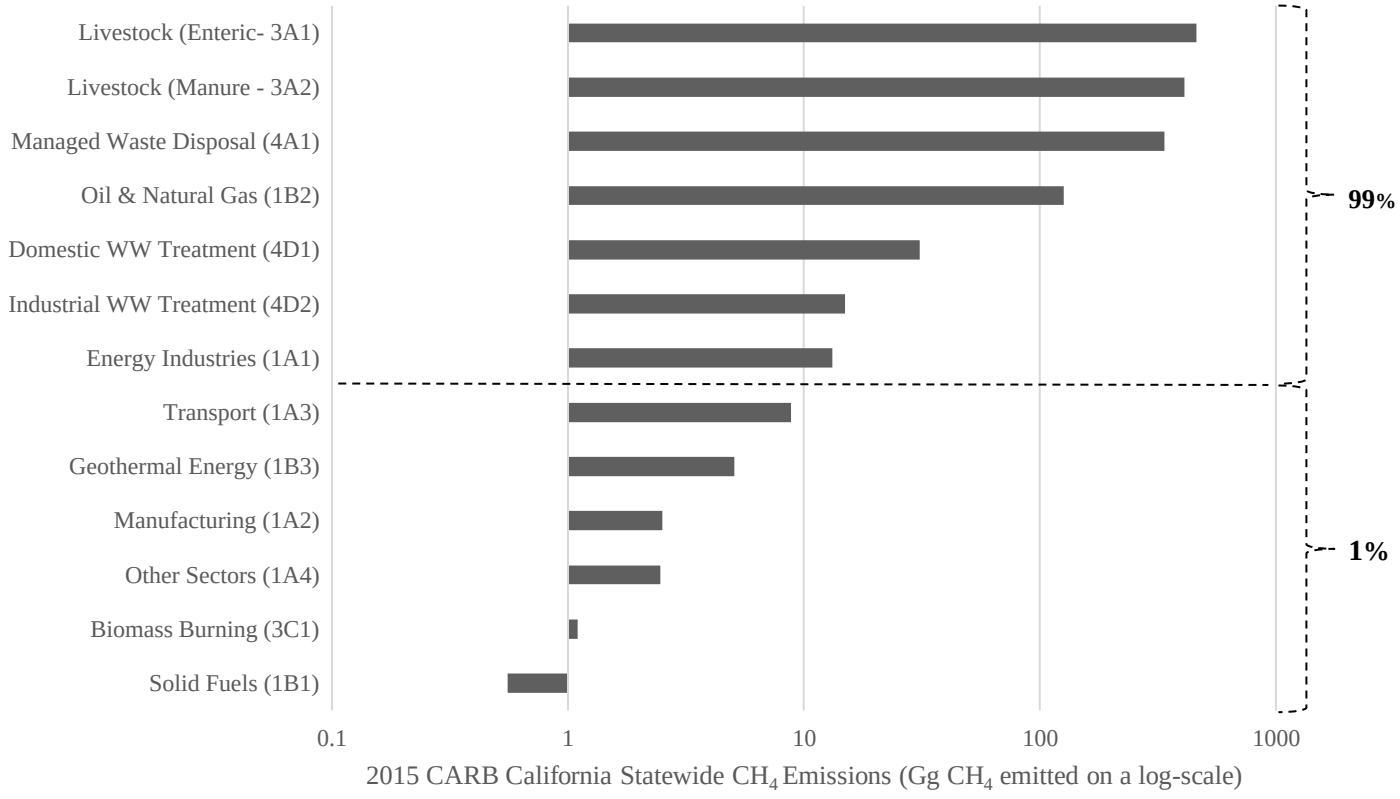

Figure 1: Ranking of inventoried California CH_4 emissions for 2015 by IPCC Level 3 categories. This graph plots the total emitted Gg of CH_4 for each IPCC sector on a logarithmic scale as calculated by CARB for the year 2015. This data is based on IPCC Level 3 categories which are indicated in parenthesis. Only the top 13 of the 35 IPCC Level 3 categories are shown for clarity. The total 2015 emission from these 13 sectors was 1,412.91 Gg CH_4. Emissions from activities that are non-existent in the South Coast Air Basin region were not considered part of the Vista-LA database and are not shown in the graph. These activities included imported electricity (IPCC − 1A1), rice cultivation (IPCC − 3C7), and coal mining (IPCC – 1A2). The top seven IPCC Level 3 categories encompass roughly 99% of California's statewide CH_4 emissions (~1,392 Gg CH_4), are also relevant to SoCAB, and are captured in the Vista-LA database. Note: WW refers to wastewater treatment plants.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

