Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-68-AC1, 2020

Open Access Earth System Science Data Discussions

Interactive comment on "Global anthropogenic CO_2 emissions and uncertainties as prior for Earth system modelling and data assimilation" *by* Margarita Choulga et al.

Margarita Choulga et al.

margarita.choulga@ecmwf.int

Received and published: 30 July 2020

Final Author Comments to the Anonymous Referee #1 and Anonymous Referee #2 Comments to the manuscript of Margarita Choulga et al. "Global anthropogenic CO2 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 expanded considerable effort to address all comments and to improve the manuscript in all its parts: text, figures, tables. 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 1 General comments The atmospheric inverse modeling community has long been waiting for an uncertainty estimate in emission inventories. The lack of such an estimate obligated to make arbitrary assumptions of the uncertainties used in inversions. Since the attribution of emissions to certain regions or processes is highly dependent on the a priori uncertainty assumed, this could lead to wrong results. Therefore, this study is very relevant and an important step into solving this problem and should be published. REPLY: We thank the reviewer for the supportive comment highlighting the relevance of the study and recommending publication. Indeed, also the Global Emissions Initiative underlined at its recent conference on 23rd June 2020 that the uncertainty assessment of gridded emissions input is urgently needed by atmospheric modellers.

However, I find the text and format can be confusing and difficult to read in certain sections (mainly in sections 1, 2, 3 and 5). I would recommend major reformatting of the text to make it more clear. My main advise would be to view each paragraph as an independent unit of information. The first sentence should give the main take home message of the paragraph. The following sentences should provide supporting information. REPLY: We took on board the comment by the reviewer and have revised the text throughout to ease the reading of the manuscript and help the readers to establish the main messages. The changes have been tracked in the revised submission and major examples are given below: For section 1: - We have included a description of the atmospheric exchanges of carbon between the biosphere, ocean and fossil sources within one single paragraph, that is introduced with the sentence summarising the CO2 growth rate variation and trend. - We have summarised the overview of global gridded anthropogenic CO2 emission datasets with their uncertainties within one paragraph, using also a new Table 1. For section 2: - We have reduced the description of EDGARv4.3.2 in section 2.1 and described the three consecutive modifications on the EDGARv4.3.2_FT2015 dataset to generate the CHE_EDGAR-ECMWF_2015 dataset

with three consecutive paragraphs. For section 3: - The core of the paper, section 3.2 has been completely reformatted using subsections to explain the different steps of the uncertainty calculation and using tables and even an example to help the reader retracing back the uncertainty results. For section 5: - We included the summary of CO2 uncertainty comparison in Table 7. - We reworked former Fig. 4 into new Fig. 6 and restructured the discussion and intercomparison of the results.

2 Specific comments âĂć In the introduction there is a lot of information but there should be more focus on what is the problem, why is it important and what solution is proposed. REPLY: We thank the reviewer for the constructive comment. We have revised the introduction and made it more concise and less dispersive. We tried to eliminate non-essential information from the Introduction and rewrote it in a clearer manner. We considerably shortened the description of the different datasets and the discussion around the base year 2015.

âĂć Why the EDGAR sector uncertainty is not purely additive? Please expand on the exemptions. REPLY: The EDGAR inventory is estimated based upon the sum of terms, each of which is a product (e.g., of emission factors and activity data). Based on the suggestion from IPCC (2006) the error propagation approach is not exact for such multiplicative terms, and corrections should be introduced.

âĂć On what basis where fuel type assumed, e.g. source or citation? Could you add the assumed fuel type for each sector in a table? REPLY: The EDGAR emission database contains highly disaggregated activity data and emission factors which account for human activity sector and subsector, fuel type, technology specifications and cover all anthropogenic emitting sources of CO2. Emission factors by fuel type are mostly derived following the IPCC (2006) guidelines. The development of the EDGAR data base is comprehensively detailed in Janssens-Maenhout et al. (2019) and references therein. We added typical fuel types for each sector in Table 2.

