Final Author Comments to the Anonymous Referee #1 and Anonymous Referee #2

Comments to the manuscript of Margarita Choulga et al. "Global anthropogenic CO_2 emissions and uncertainties as prior for Earth system modelling and data assimilation"

Dear Anonymous Referee #1 and Anonymous Referee #2, thank you for the positive evaluation and useful comments. We have addressed all comments with the aim of reaching an improved and finalised manuscript. We believe that all comments and concerns raised have now been addressed. Please find below our detailed responses to your comments.

Dear Editor, in the supplement there is the revised final version of our manuscript.

Anonymous Reviewer #1 comments and Authors reply

Substantial improvements from prior version! Credit authors for very positive changes in a new manuscript. Many comments/questions follow but overall the topic, product and description seem to definitely qualify for publication in ESSD.

We thank the reviewer for the supportive comment highlighting effort made and recommendation to publication.

Line 92: "IPCC has been addressing uncertainty from the beginning of its creation." From the beginning or from its creation but not both.

Rephrased to "IPCC has been addressing uncertainty from the beginning.".

Line 94: "emissions are considered to be fully uncorrelated" uncorrelated by type, sector, NIR? Uncorrelation represents a key assumption, reader needs better explanation. Later the authors describe situations (e.g. monthly or using IPCC definitions of fuel type) where correlation exists and benefit. Some clarity initially about what they mean by 'uncorrelated' and why that is important will help many readers.

Rephrased to "Also, the assumptions are based on IPCC (2006), so all emissions are considered to be fully uncorrelated by activity (and so by sector and by type) (i.e. all activities from IPCC (2006) are fully uncorrelated with each other), for the calculation of the uncertainty as well as of the covariance matrices.".

Lines 99-108, definitions and treatment of so-called super power plants. Clear discussion but reader needs to know if definitions derive from these authors or from other prior work. If other work, needs citation? Supplement implies the power plant distinctions arise from these authors but this reader suspects someone else prior has done similar assessment? E.g. the concept of 'super power plants' does not originate here? Not sure that 7.92 kg m-2 s-2 (grid number 30) to 7.85 kg m-2 s-2 (estimate) for grid number 31 represents a 'step function'? Seems rather more arbitrary?

Super power plant definition was derived from current authors and is based on super emitting energy generating point sources. These point sources are so important that their location, fuel type, capacity and operational parameters and actual conditions were checked thoroughly one by one to avoid a blow-up of uncertainties. Ranking emissions by the most emitting to the least emitting grid-cells, it was noticed that emissions decay in groups. A separation from the bulk of emission is found after the first 30 grid-cells, therefore those are labelled as super power plants providing a useful group for the statistical analysis presented in the study. The emissions of the power plants next the row after that are decaying more gradually.

Line 135: "emissions are on an annual national level" word missing here? Emissions are <defined>, or emissions are <reported>?

Rephrased to "emissions are reported on an annual national level".

Line 142: something wrong or missing in this second line: "$UCsector\_j, UCsector\_j<100\% \cup UCsector\_j>230\%$ "

Here mathematical symbol "∪" ("union") was used to show that this case is used when uncertainty is less than 100 % OR more than 230 %.

Line 149: "logarithmically" or log-normally?

Corrected to "log-normally".

Line 169, Figure 1: very helpful figure but again this confusion of logarithmically versus log-normally.

Corrected to "log-normally".

Line 241, Table 2: very helpful, particularly to help readers understand how authors aggregated multiple IPCC sectors into 7 groups for IFS. Should each group have a 'total' line, to show consistency / difference of EDGAR 4.3.2_FT2015 to CHE_EDGAR_ECMWF_2015? Should Table 2 include a bottom line summary for all sectors and all groups, again for comparison purposes? E.g. to show <very minor> changes? Italics not very effective to show differences. Group 7 OTHER 1.B.1.a Coal Production 0.0 vs 7.0 - not a very big difference (e.g. not worth italicizing?)