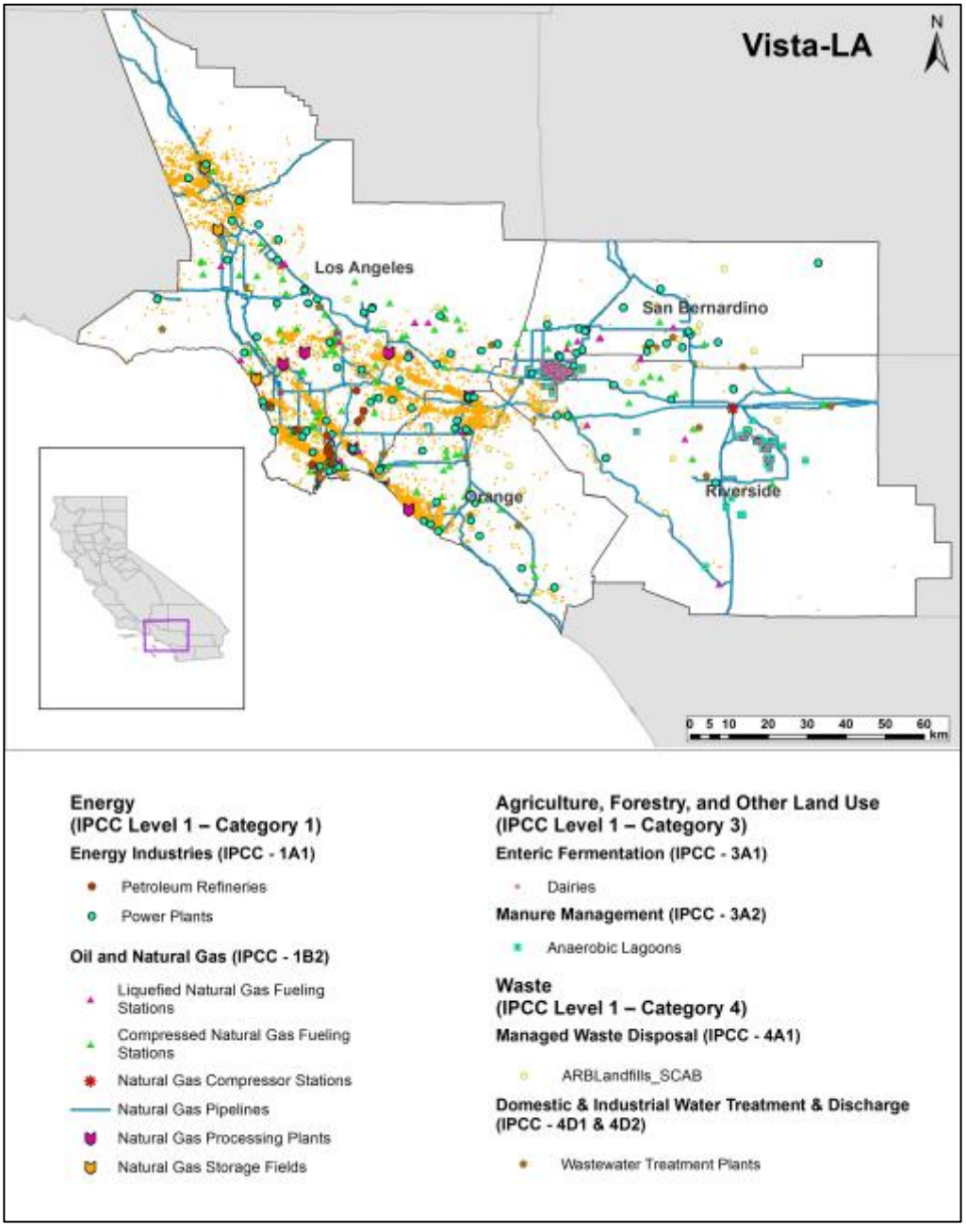

Figure 2: Overview of Vista-LA. Locations are shown for infrastructure with known or expected potential to emit CH_4 in the South Coast Air Basin (SoCAB). Vista layers are categorized by their corresponding IPCC Level 3 from the State of California GHG Inventory (see Table 1). Currently, compressed and liquefied natural gas fueling stations and natural gas storage fields are not included in the statewide GHG emissions inventory, but may be a significant source of fugitive CH_4 emissions in the SoCAB (see e.g., Conley et al., 2016; Hopkins et al., 2016b). Note: infrastructure in polygon form is difficult to distinguish from a static zoomed-out image; however, the majority of Vista layers can be viewed at the meter-level. Exceptions to this rule are for natural gas compressor stations and natural gas pipelines due to privacy and security concerns. Total: 33,353 features across 13 layers: 9 polygon layers; 3 point layers; 1 polyline layer.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

