Dear authors of the manuscript,

This manuscript describes the updated version (v4.3.2) of the EDGAR emission dataset with a focus on CO2, CH4 and N2O. The importance of the EDGAR dataset for the science community is no need for discussion. The EDGAR dataset is a unique historical gridded emissions dataset that has been extensively used in the atmospheric chemistry, carbon cycle, climate, and many other relevant atmospheric research communities. Over the years, the EDGAR team has developed and maintained the global framework that offers spatially-explicit country level sectoral emissions prepared in a systematic and consistent manner across different compounds. As the authors claimed, the EDGAR dataset should help connecting science and policy and can play a critical role in future emission verification support systems which are currently planned and studied by the research community.

I would like to note that this manuscript has been significantly improved from the previous version of the EDGAR manuscript as published as Janssens-Maenhout et al. (2017) in ESSD Discussion. While Janssens-Maenhout et al. (2017) did a great job to put the considerable amount of information regarding the data used in the EDGAR dataset development together into a manuscript, I personally felt it had some weakness as a scientific contribution to ESSD. My main concern was the level of the information provided in Janssens-Maenhout et al. (2017) was not enough to maintain the traceability of the study and promote productive scientific discussions through the use of EDGAR dataset. This manuscript seems to be much closer to what ESSD and its audience would like to read (which Dr. Carlson and myself tried to clarify in the editorial comment published as Carlson and Oda (2018)). I found the authors particularly improved the presentation in the Supplementary Information. The text newly added was helpful to understand the emission calculation and modeling processes better.

In this review, before I recommend this manuscript for publication, I (as a referee as well as one of many EDGAR emission dataset users) would like to discuss several things that I believe the authors can further address in order to improve this manuscript as a scientific contribution to ESSD. The following part will discuss more in detail. I am looking forward to receiving your response. I hope my review comments will be useful.

Sincerely,
Tomohiro Oda (tomohiro.oda@nasa.gov)

1. Detailed comments/suggestions/discussions

Data tables

First, I would like to propose to improve the data tables. In many places (in the tables and also in the main text), data source references (where you get the data) and journal references (documents that explain the data) are mixed up. Those two can be summarized in tables in a more systematic way (see Table 4. https://www.earth-syst-sci-data.net/8/605/2016/essd-8-605-2016.pdf as an example). I also found that data tables do not indicate the edition of the data used. As we are all aware, the use of different editions of data can be a major source of differences (or error and uncertainty) in resulting

emission estimates. What I request here is important for traceability and let the data users to do a fair interpretation of the EDGAR dataset.

What's new and/or different from the previous and existing EDGAR datasets?

To improve the readability of this manuscript, I would like to propose to highlight the differences from previous version of the EDGAR (v4.2). This is an updated version of the EDGAR dataset that was constructed in a very similar manner as other pollutant estimates. I would imagine many readers of this manuscript also used previous and/or existing other versions of the EDGAR dataset. Given that, it would be very helpful for the readers (especially existing data users) to highlight more the differences from the dataset they have been familiar with. In this way, this manuscript should be able to help educating the readers efficiently. I found Section 4 in the Supplementary Information (especially, Fig. S3) very useful. I believe the authors should be able to improve other parts of the manuscript in the same way.

Evaluation of gridded maps

Similar comments go to the resulting gridded map section. Because of the nature of emission spatial modeling, it is challenging to present a meaningful uncertainty analysis for gridded estimates. As pointed out by the authors, I also agree that the gridded uncertainty estimates seen in the recent literature published are often large and hard to interpret. Also, it is questionable to me if those uncertainty estimates would inform transport modeling and inverse modeling studies in a meaningful way. But I do see the benefit of those analysis as an evaluation of gridded emission estimates as we could at least inform the data users on the differences we should expect in relative to other emission datasets (or previous version of them). Especially in this study, showing the differences from the previous version of EDGAR (as done for emission estimates in the Fig. S3) should summarize and highlight the changes the authors have made to the dataset. Since the changes we would see are combinations of the changes in emission estimates and spatial modeling, they might not be easy to interpret. However, we would like to see the impact of the changed the authors made (supposedly this is the improvement from the previous version). I would be curious to see the magnitude of the emissions differences in comparison to Figure 10a for example. Such comparison would loosely tell us the sensitivity of the urban emission estimates to the modeling approach. The comparison to other studies have done by previous studies (e.g. Maasakker et al. 2016; Gately and Hutyra, 2017; Oda et al. 2018). The authors could simply use those differences to discuss the resulting gridded maps.