âĂć Emissions from Energy_A, Energy_B and manufacturing are assumed to decrease

in the summer. However, data from the US Energy Information Agency suggests that for example natural gas consumption has two seasonal peaks, with consumption patterns predominantly driven by weather. The largest peak occurs during the winter, when cold weather increases the demand for natural gas space heating in the residential and commercial sectors. A second, smaller peak occurs in the summer when air conditioning use increases demand for electric power, which can be provided by natural gas, coal or petroleum-fired generators (Bradley S., 2015 and Comstock O, 2020). REPLY: The Energy_A and Energy_B sectors as well as manufacturing are assumed to slightly decrease because of the summer holiday break. The natural gas consumption with two seasonal peaks are rather seen in the Settlements sector, which are indeed rather weather driven. We do agree that an update of the temporal profiles could be useful in a next step and would use for that the data of Crippa et al. (2020) .

3 Technical corrections âĂć line 41: Since the early 2002s -> 2000s REPLY: Corrected.

âĂć line 86: Presence of observations may should better say availability of observations and emission information. REPLY: Corrected.

âĂć line 150: lower case S in Savannah REPLY: Corrected.

âĂć line 159: What is an autoproducer? Is this an automobile manufacturer? REPLY: Autoproducers is the energy generated and used specifically for industrial purposes and manufacturing. We added this explanation to the main text.

âĂć line 235: You repeat "per activity" several times REPLY: Corrected.

âĂć Table 3 and table 7- why are lower bounds with larger uncertainty than upper bounds REPLY: Some uncertainty ranges for emission factors and/or activity data in IPCC (2006) and IPCC-TFI (2019) are not symmetrical and have higher uncertainty values for the lower bound than for the upper bound, due to expert knowledge or in-situ data available (these are the base for IPCC values), which lead to the same pattern in final prior uncertainty bounds. Tables 4 and 11 columns "Prior uncertainty bounds, %"

show values based purely on IPCC, so not yet fully corrected to lognormal distribution as for that you need budget values per country/sector – final uncertainties are shown is Tables 4 and 11 columns "Uncertainty bounds, %". We added this explanation to the main text.

âĂć Better description of ensuring log-normal distribution REPLY: We have rewritten an explanation for the yearly uncertainty calculation and added an example how uncertainties were calculated for two different countries TRANSPORT emission group.

âĂć Table S5: why '*', which indicates for residential sector only according to the table caption, on fuel types aviation fuel, motor gasoline, etc? REPLY: Unfortunate misprint. Corrected.

âĂć I find too many acronyms difficult to follow, make text confusing: AD, NIR, TFI, EF, LDS, WDS, GLB, L, U etc. REPLY: We have removed all acronyms that do not refer to international organisations or their reporting in the text. In some tables we still had to use few acronyms to save the space, every acronym is explained in table caption.

âĂć Figures 1 and 2 have text over background images and color of boxes make it difficult to read especially if printed in gray scale, much of it should be rather explained in the text. REPLY: We complemented the main text with an explanation and adopted a more transparent background colours for the figures.

âĂć Indenting or centering of equations to distinguish them better from normal text. REPLY: Unfortunately, ESSD template does not allow changes to the current format of the equations.

âĂć Could section 4.1 be largely substituted by a table and map? REPLY: We have substituted this section by Table 9.

âĂć Please consider adding section S3 to main text as it makes the log-normal distribution more clear. REPLY: Done.

4 References Bradley, S, 2015, Natural gas use features two seasonal

peaks per year, https://www.eia.gov/todayinenergy/detail.php?id=22892 Comstock, O., 2020, U.S. natural gas consumption has both winter and summer peaks, https://www.eia.gov/todayinenergy/detail.php?id=42815 REPLY: We have included the proposed references in the revised manuscript.

Anonymous Reviewer #2 comments and Authors reply The estimation of uncertainties in fossil fuel emissions inventories is an important goal. However, this paper is very difficult to follow. It needs major revision to clarify the details of the study undertaken, its results, and its context in the field. General comments are given below. Specific comments are also provided for the first few pages to give examples of the corrections needed, but the writing and presentation of the study throughout the other sections needs to be improved. REPLY: We thank the reviewer for the useful suggestions, and we have revised the flow of the paper aiming at improving clarity of exposition and description of methods. We believe the paper now reads more easily and that key messages are now easier to grasp.

