Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2017-65-RC2, 2017

Open Access Earth System Science Data Discussions

Interactive comment on "Vista-LA: Mapping methane emitting infrastructure in the Los Angeles megacity" *by* Valerie Carranza et al.

Anonymous Referee #2

Received and published: 27 September 2017

General comments: The work presented here identifies potential sources of methane emission in South Coast Air Basin of California. I believe this is the first of its kind reported in terms of details on the locations of individual facilities and additional information attached to the spatial information. I think that the authors did a tremendous amount of work and the work deserves publication after addressing the comments.

My main comments are: 1) I appreciate detailed information on how to process the data (e.g., geo-referencing) but the paper is too long. The information is useful and I recommend the authors move some of them to the supplemental. 2) I would suggest that the authors clearly state why the spatial extent (i.e., polygon delineation) is useful. As commented below, I am not convinced that the polygon is better than a simple representation as a point. This work will be useful for inversions but I don't see enough

Printer-friendly version

Discussion paper

benefits from polygons, except for a few sectors such as dairy farms. Polygon work is associated with errors and makes it difficult for the authors to update the database in the future; the authors already used a significant amount of time in manual work. 3) I realized that this work is not about mapping “emissions” when I expect some results on emission mapping. So the authors need to clarify this at the beginning. The results section is somewhat boring because it just lists datasets, not something meaningful to be absorbed. 4) There are many figures, but the authors do not link the text with figures, only providing a minimum description about the figures.

Please address the detailed comments below.

Detailed comments:

L20: essentially SoCAB?

L33: globally or California or somewhere else?

L41: very vague. “at scales relevant to actionable emissions reduction efforts”, Which scale?

L48: any reference for the urban policy scale? Why the entire city? In SoCAB, there are so many small cities. Within those small cities, policies are different from meter scales? Activities vary significantly place to place, but I don’t believe such a large variation in urban policies, in particular for GHG control.

L52 – 54: This needs more explanations. Why do we need such a fine-scale emission map? Because the sources of urban CH4 are mostly co-located with those of CO2? Here, we need to be clear. For simple inventory purposes, we may need really fine-scale maps. But as a prior model in atmospheric inversion, do we really need meter-scale maps when the transport model cannot really resolve at such a fine scale? Please add more comments. Otherwise, it sounds like “we just developed a fine-scale product because it is better.”

L54 – 56: agreed.

L58: we know the location of a lot of sources for urban areas relative to rural areas. What we don't know is the activity levels and emissions.

L63: reference? I think at least Wunch et al., 2009 mentioned this limitation.

L72 – 73: What about landfill. In SoCAB, fossil-fuel sources cannot necessarily be "dominant" given the uncertainty of recent top-down studies in the region. Furthermore, depending on the definition of the hot spot, hot spots from fossil sources alone cannot be the dominant source in SoCAB. I feel that the authors touch lots of things related to urban CH4 emissions, but I am not sure that the authors are making effective arguments towards the goal of this new study. Literature review is not only scattered here and there but also somewhat inaccurate. I would recommend that the authors revise the introduction section with more clarity and accurate statements.

L89: Did the authors clearly lay out the shortcomings of current urban inventories. Which urban inventory? Here is the problem. The authors try to deal with the "general" urban CH4. Please focus on SoCAB. The authors don't even mention which inventory is available for SoCAB. There are several spatially explicit emission maps that include California and SoCAB. What are the shortcomings of those? The authors should spare the introduction section in describing weaknesses with current inventories (CARB, EDGAR, EPA spatial (ES&T, 2016), CALGEM, etc.) and the contribution this work can make over them.

L93: For those outside California, the use of LA Megacity is confusing since the authors introduced SoCAB. For inventory purposes, SoCAB should be preferable because CARB uses this air basin for regulatory purposes.

L100: What are potential source vs. facilities/infrastructure sources?

L193: I am curious about the portion of transportation in SoCAB. It would be useful to look at the transportation emission for SoCAB from Maasakkers et. al. 2016. Although the portion of transportation relative to the total may be small, transportation seems to

Printer-friendly version

Discussion paper

be important in this high-resolution maps, in particular in this highly urbanized region.

L212: I wonder if EIA data for refinery provide spatial information (i.e., lon/lat), probably only zip code. If only zip code, then this sentence is not really correct. EPA mandatory reporting (GHGRP) may provide exact locations.