Hot spot analysis

I am a little bit confused by this analysis. The EDGAR gridded emissions are based on spatial disaggregation of country-level sectoral emissions. Thus, the emissions at grid level are obtained by scaling the proxy data with the sectoral total emissions (as shown in formula (2)). It is clear that the urban emissions are not mechanistically modeling, unlike Gurney et al. (2018). Thus, the spatial patterns and changes in emissions are only explained by changes of the proxy data and the sectoral totals. The urban hot spot is a result of emission modeling and theoretically do not show the local emission drivers. I would imagine EDGAR does a better job than the carbon cycle FF dataset like CDIAC or ODIAC, but still it is disaggregated emissions. Also, the spatial resolution is pretty coarse when looked

at urban CO2 simulations studies. I felt the authors should have touched caveats the hot spot analysis has.

Another thing I would be curious to hear from the authors is the readiness level of the gridded emissions for policy application (i.e. the Paris Agreement, PA). In my opinion, many of the gridded emission estimates including this study (and also mine) still need a lot of development and improvement to support the PA in a scientific way we wish. As mentioned earlier, many of the gridded dataset do not have a capability of showing subnational emission reductions in their emission field (no local drivers).

Be more quantitative

At least we emission data developers should do is to describe the emission modeling procedure and the data used in details. As I acknowledged earlier, the authors have greatly improved the presentation especially in the Supplementary Information. However, I would like to request the authors to be more quantitative. For example, there must have been a QA/QC process in the use of the CARMA power plant database. Careful data users already knew the power plant database has been used in the EDGAR dataset development as it has been described in the EDGAR website. What we the audience of this manuscript would like to learn from this manuscript is how differently the power plant database was used compared to other studies and the differences we expect to see. The authors might be able to tell certain thresholds for selecting the power plant entries in the CARMA. For example, as I don't believe the total emissions from CARMA and the EDGAR total sectoral emissions matches well (maybe they do), I believe there were some ad-hoc adjustments to CARMA (or EDGAR) in mapping power plant emissions. Such details should allow the data users to decide if the EDGAR dataset is a right choice to answer their research questions.

2. Line by line comments

P1, L20: "…disaggregated to IPCC-relevant source-sector levels". This does not seem to sound correct. Emissions are calculated sectoral basis and then spatially disaggregated.

P1, L25: fully traceable. I feel the authors needs a little bit of extra effort to support this.

P1, L25: IPCC-based methodology. This is confusing. IPCC (1996) is a guideline. They don't precisely define the emission calculation. Maybe IPCC-compliant would be more appropriate?

P1, L26-: Maybe the use of the term "short_cycle" would be helpful for the EDGAR users?

P3, L29: Policy impacts. I agree at a large spatial scale (e.g. country and regions, w sectors), but do you think the disaggregated gridded emissions can be used to assess the policy impact? To support this, you'd have to show that your disaggregated emissions are reasonably indicating the local change. Gridded emissions in EDGAR are disaggregated rather than mechanistically calculated. If we see a X% emission change over a city A, that changes are because of the change in the total sectoral emissions and/or changes in the proxy. I am curious to hear how the gridded emissions can be used for informing

policy. As a developer of disaggregated gridded emissions, I feel I need to acknowledge the limitation of the use of disaggregated emissions (see Oda et al. 2019).