Extra columns with the sum per "group" were added in the table, with the difference shown in *italics*, the total sum per dataset was added in the caption of Table 2. Coal production is marked to emphasise that it was missing in EDGARv4.3.2_FT2015 and added in CHE_EDGAR-ECMWF_2015.

Zenodo link works well, thank you.

Thank you.

Line 295, Figure 2: hard to compare WDS versus LDS because vertical scales differ substantially?

Figure 2 was replotted with the same vertical and horizontal scales for both countries to ease the comparison.

Line 304: "For example, by following Oda et al. (2019) to characterize spatial patterns of the" not a complete sentence?

Checked – "For example, by following Oda et al. (2019) to characterize spatial patterns of the disaggregation errors in the emission maps.".

Line 306, Figure 3: would this be more informative in relative (%) terms rather than in absolute (kg m-2 s-2) terms?

Map of uncertainty values in % might be misleading as in some places uncertainty values can be high but emissions themselves still rather low and vice versa.

Line 330: word missing here?

Rephrased to "ENERGY_A (and ENERGY_S) "group" contributes the most over power plant (and super power plant) location grid-cells (e.g. South Africa).".

Line 356: proof-readers will catch this but vice versa rather than vice a versa?

Corrected to "vice versa".

Line 390 and following: "≙ " pardon my ignorance but what does this symbol mean? Other readers may have same question?

Here mathematical symbol "≙" ("correspondence") was used to show that e.g. -25 % corresponds to $\alpha = 1.5$. The sentence was rephrased to "(e.g. -25.0 % ≙ $\alpha = 1.5$ and -124.0 % ≙ $\alpha = 2.6$; +25.0 % ≙ $\alpha = 0.8$ and +124.0 % ≙ $\alpha = 1.2$; ≙ means "corresponds to")".

Line 418, 419: "Second, try to harmonise data inclusion or omission across datasets to have more clarity in the discrepancies." Not a complete sentence?

Rephrased to "Datasets might use the same country-level information as primary input, though differences in inclusion, interpretation, and treatment of that data lead to diverse results in emissions. It is necessary to try to harmonise data inclusion or omission across datasets to have more clarity in the discrepancies.".

Line 446: "CHE_EDGAR-ECMWF_2015 the highest one." Not true? Where statistically-significant differences occur (e.g. following your upper and lower limits), CHE_EDGAR-ECMWF_2015 often lower than other three? Better to emphasize this statement (Line 456, 457) "Overall, there is quite good agreement in emission budgets and uncertainties from different sources of emission data.". Statistically, this reader accepts the latter statement but not the former?

The sentence "Out of the four different sources, usually UNFCCC and TNO_GHGco_v1.1 Tier 2 uncertainties are the lowest ones and CHE_EDGAR-ECMWF_2015 the highest one." refers to separate countries. After aggregation the CHE_EDGAR-ECMWF_2015 uncertainty values are usually the lowest ones.

The sentence "Overall, there is quite good agreement in emission budgets and uncertainties from different sources of emission data." refers to the emission and uncertainty values in general, meaning that there is no gross systematic overestimation or underestimation, and the values from these four different datasets do align well.

Line 459, Figure 6 (might apply to earlier figures as well): Use formal panel labels for graphics included within one Figure? Country labels here (e.g France, in standard text font) tend to get lost with page breaks.

Figure 6 was replotted with all needed information in the header.

Line 476: just to confirm (after reading prior paragraph): lower uncertainties = less uncertainty = improvement in reliability of the central estimate?

Yes, lower uncertainties = less uncertain = reported emission values are more certain, i.e. we have more certainty in the reported emission values.

Line 497: "(i) countries total" here you need the additional apostrophe (as you used earlier): countries' total?

Corrected to "(i) countries' total uncertainty".

Line 508: "usually quite small in Megatonne." Megatonne does not need capitalization?

Corrected to "megatonne".

Line 520: "fluxes (large-scale model BIAS mitigated by biogenic CO2 flux adjustment scheme BFAS) were considered" reader needs definition of these two new acronyms?

