Response to Anonymous Referee #1

In this paper, the authors document the methodology used to develop a new historical (1750-2014) short-lived species emissions data set for use by global chemistry-climate models for the upcoming Coupled Model Intercomparison Phase 6 (CMIP6). The paper provides detailed information on input data sets (e.g., emissions factors, activity data, population) and steps applied to generate the new emissions trends in Community Emissions Database System. The authors also compare this new dataset with existing emission inventories to place CEDS emissions in the context of existing data sets.

Emissions inventories provide crucial input data for global chemistry-climate models to simulate the spatial and temporal distributions of short-lived pollutants many of which are also climate forcers. Although gridded emissions inventories existed prior to the inventory of Lamarque et al (2010), there was a lack of consistency in the use of these products by different global modeling groups participating in multi-model intercomparisons. Development of a global gridded emissions inventory for not just the present conditions but going back in time is a major undertaking. I am sure the global chemistry- climate modeling community would be very appreciative of the service provided by the authors in not only updating the previous extensively used inventory (Lamarque et al., 2010), but also creating a consistent, transparent, and trackable process that can hopefully be sustained going forward. The paper is generally well-written and is appropriate for GMD.

My main comment on the paper is that the authors do not provide any comparisons of the spatial distribution of the gridded emissions against existing gridded products. A panel plot with maps of present day (e.g., 2010) species emissions should be included in the main text. Some discussion of how they compare with the spatial distributions in existing inventories would be helpful.

Thank you for your kind comments. A new section has been added to Section 3, titled "Gridded Emissions" which includes a panel figure of gridded emission of CEDS total emissions estimates for all 9 emission species in 2010.

Additionally a discussion of the differences between CEDS grids and Lamarqe et al (2010) grids for 1850 and 2000 have been added to CMIP5 Comparison section. A figure showing side by side gridded maps of the differences for EM for 1850 and 2000 has been added to the main text. Difference maps for all other emissions species for both 1850 and 2000 have been included in the supplemental figures document.

Below are some specific comments to help improve the paper:

Specific comments:

P2: Suggest arranging the discussion of existing emission inventories chronologically. EDGAR has a long history of developing emissions data set and was available much before the inventory of Lamarque et al (2010) (referred to as L2010 hereafter). In fact, the EDGAR informs the L2010 and the work described in this paper.

Thank you for this comment. Paragraphs discussing EDGAR and Lamarque 2010 data (L49 - 76) have been swapped so that this section discusses EDGAR and Gains, followed by Lamarque 2010 data.

L84: What is the time period for the historical data?

We have added a section in the appendix, A2.2 that explains versions of CEDS releases. The following text appears in the appendix of the manuscript:

> There have been several releases of the CEDS gridded data. The underlying emissions by country, sector and fuel have been identical in all of these releases, as are total emissions by country and gridding sector (with the exception of small changes in 1850 emissions noted below).
>
> **v2016-05-20**: Pre-industrial 1750-1850 data release
>
> **v2016-06-18**: 1851 – 2014 data
>
> **v2016-06-18-sectorDim**: Re-release of both preindustrial and 1851 – 2014 in a new netCDF format with sectors as an additional dimension in the data variable. This reformatting was necessary due to a limitation that was discovered within the ESGF system summer 2016. The reformatted data were released early Fall 2016
>
> **2017-05-18**: Re-release of entire dataset in order to correct two gridding errors discovered by users. 1) Inconsistent emission allocation to spatial grids within countries that resulted in incorrect spatial allocations and some large discontinuities in the gridded data. These issues were particularly apparent in spatially large countries such as the USA and China. 2) Minor inconsistencies in seasonal allocation, resulting largely in emissions that were too high in February. Total annual emissions within each country were not impacted by either of these issues.
>
> Emissions are also fully consistent across 1850 in this release. There were small discontinuities in 1850 between the CEDS CMIP6 preindustrial release (v2016-06-18) and the later full CEDS release (v2016-07-26) due to updates in the data system. These differences are 0.5% for all species (except NMVOC which reaches 1.5%). In absolute terms these differences are very small (relative to, for example, open biomass burning emissions) and will not have a significant impact on simulation results.
>
> A link to further examination of these issues, including comparison maps and time series comparisons, can be found at the project web site (globalchange.umd.edu/CEDS).