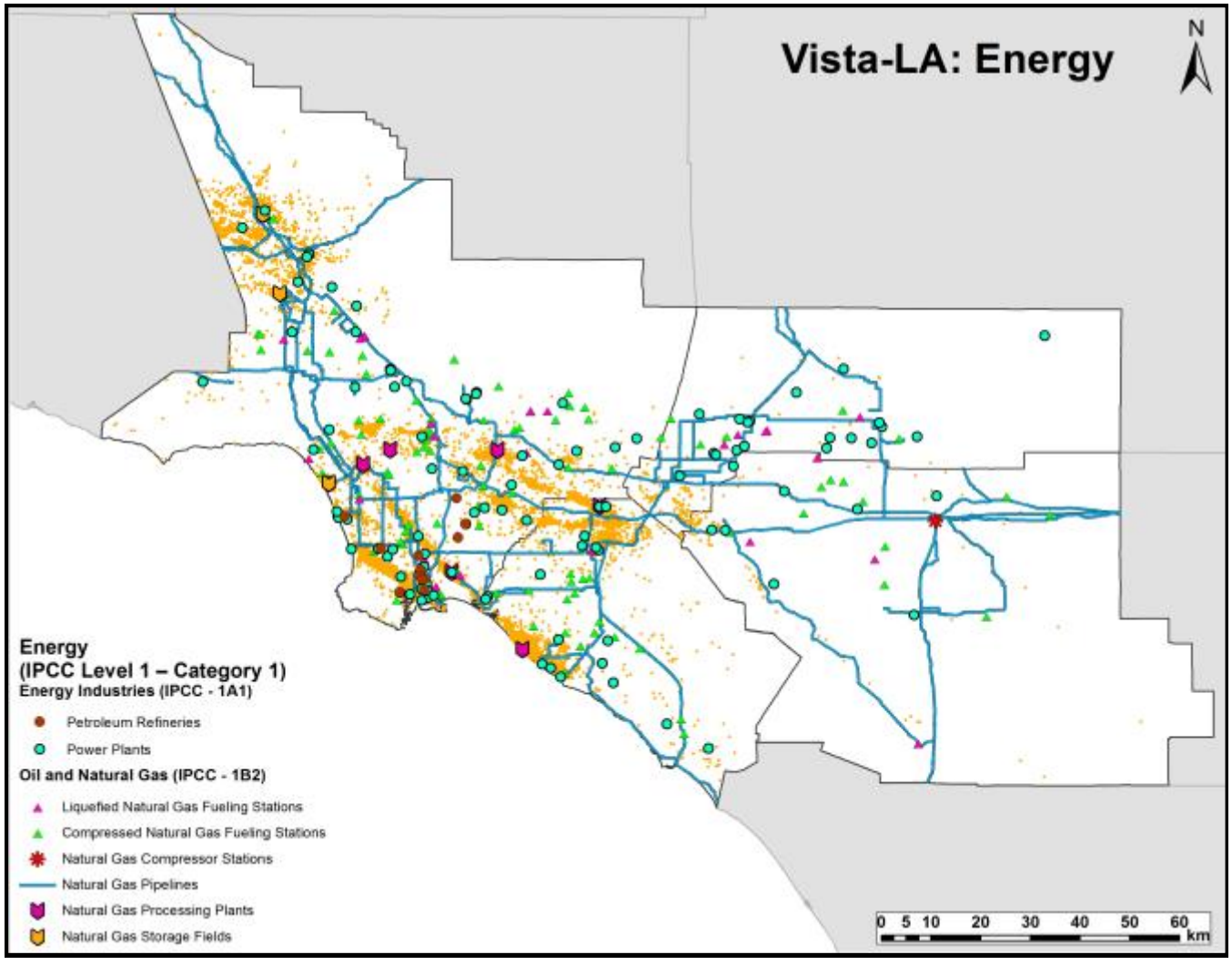

Figure 3: Energy (IPCC Level 1 - Category 1). Spatial distribution of infrastructure associated with the energy industry that emit CH_4 through fuel combustion (IPCC - 1A1) and/or fugitive emissions (IPCC - 1B2). Natural gas storage fields and compressed and liquefied natural gas fueling stations are currently not explicitly inventoried in the California GHG emissions inventory. Note: infrastructure in polygon form is difficult to distinguish from a static zoomed-out image; however, the majority of Vista layers can be viewed at the sub-meter level. Exceptions to this rule are for natural gas compressor stations and natural gas pipelines due to privacy and security concerns.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