P3, L35. Footnote 5. In my opinion, this is worth mentioning in the main text. For example, the errors in the emission ratio analysis could be dominated by the error in the air pollutants side because of this.

P4, L5: and emission disaggregation?

P5, L6: How did you deal with countries like USSR and Germany over the EDGAR period (e.g. activity data, emission estimates and spatial proxy)? Which country identifier data did the authors use?

P5, L14: So the emission seasonality is not country specific in EDGAR. Is that correct?

P5, L15: CARMA indicates only four years with their two versions of the database. Did you use the emission estimates as they are? If so, how did you match with the CARMA country total CO2 emissions with the EDGAR sectoral total? How did you use those information for intervening years? What was the quality control?

P6, L2: Just to be clear, do the authors recommend those three BB products to supplement the EDGAR emissions because of some reasons and/or compatibility?

L6, L31: I assume the authors meant to say EDGAR wants to avoid model-based estimates, but try to stick to the data based (or EDGAR way?) emissions.

P7, L5: Table S4a only provides the references for data sources. There is no narratives. So it is unclear that how the temporal profiles were constructed and what they are representing.

P7, L11: Huang et al. (2018). I could not find this reference.

P7, L11: I don't understand this. The temporal profiles presented in Andres et al. (2011) were based on a different approach than this study.

P7, L15: Not just global models, the data users are using the EDGAR dataset for regional and even local models, too.

P7, L18: Linear -> Line (?)

P7, L21: Does the area average change over the time (1970-2006, for example)? Probably no?

P7, L25: Where it is *reported*. Probably it would be better to say like "where it is likely located" as the EDGAR is based on the emission disaggregation. Point source emissions could be allocated to the geographical locations reported by databases such as CARMA. But there is no linkage between the

emission at the country, sectoral level and the power plant locations. For line and area sources, the locations are estimated (or modeled) rather than specified.

P7, L27: Just to make sure… Oda et al. (2018) also uses CARMA for point sources.

P7, L28: I think so, too! I'd suggest to add "sectoral" before "emissions". Those two emissions are both based on CDIAC country totals that are calculated based on fuel use. Thus, the emission disaggregation problem in CDIAC and ODIAC is fundamentally different than that of EDGAR.

P7, P30: I agree with the authors that the uncertainty analysis for the proxy data themselves do not provide what we want as an emission uncertainty that are useful for inverse modeling and/or data assimilation. But I do think that such sensitivity test can be done w/o transport models as demonstrated by previous studies. Also, I imagine what the uncertainty analysis that the authors have in their minds will be highly depending on the models and observations used/assumed. Aside from the uncertainty analysis, an important missing component in this manuscript is an evaluation of gridded maps that should demonstrate the improvement and/or differences/changes in emission spatial representations from existing emission datasets and previous EDGAR datasets. In my opinion, such evaluation should be done if this manuscript is going to be a scientific paper, rather than a tech report.

P7, P31: The authors should introduce the CHE project.

P8, L5: To me, this sentence does not seem to fit here as the author returned to the national level uncertainty analysis in the next sentence.

P8, L11: I am confused. The eq. (4) includes all three gases, while Table 2 (not Table 3, 4, and 5) shows uncertainty estimates for three gases.

P9, L3: So those two uncertainty estimates are not compatible.

P9, L12: Note Andres et al. (2016) limited the result by saying CASE FOR CDIAC.

P9, L17: I feel probably this needs to be elaborated a little bit. Firs of all, it is unclear that what the authors meant by saying the complete uncertainty and how it is achieved. Again, given the nature of the gridded maps and the lack of the evaluation data, it Is difficult to validate the gridded maps (e.g. Andres et al. 2016; Oda et al. 2018). It is important to combine top-down and bottom approaches and thus good error estimations are necessary. But as mentioned earlier, what ESSD and the audience of ESSD expect this manuscript to present is not necessarily the covariance matrix for data assimilation, but a reasonable sense of how much we could trust the gridded maps.