General comments It is not very clear from the abstract what the actual data product is – emissions uncertainties by sector for each country? For individual grid cells? REPLY: We have added a short description of the dataset to the abstract: “CHE_EDGAR-ECMWF_2015 consists of 11 global NetCDF files with gridded yearly and monthly upper and lower bounds of uncertainties in % and kgÂům-2Âůs-1 per each ECMWF group and their sum, and 1 Excel file with 16 spreadsheets with the same information listed per country (metadata, emissions, uncertainties, statistical parameters).”.

Why are emissions uncertainties by sector for each country needed? Does the ECMWF data assimilation system calculate posterior fluxes for individual countries? REPLY: The only source of internationally accepted anthropogenic CO2 emission uncertainty methodology is IPCC (2006), which provides guidance in estimating and reporting uncertainties associated with the national GHG inventories. National uncertainties were applied uniformly across each country to create a gridded map that later on will be used by an ECMWF data assimilation system, which is currently in the develop-

ment process to include gridded emission sectors. The resulting inversion system will provide gridded posterior fluxes which can then be aggregated for individual countries. This research is the first step and will be followed by adding spatial uncertainty of the proxies.

The paper does not address the uncertainties in spatial allocation of emissions at all, which could be much larger. REPLY: We take the point. We are aware of this important limitation but the estimation of covariances in the spatial proxy is still a steep hurdle. It requires the assessment for the spatial representativeness of the proxy data used, which varies considerably between the regions and depends on the available information (known point sources and traffic lines for energy and transport sector versus population density as proxy for settlements and other sectors for which local information on the sources are missing at global scale). We acknowledge this in the main text of subsections 3.3 and 3.4, and in the Supplementary Information Section S.3. We will devote effort to this aspect in the next step of our research. We refer to first attempts in this direction with EDGARv5.0 by Crippa et al. (2020), which started to assess uncertainties and spatial representativeness, improving the latter e.g. for the settlement sector with weather related information.

Introduction is not sufficient. It should describe - other studies that estimate emissions uncertainty, their methods and results - "the ECMWF model" (L113) and how it will use the results of this study - methods for spatial allocation of emissions to grid cells by EDGAR REPLY: We added a description of other global CO2 studies that also have calculated uncertainties and we refer to the new Table 1 for a short overview. Results of this study are used in the ECMWF data assimilation system which is documented in Bousserez (2019) . As mentioned in the revised version of the manuscript, the calculated uncertainties documented in this manuscript were already tested in McNorton et al. (2020).

The paper is not clearly organized into sections like methods, results and discussion. There is a lot of background material in the "Comparison and discussion" section. RE-

PLY: Following the reviewer's suggestion, we have reorganized several sections of the paper. We deleted most of the background information from the Comparison section.

All of section 2 is very unclear and hard to follow. It needs to be rewritten. REPLY: We have entirely revised Section 2 to make it clearer and easier to follow. We have shortened it and focused on the description of the three consecutive modifications on the EDGARv4.3.2_FT2015 dataset to generate the CHE_EDGAR-ECMWF_2015 dataset with three consecutive paragraphs.

What does it mean that "An adequate size for the inversion system of the ECMWF model is less than 50 and a covariance matrix of 7×7 has been chosen"? REPLY: At ECMWF we propose to use a 4D-Var and ensemble-based hybrid inversion system. For this reason, only an ensemble size of up to 50 members at the global scale is currently viable. Based on these technical requirements we need a reduced state vector, which requires to aggregate multiple EDGAR sectors into 7 ECMWF emission groups in order to reduce the size of the covariance matrix. We have also reformulated this sentence in the text.