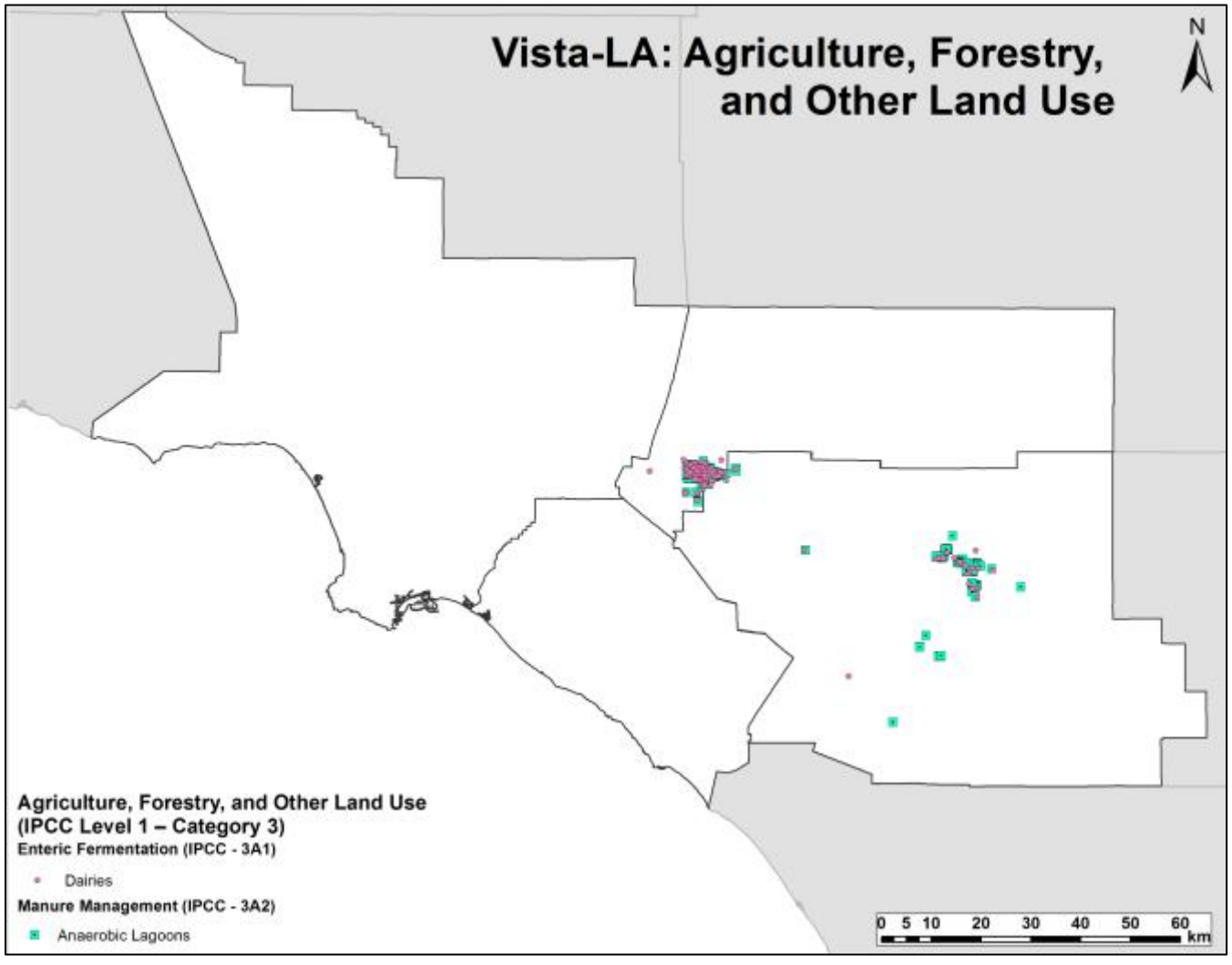

Figure 4: Agriculture, Forestry, and Other Land Use (IPCC Level 1 - Category 3). Spatial distribution of dairies and their respective manure lagoons in the South Coast Air Basin (SoCAB), encompassing enteric fermentation and manure management CH_4 sources. The largest clusters of dairies are located in the Chino and San Jacinto regions, with 110 dairies. San Jacinto Basin was home to 22 dairies and one cattle farm and Chino housed 56 dairies, 26 cattle farms and 5 other livestock farms in the year 2015. About 228 anaerobic lagoons were identified in these two clusters.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