P9, L28: TD for scientists. Maybe TD was just for scientists, but I believe it is required for policy making as we want to assure the accuracy of bottom-up emission estimates. Emission inventories are subject to

systematic biases. That is a huge problem under the Paris Agreement as the emission estimates should be consistent with what we emitted to the atmosphere (Oda et al. 2019).

P9, L35: Studies listed in Table S5 are great examples of how atmospheric modeling can inform us about the errors and biases in emission estimates. But I would like to point out that none of them seems to provide the "complete uncertainty". None of them approached to the grid level uncertainty (excepting the power plants), either. This sentence reads the use of atmospheric modeling is a perfect solution, but the authors should acknowledge that there are limitations, too. The sensitivity of the emission spatial representations should depend on the spatial resolution of the models and observation systems assumed on the top of the errors in the transport modeling (meteorology, chemistry, etc.).

P10, L1: I would imagine CO2 is a very different case compared to NOx, SO2, and CH4.

P10, L21: Elvidge et al. (2009) is based on the DMSP lights, but what the download link provides is a different data product. Elvidge et al. (2009) does not have time series if I remember correctly. Also, Elvidge et al. (2009) does not cover all the countries with gas flare emissions. How did you manage the gas flare emissions w/o nightlight location estimates? Need to explain.

P12, L14: Verify. Maybe it would be better to say like "detecting the biases in emission inventories"?

P15, L7-: As mentioned earlier, the authors should provide an evaluation of the gridded maps. This is the results of the multiplication of the total sectoral emissions and normalized proxy data. From previous studies, we expect these gridded maps should show different emission patterns from other studies. For example, Gately and Hutyra (2017) has shown the large discrepancy between their spatially-explicit urban emission estimates and EDGAR. I do not request the authors to validate these gridded maps, but help the audience of the ESSD to interpret these gridded maps in order not to over-interpret these maps beyond their limitations. It could be done by acknowledging the emission spatial differences reported by previous studies.

P16, L2: What was the QA/QC done for power plant data especially for geolocation? Are they verified? What was the threshold for emission intensity?

P18, L2: I (I am sure Dr. Carlson, too) would like this manuscript to be one of the good examples of ESSD articles!

P18, P22: This is not about EDGAR, but atmospheric models. I feel this is misplaced here (maybe I might be wrong…). I would add this is what we expect to atmospheric inversions to do, but it is not happening in the way we hope quite yet.

P19, L7: Do the authors limit to the interpretation of data from satellites, but not other platforms? The EDGAR team has experienced several success with satellite measurements for other compounds. But

data form other platforms such as ground-based sites (e.g. radiocarbon measurements) and aircrafts should not be excluded (I assume the authors did not mean to).

P19, L11: Regional specificity, too? This had to be sacrificed in order to achieve a global systematic framework. I think this is consistent with the first bullet point at L15.

P19, L17: top-down -> disaggregation or downscaling? (just to avoid a potential confusion with atmospheric top-down)

P19, P20: Would you want to add a caution that gap-fill could sacrifice the consistency in emissions that EDGAR maintained? For CO2, for example, we don't want to do a mosaic emission approach as done for air pollutants as the consistency between the global and regional budget is important.

P19, L31: I assume this is a general statement for the EDGAR dataset as differences due to the technology types do not make significant differences according to the authors.

P19, L36: Given the fact that the authors highlight the importance of the point source information, I feel the description of the point source data and modeling need to be improved a bit (see relevant comments listed earlier).

P20, L9: I am confused here. What the authors described here does make sense, but this would turn EDGAR into what we can't define as an inventory (more like a model). Is that the direction of the EDGAR development?

P20, L31: Similar comment to the above. The approach is valid, but what about the large uncertainty associated with the CO2 emission estimates? This is CO2, not SO2. Would you be happy to include those estimates as a part of EDGAR dataset? This does not seem to be very consistent with the EDGAR's basic principle mentioned earlier.