L217 – 224: It should be useful to check with EPA's mandatory GHG reporting system.

L228: Please confirm that the EIA dataset has exact point locations. Also, cross-checking with Facility Registry Service (FRS) should be useful

L252: reason for claiming "The Vista-LA power plant dataset provides accurate location and extent data"?

L372: The authors may want to verify the compressor stations in SoCAB comparing with those from the California Energy Commission.

L396: Why did the authors use the EIA pipelines instead of those of CEC? Due to security concerns although the authors do not state it (only for NPMS).

L410: Don't need repeat this geo-referencing unless there was a need to manually geolocate facilities, e.g., EIA's NG processing facilities.

L460: I think this is not a proper citation. The authors should cite the DOGGR annual report instead of online GIS datasets for this purpose. I see some confusion between the dataset and the report here and there.

L484: grammar error.

L521 – L523: I think the majority of the emission sources described so fat are "point" sources. Because the authors identified the boundary of each facilities, it does not mean they are area sources. Area sources should be much broad and sometimes, unidentified sources. Although some sources may benefit from identifying the spatial extent, I don't see how useful the spatial extent for point-scale facilities (on a map) would be. One of the source sector that can really benefit from spatial delineation of

the facility boundary is the dairy section. Unfortunately, however, the authors do not provide this although they focus a lot on facilities that, in my opinion, do not require such information on spatial extent. I'd like to hear why it is important to figure out spatial extent for point-scale facilities. Is it helpful to perform atmospheric inversion for which a typical (even state-of-the-art) transport model can be run at the kilometer scale to capture underlying processes without significant errors. For airborne sampling planning, a simple point representation should be sufficient.

L538: The sentence is not clear.

L680: As commented above, please explain why this is a major advance. Please also remove "true" here. The problem here is that the process of making polygons require a tremendous amount of manual work, as stated in many places of this manuscript. I am not sure how the polygon features will be updated in the future, in larger applications for the state or other countries as the authors claim that theoretically this method can be applied in any regions; it can be done but the efficiency is in question. For some source sectors like dairy farms, I see the utility of polygons; they are large and sometimes located across multiple pixels on gridded maps. Just using GIS techniques for making polygons cannot be regarded as a scientific advancement that can be useful to the scientific community. Working on inverse problems for a long time, I don't see such a huge benefit from this polygon feature relative to the amount of the efforts (manual work) and potential and/or unidentified errors.

L694 – 696: Simple point-scale identification should be enough even for fugitive emission sources, in particular when gridded to kilometer or sub-kilometer scales (although sub-kilometer- scale simulations are not practical for most applications). Even for use with mobile surveys, a simple representation is enough when overlaid with Google (or other similar) maps.

L696 – 699: This is really strange: 1) where is the comparison result?; 2) Vista-LA has the result as "emissions"? The answer is no; and 3) when there is no emission

product from Vista, how can it be compared with other gridded emissions to reach this conclusion? Without ∼10 km aggregation for all three (CALGEM, Vista, and EPA; because CALGEM and EPA maps are in ∼ 10 km), the authors should not conclude like this.

L838 – 847: The authors are giving too much hope for automated feature extraction. The machine learning algorithm, in general, relies on cross-validation techniques and other simple statistics (e.g., mean absolute error) to validate the classification or regression results. If you look at the result of any cross validation (typically k-fold cross validation), it is not perfect; at a certain threshold point (e.g., 80% match between predictions and observations), you have to stop. This means it is associated with a lot of uncertainties due to limited training datasets, imperfect algorithms, etc. Without using complex techniques (e.g., bootstrapping, very computationally expensive) a typical machine learning algorithm does not provide uncertainty estimates (e.g., error for regressing fit). The authors need to state that there exist equal or even more uncertainties in the machine learning approach unless they can provide an example here. If you have raw data, using raw data may reduce uncertainty rather than using a machine learning technique.

Figure 5: I hope that the authors can explain what is the benefit of delineating spatial boundaries of landfills compared with the existing spatial inventory (e.g., Maasakkers et al., 2016) where the location is a simple point. The spatial resolution of gridded inventories (to be used for atmospheric inversions) need not be in meter scales because transport model cannot be run in meter scales. Then, I think a simple representation should be enough for clearly defined facilities like landfills. If you want to look at the boundary, viewing it over a Google (or similar) map should be fine. If the authors disagree, please explain why.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-65, 2017.