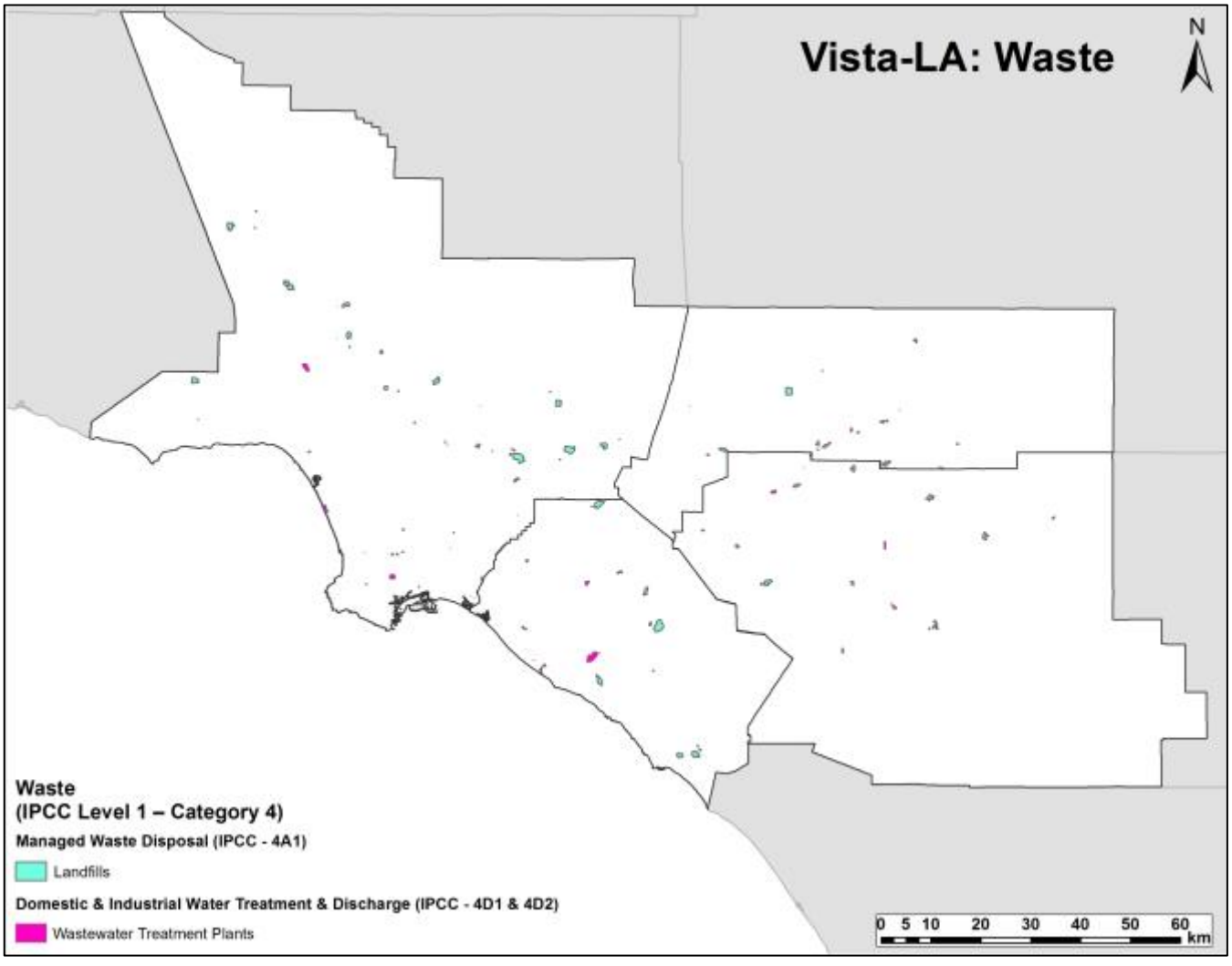

Figure 5: Waste (IPCC Level 1 - Category 4). Spatial distribution of 73 landfills and 26 wastewater treatment plants in the South Coast Air Basin (SoCAB).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