Corrected to "(large-scale model bias mitigated by biogenic CO2 flux adjustment scheme (BFAS))".

Line 542: "checked w.r.t. their spatial location" again, proof-readers will know but I suspect Copernicus journals to not allow these colloquial abbreviations.

Corrected to “with reference to”.

Additional comments:

Supplement details (in this order): super power plant definition and selection (S1); coal emissions (S2); uncertainty calculation details (S3); the CHE uncertainty tool (S4); geographic assumptions (S5); and fuel assumptions (S6). Main text refers to S1, then S3, then S4, then (not until later at Line 232) S2. Later S5 followed by S6. Does order matter here?

It was done on purpose to have some logical order of information in Supplementary Information sections too, not only main paper.

Use of the IFS 50-member ensemble proves efficiency and skill of approach. Will other modeling centres follow suit? If / where computational resources prove different (better or worse) would author recommend more (or fewer) groups, more (or fewer) ensemble members, etc? Recommendations seems to focus on future European developments (e.g. CoCO2) but authors should address a wider range of institutions and readers, at least with final recommendations?

Text added to the Section 5 Recommendations and conclusion: “The use of ensemble technique to estimate CO_2 uncertainties is recommended. The optimal number of ensemble members is bounded by practical considerations on computational costs. Leutbecher (2018) found a minimum of 8-member ensemble can mimic some of the skill of larger ensembles, with a 20-member ensemble being a typical value used by several modelling systems and with 50-member being a desirable target. Further grouping of anthropogenic emissions into e.g. one to reduce the dimensions of the problem is also possible with the tool CHE_UNC_APP (Choulga et al., 2021).”.

Anonymous Reviewer #2 comments and Authors reply

The paper has been throughly restructured and made much clearer. As previously stated, it will be a very useful publication for the atmospheric inverse modeling field. I suggest a few corrections:

We thank the reviewer for the supportive comment highlighting the effort made and the relevance of the study.

General remarks:

• Please have the native english-speaking co-authors review the language style.

Paper has been proof-read by several English-speaking co-authors and one external proof-reader.

• Take care of formatting in the tables, e.g. capitalizations and indenting.

Paper has been checked for formatting in tables and figures.

I also have the following specific remarks:

• line 19: ”though often limited for bottom-up anthropogenic CO2 emission” not clear for reader. Better say that it is not often known or available.

Rephrased to “though often not available for bottom-up anthropogenic CO_2 emissions”.

• line 30: ”sensitivity studies”, experiments better than studies.

Corrected to “sensitivity experiments”.

• Abstract: the main result of sensitivity experiments 1 and 2 should also be included.

Rephrased to “Several sensitivity experiments are performed to check: 1) the country dependence – by analysing the impact of assuming either a well- or less well-developed statistical infrastructure, 2) the fuel type dependence – by adding explicit information for each fuel type used per activity from the Intergovernmental Panel on Climate Change, and 3) the spatial source distribution dependence – by aggregating all emission sources and comparing the effect against an even redistribution over the country. The first experiment shows the SETTLEMENT group (of energy for buildings) uncertainty changes the most when development level is changed. The second experiment shows that fuel specific information reduces uncertainty in emissions only when a country uses several different fuels in the same amount, when a country mainly uses globally most typical fuel for an activity uncertainty values computed with and without detailed fuel information are the same. The third experiment highlights the importance of spatial mapping.”.

• lines 41, 42, repeats too much the phrase ”for example” (could be omitted in some cases).

Rephrased to “(in-situ from, for example, the Integrated Carbon Observation System, ICOS, air-borne e.g. aircraft campaigns, or space-borne e.g. the Orbiting Carbon Observatory, OCO-2, and the Greenhouse gases Observing Satellite, GOSAT)”.