What is the motivation for separating the super emitting power plants? What is an autoproducer? REPLY: The reasoning behind is that the large power plants are operating usually at their maximum capacity, where standard power plants operate on day-to-day basis; also large power plants are large CO2 point sources, generating CO2 plumes that can be directly observed by in-situ or space borne measurements and these CO2 "base-load" emissions contribute a considerable and constant share to the national total. Therefore, their uncertainty is different. All the super power plants that were identified were also verified on their location, so that the spatial representativeness is no issue. For their uncertainty, we assume that they operate at full capacity and maximum availability. According to expert knowledge the upper bound of uncertainty for such supper power plants is smaller (+3.0 %) than for standard power plants, which operate based on day-to-day needs. The manuscript has been updated to clarify better this choice. Autoproducers are defined by IEA energy statistical office and include the

energy (electricity and heat) generated by an industry for its own use, mostly for the manufacturing. We added this explanation to the main text.

In section 3 it is confusing to discuss Tier 1 calculations because it seems like the emissions themselves have already been specified. Are the emissions calculations also Tier 1? REPLY: IPCC uses a tiered approach to calculate uncertainty and to estimate emission factors. Section 3 discusses uncertainty calculations according to Tier 1 (sometimes Tier 2 – fuel specific) approach from IPCC (2006) guidelines. At the same time EDGAR emissions were also calculated according to Tier 1 (sometimes Tier 2) approach., such that the uncertainty calculation is completely consistent with the bottom-up emission calculation. We have tried to explain these aspects in the main text.

Section 3.2 is hard to follow. Can the authors give an example, and specify which sectors are corrected? REPLY: We have reorganized this section and added in consecutive subsections all information on the corrective steps for the uncertainty calculations which we illustrate with an example for the TRANSPORT emission group in two countries with different statistical infrastructure.

Doesn't Equation 1 assume Gaussian uncertainties? What does it mean that "calculations were performed for upper and lower uncertainty limits separately"? REPLY: We performed all calculations separately for upper and lower uncertainty bounds, because IPCC (2006) guidelines provide non-symmetrical ranges, which we wanted to preserve. Tier 1 suggests using higher uncertainty values to create symmetrical ranges, but it was significantly inflating uncertainties and therefore we refined the calculation as recommended by IPCC (2006).

For equation 2, the propagation of uncertainties for sums should not be in percent but in absolute units. REPLY: According to explanation for Equation 3.1 in Volume 1, Chapter 3, IPCC (2006) values should be expressed as a percentage.

In Table 3 it appears that the lower limits for manufacturing are larger than the upper

limits. REPLY: Some uncertainty ranges for emission factors and/or activity data in IPCC (2006) and IPCC-TFI (2019) are not symmetrical and have higher uncertainty values for the lower bound than for the upper bound, due to expert knowledge or in-situ data available (on which IPCC (2006) default values are based), which lead to the same pattern in final prior uncertainty bounds. Tables 4 and 11 columns "Prior uncertainty bounds, %" show values based purely on IPCC (2006), not yet corrected to lognormal distribution. For the correction we need budget values per country/sector and final uncertainties are shown is Tables 4 and 11 columns "Uncertainty bounds, %". We added this explanation to the main text.

Section 3.4 can be deleted. REPLY: We have deleted Section 3.4 and moved the information on the covariance matrix to the Supplementary Information, to which refer a few sentences at the end of Section 3.3.

Figure 3. It is impossible to read the numbers on the graphs. Why are all the countries shown here WDS countries? REPLY: The manuscript will be provided with high resolution figures to better see all the details. We appreciate the interest and suggestion of the reviewer to show also a LDS country and added the Russian Federation as LDS country in Fig.3.

Text on page 16 should be rewritten more clearly and not in bullet point form. REPLY: We have substituted this section by Table 9 for more clarity.

In Figure 4, the authors should add an additional bar to the chart representing total emissions because there is too much text at the top of each panel. Aren't all the datasets omitting biofuels? Why does the "other" category have so much higher uncertainty in CHE, also shown in Table 9? The sentence explaining this graph is very long and confusing. REPLY: We have rearranged Figure 6 in the updated version of the manuscript. All datasets, except ours take biofuels (e.g. blended within the fossil oil) into account. The OTHER emission group has several extremely high uncertain activities. Since we have only the sum of all these activities we have to assume that

all of them are emitted in the same proportion; other datasets have more detailed information and can skip activities with very high uncertainties if their emissions were zero.

