Response to **A. Sellar (Referee)**

General comments:

Firstly I congratulate the CEDS team on what has clearly been a monumental effort, and one which will benefit the climate modelling community greatly. On this note, it would be nice to say a little more about the CEDS project itself in the introduction: e.g. what is the "community" aspect, what is the formal project goal.

Thank you for your kind comments. The following text on "community" has been added to the summary section:

With release of this data set, and soon the entire data system, it is our intention that further improvements will be made through feedback from the global emissions inventory community. 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. We welcome comments, including notes on any potential inconstancies or relevant new data sources, so that that these data can be improved in future releases.

The manuscript gives a detailed and thorough account of the methodology. Some of the detail could perhaps be moved to the supplementary material to reduce the length of section 2. This is merely a suggestion: I leave this at the authors' discretion.

Thank you for the suggestion. We agree that the paper is a bit long, and we have moved a few secondary points to the supplement, but during our internal review process found that we received many methodology questions, as many readers (unless motivated by a specific highly detailed questions) are unlikely to page through the extensive supplement.

Please include some comparison of the spatial distribution of emissions against Lamarque et al. in section 3, at least for the species totals, focusing on 1850 and 2014 since these are important years for CMIP6.

We have added a section showing gridded data, as well as comparisons of gridded data between Lamarque et al 2010 and CEDS data. We have included additional gridded figures in the SI as well.

The discussion on uncertainty is very useful, and I am pleased to see that "quantitative uncertainty analysis" and emissions ensembles will be included with future data releases. This will assist in understanding sources of uncertainty in historical radiative forcing due to composition changes; given that this is a major outstanding question in climate modelling I would urge the CEDS project to place high priority on this development.

Thank you. We agree. After the upcoming public release of the data system, uncertainty is a high priority.

Finally, given the problems in the 2016 data release which emerged after this discussion paper was published, the manuscript should be updated to refer to the methodological changes which have been applied for the 2017 release, and to summarize the impact of these changes.

We've added a section in the appendix detailing the changes in the various releases of the gridded data.

Specific comments and suggestions:

P2 L53: "(sometimes also as RCP historical data)" this is not a particularly meaningful phrase. I would suggest using only the name "CMIP5 dataset" for the collective historical and future dataset.

This text has been changed to:
"This data is also used as the historical starting point for the Representative Concentration Pathways (RCP) scenarios (van Vuuren et al., 2011) and in some research communities is referred to as the RCP historical data. In this article it is referred to as the CMIP5 data set."

P2 L83: "Preindustrial data (CEDS-v2016-06-18), 1750 – 1850, were released in June 2016 and CMIP6 historical data". Insert date range 1850-2014 for 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).

P3 L97: This list of 6 phases would be clearer as numbered bullet points (i.e. an {enumerate} environment in LaTeX).

This change has been made. Thank you.

P4 L113: This seems like a key methodological difference from Larmarque et al: it would be good to say something about the impact this difference has on the resulting dataset.

This is a large change in methodology. First, it allows us to extend estimates forward over recent years by using recently updated energy data. Second it allows us to use more detail in historical years by modeling fuel use and EFs separately, which has had an impact on NOx and CO emissions from residential biomass burning for example. While these impacts are not explicitly noteed in this section of the paper, we feel the rest of the paper has highlighted these differences, specifically the section on comparisons to the CMIP5 (*Larmarque et al*) dataset. A major advantage of this method is also that it can more consistency capture trends over time, mentioned here in the text, and which results in some changes in recent trends discussed later in the comparison with the CMIP5 dataset.

P5 L160: Suggest "available" -> "documented" or "detailed"

Thank you. The following text now appears in the manuscript:

Mapping of IEA products to CEDS fuels is **provided** in Sect. A3.

P6 L180: "Several other changes were made, such as". Clearly this paper cannot list all such changes, but from a methodological perspective, where is the full set of changes documented? In the CEDS code, or accompanying documentation?

Most, if not all of these assumptions are detailed in the Data and Assumption Supplement. All detailed methods (e.g. code), assumptions, and data will also be available with the open source release of the system. This is detailed in the manuscript in additional text describing the release of the system, as well as existing text describing the substantial supplemental information available to download with the manuscript.

P10 L345: "with a time and sector specific options ...". Delete "a".

This change has been made. Thank you

P12 L396: "Emissions from mineral and manure emissions are often inconsistently reported;

3B_Manuremanagement and 3D_Soil-emissions together, so CEDS total estimates should be reliable". I think some text is missing here.

This has been fixed. Thanks.

P13 L408: "Gridded emissions are aggregated to 9 sectors for final distribution". Does "final distribution" refer to the temporal distribution described in the next sentence? Please make this clearer. Also, why use the intermediate sectors for spatial distribution and the 9 sectors for temporal distribution?

Emissions are distributed over space using the intermediate gridding sectors. Seasonality is added using the 9 final gridding sectors. These steps in aggregation are determined by the level of detail of the proxy data and seasonality profiles. We've rearranged and added some language to make this more clear in the manuscript. Thank you.

P14 L424: This seems to be repeating what was said on P13 L 411.

Thanks. We've moved some text around so it is less repetitive.

P18 L522: missing "due" after comma?

Thanks. This change has been made.

P23 L690: "In future versions of CEDS, quantitative uncertainty analysis will be included for all time periods, but is not complete as of the CMIP6 data version." Does this mean that there is partial uncertainty information in the CMIP6 data version, or none because you will wait for complete information before publishing any? If the former, please say something about the quantitative methodology.

There is no quantitative uncertainty analysis at this time other than what is already in the literature. (The text has been revised to refer to an existing literature summary.)

P23 L694: "emissions concentrations are observed". Would "near-source concentrations" be a more accurate description?

We would like to note that comparison to both near-source and far field observations can be helpful in comparisons such as these, and we have compared to both in the paper (Hassler et al. 2016 and Kanaya et al. 2016).