• line 43: ”All measurements are assimilated by global tracer transport models to infer atmospheric CO2 changes, or by flux inversion systems to estimate the large-scale surface CO2 fluxes. ” Is not correct because of the following reasons:

1. The atmospheric transport models do not assimilate the measurements, the inversion systems assimilates the measurements (model mole fractions from transport models can be compared to observations for manual analysis)

2. Not all measurements can be assimilated by models, it depends on the model. Some models are not able to represent certain measurements accurately because they are too coarse.

3. It is not just global model but regional models as well.

Rephrased to "Atmospheric measurements of CO_2 and co-emitted species can be assimilated into flux inversion systems to provide top-down estimates of CO_2 fluxes at multiple spatiotemporal scales.".

• Line 48: The global transport models require an initial best estimate of the CO2 emission fields with uncertainties, the so-called prior information. This Is not accurate. It is not the transport model, but the inversion that requires a prior to stabilize the calculation. Using the initial best estimate is an approach. There are other approaches such as using a yearly average, mask of emission regions, a linear model, etc.

Rephrased to "The European Centre for Medium-Range Weather Forecasts (ECMWF), for example, aims to develop an operational inversion system to estimate CO_2 fluxes using observed atmospheric concentrations of CO_2 and other relevant species.".

• In page 2, there is too much use of "bottom-up", sometimes "emission inventories" would be enough

Rephrased.

• Section 2.1 is better in the introduction except for parts of the last paragraph.

Section 2.1 contains all information on IPCC and how it was used in the current study. First paragraph is more generic and in theory could be moved to the introduction, but we think for the flow of the information it is better to keep all IPCC related information together.

• Section 2.2.1: Not very clear what UC and AD stand for, you may consider want to consider σ as a standard variable for uncertainty

In the Eq. (1) explanation UC stands for combined uncertainty, AD – for activity data uncertainty.

• Section 2.2.2:

– One more reason why sectors are merged is that some sectors have very low emissions, which are not distinguishable from a global or large regional modeling perspective.

Rephrased to "Usually, there are computational restrictions for operational modelling: the number of emission input fields read by the model can't be too large or emission values are too low to be distinguishable from a global or large regional modelling perspective, so some "sectors" need to be merged.".

– It is not clear why activity and emission factor uncertainties are not log-normal themselves

Most of uncertainty bound values are expert based and sometimes are higher than 50 % and even 100 % which leads to negative emission sampling.

• Table 3 could be included in supplemental information and further deisaggreated into different tables to make it more readable.

Table 3 is assumed to be rather important part of current study which shows results of the tool presented in the paper. We think Table 3 should stay in the main text.

• Section 3.4 is more a result

Section 3.4 is specially designed to show the flow of the tool presented in the paper – shows the intermediate results and helps to understand the tool.

• A transposed presentation of tables 4a- could make the comparison between the countries easier, as well as having the curves in the same plot in figure 2.

Tables 4a-c were transposed to ease the comparison between the countries. Figure 2 was replotted with the same vertical and horizontal scales for both countries, curves remained in separate plots to show additional information per curve in the header of each plot.

• Table 5 highlight this study so we can know the relevance of this study just by looking at the table

Current study is put in *italic*.

• Table 6 could be replaced by map in which the countries are color coded according to type.

In order not to expand the paper we think information in Table 6 should remain as a table.

• Section 4.2: figure 5 because we also talk of atmospheric inversions, it might not be wise to use of "inverse" type because it can lead to confusion. Maybe with instead "inverse" type use "switched"

Rephrased to "switched type".

• Section 4.3 why such an arbitrary boosting factor and not simply uncertainty propagation? When aggregating the monthly emissions to yearly, the aggregated emissions should have the same uncertainty.

Boosting factor is based on simple uncertainty propagation – if month uncertainty is combined with error propagation method yearly uncertainties are obtained.

• Tables 9 and S8 could be clearer as 3 maps:

1. EDGAR-JRC

2. CHE EDGAR-ECMWF

3. Difference between both

Tables 9 and S8 represent countries from different locations all over the world, each country has upper and lower uncertainty bounds for 7 different “groups”. In order not to expand the paper we think information in Tables 9 and S8 should remain as tables.