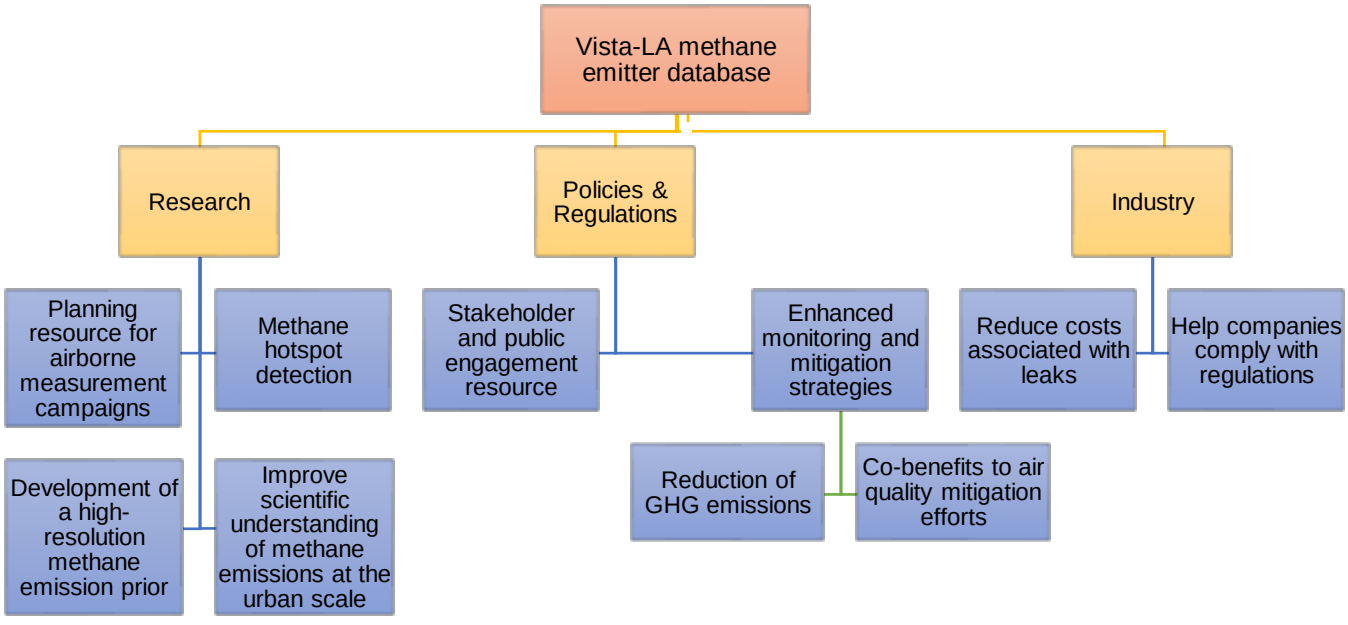

Figure 6: Overview of applications of the Vista-LA CH_4 emissions mapping tool. Vista-LA can provide numerous applications and benefits to research, policy, and industry.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