Table 8 should include references. REPLY: An extra column has been added to the table that includes the main references Andrew (2020) , Hoesley et al. (2018), Janssens-Maenhout et al. (2019), Andres et al (2016), Friedlingstein et al. (2019).

Section 4.5 should be removed. Figure 6 is extraneous to this study and the simulations are not described at all, and Figure 7 appears to be already published in McNorton et al. 2020. REPLY: We have adjusted this section accordingly – deleted Figure 7 and text referring to it, yet we think that it is important to show what impact detailed source distribution has and stress that this is very important to collect detailed information on emission source allocation. The simulations are explained in detail in McNorton et al. (2020). In the revised manuscript we just give a small summary and refer to the McNorton et al. (2020) paper.

In the conclusions it says that "The CHE_EDGAR-ECMWF_2015 represents the 2015 fossil CO2 emissions prior at 0.1âŰę×0.1âŰę resolution that has been for the first time to our knowledge completed with full uncertainty information with global coverage." This is not true because the uncertainty in spatial allocation of emissions has not been considered. And what about the other datasets that report uncertainties listed in Table 1? Furthermore, there is not even a description of how uncertainties are specified at the grid cell level in this paper – the uncertainties seem to be only given for country totals. The dataset is only described in the Conclusions, but it should be described earlier in the paper with all the details on how grid cell values are specified. REPLY: We agreed to the need for a more refined description of the CHE_EDGAR-ECMWF_2015 dataset and its strength and rephrased it as follows: "The CHE_EDGAR-ECMWF_2015 represents the 2015 global fossil CO2 emissions at 0.1âŰę×0.1âŰę resolution that has been for the first time to our knowledge bridging the inventory community and the atmospheric modelling community. In fact, the uncertainty calculations fully respect the

detailed error propagation approach recommended by IPCC (2006) guidelines for GHG inventories while the input datasets were processed such that the uncertainty information could be fully taken up by the ECMWF model IFS." Moreover, we emphasised in the main text that currently calculated national uncertainties are applied uniformly across each country to create a gridded map, and that these uncertainties do not take into account spatial allocation, which would be the next step of our research. We have added extra information on other global CO2 datasets that also provide uncertainty information – CDIAC, ODIAC, FFDAS and PKU-FUEL. We have added a description on how the calculated uncertainties were specified at the grid-cell level. The dataset description has also been relocated to a new Section 4. It is also worth noting that the Global Emissions Initiative underlined at its recent conference on 23rd June 2020 the need to address the uncertainty assessment of gridded emission inputs as a crucial piece of information for atmospheric modellers and that this still requires further research efforts.

For the actual datasets, users should be able to download these individually as needed rather than having to download everything in a large zipped folder. REPLY: This has been corrected and the big folder has been split in more clear subfolders. As such, a new Zenodo link is introduced.

Specific comments from the first few pages Title – the results of this work are the uncertainties only, right? The emissions themselves are already reported by EDGAR? How much different are they from EDGAR? How will the uncertainties be used "as prior for Earth System Modelling"? REPLY: The main result of this work are the uncertainties and reprocessing of the EDGARv4.3.2_FT2015 dataset as prior input for the ECMWF atmospheric model. The update of EDGARv4.3.2_FT2015 and difference with CHE_EDGAR-ECMWF_2015 dataset has been clearly described (incl. Table S3 in the Supplementary). Uncertainties will be used in the data assimilation part of the ECMWF IFS model.

L12 How do emissions raise awareness? Rephrase. REPLY: This has been rephrased

as “For an increased understanding of the CO2 emission sources, patterns and trends, a link between the emission inventories and observed CO2 concentrations is best established via Earth system modelling and data assimilation.”.

L15 prior should be defined. The word prior is probably unnecessary here because the results could have more uses than just as a prior. REPLY: This has been rephrased, avoiding the term “prior”.