L85-L87: The chemical formulas for these species are first used on L55 without defining them. Suggest moving the full names closer to where the formulas are used for the first time in the text.

Thank you for pointing this out, Appropriate explanations of chemical formulas were moved to L55 where they first appear.

L102-103: At what point is the seasonality added?

The following text was changed to specify that seasonality is added to gridded data in the final step: "…6) gridded emissions with monthly seasonality are produced from aggregate estimates using proxy data.." There is additional explanation of this process in section 2.6 Gridded Emissions.

Figure 1, captions: replace "produces" with "products" on P4.

Thank you. This change was made

L116: What does "energy balance statistics" mean here?

We are refereeing to detailed IEA energy statistics. The following change has been added for clarification:

"energy statistics"

L117: A quick search on google tells me that there are 196 countries in the world (not considering Taiwan separately would bring the count down to 195). Am I missing some- thing? Could the authors please provide a color coded map of countries considered in the work?

We clarify in the revised manuscript that we consider a number of regions whose exact status might not be clear. The most definitive categorization is given by the UN, and we use UN population data as the basis of our current "country" disaggregation. As noted in the manuscript, we are using the term "country" regardless of the exact status of any particular entity.

The supplemental information includes a csv file which contains a list of all the "countries" used here along with their common name, ISO code, and mapping to countries and regions from other data sets such as the IEA energy data. Many of these "countries" would not be visible on a global color coded map as they are small islands and territories.

The following change has been added to the manuscript for clarification:

"CEDS estimates emissions for 221 **regions** (and a global region for international shipping and aircraft), ... **"Regions" refers to countries, regions, territories, or islands and are listed, along with mapping to regions and ISO codes in the supplemental files; they will henceforth be referred to as "countries"."**

L123: A reference is needed for the IPCC guidelines and Nomenclature for reporting document.

This reference has been added. Thank you.

L137: What kind of "additional effort" would be needed? Please elaborate.

These efforts are briefly described in the future work section of the manuscript. The following text has been added to the manuscript to clarify:

Greater disaggregation for these sectors would improve these estimates, but will require additional effort**, described in Sect.5 Limitations and Future work.**

L140-141: Please elaborate on the "confidentiality issues." As I understand, sector level emissions are provided in the gridded files, so I am confused by this statement.

The text has been clarified on this point. We note that other global emissions data providers, such as EDGAR, are subject to a similar limitation in terms of releasing emissions data at the level of fuel, sector, and country.

The core outputs of the CEDS system are country-level emissions aggregated to the CEDS sector level. Emissions by fuel and **detailed CEDS** sector are also available within the system for analysis, although these are not released because this could violate the terms of our use of the IEA energy statistics. (This is the same reason other global inventory data, such as EDGAR, also do not release data by sector, country, and fuel). Emissions are further aggregated and processed to provide gridded emissions data with monthly seasonality, detailed in Sect. …

P4L143: What is meant by "emission control degradation?"

"Emission control degredation" refers to certain emission control technologies that may become less effective overtime, for example an old catalytic converter in an old car may be less effective than a new catalytic in an old car. The following text has been changed in the manuscript for clarification:

For example, CEDS does not include a representation of vehicle fleet turnover and emission control degradation **(*e.g.* the effectiveness of catalytic converters over time)** or multiple fuel combustion technologies that are included in more detailed inventories.

L161-162: Does the population data used to disaggregate energy data for CEDS countries change with time? Describe any assumptions made. Also, please provide a reference for the population data set used.

As described in section 2.2.2 and further the supplement, a full time series over 1750-2014 of population estimates for all CEDS "countries" is developed by merging several data sources.

However, this explanation incorrectly describes the methodology we used for disaggregating IEA aggregate regions. Aggregate IEA data were disaggregated using CDIAC data, not population data. The following text now appears in the manuscript:

Data for a number of small countries provided by IEA only at an aggregate level, such as "Other Africa" and "Other Asia", are disaggregated to CEDS countries using historical CO_2 emissions data from the Carbon Dioxide Information Analysis Center (CDIAC) (Andres et al., 2012; Boden et al., 1995).