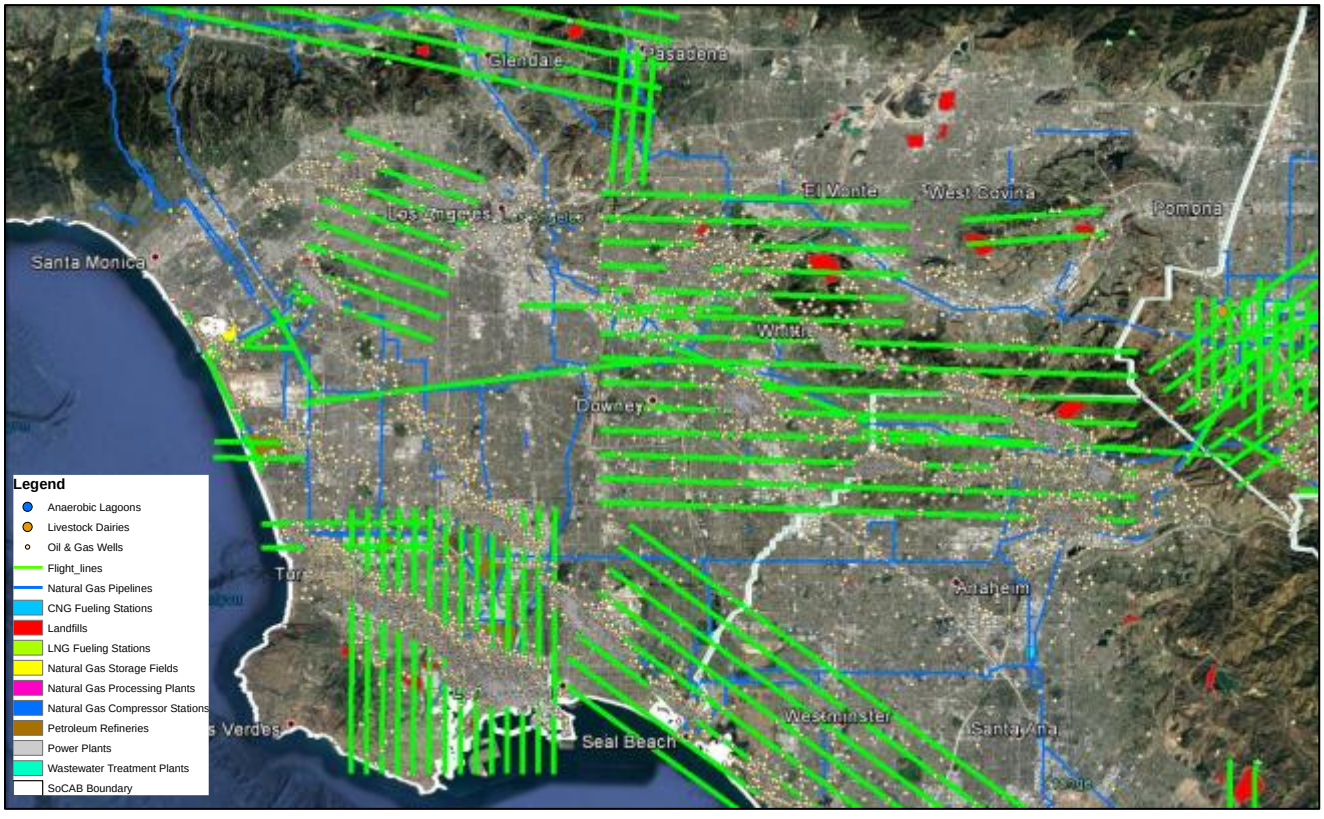

Figure 7: Vista-LA as a flight planning resource. The flight path of an airborne CH_4 remote sensing campaign to survey CH_4 point source emissions, shown as green lines, was optimized to include CH_4 emitting infrastructure for key sources shown in Vista-LA.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

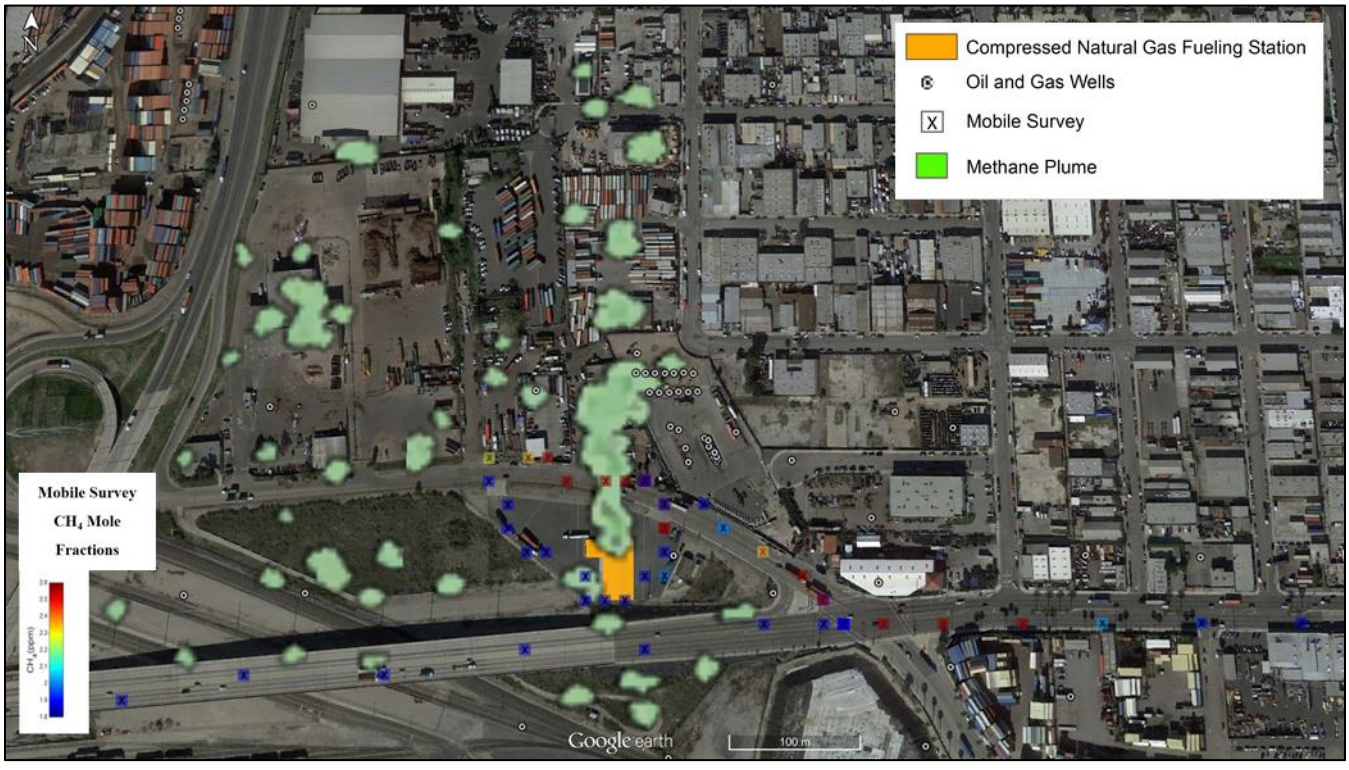

Figure 8: Application of Vista as a source attribution tool for CH_4 hotspots. Vista-LA can be used to detect and determine the source of CH_4 hotspots from on-road mobile surveys and aircraft measurements. In this example, Vista-LA layers are shown with CH_4 plumes from airborne imaging by the HyTES instrument and atmospheric CH_4 levels from a mobile survey in the Port of Long Beach, California. The co-location of the green colored CH_4 plume from HyTES (July 2014) and the red point observations of enhanced CH_4 levels from the mobile survey (June 2013) suggest that the CNG fueling station, shown as an orange polygon, is the source of observed CH_4 emissions.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

Tables

Table 1: Summary of Vista-LA layers. Vista-LA layers, representing CH_4 sources corresponding to IPCC Level 3, are shown organized by IPCC greenhouse gas emission reporting taxonomy. The source and year of the raw datasets, the maximum spatial coverage, number of features and format are also given for each Vista-LA layer.