P21, L1-: Data availability section needs to be improved.

P40, Table 3. Note the latest version of the ODIAC data product (ODIAC2018) has been available. I suspect that the authors might be comparing the CO2 estimates based on the different year edition of the statistical data which do not allow us to do a fair comparison.

P48, Figure 10a and 10b: The quality of the figure needs to be improved. The numbers of the main panel are hard to read even enlarged.

P49: Figure 11a and 11b: I'd suggest to do the v4.3.2 minus v4. It is not going to be a clean comparison. But it should demonstrate the improvement and/or changes over different versions of EDGAR (which is the main claim of this manuscript). The same comment goes to CH4 and N2O maps, too.

Supplement Information

I could not find the information about technology specific proxy data. Are those only used for air pollutants?

P2, L2: How did the authors deal with countries such as Germany and Former Soviet Union in the EDGAR calculation and mapping?

P14, P4: Did the authors linearly interpolate the population in time?

P14, L13: How the geolocation errors were managed in the QA/QC process? Did the authors consider commission/decommission of plants?

P15, L4: NOAA gas flaring nightlight data used in Oda et al. (2018) only convers 60+ countries and thus the emissions needed to be distributed as a part of area sources in ODIAC. How was it done in EDGAR?

P15, L12: Are this weighting factors listed somewhere? Did you use population to get an additional spatial weighting function?

P15, L24: km 101 until the last 101 km?

P15, L32: Is Friedrich and Reis (2004) accurate reference for this? Friedrich and Reis (2004) was a compilation of air pollutants studies. So I assume the authors used a temporal profile of an air pollutant(s) as a proxy. I also could not find residential emission temporal profile plot in Friedrich and Reis (2004).

P15, L33: I think the authors should at least explain how the temporal profiles were constructed. Are they averages for multiple year data (if so, error bar?) or single year? Where did the underlying data come from (only from Europe)?

P21: L2: CARMA base years should be 2000 & 2006 (earlier version) or 2004 & 2009 (for v3.0). No?

P21, L12: This is a bit stretch… the authors should carefully read Nassar et al. (2017). I am curious to ask the EDGAR team is willing to incorporate the satellite-based power plant estimates in the emission dataset. It seems to be ok for SO2 (large errors), but I am not sure about CO2.

P21, Figure S2. Is it long term averages of European sectoral temporal profiles? What do we see here?

P22, Fig S2b: The authors need to describe how each one of them was constructed and explain the differences in terms of modeling approach and data used.

P22, L6: We need these for gridded maps.

P23, L15: Were the emissions CARMA indicated use as they were? The linkage between CARMA and the EDGAR sectoral total emission as the two should not match perfectly w/o any adjustment.

References:

- Andres R. J., Gregg J. S., Losey, L., Marland, G., and Boden, T. A.: Monthly, global emissions of carbon dioxide from fossil fuel consumption, Tellus B, 63, 309-0327, https://doi.org/10.1111/j.1600-0889.2011.00530.x, 2011.
- Gately, C. K., & Hutyra, L. R. (2017). Large uncertainties in urban-scale carbon emissions. Journal of Geophysical Research: Atmospheres, 122, 11,242–11,260. https://doi.org/10.1002/2017JD027359
- Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U., and Petrescu, A. M. R.: EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970–2012, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-79, 2017.
- Maasakkers et al. Gridded National Inventory of U.S. Methane Emissions, *Environ. Sci. Technol.*, 2016, *50* (23), pp 13123–13133, DOI: 10.1021/acs.est.6b02878
- Oda, T. and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543-556, https://doi.org/10.5194/acp-11-543-2011, 2011.
- Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO_2, version 2016 (ODIAC2016): a global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth Syst. Sci. Data, 10, 87-107, https://doi.org/10.5194/essd-10-87-2018, 2018.
- Oda, T. et al.: Errors and uncertainties in a gridded carbon dioxide emissions inventory, accepted for *Mitigation and Adaptation Strategies for Global Change*