L170: Is the BP data freely available?

Yes. BP data is publically available online. The following text has been changed for clarification

IEA energy statistics were extended to 2014 using BP Statistical Review of World Energy (BP,

2015), **which is freely available online** and provides annual updates of country energy

L180: Please provide a reference for the MEIC inventory.

The MEIC citation has been added. Thank you.

L195-196: This statement conveys ambiguity in the use of population data for generating emission trends. Please clarify.

This text has been changed to:

"While non-combustion emissions use population as an "activity driver" in calculations, emissions trends are determined by a combination of EDGAR and country level inventories. Final emissions estimates, therefore, reflect recent emissions inventories where these are available, rather than population trends."

L214: I feel that this equation can be moved up near the beginning of section 2.1 as it describes clearly how activity data and emissions factors are combined to obtain emissions.

Thank you for your suggestion. We've kept the formal equation where it is to avoid restructuring a too much text, but added the simple "emissions = driver x emission factor" phrase to section 2.1.

P9L310-311: To clarify, is the "value" of the scaling factor limited to greater than 1/100 and less than 100? If so, please rephrase the sentence.

Yes, that is correct. Thank you for this comment. The following change has been made in the manuscript:

Calculated scaling factors are **limited to values between 1/100 and 100. Scaling factors outside this range** may result from…

L344: Replace does with do.

This change has been made, thank you.

L396-399: Specify that the discussion in paragraph is pertinent to soil NH3 and NOx emissions.

"Emissions from mineral and manure emissions…" has been changed to "NH_3 and NO_X emissions from mineral and manure are often…"

L413: Which CEDS sectors are the authors referring to here? The 55 working sectors, the 16 intermediate sector or the 9 aggregated sectors?

By "in most sectors" we mean, for most of the data, which could mean any sector aggregation. To avoid confusion "Proxy data used for gridding in most CEDS sectors are primarily gridded emissions …" has been changed to "Proxy data used for gridding are primarily gridded emissions from…" Thank you for this comment

L414: Please clarify what sector (of the 9 aggregated) is the flaring emission relevant for.

Flaring emissions are one of the intermediate gridding sectors within the energy sector for final gridding. Final gridding sectors have been added to table 6 for clarification.

P15, Figure 2: The figure caption says that aviation emissions are not included but the color bar shows "Air" as an option. Please clarify

Even though "Air" was in the legend, the figure did not show any emissions, as they were not included in the graph. The Air sector has been removed from the legend in these figures for clarity.

P16, Figure 3: The label "International" to describe international aviation and shipping is misleading. Suggest replacing it with Air_ship (or some combination of air and ship) so that it is clear the authors are referring to combined aviation and shipping emissions.

The region "International" has been changed to "International Air-Ship" throughout the paper.

L451: Please clarify "anthropogenic emissions" from which inventory are being referred to here. Are the CEDS anthropogenic emissions 20-30% of the total global emissions for BC, OC, NMVOC, and CO?

The following paragraph has been rearranged to read:

"In 1850, the earliest year in which most existing data sets provide estimates, anthropogenic emissions are dominated by residential sector cooking and heating and therefore products of incomplete combustion for BC, OC, CO, and NMVOC. In 1850, anthropogenic emissions (sectors included in this inventory), make up approximately 20 – 30% of total global emissions (which also include grassland and forest burning, estimated by Lamarque et al. (2010)) for BC, OC, NMVOC, and CO but only 3% of global NO_X emissions."

L459: replace "in 1950" with "post 1950"

This change has been made.

L460: Insert a reference to Figure 3 at the end of the sentence.

This change has been made.

L480: Can the authors postulate any specific reasons for the flat residential biomass emissions in latin America despite a growing rural population, and flatter China emissions than rural population after 1990?

The following text was added:

"While rural population in China continually grows, residential biomass use flattens in 1990 as both the share of urban population in China increases and rural residential per capita biomass use

decreases."

L496: replace 'species of emissions" with "species emissions"

This change has been made

L506: Please refer to a figure to support the statement "Global CO emissions flatten".