CH_4 Sector	CH_4 Source Type	Vista-LA Layers (CH_4 Source)	Data Source (Year)	Raw Data Spatial Coverage (Data Source)	Vista-LA No. of Features	Vista-LA Data Format
IPCC Level 1	**IPCC Level 2**	**IPCC Level 3**				
1. Energy	**1A Fuel Combustion Activities**	**Energy Industries (IPCC - 1A1)**				
		Petroleum Refineries[a]	EIA (2016) SCAG (2005, 2012)	CONUS (EIA) California (SCAG 2005, SCAG 2012)	12	polygons / kmz
		Power Plants[a]	EIA (2016) SCAG (2005, 2012)	CONUS (EIA) California (SCAG 2005, SCAG 2012)	109	polygons / kmz
	1B Fugitive Emissions From Fuels	**Oil and Natural Gas (IPCC - 1B2)**				
		Compressed Natural Gas (CNG) Fueling Stations[b]	U.S. DOE AFDC (2017)	CONUS	109	polygons / kmz
		Liquefied Natural Gas (LNG) Fueling Stations[b]	U.S. DOE AFDC (2017)	CONUS	27	polygons / kmz
		Natural Gas Compressor Stations[c]	EPA FLIGHT Tool (2016)	CONUS	1[c]	polygons / kmz
		Natural Gas Pipelines[d]	CEC (2012)[d] EIA (2017)	California (CEC) CONUS (EIA)	N/A 111	N/A polylines / kmz
		Natural Gas Processing Plants	EIA (2014)	CONUS	6	polygons / kmz
		Natural Gas Storage Fields	DOGGR (2016) EIA (2016)	California (DOGGR) CONUS (EIA)	3	polygons / kmz
		Oil and Gas Wells	DOGGR (2016)	California	32,537	points / kmz
3. Agriculture, Forestry & Other Land Use	**3A Livestock**	**Enteric Fermentation (IPCC - 3A1)**				
		Dairies	RWQCB (2015)	Chino, Ontario, Riverside Areas	110	points / kmz
		Manure Management (IPCC - 3A2)				
		Anaerobic Lagoons	NASA JPL-Caltech\RWQCB (2015)	Chino, Ontario, Riverside Areas	228	points / kmz
4. Waste	**4A Solid Waste Disposal**	**Managed Waste Disposal (IPCC - 4A1)**				
		Landfills	CARB (2014) CalRecycle (2015) SCAG (2005, 2012)	California (CARB) California (CalRecycle) California (SCAG 2005, SCAG 2012)	73	polygons / kmz
	4D Wastewater Treatment & Discharge	**Domestic and Industrial Water Treatment & Discharge (IPCC - 4D1 and 4D2)**				
		Wastewater Treatment Plants	CARB (2016) SCAG (2005, 2012)	California (CARB) California (SCAG 2005, SCAG 2012)	26	polygons / kmz

[a]Sources may also include fugitive emissions that fall under IPCC source type 1B
[b]Source not currently included in the California Air Resources Board's 2010-2015 GHG Inventory.
[c]Only includes reporting facilities
[d]CEC pipeline data only available as a static representation in Figures 2 and 3.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

NOTE:
CalRecycle = California Department of Resources Recycling and Recovery
CARB = California Air Resources Board
CEC = California Energy Commission
CONUS = Contiguous United States Region
DOE = U.S. Department of Energy
DOGGR = California Department of Conservation, Division of Oil, Gas, and Geothermal Resources
EIA = U.S. Energy Information Administration
EPA FLIGHT Tool= U.S. Environmental Protection Agency Facility Level Information on GHG Tool
EPA FRS = U.S. Environmental Protection Agency Facility Registry Service
NPMS = National Pipeline Mapping System
RWQCB = California EPA Regional Water Quality Control Board, Santa Ana Region
SCAG = Southern California Association of Governments

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 19 July 2017

Open Access Earth System Science Data Discussions

Appendix

Figure A1: Comparison between GHG emissions inventory reporting structure for the State of California vs. the United States. Vista-LA complies with the State of California's GHG emissions inventory structure, but can be adapted to different regions, such as for the national GHG emissions inventory of the United States. Arrows indicate links between sector levels of the two GHG inventories.