L15 Are power and energy different? If not, the same word should be used. REPLY: This has been corrected: we use energy production consequently throughout the paper.

L17 Here and elsewhere (L25, Section 3.4) it seems misleading to say covariance and covariance matrices estimated when actually covariances are just assumed to be zero. REPLY: This assumption is suggested by IPCC (2006) guidelines and it is currently used in our research. The main text was updated to better represent work done concerning covariance matrices.

L18 Are the CO2 emissions really going to be included in IFS? I suspect they will be used with IFS in the CAMS reanalysis. REPLY: These CO2 emissions and uncertainties have been used in the ECMWF IFS CO2 ensemble simulations (McNorton et al., 2020) and in the CHE tier-2 high resolution nature run (https://www.che-project.eu/sites/default/files/2020-01/CHE-D2-6-V1-0.pdf, Agusti-Panareda et al., in preparation). In the near future these CO2 emissions and uncertainties will be used by the CAMS inversion system (currently under development).

L21 How large are these changes to EDGAR emissions? REPLY: Updated improved apportionment of the energy sector decreased emissions by 8 %, and the energy usage for manufacturing increased by 18 %. The extra emission source of the diffusive CO2 emissions from coal mines added 7 Mtons globally but localised to few regions.

L26-31. Hard to understand. Please give some values on the uncertainties, and de-

scribe the sensitivity tests a bit more. REPLY: The text is revised, and values and some extra explanation on sensitivity tests are provided.

L36 I think you mean to say “climate change” rather than “the Earth’s radiative balance and climate stability”. REPLY: This has been rephrased accordingly.

L37 ‘long carbon cycle” should be replaced by “fossil fuel” throughout. This definition is unclear and it would include wood. REPLY: We have changed it to long-cycle carbon and added clear definition with reference. This is also consistent with the definitions in the paper of Janssens-Maenhout et al. (2019) on EDGARv4.3.2.

L41 “early 2002s”? REPLY: This has been corrected.

L43-8. Sentence needs to be revised or deleted. REPLY: This has been revised.

L51 Observation not Observatory REPLY: This has been corrected.

L61 emissions not concentrations REPLY: This has been corrected.

L63 What is the Mitchell 1984 reference, and why is it cited when referring to the year 2018? REPLY: Mitchell et al. (1984) had conversion factors for different emission units, as this can be rather easily recalculated this citation was deleted.

L68 Andrew 2020 is not in the reference list REPLY: The reference is added to the list.

Table 1. FFDAS says resolution is annual, then in “Note” it says hourly. In general, the information given in “Note” for each dataset seems random. REPLY: Table 1 was updated with more precise wording.

L73 Global emissions are the same in CDIAC and ODIAC REPLY: This has been corrected accordingly in the text.

L75 3 of the datasets in Table 1 include uncertainties, according to “Note” REPLY: Description of these three datasets and short explanation how their uncertainties were calculated are added to the main text.

L77-8 Unclear REPLY: We found this sentence not relevant for the explanation of our research and deleted it.

L80 delete “with long carbon cycle” REPLY: This has been deleted.

L81-3 Uncertainties on a 0.1degree grid? What about your revised estimates of emissions? REPLY: National sectoral uncertainties and revised emissions are both uniformly mapped onto a regular latitude/longitude $0.1^\circ \times 0.1^\circ$ resolution grid.

L86, 90 Incomplete sentences, should start with “the” REPLY: Theses have been corrected.

L89 Delete – it’s not true that there was a stagnation since 2015, it has increased since then. REPLY: This sentence has been deleted.

L100 The Paris Agreement limit is not really 1.5C REPLY: We have rephrased the abstract.

Page 6. This entire page is difficult to understand. REPLY: We have revised the entire Section 2 for more clarity.

L165-6 Needs reference REPLY: The text was updated and the following reference was added: Beamish, B.B., and Vance, W.E.: Greenhouse gas contributions from coal mining in Australia and New Zealand, Journal of the Royal Society of New Zealand, 22:2, 153-156, doi:10.1080/03036758.1992.10420812, 1992.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-68, 2020.