The following text has been added to the sentence: "… shown in Figure 2 and in more detail in the Supplemental figures and tables."

L514-515: Please clarify the sentence: "offset by international shipping emissions grow then decrease. . ."

Thank you for pointing out this poorly phrased sentence. This sentence has been changed to:

"Global NO_X emissions rise then flatten around 2008. The growth in industrial emissions after 2000 is offset in 2007 by the decrease in international shipping emissions, while global emissions in other sectors stay flat."

L516: Is it possible that the decline in North American NOx emissions is driven by the decreases in US NOx emissions in response to the NOx control regulations implemented in the US (NOx SIP call). This is fairly well documented and literature should be referenced here as this lends confidence to the trends in NOx emissions derived from CEDS.

Thank you. A reference was added.

L521: A reference is needed for "more stringent emission standards for power plants"

Thank you. A reference was added.

L554-L559: Please refer to a specific figure in the Figures and Tables Supplement for this comparison (e.g., Figure S40). I would also suggest doing the same for other species in the paragraphs below as it is cumbersome to sift through the many plots.

Thanks. Specific references to supplemental figures have been added strategically in the results sections of the manuscript.

Section 3.4: It would be very helpful to have a table with published level of uncertainties in emissions for each species (CO2, SO2, CO, NOx, NMVOC, BC, OC) and specific sectors. Much of the information is contained in this section and can be pulled into a summary table that will come in handy when uncertainties in CEDS emissions are determined.

While we agree this would be useful, we will refer the reader to the literature summary in IPCC AR5 for now (we're not aware of a more up to date general summary) as more significant effort to collect uncertainty estimates will need to wait for future work.

L622: This sentence can be rephrased to " Emissions uncertainties for CO, NOx, and NMVOCs

typically lie between those of carbonaceous aerosols. . .."

This change has been made. Thanks.

L626-627: Hassler et al (2016) should also be cited here.

This citation has been added.

L637-639: Paulot et al can be cited here as an example of detailed modeling of agricultural NH3 emissions.

This citation has been added.

Section 5. The ultimate test of an emission inventory is comparison of species concentrations simulated by a model driven by an inventory against observations (e.g., Parrish et al., 2014; Hassler et al., 2016). If the model is able to capture the distribution and trends then the said inventory is considered to represent the real conditions well. I think a case could be made for better coordination between modelers and emission inventory developers so that a two-way interaction can help improve both models and emissions inventories.

We agree in general, although note that models are not necessary in all cases: Hassler et al., for example, use observations directly. There are many complications of course: models are imperfect or incomplete, observations are often not available at the same scale as model results, and inventories are often not available at scale of observations (e.g. Wang et al. doi/10.1073/pnas.1318763111).

We have added a comment on this in the future work section.

An outline of long-term plans for the CEDS database is needed in the summary section to build confidence in its sustainability. Modelers would like to know if they can rely on the CEDS system working even after CMIP6. What are the plans for maintenance of the back-end software, frequency of updates to the input data and for maintaining funding for CEDS?

Thanks. The following text has been added to the manuscript detailing future plans for the System and community engagement.

"The CEDS data system, including R code and all input data other than the IEA energy balances, is being prepared for public release in fall 2017 through the gitHub collaboration website. This will facilitate community comment, and direct contributions to improving these emissions data. The next data release is planned for Fall 2017, which will extend the time series to 2016 and correct, to the extent possible, any known issues with the dataset. We aim to continue annual updates in subsequent years."

References:

Paulot, F., D. J. Jacob, R. W. Pinder, J. O. Bash, K. Travis, and D. K. Henze (2014), Ammonia

emissions in the United States, European Union, and China derived by high- resolution inversion of ammonium wet deposition data: Interpretation with a new agri- cultural emissions inventory (MASAGE_NH3), J. Geophys. Res. Atmos., 119, 4343– 4364, doi:10.1002/2013JD021130.

Hassler, B., et al. (2016), Analysis of long-term observations of NOx and CO in megac- ities and application to constraining emissions inventories, Geophys. Res. Lett., 43, 9920–9930, doi:10.1002/2016GL069894.